*Midwest Consortium for*
*Computational Pathology Workshop*
*January 26th, 2021*

# Computational Pathology and Improving Predictive Analytics

**Michael J. Becich, MD PhD - becich@pitt.edu**
**Chairman and Distinguished University Professor, Department of Biomedical Informatics**
**http://www.dbmi.pitt.edu**
**University of Pittsburgh School of Medicine**

# Disclosures of COI for 2021 for MJB

Startup/Public Companies Royalties/Licensing, Equity):

- De-ID Data Corp – de-identification software (licensing agreement) http://www.de-idata.com/
- Nexi – Newco by Rebecca Jacobson/TIES/TCRN team (royalties to my Department) to support pharma/biotech – http://www.nexihub.com
- SpIntellx (formerly SpDX) – Spatial Intelligence for Cancer Diagnostics (founder equity)

Consultancy (honoraria)

- Cancer Center Consulting – Baylor, CINJ/Rutgers, U Colorado, U NM
- CTSA Consulting – MCW, Northwestern, UC Davis, U Chicago, U IN, UC Davis, U NM, U WI
- Biomedical Informatics Consulting – Northwestern, Rockefeller, UC Davis, U Chicago, U FL

Federal Grants

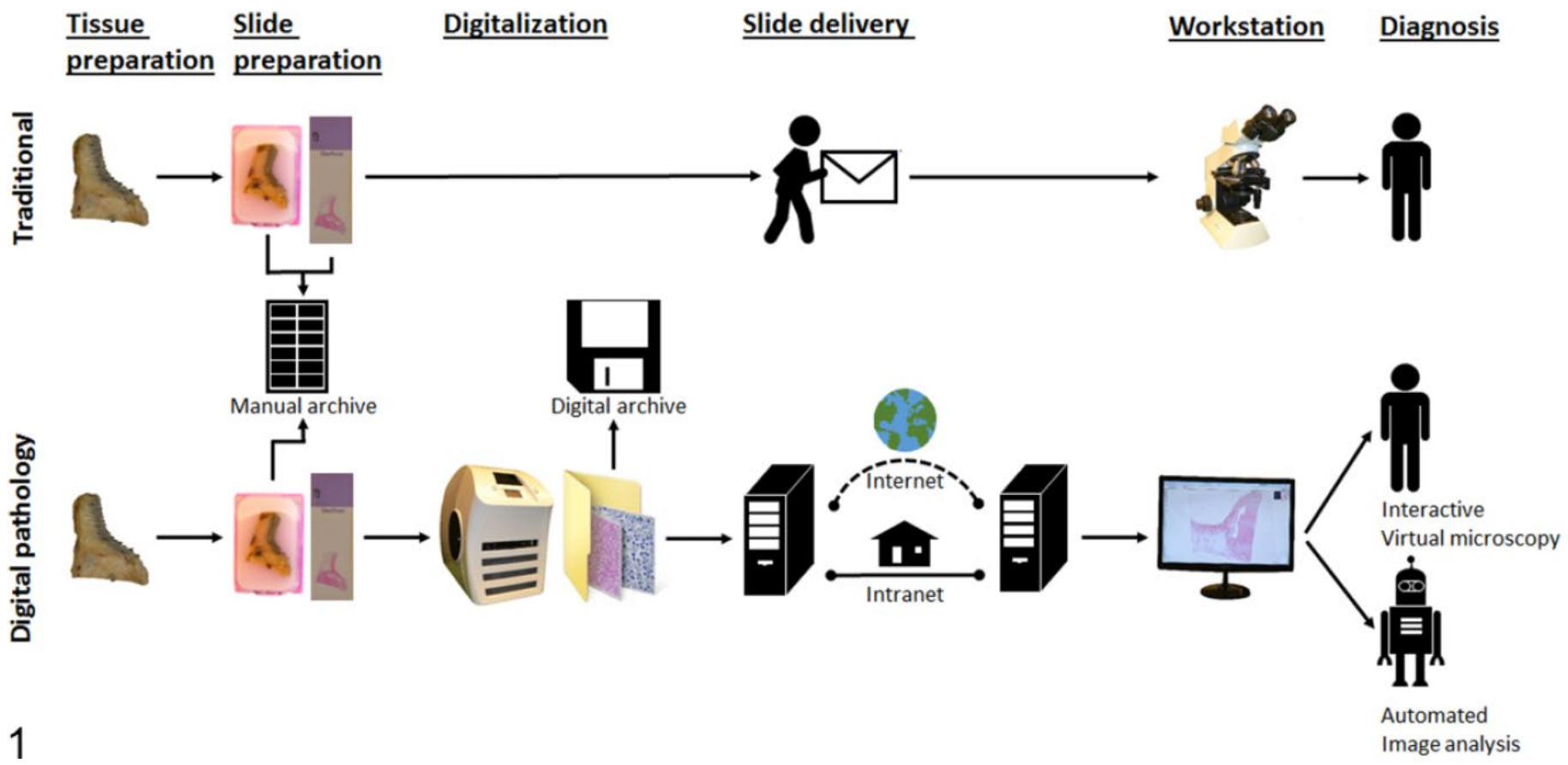
- CDC NIOSH – National Mesothelioma Virtual Bank
- NCI - CCSG – Cancer Bioinformatics Shared Facility
- NCATS - CTSA and ACT
- NLM - BMI TP
- PCORI – PaTH CRN

***Disclaimer: I am a member of NCI's Board of Scientific Advisors and Frederic National Labs Advisory Board Tech WG***

# Take Home Messages

- **FDA approval of whole slide imaging is driving Computational Pathology and AI in Pathology**
- **Informatics is key to impact of Comp Path & predictive analytics in clinical practice**
- **Pathology and Radiology partnerships need to be enabled**
- **Key Impact Areas – Predictive Analytics for health and discovery science via Comp Path!**

# Whole Slide Imaging (WSI) Workflow




**From Bertram and Kopfleish, Vet Path 2017**

# FDA Approval for Whole Slide Imaging

## WSI Regulatory Timeline

| Date | Event | Consequence |
|---|---|---|
| 2000 | Commercial WSI devices | Digital Pathology trend started<br>Validation of diagnostic applications |
| 2009 | FDA advisory panel | High risk (class III) device<br>Non-clinical use cases expanded<br>Non-US regulatory approval<br>DPA & select vendor discussions with FDA |
| 2015 | TPA* guideline | Assures manufacturers follow same standards |
| 2017 | FDA approval of WSI | For primary diagnosis in surgical pathology |


Feldman M et al A Large Multicenter, Retrospective Non-Inferiority Study to Evaluate Diagnostic Concordance Between Optical vs Digital Microscopic Diagnoses in 2000 Surgical Pathology Cases. Mod Pathol 2017; 30(S2):395A

