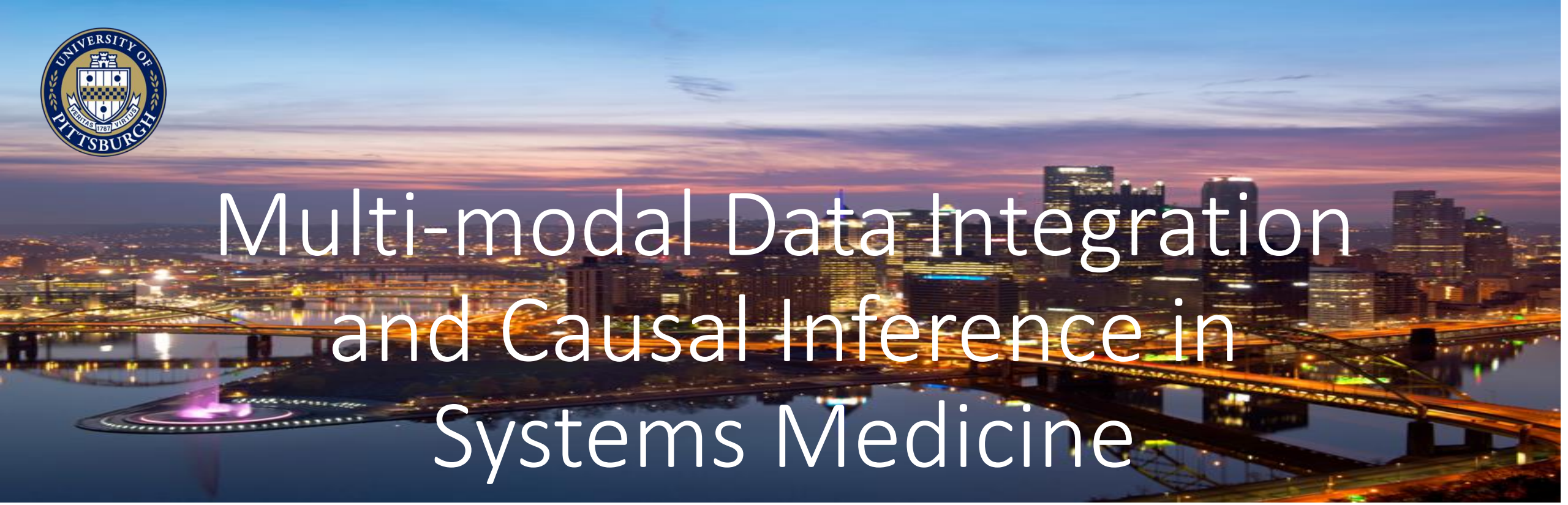


Multi-modal Data Integration and Causal Inference in Systems Medicine

University of Utah

Salt Lake City, Utah, September 2019

Takis Benos, PhD
Professor and Vice Chair
Department of Computational and Systems Biology
University of Pittsburgh School of Medicine

E-mail: benos@pitt.edu
Lab URL: http://www.benoslab.pitt.edu

From Basic to Translational Science:

What is Causal Inference and can it help my research?

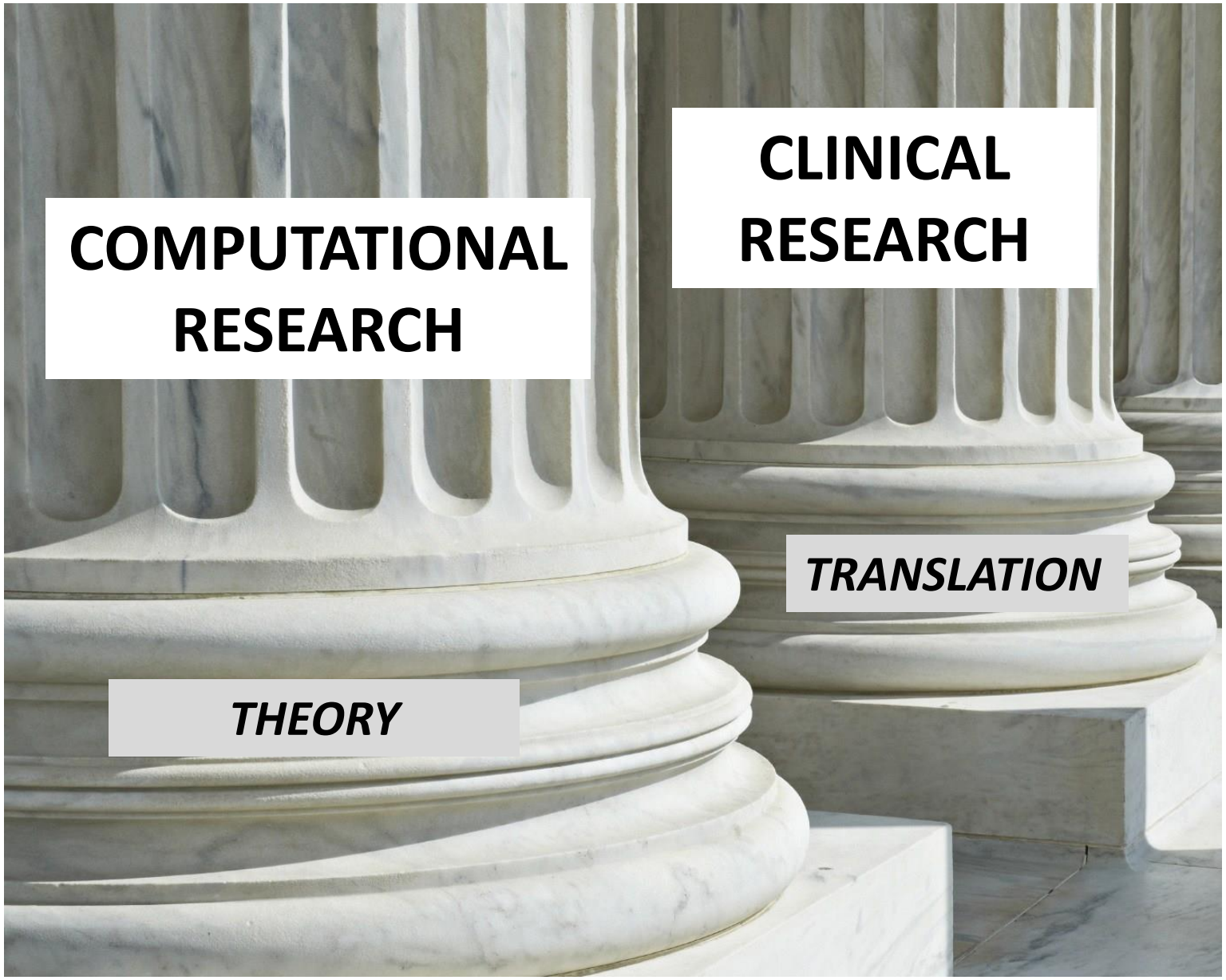
University of Utah

Salt Lake City, Utah, September 2019

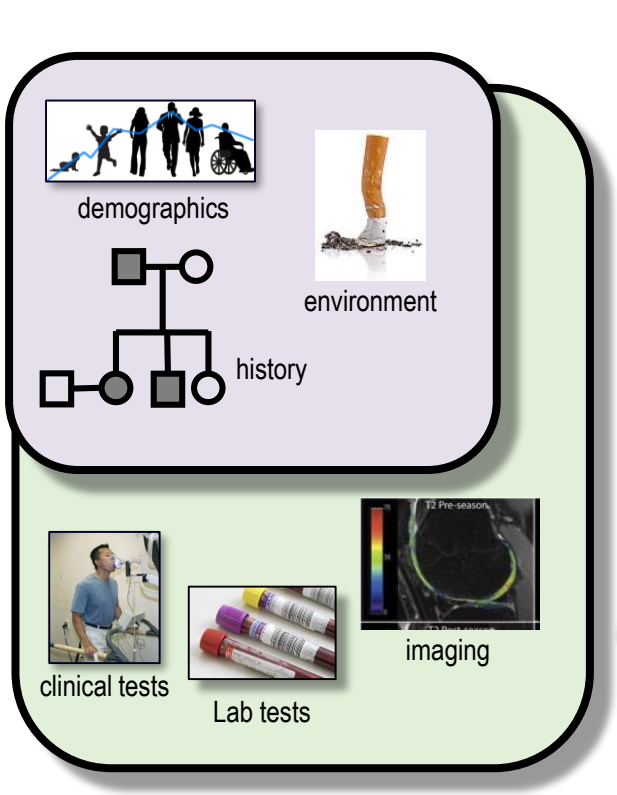
Takis Benos, PhD
Professor and Vice Chair
Department of Computational and Systems Biology
University of Pittsburgh School of Medicine

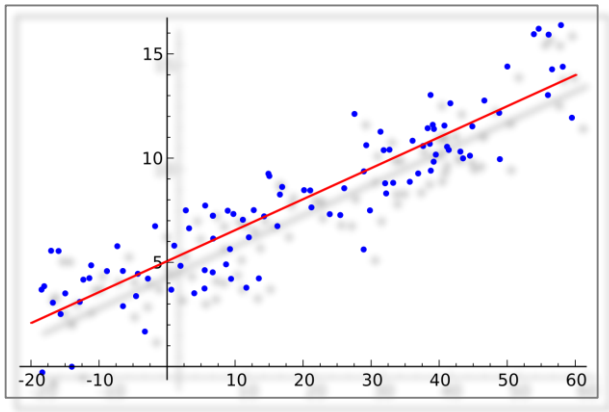
E-mail: benos@pitt.edu
Lab URL: http://www.benoslab.pitt.edu

BENOS' LAB

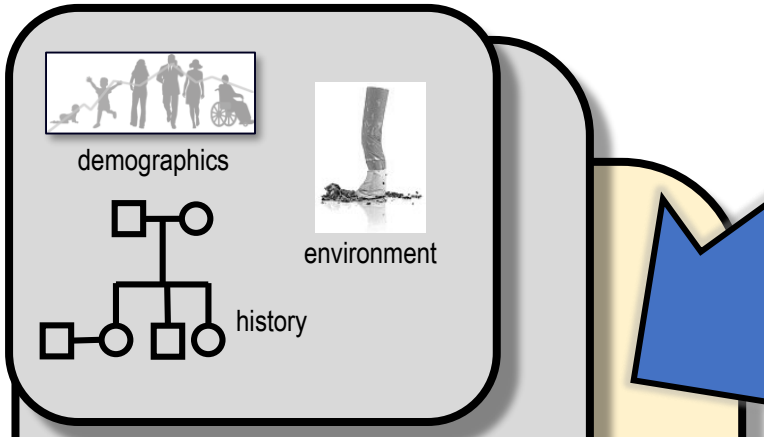


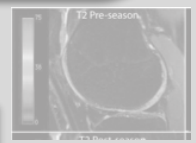
“Top-down” approach of investigating a disease

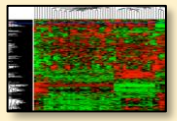


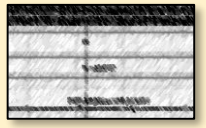


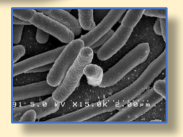
“Bottom-down” approach of investigating a disease

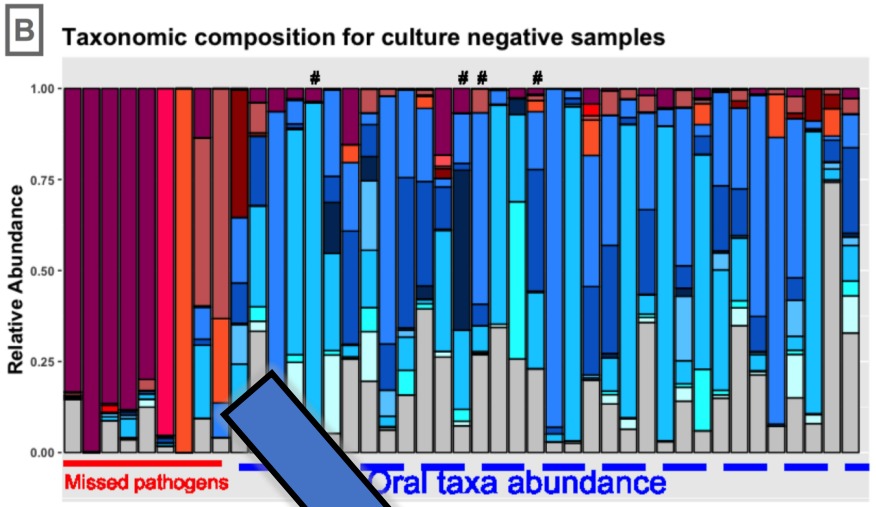


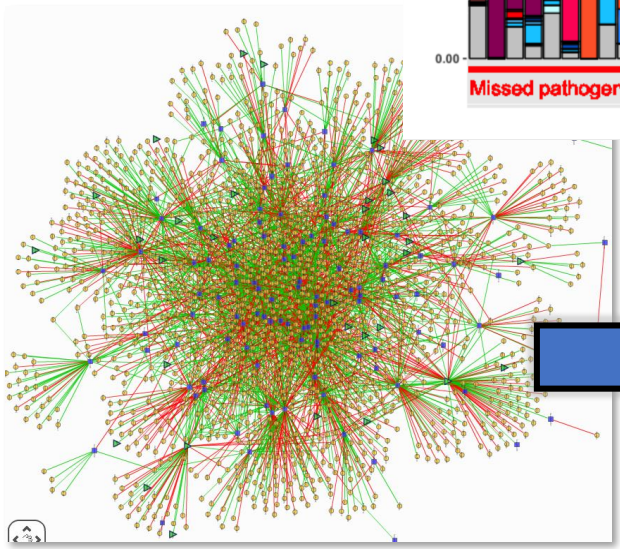


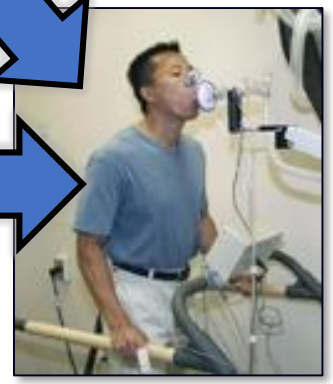






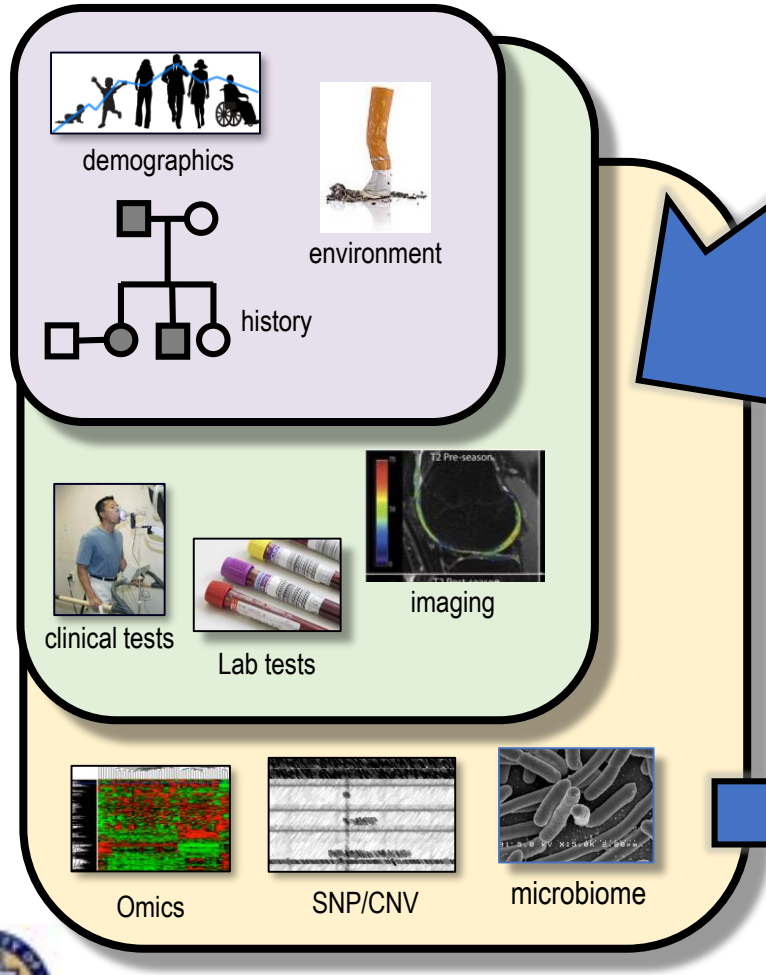


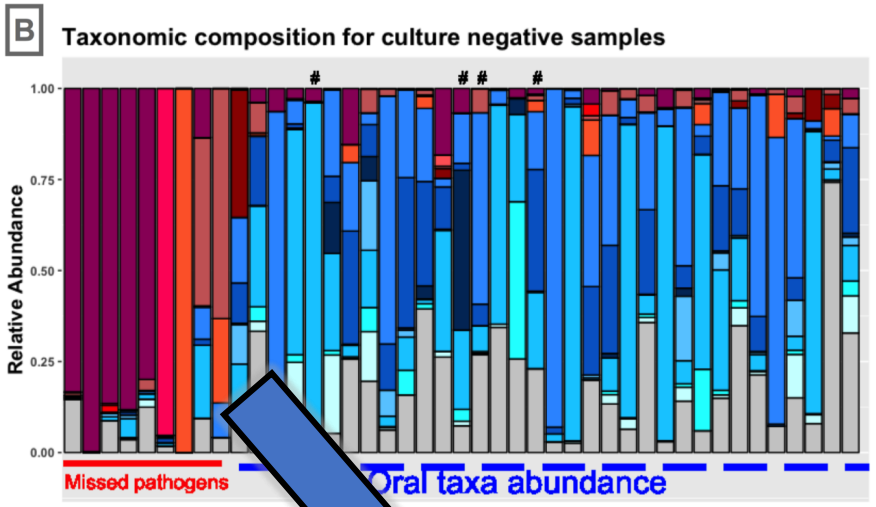


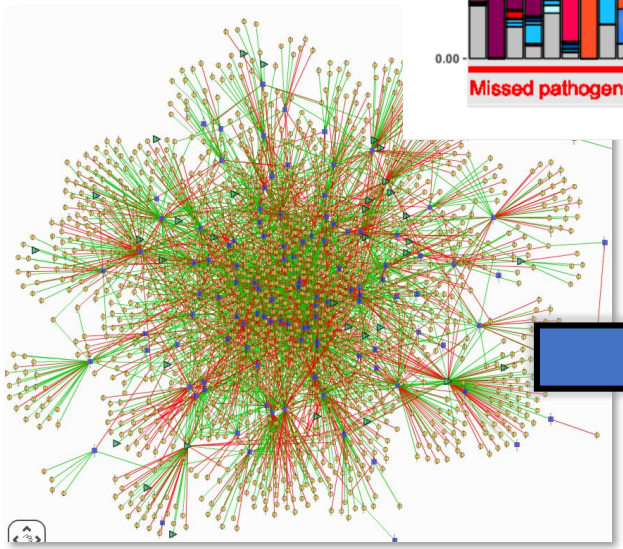


“Bottom-down” approach of investigating a disease

MULTI-SCALE DATA





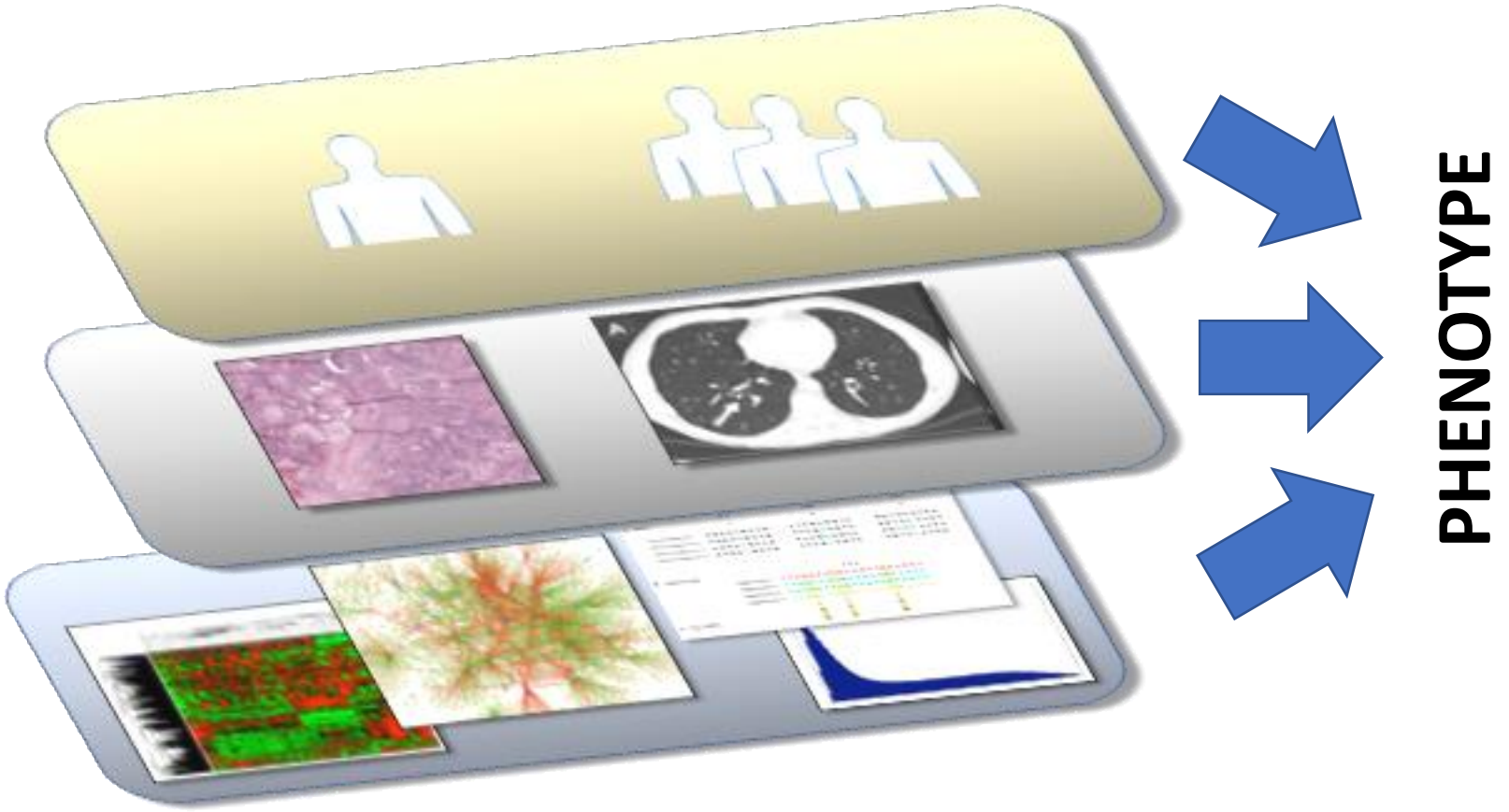


Systems Medicine: integrative systems-level analytics for individualized treatments

