COLLEGE of AMERICAN PATHOLOGISTS
Education

CAP19
Knowledge
Relationships
Expertise

S1953S

A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD

Agenda

- Brief comments
  - Introduce an interesting reference (Eric Topol, *Deep Medicine*)
- Introduce the speakers
  - Anand S. Dighe, MD, PhD (Massachusetts General Hospital)
  - Michael D. Feldman, MD, PhD (University of Pennsylvania)
  - Andrew R. Janowczyk, PhD (Case Western Reserve University)

Will artificial intelligence replace doctors?

Several new studies have shown that computers can outperform doctors in cancer screenings and disease diagnoses. What does that mean for newly trained radiologists and pathologists?

The outlandish expectations for AI in healthcare, a partial list (Table 1.1 in Deep Medicine by Eric Topol)

- Outperform doctors at all tasks
- Diagnose the undiagnoseable
- Treat the untreatable
- See the unseeable on scans, slides
- Classify the unclassifiable
- Eliminate workflow inefficiencies
- Eliminate hospital admissions and readmissions
- Eliminate the surfeit of unnecessary jobs
- 100% medication adherence
- Zero patient harm
- Cure cancer

Table 1.2 from Deep Medicine: changes in Medicine from 1975 to now

Table 1.2

Metric	1975	Now
# healthcare (HC) jobs	4 Million	>16 Million
HC spending per person per year	$550	>$11,000
Time allocated for office visits	**60 min. new 30 min. return**	**12 min. new 7 min return**
% GDP for HC	< 8	18
Miscellaneous	None of these	RVUs, EHRs, PBMs, “health systems”

Future of AI (from Deep Medicine)

- Deep phenotyping (Ability to deeply define each individual using all relevant data (medical, social, behavioral, family history, anatomy, physiology and environment))
- Deep learning (wide range of conditions)
- Deep empathy and connection – greatest opportunity

Where AI lives the hype

- Specific pattern based recognition in specific fields
  - Needs careful training and validation (more coming)

COLLEGE of AMERICAN PATHOLOGISTS

COLLEGE of AMERICAN PATHOLOGISTS
Education

CAP19
Knowledge
Relationships
Expertise

S1953S

A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD

Clinical Laboratory Informatics: New Tools to Improve Quality and Enhance Value

Director, MGH Core Laboratory
Director of Clinical Informatics
Clinical Lead, Partners Enterprise Pathology and eCare
Associate Professor, Harvard Medical School
Massachusetts General Hospital
Boston, MA

MASSACHUSETTS GENERAL HOSPITAL
PATHOLOGY

COLLEGE of AMERICAN PATHOLOGISTS
Education

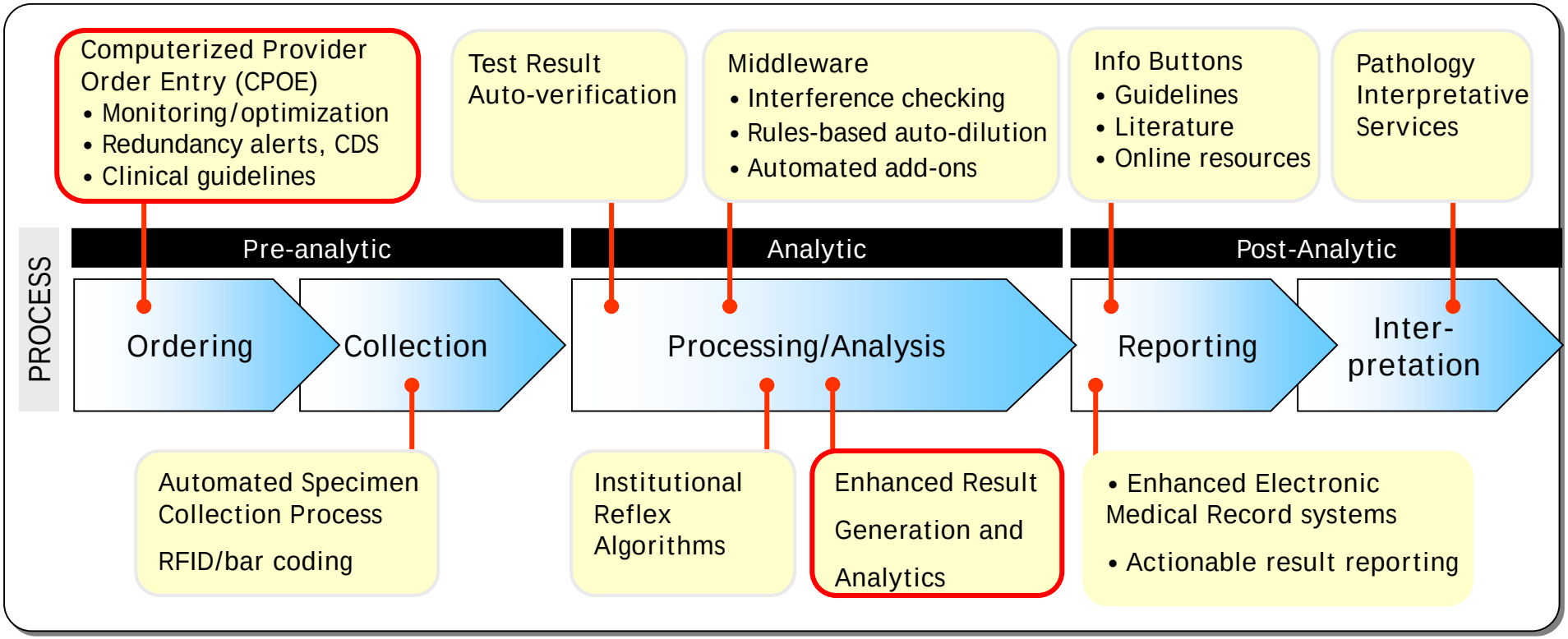
CAP19
Knowledge
Relationships
Expertise

Course S1953S

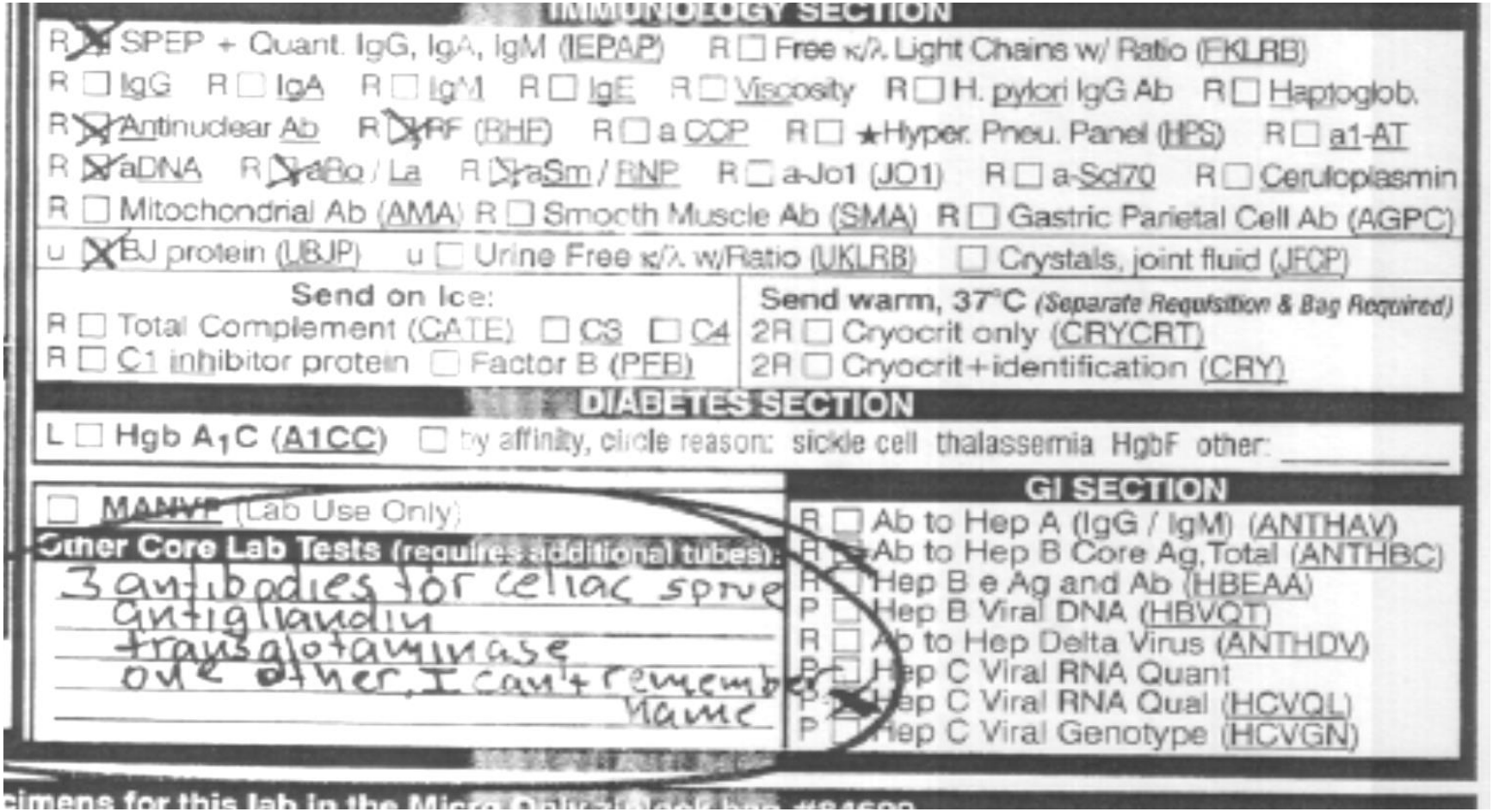
Outline

Role of informatics in the redesign of two key pathology/healthcare processes:

1. Test ordering
2. Result generation/interpretation



“The Dark Ages” (1990s)



IMMUNOLOGY SECTION

R ☒ SPEP + Quant. IgG, IgA, IgM (IEPAP) R ☐ Free κ/λ Light Chains w/ Ratio (FKLRB)

R ☐ IgG R ☐ IgA R ☐ IgM R ☐ IgE R ☐ Viscosity R ☐ H. pylori IgG Ab R ☐ Haptoglob.

R ☒ Antinuclear Ab R ☒ RF (RHF) R ☐ a CCP R ☐ ★Hyper. Pneu. Panel (HPS) R ☐ a1-AT

R ☒ aDNA R ☒ aRo / La R ☒ aSm / RNP R ☐ a-Jo1 (JO1) R ☐ a-Scl70 R ☐ Ceruloplasmin

R ☐ Mitochondrial Ab (AMA) R ☐ Smooth Muscle Ab (SMA) R ☐ Gastric Parietal Cell Ab (AGPC)

u ☒ EJ protein (UBJP) u ☐ Urine Free κ/λ w/Ratio (UKLRB) ☐ Crystals, joint fluid (JFCP)

Send on Ice:

R ☐ Total Complement (CATE) ☐ C3 ☐ C4

R ☐ C1 inhibitor protein ☐ Factor B (PFB)

Send warm, 37°C *(Separate Requisition & Bag Required)*

2R ☐ Cryocrit only (CRYCRT)

2R ☐ Cryocrit + identification (CRY)

DIABETES SECTION

L ☐ Hgb A_1C (A1CC) ☐ by affinity, clicle reason: sickle cell thalassemia HgbF other: _____

☐ MANVP (Lab Use Only)

Other Core Lab Tests (requires additional tubes):

3 antibodies for celiac sprue
antigliadin
transglotaminase
one other. I can't remember
Name

GI SECTION

R ☐ Ab to Hep A (IgG / IgM) (ANTHAV)

R ☐ Ab to Hep B Core Ag, Total (ANTHBC)

R ☐ Hep B e Ag and Ab (HBEAA)

P ☐ Hep B Viral DNA (HBVQT)

R ☐ Ab to Hep Delta Virus (ANTHDV)

P ☐ Hep C Viral RNA Quant

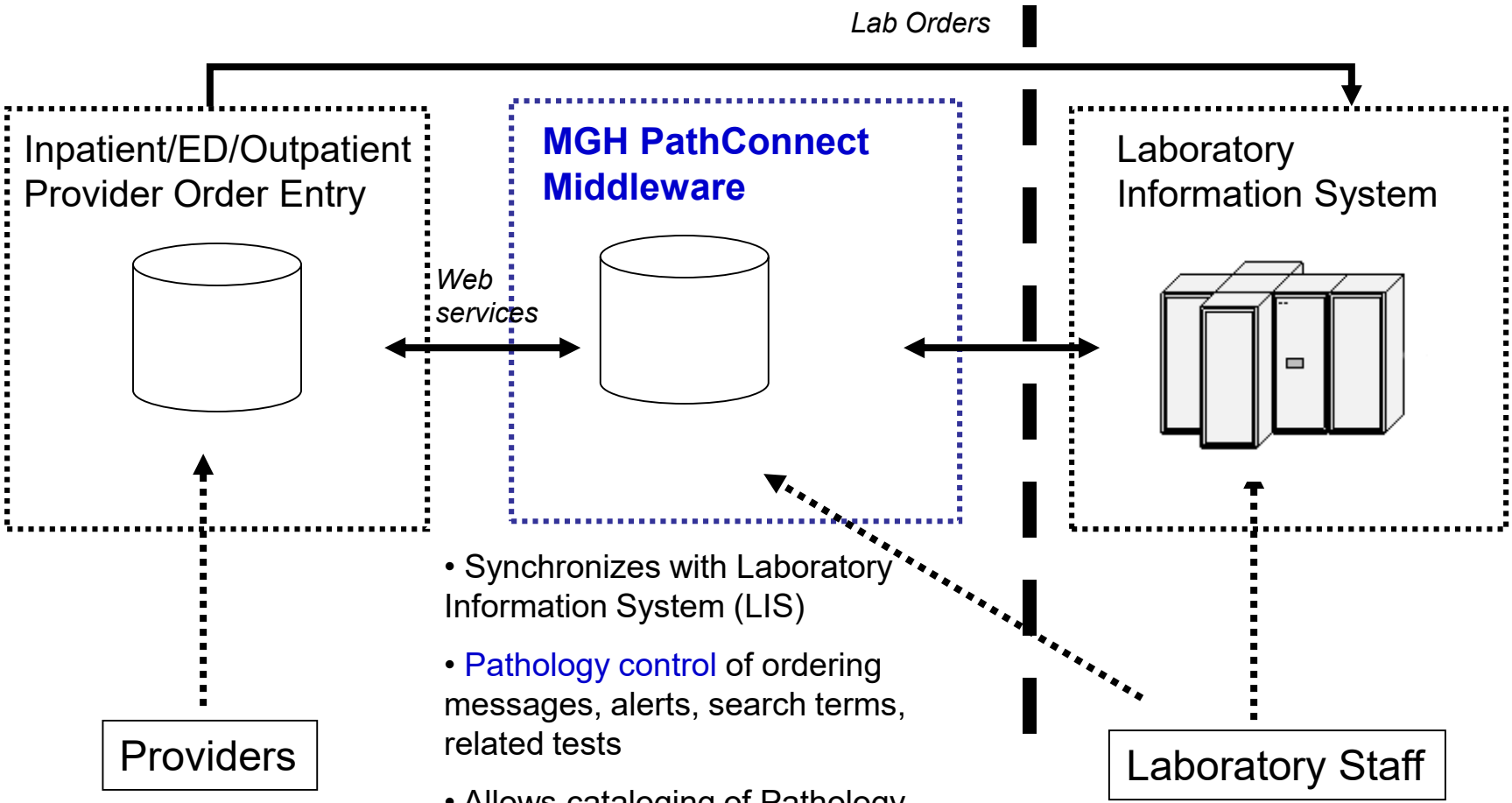
P ☒ Hep C Viral RNA Qual (HCVQL)

P ☐ Hep C Viral Genotype (HCVGN)

cimens for this lab in the Micro

The "Good Old Days" (2000-2015: homegrown custom lab software)

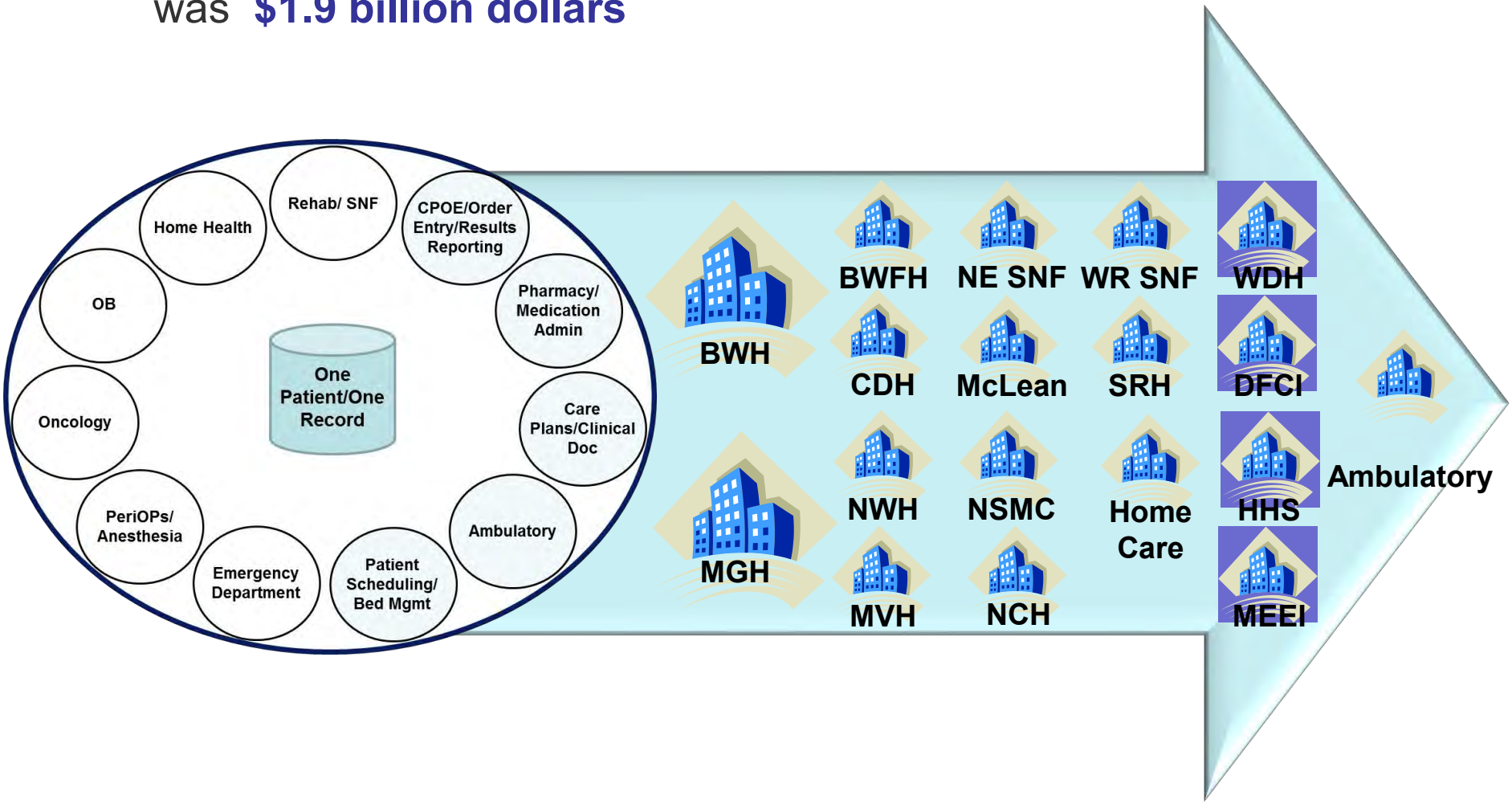
Permits Pathology to have control over Provider Order Entry screens



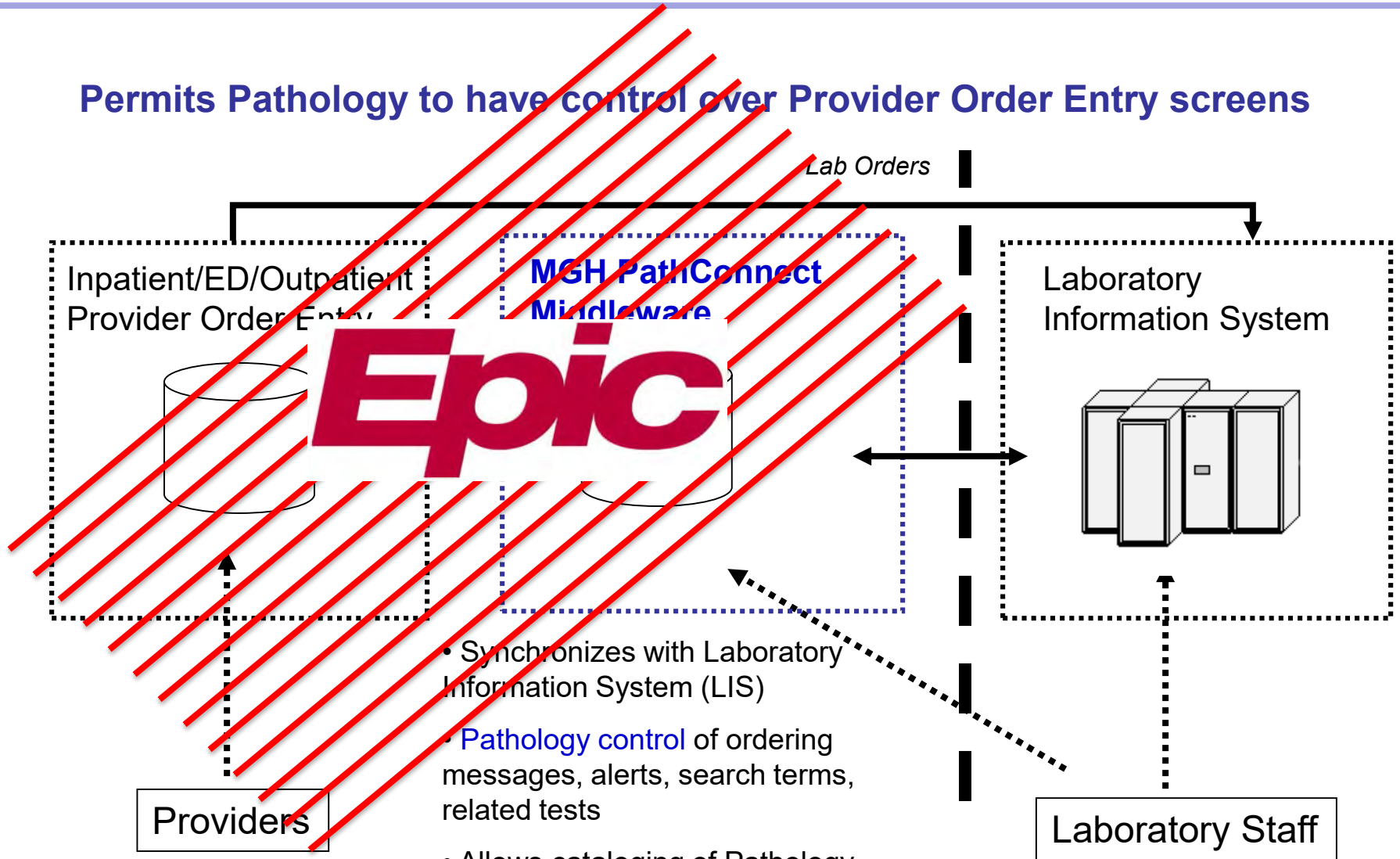
- Synchronizes with Laboratory Information System (LIS)
- Pathology control of ordering messages, alerts, search terms, related tests
- Allows cataloging of Pathology data such that it can be shared via web services

Am J Clin Pathol. 2011;135:108
Am J Clin Pathol. 2010;133:860

- After extensive evaluation from Partners selected Epic as the vendor for the enterprise-wide implementation of Partners eCare
- Total cost of the 7 year implementation (2013-2019) across Partners was **$1.9 billion dollars**



The "Good Old Days" (2000-2015: homegrown custom lab software)



“If you're yearning for the good old days, just turn off the air conditioning.”

- Griff Niblack

The “Good Old Days:”

- Limited ability for Pathology to intervene in care pathways
- Limited ability for Pathology to interact with provider orders
- Limited ability to assess outcomes
- Poor infrastructure in place for decision support
- Challenges in obtaining data for projects
- Computational techniques not widely used in Pathology

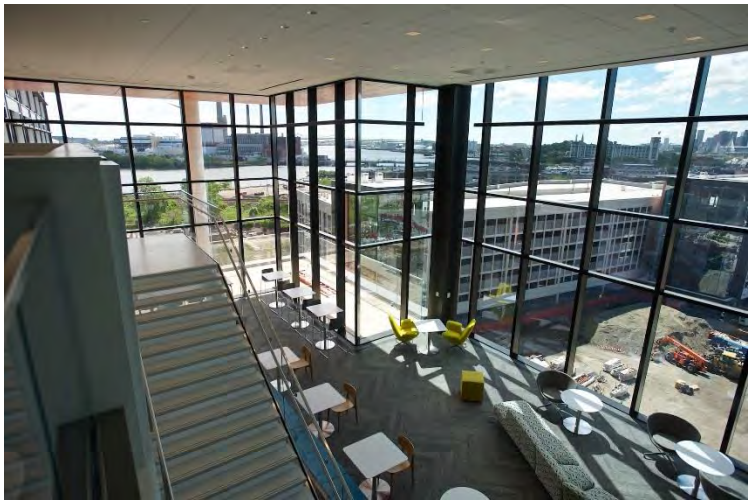
Enterprise Lab Information System Governance

Pre-Epic

- Local LIS teams at each hospital (large teams at MGH and BWH)

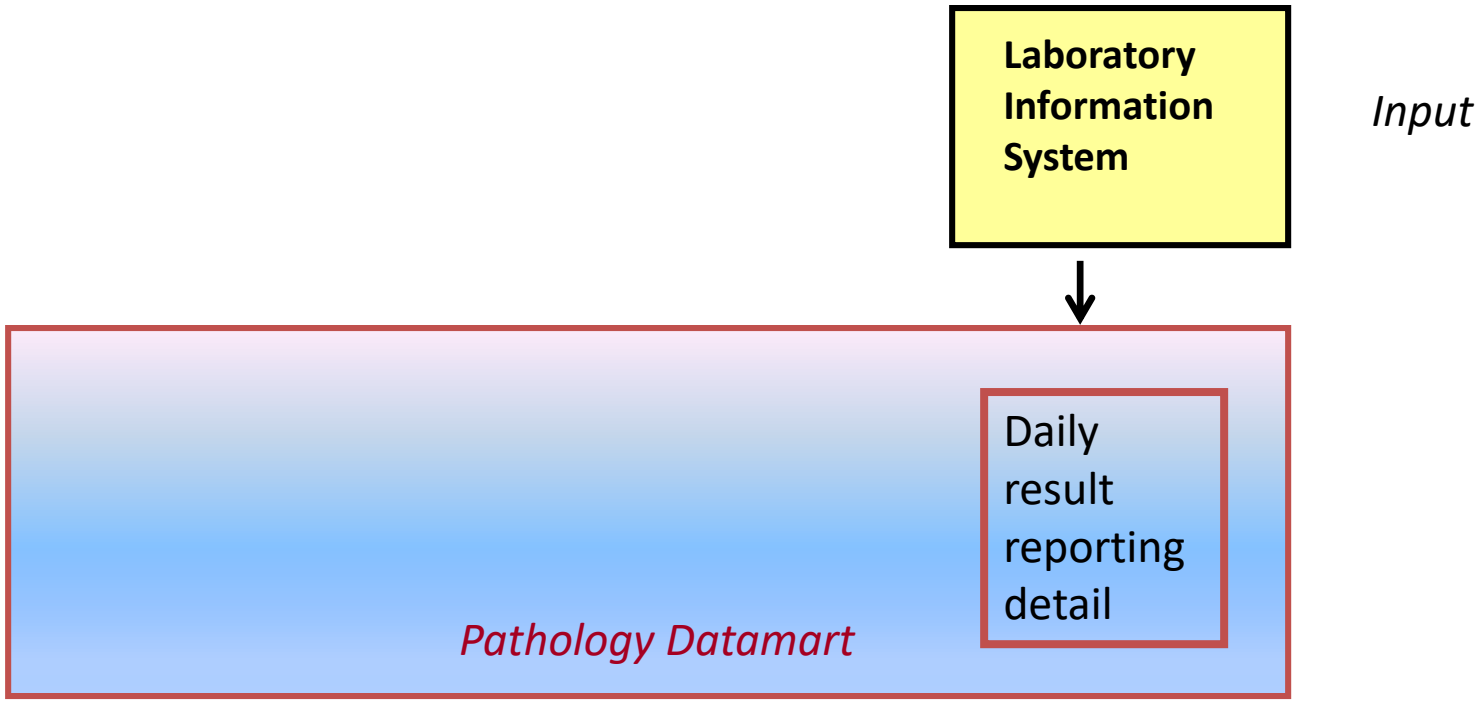
Post-Epic

- **Single enterprise LIS team** for all enterprise lab functions
- Enterprise LIS team **tightly integrated with the Epic orders team**
- **Shared tools, org chart, and infrastructure**
- EHR **enterprise lab leadership** structure



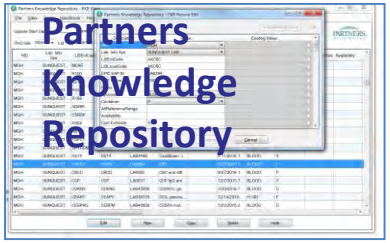
All IT and LIS teams co-localized in new facility

Pathology Datamart (Pre-EHR)



Pathology Datamart (Post-EHR)

Epic



Laboratory Information System

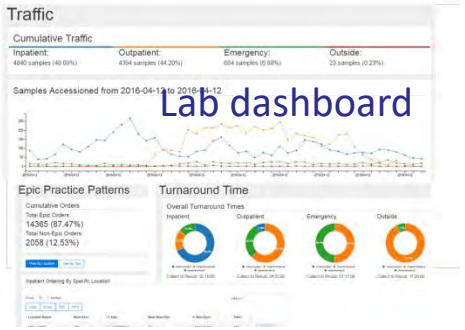
Inputs

Daily reports of orders and result messages

Expanded test profiles

Daily result reporting detail

Pathology Datamart



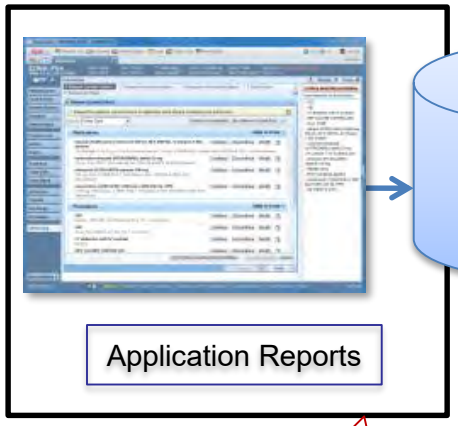
Lab dashboard

Online lab handbook

Key Epic/Lab reports

Outputs

Epic Reporting Tools



Chronicles (PRD)

Reporting Shadow

ETL Process

Clarity

Data mart

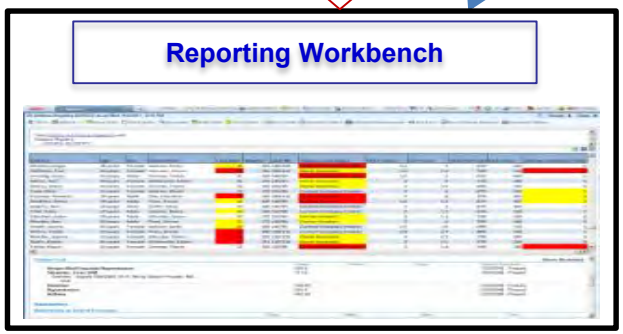
- Quick, real time data
- Little flexibility in customization
- Actionable to access patient record

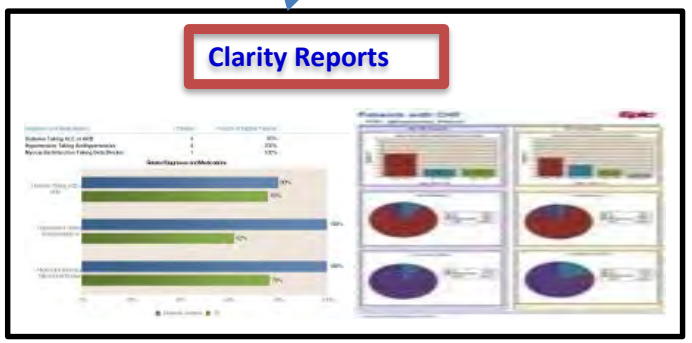
- Real time & small data set
- More flexibility in customization
- Actionable to patient record

Daily reports available of all Epic orders including problem list, diagnosis, order source (order set or preference list)

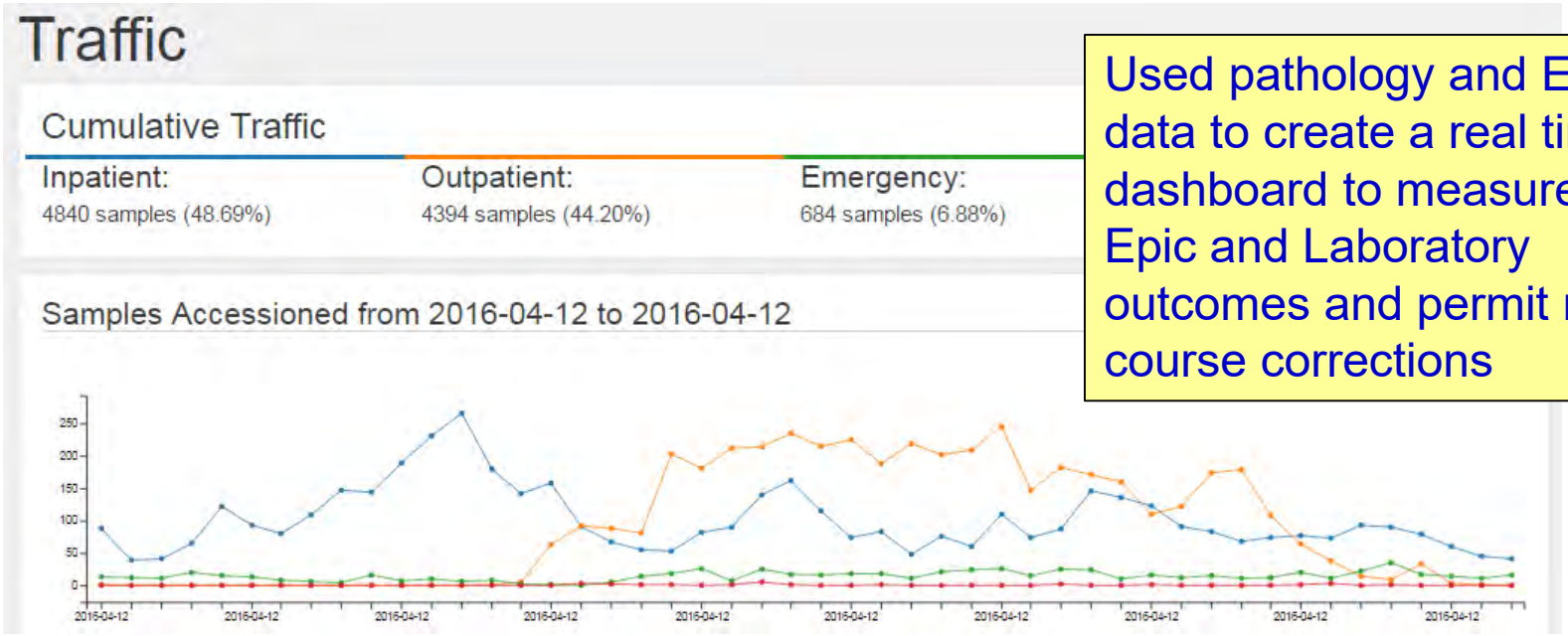
Daily monitoring needed to assess for utilization issues

- Refreshed every night
- Analytical trending data
- Highly customizable

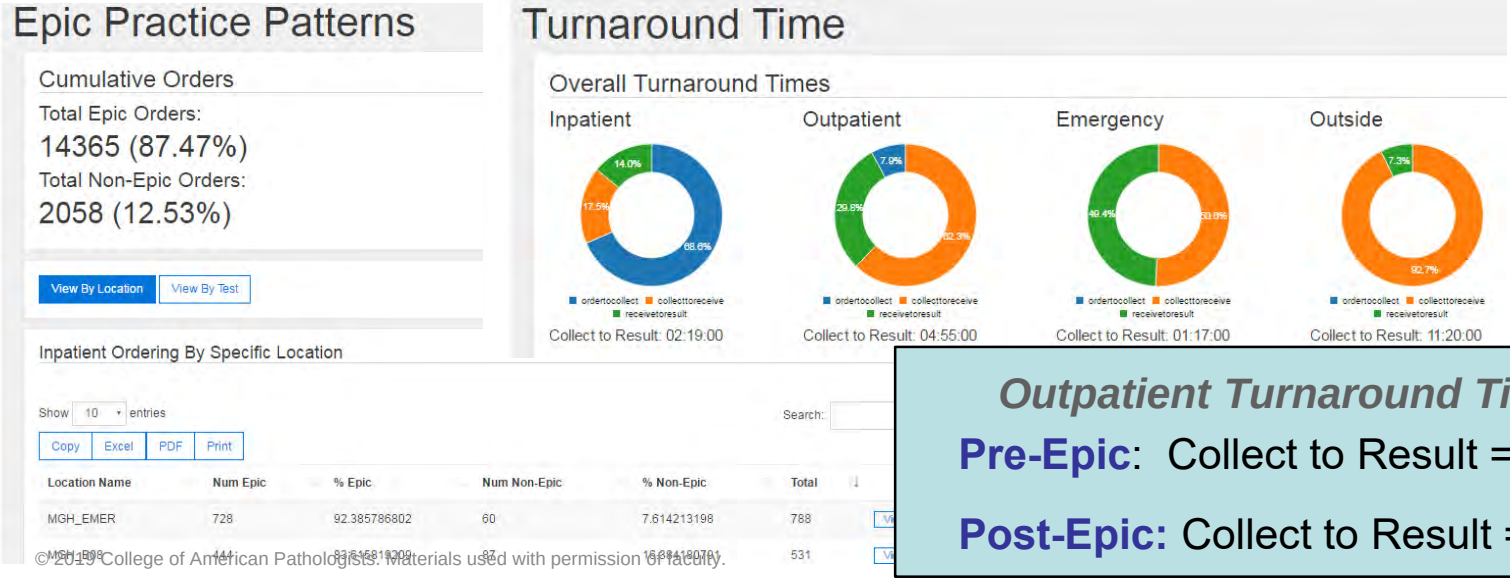




Laboratory Metrics: Online Dashboard



Used pathology and Epic data to create a real time dashboard to measure Epic and Laboratory outcomes and permit rapid course corrections