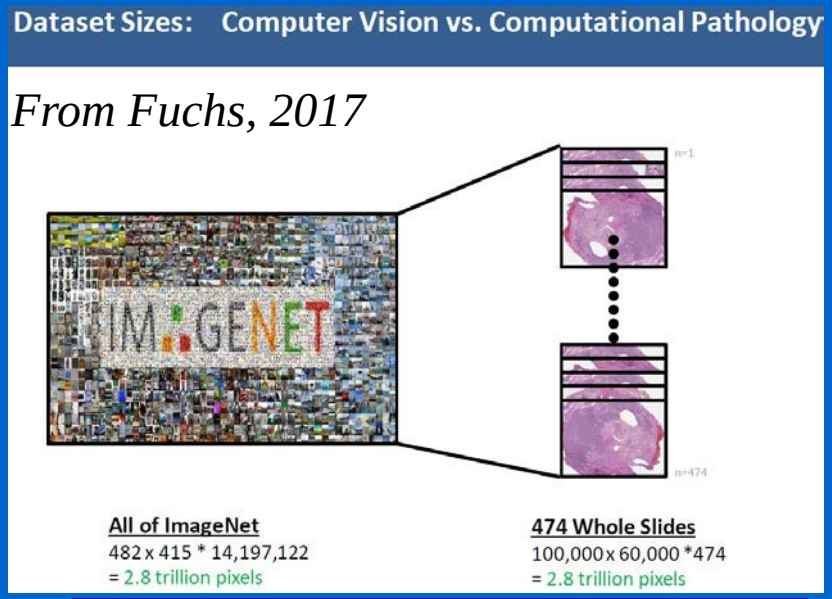
*TPA = Technical Performance Assessment

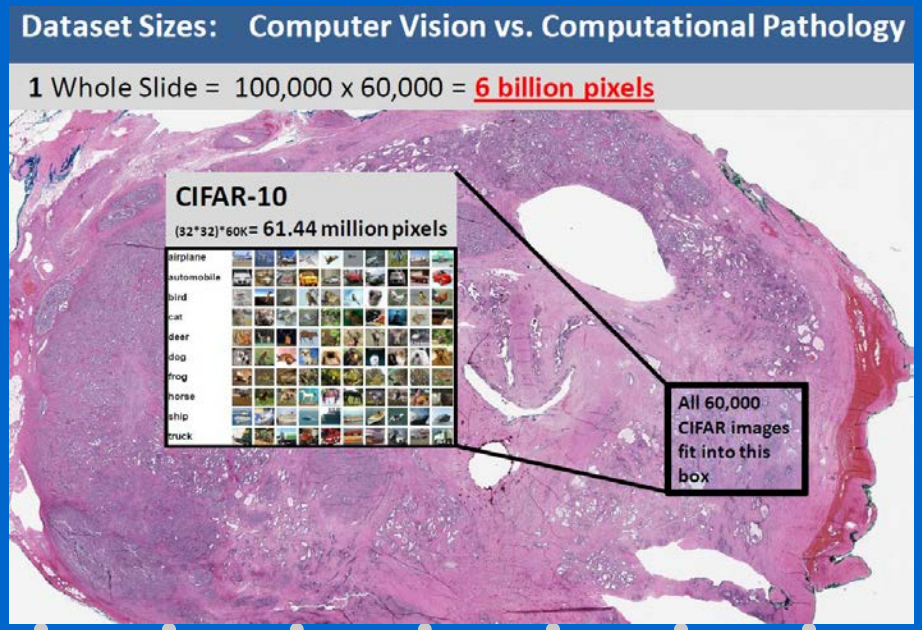
*From Pantanowitz, Path Info Summit, 2017*

# Computational Pathology – What is it?

**Computational Path = Big Data Science Meets Pathology**

- **Massive Increases in Volume of Digital Data Generated from WSI PLUS genomic sequencing data –**
- **Heralds the rise of computational pathology**
- **Critical for Personalized Medicine, Learning Health Systems, Basic Research and "Big Data/Data Science"**



# Introduction to Whole Slide Imaging

## Pipeline for Whole Slide Feature Characterization

- $10^{10}$ pixels for each whole slide image: $10^5 x 10^5$
- 10 whole slide images per patient
- $10^8$ image features per whole slide image
- $10^{15}$ pixels
- $10^{13}$ features

From Saltz, NLM circa 2011

CAMELYON17

1000 scanned slides

3.03 Tbyte

444 registrants

# Definition of Computational Pathology

- **An approach to diagnosis that incorporates multiple sources of data (H&E, IHC, IF & genomic data)**
- **Presents clinically actionable knowledge (big data to knowledge)**
- **Advanced decision support for precision (personalized) medicine**
- **Helps to redefine Pathology from an observational to knowledge engineering discipline hence critical to healthcare data science**

(Louis et al Arch Path Lab Med 2014)

# Support for Comp Path – Integration Needed

**Data Types Critical to Success for Predictive Analytics**

- **Imaging – Rich source of computable information**
  - Need to De-ID WSI and then “deeply” annotate
- **Phenotype – From Anatomic Pathology Lab Info Sys**
  - Structured data from synoptic reports
  - Unstructured data via NLP (Text Info Extract Sys – TIES)
- **Computational Pathology Annotation – more later**
- **Outcomes Data – From Cancer Registry Systems**
- **Data to Integrate – Biobanks, Clinical Pathology & Molecular Pathology**

# Computational Pathology and Informatics Science

- A Research Data Warehouse (RDW) and robust Biorepository Supported by Informatics Science are key enablers!!!
  - This will require biomedical informatics expertise:
    - Expertise in **database design and query** – ***PCORnet and PaTH***
    - **Natural language processing** of text (H&P, Consults, Discharge, Pathology Reports, etc..) – ***TIES/TCRN and CDP***
    - Structured capture of key medical data will require controlled **vocabularies** and implementation of **ontologies** - ***TIES/TCRN & CDP***
    - **De-identification** of text for sharing with researchers (De-ID Data Corp) – ***TIES/TCRN & CDP***
    - Implementation of **Bayesian** algorithms to make genomic data "actionable" via **Causal Modeling and Discovery** – ***BD2K – Center for Causal Modeling and Discovery***
      - *– Pitt just funded for four years and $11M with CMU, UPMC, Yale*
  - Personalized Medicine critically requires **biorepositories***
    - ***TCRN***

*(U Pitt is #1 contributor to The Cancer Genome Atlas – see subsequent slides)

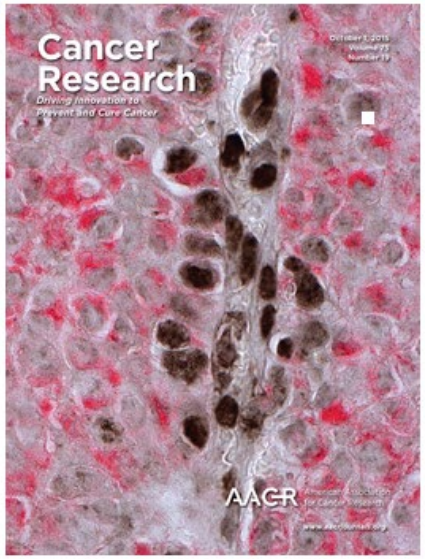
# What is the Text Info Extraction System (TIES)?

- An NLP and Information Retreival system for de-identifying, annotating, storing and retrieving pathology and radiology documents
- A system for indexing research resources (FFPE, FF, WSI) with document annotations
- An GUI for querying large repository of annotated documents and obtaining resources locally, using an honest broker model
- A platform to support data, biospecimen and both Pathology and Radiology images for sharing among networks of cancer centers and other institutions

# TCRN Pubs

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Home | OnlineFirst | Current Issue | Past Issues | Subscriptions | Alerts | Feedback | AACR Publications | CME | AACR Home




- View the new Impact Factor
- View the Most-Cited Articles of *Cancer Research*

Cancer Research

Resource

## A Federated Network for Translational Cancer Research Using Clinical Data and Biospecimens

Rebecca S. Jacobson[1], Michael J. Becich[1], Roni J. Bollag[2], Girish Chavan[1], Julia Corrigan[1], Rajiv Dhir[1], Michael D. Feldman[3], Carmelo Gaudioso[4], Elizabeth Legowski[1], Nita J. Maihle[2], Kevin Mitchell[1], Monica Murphy[4], Mayurapriyan Sakthivel[4], Eugene Tseytlin[1], and JoEllen Weaver[3]

### Abstract

Advances in cancer research and personalized medicine will require significant new bridging infrastructures, including more robust biorepositories that link human tissue to clinical phenotypes and outcomes. In order to meet that challenge, four cancer centers formed the Text Information Extraction System (TIES) policies, and procedures, enable regulatory compliance. The TIES Cancer Research Network now provides integrated access to investigators at all member institutions, where multiple investigator-driven pilot projects are underway. Examples of federated search across the network illustrate the potential impact on

TIES

http://ties.dbmi.pitt.edu/

# Computational Pathology and Informatics Science

- A Research Data Warehouse (RDW) and robust Biorepository Supported by Informatics Science are key enablers!!!
  - This will require biomedical informatics expertise:
    - Expertise in **database design and query** – ***PCORnet and PaTH***
    - **Natural language processing** of text (H&P, Consults, Discharge, Pathology Reports, etc..) – ***TIES/TCRN and CDP***
    - Structured capture of key medical data will require controlled **vocabularies** and implementation of **ontologies** - ***TIES/TCRN & CDP***
    - **De-identification** of text for sharing with researchers (De-ID Data Corp) – ***TIES/TCRN & CDP***
    - Implementation of **Bayesian** algorithms to make genomic data "actionable" via **Causal Modeling and Discovery** – ***BD2K – Center for Causal Modeling and Discovery***
      - *– Pitt just funded for four years and $11M with CMU, UPMC, Yale*
  - Personalized Medicine critically requires **biorepositories***
    - ***TCRN***