Organism / Population

Tissue / Organ

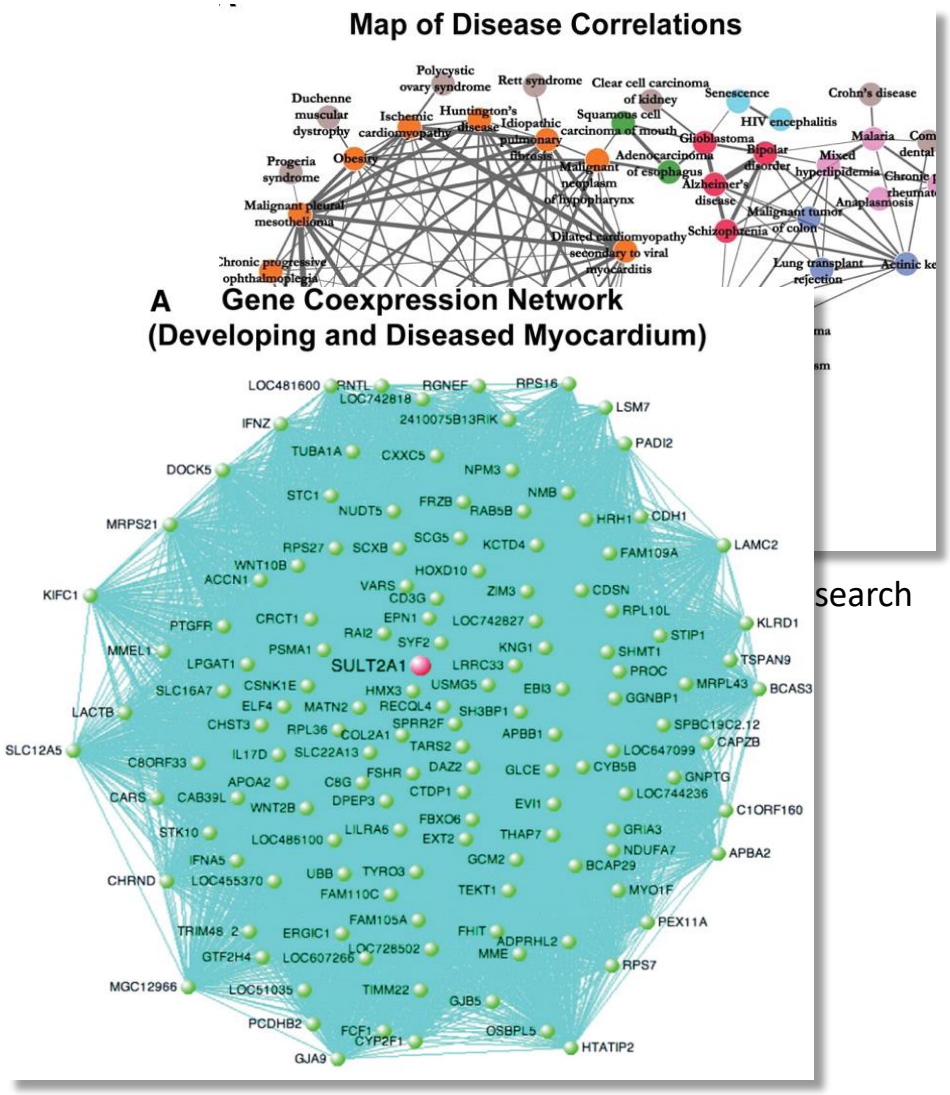
Molecular / Cellular



Correlation-based methods

- They are simple and thus very attractive

- They tend to overestimate the number of true connections
  - So we need to use prior or expert information to find testable hypotheses



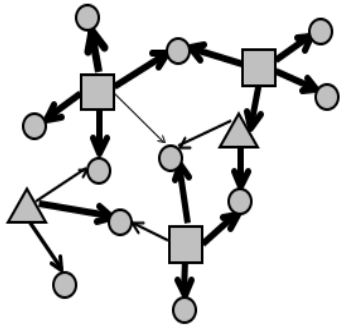
search

Chan and Loscalzo, 2012, Circulation Research

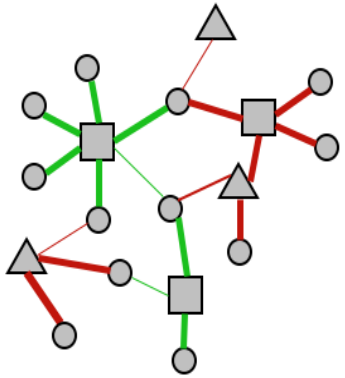
mirConnX: correlations with priors for miRNA:mRNA networks

Grace Huang
PhD

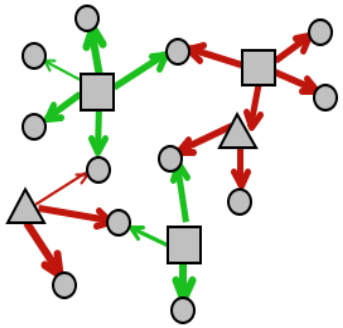
Static network

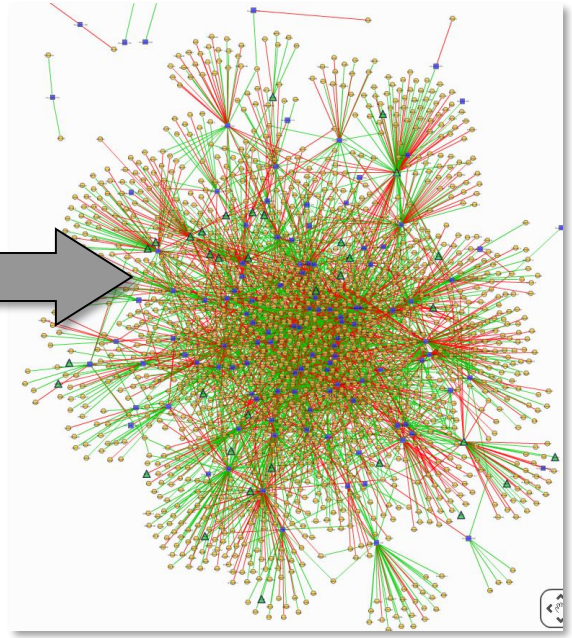


Dynamic network



Integration function



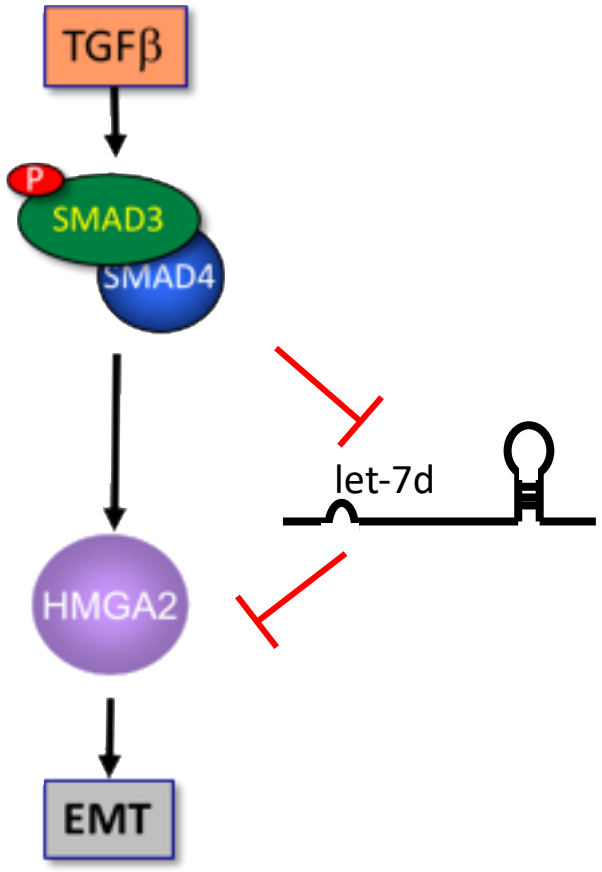


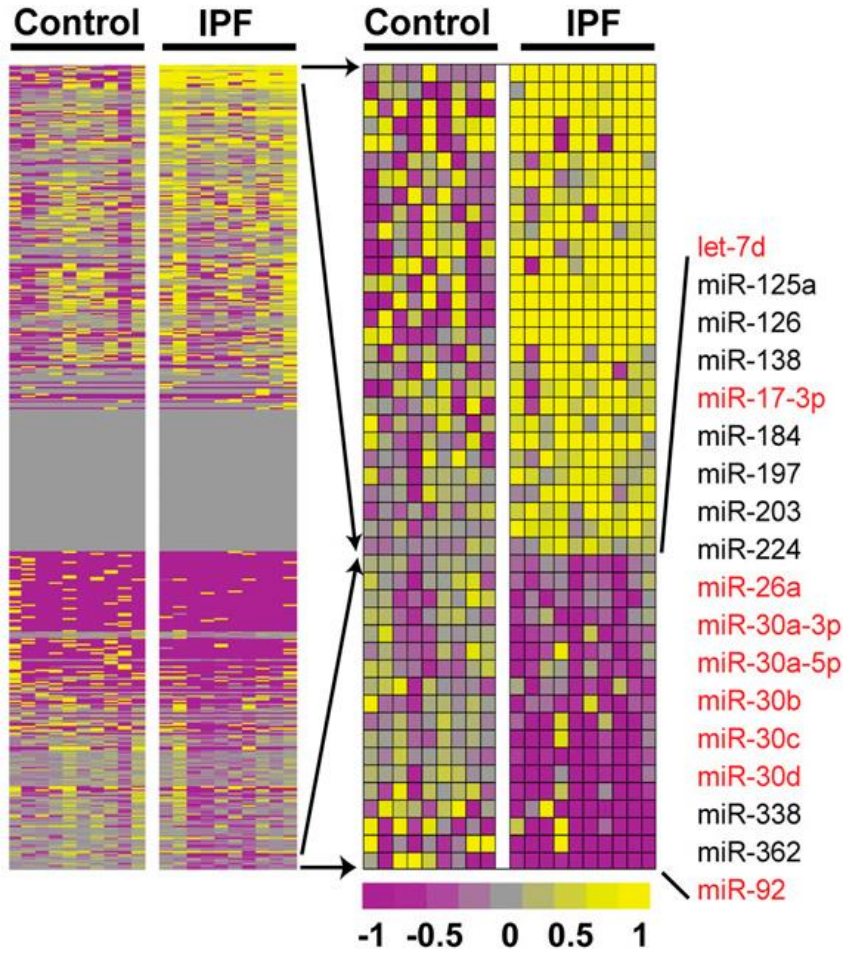
Transcription factor
microRNA
Other gene

Huang, Athanassiou, Benos, 2011, *Nucl Acids Res*

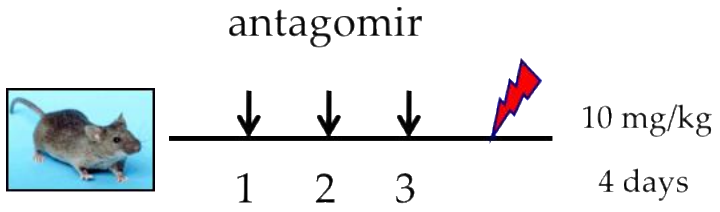
Discovery of important network module in Idiopathic Pulmonary Fibrosis (IPF)

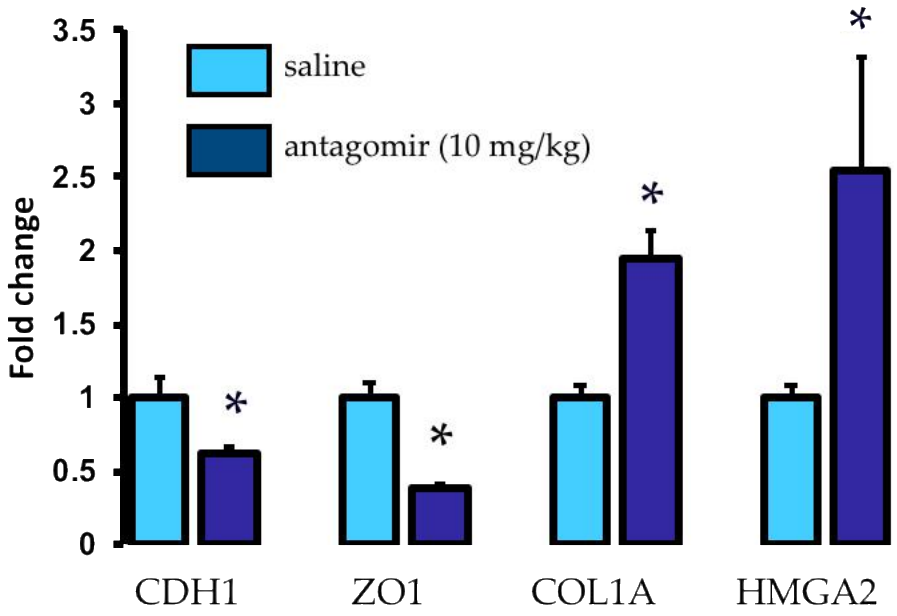
Downregulation of let-7d in IPF patients and in mice

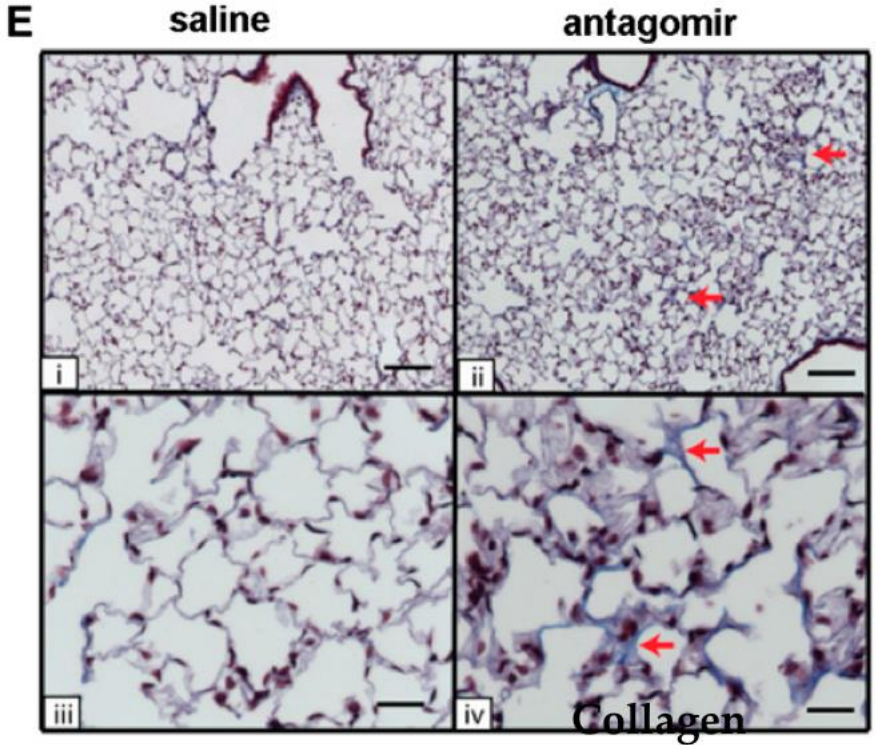




Downregulation of let-7d in IPF patients and in mice







Pandit, Corcoran, …, Benos, Kaminski, 2010, *Am J Resp Critic Care Med*

Correlations: what can and can not do

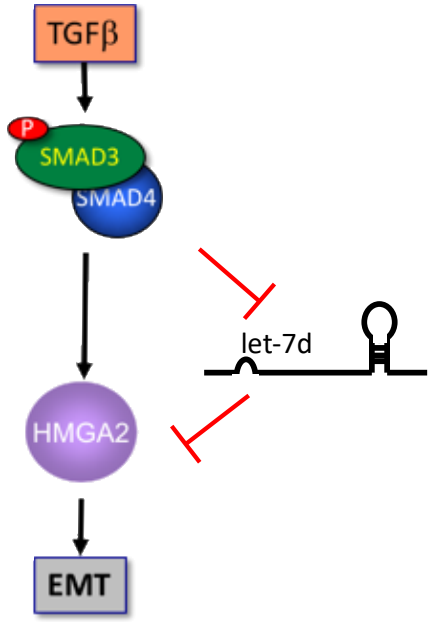
✓They are easy to calculate and intuitive and can be very useful

✓Provide all variables possibly related to our target variable

- …and then some

X Generate many ”false positive” edges

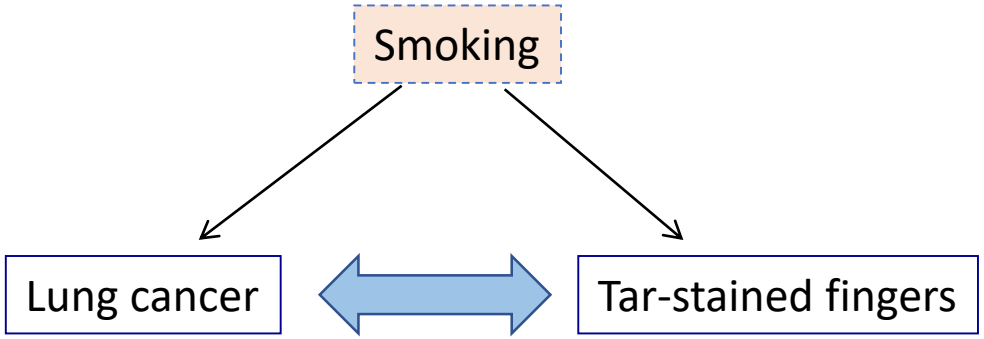
- In the previous example, TGF-β and EMT were also correlated (pairwise) to let-7d
- We needed prior biological knowledge to guide experiments



- Correlation vs causation
  - Causation → Correlation
  - Correlation **does not** prove causation (intervening experiments)
  - Example: smoking in the 50s

Correlation does not (always) imply causation

- A physician in the 50s may have noticed



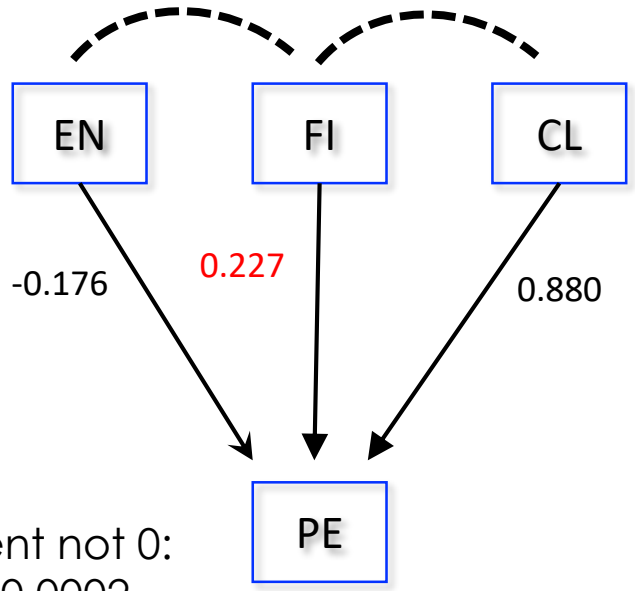
These is no causal link between these variables!

Regression models... (should be used with caution)

$$\hat{Y} = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \varepsilon$$

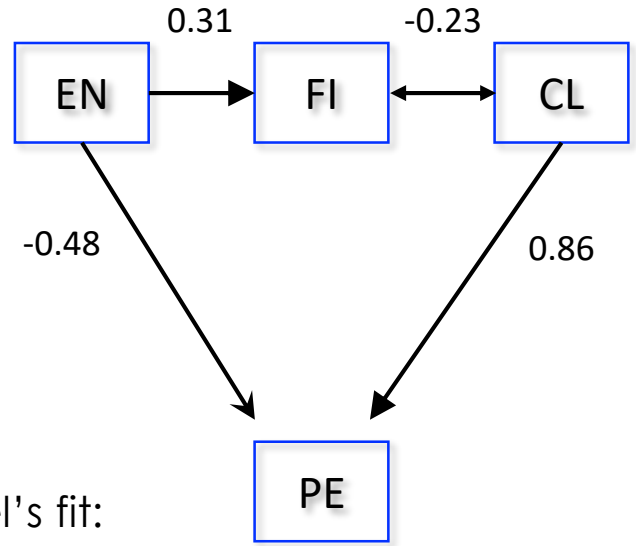
$$\hat{Y} = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \beta_{age} x_{age} + \beta_{smk} x_{smk} + \cdots + \varepsilon$$

REGRESSION MODEL



Coefficient not 0:
p-value=0.0002

CAUSAL MODEL



Model's fit:
p-value=0.94

Data from: *American Sociological Review*, 1984, vol 49, pp. 141-146

Slide modified from Richard Scheines

Regressions: what can and can not do

✓ They are intuitive and flexible

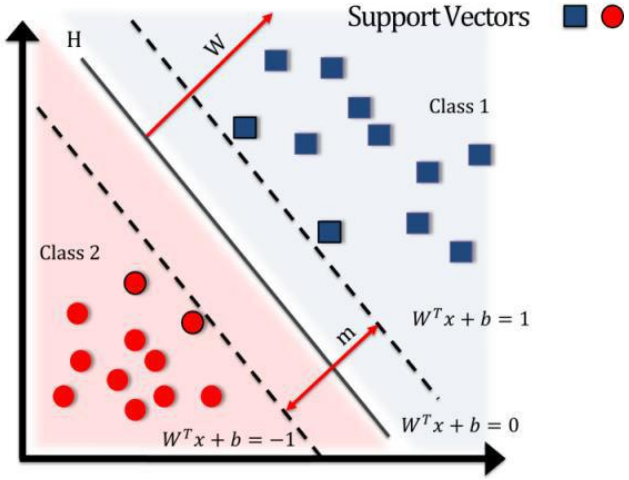
✓ Relatively fast to calculate

✓ Provide relative contributions of all predictors to the target variable

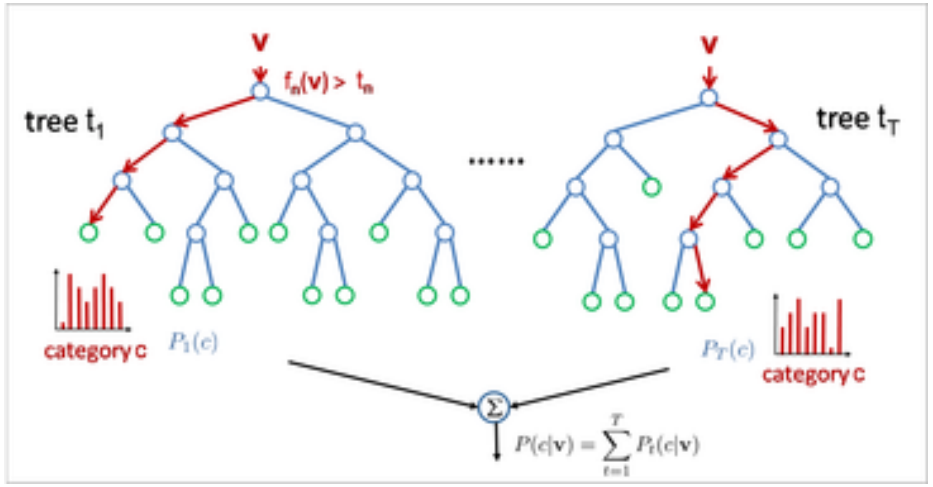
X In practice, it is not easy to implement interactive terms on predictors when number of predictors is large