Outpatient Turnaround Time

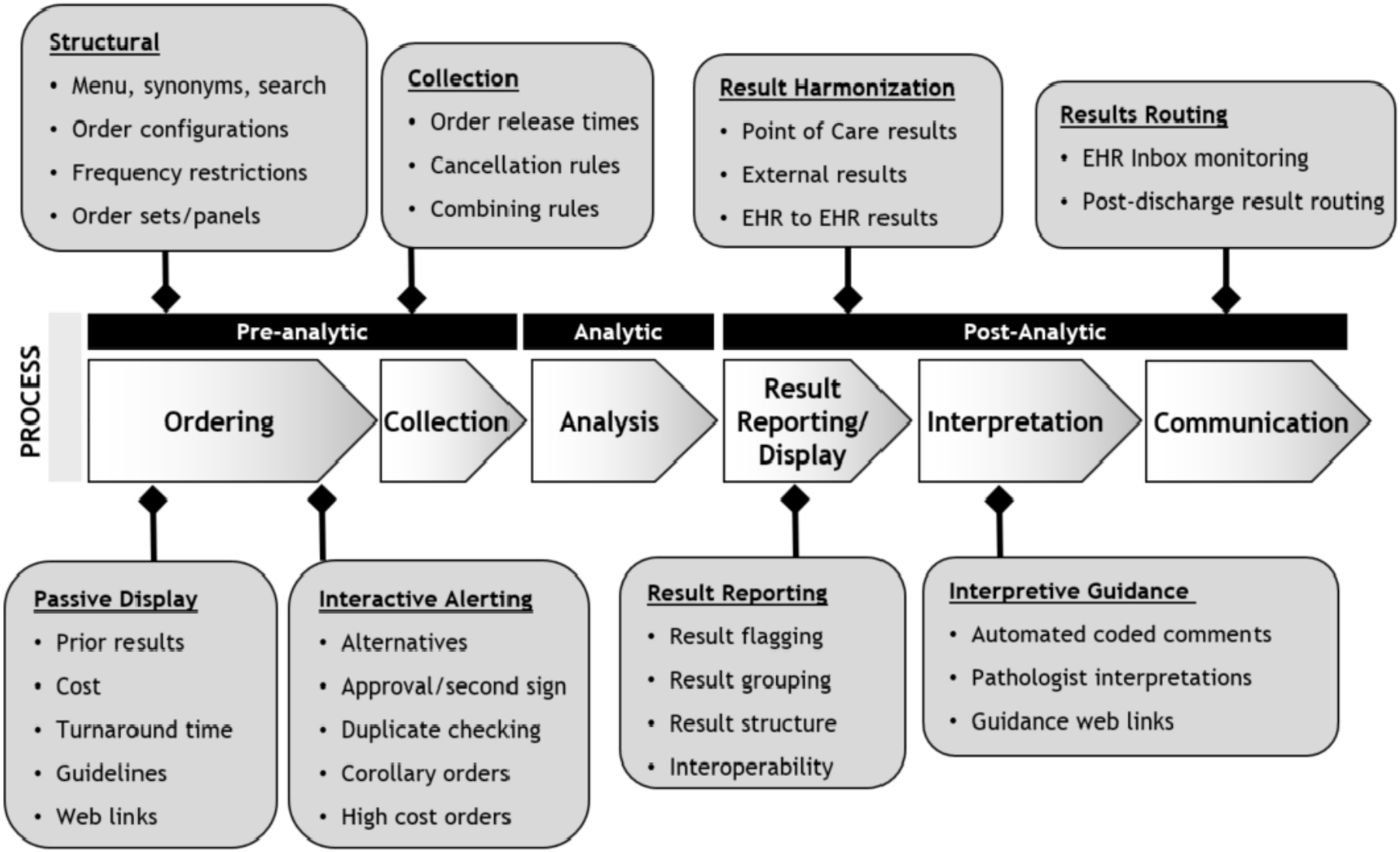
Pre-Epic: Collect to Result = **237 min**

Post-Epic: Collect to Result = **143 min**

Adapting to Enterprise Information Systems

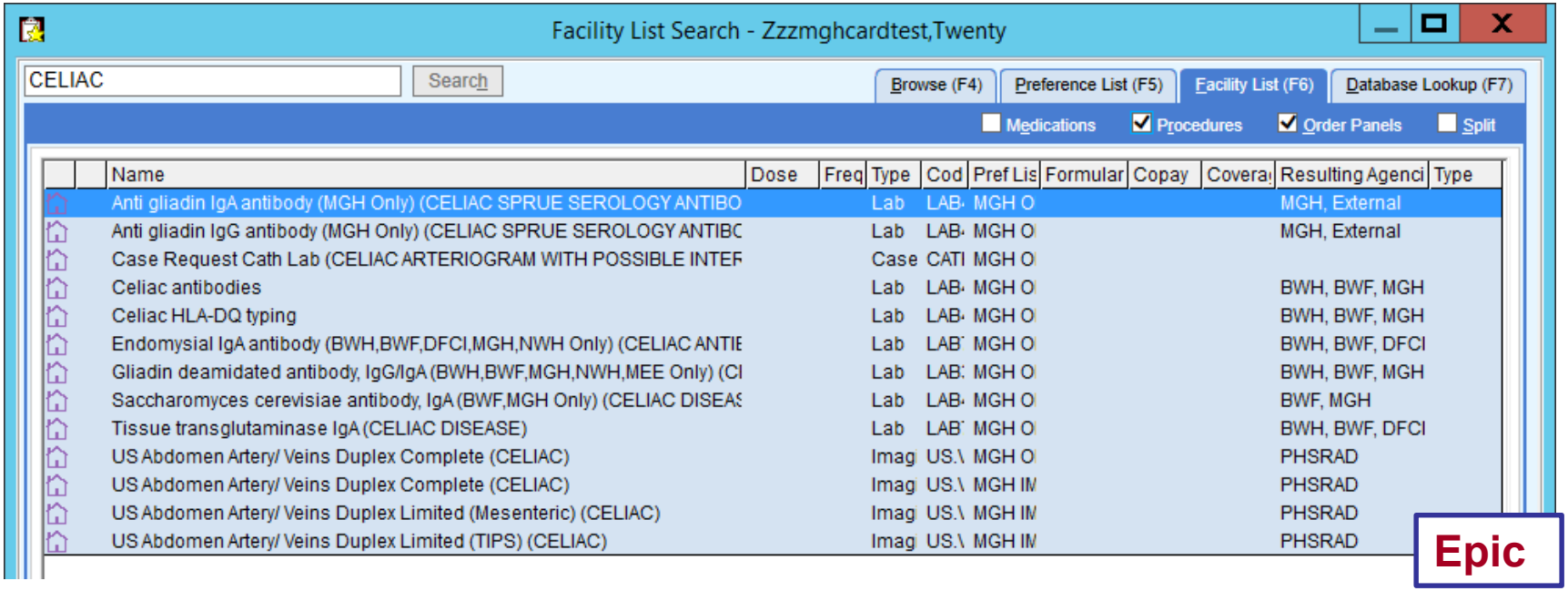
	Homegrown system	Enterprise system (Epic)
Menu size	Limited (95%)	**Most tests available (99%)**
In lab processing	Manual steps, slow	**Rapid, efficient**
Lab test search	Provides decision support, CDS visible when searching	**Search capabilities primitive, does not store search results or provide visible CDS when searching**
Ordering favorites	Not permitted	**Allowed**
Order sets	Reviewed by lab	**Uncommonly reviewed by lab**
Collection process	Simple (but manual)	**Complex** (but electronic)
Decision support availability	Custom, fast, not requiring programming	**Extensive possibilities but requires many levels of approvals, implementation complex**
Utilization control	Able to easily manipulate menu, CDS to influence testing	**More challenging**

New Responsibilities for Pathologists in the EHR Era



Clinics in Lab Medicine, 39, 2019

EHR Lab Order Entry

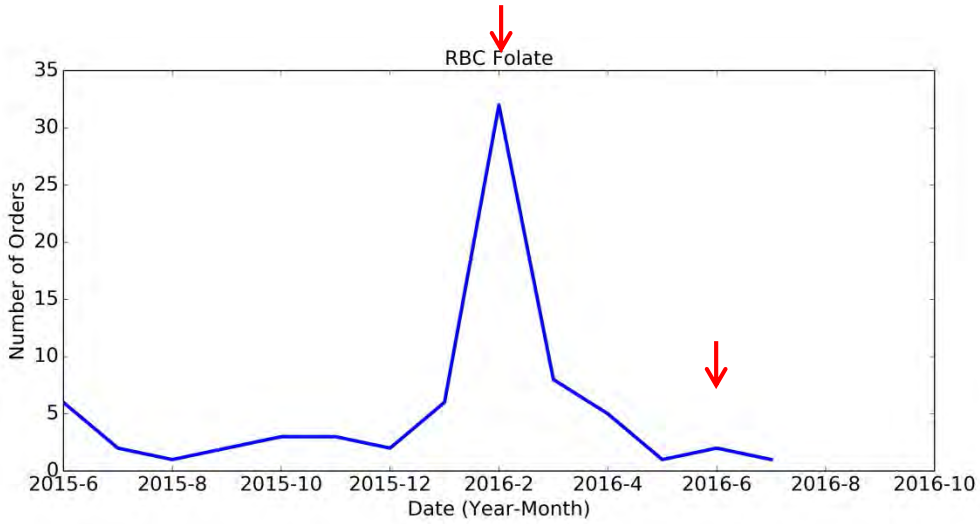


- EHR order generation is complex: lab tests can be ordered via a facility list, departmental list, personal preference list, order set, therapy plan, or decision support rule
- Team oriented medicine means many providers (resident, attending, medical assistant) may be involved in a single order

→ **Understanding EHR test ordering pathways is essential to control utilization**

Example: Rapid Responses to Volume Spikes

RBC Folate



- Large spike in RBC folate orders noted during routine monitoring
- With our Epic dashboard we were quickly able to localize the orders to a single **enterprise anemia order set**
- We first swapped out RBC folate for serum folate and later entirely removed RBC folate from the menu

I'm reaching out in regards to your ticket about switching the folate lab available in the anemia panel. I've chatted with our clinical content lead, and he wanted me to send along this screenshot for review:

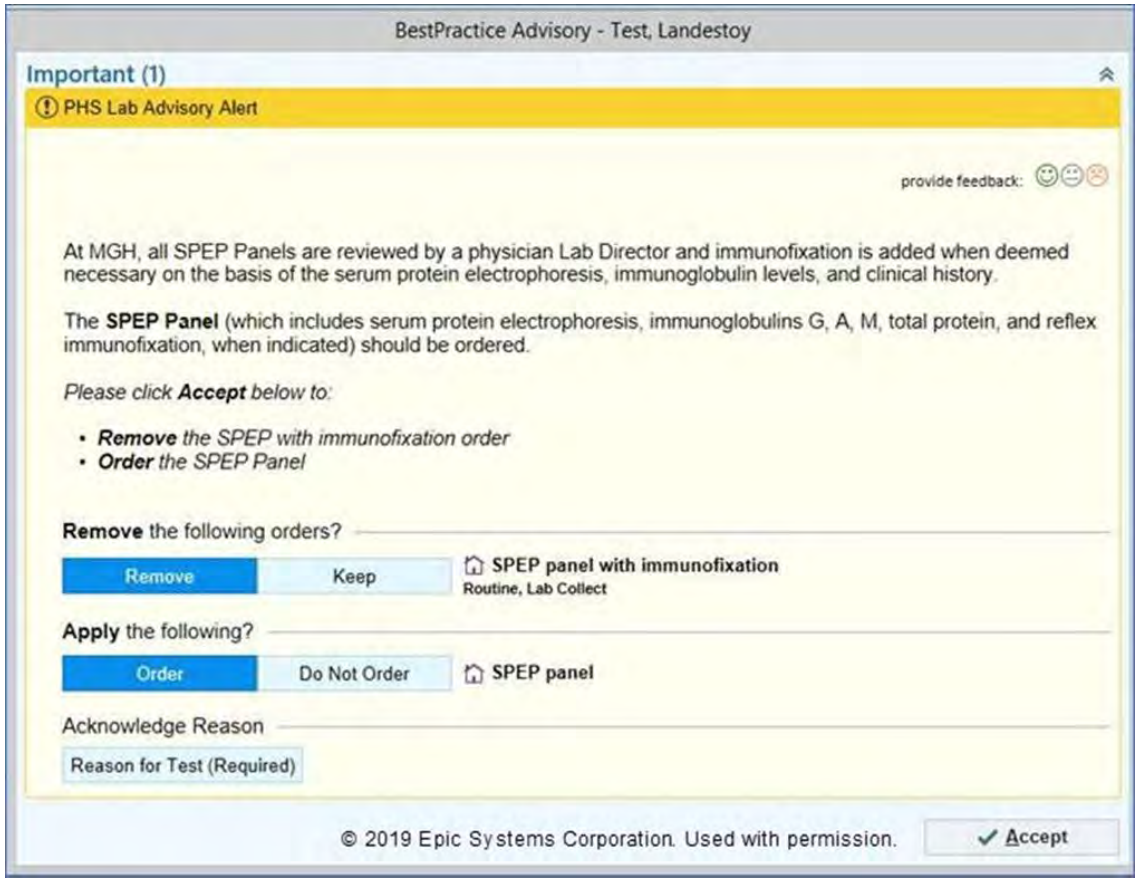
> Folate levels in the liquid portion of blood (serum) can vary based on a person's recent diet. Because red blood cells store 95% of circulating folate, a test to measure the folate level within RBCs may be used in addition to or instead of the serum test. Some health practitioners feel that the RBC folate test is a better indicator of long-term folate status and is more clinically relevant than serum folate, but there is not widespread agreement on this.

Appl Clin Inform. 2018,9:519

Agile Approaches to Clinical Decision Support

- Laboratory is an integral part of the enterprise decision support team
- CDS team meets weekly to reset priorities for two week software sprints
- Has reduced the cycle time for decision support alerts to weeks instead of quarters
- Automation of monitoring with interactive, real time CDS dashboards

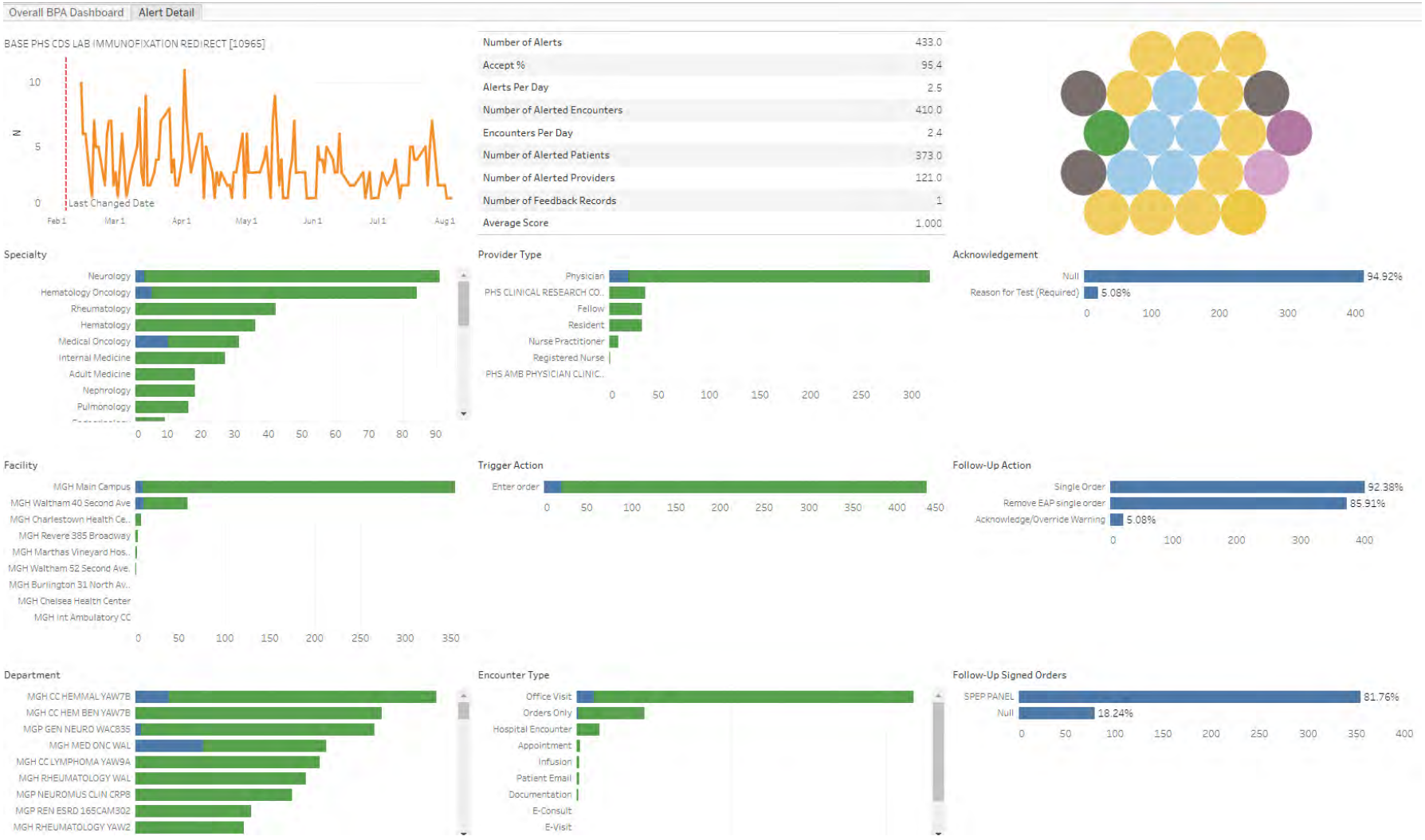
Example: CDS Alert to Discourage Immunofixation Ordering



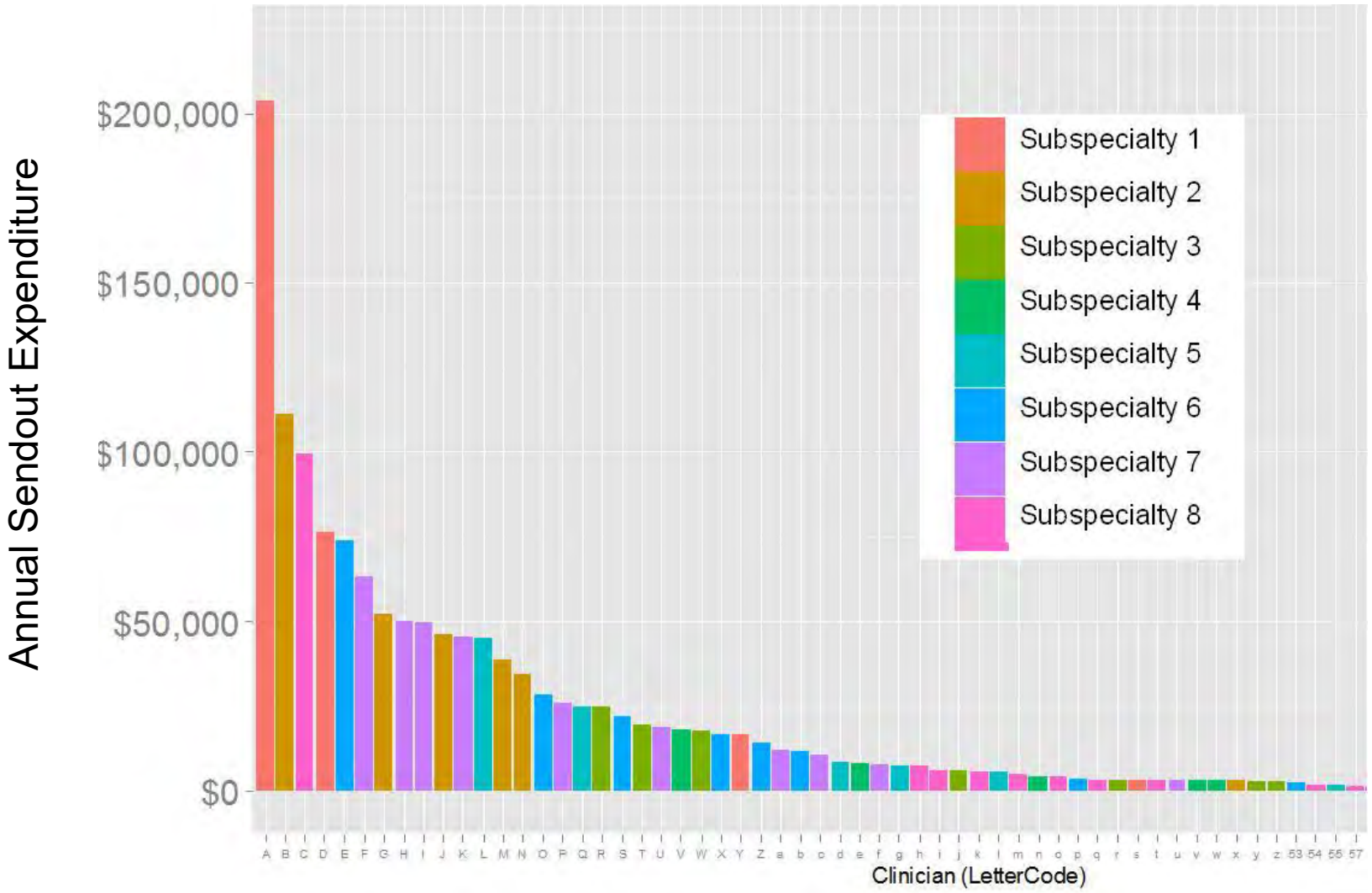
- Due to leveraging our existing EHR infrastructure the alert took < 5 hours of total effort including design, build, testing, monitoring and change management
- 85% of the time providers accept the laboratory's advice
- Annual savings of over $40,000 per year at a single hospital with this one alert

Example: Automation of Data Collection and Monitoring

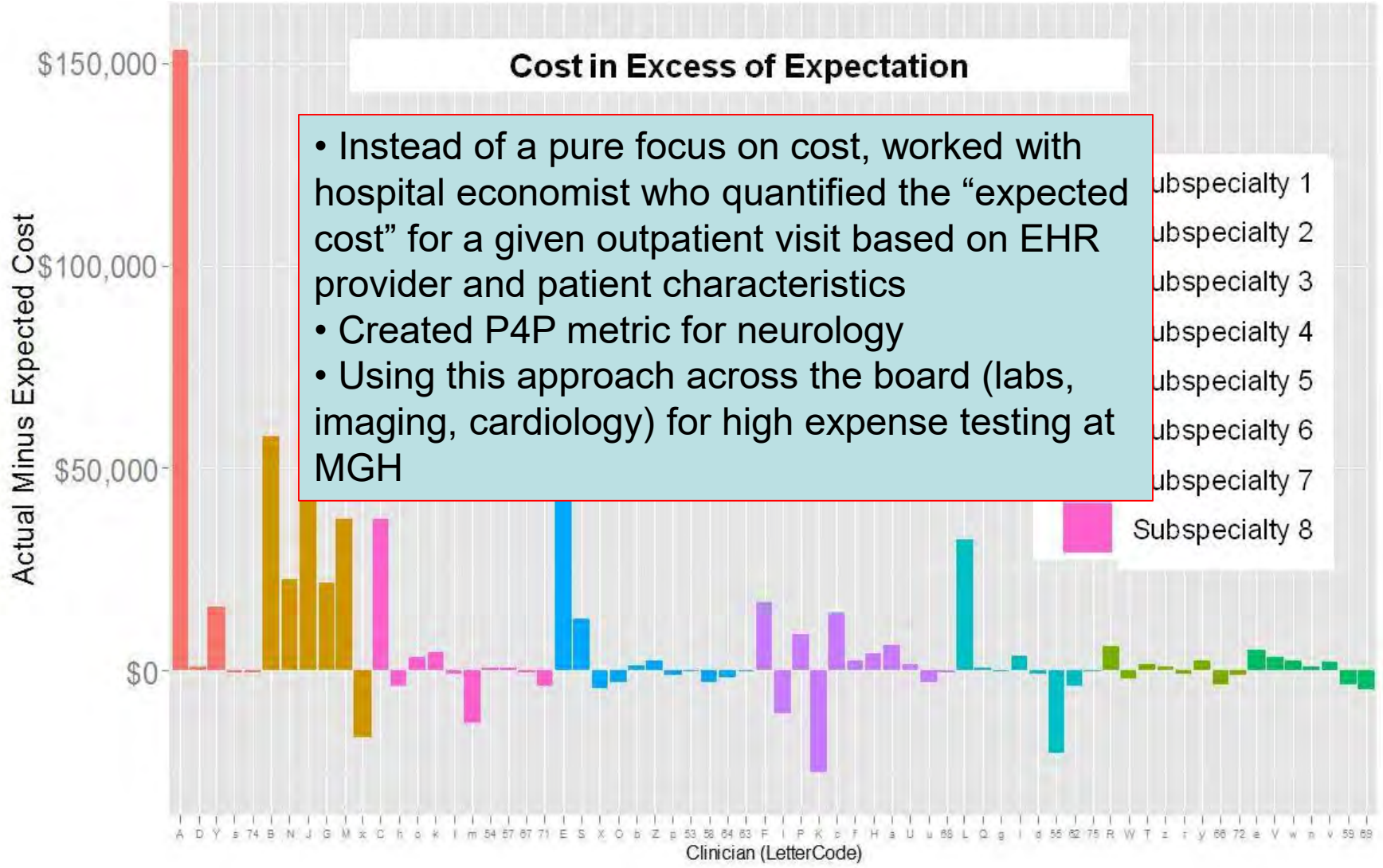
Reusable dashboards and data models that are interactive and shorten the build cycle (<u>none</u> built by the laboratory but many now heavily used by lab)



The Old Days: Neurology: Inter-Specialist Variation in Sendout Costs

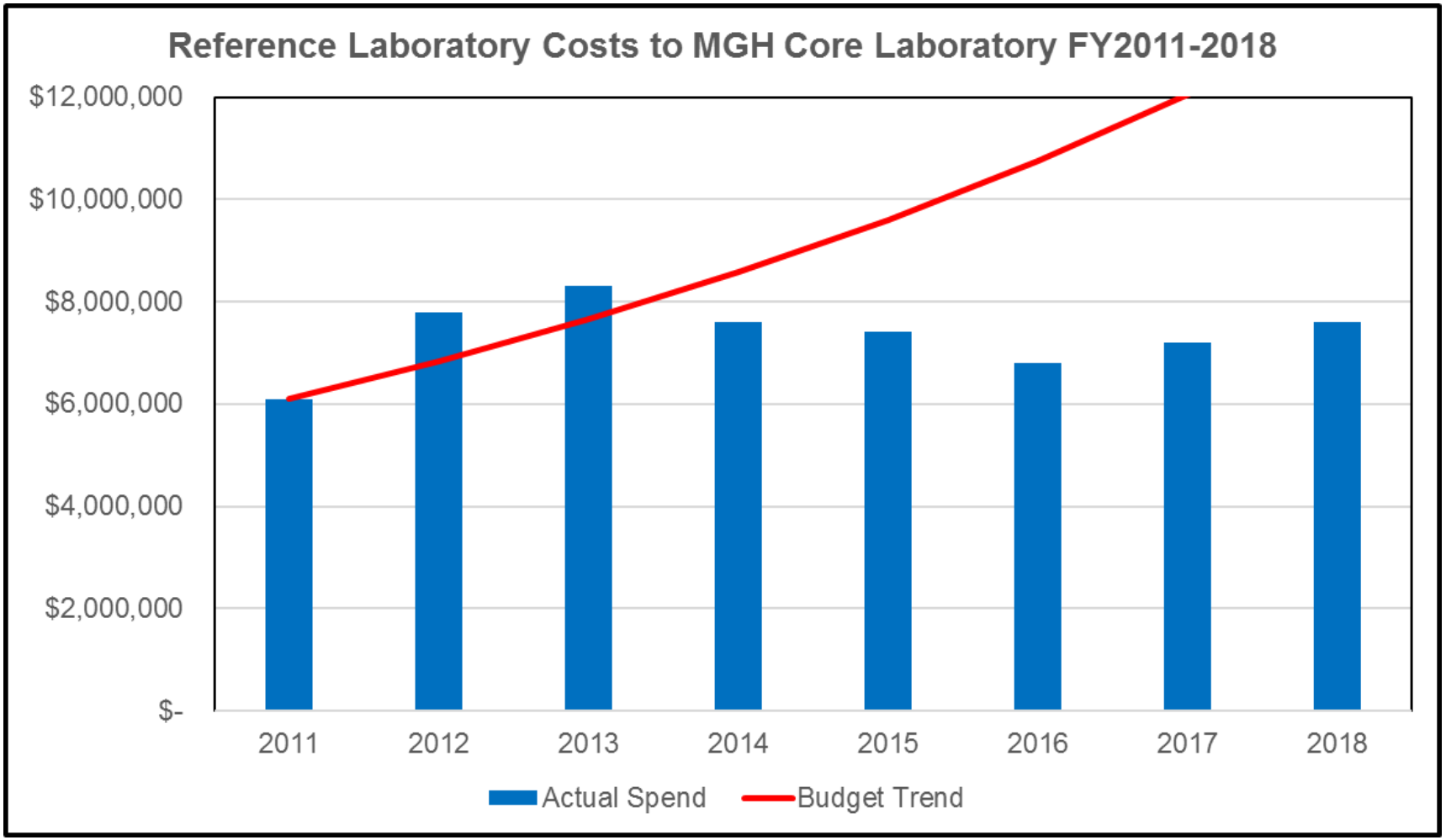


New Approach: Sendout Adjusted Expenditure, Accounting for Diagnosis

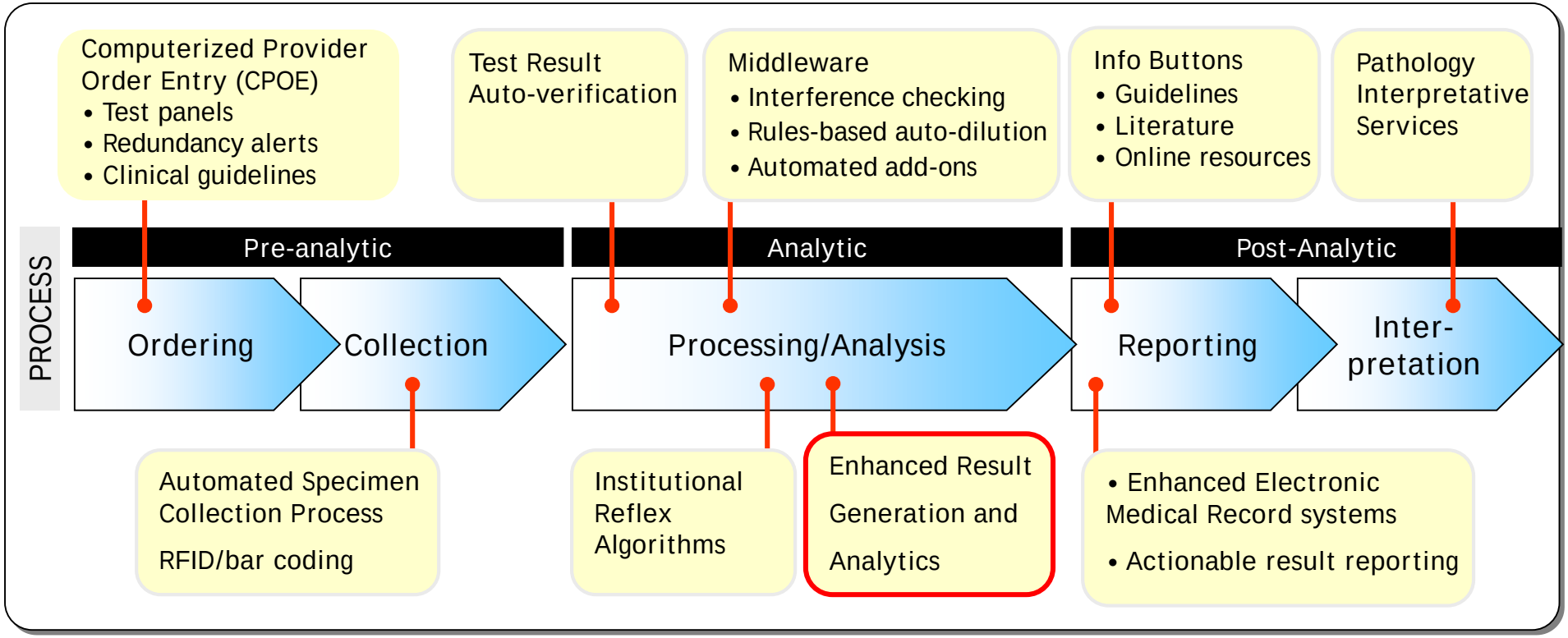


- Instead of a pure focus on cost, worked with hospital economist who quantified the "expected cost" for a given outpatient visit based on EHR provider and patient characteristics
- Created P4P metric for neurology
- Using this approach across the board (labs, imaging, cardiology) for high expense testing at MGH

MGH Reference Laboratory Testing 2011-2018

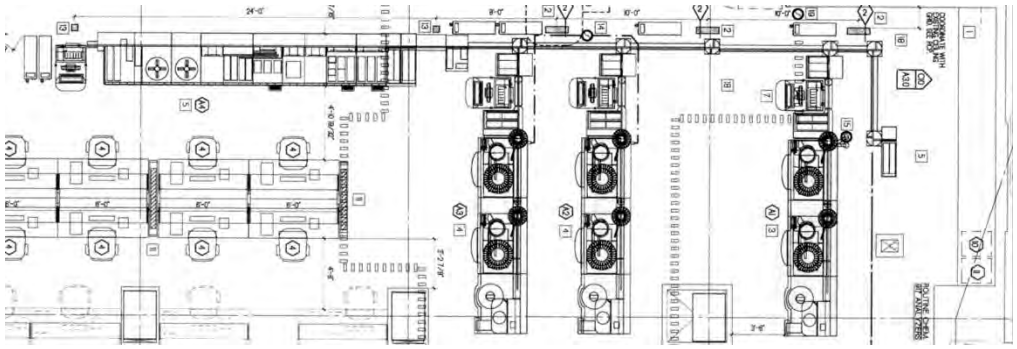


Information Processing in the Laboratory Testing Process



The Need for Informatics

- Millions of results per year
- Reported 24/7 with high levels of auto-verification
- Rate of *data* production exceeds capacity of clinicians, pathologists and technologists to generate *information*
- The human brain is not well equipped to process high dimensional data
- **Computational techniques needed**



Desired Black Box

	2/2/2018 6:22
Glucose	652
Potassium	4.2
Chloride	98
CO2	25
Sodium	133
Anion gap	10
Mean glucose	252
Creatinine	1.21
BUN	12

Release to medical record

	2/2/2018 6:22
Glucose	652
Potassium	4.2
Chloride	98
CO2	25
Sodium	133
Anion gap	10
Mean glucose	252
Creatinine	1.21
BUN	12

OR

	2/2/2018 6:22
Glucose	652
Potassium	4.2
Chloride	98
CO2	25
Sodium	133
Anion gap	10
Mean glucose	252
Creatinine	1.21
BUN	12

Refuse or add comment

Creating Decision Trees with Supervised Machine Learning

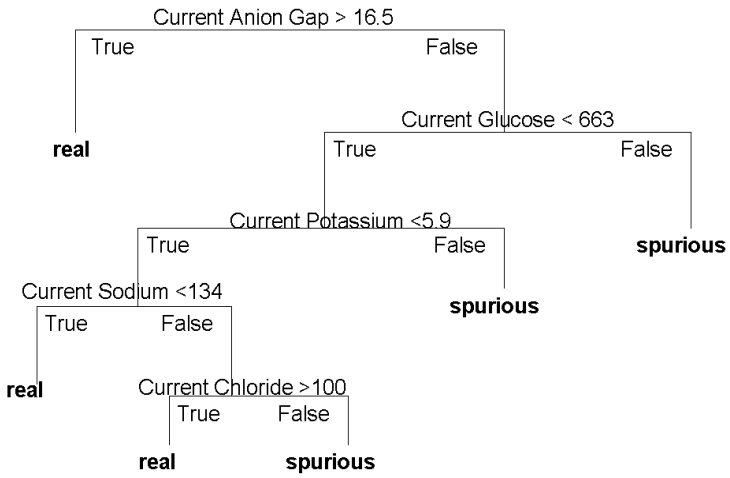
Compile and Preprocess Data

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, ..., x_k, Y)$$

	B	C	D	E	F	G	H	N
1	cur_na	cur_k	cur_cl	cur_CO3	cur_glucos	cur_anion	avg_glucose	spurious
23	132	2.8	95	18.2	654	18.8	105.8070175	y
24	128	6.6	95	22.8	585	10.2	116.1451613	y
25	130	3.1	82	48.5	751	-0.5	123.7222222	y
26	130	4.1	94	21	619	15		n

$$I_G(f) = \sum_{i=1}^{m} f_i(1 - f_i) = \sum_{i=1}^{m} (f_i - f_i^2) = \sum_{i=1}^{m} f_i - \sum_{i=1}^{m} f_i^2 = 1 - \sum_{i=1}^{m} f_i^2$$

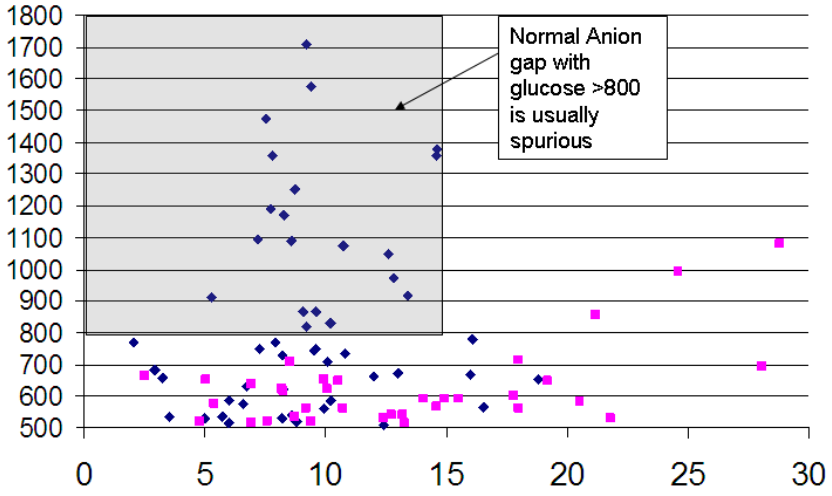
Recursive partitioning to “purify” data and generate intuitive decision trees



Combining Decision Tree Output With Pathology Knowledge

Pre-intervention, technologist judgment identified spurious results only 9% of time (**9% sensitivity**)

Parameters Supplied in Building Tree	Data Set	Spurious Correctly Classified	Total Spurious	Real Correctly Classified	Total Real	Sensitivity (95% CI) %	Specificity (95% CI) %
Current Na, K, Cl, Bicarb, Anion Gap, Glucose	Training	57	64	84	92	89 (79-95)	91 (84-96)



Implemented algorithm performed prospectively on real patients with **74% sensitivity and with 100% specificity**

Am J Clin Pathol 2015 138:406

Machine Learning, Step by Step

Step 1: Frame Problem and Setup Data

PT	Predictor A (e.g., BP)	Predictor B (e.g., Creat)	..	Clinical Outcome (e.g. progression to CKD)
1	137	1.36	..	No
2	168	1.19	..	Yes
…	…	…	..	…

Step 2: Clean Data

Step 3: Partition into training and testing

Mask outcomes from test partition.