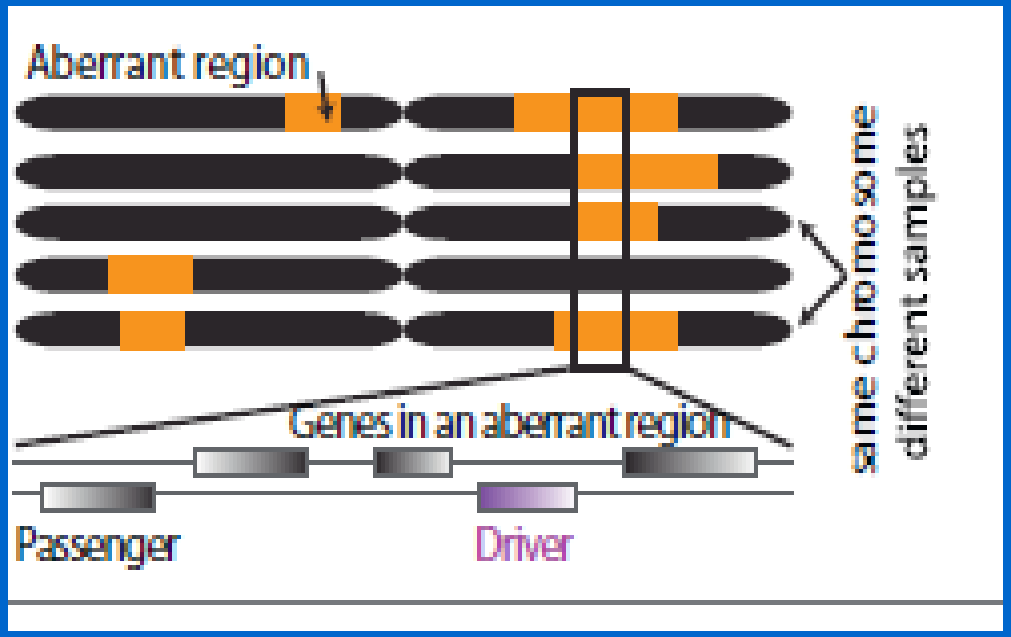
*(U Pitt is #1 contributor to The Cancer Genome Atlas – see subsequent slides)

# Pitt Center for Causal Modeling and Discovery

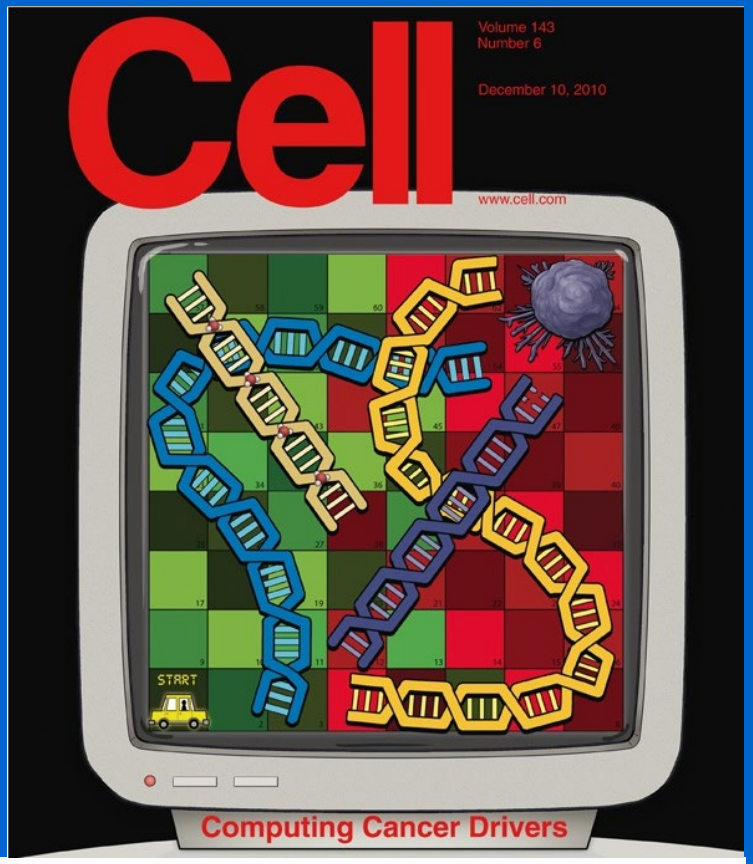
- Led by Greg Cooper, MD PhD (Biomedical Informatics), Ivet Bahar, PhD (Computational and Systems Biology) and Jeremy Berg, PhD (Director of Institute for Personalized Medicine) involves Becich/Crowley – TCGA/PGRR/Computable Phenotypes
- Theme: Modeling and discovery of causal networks from genome and phenome (patient EHR/i2b2 RDW) biomedical datasets including those from clinical trials and other controlled biospecimen based efforts
- Aims
  - ***Represent*** causal knowledge within a unified, formal framework
  - ***Discover*** causal knowledge from biomedical data (both observational and experimental) and background knowledge (e.g., from the literature) using efficient algorithms
  - ***Apply*** causal knowledge to support browsing, answering causal queries, simulating causal processes, and designing experiments to resolve causal uncertainties
- Driving Biological Problem areas: Signaling Pathways in Cancer/TCGA (**Computational Pathology and Genomics/TCGA**), Idiopathic Pulmonary Fibrosis (**Computational Pathology Imaging**) and "Connectome" via fMRI brain imaging

# Causal Network Discovery = Computational Pathology

**A probabilistic network approach to uncover genetic drivers of melanoma using data on copy number variation and gene expression***



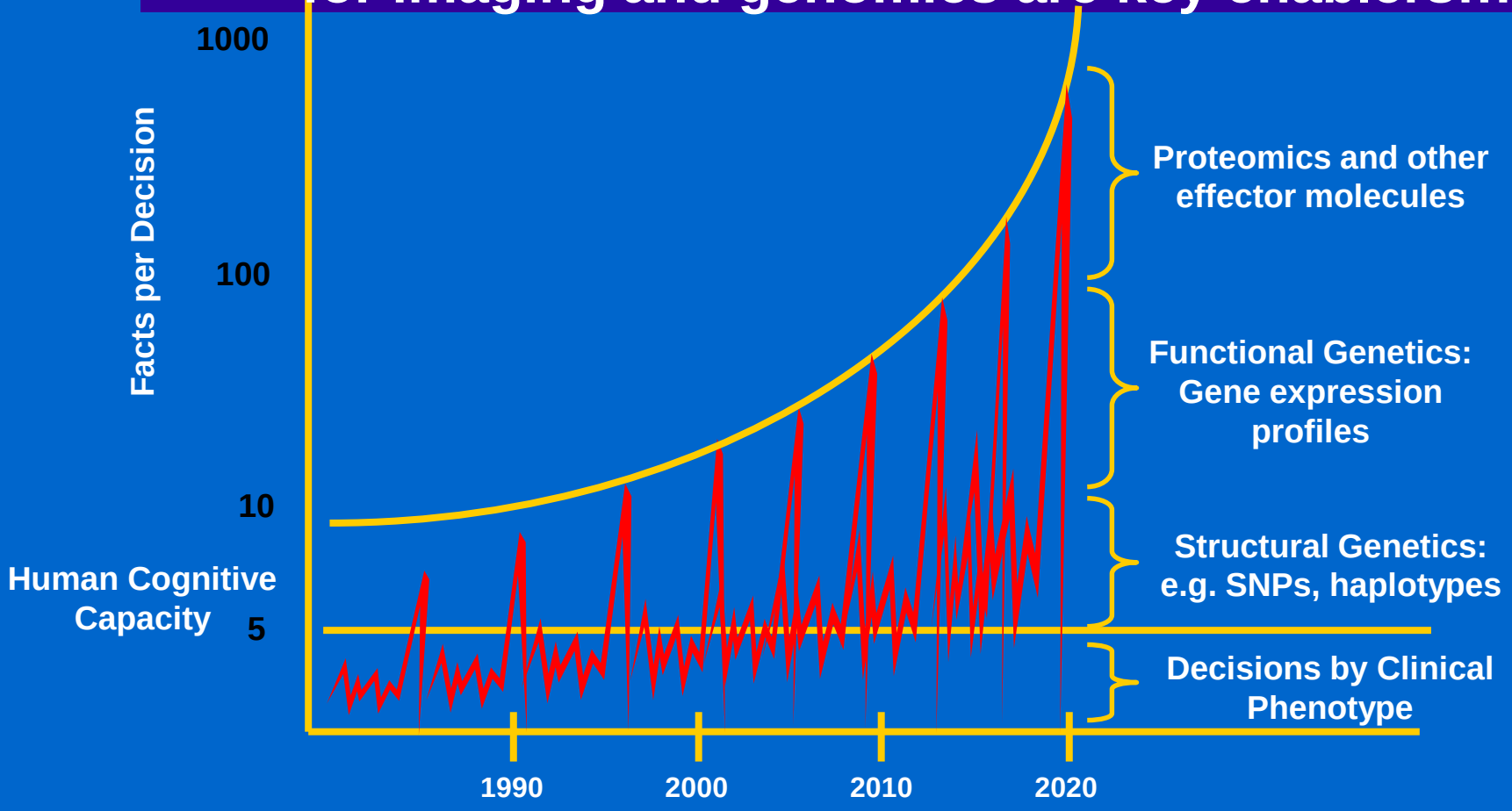

Akavia UD , et al. *Cell* 143 (2010) 1005-1017.
(The figure above appears in this paper)