- This may result in misleading coefficients

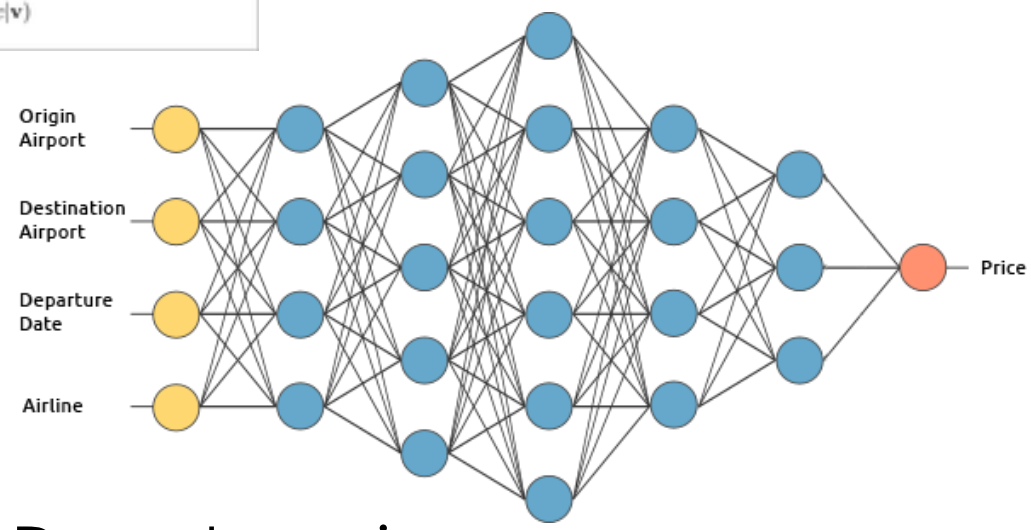
Some machine learning methods



SVM



RF



Deep Learning

ML “black box” methods: what can and can not do

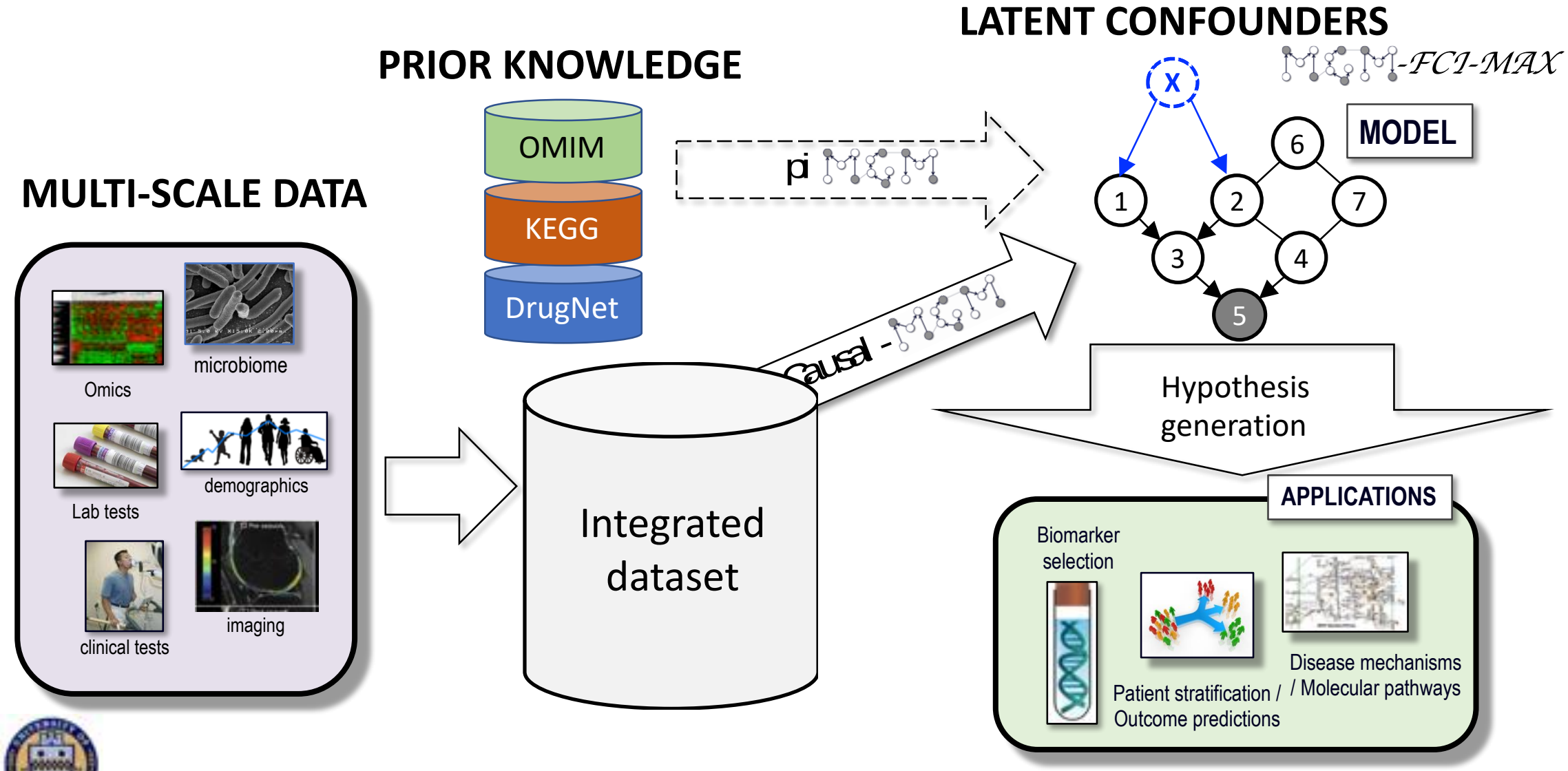
✓ Can model non-linear effects

✓ Very good for classification purposes (given enough data)

X They typically require large amounts of data

X Interpretability is not straightforward

Researcher dream analysis pipeline



Overview of the talk

- Discuss the probabilistic graphical models (PGMs) approach
  - What PGMs are / does it matter what type of variables I have?
  - How can we train them and interpret the results (*with caution!)*
  - How can we incorporate prior information

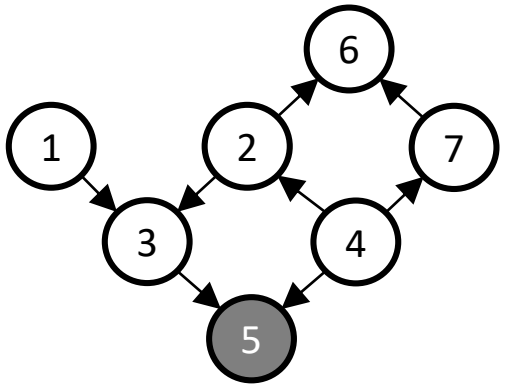
- Applications of graphical models in biomedical and clinical research
  - Clinical: Predicting lung cancer from low-dose CT scan and clinical data
  - Personalized medicine: A SNP that predicts response to chemotherapy
  - Clinical: Determinants of longitudinal lung function decline in COPD patients
  - Microbiome: Microbiota and clinical variables that predict culture positivity in lung ICU patients

What PGMs are: some definitions

- A **graph** consists of a set of nodes (variables), some of which are connected through edges
  - Edge connections imply information transfer
  - Two variables are connected when they have unique information for each other, not present in any other variable

- **Probabilistic graphical model** (**PGM**) is a model of the data in which a graph represents the conditional (in)dependencies between variables
  - PGMs can be undirected or directed
  - Undirected: easier to calculate, but contain FP edges

- **Causal graphs** are directed acyclic graphs (DAGs)

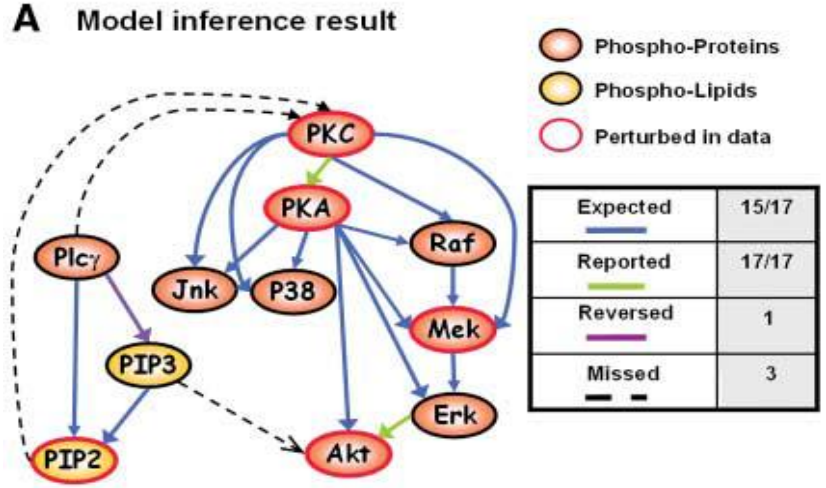


History of PGMs and past successes

- The development of PGMs started in mid-90s
- First books published in 2000

- Application of Bayesian networks to infer gene regulatory networks in yeast. [Friedman, Science, 2004]

- Application of causal learning methods proteomics data [Sachs et al, 2005]



Inferring Cellular Networks Using Probabilistic Graphical Models

Nir Friedman

...be cellular networks from ...ology. Probabilistic graph- ...insights from the resulting ...of complex cellular net- ...ased on well-understood ...odel-based methodology ...abilities are illustrated by

...erates predictions of system behavior under different conditions (as reflected by observations) and illuminates the roles of various system components in these behaviors. We focus on probabilistic models, which use stochasticity to account for measurement noise, variability in the biological system, and aspects of the system that are not captured by the model.

Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data

Karen Sachs,[1]* Omar Perez,[2]* Dana Pe'er,[3]* Douglas A. Lauffenburger,[1]† Garry P. Nolan[2]†

Machine learning was applied for the automated derivation of causal influences in cellular signaling networks. This derivation relied on the simultaneous measurement of multiple phosphorylated protein and phospholipid components in thousands of individual primary human immune system cells. Perturbing these cells with molecular interventions drove the ordering of connections between pathway components, wherein Bayesian network computational methods automatically elucidated most of the traditionally reported signaling relationships and predicted novel interpathway network causalities, which we verified experimentally. Reconstruction of network models from physiologically relevant primary single cells might be applied to understanding native-state tissue signaling biology, complex drug actions, and dysfunctional signaling in diseased cells.

History of PGMs and past successes

- Eric Schadt applies causal graphs for identification of causal SNPs [Schadt et al, Nat Genet, 2005]

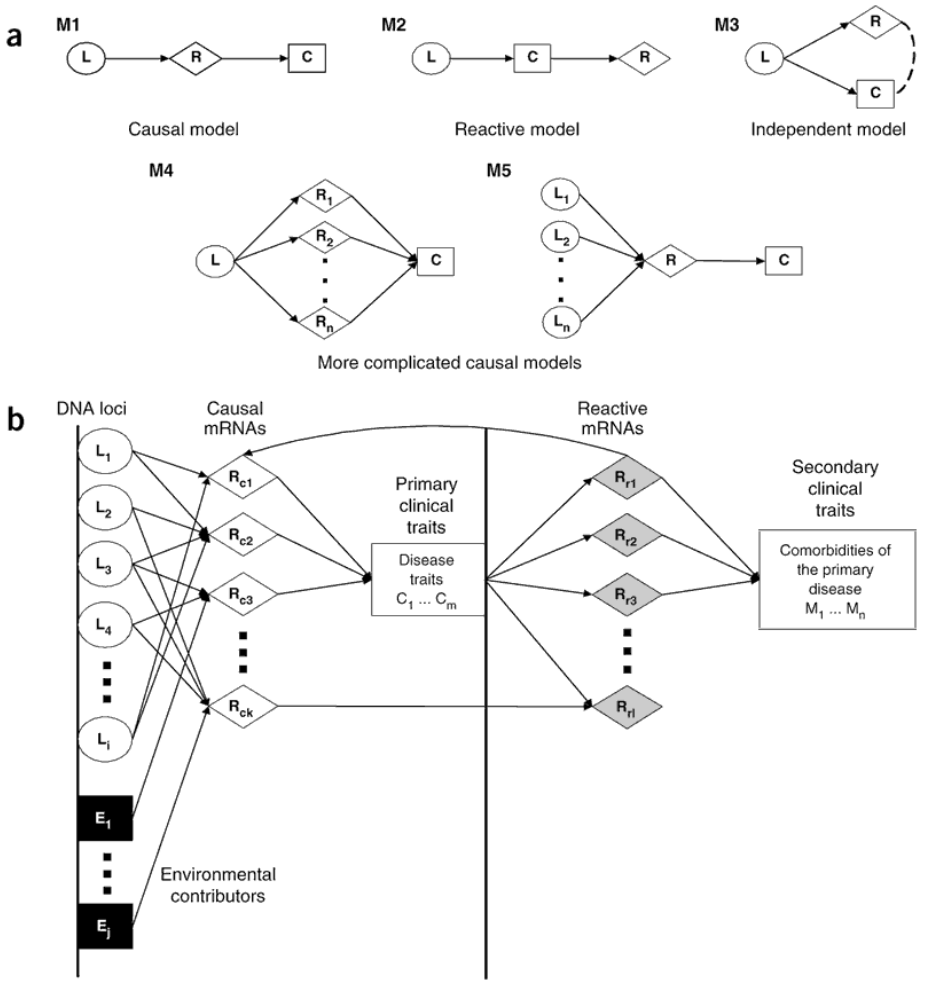
ARTICLES

nature genetics

An integrative genomics approach to infer causal associations between gene expression and disease

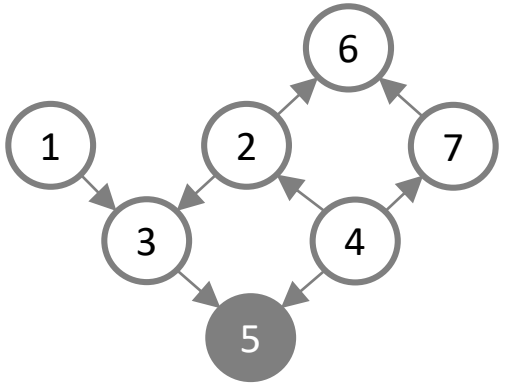
Eric E Schadt[1], John Lamb[1], Xia Yang[2], Jun Zhu[1], Steve Edwards[1], Debraj GuhaThakurta[1], Solveig K Sieberts[1], Stephanie Monks[3], Marc Reitman[4], Chunsheng Zhang[1], Pek Yee Lum[1], Amy Leonardson[1], Rolf Thieringer[5], Joseph M Metzger[6], Liming Yang[6], John Castle[1], Haoyuan Zhu[1], Shera F Kash[7], Thomas A Drake[8], Alan Sachs[1] & Aldons J Lusis[2]

A key goal of biomedical research is to elucidate the complex network of gene interactions underlying complex traits such as common human diseases. Here we detail a multistep procedure for identifying potential key drivers of complex traits that integrates DNA-variation and gene-expression data with other complex trait data in segregating mouse populations. Ordering gene expression traits relative to one another and relative to other complex traits is achieved by systematically testing whether variations in DNA that lead to variations in relative transcript abundances statistically support an independent, causative or reactive function relative to the complex traits under consideration. We show that this approach can predict transcriptional responses to single gene–perturbation experiments using gene-expression data in the context of a segregating mouse population. We also demonstrate the utility of this approach by identifying and experimentally validating the involvement of three new genes in susceptibility to obesity.

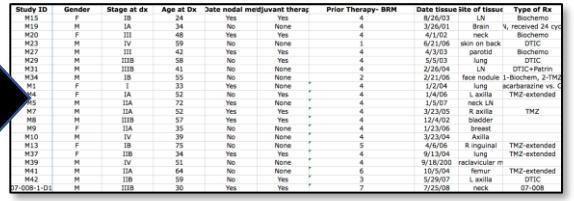


PGM underlying assumption: a causal graph generates the data

DATA GENERATING GRAPH

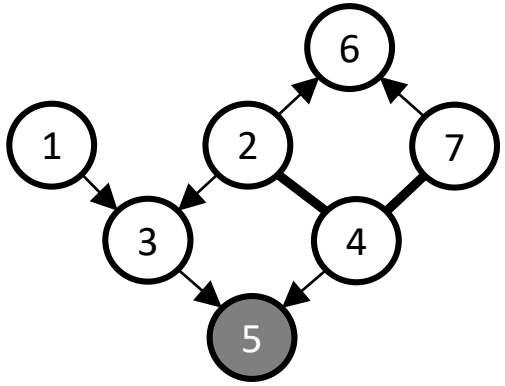


DATA (OBSERVED)



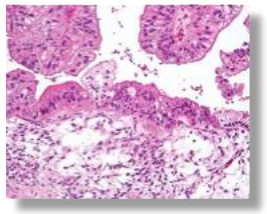
Study ID	Gender	Stage at dx	Age at Dx	Date nodal med	Juvant therap	Prior Therapy- BRM	Date tissue	Site of tissue	Type of Rx
M15	F	IB	24	Yes	Yes	4	8/26/03	LN	Biochemo
M19	M	IA	34	No	None	4	3/26/01	Brain	¼, received 24 cyc
M20	F	III	48	Yes	Yes	4	4/1/02	neck	Biochemo
M23	M	IV	59	No	None	1	6/21/06	skin on back	DTIC
M27	M	III	42	Yes	Yes	4	4/3/03	parotid	Biochemo
M29	M	IIIB	58	No	Yes	4	5/5/03	lung	DTIC
M31	M	IIIB	41	No	None	4	2/26/04	LN	DTIC+Patrin
M34	M	IB	55	No	None	2	2/21/06	face nodule	1-Biochem, 2-TMZ
M1	F	I	33	Yes	None	4	1/2/04	lung	acarbarazine vs. C
M4	F	IA	52	No	Yes	4	1/4/06	L axilla	TMZ-extended
M5	M	IIA	72	Yes	None	4	1/5/07	neck LN	
M7	M	IIA	52	Yes	Yes	4	3/23/05	R axilla	TMZ
M8	M	IIIB	57	Yes	Yes	4	12/4/02	bladder	
M9	F	IIA	35	No	None	4	1/23/06	breast	
M10	M	IV	39	No	None	4	3/23/04	Axilla	
M13	F	IB	75	No	None	5	4/6/06	R inguinal	TMZ-extended
M37	F	IIB	34	Yes	Yes	4	9/13/04	lung	TMZ-extended
M39	M	IV	51	No	None	4	9/18/200	raclavicular m	
M41	M	IIA	64	No	None	6	10/5/04	femur	TMZ-extended
M42	M	IIB	59	No	Yes	3	5/29/07	L axilla	DTIC
07-008-1-D1	M	IIIB	30	Yes	Yes	7	7/25/08	neck	07-008

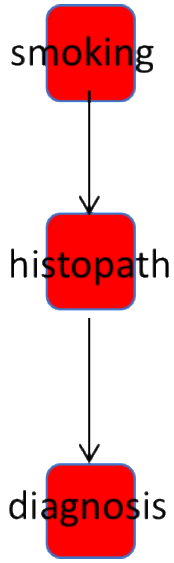
LEARNED GRAPH

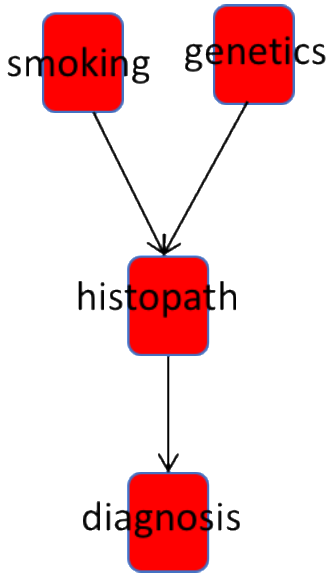


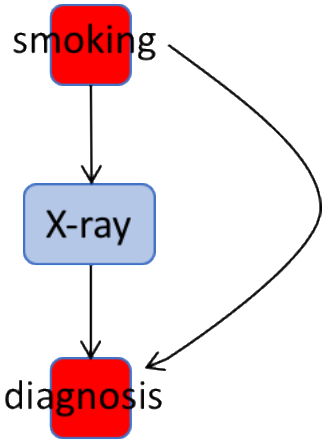
Nodes = variables
Edges = direct (causal) associations between variables

Graph adjacency learning using conditional independencies

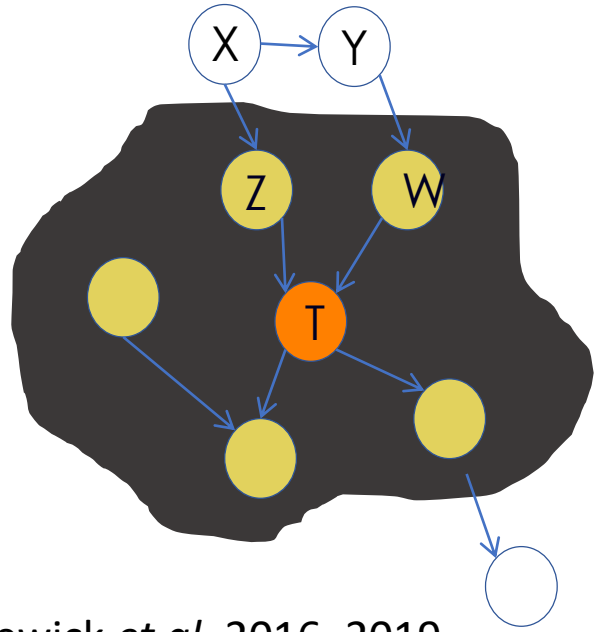








Properties and Drawbacks of Graphical Models

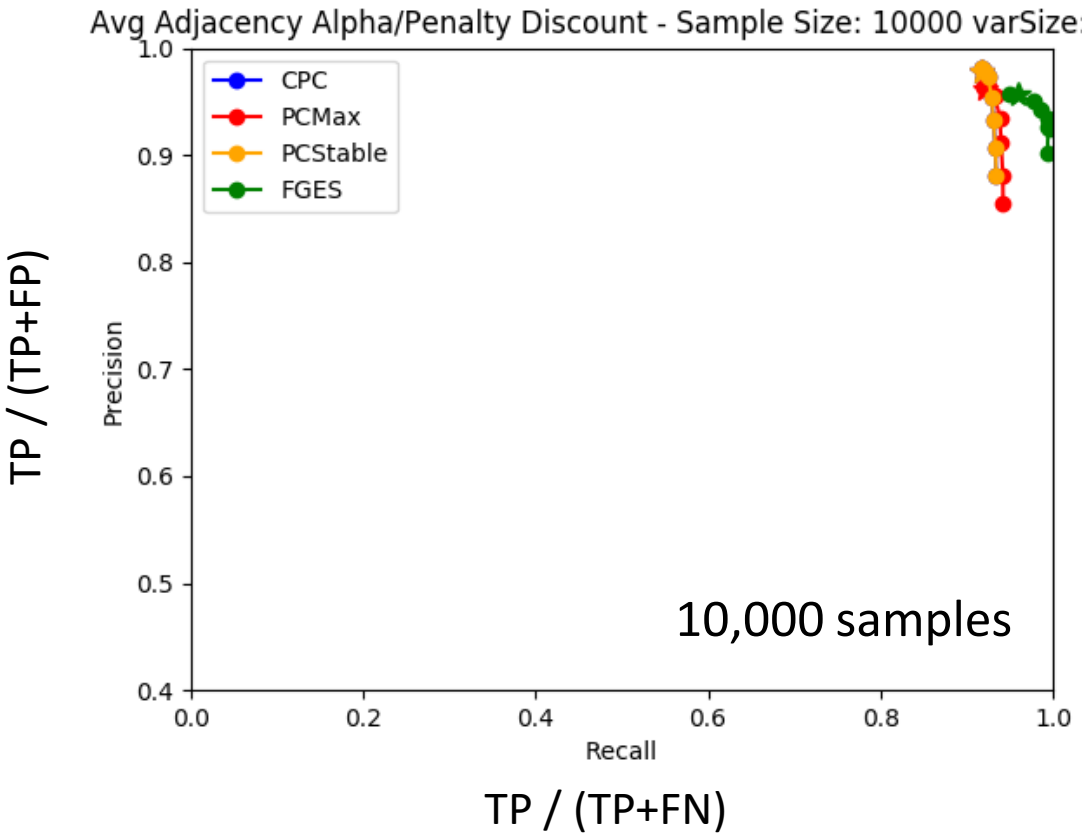


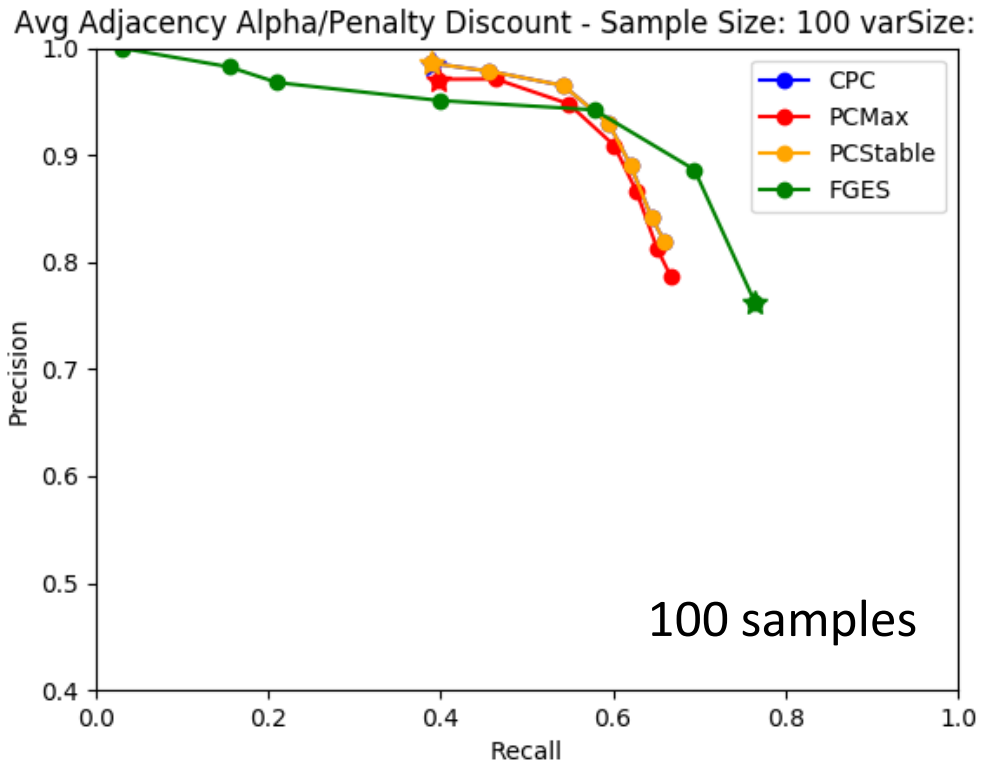
- They can distinguish between direct and indirect effects
- They are *asymptotically* correct. 😊
- The output graph can be used for predictive models