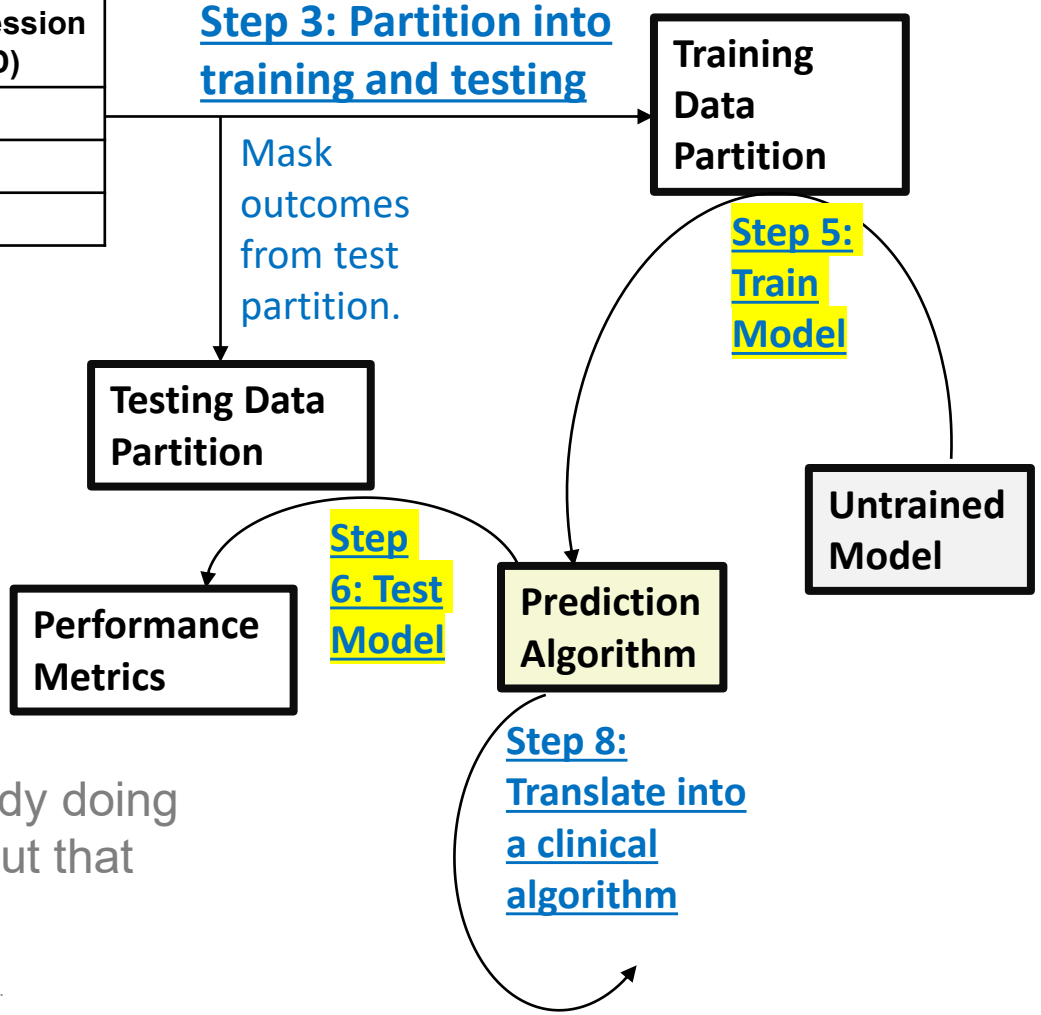
Step 4: Select type of model

Step 5: Train Model

Step 6: Test Model

Step 7: Iterate

Step 8: Translate into a clinical algorithm



Steps 1-3, 8 are what you are already doing
Steps 4-7 may require a resource but that resource can be YOU!

Model Selection

Selected Options

Linear methods

- Ordinary least squares regression
- Logistic regression
- Perceptrons

Decision trees

- Recursive portioning trees
- Ensemble methods (random forest)

Artificial neural networks

Support vector machines

Neural Networks

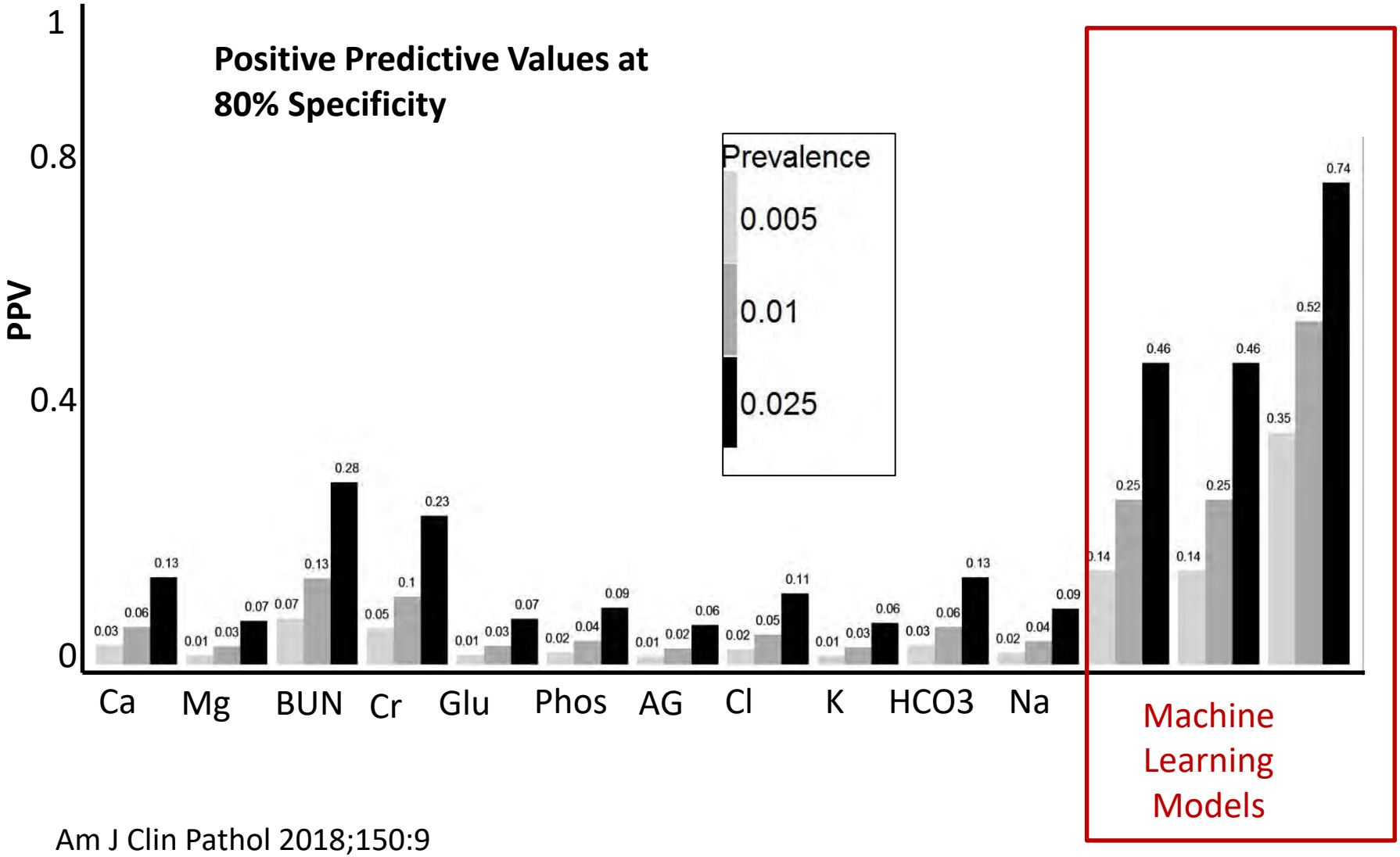
Considerations

- Data types
- Classification vs. regressions
- Complexity vs. data size
- Intuitiveness of output

- In practice often try multiple models
- Sometimes use an aggregate across various model types as final output

In prior example, we used recursive partition to produce interpretable decision trees

Machine Learning Based Multi-analyte Delta Checks Outperform Individual Analytes for Wrong Blood in Tube (WBIT) Detection



Am J Clin Pathol 2018;150:9

Machine Learning

- Clinical protocols developed using machine learning techniques have improved the laboratory's identification and annotation of spurious results, anomaly detection, and WBIT errors.
- We have used similar techniques to develop EHR implementable rules for ordering alerts
  - e.g. Used machine learning to develop algorithm to suggest discontinuing peripheral blood flow cytometry orders when not indicated
- In addition to test result interpretation, machine learning can be used to **predict** test results

Why Predict Lab Results?

1. **Eliminate redundant testing:** Tests that can be accurately predicted to be normal or abnormal may not be needed (improve utilization)
2. **Detect anomalies:** When the actual results are discrepant from predicted results, investigate and may report test results with a comment (improve interpretation)
3. **Avoid overlooked diagnoses:** Alert clinicians when tests are not ordered but are predicted to be abnormal (avoid missed diagnosis)

Ferritin Prediction

We used Ferritin in a proof-of-concept for test prediction

- **Ferritin**
  - A marker of iron stores
  - Used in the diagnosis of iron deficiency
  - Must be interpreted in the setting of other clinical and laboratory data
    - Decreased in iron deficiency
    - Increased in inflammation

"(Lab test) prediction is difficult, especially about the future"

Ferritin Methods Overview

Raw Data (3 months outpatient ferritin values)

- Transform ferritin values
- Divide at random into training and test partitions
- Mask ferritin results for test partition

Raw Data with ferritin masked for test patients

- Impute missing data
- Use 4 different imputation methods

"Completed" dataset

- Predict ferritin results
  - 4 regression methods
  - 1 classification method
  - Pair each with each imputation method

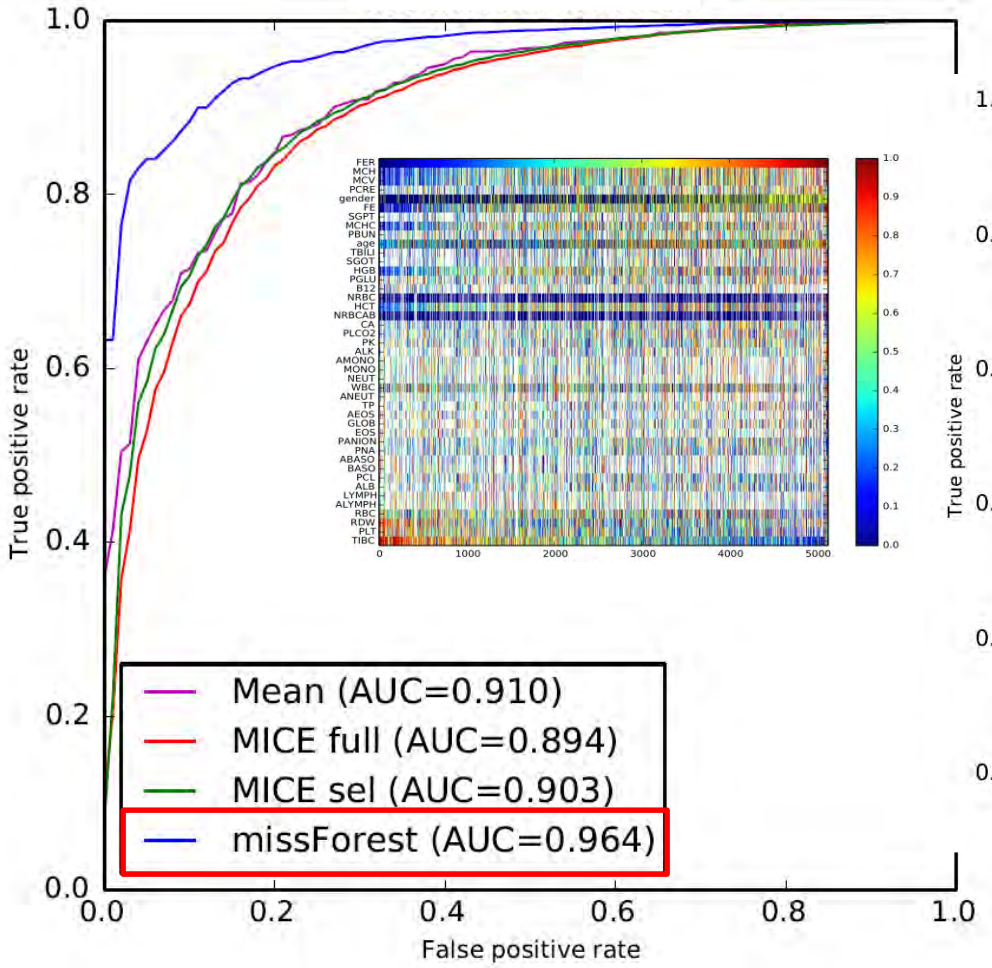
Predicted ferritin results and classifications

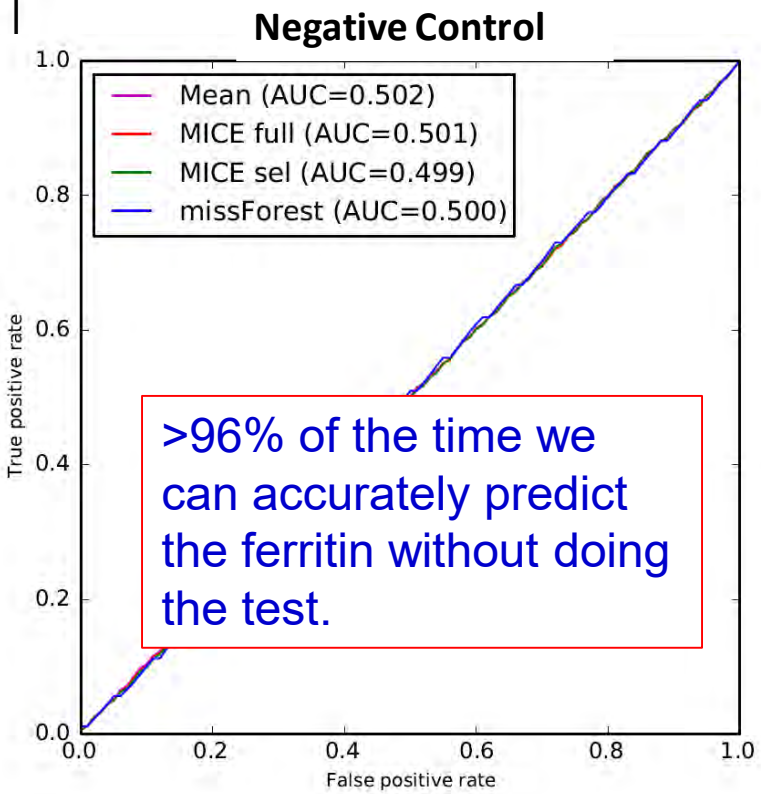
- Compare predicted ferritin results to measured results
- Compare predicted to "masked" results for test partition
- Review selected cases to determine clinical significance

Performance Metrics

Ferritin Classification Performance

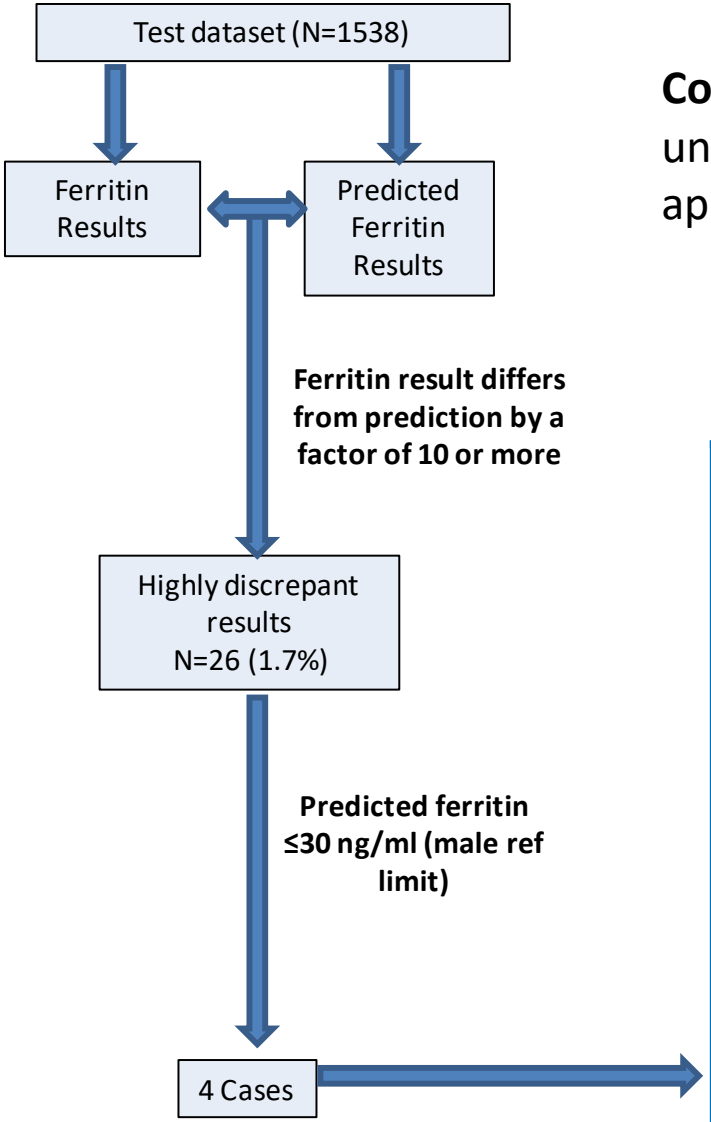
Accuracy of Predicted Ferritin Results





Am J Clin Pathol. 2016 :778.

Ferritin: Case Review



Conclusion: Predicted ferritin may more accurately reflect underlying iron status in some patients→ potential application to clinical decision support

Case	Ferritin	Predicted Ferritin	Impression	Comment
1	230	21	Iron deficiency, not clinically identified	Ferritin increased secondary to inflammation
2	197	19	Recovering iron deficiency	Receiving IV iron therapy
3	1768	9	Limited predictive data	•Only two predictor tests available •Decision support will likely require a minimum number of predictor tests
4	197	19	Complex hematologic picture	Referral to hematology would have likely been useful had the testing been ordered by a non-specialist

Am J Clin Pathol. 2016 :778.

Ferritin Prediction: Potential Applications

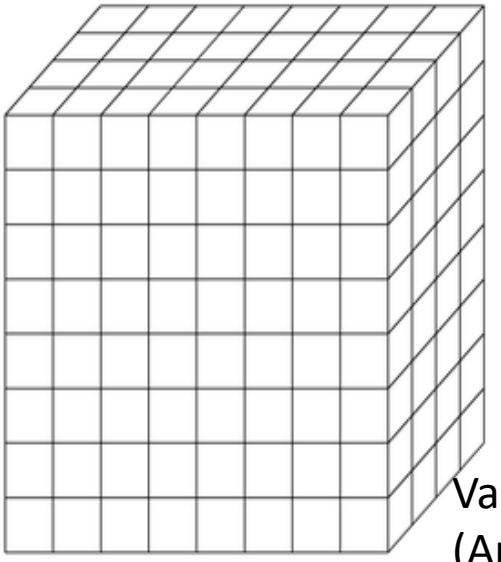
1. **Detecting anomalies:** Report test results with a comment when discrepant from predictions
2. **Avoiding overlooked diagnoses:** Alert clinicians when tests are not ordered but predicted to be abnormal
3. **Eliminating redundant testing:** Tests that can be accurately predicted may not be needed

This ferritin result is inconsistent with other testing. Do not exclude a diagnosis of iron deficiency on the basis of the ferritin alone.

Test results indicate the possibility of iron deficiency. Consider ordering ferritin if clinically indicated.

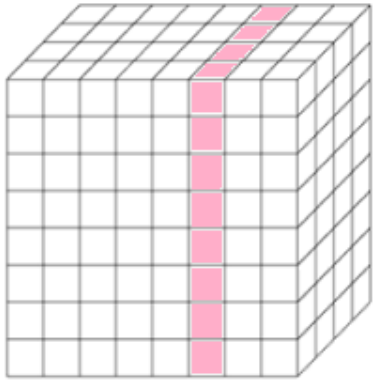
But what about the temporal component?

Real Data



VS.

Single Time Slice



Pt	Age	Gen	Hct	Plt	…	K	Fer
1	38	F	37.2	437	…	3.9	**8**
2	76	M	29.1	68	…	4.1	**Test pt**
…					…		
N	48	F	41.3	212	…	4.1	**221**

Time is a Challenge

- Many existing methods for analyzing time series are not well-adapted to laboratory data
- Clinical laboratory data is
  - "Sampled" irregularly and often sparsely
  - Of high dimensionality
  - Not sampled at random

Current Efforts

Develop the methods and models to accomplish the following:

INPUTS

For a given patient, time point and "analyte of interest"

- Prior values for the analyte of interest and other analytes
- Values for other analytes at the same time
- Desired confidence level (α)

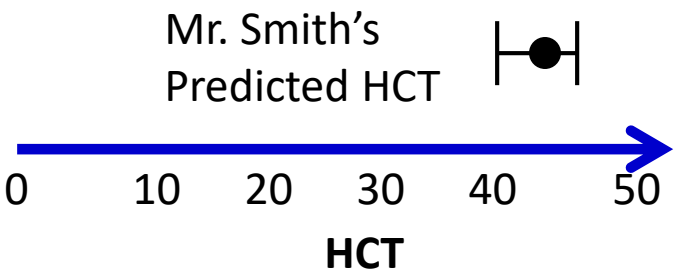
MODEL

OUTPUT

- **Point estimate** and **confidence interval** for the analyte of interest

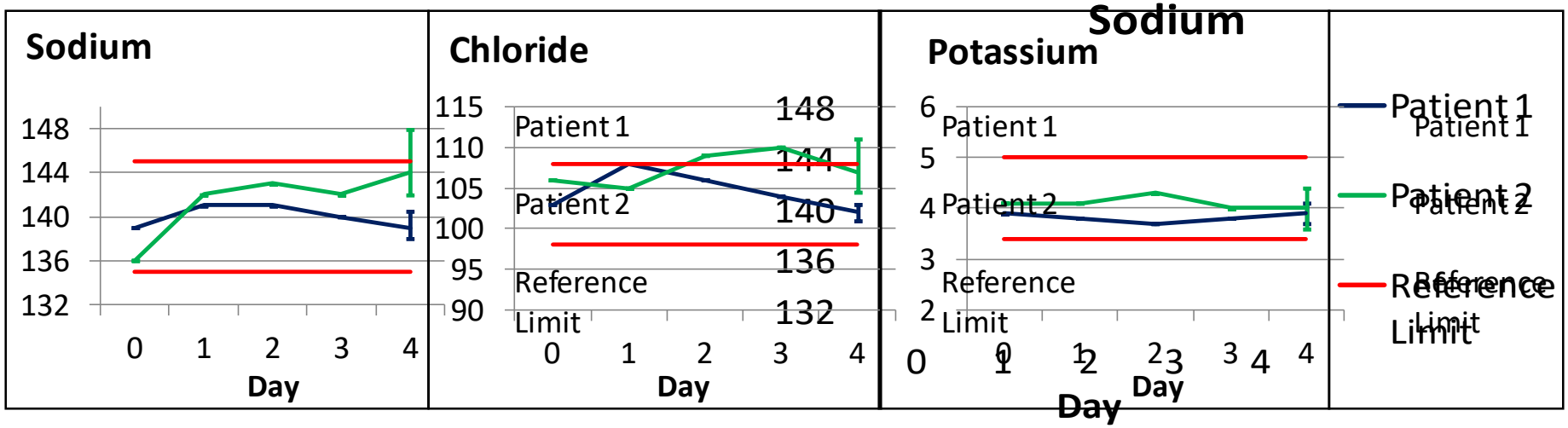
Example

Mr. Smith's prior test results → HCT Model →



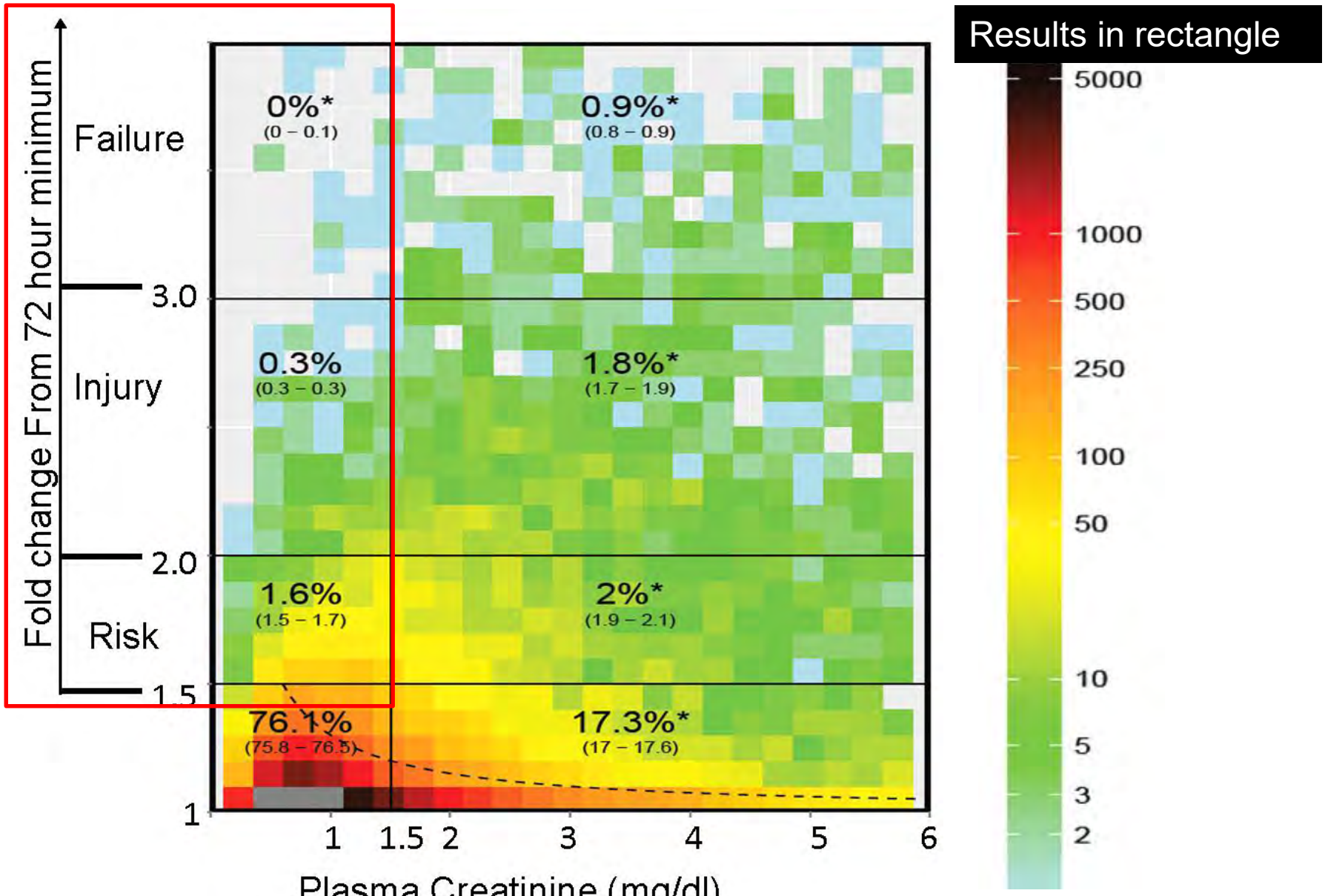
Example Application: Daily Electrolyte Testing

2 patients with Na, K, Cl measured on days 0-3



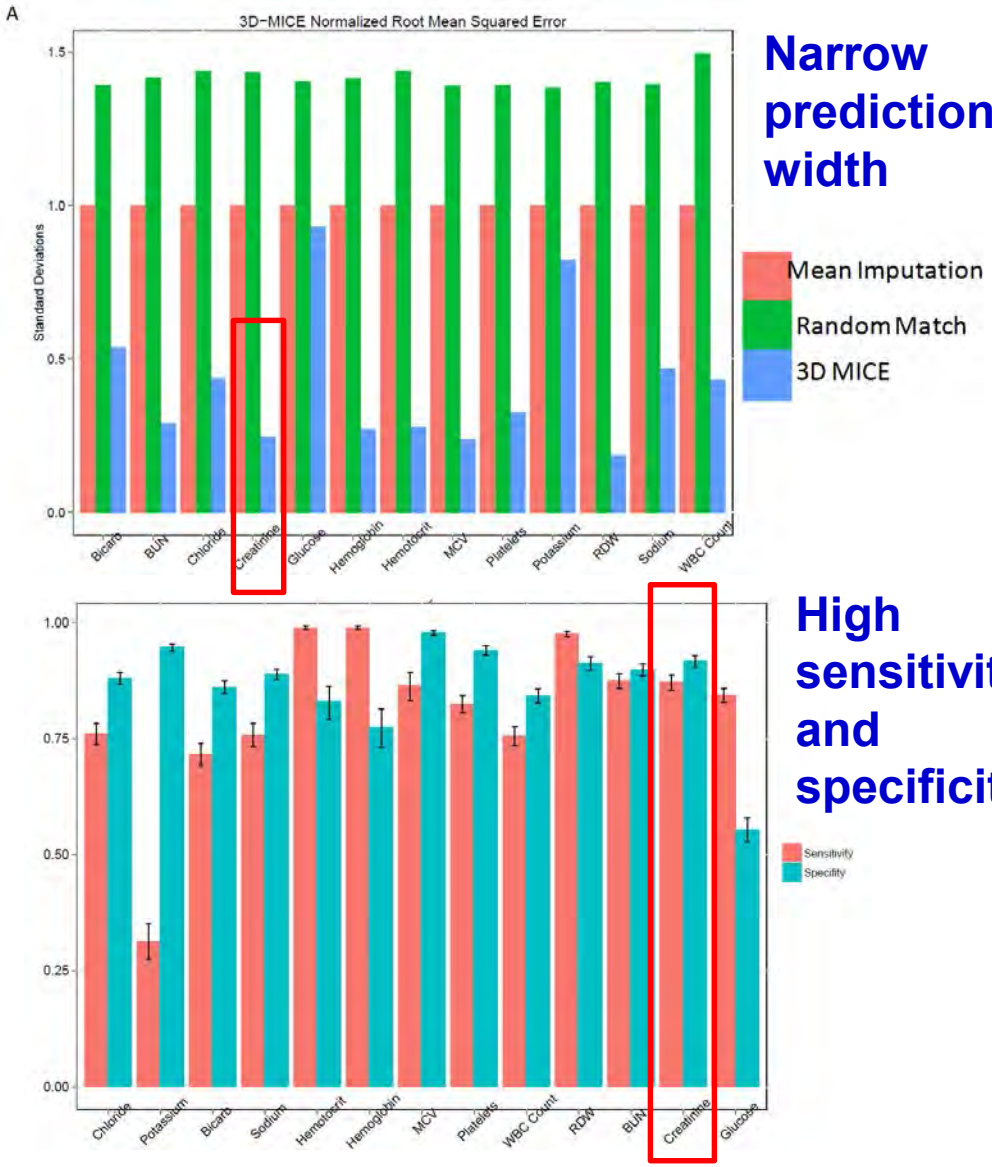
- **Day 4 values = imputations with confidence intervals**
- **Patient 1 might not need this testing**
- **Patient 2 needs day 4 electrolyte testing since values may be abnormal**

Traditional Reference Ranges are Inadequate for AKI Detection



Am J Clin Pathol. 2015, 143:42.

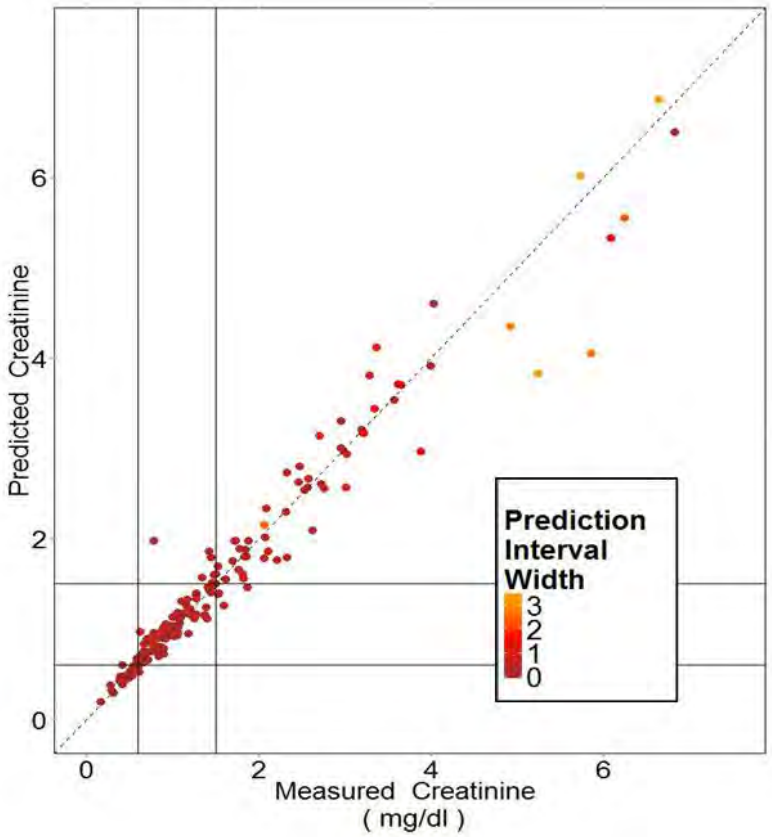
Using 3D-MICE, Future Lab Tests Can Be Predicted With High Accuracy Using Trends and Ancillary Data



Narrow prediction width

High sensitivity and specificity

"Tomorrow's creatinine"

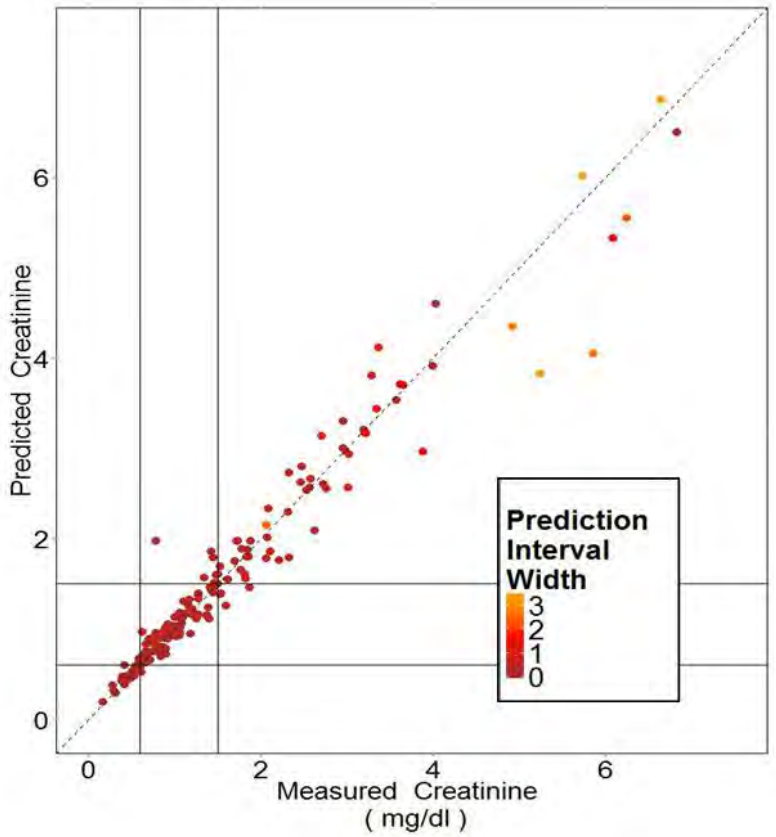


JAMIA 2018: 25(6):645-653.

Accurate Prediction Can Permit Earlier Intervention

- Instead of alerting providers that their patients are **now in early acute kidney injury** (AKI), alert providers that **in 24 hours their patients will likely satisfy criteria for AKI**
- Alert suppressed if calculation of prediction interval width is too wide
- Requires real time data access for data analysis

"Tomorrow's creatinine"

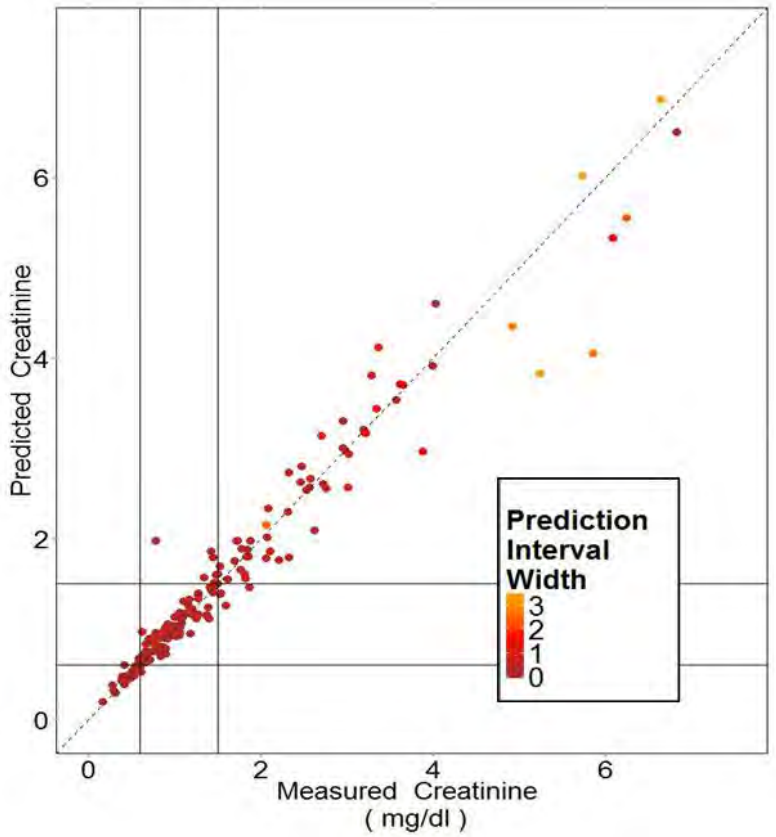


JAMIA 2018: 25(6):645-653.