**10 December, 2010 Volume 143, Issue 6**

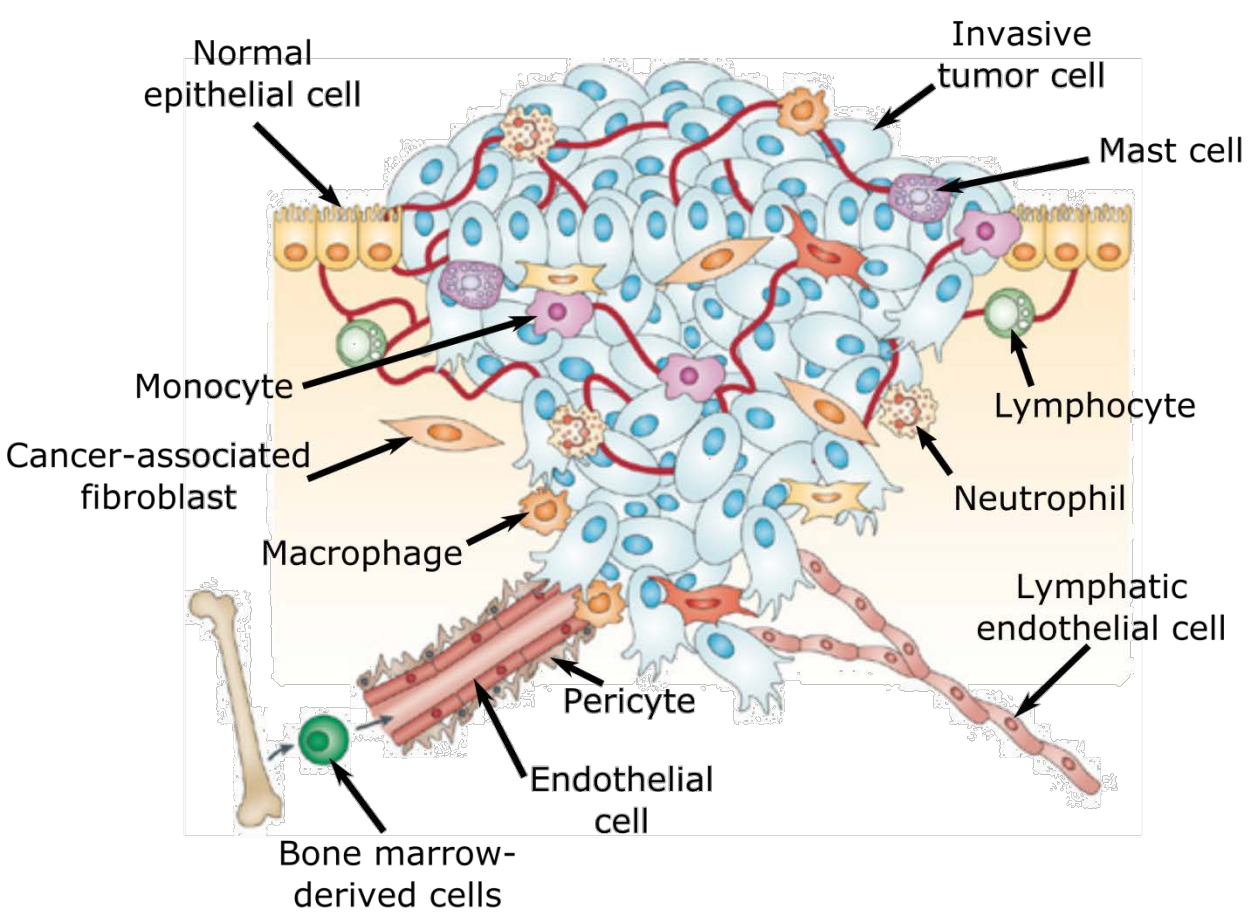
On the cover: In cancer, heterogeneous genetic aberrations frequently result in similar phenotypic outcomes. In this issue, Akavia et al. (pp. 1005–1017) report a computational algorithm that identifies driving mutations and links them with their downstream transcriptional effects. The approach thus allows the authors to begin disentangling the complex mechanisms by which genetic aberrations drive transformation. The cover shows a driver with many DNA-mediated paths to a tumor.

# Computational Pathology Algorithms like those in CCD for imaging and genomics are key enablers!!!



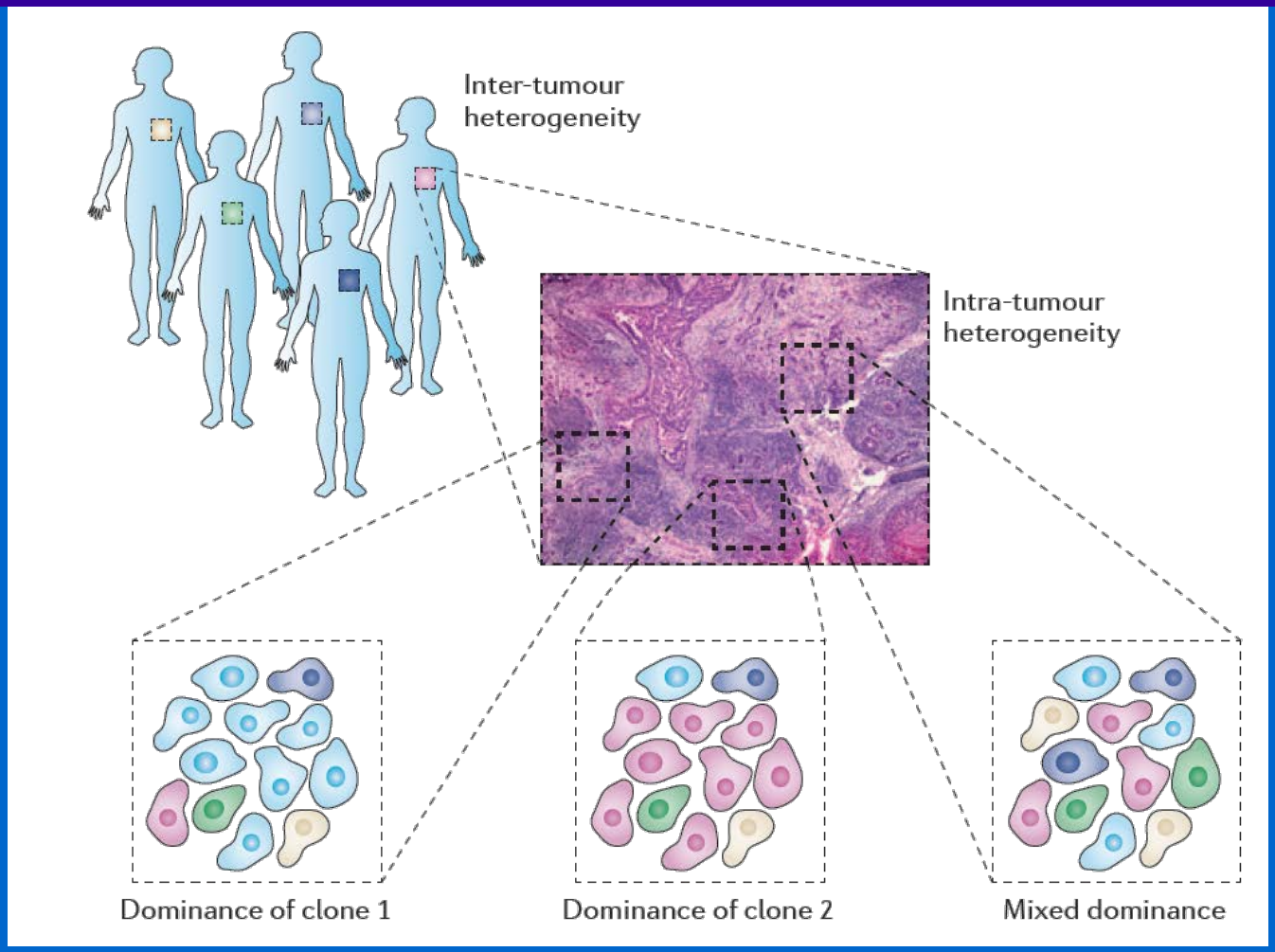

From William Stead: http://courses.mbl.edu/mi/2009/presentations_fall/SteadV1.ppt & http://www.mbl.edu/education/courses/special_topics/pdf/med_sched09_fall.pdf