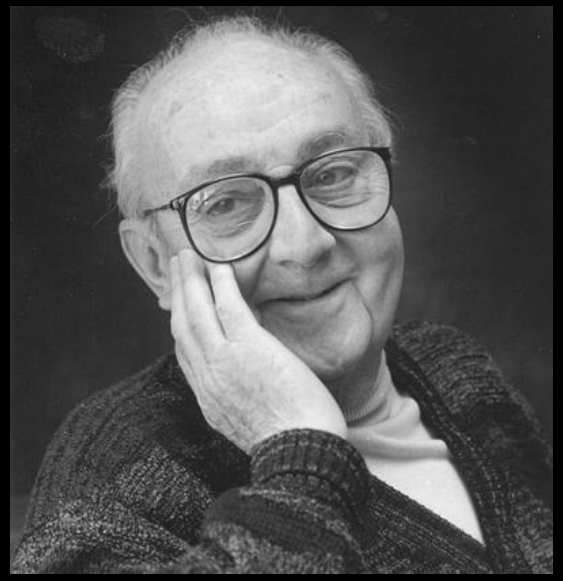
- They have some non-realistic assumptions (but they can be relaxed)
  - Variables are either all continuous or all discrete ⟸ Sedgewick *et al*, 2016, 2019
  - All common causes are measured (no latent confounders) ⟸ Raghu *et al*, ACM SIGKDD 2017
  - All continuous variables should be normally distributed
  - There are no cycles in the graph

- Additional considerations
  - Relatively slow (heuristics are needed) ⟸ Sedgewick *et al*, 2016, 2019
  - Parameter setting ⟸ Raghu, Poon, Benos, ACM SIGKDD 2018; Raghu *et al*, ACM SIGKDD 2019
  - Incorporating priors ⟸ Manatakis, Raghu, Benos, 2018

Edge prediction accuracy in DAGs (100 nodes, Gaussian)



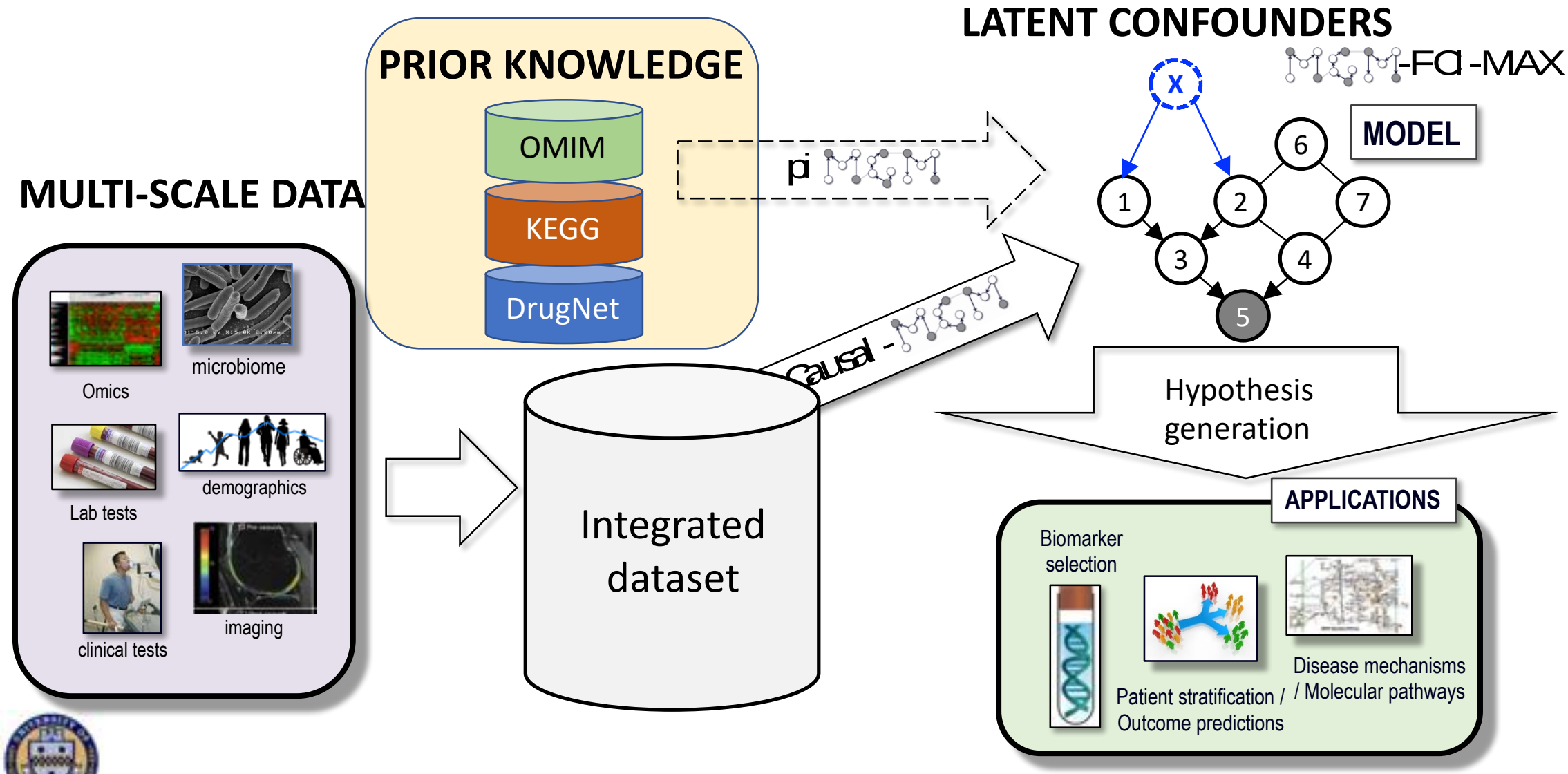




"Essentially all models are wrong,
but some are useful"

George E.P. Box

Researcher dream analysis pipeline

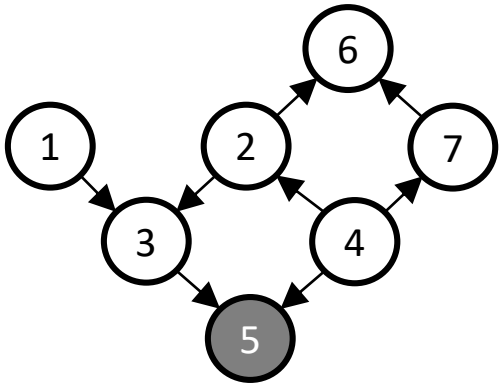


piMGM: MGM with prior information

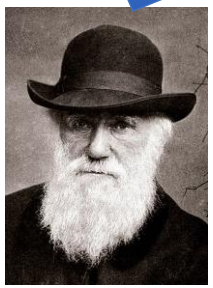
Goals

1. Estimate Reliability of each “expert”
2. Construct a properly weighted combined prior
3. Learn an informed undirected model using this prior

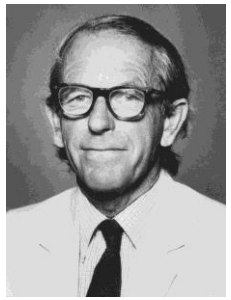
DATA GENERATING GRAPH



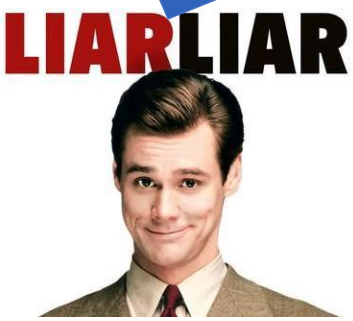
{6,7, 80%}
{4,7, 95%}



{1,3, 100%}
{5,4,90%}



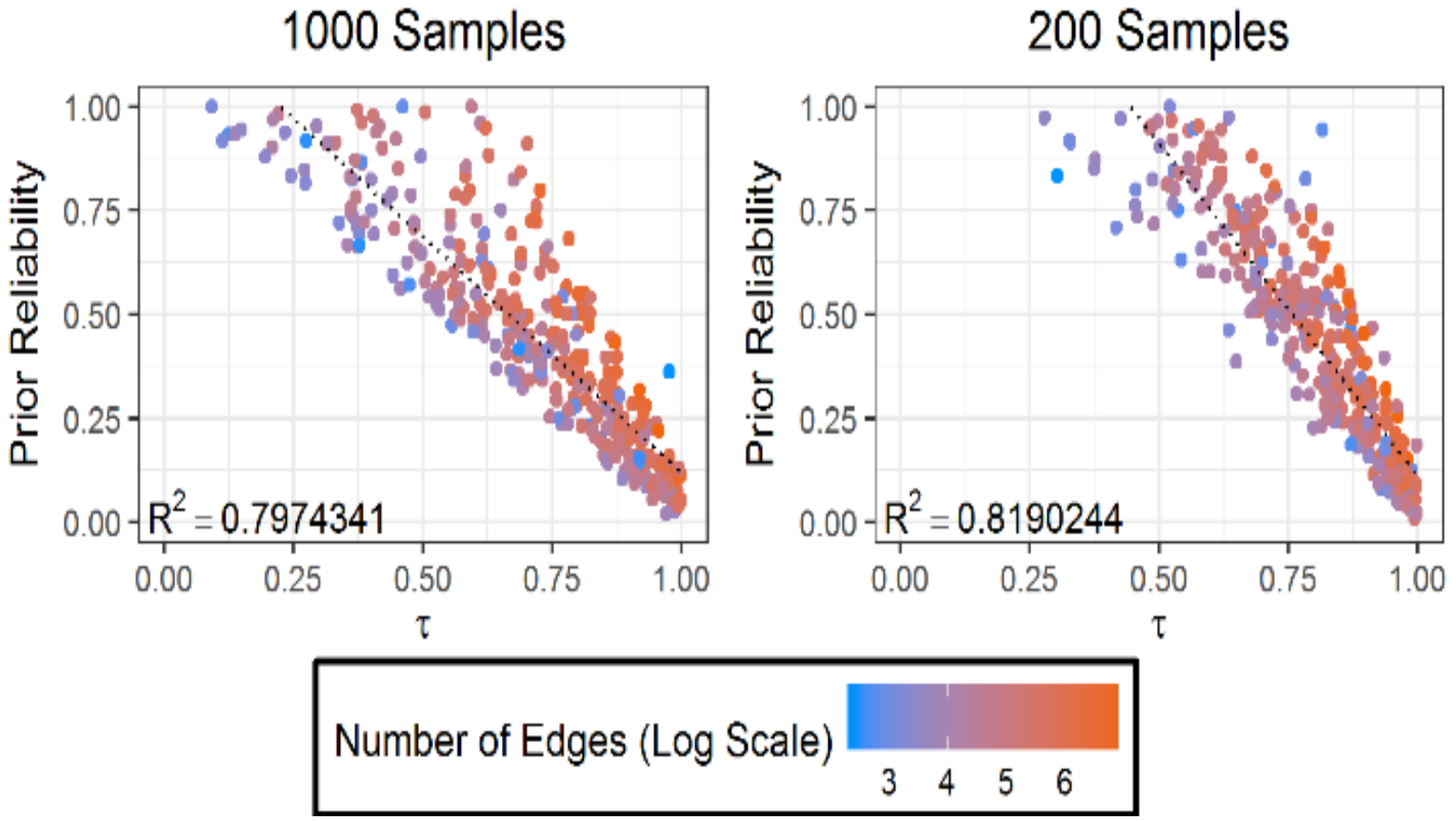
{5,7, 95%}
{1,6,90%}



Slide adapted from Vineet Raghu, Benos Lab

piMGM Correctly Evaluates the Reliability of Experts

Vineet Raghu



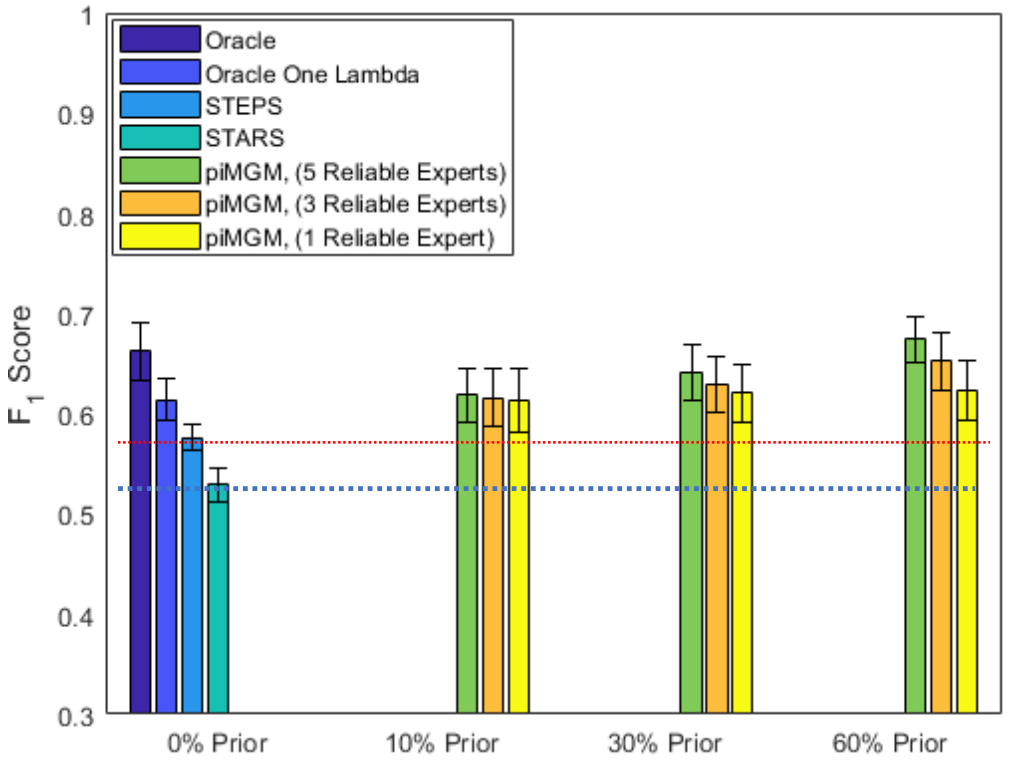
Manatakis*, Raghu*, Benos, 2018, *Bioinformatics*.

Slide courtesy of Vineet Raghu, Benos Lab

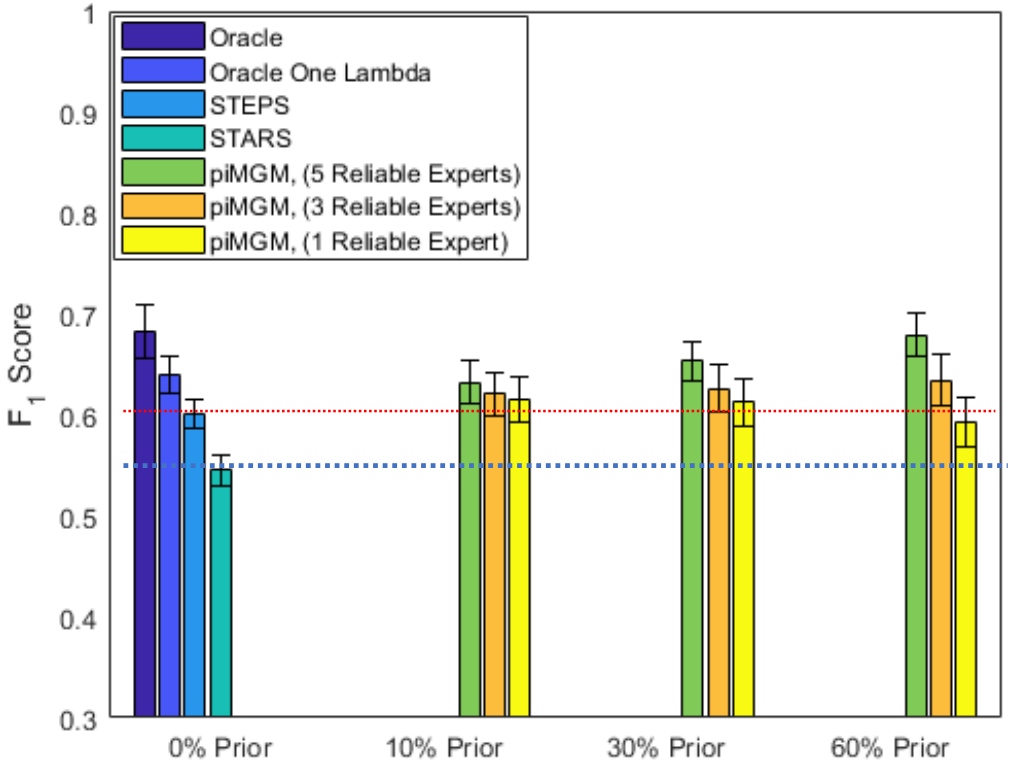
piMGM Overcomes Unreliable Priors

Vineet Raghu

Experts Give Prior for Real Edges



Experts Give Prior for Any Edge



Manatakis*, Raghu*, Benos, 2018, *Bioinformatics*.

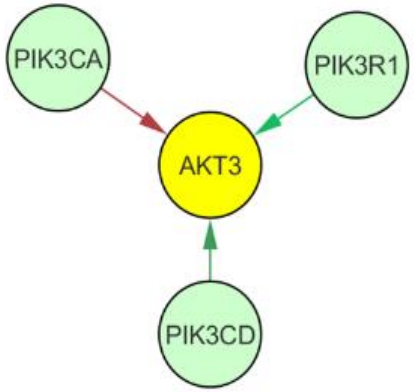
Slide courtesy of Vineet Raghu, Benos Lab

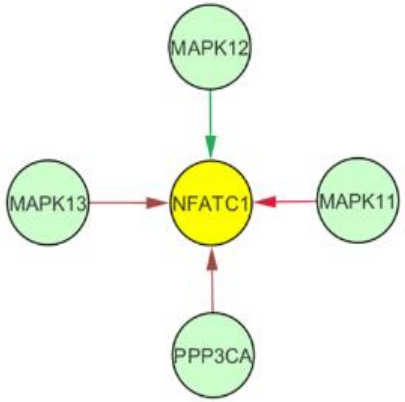
Use “expert evaluation” as a way to evaluate pathway significance

- Use the expert evaluation method to:
  - Identify active pathways in disease (by evaluating *edge* presence)
  - Learn high confidence gene-gene interactions

- Example: breast cancer (TCGA), ER+ and ER- cases

Pathway	p-value (ER+)	p-value (ER-)	Reference
Glutathione Metabolism	0.507	0.091	(Lien, et al., 2016)
Glycolysis	**0.000**	0.129	(Schramm, et al., 2010)
Neurotrophin signaling	0.702	0.074	(Patani, et al., 2011)
Notch signaling	**0.000**	0.223	(Hossain, et al., 2017)
Pentose Phosphate	**0.025**	0.239	(Cha, et al., 2017)
B Cell Receptor signaling	0.141	**0.004**	(Hill, et al., 2011)
Insulin signaling	0.098	0.384	
T cell receptor signaling	0.507	0.058	





Manatakis*, Raghu*, Benos, 2018, *Bioinformatics*.

Summary of piMGM results

- ✓piMGM can accurately determine the reliability of prior information sources on simulated and real data
- ✓piMGM is resilient to unreliable priors when learning network structure
- ✓The benefits of using prior information to learn network structure are greatest in high-dimensional, low sample size cases

Manatakis, D.*, Raghu, VK.*, and Benos, PV, 2018, *Bioinformatics*.