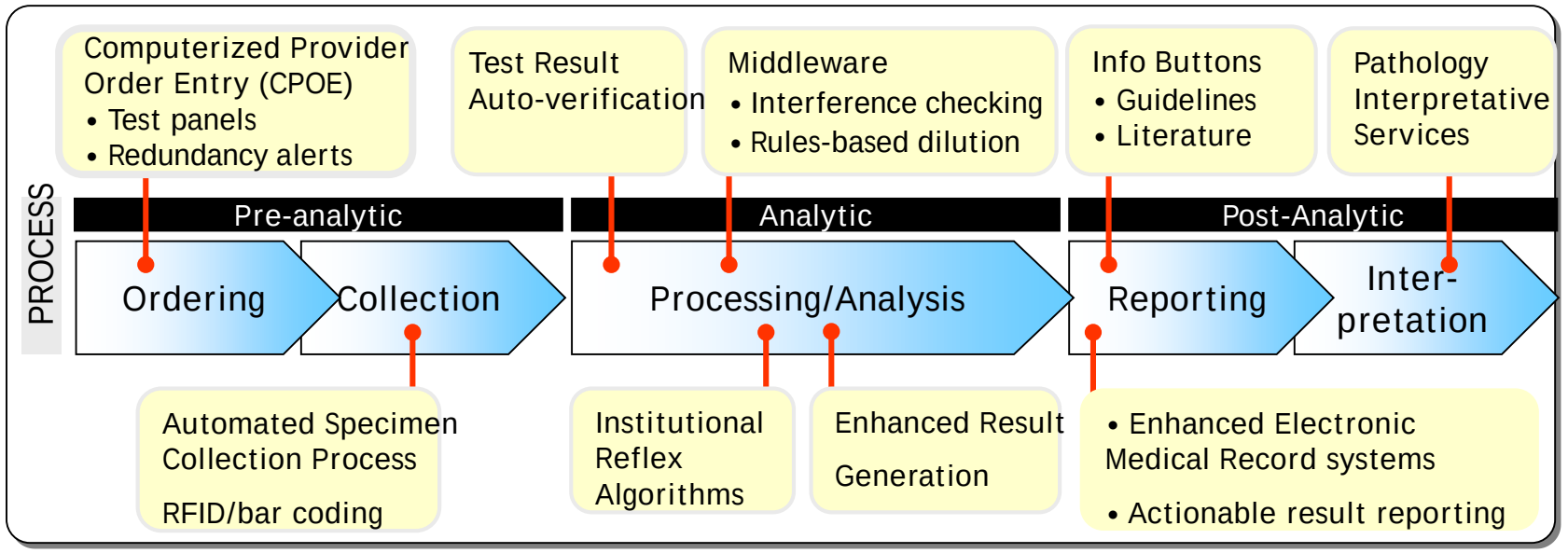
AKI Prevention Trial: Combining Analytics with Traditional Laboratory Evaluations

- New biomarkers for acute kidney injury (e.g. NGAL) may be accurate but are high cost and indications are unclear
- If we add to the EHR inpatient menu we can guarantee ourselves a $300,000 cost for testing without clear evidence of improved outcomes
- **Hypothesis:** One way of efficiently using new biomarkers may be to add them on as a reflex test when prediction criteria have been satisfied
- In the current study when the algorithm predicts a patient will likely meet renal failure criteria in 24 hours → Add NGAL to assess for kidney injury now

"Tomorrow's creatinine"



Informatics to Improve Clinical Outcomes



Informatics can provide the tools to address a wide variety of clinical and operational issues

- Pathologists should engage with and become educated in the capabilities of their EHR
- Most groups already have the infrastructure, pathologists need to step up to the EHR table
- Partnering with computational colleagues can provide mutual benefit but learning the basics yourself is important to be able to frame the questions you want to ask

Thank You

MGH Pathology

- Jason Baron, MD
- Kent Lewandrowski, MD
- Matt Rosenbaum, MD
- Joe Rudolf, MD
- Peter McCaffrey, MD
- Danielle Kurant, MD

MGH Medicine

- Hasan Bazari, MD
- Xingxing Cheng, MD
- Ishir Bhan, MD

MIT Computer Science and Artificial Intelligence (CSAIL)

- Yuan Luo, PhD
- Peter Szolovits, PhD

MGPO

- Jeff Weilburg, MD

Partners Healthcare

- Irina Kamis
- Balaji Singh
- Dustin McEvoy
- Sayon Dutta, MD

COLLEGE of AMERICAN PATHOLOGISTS
Education

CAP19
Knowledge
Relationships
Expertise

COLLEGE of AMERICAN PATHOLOGISTS
Education

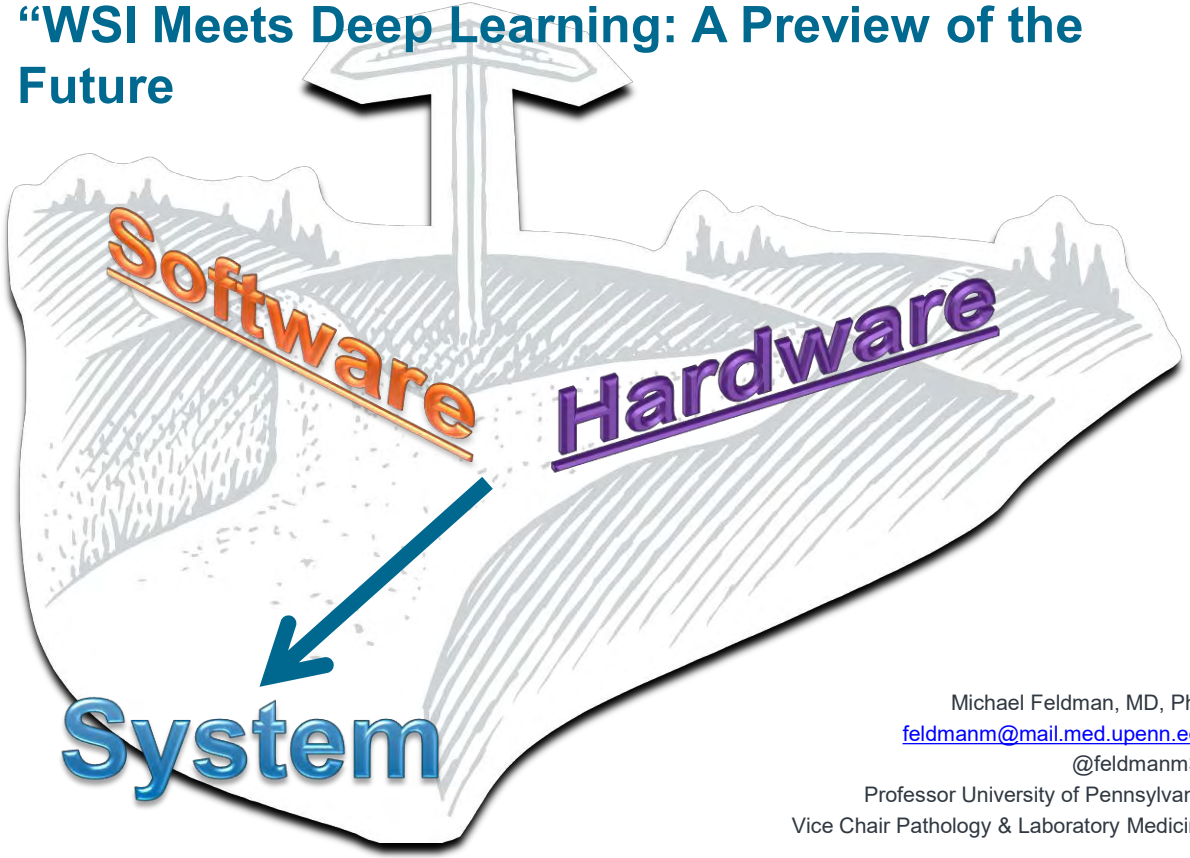
CAP19
Knowledge
Relationships
Expertise

S1953S

A Glimpse into the Future Practice: Transitioning to an Effective Pathology Practitioner in the Age of Machine Learning

Anand S. Dighe, MD, PhD
Barbara S. Ducatman, MD, FCAP
Michael D. Feldman, MD, PhD
Andrew R. Janowczyk, PhD

“WSI Meets Deep Learning: A Preview of the Future



Michael Feldman, MD, PhD
feldmanm@mail.med.upenn.edu
@feldmanm30
Professor University of Pennsylvania
Vice Chair Pathology & Laboratory Medicine

Digital Workflow and ML/AI

Microscope | Res/Fellows Attending | Molecular IHC | Manual | Today

Tissue | Slide | Screening Rescreening | Predictive | Quantitation IHC | Dx

HistoQC

Event finding:
- Ca in Nodes
- Ca in tissue
- Mitoses
- Grading

IBRIS:
- Breast
- Lung
- Prostate
- ENT

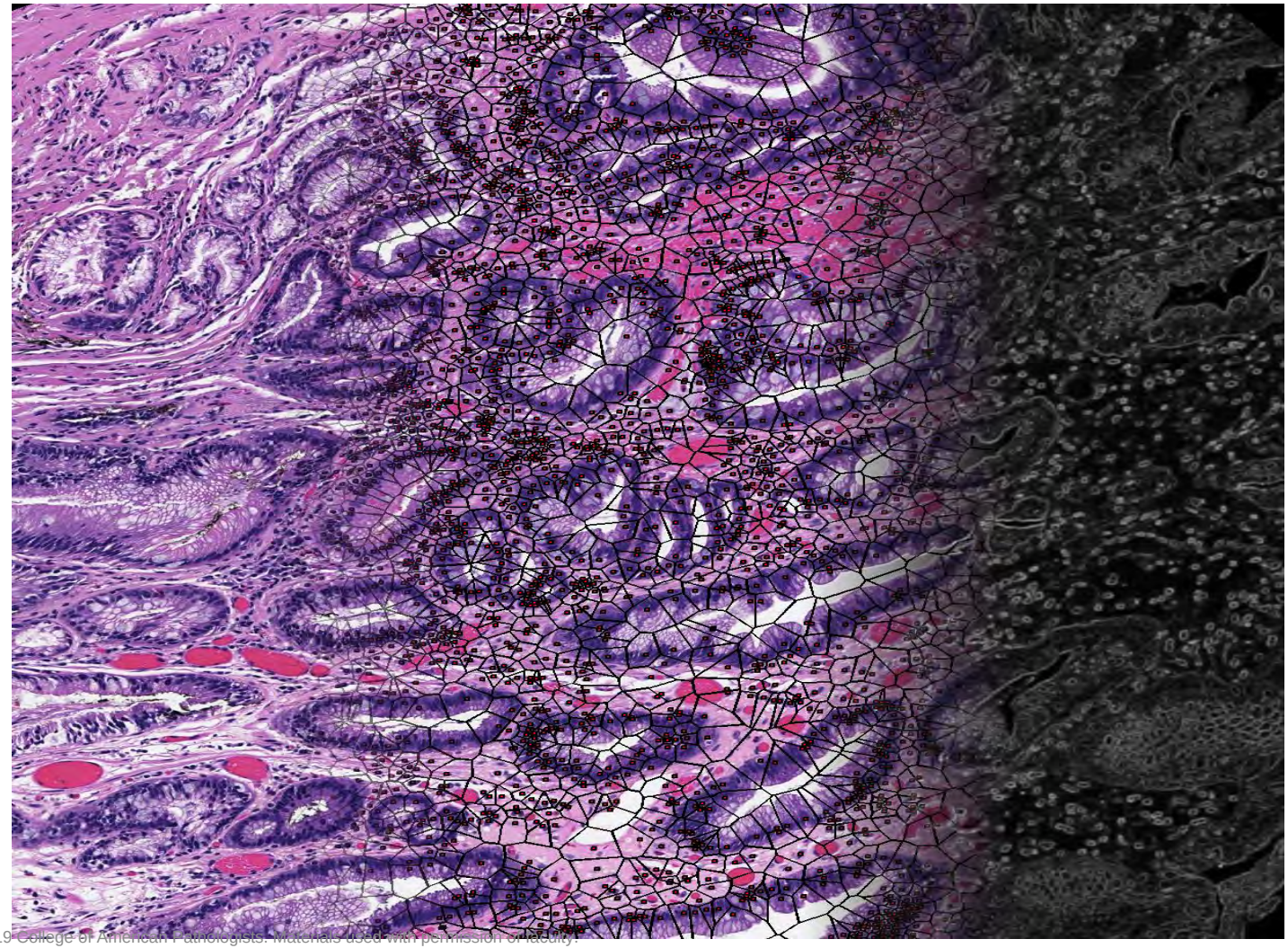
Automated quant
- Companies

Rescreening

EHR

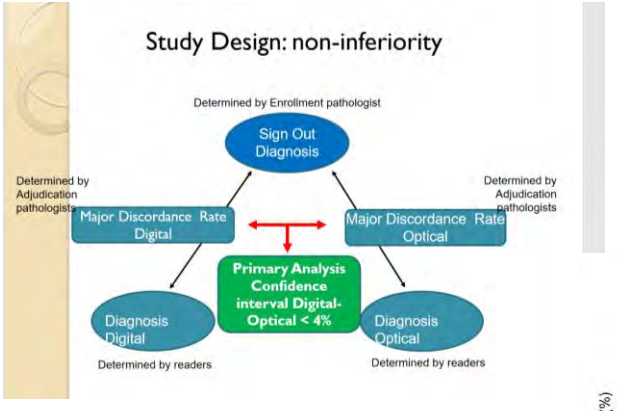
Missed events:
Neg Bx - relook

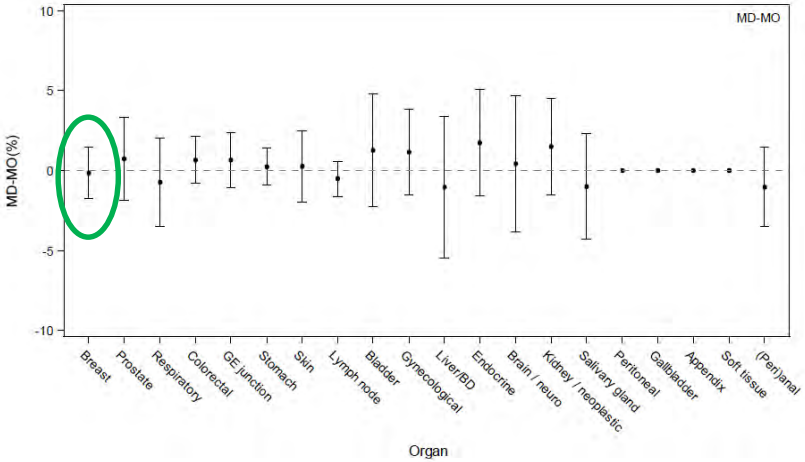
Digital fellow



Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study)

The American Journal of Surgical Pathology: November 2017





Primary analysis

Difference in Major Discordance Rate = Digital minus Optical

Left CI (95%)	Average	Right CI (95%)
-0.31%	0.35%	1.00%

Acceptance criterion <4%

MSK Whole slide imaging equivalency and efficiency study: experience at a large academic center
Mod Path Feb 2019, Sirintrapun et al

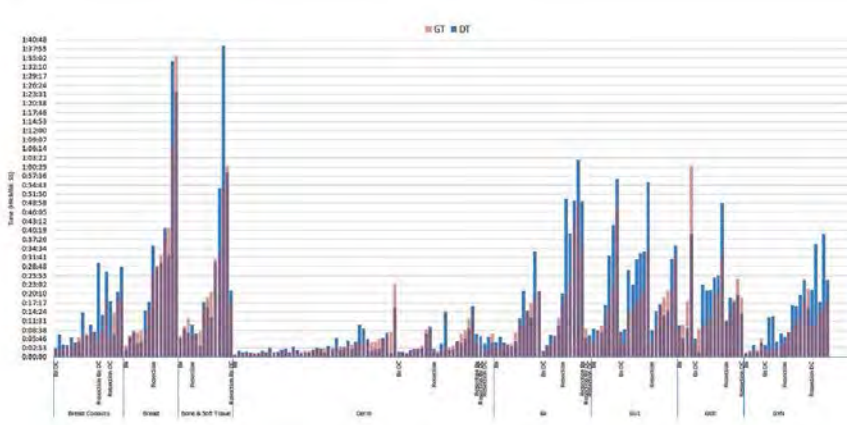
Table 3 List of specimens in each respective subspecialty

Breast		Genitourinary		Gynecologic		Gastrointestinal	
Breast	80	Prostate	119	Cervix	13	Stomach	12
Lymph node	28	Bladder	19	Uterus	10	Colon	10
Other	6	Kidney	8	Fallopian tube	9	Rectum	10
		Ureter	1	Ovary	4	Gallbladder	5
Bone and soft tissue		Urethra	2	Vulva	2	Liver	5
Soft tissue	12	Testis	4			Pancreas	4
Bone	9	Lymph nodes	4	**Dermatopathology**		Esophagus	3

Table 5 Summary of intraobserver concordance and turnaround times of glass and digital reporting

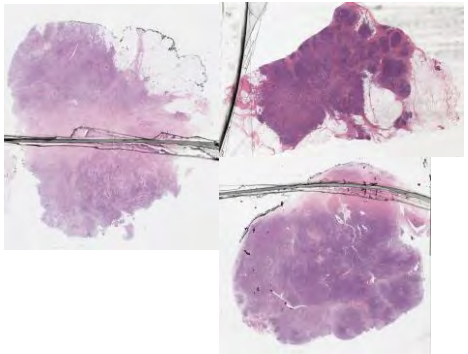
Equivalence		Efficiency[a]		
Concordance			Glass (HH:MM:SS)	Digital (HH:MM:SS)
Diagnosis	99.3%	Pathologist A	6:17:12	8:48:23
Size	4.1% larger by digital measurements	Pathologist B	2:41:15	3:37:46
Grade	94.10%	Pathologist C	4:30:57	4:09:05
Margin	100%	Pathologist D	5:05:01	5:18:49
LVI/PNI	83.80%	Pathologist E	3:06:29	5:00:11
pT	97.30%	Pathologist F	7:19:07	8:22:05
pN	97.10%	Pathologist G	5:49:31	7:37:20
		Pathologist H	5:02:30	5:53:12

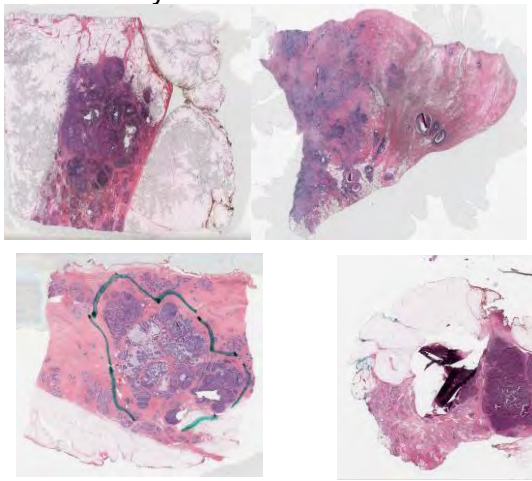
[a]Median time difference was 19 s longer per whole slide image and 107 s longer per digital case



Unmet Need

- Transition to digital pathology workflows
  - Digital Quality Control is paramount
  - Recut and rescan slides immediately before getting to a pathologist
  - Cost and efficiency savings
- Previously not insurmountable
  - Increasingly too time consuming to do manually
  - Non-reproducible



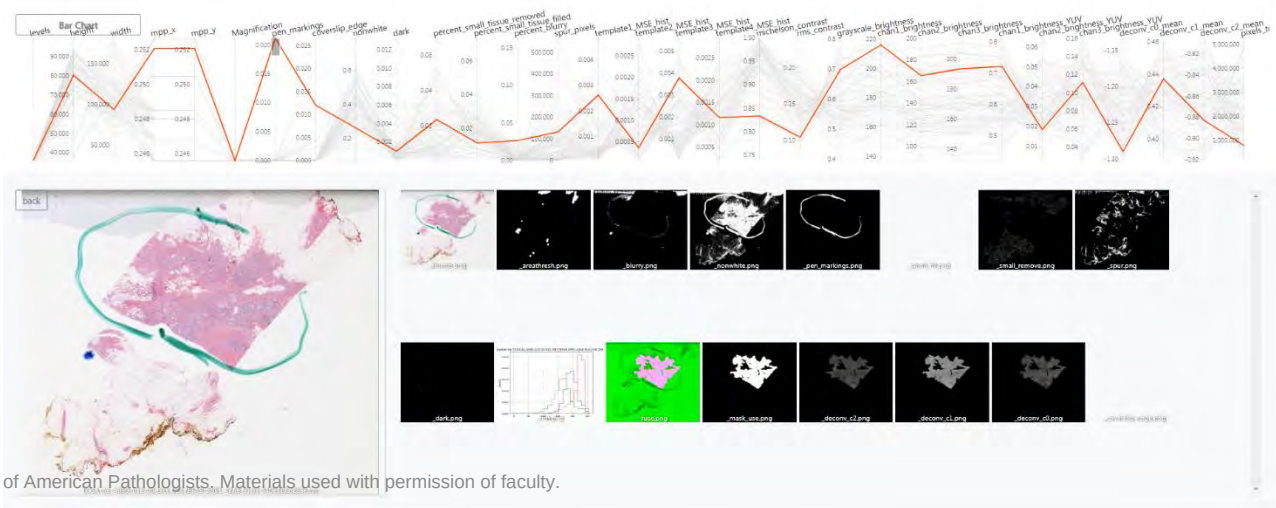


Slides taken from diagnostic cohort of TCGA-BRCA

We need better quality control of our slides!

HistoQC Properties

- Fast: n=1,143 in 466 minutes (24s/slide) using 6 cores (1.1TB)
- Easy “install” (git clone) with minimum dependencies:
  - python-openslide, matplotlib, numpy, scipy, skimage, sklearn
- User interface is a single local html5 + JS file, no hosting
- No specialized hardware
- No internet connection required
- Designed to be modular and easily extendible



HistoQC…Your Pixels Matter

HistoQC: reproducible slide quality metrics with artifact localization

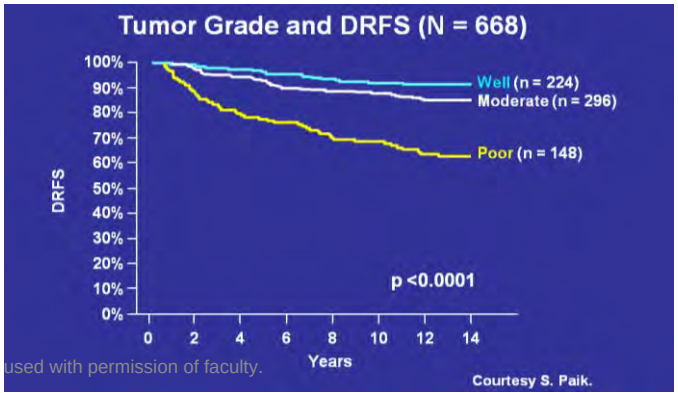
github.com/choosehappy/HistoQC

andrew.janowczyk@case.edu
andrewjanowczyk.com

HistoQCRepo.com

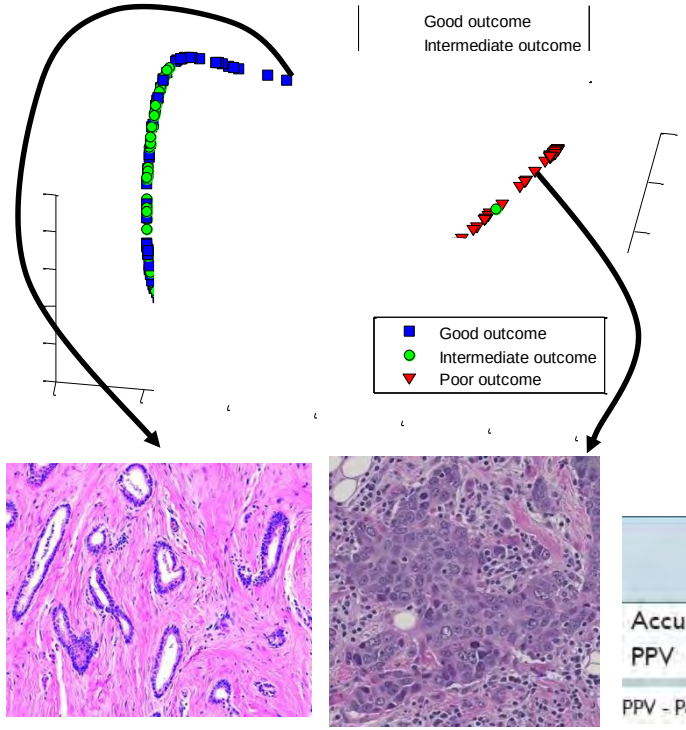
Gold in the hills…
Role of Tumor Morphology ER+ Breast Ca

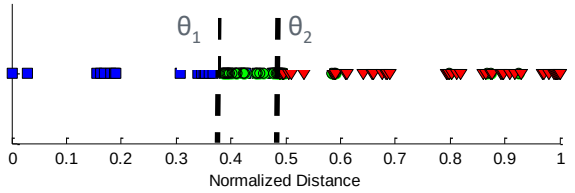
- Modified Bloom-Richardson (mBR) grading (Elston and Ellis,1991)
  - Tubule formation, nuclear pleomorphism, mitotic activity
- mBR identifies tumors as low, intermediate and high grade.
- Correlation between tumor Grade and outcome
- Visually determined, **qualitative**
- High inter- and intra-observer variability
  - Among 7 pathologists: *k = 0.50 – 0.59* (Meyer et al., 2005)
  - Between pathology departments: *k = 0.51 to 0.54* (Boisen et al., 2000)
- <u>Suboptimal treatment</u> can result from incorrect grading (Dalton et al., 2000)



IbRiS: Comparing against Oncotype Dx RS

~450 feature data space
Hand crafted features
Inspired by Pathology expertise



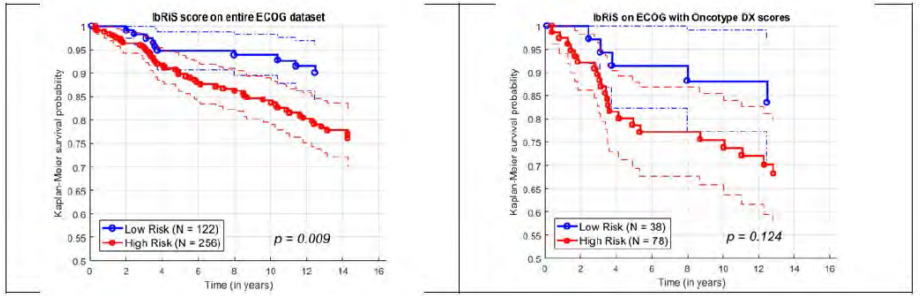


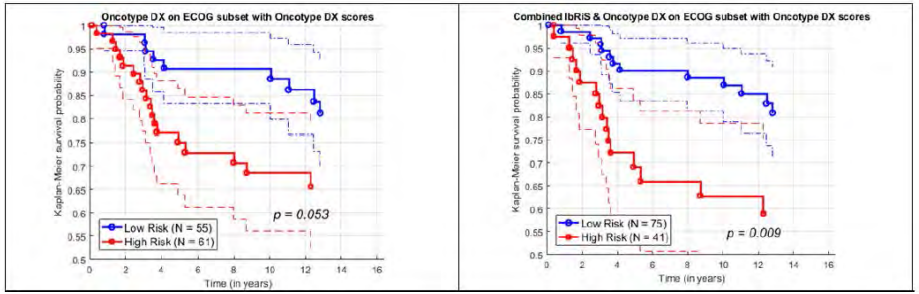
	Good vs. poor	Good vs. intermed	Intermed. vs. poor
Accuracy	0.91 ± 0.022	0.76 ± 0.051	0.83 ± 0.076
PPV	0.94 ± 0.10	0.85 ± 0.11	0.92 ± 0.13

PPV - Positive predictive values

Journal Path Informatics 2011, Madabhushi, Feldman et al

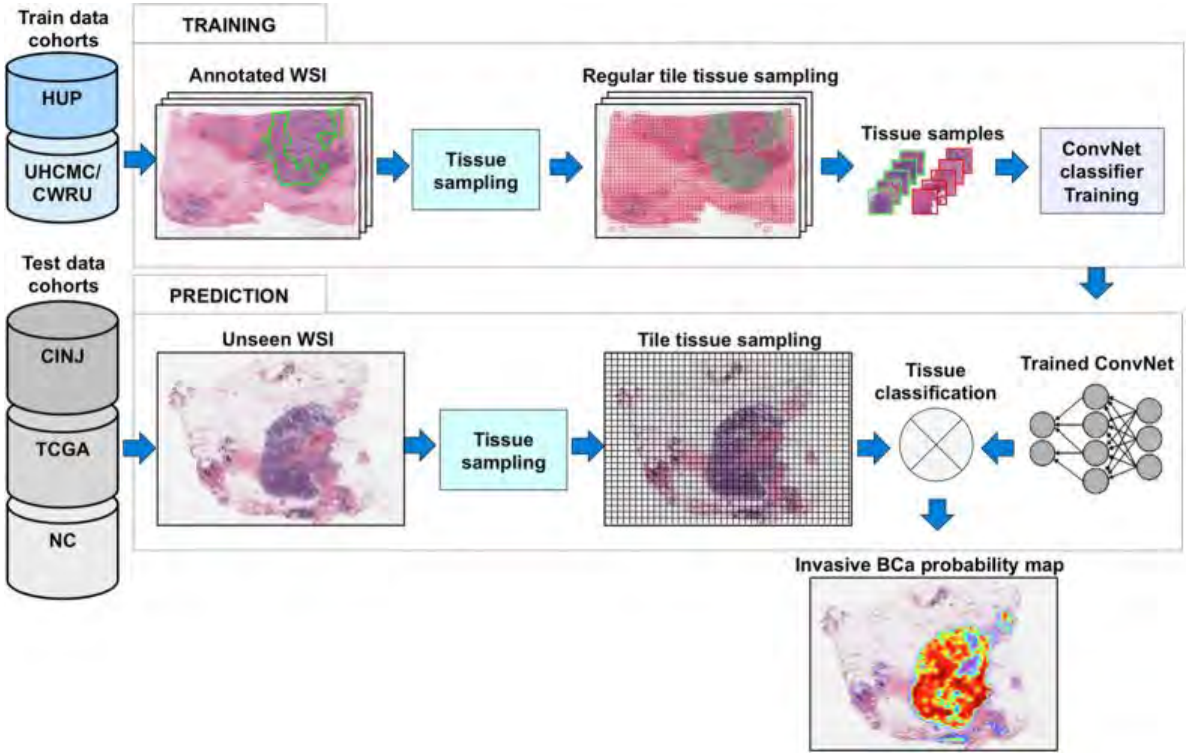
Ibris and outcomes ECOG 2197





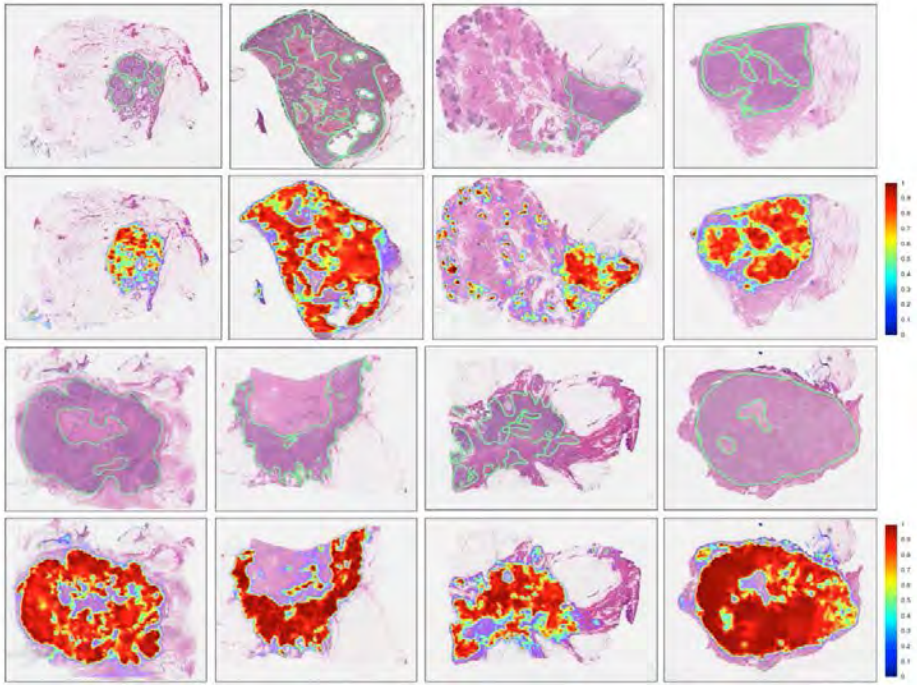
Assay	% of patients with no recurrence after 10 years classified as low-risk	10-year recurrence rate in low-risk group
IbRiS	37.5%	17.2%
ODx	56.3%	20.0%
IbRiS + ODx	75.0%	20.0%

Findings Cancer – Deep learning



Nature Sci Reports 2017 Apr 18:7:46450; Madabhushi et al

Breast cancer maps



False positives
a. DCIS
b. Sclerosing lesions

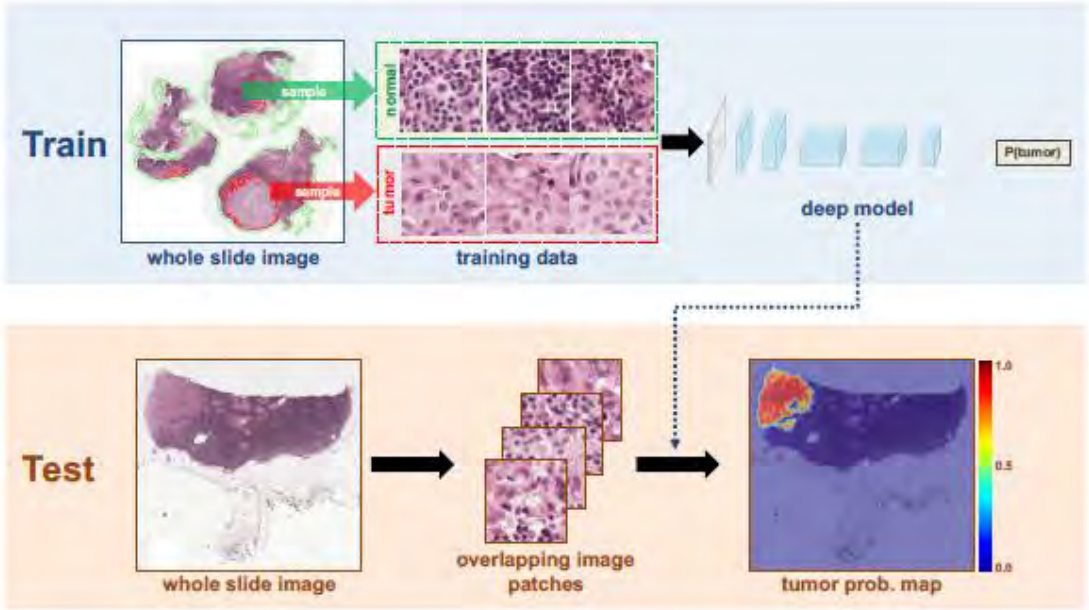
False negatives
a. Small areas of invasion

Data set	Dice	PPV	NPV	TPR	TNR	FPR	FNR
TCGA	0.7586 ± 0.2006	0.7162 ± 0.2204	0.9677 ± 0.0511	0.8691 ± 0.1582	0.9218 ± 0.0764	0.0782 ± 0.0764	0.1309 ± 0.1582
NC	N/A	N/A	1 ± 0	N/A	0.9964 ± 0.0110	0.0036 ± 0.0110	N/A

Table 1. Performance measures for the ConvNet classifier on the TCGA (pathological, N = 195) and NC (normal, N = 21) data cohorts. The measures included Dice, PPV, NPV, TPR, TNR, FPR and FNR. Note that for the normal cases considered, not all the performance measures are shown because the NC data cohort did not have cancer annotations.

Nature Sci Reports 2017 Apr 18;7:46450; Madabhushi et al

Finding Lymph node mets ISBI 2016 CAMELYON 2016/17



Lymph node detection breast cancer

Table 1. Characteristics of the Whole-Slide Images and Glass Slides in the Data Sets Used in the CAMELYON16 Challenge

Data Set (N = 399 Slides and Images)[a]	Hospital Providing the Slides and Images	Primary Tumor Histotype[b]		Slides Containing Metastases, No.			No. of Lesions per Slide or Image, Median (Range)	Total Slides or Images
		IDC	Non-IDC	None	Macro	Micro		
Training (n = 270 images)	RUMC	54	16	100	35	35	2 (1-20)	170
	UMCU	30	10	60	26	14	3 (1-27)	100
Test (n=129 slides and images)	RUMC	23	6	50	14	15	2 (1-14)	79
	UMCU	15	5	30	8	12	3 (1-25)	50

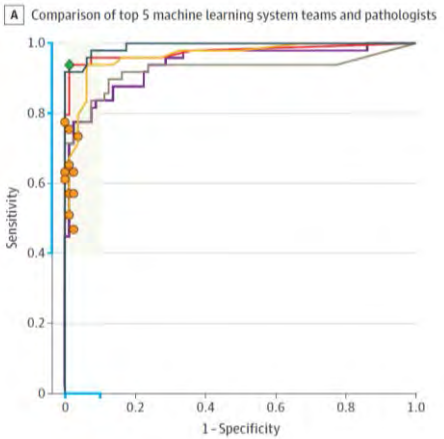
Abbreviations: CAMELYON16, Cancer Metastases in Lymph Nodes Challenge 2016; IDC, infiltrating ductal carcinoma; RUMC, Radboud University Medical Center; UMCU, University Medical Center Utrecht.

[a] All analyses in the training set were determined with whole-slide images.

Analyses in the test were determined with whole-slide images by the algorithms and with glass slides by the panel of 11 pathologists (because diagnosing is most commonly done using a microscope in pathology labs).

[b] Primary tumor histotypes included IDC and other histotypes (non-IDC).

11 Pathologist with time constraint or without time constraint compare to different machine learning algorithms (23 teams, 32 models)



Machine vs person	Performance (ROC)
Pathologist	0.966 (3.4% misses) WOTC
Harvard (Best algorithm)	0.925 (7.5% misses)
Combination Man + Machine	0.994 (0.6% misses)
Dual Neural net	0.994

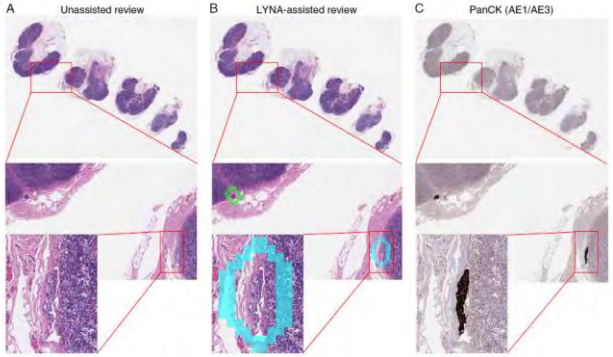
Input & model size	Validation FROC	Validation @8FP	Validation AUC	Test FROC	Test @8FP	Test AUC
40X	98.1	100	99.0	**87.3** (83.2, 91.1)	**91.1** (87.2, 94.5)	**96.7** (92.6, 99.6)
40X-pretrained	99.3	100	100	**85.5** (81.0, 89.5)	**91.1** (86.8, 94.6)	**97.5** (93.8, 99.8)
40X-small	99.3	100	100	**86.4** (82.2, 90.4)	**92.4** (88.8, 95.7)	**97.1** (93.2, 99.8)
ensemble-of-3	-	-	-	**88.5** (84.3, 92.2)	**92.4** (88.7, 95.6)	**97.7** (93.0, 100)
20X-small	94.7	100	99.6	**85.5** (81.0, 89.7)	**91.1** (86.9, 94.8)	**98.6** (96.7, 100)
10X-small	88.7	97.2	97.7	79.3 (74.2, 84.1)	84.9 (80.0, 89.4)	**96.5** (91.9, 99.7)
40X+20X-small	94.9	98.6	99.0	**85.9** (81.6, 89.9)	**92.9** (89.3, 96.1)	**97.0** (93.1, 99.9)
40X+10X-small	93.8	98.6	100	82.2 (77.0, 86.7)	87.6 (83.2, 91.7)	**98.6** (96.2, 99.9)
Pathologist [1]	-	-	-	73.3*	73.3*	**96.6**
Camelyon16 winner [1,23]	-	-	-	80.7	82.7	**99.4**

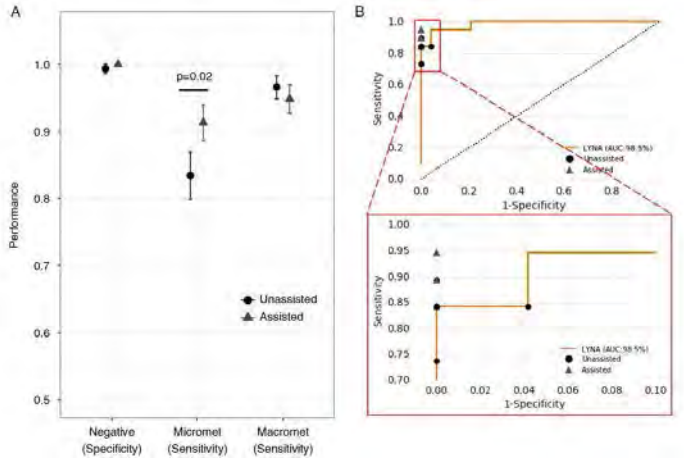
Table 1. Results on Camelyon16 dataset (95% confidence intervals, CI). Bold indicates results within the CI of the best model. "Small" models contain 300K parameters per Inception tower instead of 20M. -: not reported. *A pathologist achieved this sensitivity (with no FP) using 30 hours.