# 1. Intra-tumoral spatial heterogeneity




Nature Reviews Cancer (2009) 9:239–252

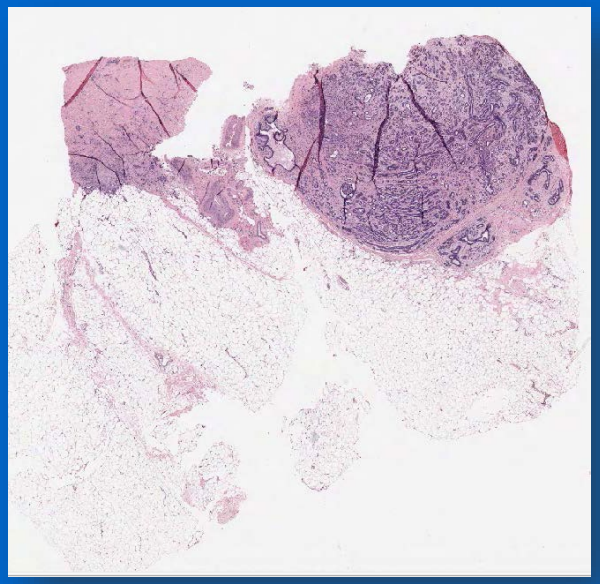
# 1. Quantify Heterogeneity for Diagnostics, Prognostics & Immunotherapy




Marusyk A. Nat Rev Cancer. 2012 Apr 19;12(5):323-34

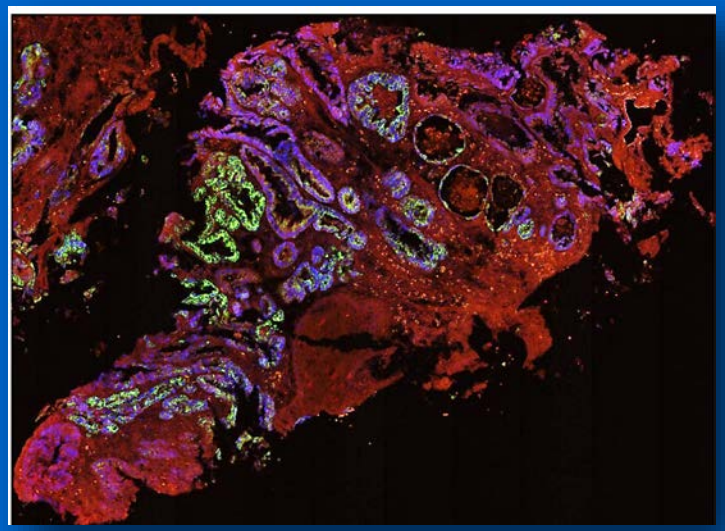
# 2. Integration of H&E, IHC and IF

**H&E in transmitted light**



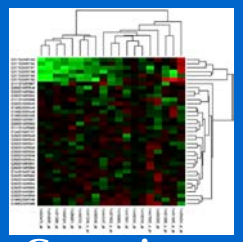
- Universal method
- Limited molecular measurements in transmitted (IHC)
- Complex tissue “scenes”

**Multi to hyperplexed**

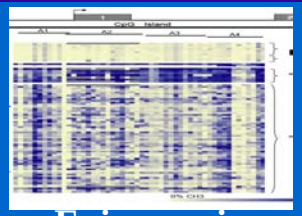


- Emerging method with potential to measure many DNA/RNA/proteins in the same tissue section/TMA
- Structural biomarkers
- Quantify biomarker expression levels and tissue based spatial relationships.

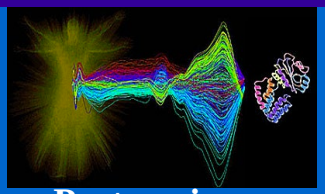
# 3. Histopatholomics: Spatial ITH



Genomics

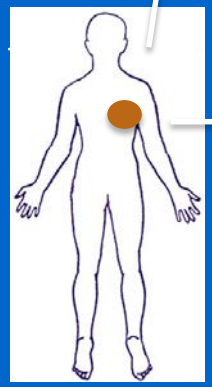


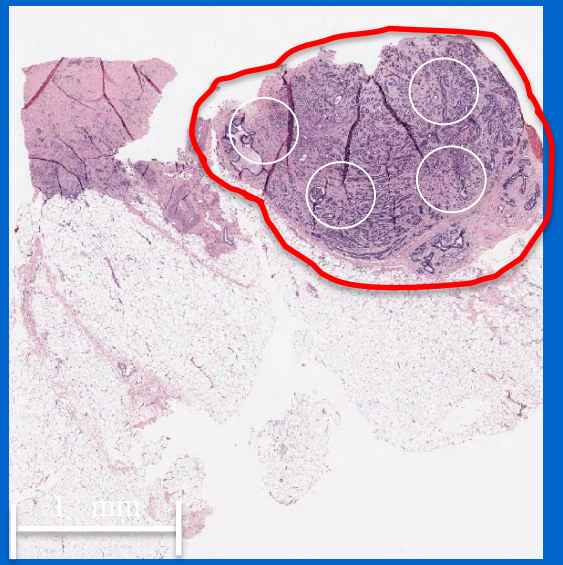
Epigenomics



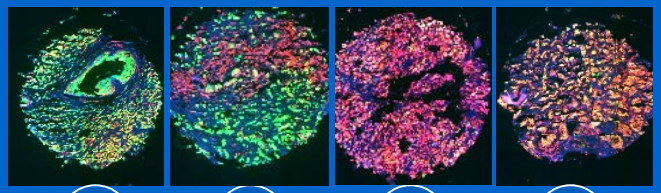
Proteomics

Metabolomics

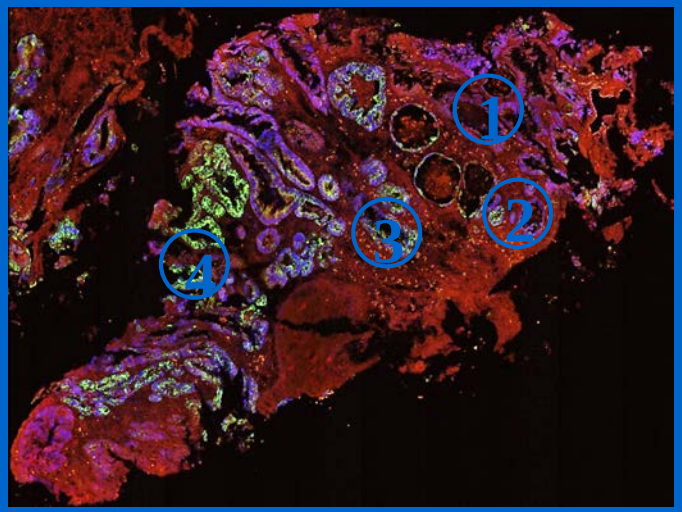





H&E stained whole tissue section from FFPE tumor sample



4 3 2 1




Multi to hyperplexed fluorescence imaging of whole section for higher spatial resolution and tissue context