BENOS' LAB

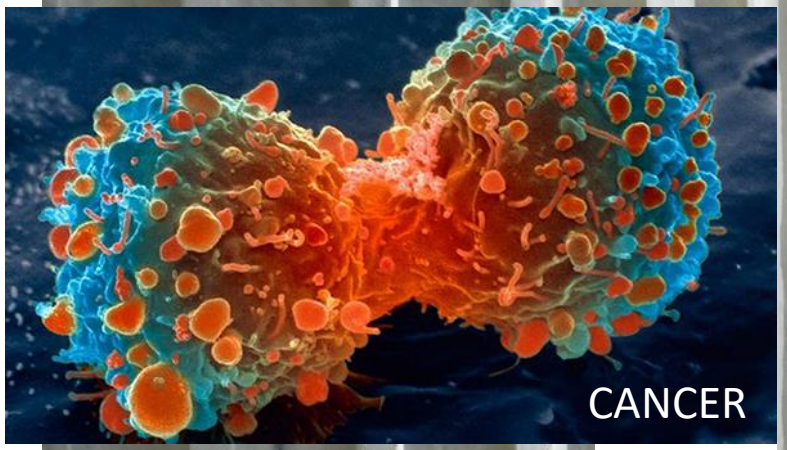


Image: NCI

Image: Stanford

CLINICAL RESEARCH

TRANSLATION

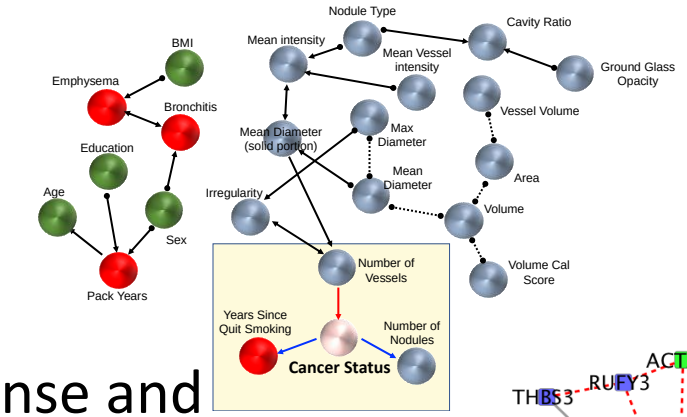
Overview of the talk

- Discuss the probabilistic graphical models (PGMs) approach
  - What PGMs are / does it matter what type of variables I have?
  - How can we train them and interpret the results (*with caution!)*
  - How can we incorporate prior information

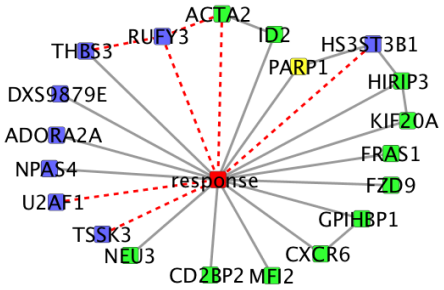
- **Applications** of graphical models in biomedical and clinical research
  - Clinical: Predicting lung cancer from low-dose CT scan and clinical data
  - Personalized medicine: A SNP that predicts response to chemotherapy
  - Clinical: Determinants of longitudinal lung function decline in COPD patients
  - Microbiome: Microbiota and clinical variables that predict culture positivity in lung ICU patients

Applications of ***CausalMGM*** in (Bio)Medicine

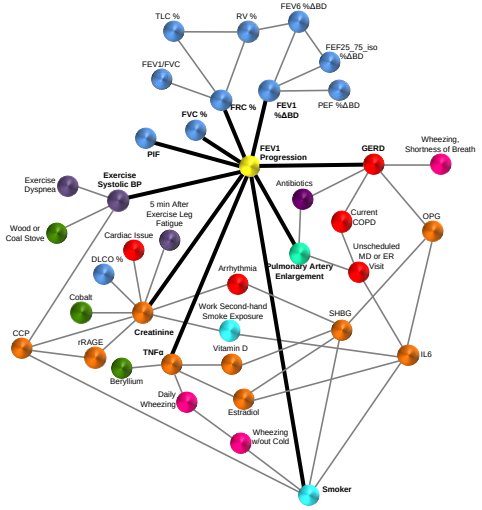
- Early disease diagnosis
  - Lung cancer detection (LDCT scans + comorbidities)



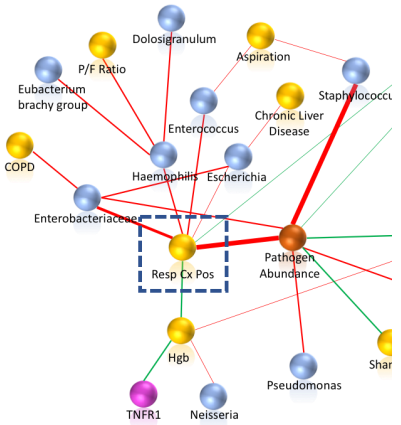
- Identifying biomarkers indicative of treatment response and alternative treatments
  - Melanoma chemotherapy (multi-omics data)



- Identifying factors affecting disease progression
  - FEV_1 decline in COPD patients (clinical variables)



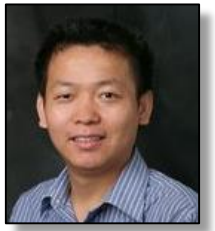
- Disease diagnosis
  - Pneumonia detection in ICU (microbiome + clinical data)



Vineet Raghu

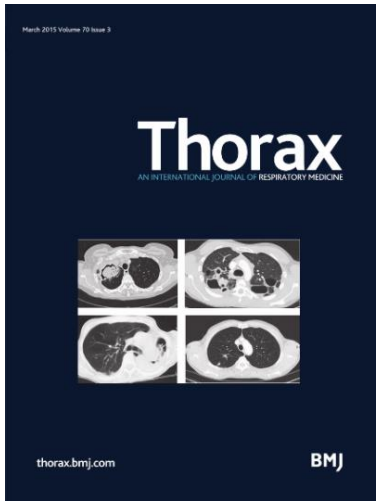
In collaboration with:

David Wilson MD



Jiantao Pu PhD

Factors determining malignancy of a lung nodule from low-dose CT scan and clinical data



Feasibility of lung cancer prediction from low-dose CT scan and smoking factors using causal models

Vineet K. Raghu[1,2], Wei Zhao[3], Jiantao Pu[3], Joseph K. Leader[3], Renwei Wang[4], James Herman[5], Jian-Min Yuan[4,6], Panayiotis V. Benos[1,2*], David O. Wilson[7]

[1]Department of Computer Science, University of Pittsburgh, Pittsburgh, PA
[2]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA
[3]Department of Radiology, University of Pittsburgh, Pittsburgh, PA
[4]UPMC Hillman Cancer Center, Pittsburgh, PA
[5]Division of Hematology, Oncology, Department of Medicine, University of Pittsburgh, Pittsburgh, PA
[6]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA
[7]Division of Pulmonary, Allergy and Critical Care Medicine, School of Medicine, University of Pittsburgh, Pittsburgh, PA

[*]Corresponding author

Low dose CT scan screening reduces lung cancer mortality

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 AUGUST 4, 2011 VOL. 365 NO. 5

Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening

The National Lung Screening Trial Research Team*

ABSTRACT

BACKGROUND

The aggressive and heterogeneous nature of lung cancer has thwarted efforts to reduce mortality from this cancer through the use of screening. The advent of low-dose helical computed tomography (CT) altered the landscape of lung-cancer screening, with studies indicating that low-dose CT detects many tumors at early stages. The National Lung Screening Trial (NLST) was conducted to determine whether screening with low-dose CT could reduce mortality from lung cancer.

The members of the writing team (who are listed in the Appendix) assume responsibility for the integrity of the article. Address reprint requests to Dr. Christine D. Berg at the Early Detection Research Group, Division of Cancer Prevention, National Cancer Institute, 6130 Executive Blvd., Suite 3112, Bethesda, MD 20892-7346, or at bergc@mail.nih.gov.

RESULTS

The rate of adherence to screening was more than 90%. The rate of positive screening tests was 24.2% with low-dose CT and 6.9% with radiography over all three rounds. A total of 96.4% of the positive screening results in the low-dose CT group and 94.5% in the radiography group were false positive results. The incidence of lung cancer was 645 cases per 100,000 person-years (1060 cancers) in the low-dose CT group, as compared with 572 cases per 100,000 person-years (941 cancers) in the radiography group (rate ratio, 1.13; 95% confidence interval [CI], 1.03 to 1.23). There were 247 deaths from lung cancer per 100,000 person-years in the low-dose CT group and 309 deaths per 100,000 person-years in the radiography group, representing a relative reduction in mortality from lung cancer with low-dose CT screening of 20.0% (95% CI, 6.8 to 26.7; P=0.004). The rate of death from any cause was reduced in the low-dose CT group, as compared with the radiography group, by 6.7% (95% CI, 1.2 to 13.6; P=0.02).

CONCLUSIONS

Screening with the use of low-dose CT reduces mortality from lung cancer. (Funded by the National Cancer Institute; National Lung Screening Trial ClinicalTrials.gov number, NCT00047385.)

- Follow-up CTs
- Unnecessary invasive biopsies
  - with potential serious complications
- Anxiety
- Increased healthcare costs

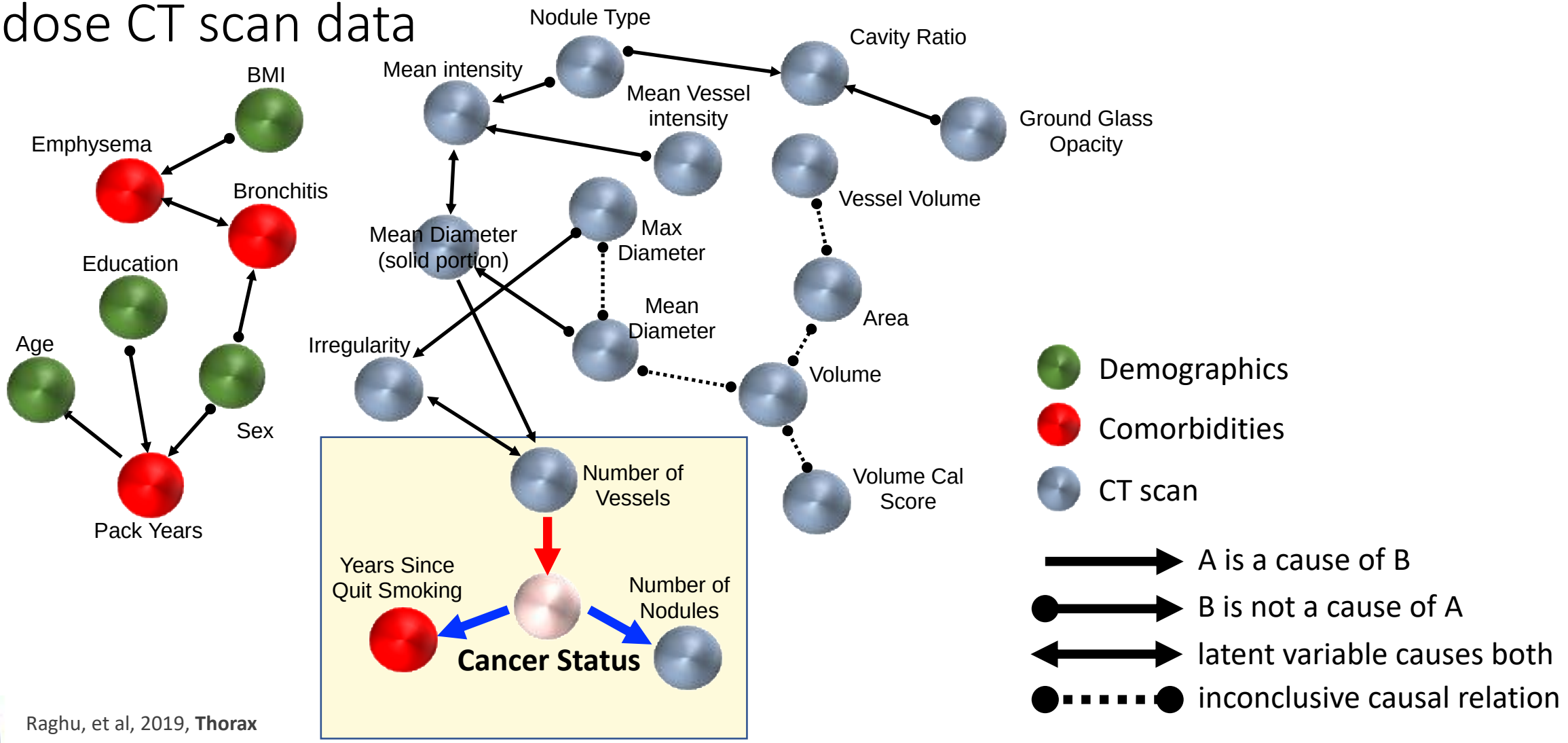
Pittsburgh Lung Screening (PLuSS) cohort

A. Training n=92	Lung cancer (n = 50)	Benign nodules (n = 42)	P value†
Male, n (%)	25 (50)	28 (67)	0.162
Age (years), mean (SD)	63.6 (7.1)	65.2 (6.9)	0.261
Current smoker, n (%)	32 (64)	19 (45)	0.111
Pack-Years, mean (SD)	60.35 (24.11)	61.81 (22.81)	0.766
Years since quit smoking, mean (SD)	1.52 (2.88)	3.25 (3.95)	0.020
Nodule size in diameter (mm), mean (SD)	13.43 (6.14)	9.74 (6.69)	0.007
Nodule number, n (%) °			0.203
Solid	28 (56)	34 (81)	
Non-solid/mixed	22 (44)	8 (19)	
Vessel number, mean (SD)	9.22 (9.48)	2.26 (2.21)	<0.0001

B. Validation (PLuSS-X) n=126	Lung cancer (n = 44)	Benign nodules (n = 82)	P value†
Male, n (%)	23 (52)	48 (59)	0.626
Age, mean, years (SD)	65.23 (9.62)	66.93 (7.54)	0.313
Current smoker, n (%)	37 (84)	36 (44)	<0.0001
Pack-Years, mean (SD)*	49.41 (22.79)	49.49 (22.0)	0.985
Years since quit smoking, mean (SD)	0.477 (1.50)	3.037 (4.33)	<0.0001
Nodule size in diameter (mm), mean (SD)	18.86 (7.12)	11.57 (5.76)	<0.0001
Nodule number, n (%) °			0.981
Solid	28 (78)	54 (68)	
Non-solid/mixed	8 (22)	25 (32)	
Vessel number, mean (SD)	18.57 (5.21)	3.02 (3.98)	<0.0001

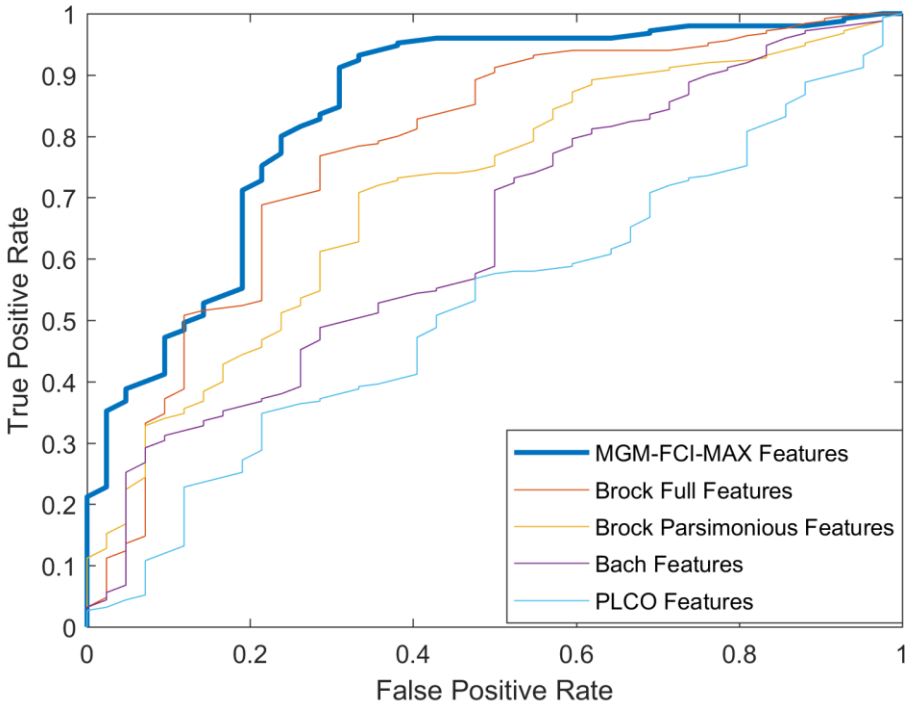
Raghu, et al, 2019, Thorax, *in print*

LCCM: a CausalMGM-based lung cancer predictor from low-dose CT scan data



Raghu, et al, 2019, **Thorax**

LCCM outperforms existing lung cancer predictors (cross-validation)

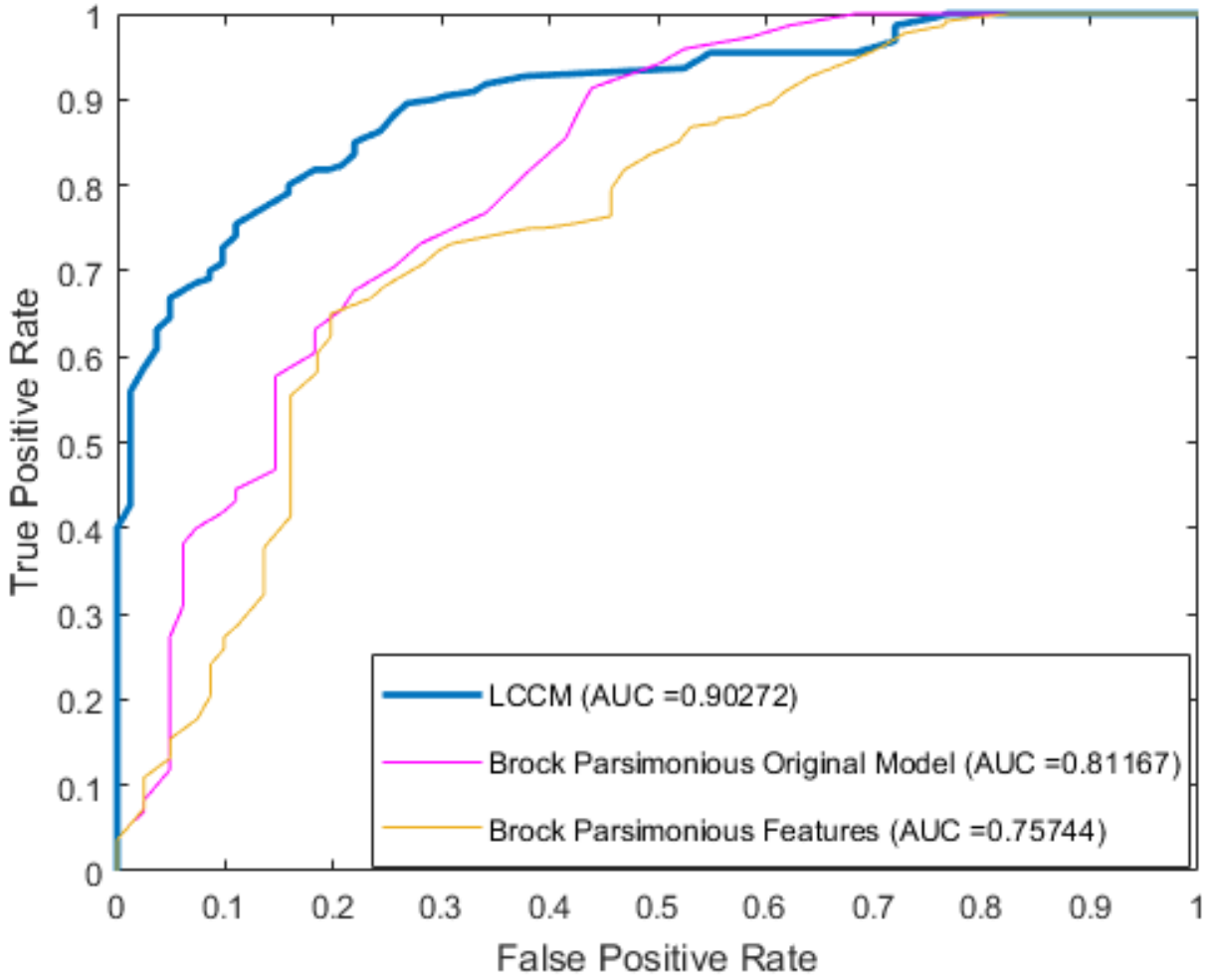


Model	No. of Features	AUC (95% CI)	p-value	Features Used
IGM-FCI-MAX eatures	3	0.882 (0.789, 0.975)	-	**Smoking:** Years Quit **Radiographic:** Nodule Count, Vessel Number
rock Full Features	8	0.792 (0.699,0.885)	0.16	**Demographics:** Age, Sex, Family History Ca **Comorbidities:** Emphysema **Radiographic:** Nodule Size, Nodule Type, Nodule Location, Nodule Count
rock Parsimonious eatures	3	0.700 (0.607,0.793)	0.01	**Demographics**: Sex **Radiographic:** Nodule Location, Nodule Size
ach Features	5	0.722 (0.629,0.815)	0.02	**Demographics**: Age, Sex **Smoking:** Cigarettes Per Day, Smoke Duration, Years Quit
LCO Features	10	0.5613 (0.412,0.701)	<0.001	**Demographics**: BMI, Education, Family History Ca, Race **Comorbidities**: Ca History, COPD **Smoking**: Duration, Intensity, Smoking Status, Years Quit

Predictors	Coefficient (95% CI)	*p*-value
Years since quit smoking	-0.178 (-0.349, -0.007)	0.041
Number of Vessels	0.238 (0.074, 0.510)	0.009
Number of Nodules	-0.203 (-0.325, -0.081)	0.001
Model Intercept	1.053	

Raghu, et al, 2019, Thorax, *in print*

LCCM outperforms existing lung cancer predictors (external cohort)

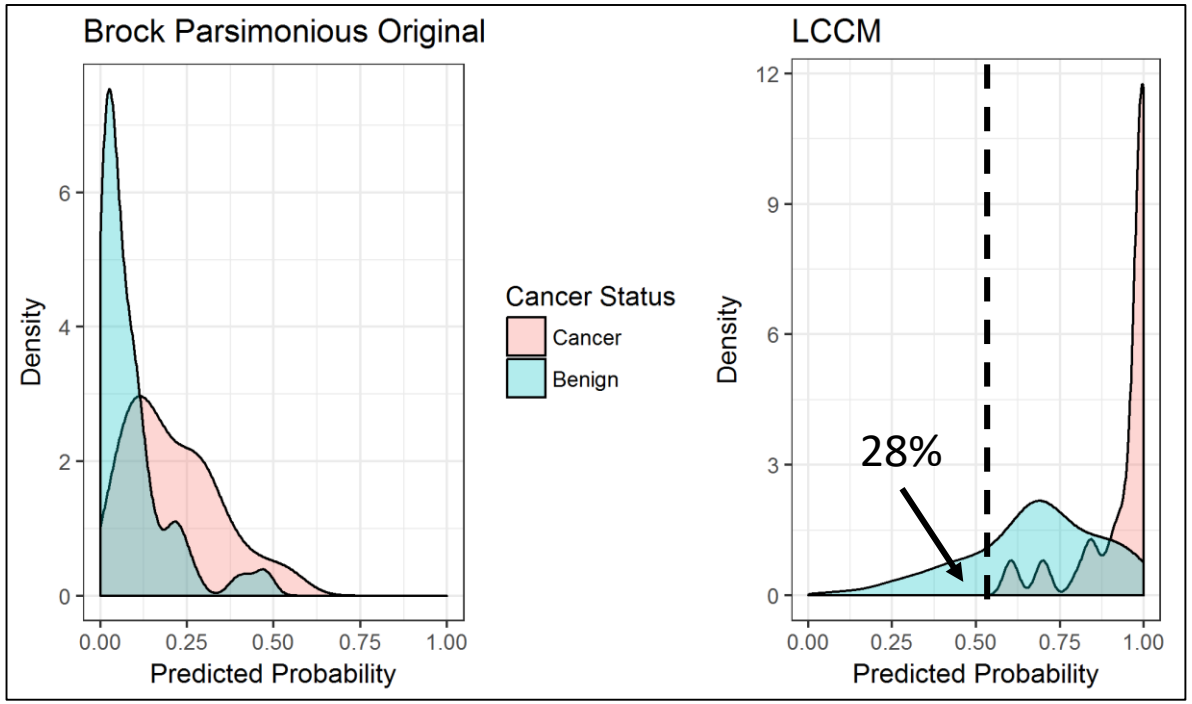


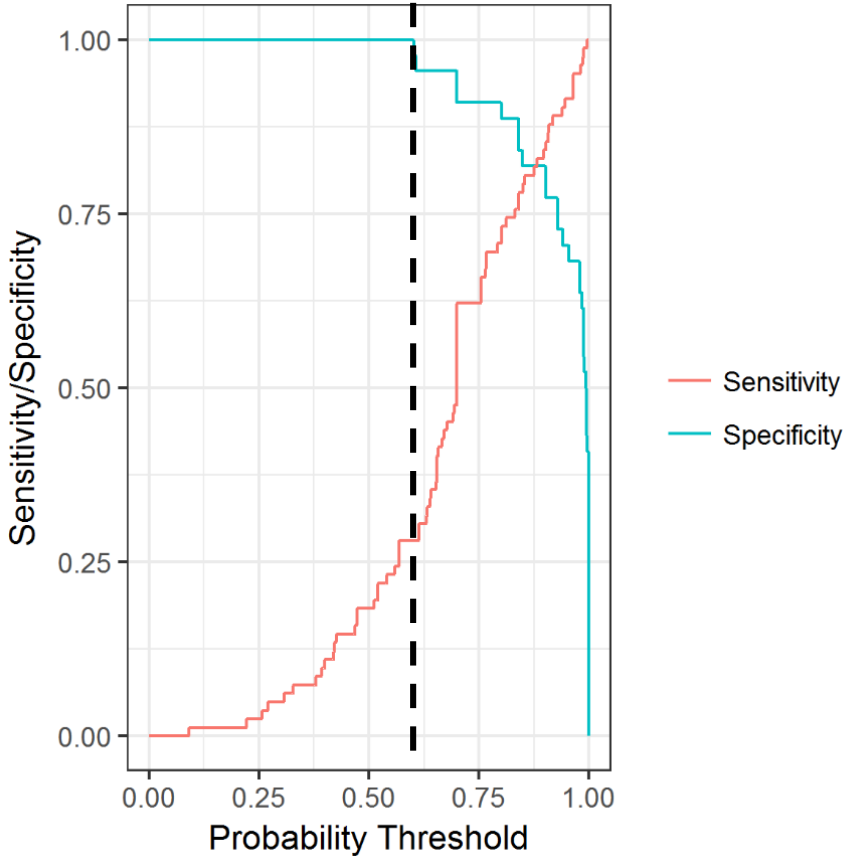
$p = 0.018$

$p < 0.01$

Raghu, et al, 2019, Thorax, *in print*

LCCM can help reduce unnecessary follow up screenings





Raghu, et al, 2019, Thorax, *in print*

Making some noise…

90.5 WESA Pittsburgh's NPR News Station — Donate

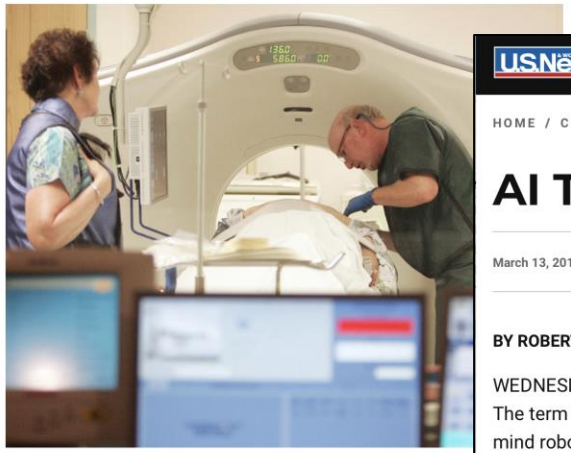
90.5 WESA Marketplace

Latest Popular News Series & Podcasts Schedule About Support Search

Artificial Intelligence Could Reduce False Positives In Lung Cancer Screenings

By KATHLEEN J. DAVIS • MAR 13, 2019

Tweet Share Google+ Email



In this June 3, 2010, file photo, Dr. Steven Birnbaum works with a patient in a CT scanner at Hampshire Medical Center in Nashua, N.H.