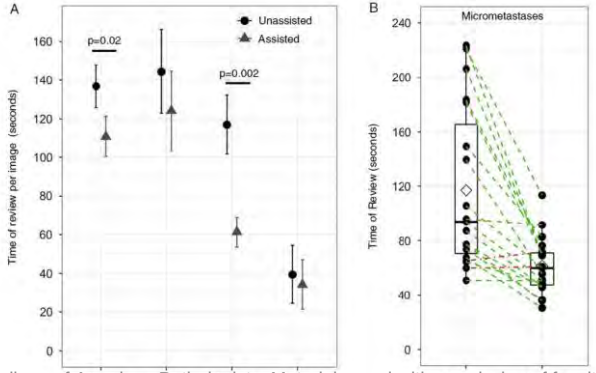
FROC=sensitivity at various FP rates
@8FP is FN rate at 8FP per slide

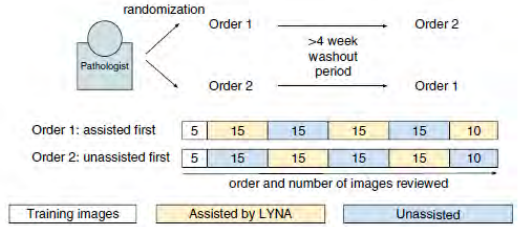
JAMA December 12, 2017 Volume 318, Number 22 and arXiv 1703.02442v2 March 2017

Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer
AJSP 42(12): 2018 Stumpe et al

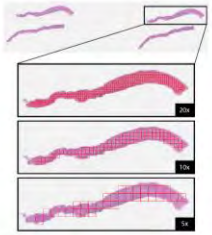




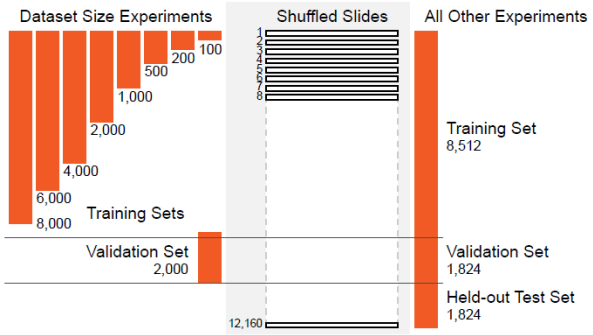


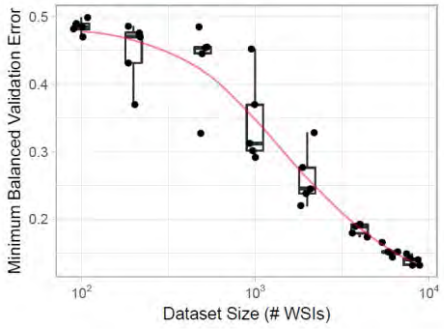


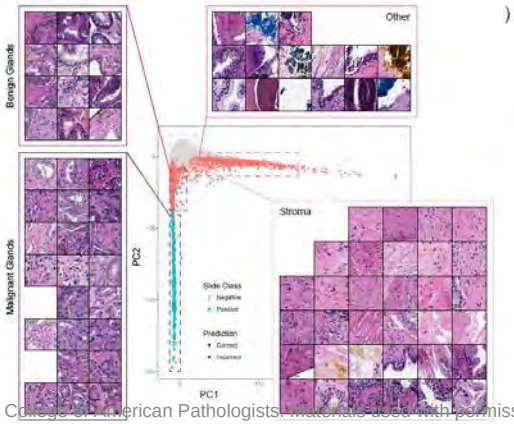
Terabyte-scale Deep Multiple Instance Learning for Classication and Localization in Prostate Pathology



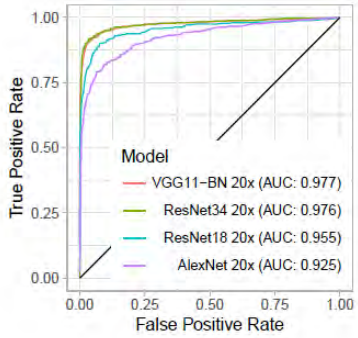
12,000 WSI



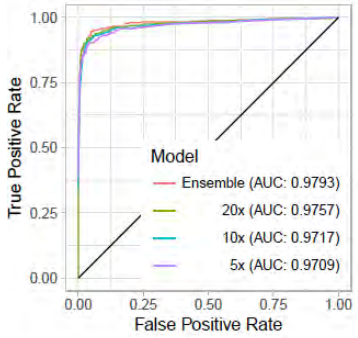




)



b)



Fuchs et al ArXiv May 2018

Rescreening for prostate cancer
Ibex inc.

Maccabi Healthcare Services is a large healthcare provider with a centralized pathology institute - 120,000 histology accessions per year

- ~700 prostate core needle biopsies (PCNBs)
- Roughly 40% of the PCNBs are diagnosed with cancer.

IBEX Medical Analytics, whole slide images of PCNBs, including cancerous glands (of Gleason patterns 3, 4 and 5), high-grade PIN and inflammation. The algorithm utilizes state-of-the-art Deep learning CNN, trained on many thousands of image samples, taken from hundreds of PCNBs from multiple institutes, and manually annotated by senior pathologists.

Small study shown at ECDP 2018 in Helsinki – 100 retrospective cases that had been diagnosed as benign, and found two three errors

- In two cases, the algorithm identified small foci of Gleason 3. Placed into watchful waiting groups. Two years later, both patients were diagnosed with higher grade cancer and underwent radical prostatectomy.
- Third case was a larger focus of pseudo-hyperplastic CAP, resection showed a CAP(4+3) confined to prostate.

System now used to rescreen all negative core prostate biopsies

- New workflow, AI has 30-40% false positive rate – pathologist then reviews specific cores identified by hotspots to decide if any lesion needs further workup or staining

Problems with AI

Brittle

- Most published papers are small data sets that don't always validate well
  - Deep Mitoses – training vs validation very different

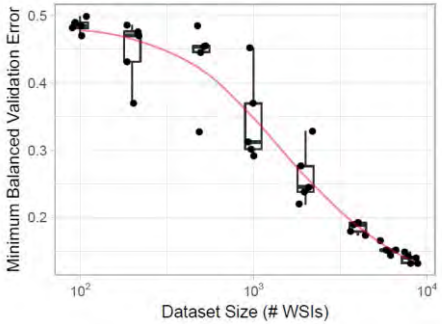
Table 4
Performance comparison of DeepDet with other competing approaches on 2012 MITOSIS test set.

Method	Precision	Recall	F-score
DeepDet	0.854	0.812	**0.832**
RRF (Paul et al., 2015)	0.835	0.811	0.823
CasNN (Chen et al., 2016a)	0.804	0.772	0.788
HC+CNN (Wang et al., 2014)	0.84	0.65	0.735
IDSIA (Cireşan et al., 2013)	0.886	0.70	0.782
IPAL (Irshad et al., 2013)	0.698	0.74	0.718
SUTECH (Tashk et al., 2013)	0.70	0.72	0.709
NEC (Malon et al., 2013)	0.75	0.59	0.659

Table 5
Performance results of our methods on 2014 MITOSIS validation set.

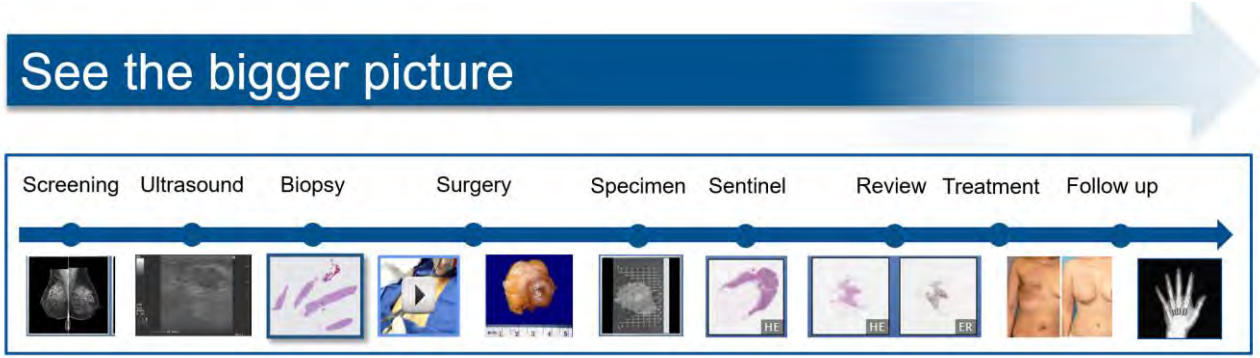
Method	F-score
DeepDet (fixed)	0.489
DeepDet+Seg	0.505
DeepDet+Seg+Ver(c)	0.559
DeepDet+Seg+Ver(f)	**0.572**

Medical Image Analysis 45 (2018) 121–133



Fuchs et al ArXiv May 2018

Path Rads integration



Central PACS

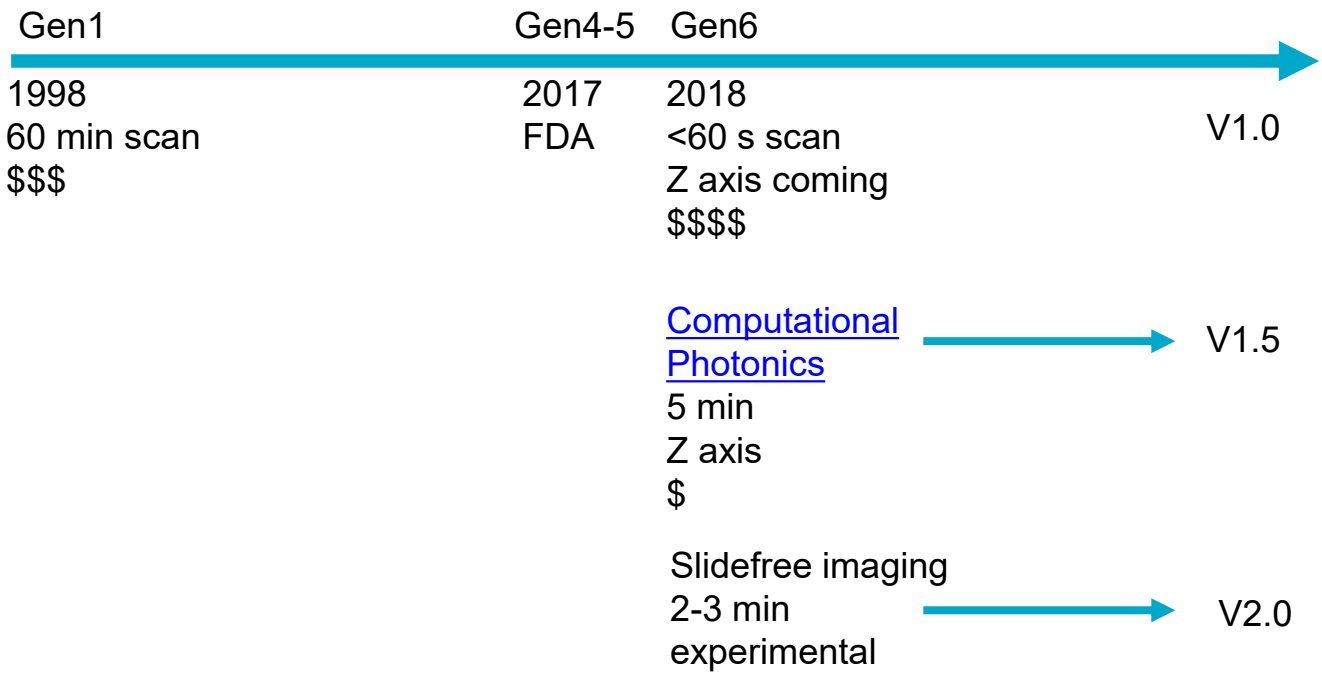
High volume
- Workload balancing
- Access to subspecialists

Image Exchange Portal

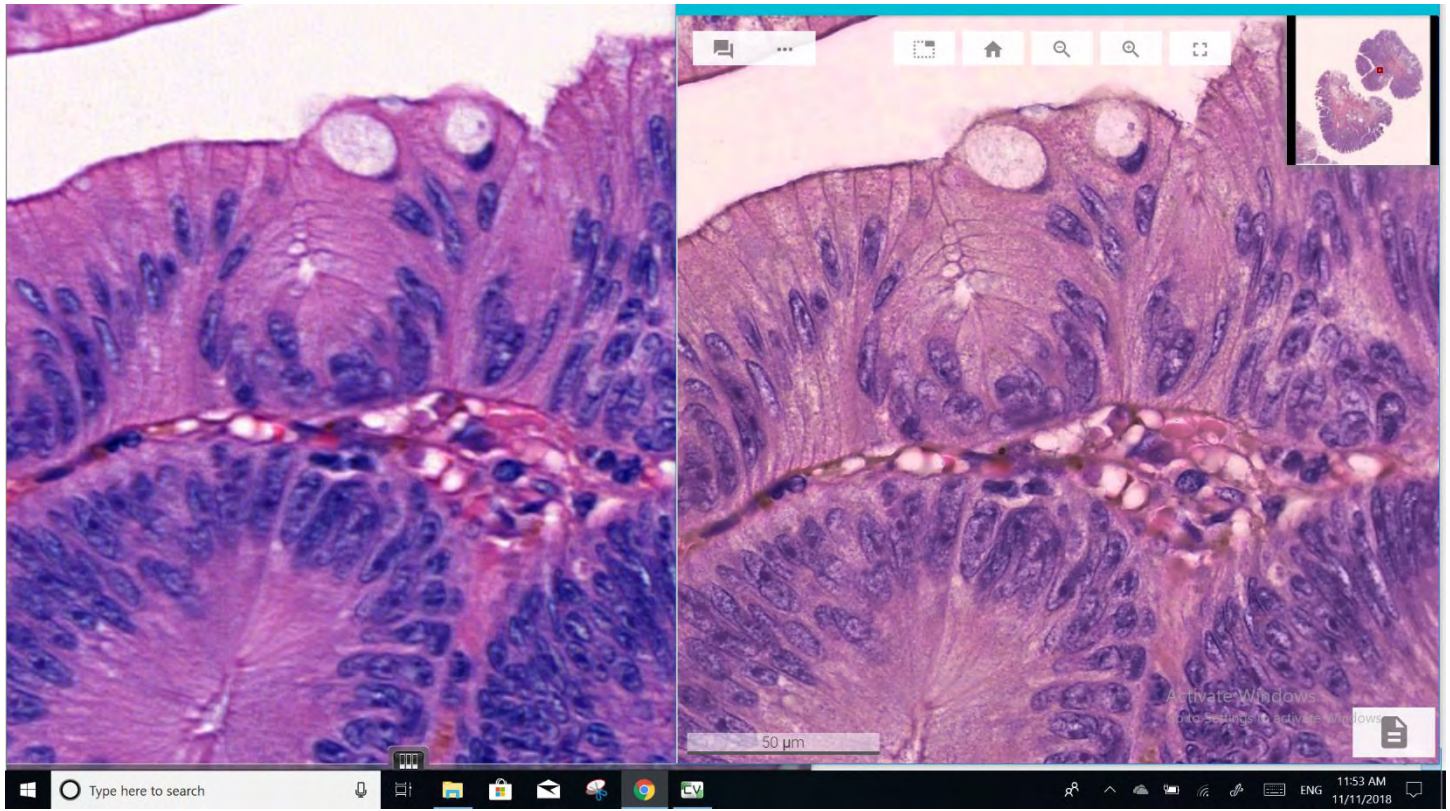
Low volume
- Consultations to anyone
- Second opinion

<u>Common Workflow</u>
- Synoptics
- Data from Image
- Compute on Image
- FTE
- Server/Storage
- EMR/LIS interface
- AI/ML

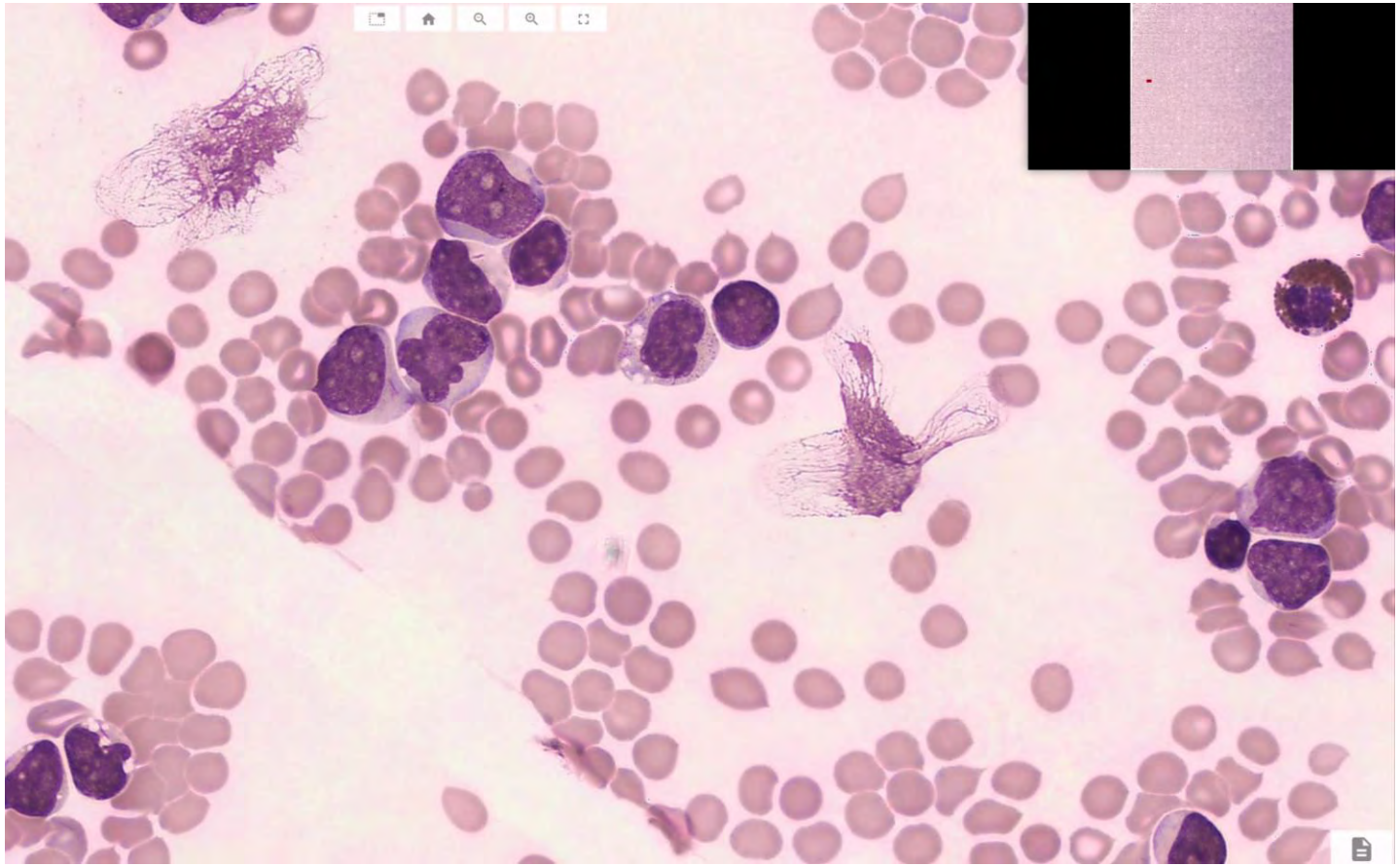
Industry timeline digital pathology



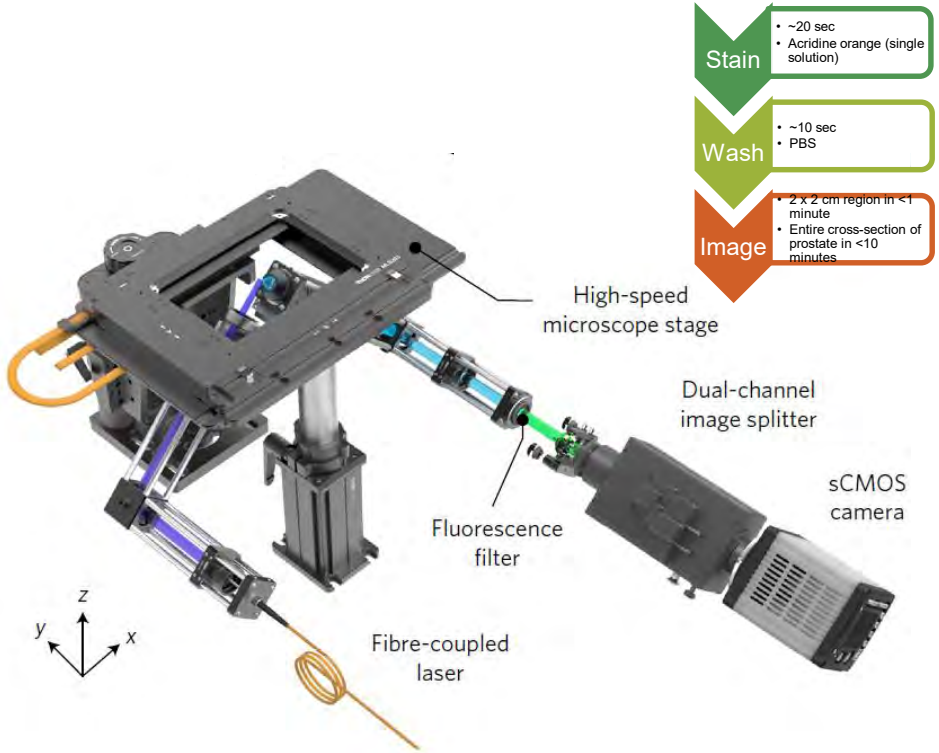
New imaging modalities coming

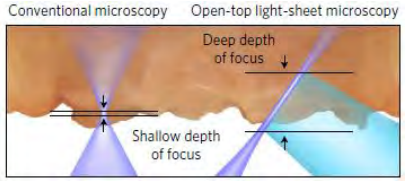


0.75NA 1.2 NA



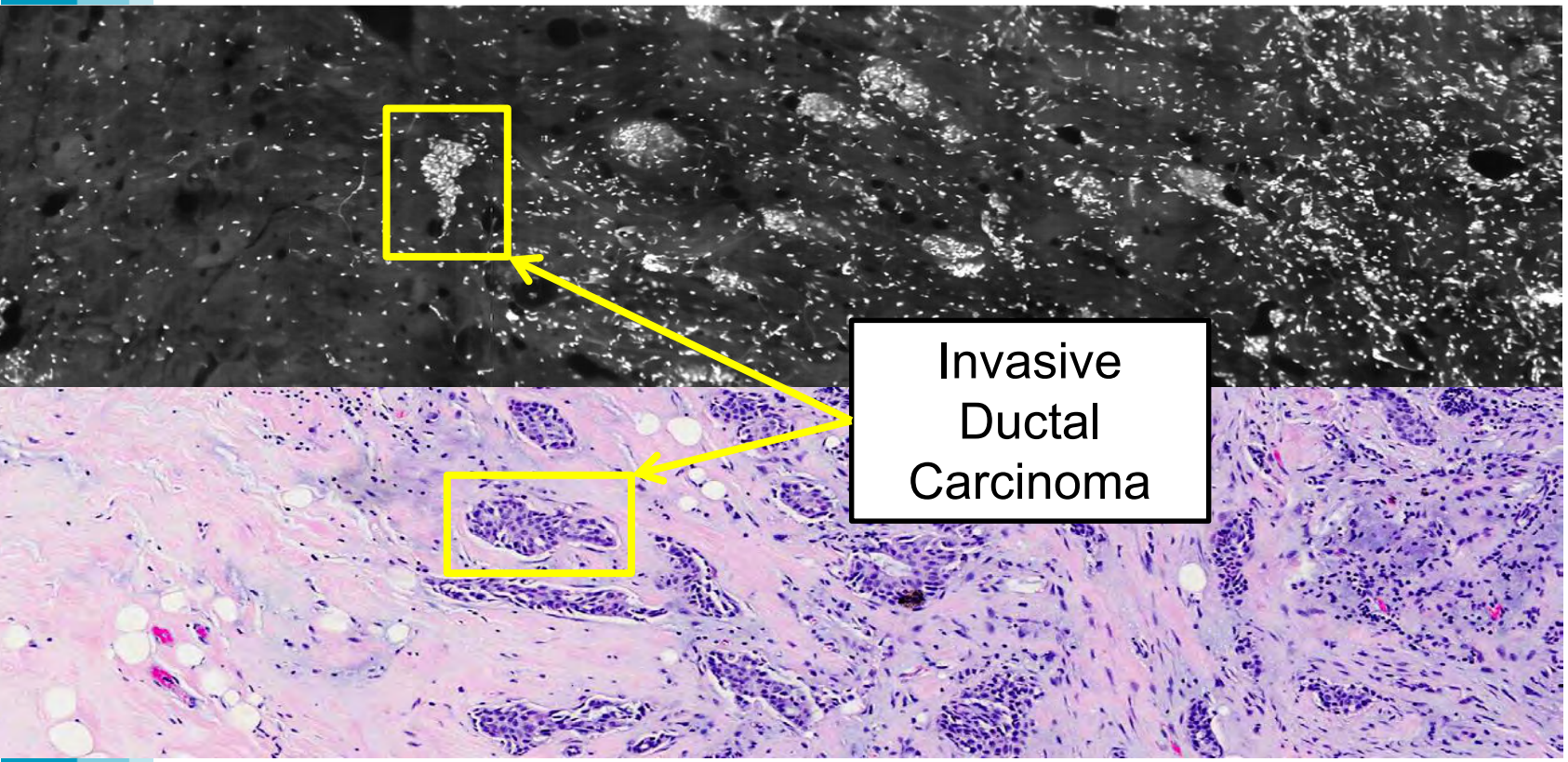
Light-sheet Fluorescence Microscopy





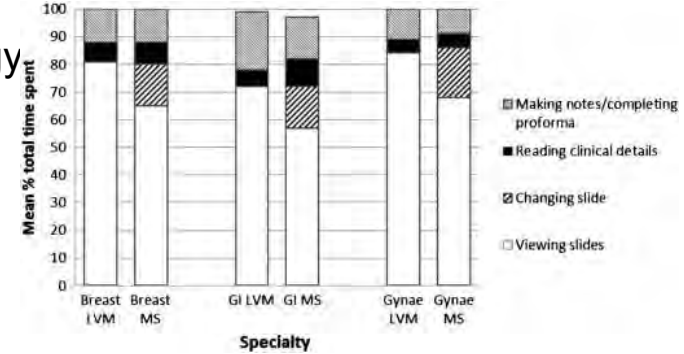
- Separate excitation and emission paths
- Sheet of excitation created by deep depth of focus
- Increased speed with imaging a plane rather than single point
- Able to reconstruct optical sections

Liu et al, Nature Biomed 2017



Digital Pathology is not just images

- Treanor et al 2014 time and motion studies
  - 1/4-1/2 of time is looking through medical record for data
  - How do we make data gathering more precise, more focused and faster?
  - Center healthcare Innovation EHR extensions
    - Yevgeniy Gitelman, Katherine Choi
    - Oncology, Pathology, Radiology



Hum Path Vol45(10) Oct2014: 2101-6

Exploring EHR Extensions

Contextual Inquiry → Scope + Data Feasibility → Prototype + Design → Build + Iterate

Focused Snapshots: Breast Oncology

Reduce Hunting for Information

Baseline: Time to prep chart before visit = 30 min / new patient referral

Contextual Inquiry

Notes serve many purposes

For future self:
Avoid the pain of recreating the picture of the patient from scratch

For others:
Show what's relevant at time of note being written

Gathering relevant data to form a picture of the patient is effortful

- Too much screen switching between data types
- Drilldown is inefficient: reports and links are buried
- Difficult to jump into timeline where most relevant
- Gaps: Others' notes focus on diff parts of the story
- Gaps: Feeling there is a clue in the chart you didn't see
- Lots of reformatting
- Unclear what data other providers need

Exploring EHR Extensions

Scope + Data Feasibility

Prototype + Design

Notes, **Imaging**, and **Surgical Pathology** on one timeline chronology

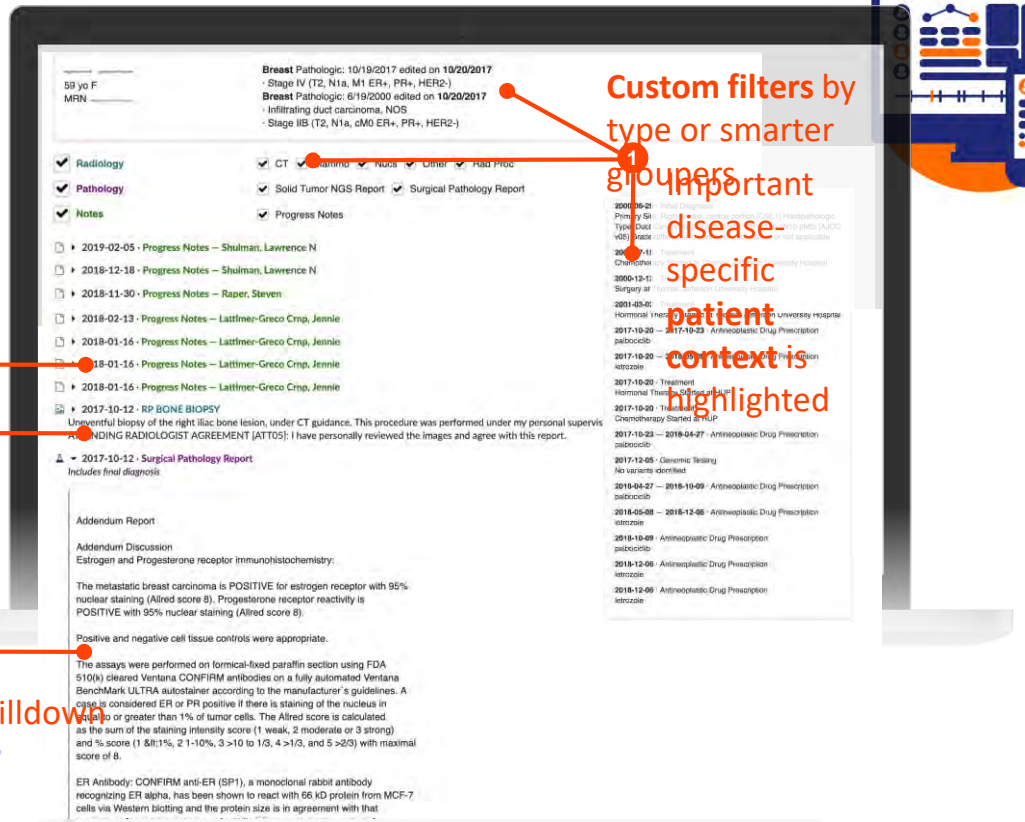
Focus on the key 30%

Smart previews show "Impression" for quick scanning

Focus on the key 5%

Full reports for drilldown

Eliminate non-clinical 50%



Custom filters by type or smarter groupers

Important disease-specific **patient context** is highlighted

Conclusions

- FDA clearance is only a beginning
- Machine learning will accelerate
  - Targeted review
    - Rare event detection
    - Tumor finding
    - Feature classification
    - Grading
    - Screening/Rescreening
  - Outcome prediction
  - qIHC, qMultiplex
- Large well curated and annotated datasets are platinum
- Data Science is our future
  - AI and ML are key to unlock our data
  - Path-Rads integration is our future
  - New technologies coming
- Business – new models of practice

Case Western Lab Director:
Anant Madabhushi, PhD

Postdocs:
James Monaco, PhD
Gaoyu Xiao, PhD
Jun Xu, PhD
Andrew Janowczyk, PhD

Graduate Students:
Jonathan Chappelow
Scott Doyle
Satish Viswanath
Pallavi Tiwari
George Lee
Shannon Agner
Ajay Basavanhally
Rob Toth
Andrew Janowczyk

Undergraduate Students
Jay Naik
Hussain Fatakdawala
Amod Jog

Penn Clinical Collaborators
Mitch Schnall, MD, PhD
David Roth, MD, PhD
John E Tomaszewski, MD
William Lee, MD
Natalie Shih, MD

Clinical Collaborators
Shridar Ganesan, MD, PhD

Penn Center Clinical Innovation
Roy Rosin
Yevgeniy Gitelman, MD
Katherine Choi

Introduction To Machine Learning Using Examples From Anatomic Pathology

September 24, 2019

Andrew Janowczyk, PhD
Assistant Research Professor

CASE SCHOOL
OF ENGINEERING
CASE WESTERN RESERVE
UNIVERSITY

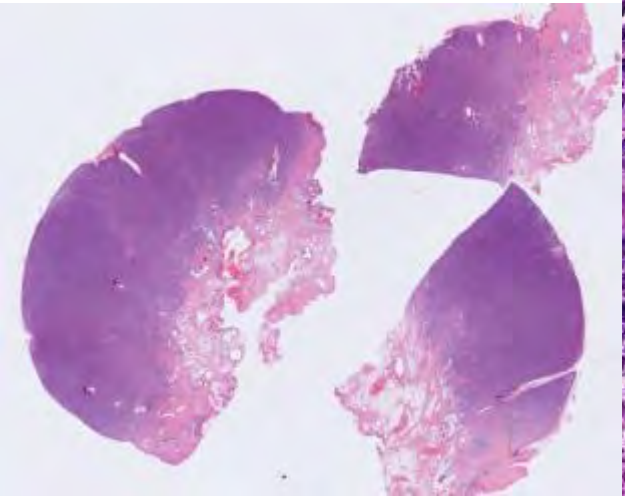
Center for Computerized Imaging and
Personalized Diagnostics

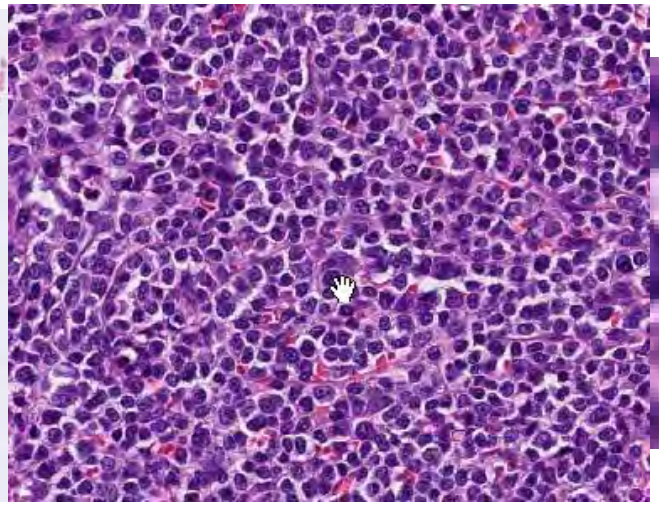
Outline

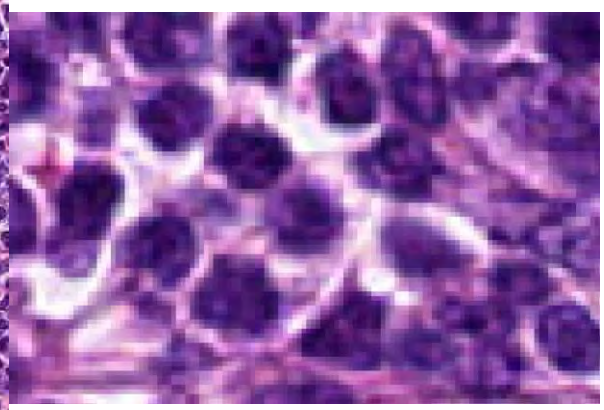
- What are images?
- What can be done with them?
- Feature extraction
- Intuition behind Classifiers
  - Real world examples
- Important considerations
  - Types of annotations
  - Batch effects
  - Quality Control

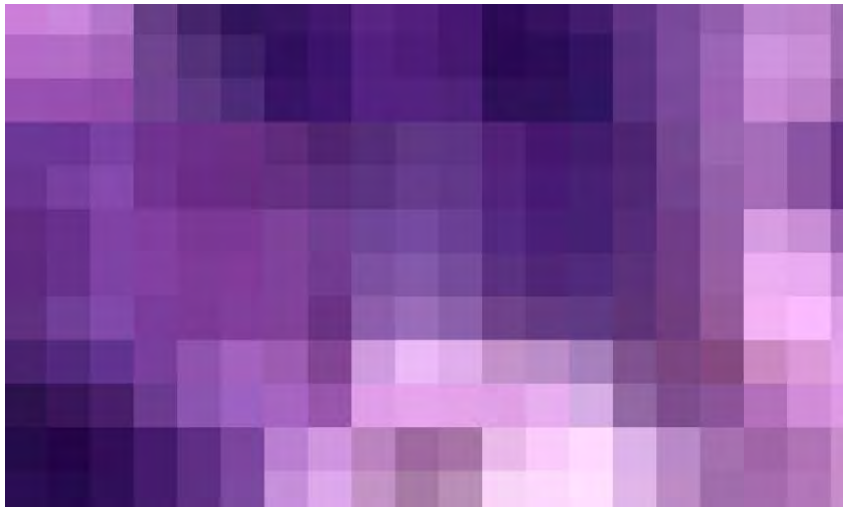
- If anything is unclear, let me know!