Thanks to Dr. Joe Ayoob for HPlomics

# 4. Ground-truth & Annotation

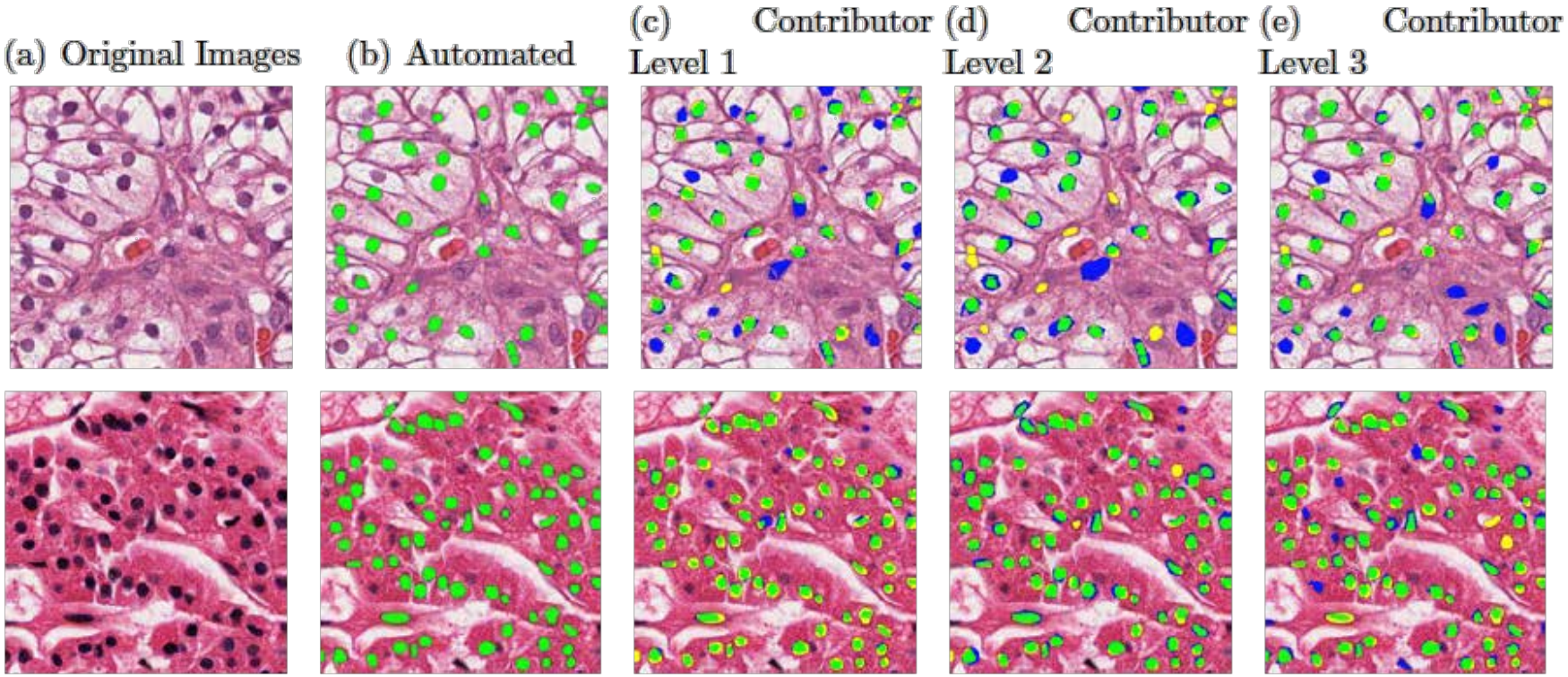
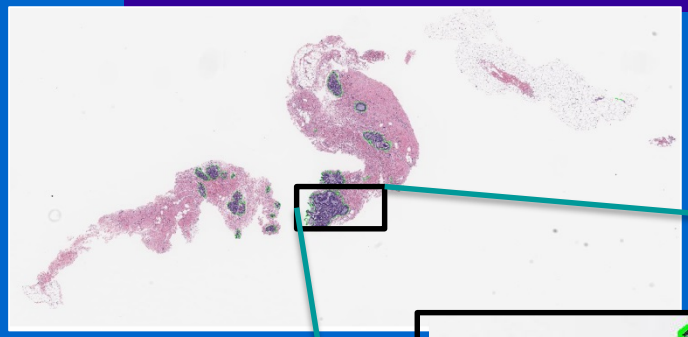
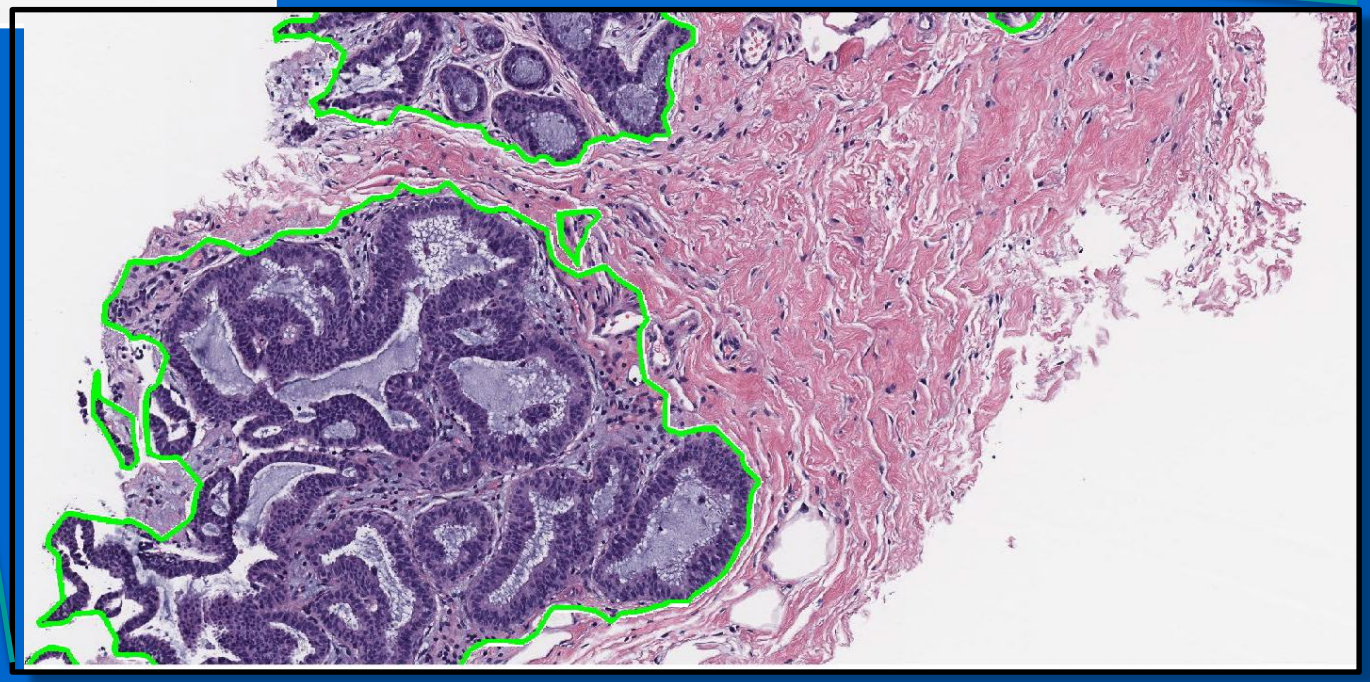



Fig. 2. Examples of nuclear segmentation using an automated method and increasing contributor skill level, ranging from 1 to 3. (Green region indicates TP region, yellow region indicates FN region and blue region indicates FP region). The automated nuclei segmentation used as ground truth.