JIM COLE / AP

ScienceDaily®

Your source for the latest research news

Follow

Just In: Rise of Rel

ADVERTISEMENT

Free CRISPR Analysis Guide - Analyze Gene Editing Results

Step-by-Step guide for analyzing CRISPR editing results with ICE from Synthego. synthego.com

OPEN

S_D Health Tech Enviro Society

Science News

U.S.News CIVIC » Best Countries Best States Healthiest Communities Cities The Civic Report

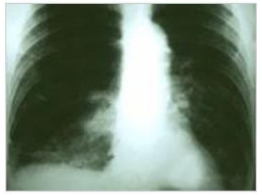
HOME / CIVIC / HEALTH NEWS

AI Takes Aim at Lung Cancer Screening

March 13, 2019, at 9:00 a.m.

BY ROBERT PREIDT, *HealthDay Reporter*

WEDNESDAY, March 13, 2019 (HealthDay News) -- The term artificial intelligence (AI) might bring to mind robots or self-driving cars. But one group of researchers is using a type of AI to improve lung cancer screening.



(HEALTHDAY)

Screening is important for early diagnosis and improved survival odds, but the current lung cancer screening method has a 96 percent false positive rate.

HEALTH IT ANALYTICS

xtelligent HEALTHCARE MEDIA

Home News Features

Population Health Precision Medicine Quality & Governance Tools & Strategies Analytics in Action

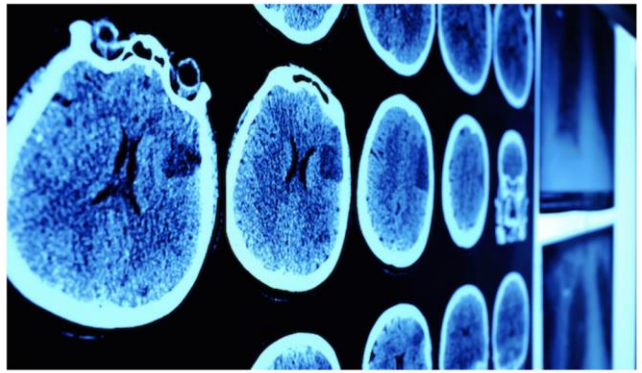
TELEHEALTH SUMMIT 2019 — June 6-7, 2019, Atlanta, GA, W Atlanta - Midtown

KEYNOTE: Ann Mond Johnson, CEO, American Telemedicine Association

TOOLS & STRATEGIES NEWS

UPMC Uses Machine Learning to Cut Lung Cancer False Positives

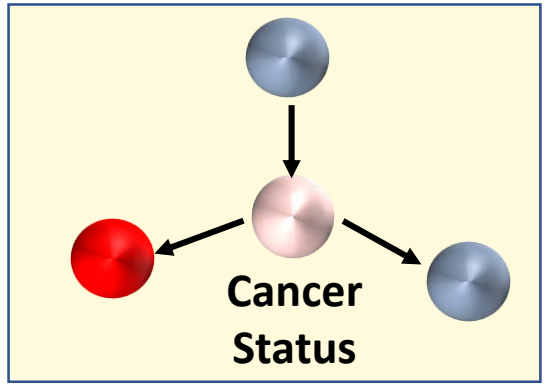
Using machine learning, researchers have significantly reduced the false positive rate for lung cancer diagnoses.

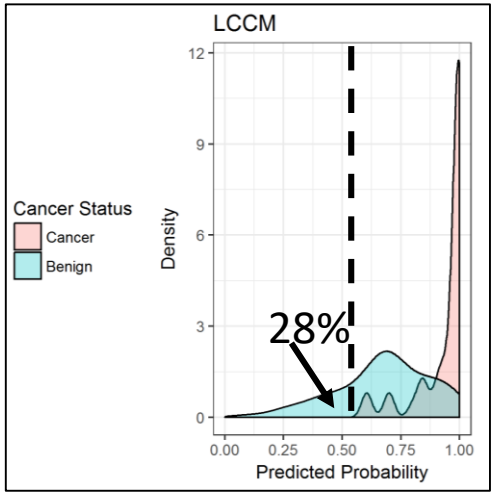


What we learned from the LCCM study?

- Vasculature around a nodule and total number of nodules are important discriminants of nodule status

- LCCM in the future may help reduce unnecessary follow up screens for 28% of the benign nodule subjects





AJ Sedgewick PhD

Disclosure:
US Patent Application No. 15/524,242, filed May 3, 2017

In collaboration with:

Hussein Tawbi MD

A SNP that predicts response to chemotherapy and suggests new combination therapy

SCIENTIFIC REPORTS

Article | OPEN ACCESS | Published: 01 March 2019

PARP1 rs1805407 Increases Sensitivity to PARP1 Inhibitors in Cancer Cells Suggesting an Improved Therapeutic Strategy

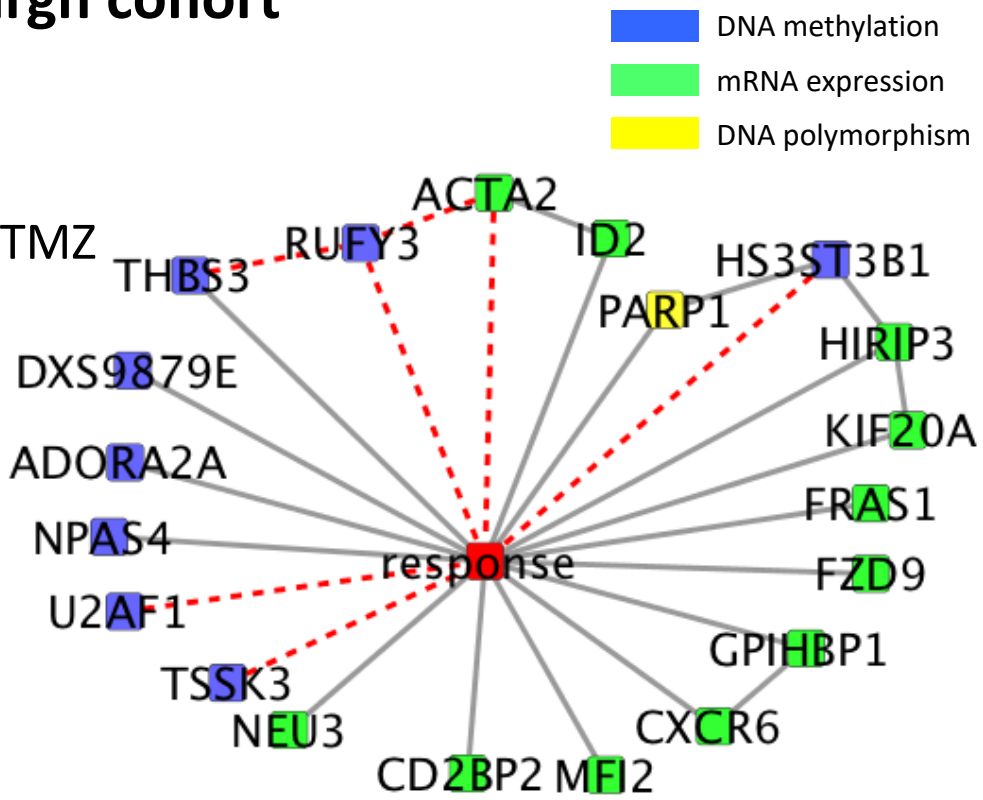
Irina Abecassis, Andrew J. Sedgewick, Marjorie Romkes, Shama Buch, Tomoko Nukui, Maria G. Kapetanaki, Andreas Vogt, John M. Kirkwood, Panayiotis V. Benos & Hussein Tawbi

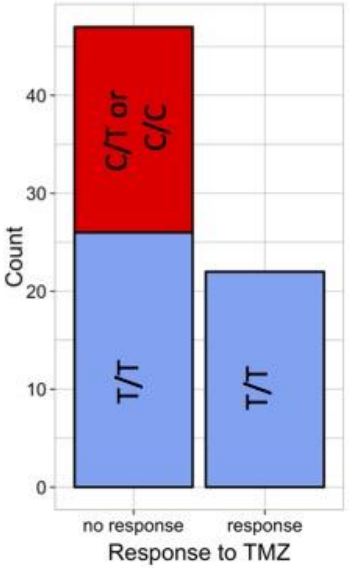
Scientific Reports **9**, Article number: 3309 (2019) | Download Citation

Identify cancer chemotherapy biomarkers

Hussein Tawbi MD

- **Metastatic melanoma Pittsburgh cohort**
- Subjects:
  - 69 subjects
  - Demographics and response to TMZ treatment
- **Data acquisition from tumor:**
  - Gene expression
  - miRNA expression
  - DNA methylation
  - SNP assay (selected SNPs)



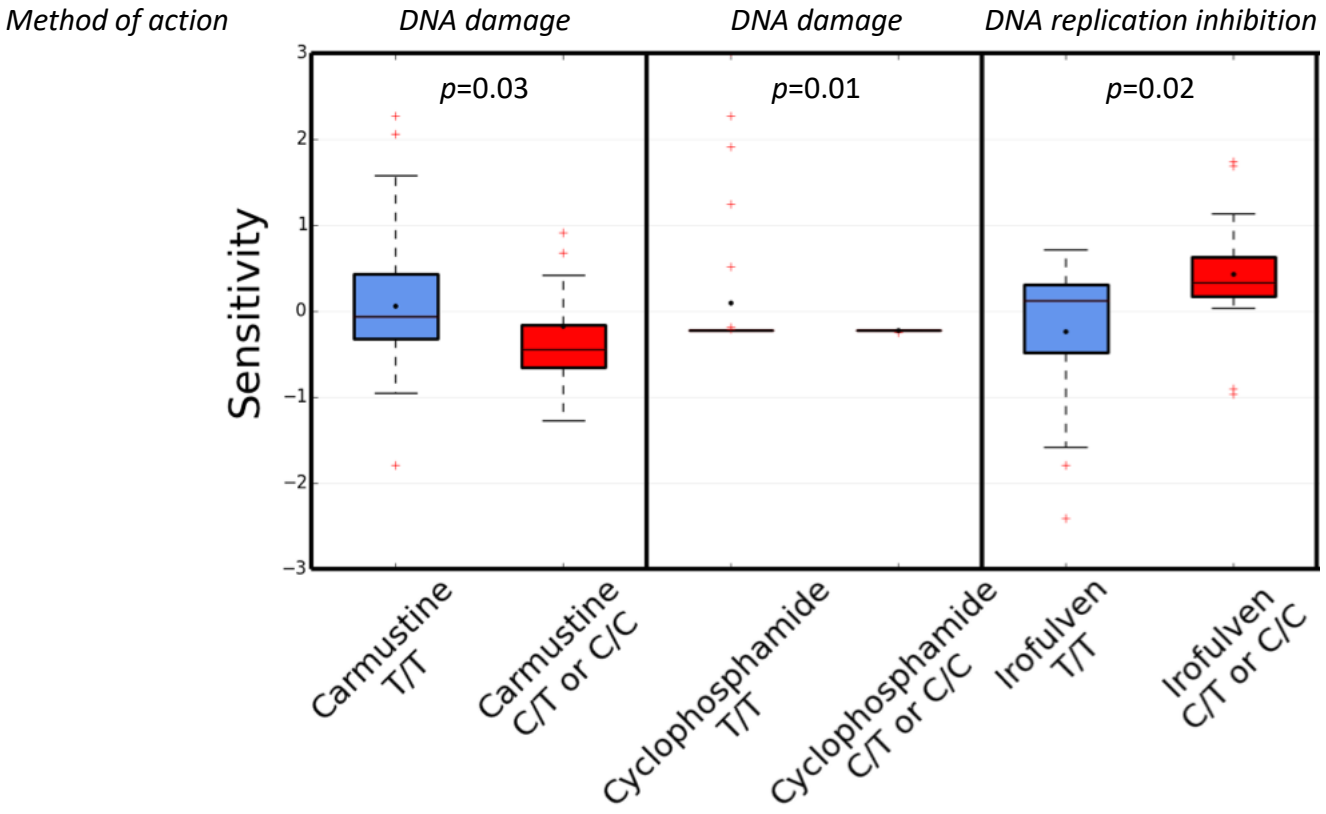


$p=10^{-5}$

Abecassis*, Sedgewick*, …, Benos¶, Tawbi¶, 2019, *Sci Rep*, **9**:3309

Alkylating agents induce the strongest changes in drug sensitivity between carriers/non-carriers

AJ Sedgewick PhD



Abecassis*, Sedgewick*, …, Benos¶, Tawbi¶, 2019, *Sci Rep*, **9**:3309

Hypothesis (testable)

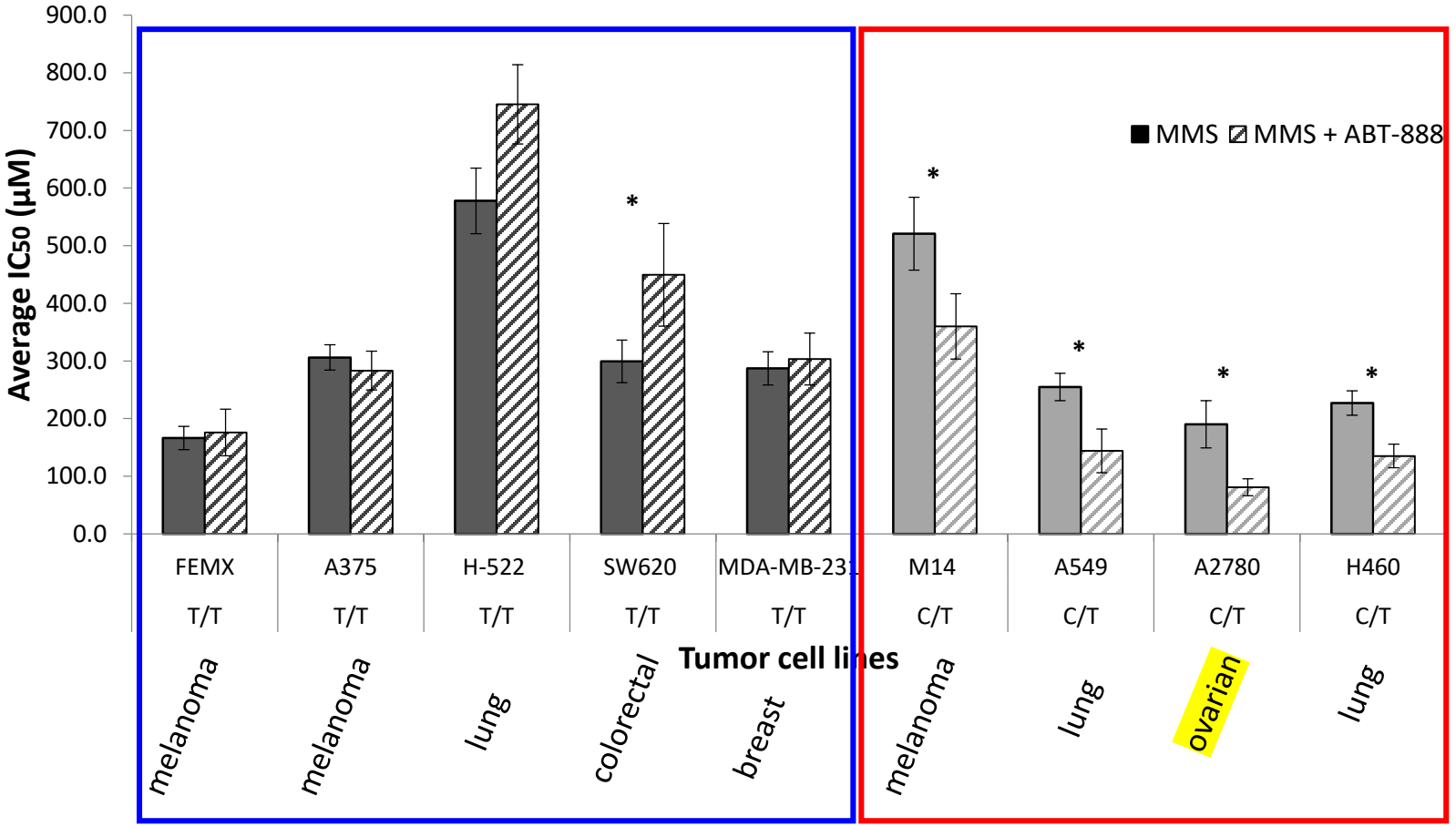
- The PARP1 SNP is directly related to improved DNA damage repair
  - Improved DNA damage repair ➔ worse response to chemotherapy

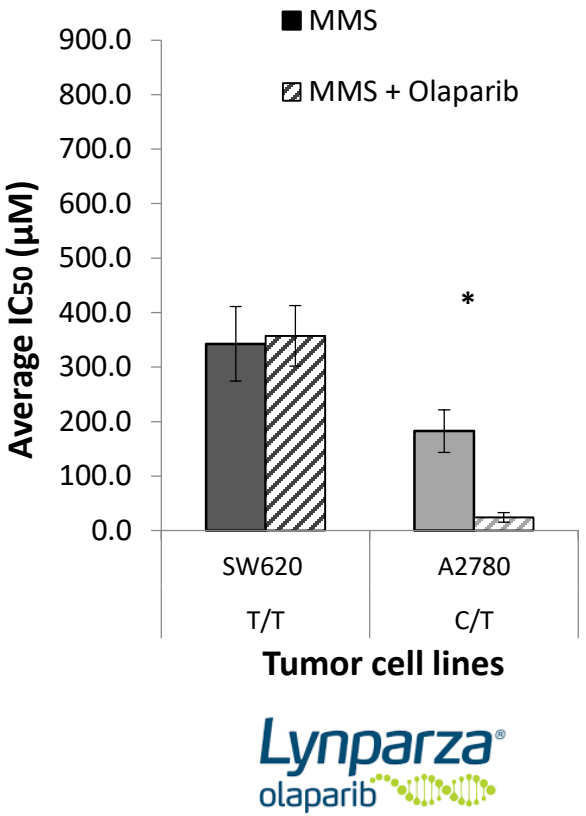
- Testing:

  Treat cells with PARP inhibitor (PARPi) ➔ do SNP cells require lower doses of alkylating agent than WT cells? (lower IC_{50})

PARP-1 inhibition increases chemo efficiency to cell lines with the SNP

Hussein Tawbi MD





Lynparza® olaparib

Abecassis*, Sedgewick*, …, Benos¶, Tawbi¶, 2019, *Sci Rep*, **9**:3309

PARP-1 inhibition increases chemo efficiency to cell lines with the SNP

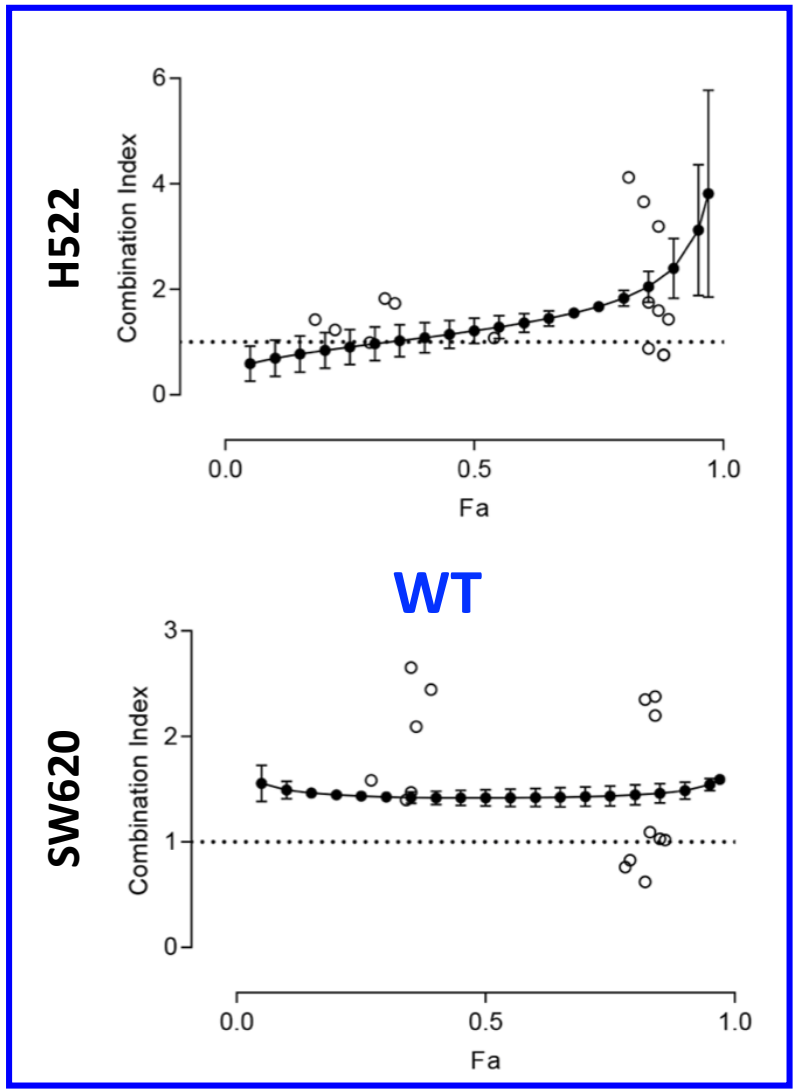
Andreas Vogt PhD

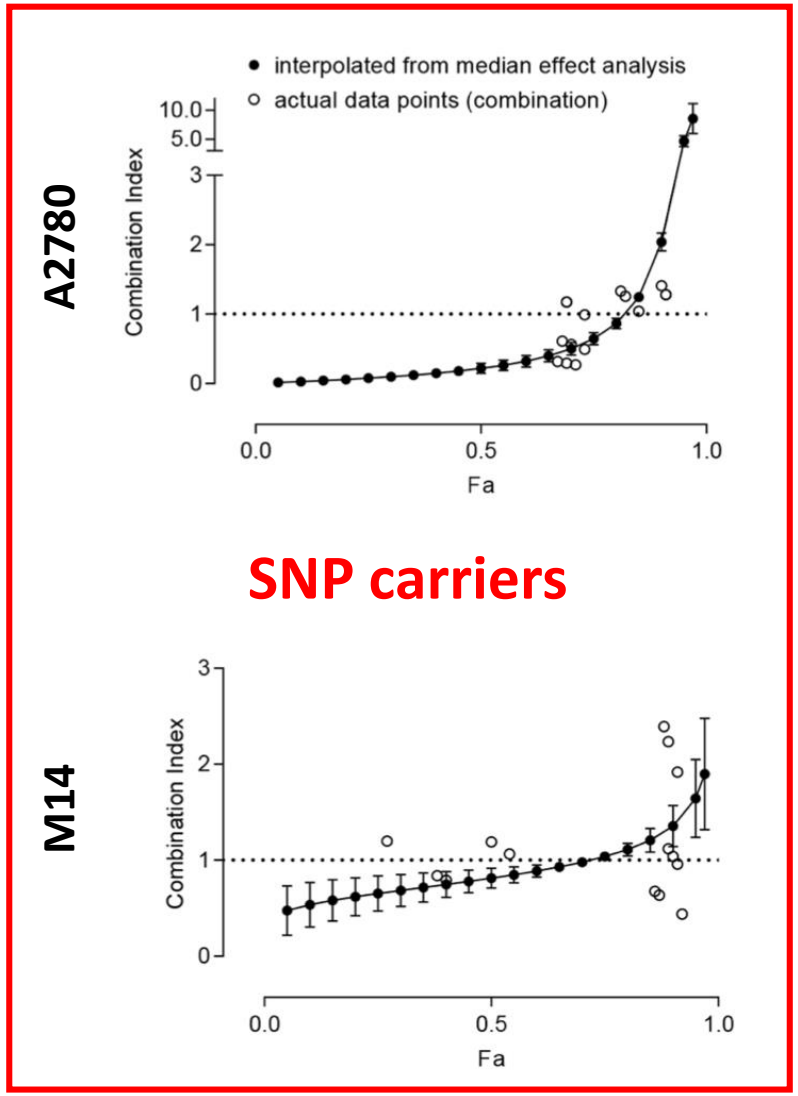
additive

antagonism

synergy

synergy





Abecassis*, Sedgewick*, …, Benos¶, Tawbi¶, 2019, *Sci Rep*, **9**:3309

Hypothesis (testable)