Digital pathology images are pixels



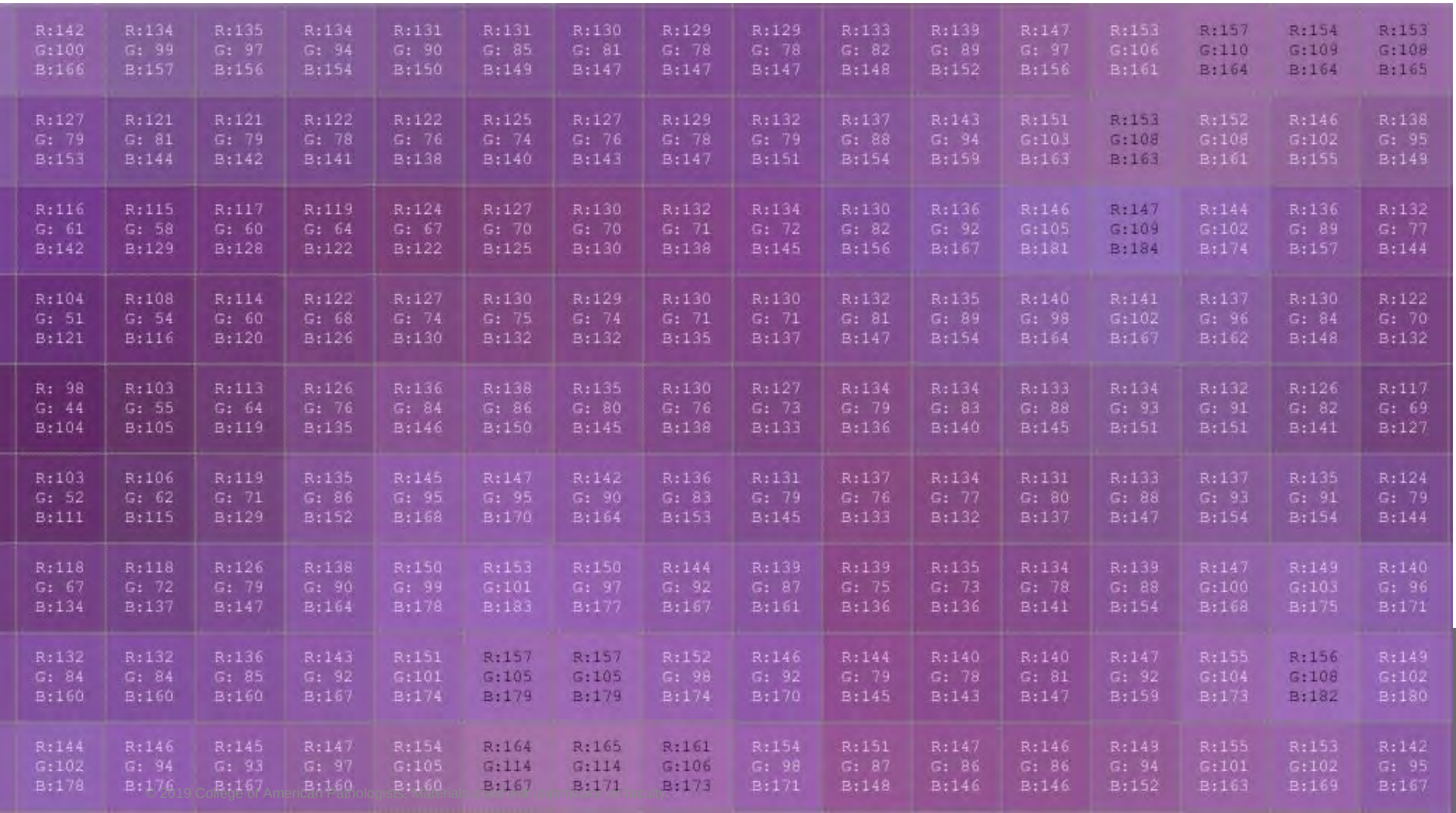






Images are 3D matrices of Pixels

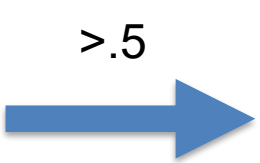
Width x Height x [Red Green Blue]



Pixels can have operations done on them

- Processing and Filtering:
  - Taking a pixel
  - Perform a function
  - New Value

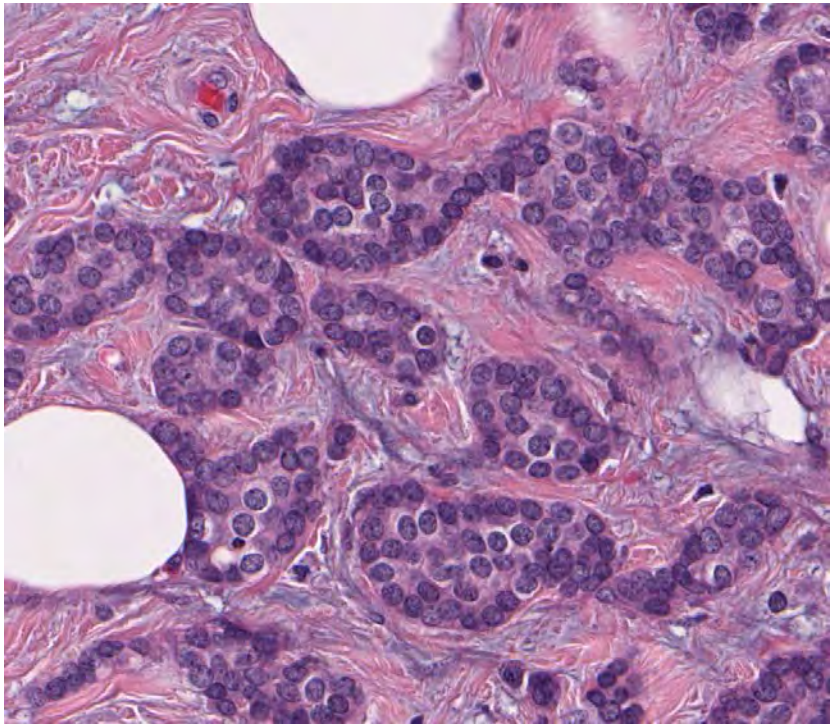
.5	.3	.5
.3	1	.2
.6	.3	.9

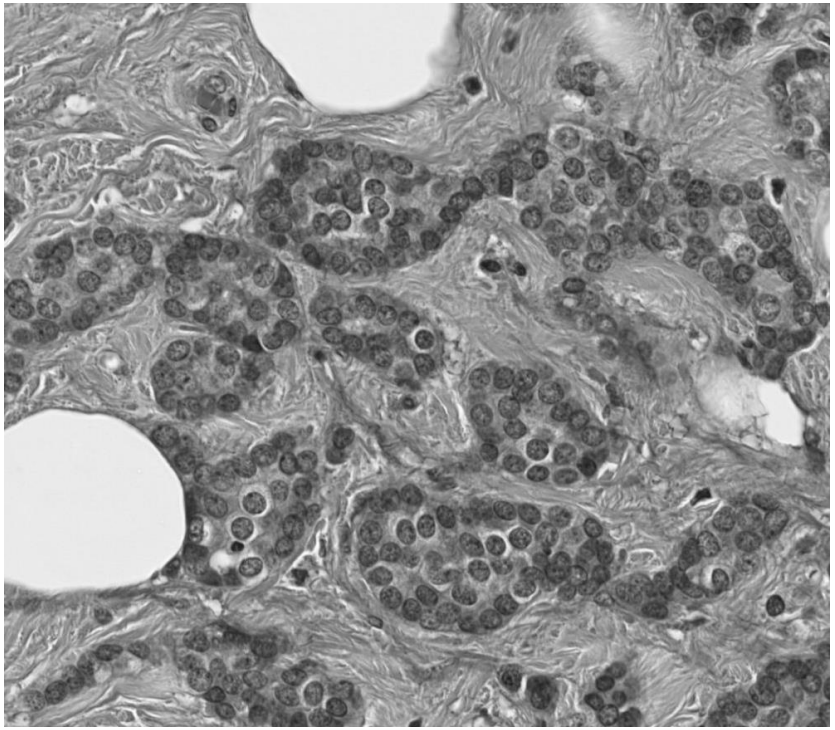


0	0	0
0	1	0
1	0	1

Can convert from RGB to Grayscale

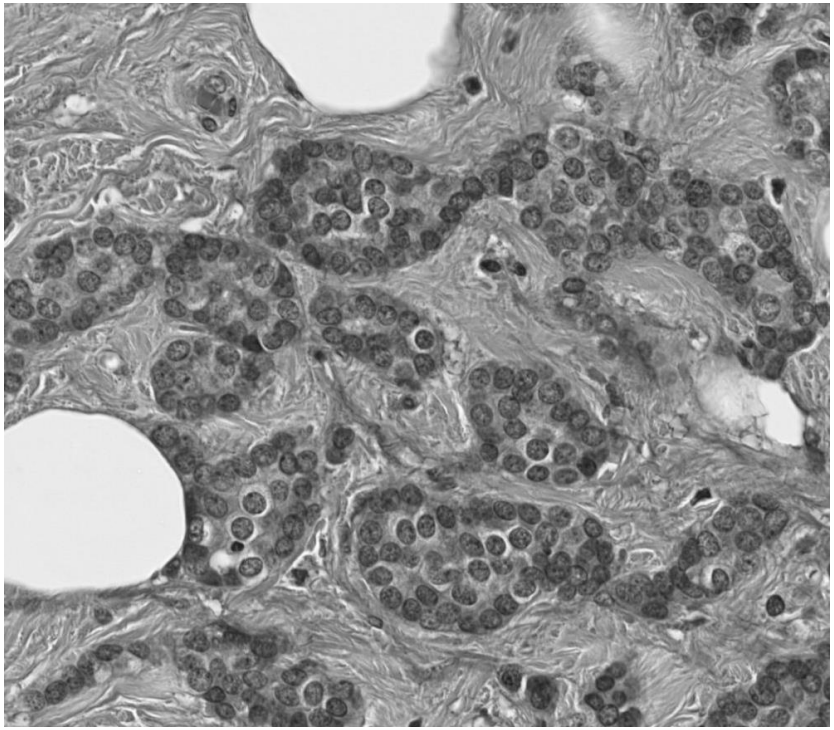
- Pixel by pixel - Linear equation
- Gray = 0.2989 * R + 0.5870 * G + 0.1140 * B

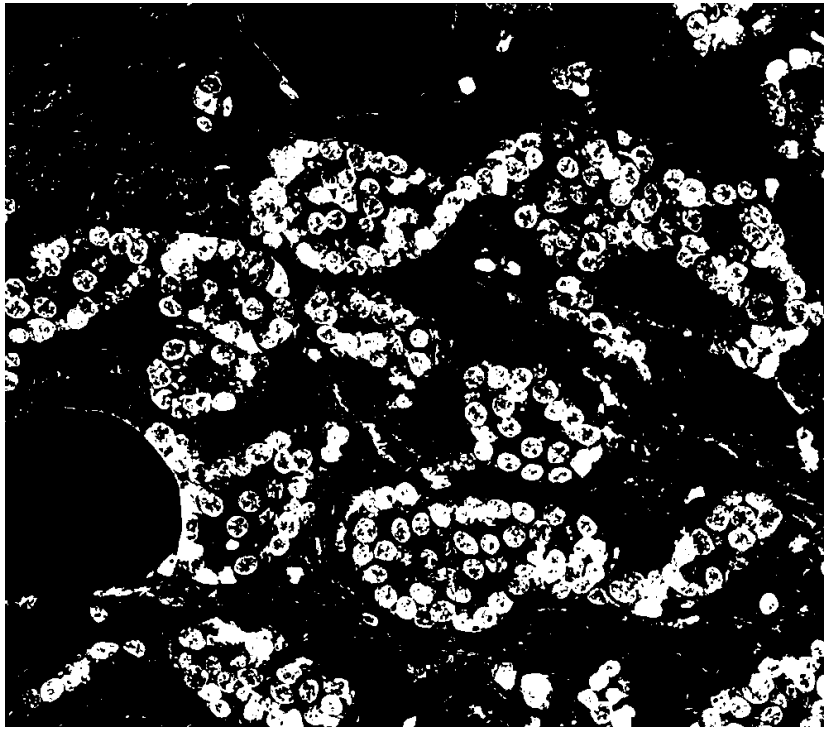




Can apply a threshold

- Images range from [0 = black ,1 = white]
- Values < .5





Pixels can have operations done on them

- Processing and Filtering:
  - Taking a pixel
  - Look at its neighbors
  - Perform an operation

2	3	6
7	1	2
8	6	5

Sum

	40	

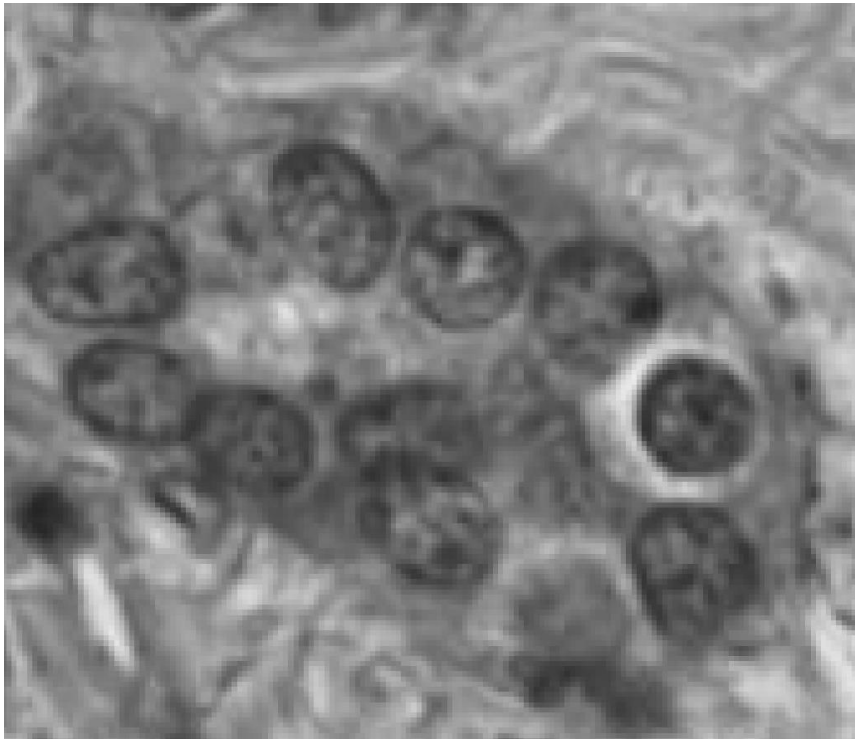
2	3	6
7	1	2
8	6	5

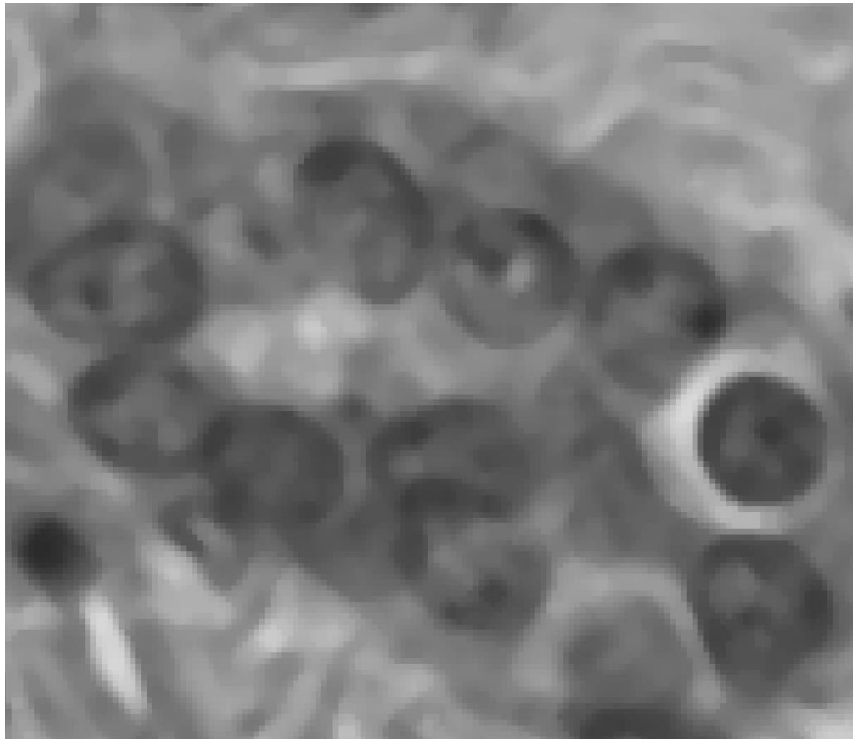
Subtract Left

?	-1	-3
?	6	-1
?	2	1

Can reduce noise by smoothing

- Each pixel is replaced by the mean value around it





2	3	6
7	1	2
8	6	5

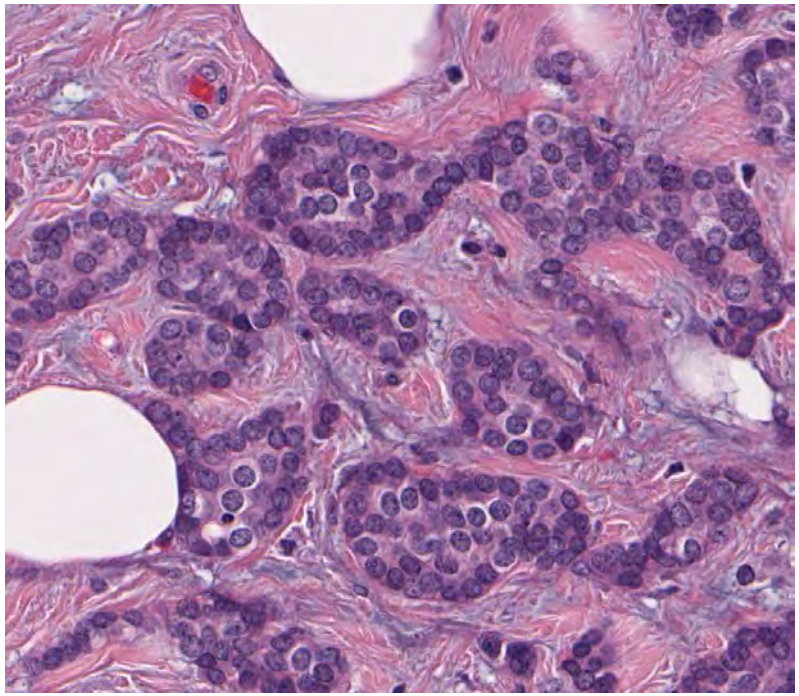
	4.4	

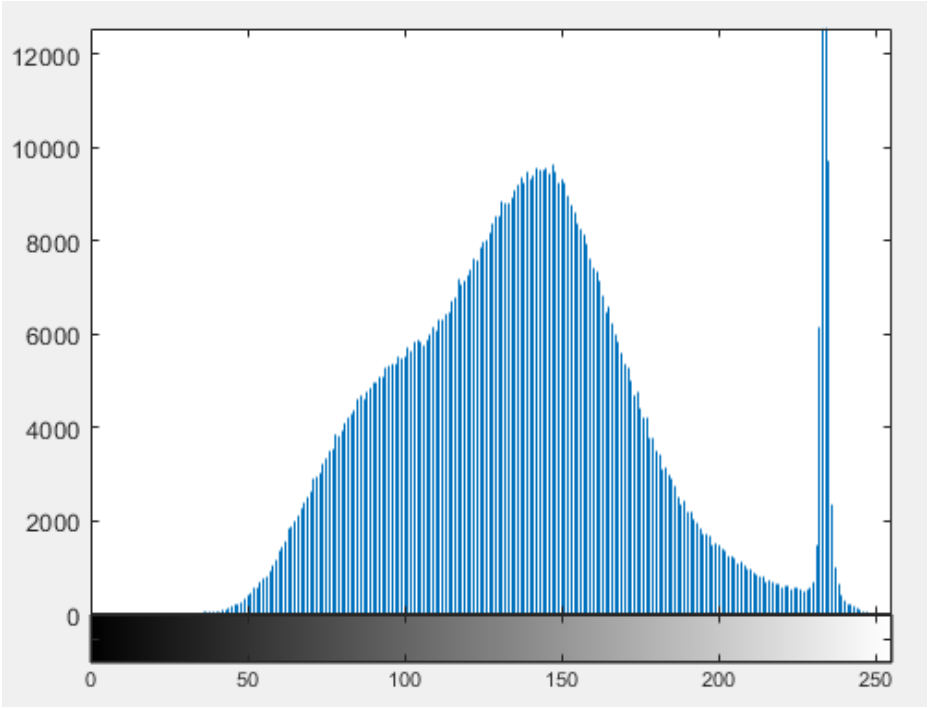
Cente aging and Personalized Diagnostics

CASE WESTERN RESERVE UNIVERSITY

Can look at values as a collection

- Precisely quantify familiar image properties

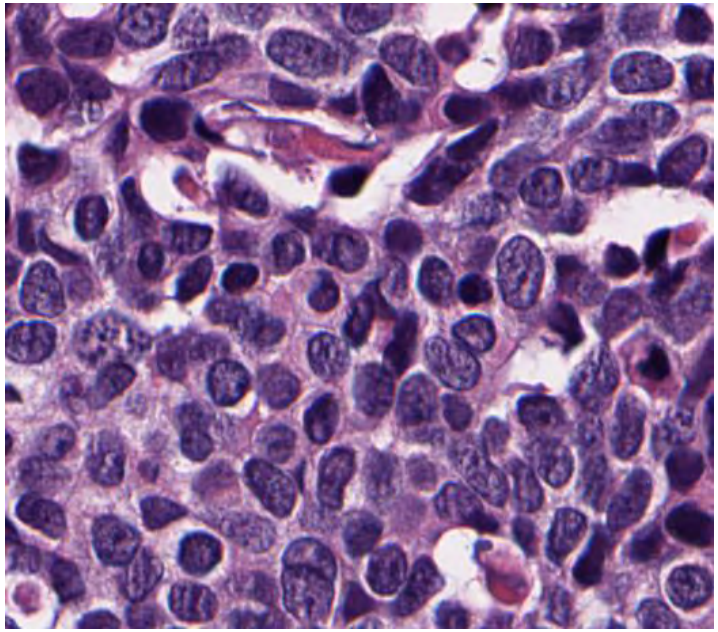


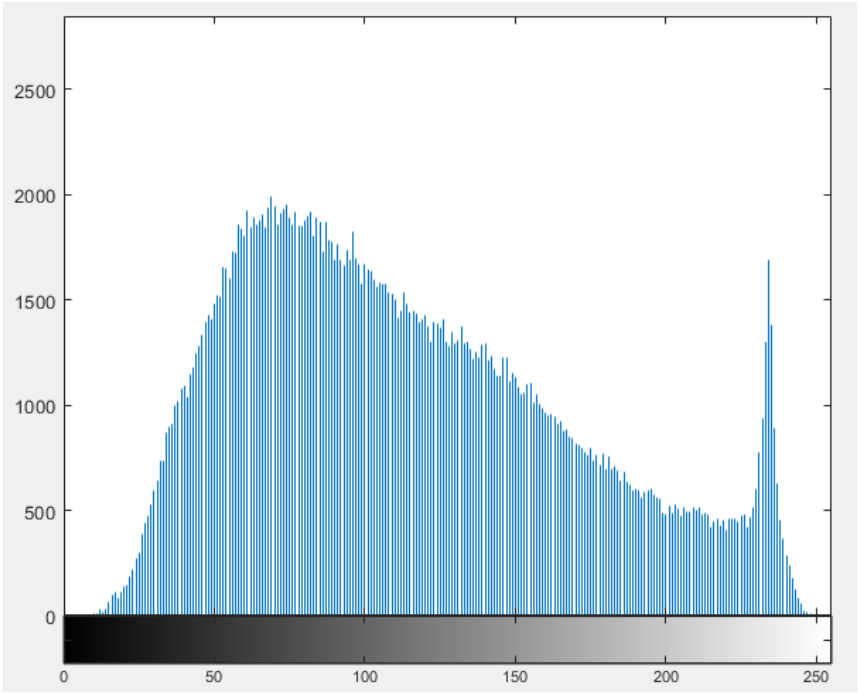


Average Brightness: 140.7963

Can look at values as a collection

- Precisely quantify familiar image properties

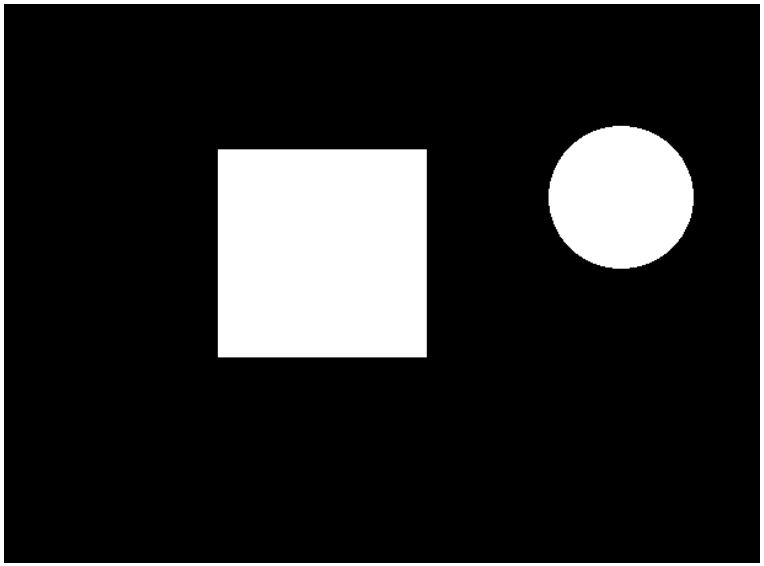


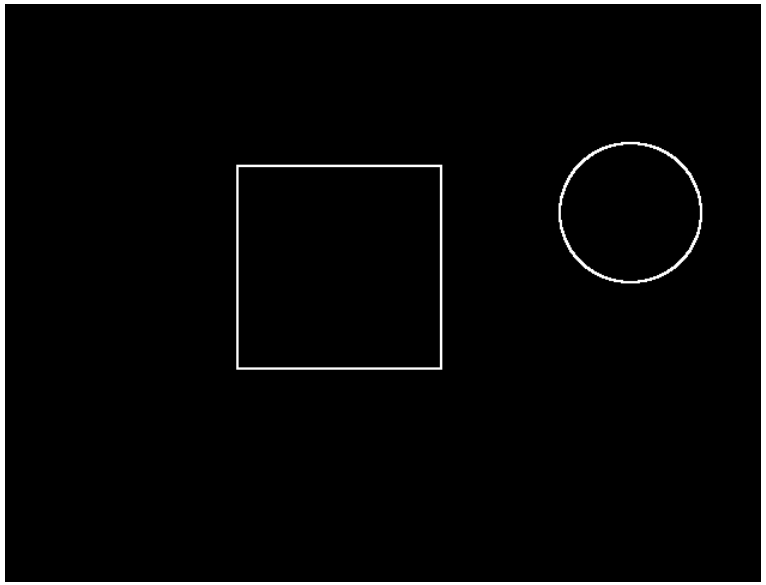


Average Brightness: 114.5927

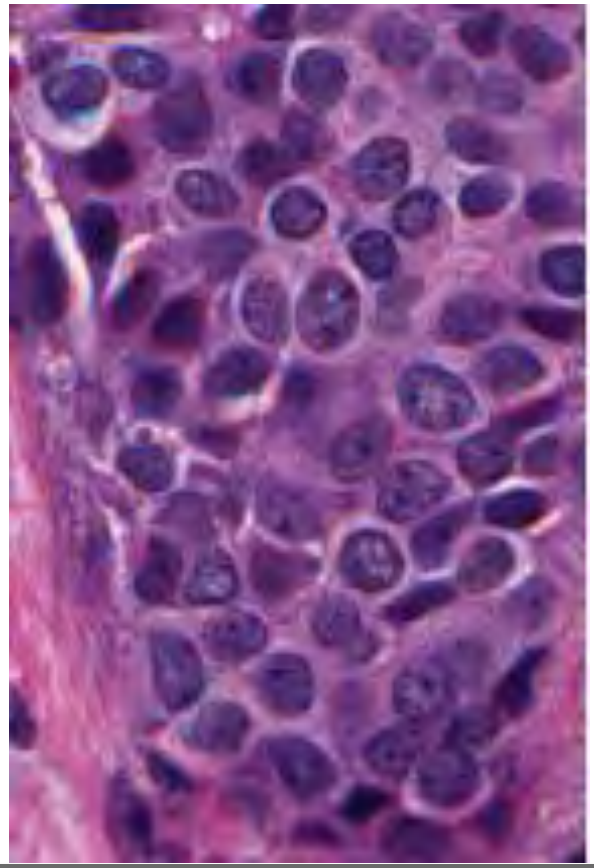
Filtering and Processing using operations

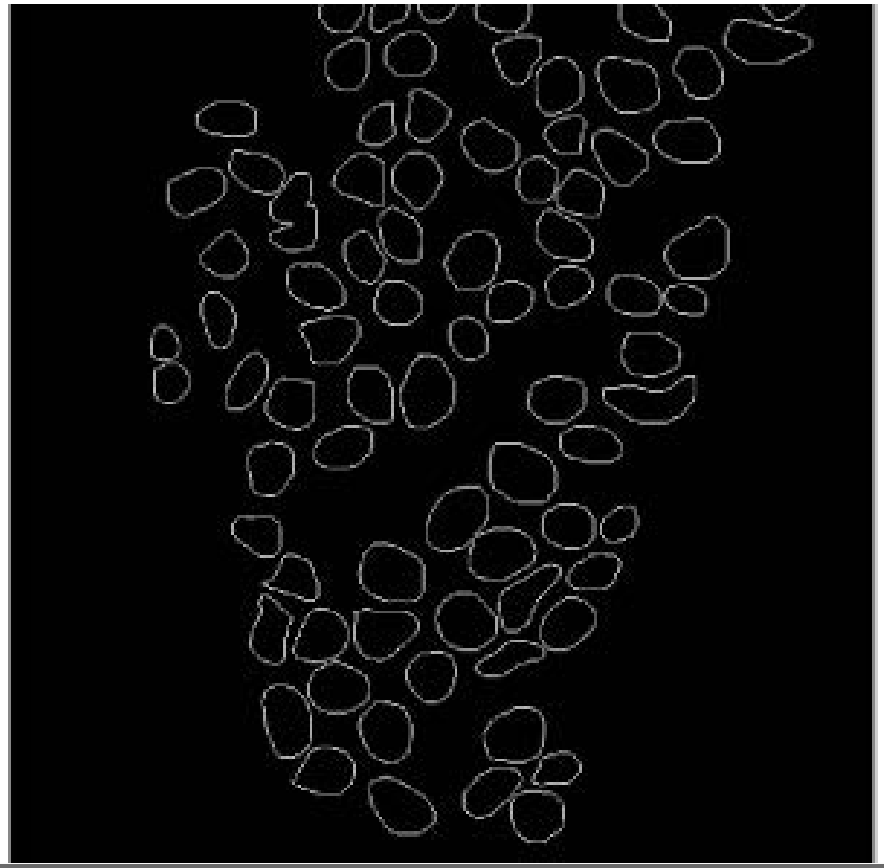
- For example, can identify edges by subtracting adjacent pixels
- 0 – 0 = 0, 1 - 1 =0, 1 - 0 = 1, 0 - 1 = 1





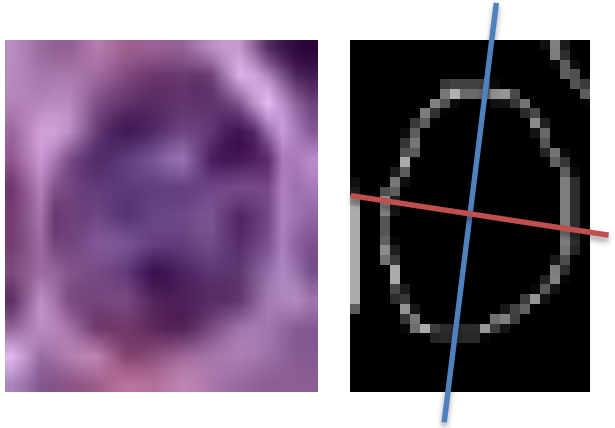
Example edge detection in DP space





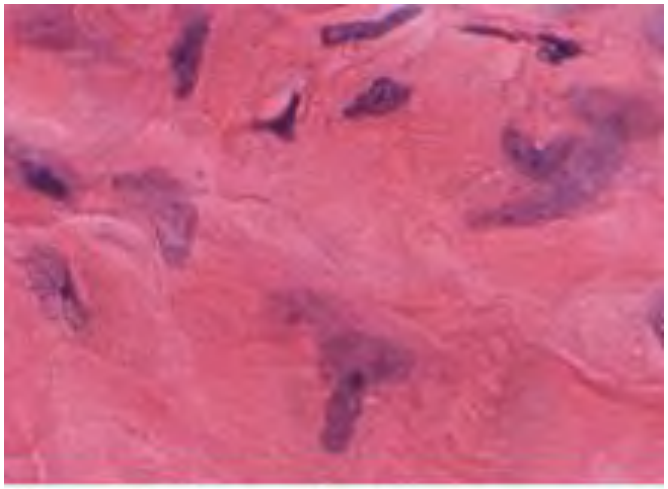
These outputs lead to features

- Domain relevant
- Used in grading schemes
- Properties
  - Size
  - Shape
  - Texture
- More precise quantification as features
  - Eccentricity (how circular)
  - Length major/minor axis
  - Orientation
  - Staining intensity
  - Smoothness
  - Entropy

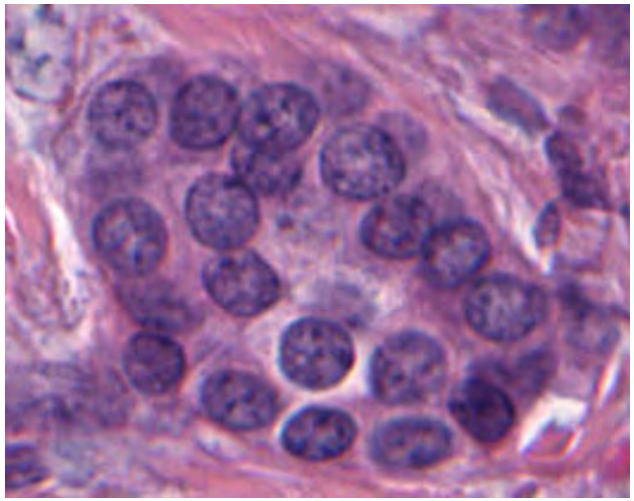


Texture Features

- Stroma region is “smooth”
- Measure of homogeneity
- Neighboring pixels are similar
- Small gradient between them

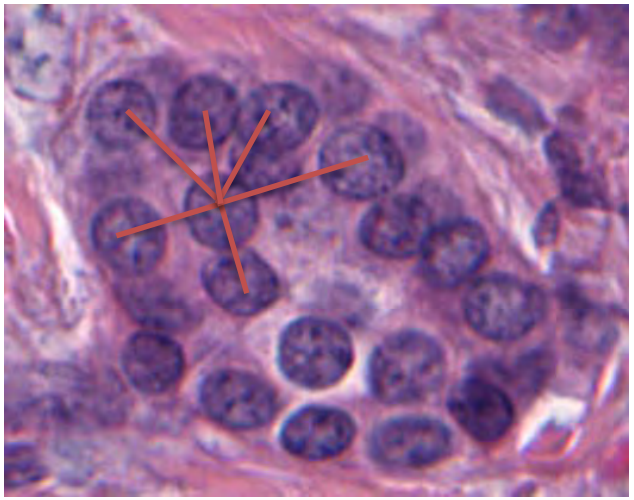


- Other regions more “rough”
- Measure of heterogeneity
- High amounts of entropy
- Larger and unpredictable gradients



Graph Features

- Can think of it in terms of connectivity
- At the object level:
  - How many neighbors do I have in a defined *radius*?
  - Average length away?
- Measurement for infiltration
  - How far am I away from a boundary?
  - How far am I away from a cancer cell?

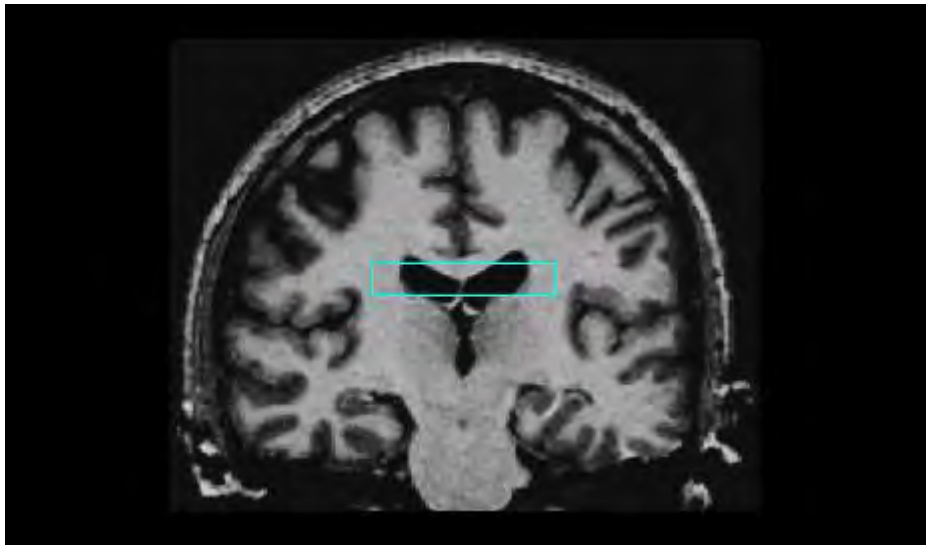


Remarks on features

- Try to avoid throwing the “kitchen sink” at a problem
- Start with a “reasonable” subset
- Based on:
  - Domain expertise
  - Grading schemes

Approaches using features

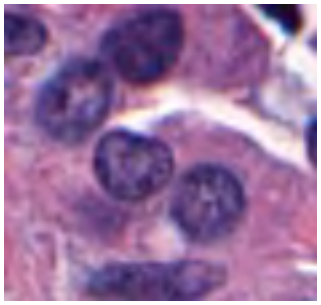
- Active contours
- Keep expanding boundary of an initial box while:
  - Inside is homogenous
  - No edge detected

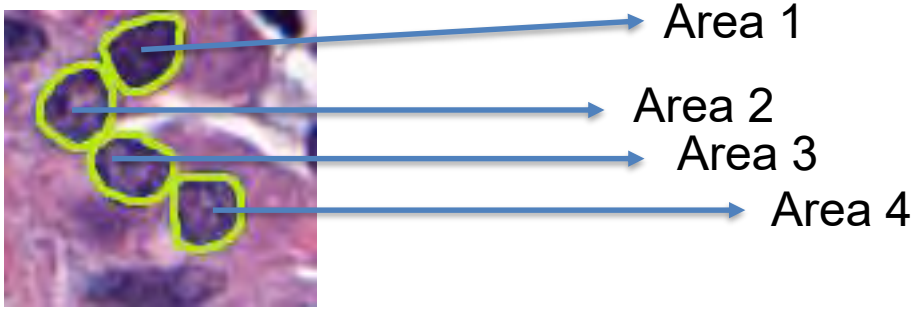


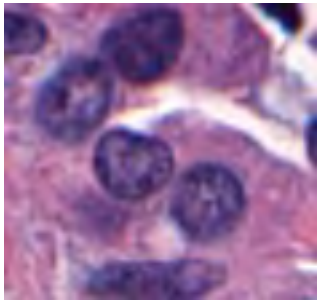
What can we do with features?