# 5. ”Callable” ID of Histologic Features

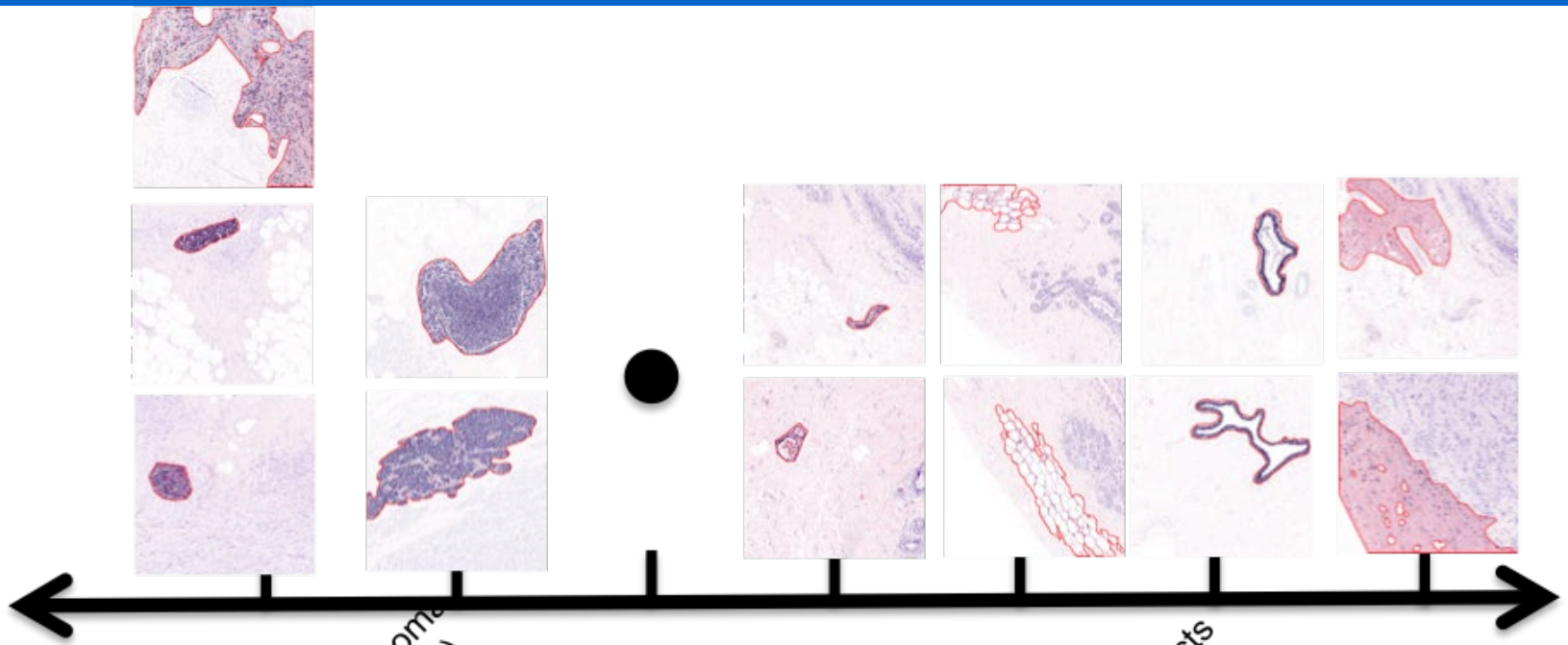


SpIntellx™



# 6. Triaging ROIs

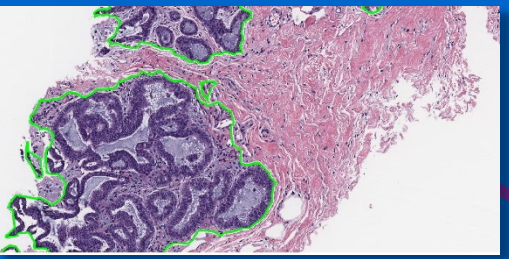
SpIntellx™

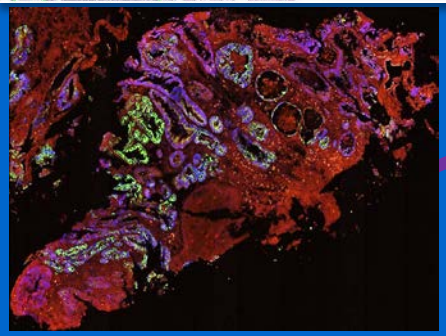


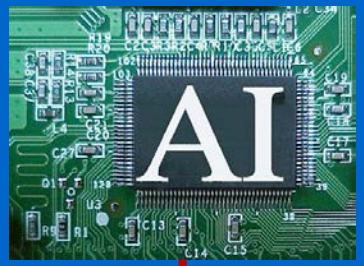

Most suspicious

Least suspicious

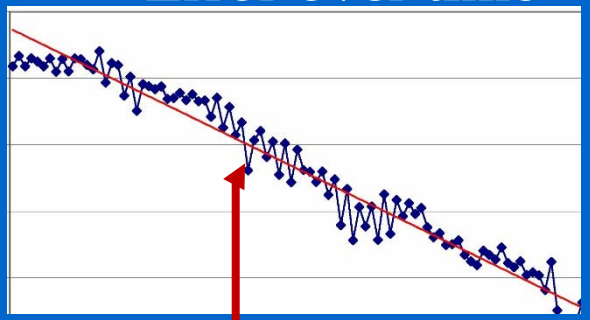
# 7. Computational Pathology and AI



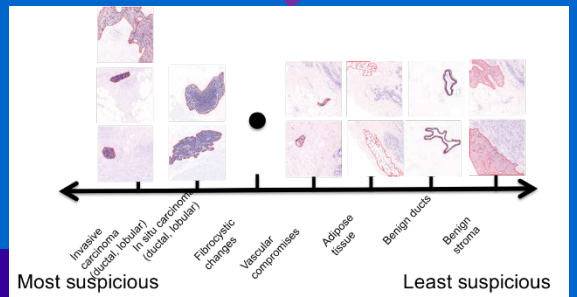





Error over time



Intervention

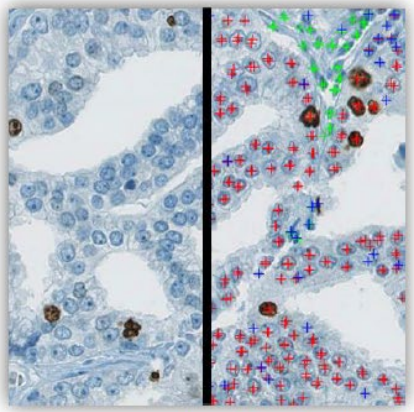


# Pathology and Radiology Must Partner!

## A Joint Effort for Personalized Medicine

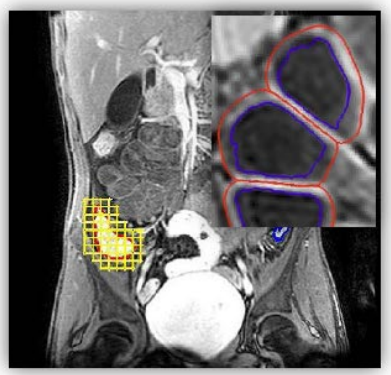
*From Fuchs, 2017*

**Pathology**



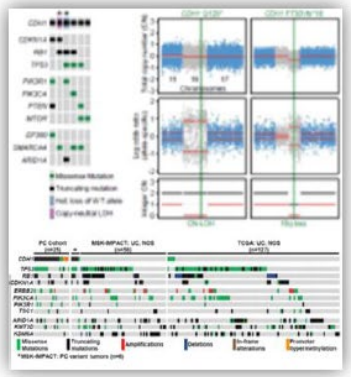
Computational Pathology

**Radiology**



Radiomics @ MSKCC

**Genomics**



cBioPortal @ MSKCC

Combining quantitative analyses from pathology, radiology and genomics facilitates true **personalized medicine.**

- End of Talk – e-mail me at becich@pitt.edu if you have questions/clarifications not covered in the discussion.

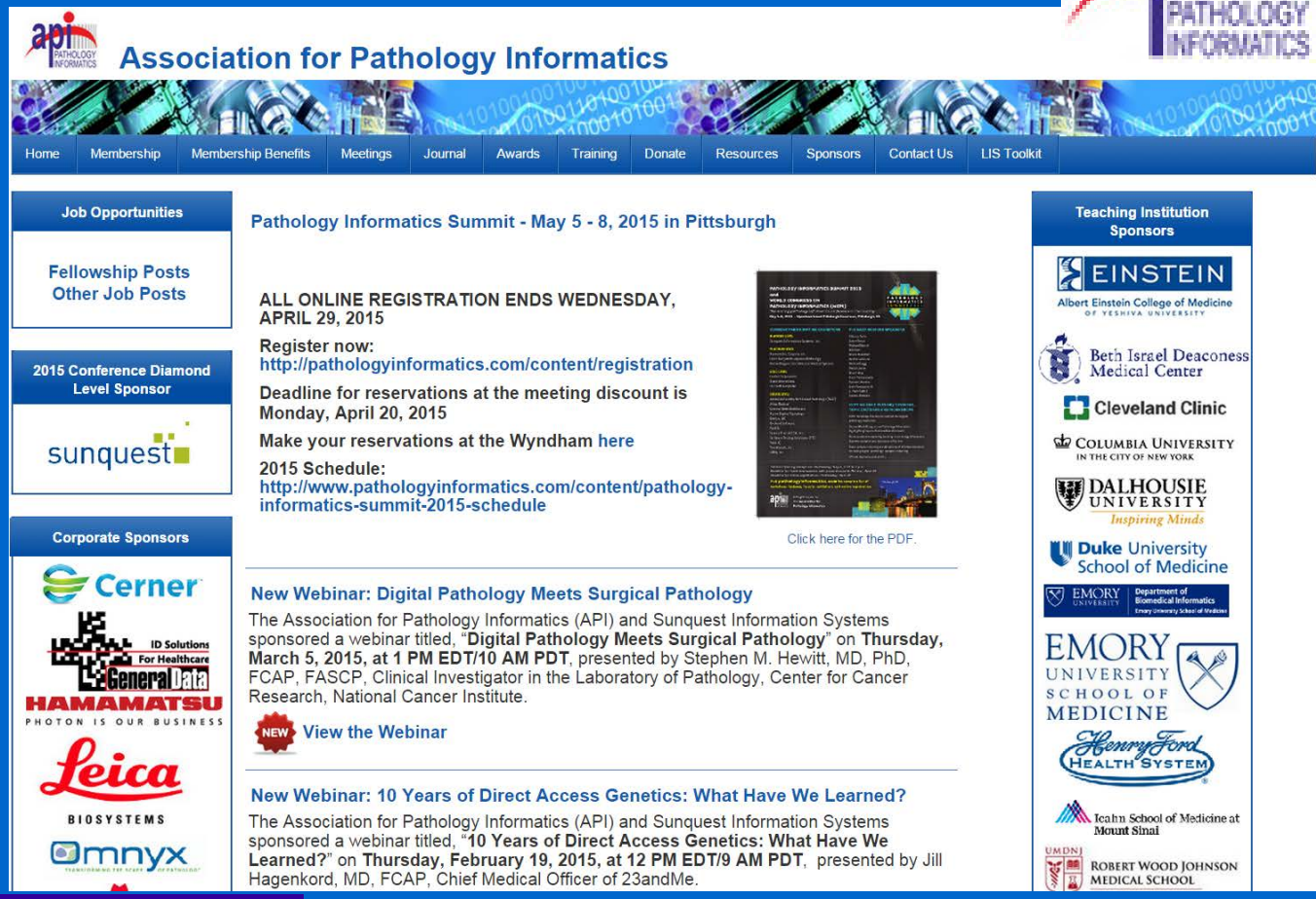
NOTE: E-mail me if you want PDFs of articles or presentation.