- The PARP1 SNP is directly related to improved DNA damage repair
  - Improved DNA damage repair ➔ worse response to chemotherapy

- Testing:

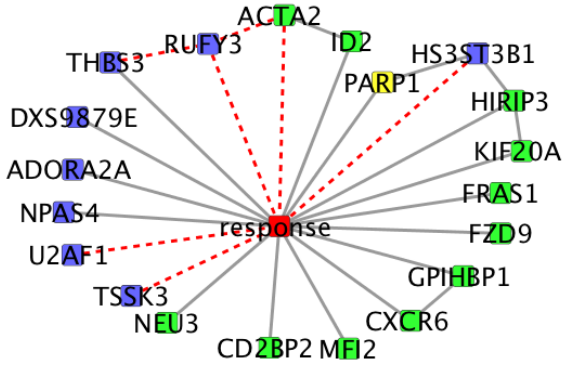
  Treat cells with PARP inhibitor (PARPi) ➔ do SNP cells require lower doses of alkylating agent than WT cells? (lower IC_{50})

- **Result:**

  A PARP1 SNP may be suitable for patient stratification and deciding optimal therapeutic intervention
  - SNP carriers ➔ combination therapy w/ FDA-approved olaparib
  - wt patients ➔ no PARP1 inhibitor

What we learned from the PARP1 study?

- PARP1 SNP rs1805407 is linked to poor response to chemotherapy
- PARP1 inhibitors and alkylating agents act synergistically on SNP carrier cell lines
- PARP1 inhibitors make SNP carrier cell lines more sensitive to chemotherapy, indicating potential new therapeutic strategy



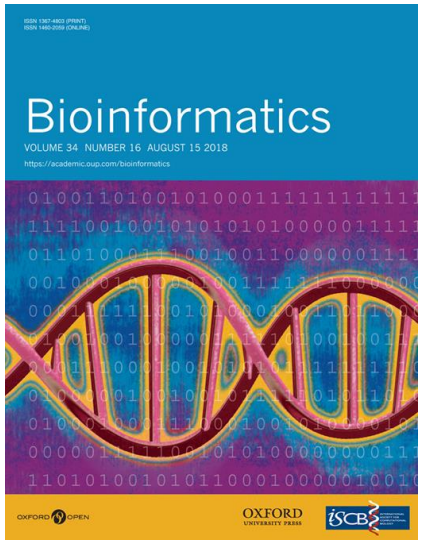
Kristina Buschur

Ivy Shi

In collaboration with:

Frank Sciurba MD

Determinants of longitudinal lung function decline in COPD patients



Systems Biology

Mixed Graphical Models for Integrative Causal Analysis with Application to Chronic Lung Disease Diagnosis and Prognosis

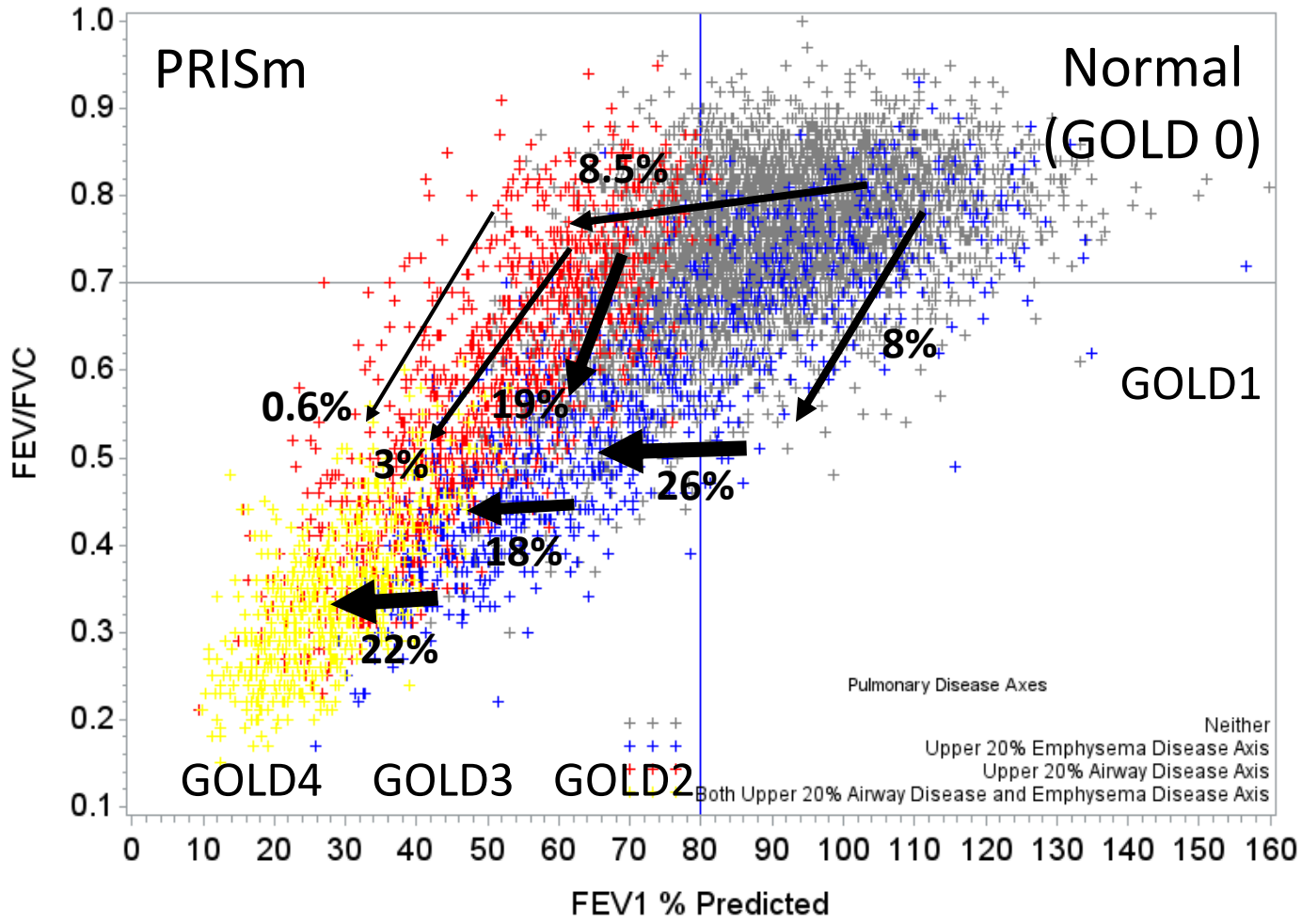
Andrew J Sedgewick[1,2], Kristina Buschur[1,2], Ivy Shi[3], Joseph D. Ramsey[4], Vineet K. Raghu[5], Dimitris V. Manatakis[1], Yingze Zhang[6], Jessica Bon[6], Divay Chandra[6], Chad Karoleski[6], Frank C. Sciurba[6], Peter Spirtes[4], Clark Glymour[4], Panayiotis V. Benos[2,3,*]

[1]Department of Computational and Systems Biology, [3]Department of Bioengineering, [5]Department of Computer Science, [6]Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [4]Department of Philosophy, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

[2]Joint CMU-Pitt PhD Program in Computational biology

*To whom correspondence should be addressed.

COPD progression (COPDGene® cohort)



Red = Upper 20% Airway Dis Axis
Blue = Upper 20% Emphysema Dis Axis
Yellow = Both
Gray = Neither

Graph source: COPDGene®

FEV1 progression in COPD patients (SCCOR cohort)

Frank Sciurba MD

- **SCCOR** (Pittsburgh Specialized Center of Clinically Oriented Research)
- Subjects:
  - 762 subjects (community-based, tobacco-exposed cohort)
  - 385 subjects returned for a 2-year follow-up evaluation

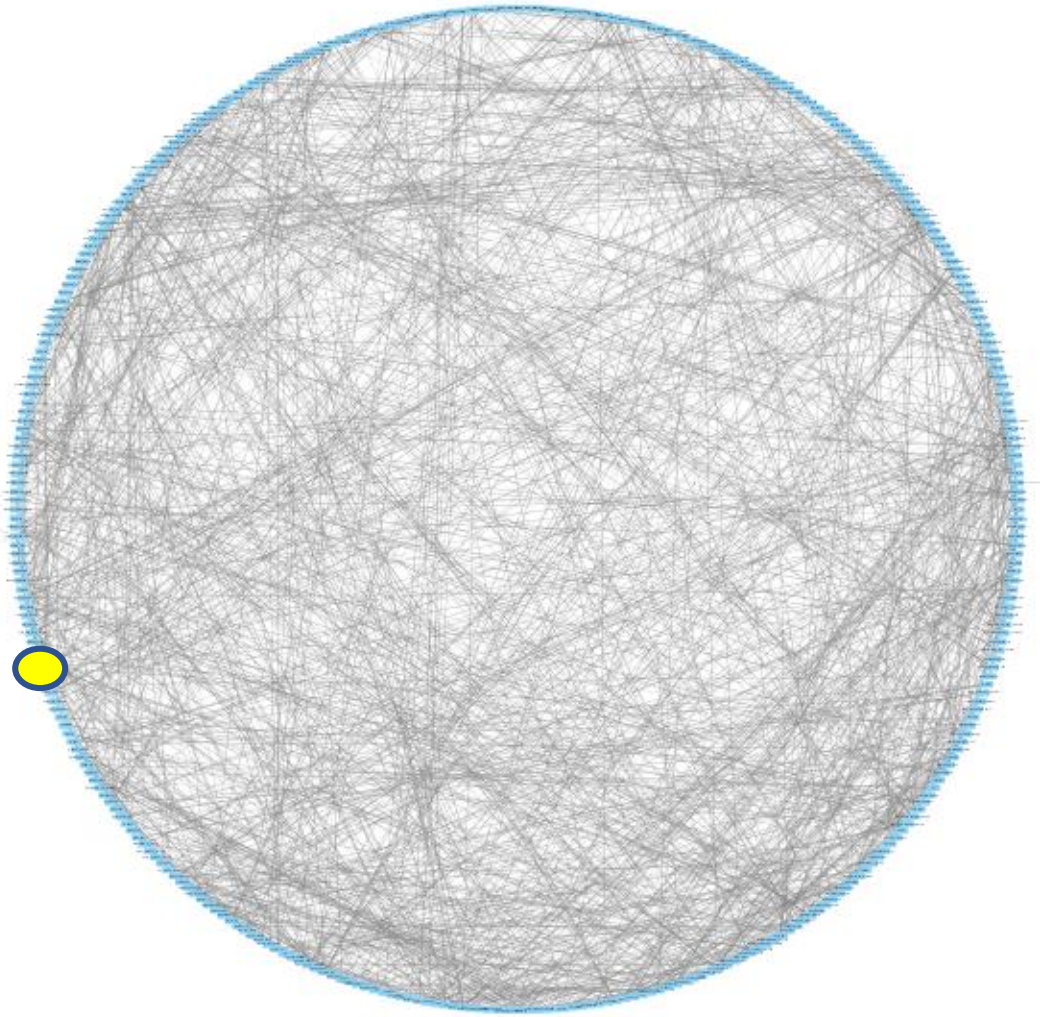
- **Data acquisition in visit-1:**
  - Demographics
  - Spirometry (pre- and post-bronchodilators)
  - Semi-quantitative visual and quantitative MDCT
  - Blood biomarkers
  - Exercise testing
  - Questionnaire

Questionnaire:

- Patient's history of other diseases (asthma, etc)
- Environmental (asbestos, arsenic, etc)
- Symptoms (coughing, dyspnea, etc)
- Psychological

Integrating multi-modal datasets with probabilistic models

Ivy Shi



All baseline variables + ΔFEV1

Clinical and blood biomarker variables linked to FEV_1 decline in COPD

Kristina Buschur

- Spirometry variables
- Smoking-related
- Co-morbidities
- Environmental factors
- Blood biomarkers
- Other conditions

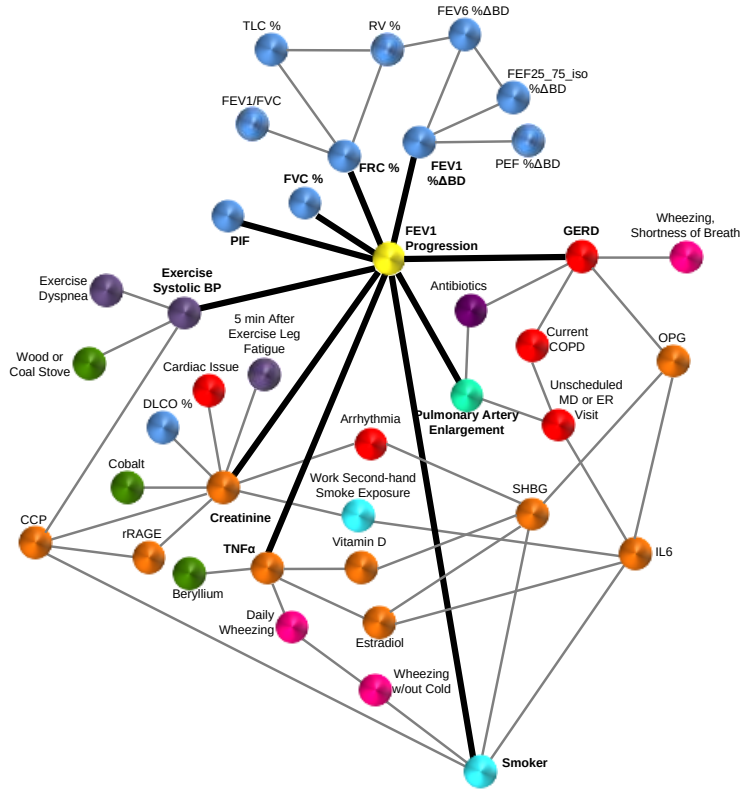
1st and 2nd neighbors of FEV1 progression

Sedgewick *et al*, Bioinformatics, 2018.

What we learned from the COPD study?

- Creatinine and TNF-α are directly linked to longitudinal lung function decline in COPD patients
  - Creatinine may be linked to muscle loss
  - TNF-α is linked to inflammation: can inflammation reduction help delay lung function decline?

- Reducing GERD exacerbations may help delay lung function decline



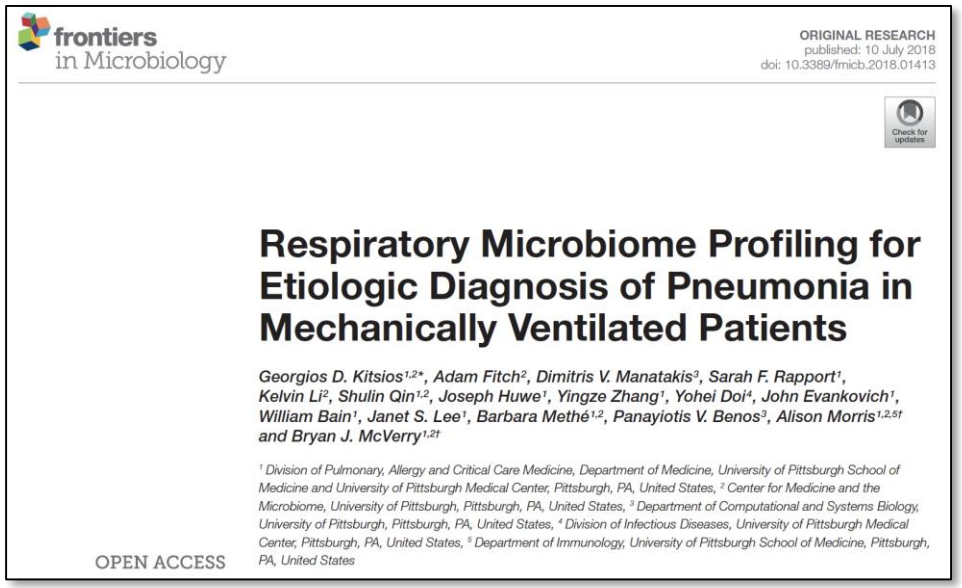
Dimitris Manatakis PhD

In collaboration with:

George Kitsios MD

Alison Morris MD

Microbiota and clinical variables that predict culture positivity in lung ICU patients



frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 July 2018
doi: 10.3389/fmicb.2018.01413

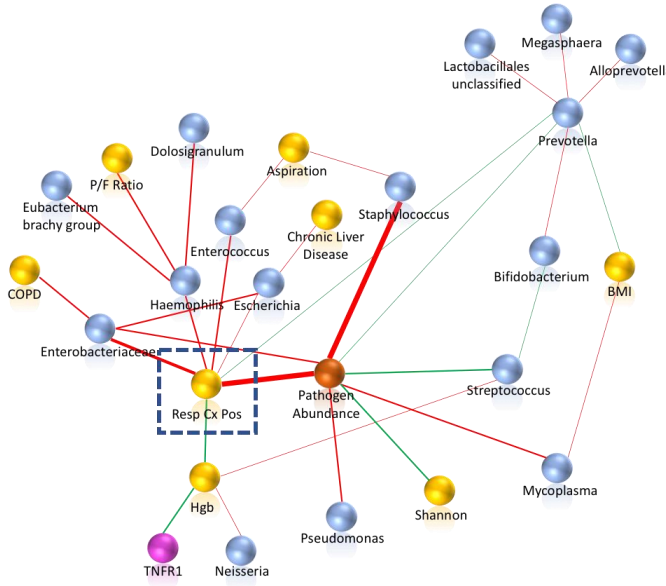
Check for updates

Respiratory Microbiome Profiling for Etiologic Diagnosis of Pneumonia in Mechanically Ventilated Patients

Georgios D. Kitsios[1,2], Adam Fitch[2], Dimitris V. Manatakis[3], Sarah F. Rapport[1], Kelvin Li[2], Shulin Qin[1,2], Joseph Huwe[1], Yingze Zhang[1], Yohei Doi[4], John Evankovich[1], William Bain[1], Janet S. Lee[1], Barbara Methé[1,2], Panayiotis V. Benos[3], Alison Morris[1,2,5†] and Bryan J. McVerry[1,2†]*

[1] Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of Medicine and University of Pittsburgh Medical Center, Pittsburgh, PA, United States, [2] Center for Medicine and the Microbiome, University of Pittsburgh, Pittsburgh, PA, United States, [3] Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, United States, [4] Division of Infectious Diseases, University of Pittsburgh Medical Center, Pittsburgh, PA, United States, [5] Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

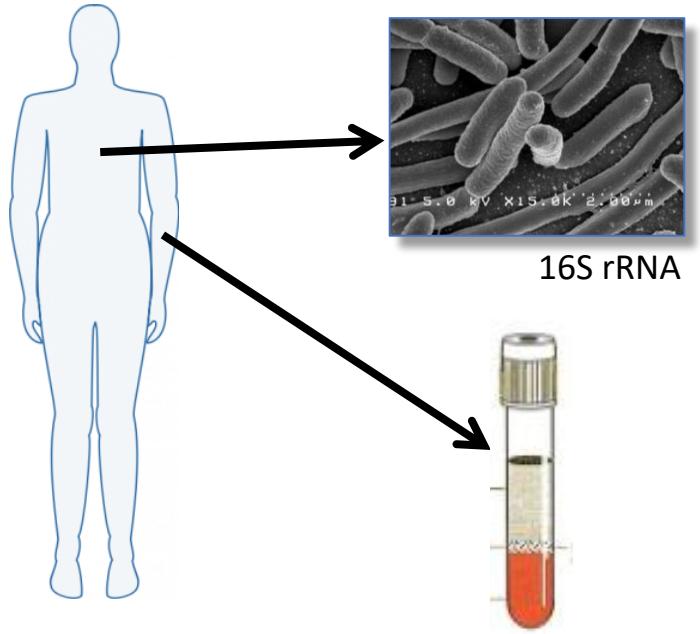
OPEN ACCESS



Can we predict Cx positivity in ICU patients from lung 16S microbiome?