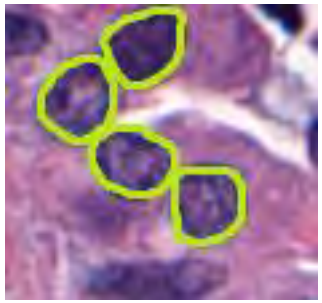
- Measure difference between classes
- Quantify differences:
  - Benign vs Malignant
  - Subtypes
  - Outcome
  - Therapy Response

Feature extraction

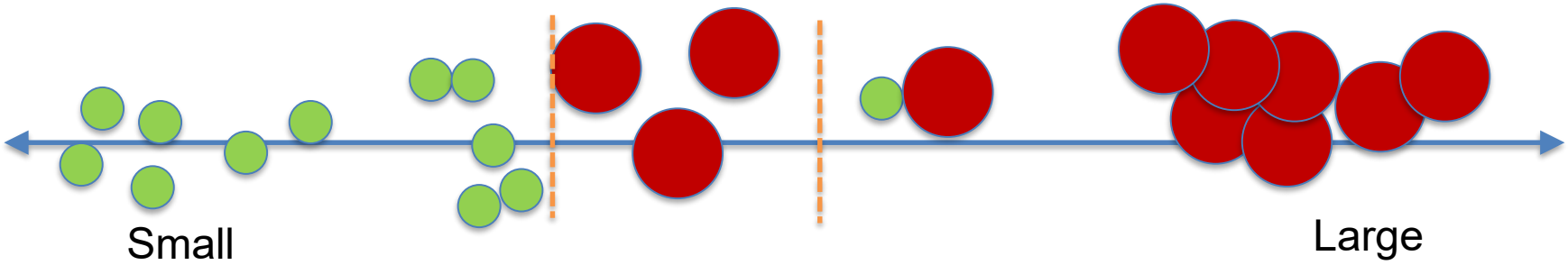








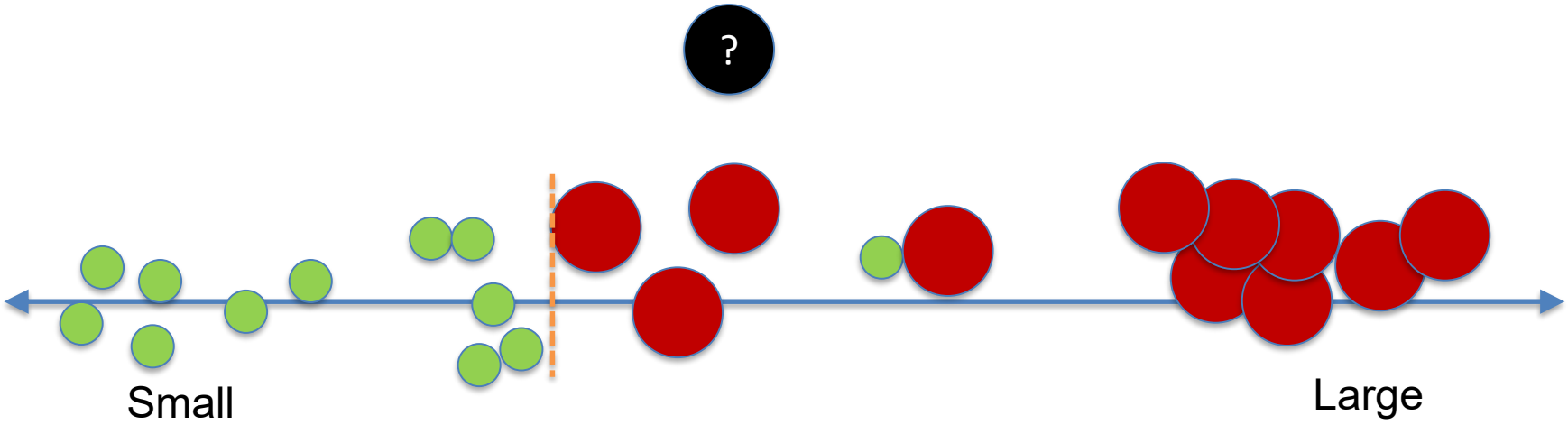
Area

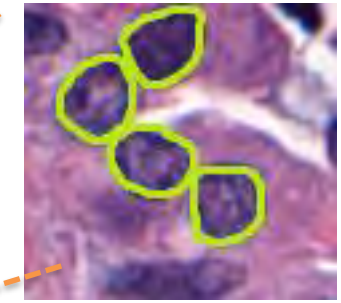


CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Center for Computerized Imaging and Personalized Diagnostics

- Goal of the model
  - Fit training data well
  - Generalize to Testing data
- If identified something biologically relevant and “true”
  - should be consistent
- Can we improve the model by adding dimensions?





??

Circularity

Small

Area

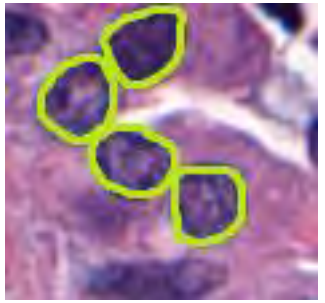
Large

CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Center for Computerized Imaging and Personalized Diagnostics

This problem only gets worse

- The more dimensions, the larger the solution space is
- A lot of noise as well
  - Measurements
  - Labels
- Optimization is hard and time consuming
- Is there anything we can do to help?



Circularity

??

Hopefully doesn't occur "in nature"

Small

Area

Large

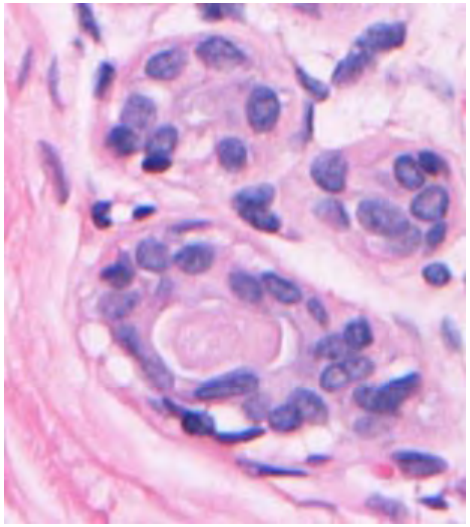
Difference Machine vs Deep Learning?

- Machine Learning:
  - Explicitly provide feature measurements: e.g., Area, Circularity
- Deep Learning:
  - Self-discover the features
- Both approaches require good quality examples

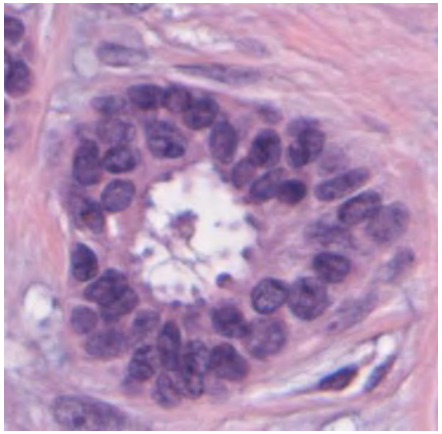
What kind of examples are best?

- Near to decision boundary!
- Information rich
- Cancer vs non-cancer

Not cancer
Not informative



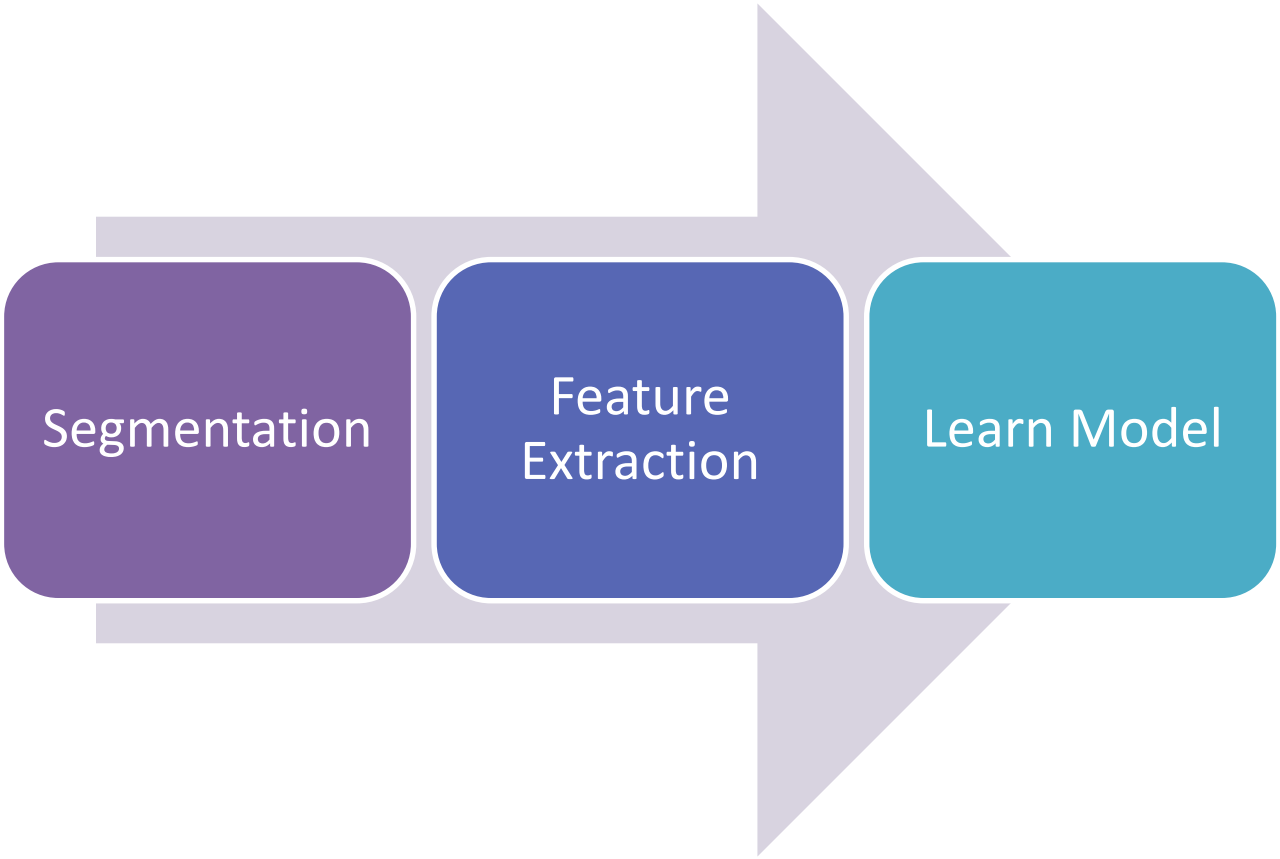
Not cancer
Informative



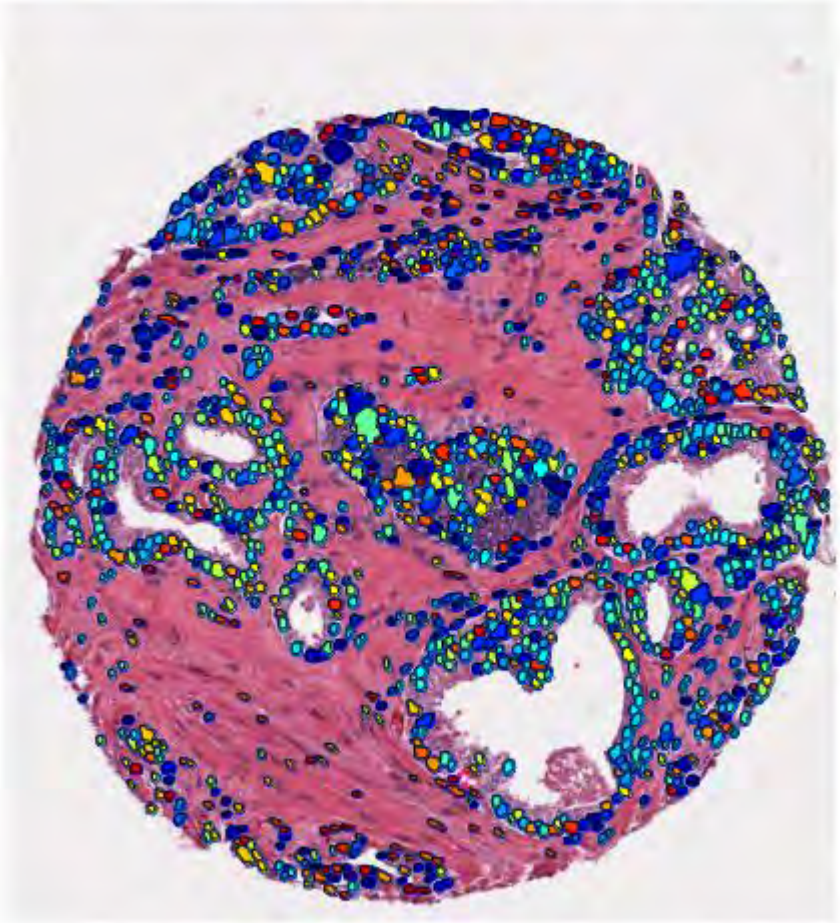
Cancer
Informative

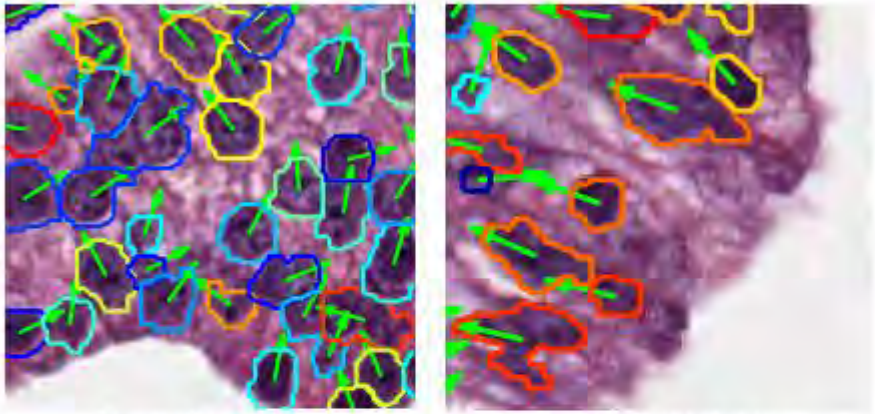
Don’t need 10000s of these!!!

Examine use cases

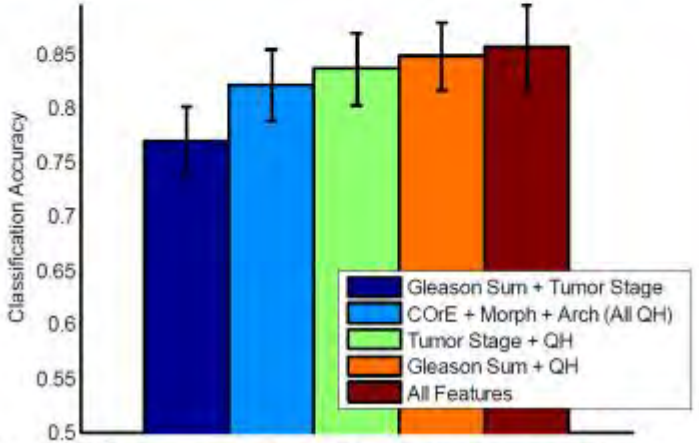


Cell Orientation Entropy (COrE) Features Stratify More and Less Aggressive Prostate Cancer on Tissue Microarrays





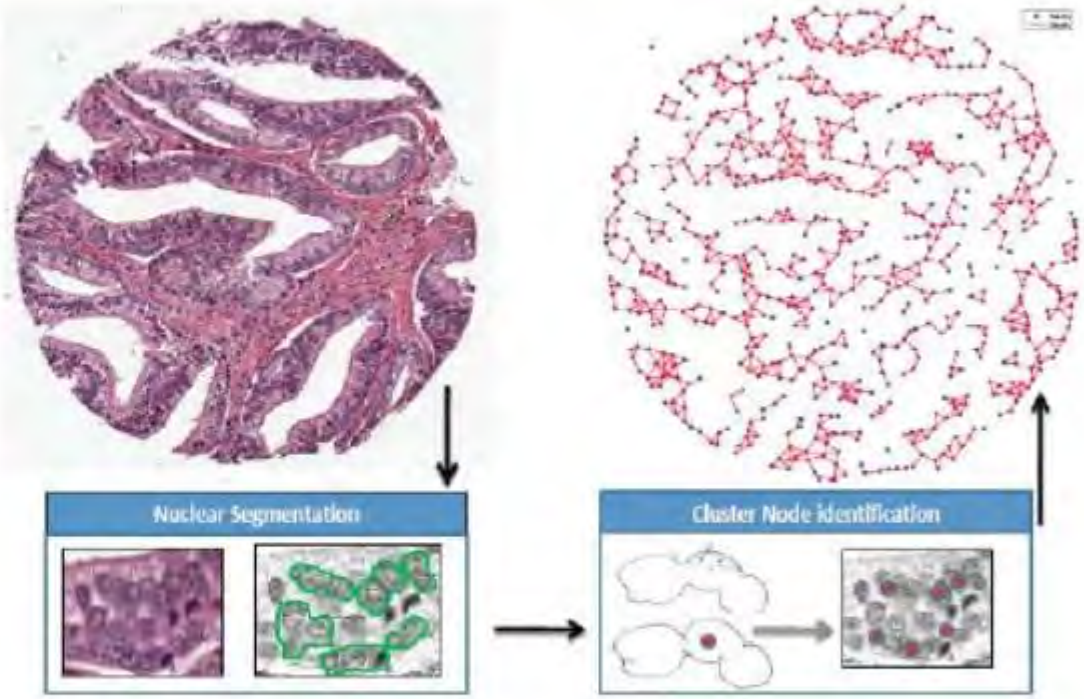
Aggressive cancer (left) shows more disorder in orientation of the nuclei compared to less aggressive cancer (right)



Lee, G, Ali, S, et al., "Cell Orientation Entropy (COrE): Predicting Biochemical Recurrence from Prostate Cancer Tissue Microarrays", In Proc of Medical Image Computing and Computer Assisted Interventions (MICCAI), vol. 3, pp. 396-403, 2013.

Cell Cluster Graph for Prediction of Biochemical Recurrence in Prostate Cancer Patients from Tissue Microarrays

- Novel Cell Cluster graph (CCG) that can quantify tumor morphology
- Extracted features from CCG can predict Biochemical recurrence in Prostate Cancer in 80 patients.

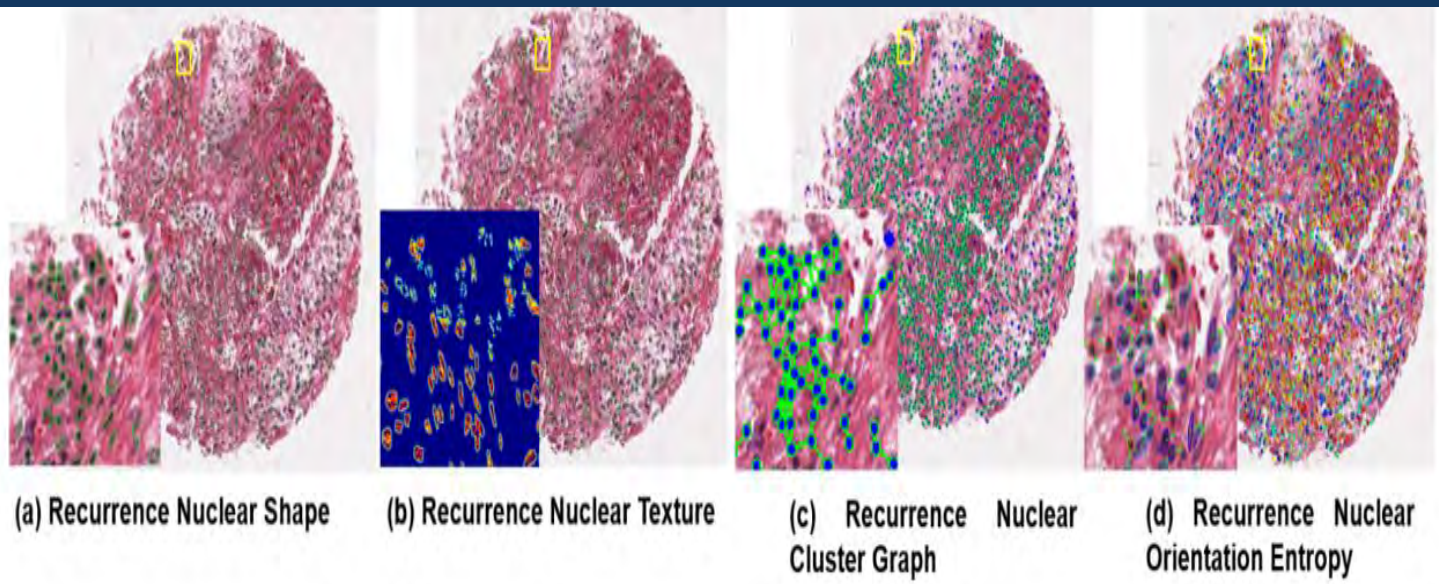


Voronoi	Delaunay	CCG
$67.1 \pm 1.8\%$	$60.7 \pm 0.9\%$	$83.1 \pm 1.2\%$

Table 2. Comparison of CCG against other graph based methods in predicting biochemical failure.

Ali, Sahirzeeshan, Veltri, Robert, Epstein, Jonathan, and Madabhushi, Anant "Cell Cluster Graph for Prediction of Biochemical Recurrence in Prostate Cancer Patients from Tissue Microarrays", SPIE Medical Imaging, 2013.(In Press)

Computerized nuclear features predict recurrence in lung cancer

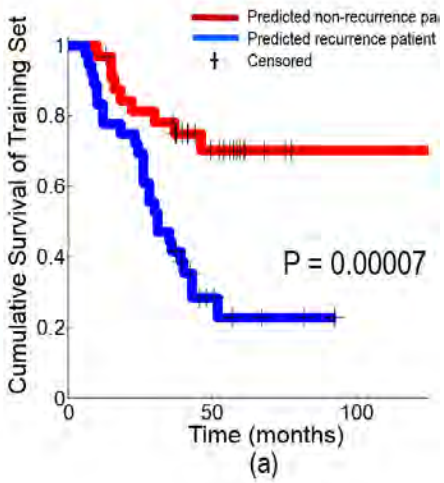


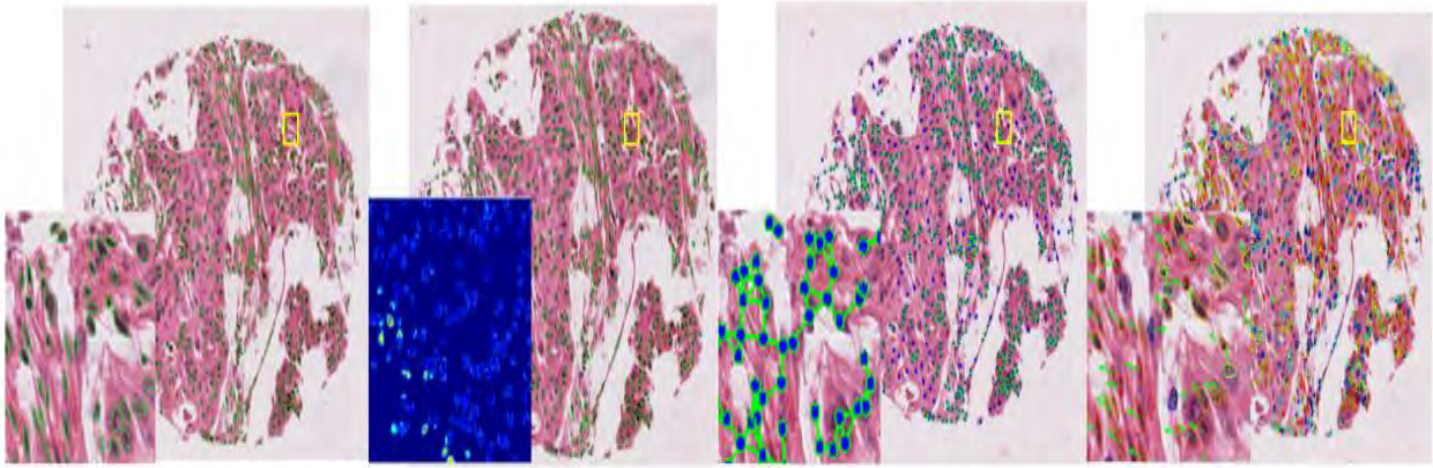
(a) Recurrence Nuclear Shape

(b) Recurrence Nuclear Texture

(c) Recurrence Nuclear Cluster Graph

(d) Recurrence Nuclear Orientation Entropy



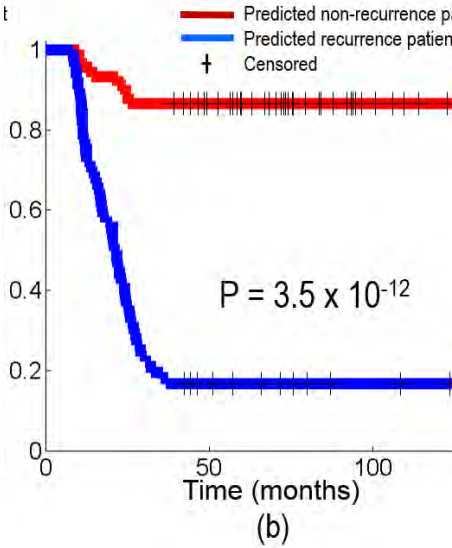


(e) Non-recurrence Nuclear Shape

(f) Non-recurrence Nuclear Texture

(g) Non-recurrence Nuclear Cluster Graph

(h) Non-Recurrence Nuclear Orientation Entropy



Xiangxue Wang, Andrew Janowczyk, Yu Zhou, Sagar Rakshit, Vamsidhar Velcheti, Anant Madabhushi*, “Computer extracted features of nuclear morphology from digital H&E images are predictive of recurrence in stage I and stage II non-small cell lung cancer” Nature Scientific Reports 2017

Spatial arrangement of tumor infiltrating lymphocytes (TILs) predict response to Nivolumab in non-small cell lung cancer (NSCLC)

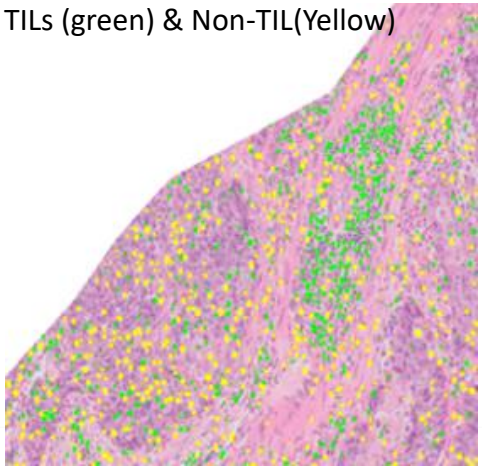
Hypothesis: Spatial arrangement of TILs and local density variance are highly correlated to the patient response.

Data sets:

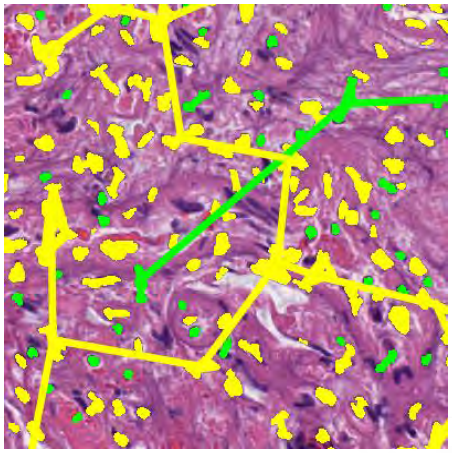
Two independent data (whole slide image) acquired from UPenn (32) and CCF (24)y

TIL detection and image feature extraction

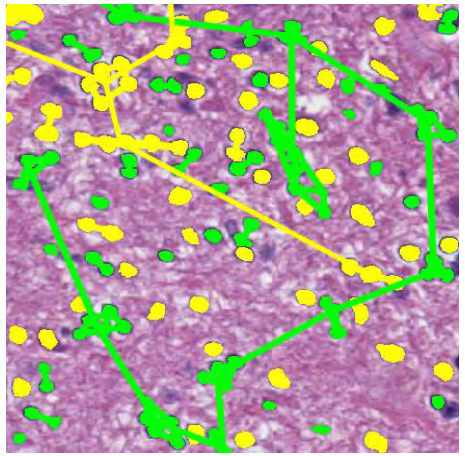
TILs (green) & Non-TIL(Yellow)



Non-Responder

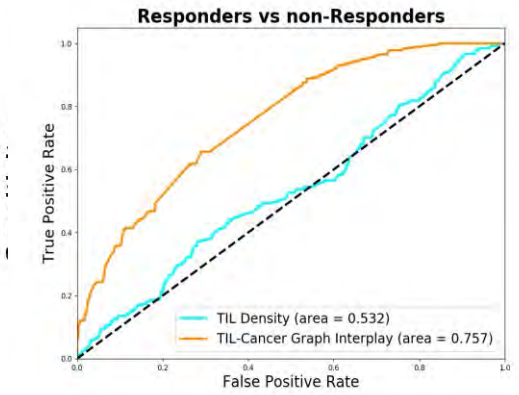


Responder



Top 5 most significant features obtained by feature selection

1. Median of TILs formed areas
2. Ratio of Cancer cells to TILs cells
3. Cancer cell averaged Density
4. Density of TILs
5. Median of Cancer cell formed areas



A QDA classifier was trained using a Training set (n=32) and a independently validation set from a different institution (n=24).

Wang, X, Barrera, C, Velu, P, Bera, K, Prasanna, P, Khunger, M, Khunger, A, Velcheti, V, Madabhushi, A, "Computer extracted features of cancer nuclei from H&E stained tissues of tumor predicts response to Nivolumab in non-small cell lung cancer", American Society for Clinical Oncology (ASCO) Annual Meeting (Poster), Chicago, IL, 2018

Outline

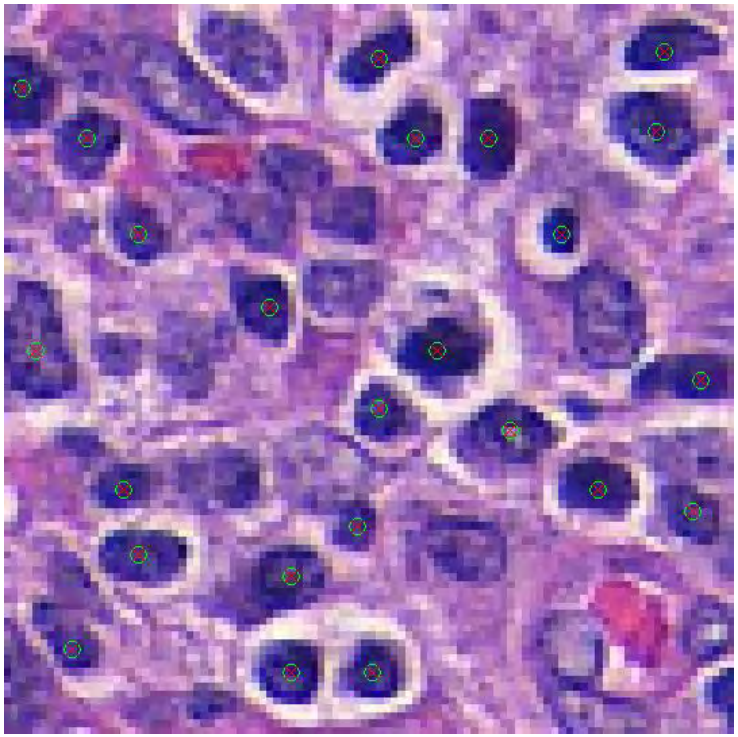
- ~~What are images?~~
- ~~What can be done with them?~~
- ~~Feature extraction~~
- ~~Intuition behind Classifiers~~
- Important considerations
  - Types of annotations
  - Batch effects
  - Quality Control

Types of annotations and tasks

- Detection
- Bounding box
- Segmentation

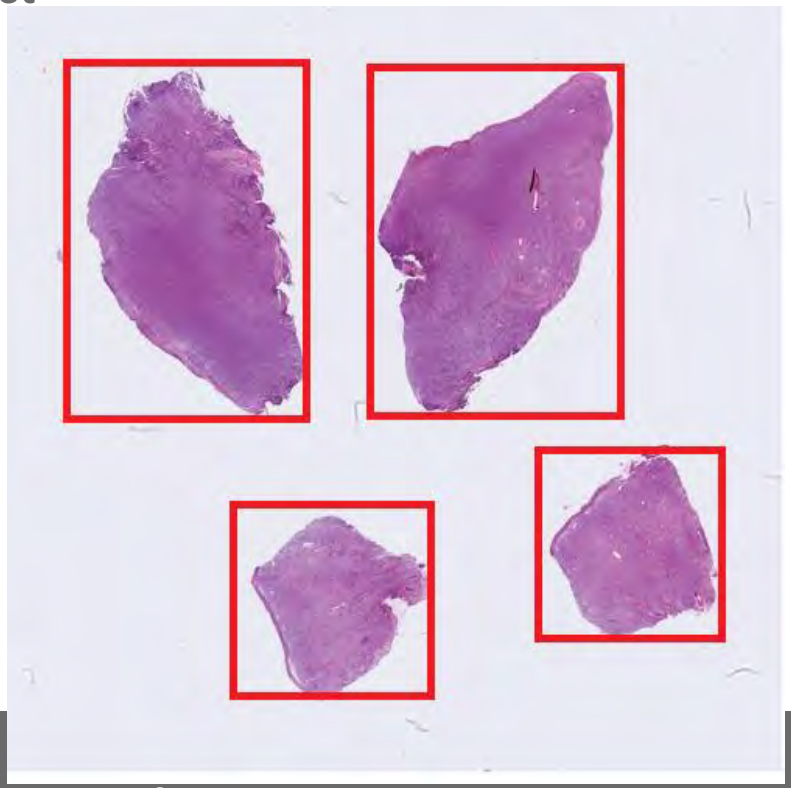
Types of annotations and tasks

- Detection – Where is it?
- Typically place dot in center
- Pros:
  - Fast and easy
  - Easy to score
- Cons:
  - No size information
  - No shape information
- Use cases:
  - Mitosis detection
  - Lymphocyte detection



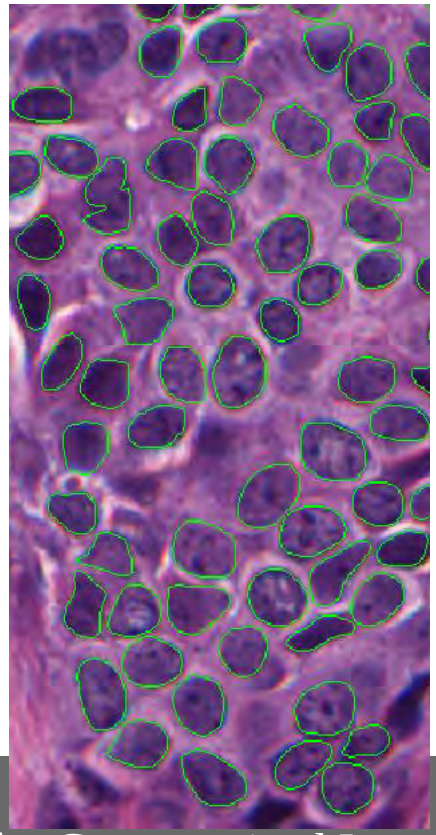
Types of annotations and tasks

- Bounding box – Where and about how big is it?
- Smallest box which will surround object
- Pros:
  - Give size information
  - Faster than segmentation
- Cons:
  - Still no shape information
- Use cases:
  - ROI identification



Types of annotations and tasks

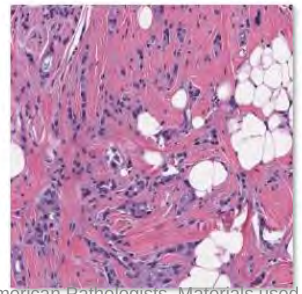
- Segmentation – What are its borders?
- Precisely circle object at a pixel level
- Pros:
  - All types of analysis are possible
  - Morphological analysis
- Cons:
  - Very time consuming
  - Lots of noise: human error + ambiguity
- Use cases:
  - Nuclei segmentation

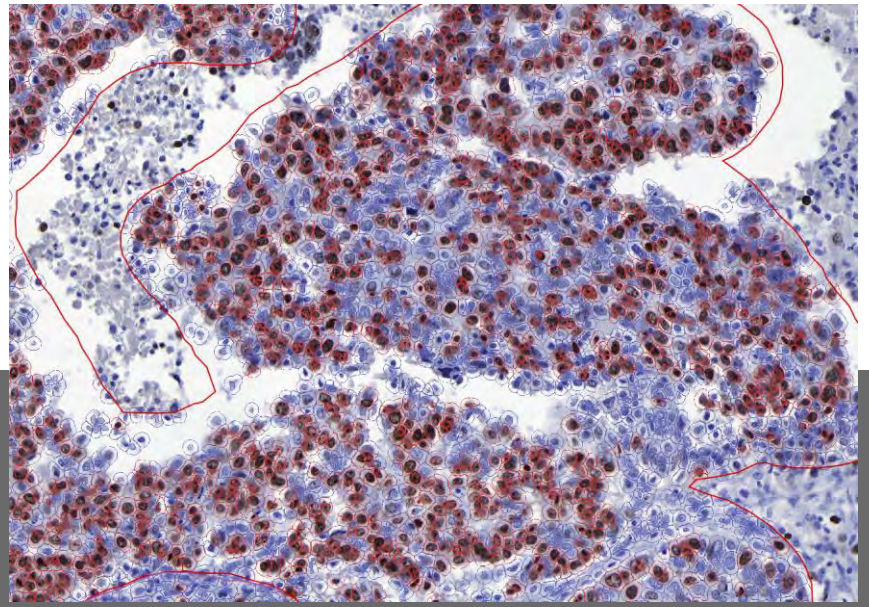


60 minutes to gather annotations, how?

- Many images sparsely annotated
  - High image number, low density
- A few images fully annotated
  - Low image number, high density
- Multiple images with selected ROIs fully annotated
  - Modest image, modest density
- Last approach is the best!
  - More patient diversity
  - More region diversity
  - Likely to be “different” and informative

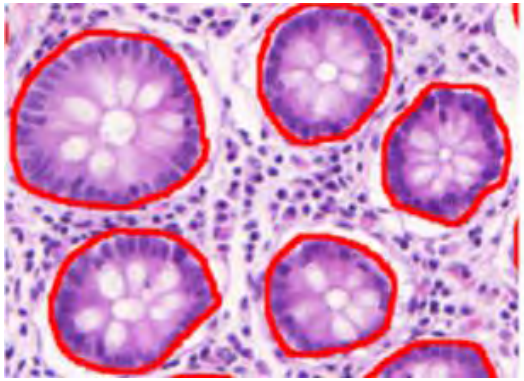

8865_500_f00003_mask.png

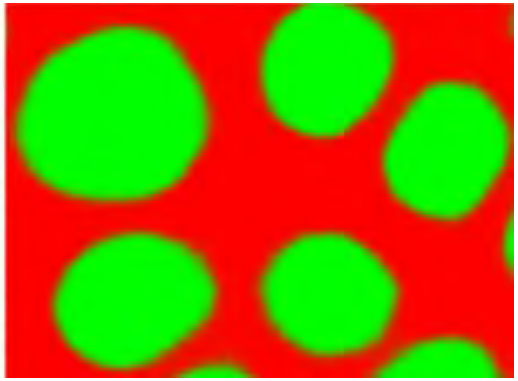

8865_500_f00003_original.tif



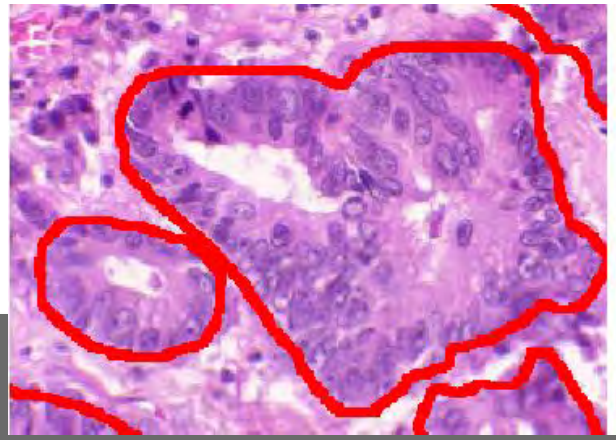
How to select the ROIs?