**Thank you for organizing this MCCP Workshop!**

# Association for Pathology Informatics (API)

http://www.pathologyinformatics.org

*"…to advance the field of pathology informatics as an academic and a clinical subspecialty of pathology…"*



# Journal of Pathology Informatics

Co-Editors Liron Pantanowitz, MD PhD and Anil Parwani, MD PhD

Please support JPI, API and Pathology Informatics as the Home for Digital Pathology - Great Academic and Strategic Partnership with Multiple Benefits!!!

# Thanx to the Pitt team!

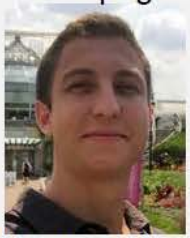
SpIntellx™


Dan Spagnolo


Dr. Shikhar Uttam


Dr. Burak Tosun


**Lương Nguyên**


Maurice Marx


Dr. Om Choudhary


Dr. Filippo Pullara


Akash Parvatikar


Nathan Ong


Chakra


Alex Andonian


Tammy Ma

Drug Discovery Institute/CSB
Dr. Lansing Taylor
Dr. Andy Stern
Dr. Bert Gough
Dr. Tim Lezon

**UPMC/Magee/Pathology**
Dr. Jeffrey Fine

Center for Commercial Applications of
Health Care Data/DBMI
Dr. Michael Becich
Dr. Don Taylor

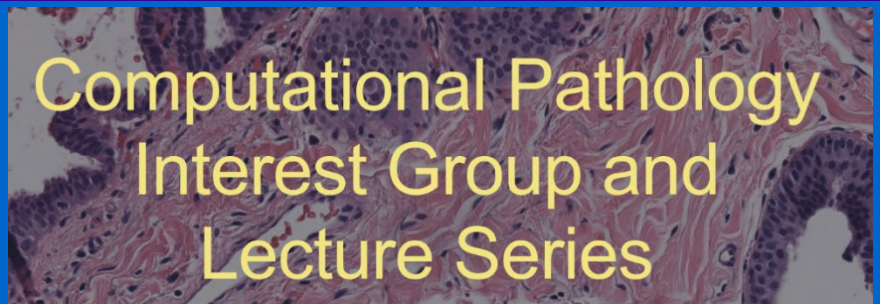
GE Global Research
**Brion Sarachan**
**Dr. Yousef Al-Kofahi**
**Dr. Fiona Ginty**
Magee-Womens Research Institute
Dr. Adrian Lee
Dr. Rekha Gyanchandani
CSB
Dr. Takis Benos
**Dr. Dimitrios Manatakis**
**Dr. Joe Ayoob**


- UPMC & Center for Commercial Applications of Healthcare Data Grant #711077
- NIH-NCI U01 CA204826
- NHGRI BD2K U54HG008540 (Lung DBP)


SPDx

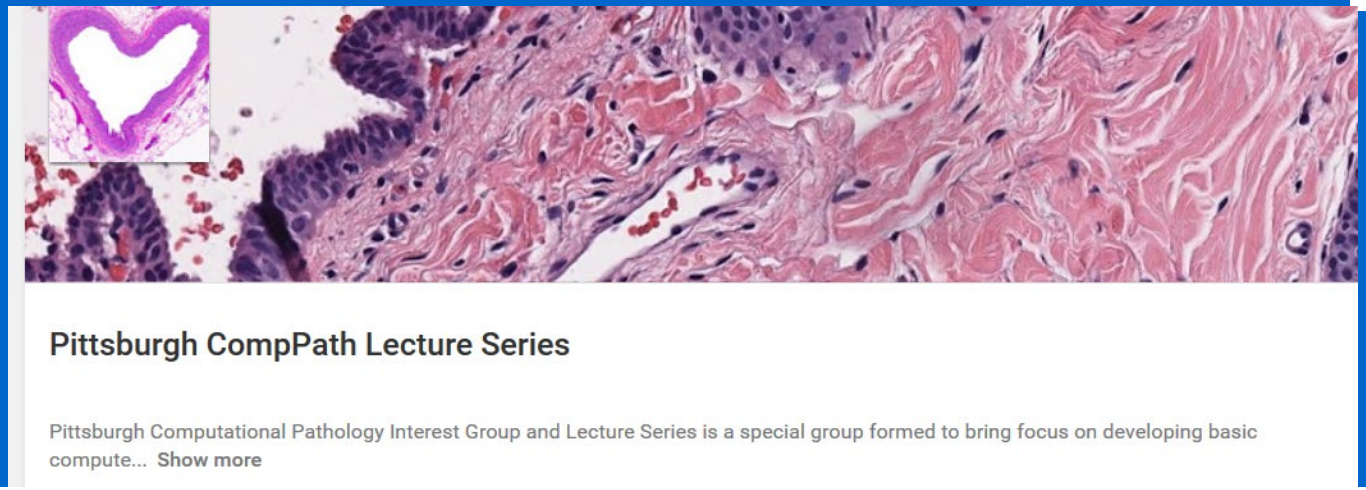
Spatial Pathology Powers Precision Oncology

# Thanx to Chakra Chennubhotla, PhD & Burak Tosun, PhD




http://www.csb.pitt.edu/comppath/

Youtube Channel for Lectures

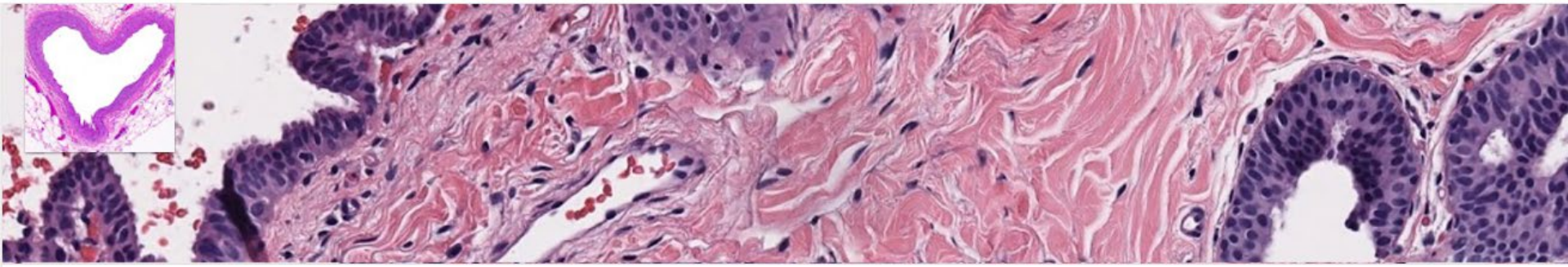



https://www.youtube.com/watch?v=Xrvxc0YNcAM&feature=youtu.be

# Comp Path Team Profile - Pittsburgh

- **Special Interest Group Members (n=93):**
  - Bioengineering (5%)
  - Biomedical Informatics (10%)
  - Computational and Systems Biology (10%)
  - Computer Science and Machine Learning (20%)
  - Industry and Entrepreneurs (20%)
  - Medicine (10%)
  - Pathology Informatics (10%)
  - UPMC (15%)

***Participation from Carnegie Mellon & Dusquene Universities***



## Pittsburgh Computational Pathology Lecture Series

Subscribe 32

Pittsburgh Computational Pathology Interest Group and Lecture Series aims to bring focus on developing basic computer vision and and machine learning algorithms for transmitted light and fluorescence microscopy imaging data from histopathology and on highlighting the translation role of computational pathology research in interfacing with genomics, personalized medicine and microbiomics.

Our goal is to promote a vibrant community of computational pathology research and education practices here in Pittsburgh by engaging researchers and clinicians across Pitt, CMU and Duquesne campuses, and the local industrial partners.

http://www.csb.pitt.edu/comppath/
Show less

### Uploads


CompPath Dr. Shikhar Uttam, Ph.D. Lecture (April 14th 2017)
21 views • 1 month ago


CompPath Dr. Shandong Wu, Ph.D. Lecture (March 14th 2017)
24 views • 1 month ago


CompPath Dr. Jeffrey L. Fine, M.D. Lecture (January 19th 2017)
56 views • 4 months ago


CompPath Dr. Kayhan Batmanghelich, PhD Lecture
48 views • 4 months ago


CompPath Dr. Joshua T Mattila, PhD Lecture (November 14th
41 views • 6 months ago

Computational Pathology Lecture Archive -
**https://www.youtube.com/channel/UCWfBS3PLWHTIACeccm2CIog**