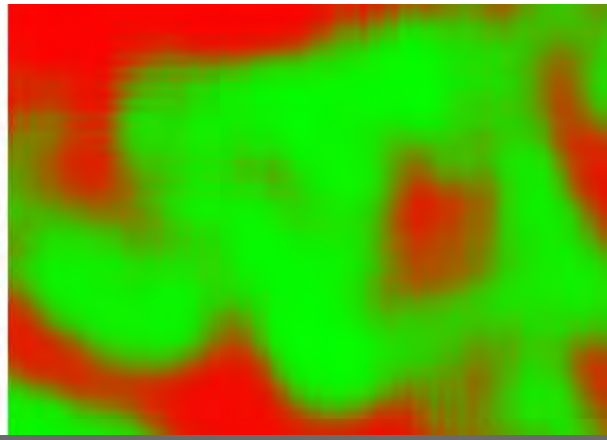
- Exploit domain knowledge
  - I know nuclei + lymphocytes appear similarly, try to find ROIs which have both present to challenge the classifier
- False positive sampling
  - Train a model
  - Use it on training data
  - See where errors occur
  - Hyper-sample those types of regions





- Ultimately, the classifier can tell *you* where its struggling by displaying poor performance!
- Target those types
  - Similar to teaching a student



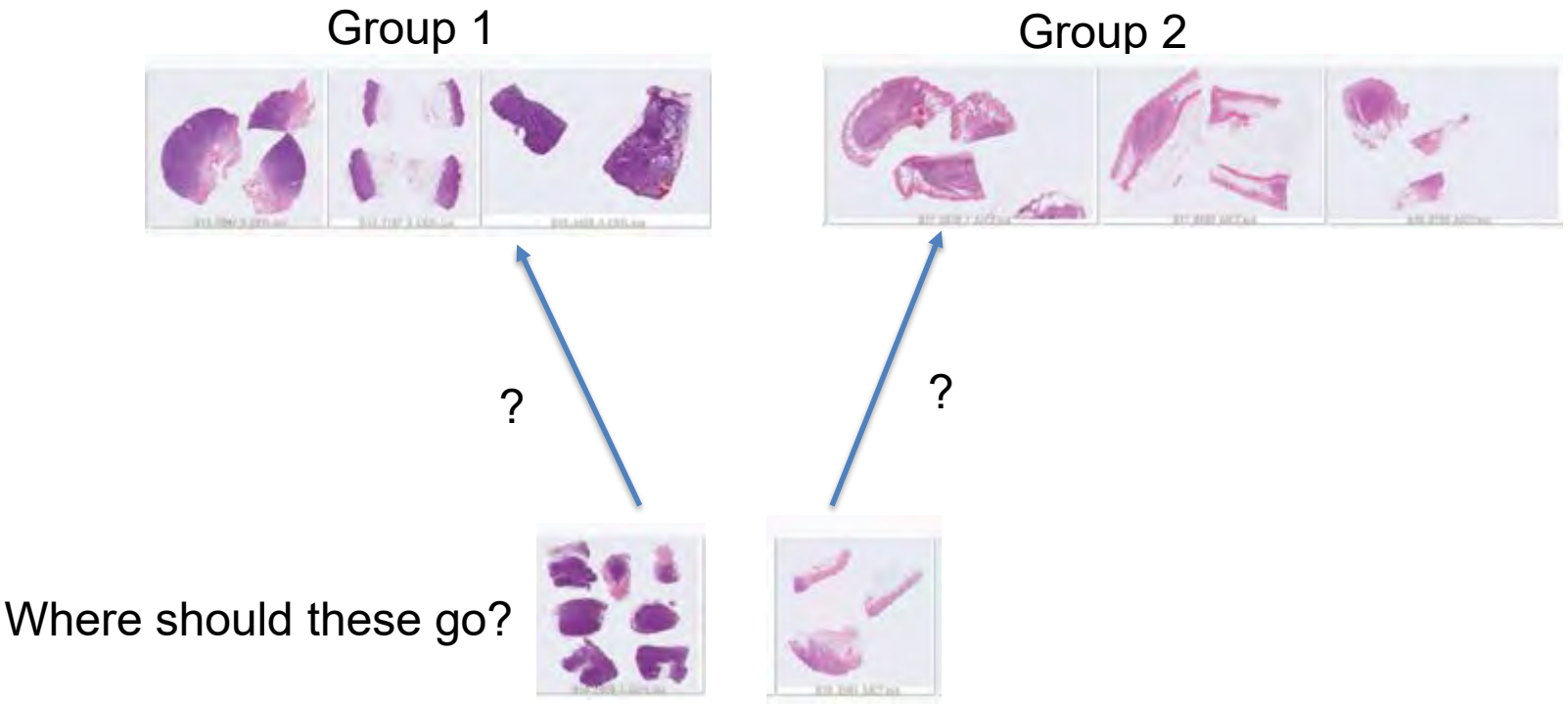


Outline

- ~~What are images?~~
- ~~What can be done with them?~~
- ~~Feature extraction~~
- ~~Intuition behind Classifiers~~
- Important considerations
  - ~~Types of annotations~~
  - Batch effects
  - Quality Control

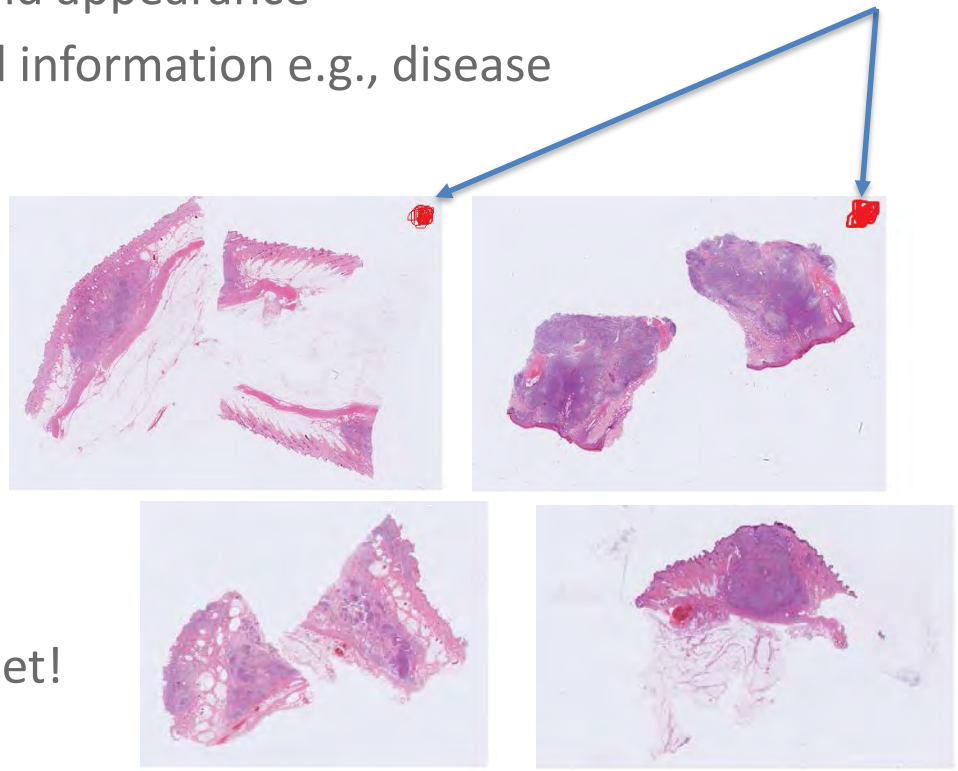
Batch Effects

- Confounding of non-biological signal with biological signal



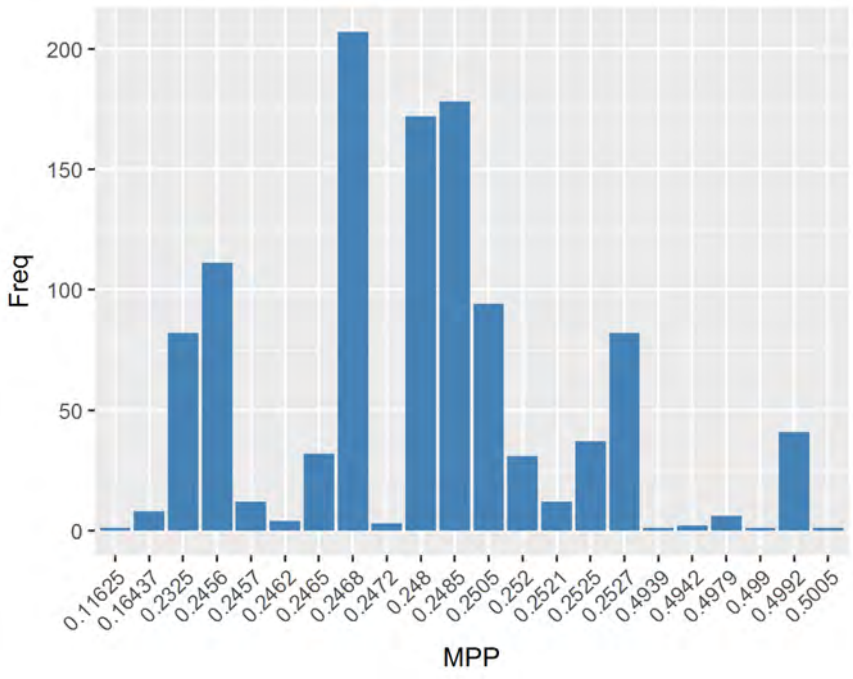
Batch Effects 2

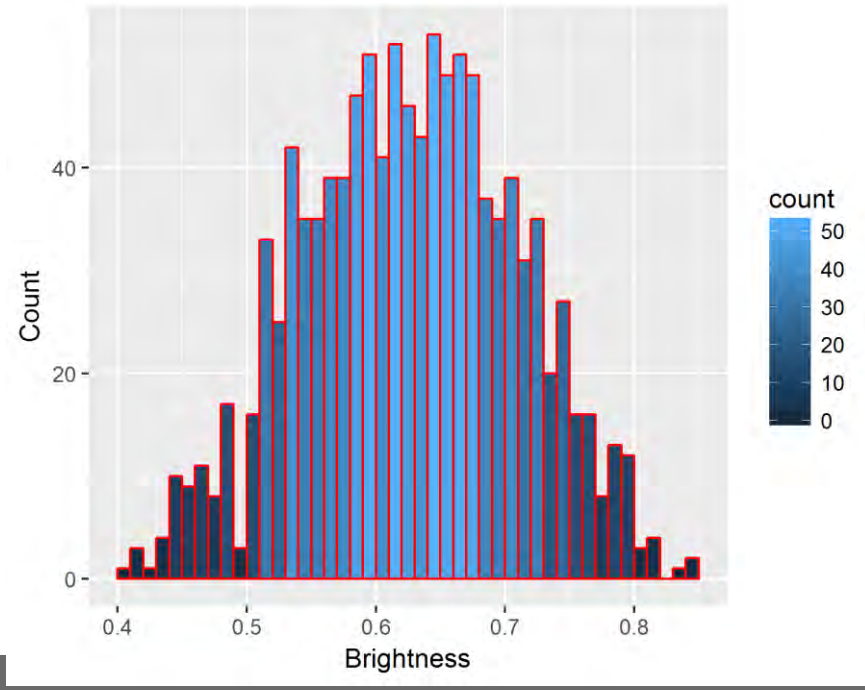
- You've done this grouping by color and appearance
- Not taken into account any biological information e.g., disease presentation
- Examine in practice
- Pathologist marked slide with dot
- Identify/scan (rare) samples
- Add in "undotted" samples
- Classifier learned to focus on dot
- Great performance on test set!
- Very poor performance on external set!



How would Batch Effects present in DP

- Pre-scanning: stain intensity, thickness
- Scanning: brightness, compression, microns per pixel



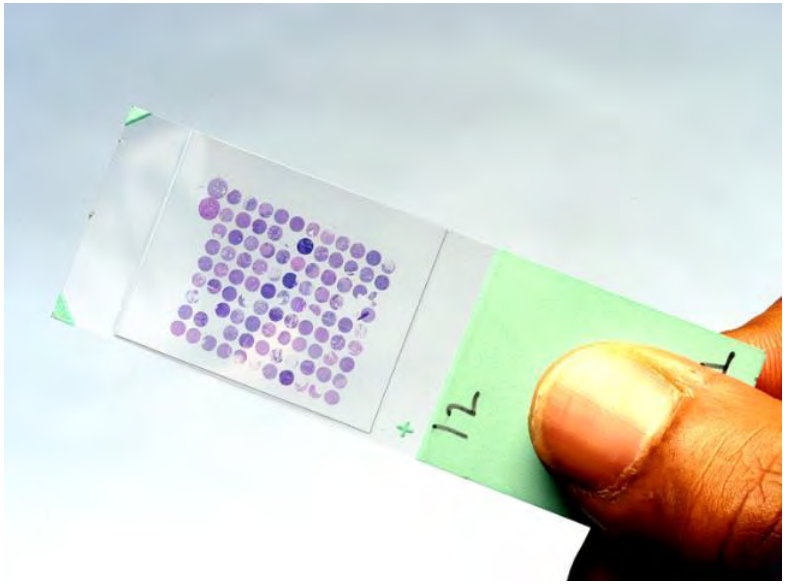


CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Center for Computerized Imaging and Personalized Diagnostics

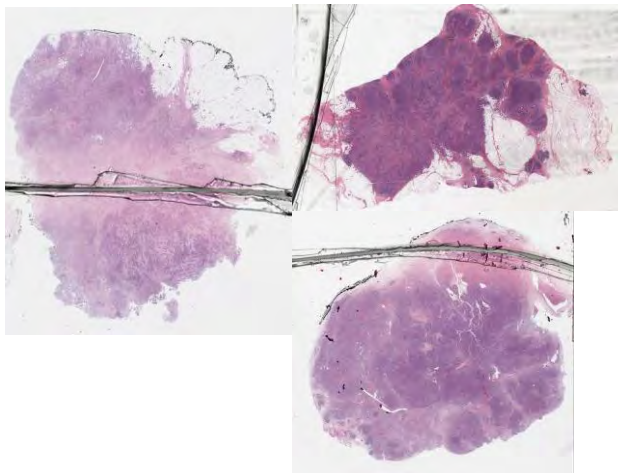
Potentially Likely (Worst) Situation

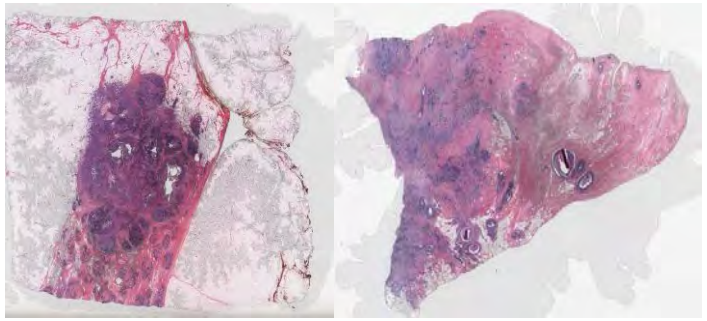
- TMA created containing only high-risk patients
- TMA created containing only low-risk patients
- *Any* artifact will perfectly separate groups

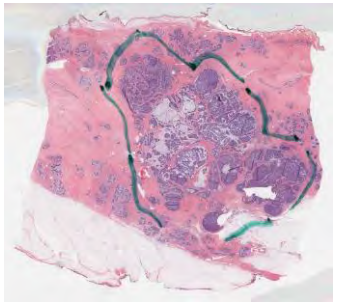


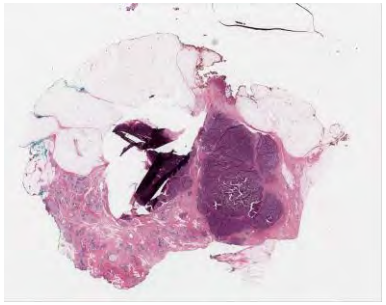
Unmet Need For Quality Control

- Transition to digital pathology workflows
  - Digital Quality Control is paramount
  - Recut and rescan slides immediately before getting to a pathologist
  - Cost and efficiency savings
- Previously not insurmountable
  - Increasingly too time consuming to do manually
  - Non-reproducible









Slides taken from diagnostic cohort of TCGA-BRCA

We need better quality control of our slides!

What is HistoQC?

- Open source reproducible slide quality metrics with artifact localization
- Python backend
  - identify artifacts and produce binary masks of “good” tissue
  - compute *actionable* quality scores and metrics
- HTML5 front end for visualizing and investigating results
- Received innovation award at European Congress on Digital Pathology 2018
- JCO CCI April 2019, Available: http://github.com/choosehappy/HistoQC

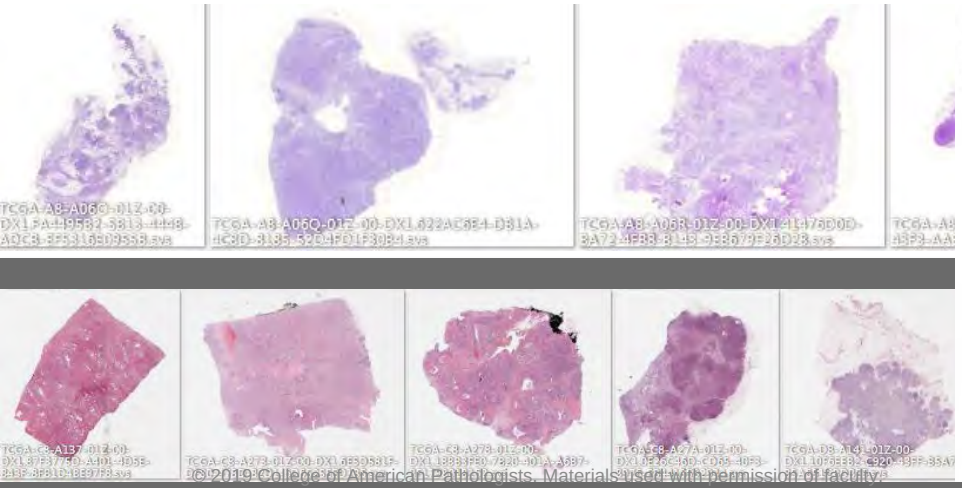
config.ini | MorphologyModule.py | qc_pipeline.py | colorconv.py | Base

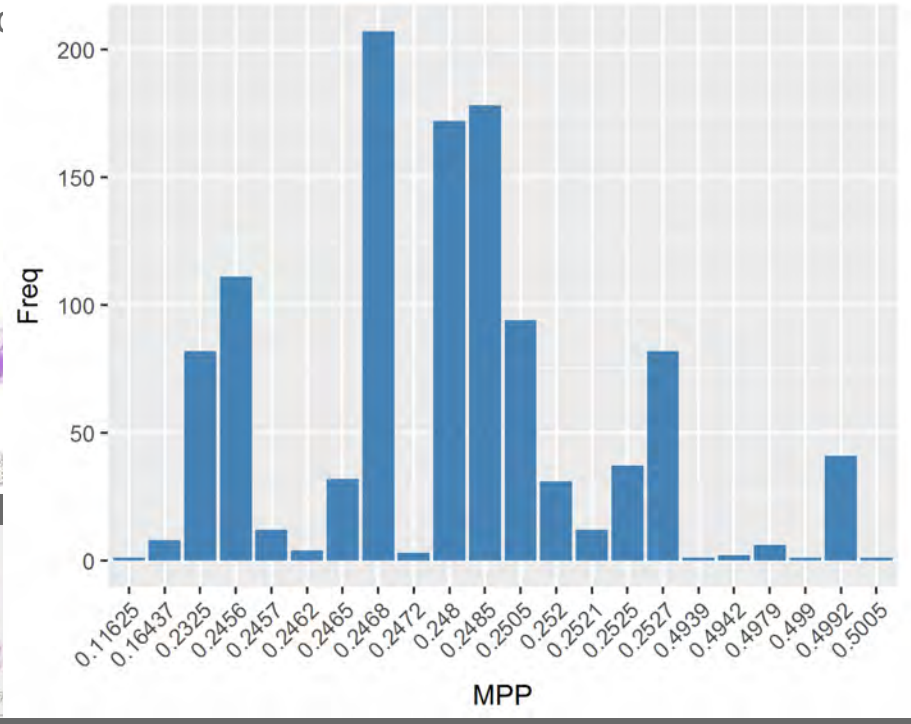
```
[pipeline]
steps= BasicModule.getBasicStats
    BasicModule.getMag
    ClassificationModule.byExampleWithFeatures:pen_markings
    #ClassificationModule.byExampleWithFeatures:pen_markings_red
    ClassificationModule.byExampleWithFeatures:coverslip_edge
    #LightDarkModule.getIntensityThresholdPercent:bubble
    LightDarkModule.getIntensityThresholdPercent:tissue
    #BubbleRegionByRegion.pixelWise
    LightDarkModule.getIntensityThresholdPercent:darktissue
    MorphologyModule.removeSmallObjects
    MorphologyModule.fillSmallHoles
    BlurDetectionModule.identifyBlurryRegions
    BasicModule.finalProcessingSpur
    BasicModule.finalProcessingArea
    HistogramModule.compareToTemplates
    HistogramModule.getHistogram
    BrightContrastModule.getContrast
    BrightContrastModule.getBrightnessGray
    BrightContrastModule.getBrightnessByChannelinColorSpace:RGB
    BrightContrastModule.getBrightnessByChannelinColorSpace:YUV
    DeconvolutionModule.seperateStains
    SaveModule.saveFinalMask
    SaveModule.saveThumbnails
    Basi
```

Center for Computerized Imaging and Personalized Diagnostics

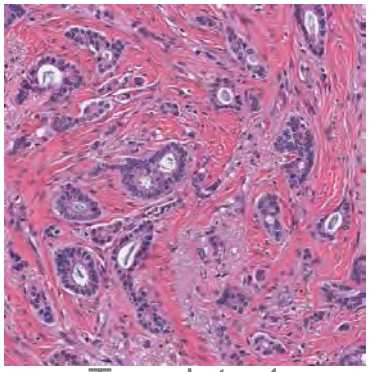
End Users of HistoQC

- Pathology Departments
  - Real-time rolling average of metrics
  - Identify issues early
- Repositories + Computational pathologists
  - Identify and avoid artifacts and outliers for better datasets
    - Stain variances
    - Micron per pixel (MPP) heterogeneity
    - Batch effect presence
  - Explicitly define acceptable tolerance

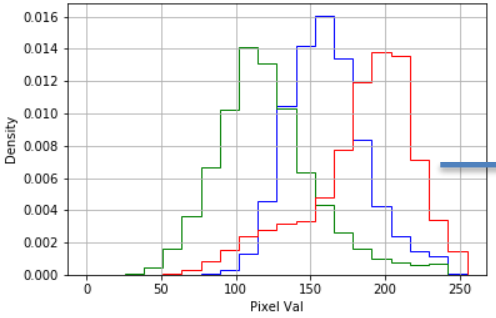


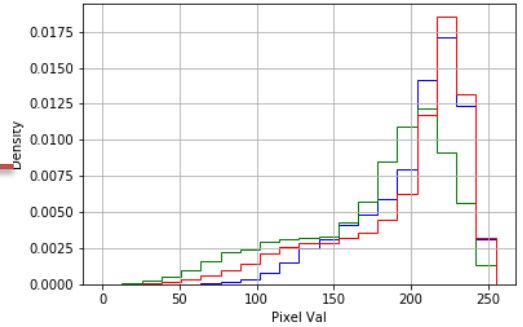


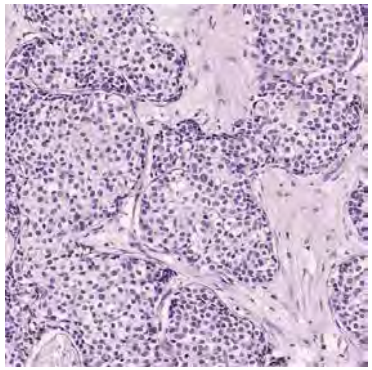
Example Feature: Template Matching



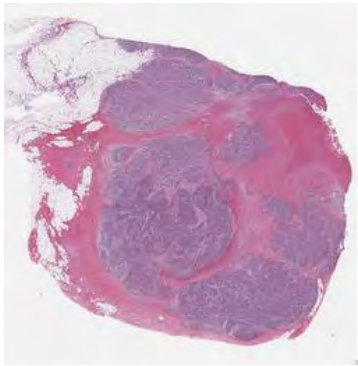
Template 1

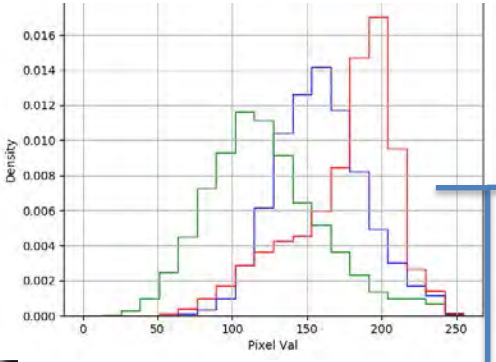


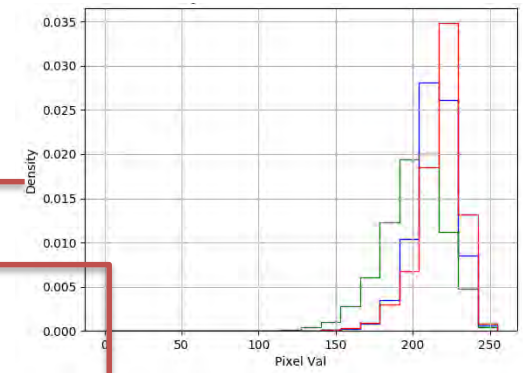


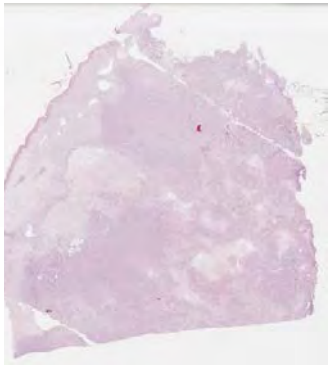


Template 2









	Template 1	Template 2
Left Slide	0.00009	0.00212
Right Slide	0.00485	0.00092

Is slide within tolerances?
Will my algorithm work?

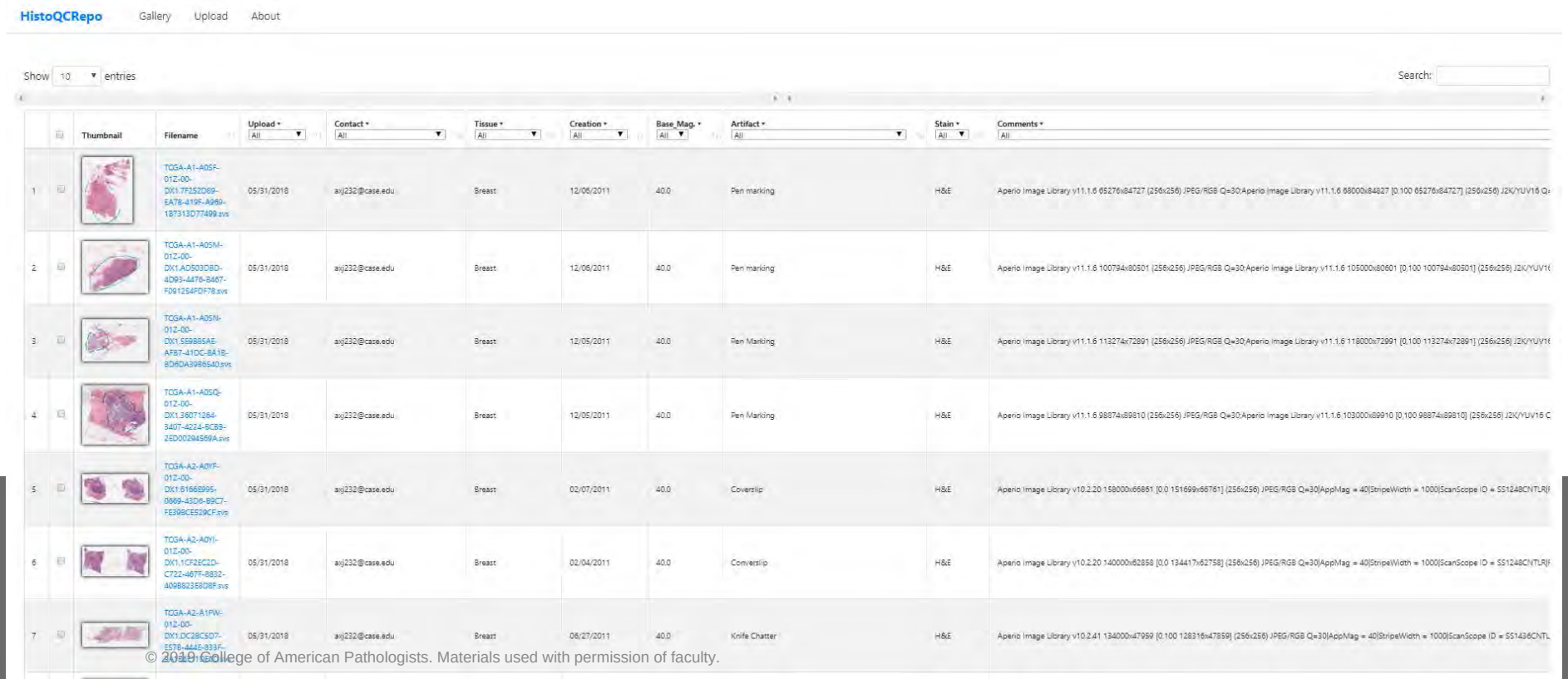
- Precisely specify ranges

CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Center for Computerized Imaging and Personalized Diagnostics

Quality Control Slide Repository

- Created website to host the “greatest hits”
- Slide and associated metadata (e.g., artifact type)
- Useful as a didactic tool for new pathologists
- Benchmark algorithms
  - Detecting artifacts
  - Measure algorithm robustness to artifacts
- Currently available: http://www.histoqcrepo.com



Final Call To Action - Download HistoQC

- Try on your data
- Submit pull requests for new modules

HistoQC: reproducible slide quality metrics with artifact localization

github.com/choosehappy/HistoQC

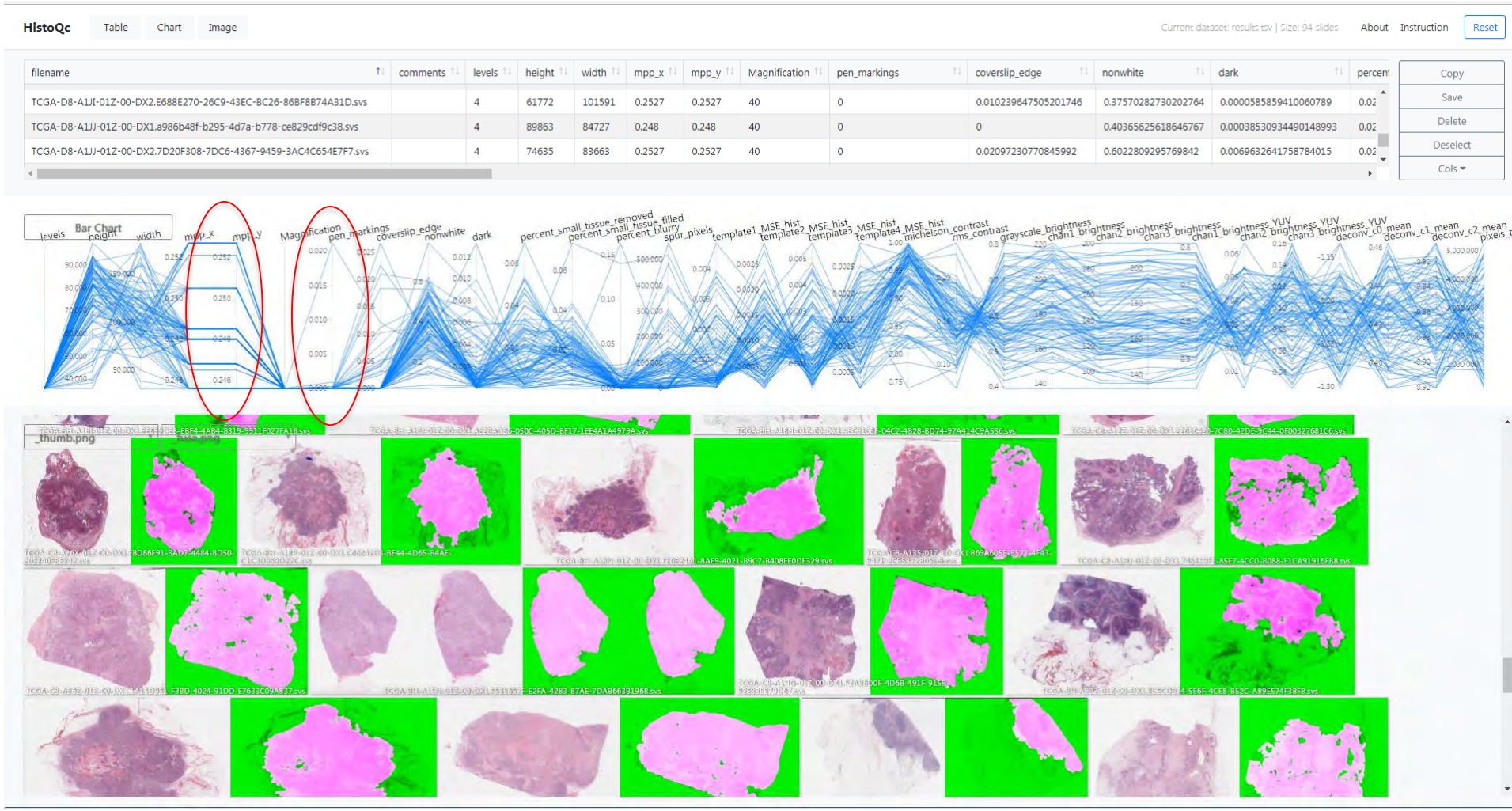
- Upload/download artifact containing slides to/from repository:

HistoQCRepo.com

Thank you!

andrew.janowczyk@case.edu

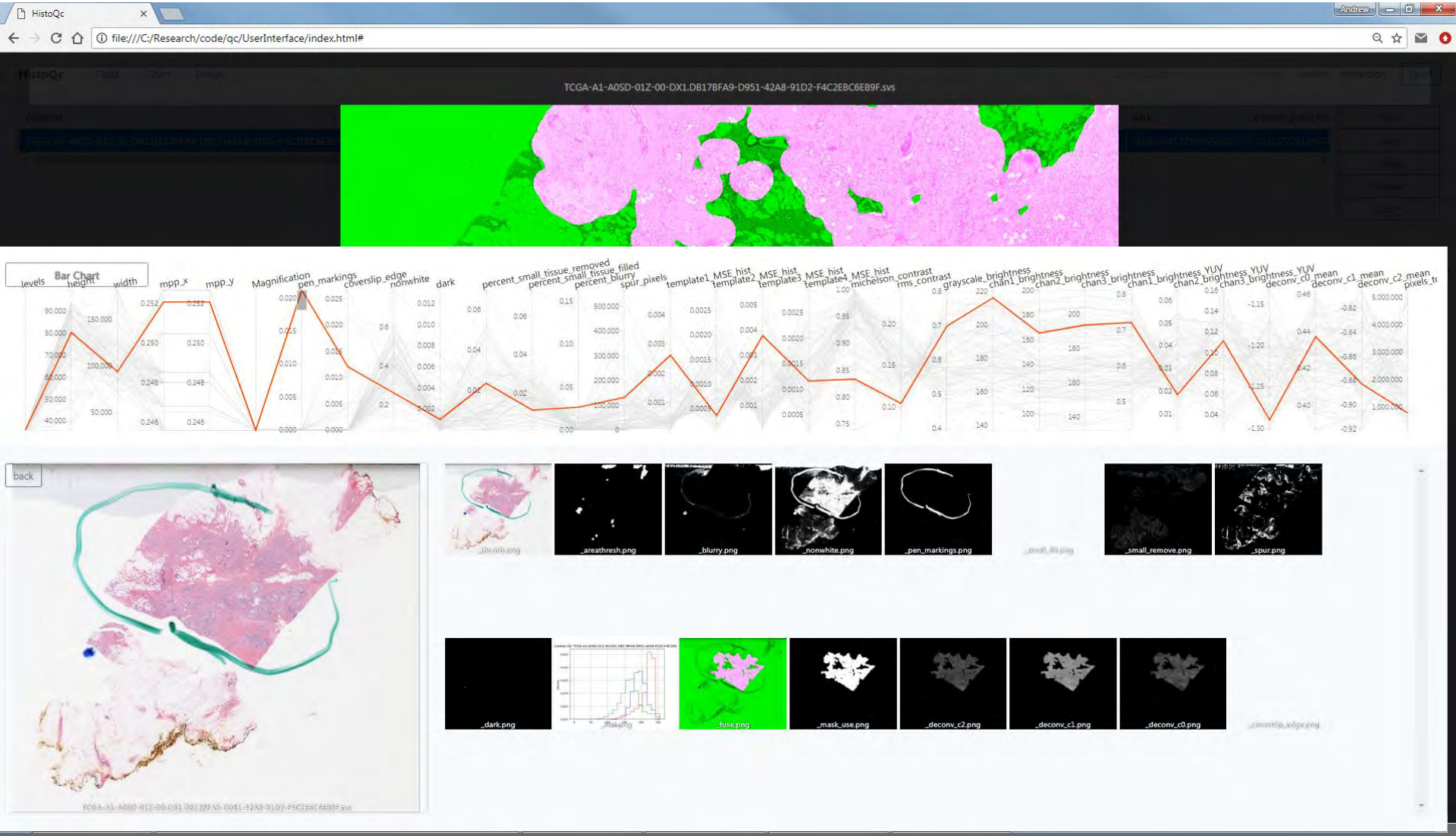
andrewjanowczyk.com



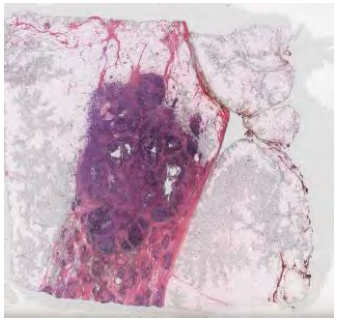
CASE SCHOOL OF ENGINEERING
CASE WESTERN RESERVE UNIVERSITY

Center for Computerized Imaging and Personalized Diagnostics

Visualizing Individual Results



What to do with outliers?



- Make a decision
  1. Remove entirely from dataset is image is really bad
  2. If rest of the image is okay, make sure to avoid that bad region
  3. Can (and should) use HistoQC output mask to select regions to sample from!