George Kitsios MD

Variable	All	Culture-positive	Culture-negative^	P-value
N	**56**	**12**	**44**	
Age, mean (SD), yrs	55.9 (15.3)	54.7 (17.2)	56.2 (14.9)	0.88
Males, N (%)	34 (61)	5 (42)	29 (66)	0.18
BMI, mean (SD)	32.2 (10.2)	28.8 (7.1)	33.1 (10.8)	0.19
History of diabetes, N (%)	25 (45)	6 (50)	19 (43)	0.75
History of COPD, N (%)	17 (30)	5 (42)	12 (27)	0.47
Sepsis, N (%)[#]	50 (89)	12 (100)	38 (86)	0.32
ARDS, N (%)[$]	21 (38)	7 (58)	14 (32)	0.11
High clinical index for pneumonia[&]	34 (61)	12	22 (50%)	**0.002**

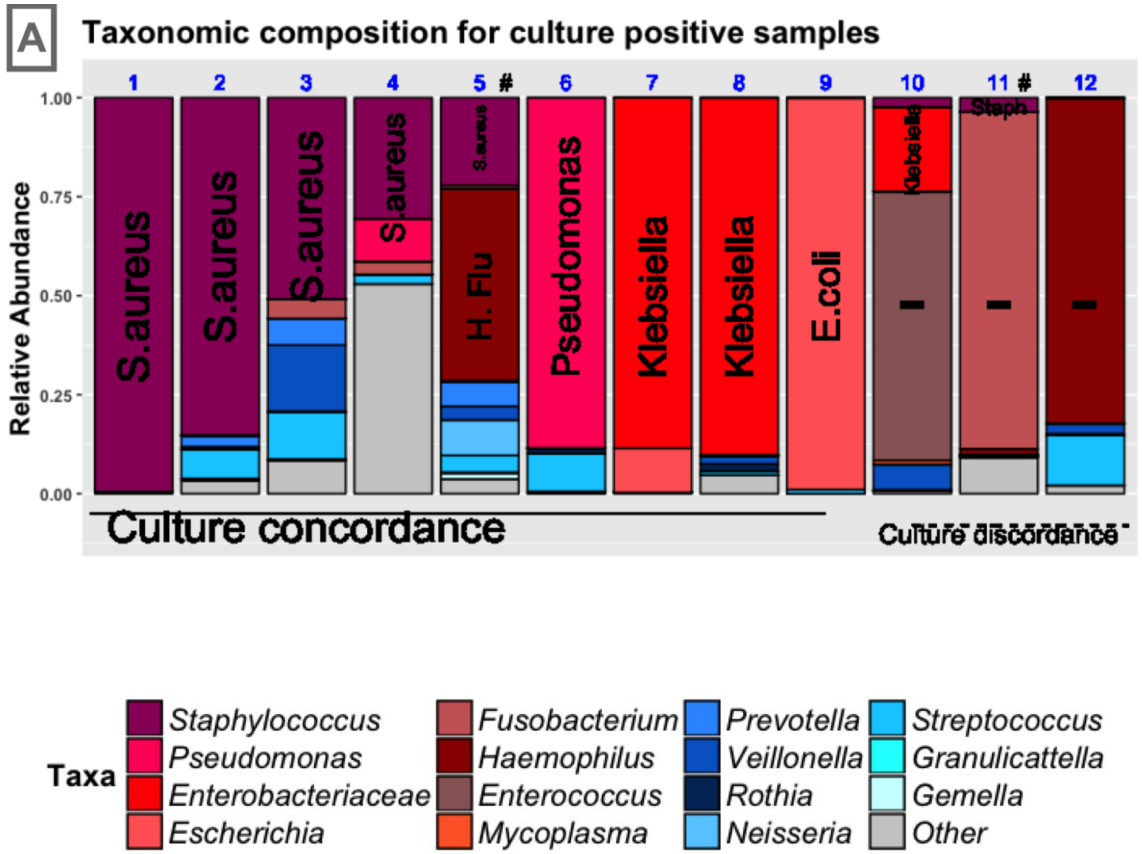


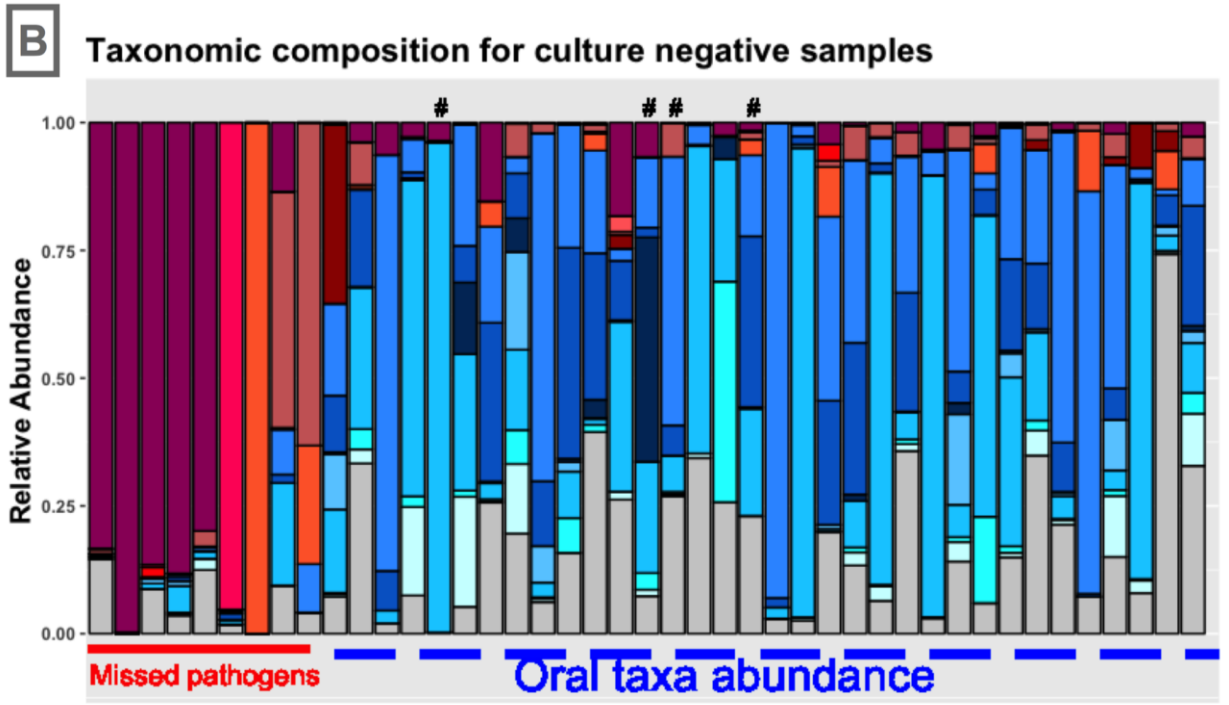
16S rRNA

Kitsios *et al*, "Respiratory microbiome profiling for etiologic diagnosis of pneumonia in mechanically ventilated patients", 2018, ***Frontiers in Microbiol***

Lung ICU patient cohort: microbiome profiles and Cx positivity

George Kitsios MD





Kitsios *et al*, 2018, *Frontiers in Microbiol*

Variables directly linked to ICU patient culture positivity

Dimitris Manatakis PhD

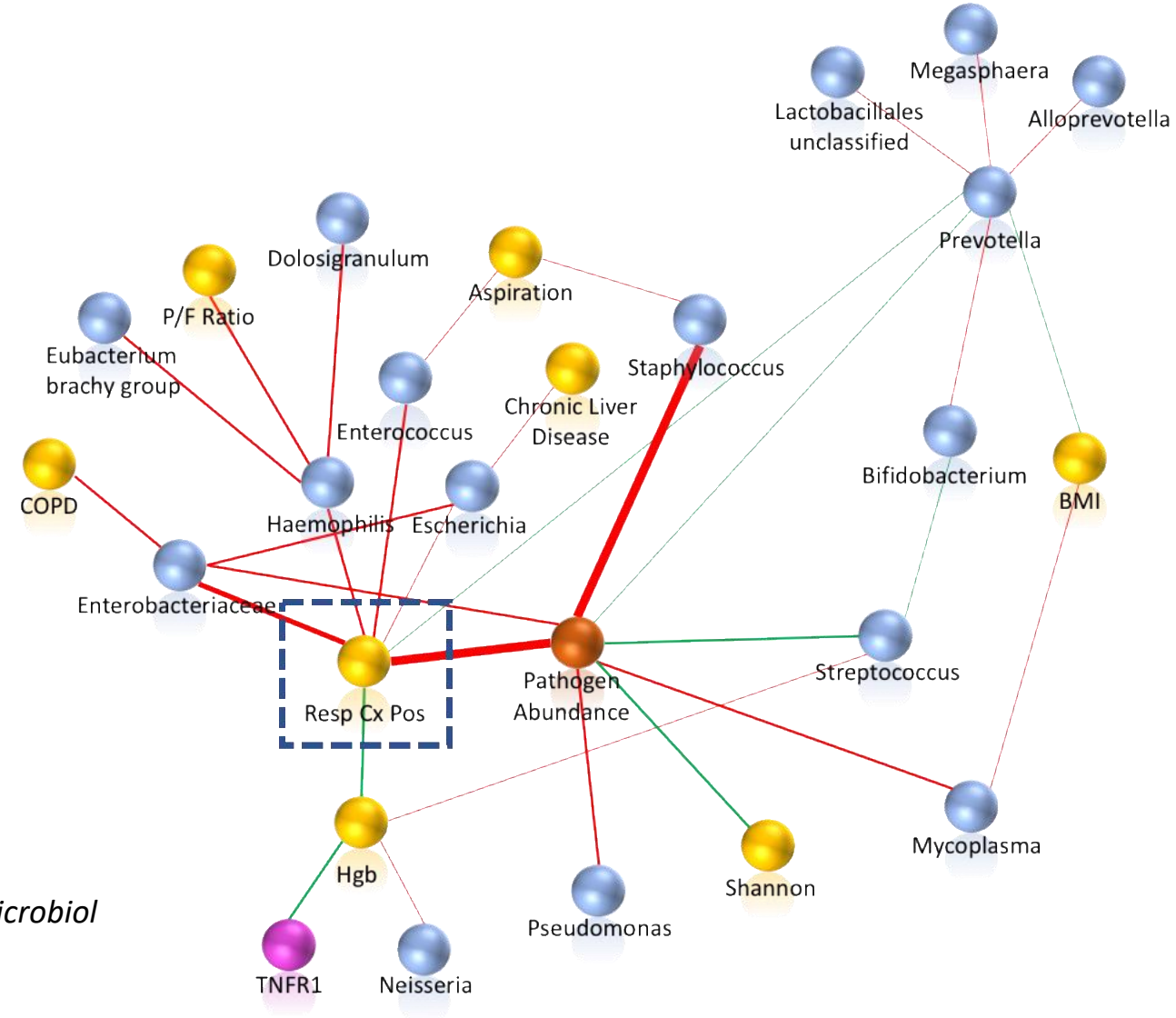
clinical

16S taxa

biomarkers

Positive assoc

Negative assoc



Kitsios *et al*, 2018, *Frontiers in Microbiol*

Variables directly linked to ICU patient culture positivity

Dimitris Manatakis PhD

clinical

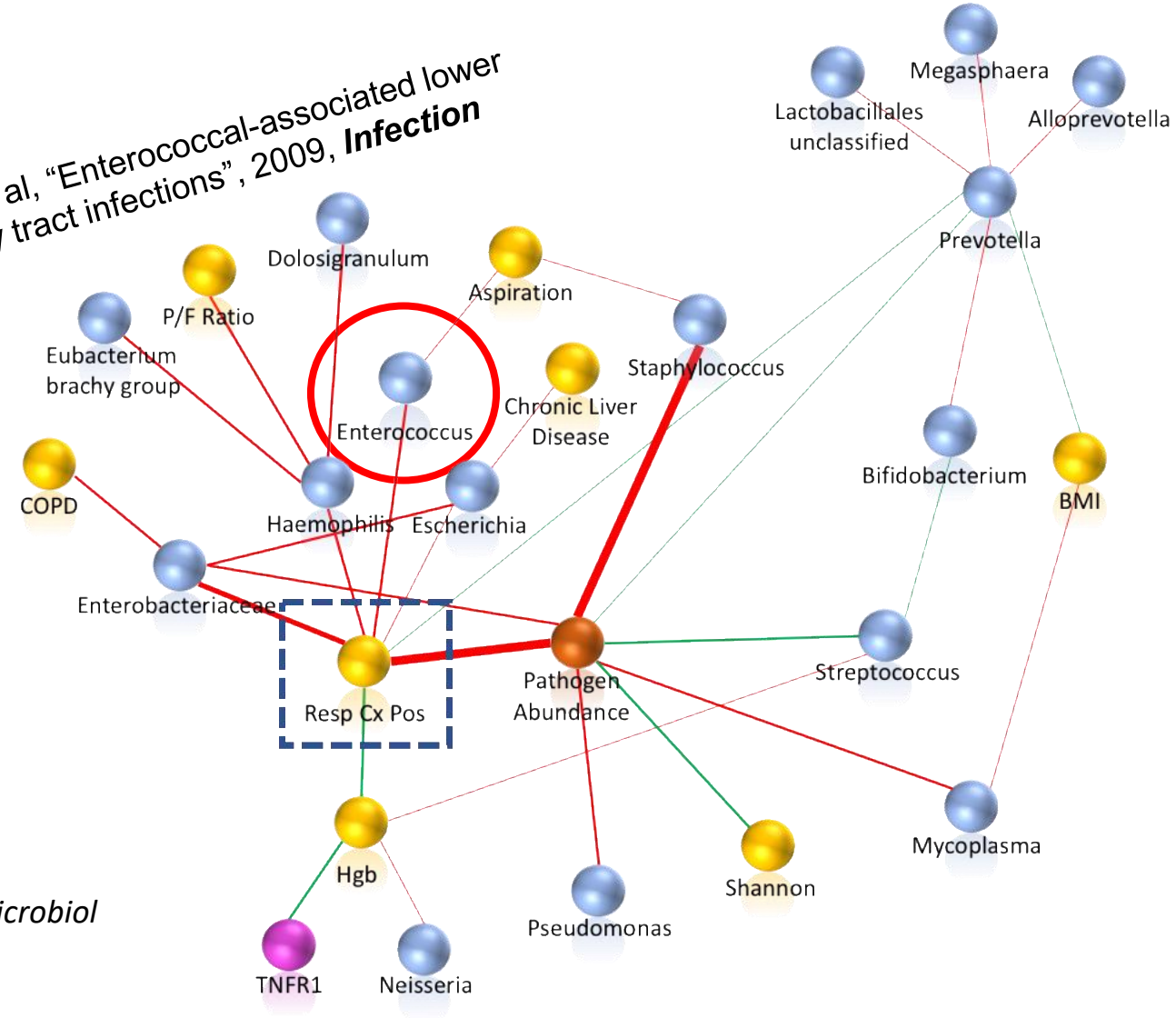
16S taxa

biomarkers

Positive assoc

Negative assoc

Grupper et al, "Enterococcal-associated lower respiratory tract infections", 2009, ***Infection***



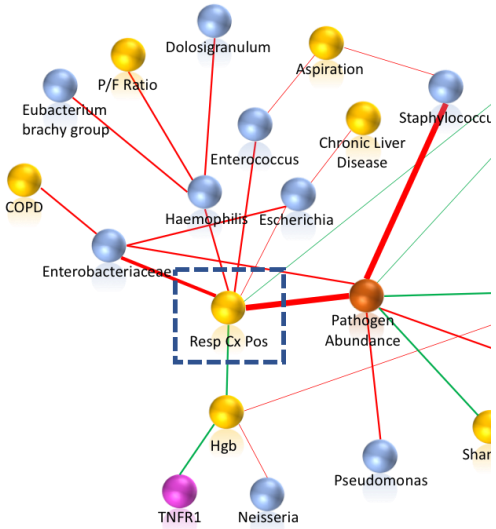
Kitsios *et al*, 2018, *Frontiers in Microbiol*

Some results from the ICU culture positivity study

- The microbial communities in 20% (9/44) of culture negative patient samples are dominated by pathogenic taxa (*Staphylococcus*, *Pseudomonas*)

- Using the network model we can predict culture positivity with an average accuracy of 83% (±7%)

- The 16S method is promising for prediction culture positivity in ICU patients



Kitsios *et al*, 2018, *Frontiers in Microbiol*

Take home messages

- ✓ *CausalMGM* is a highly flexible framework that can be used to analyze multi-modal and multi-scale data

- ✓ *CausalMGM* has the ability to efficiently incorporate prior information to learn more accurately graphs in high-dimensional data

Take home messages

- ✓ *CausalMGM* has been successfully applied to a variety of medical problems:
  - ✓ We developed a new accurate predictor of lung cancer from clinical and LDCT scan data, which has the potential of reducing unnecessary procedures in subjects with benign nodules
  - ✓ We identified a PARP1 SNP that is a marker for no response to chemotherapy and we've shown evidence to suggest that the SNP carriers may benefit from combination therapy (chemo + PARP1 inhibitors)
  - ✓ We identified blood biomarker proteins and comorbidities that are directly linked to longitudinal lung function decline in COPD patients (creatinine, TNF-α, GERD, etc)
  - ✓ We identified microbiome taxa and clinical variables that are indicative of culture positivity in ICU patients

Acknowledgements: Some current collaborations in Pittsburgh

Causal modeling on mixed data (NLM R01)

Clark Glymour, PhD – Philosophy, CMU

Peter Spirtes, PhD – Philosophy, CMU

Joe Ramsey, PhD – Philosophy, CMU

Cloud interfaces (NHLBI U01)

Panos Chrysanthis, PhD – Computer Science, Pitt

FUNDING

NHLBI, NLM, NCI, NHGRI (BD2K)

NLM: **R01 LM012087 (Benos/Glymour)**

NHLBI: **U01 HL145550 (Rojas/Benos/et al)**

NHLBI: **R01 HL140963 (Morris/Benos/Chan)**

NHLBI: **U01 HL137159 (Benos/Sciurba)**

NHLBI: P01 (Gladwin/Morris)

NCI: R35 (Finn)

Early detection of lung cancer

David Wilson MD – Medicine, Pitt / UPMC

Jiantao Pu PhD – Radiology, Pitt

Biomarkers for cancer treatment response (NLM R01)

John M. Kirkwood MD –Medicine, Pitt / UPMC

Hussein Tawbi MD – MD Anderson

COPD progression & subtyping (NHLBI U01)

Frank Sciurba, MD – Pulmonary Medicine, Pitt / UPMC

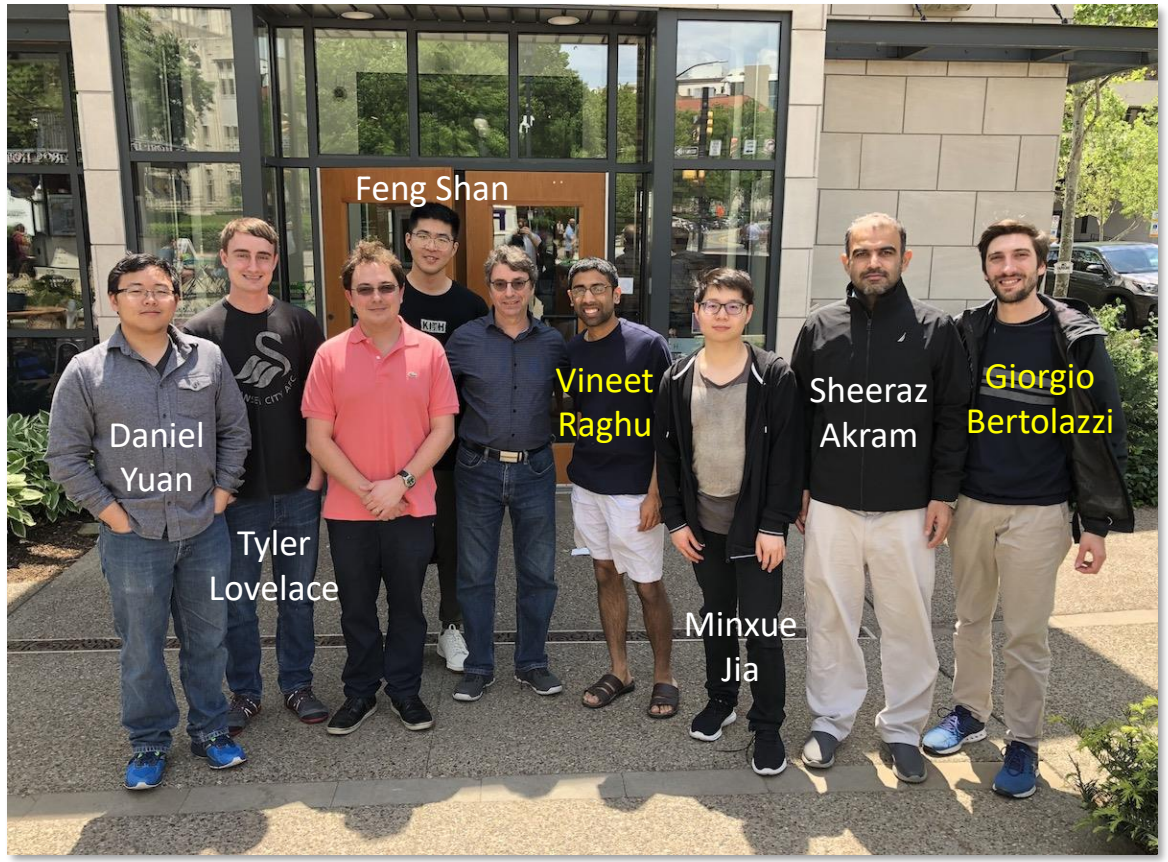
Craig Riley, MD - Pulmonary Medicine, Pitt / UPMC

Microbiome in chronic lung diseases and ICU (NHLBI P01)

Alison Morris MD – Pulmonary Medicine, Pitt / UPMC

George Kitsios MD – Pulmonary Medicine, Pitt / UPMC

Benos' laboratory



MD FELLOW

Craig Riley MD
(co-advised: Frank Sciurba)

MD STUDENT

Grace Zhang

Electronic contacts:

benos@pitt.edu

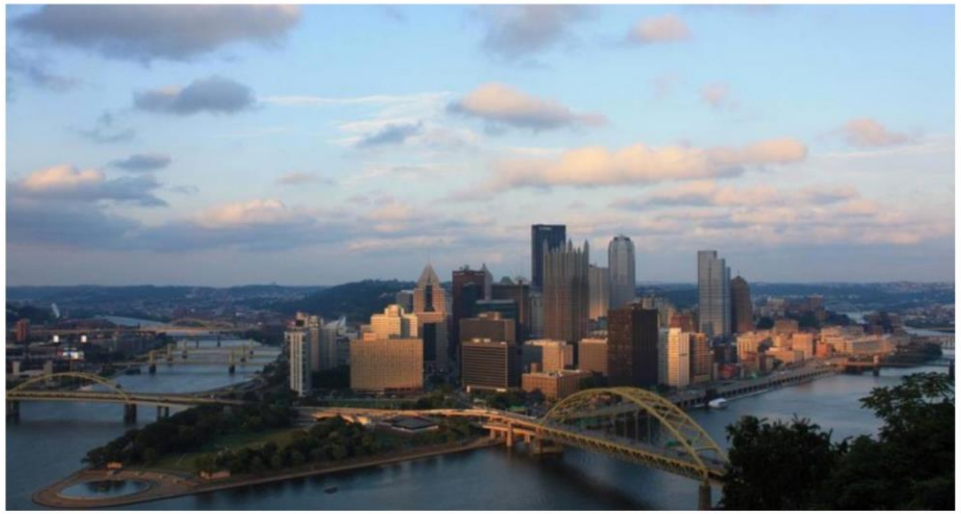
http://www.benoslab.pitt.edu

Postdoc positions available in Benos' Lab

Developing causal graphical models for integrating biomedical and clinical Big Data

Takis Benos (benos@pitt.edu**)**

Department of Computational and Systems Biology



The Economist names Pittsburgh the Most Livable City (on the mainland) again

Deb Smit · August 25, 2014 · Business & Tech News

Many thanks to…

Georgios Deftereos, MD

THE UNIVERSITY OF UTAH

ARUP® LABORATORIES

PATHOLOGY

Questions???

Electronic contacts:

benos@pitt.edu

http://www.benoslab.pitt.edu

References:

- Sedgewick et al, “Learning mixed graphical models with separate sparsity parameters and stability-based model selection”, 2016, ***BMC Bioinformatics***
- Sedgewick et al, “Mixed Graphical Models for Integrative Causal Analysis with Application to Chronic Lung Disease Diagnosis and Prognosis”, 2019, ***Bioinformatics***
- Manatakis*, Raghu*, Benos, “piMGM: Incorporating Multi-Source Priors in Mixed Graphical Models for Learning Disease Networks”, 2018, ***Bioinformatics***
- Raghu et al, “Feasibility of lung cancer prediction from low-dose CT scan and smoking factors using causal models”, 2019, ***Thorax***
- Kitsios et al, “Respiratory microbiome profiling for etiologic diagnosis of pneumonia in mechanically ventilated patients”, 2018, ***Frontiers in Microbiology***
- Abecassis, Sedgewick, et al, “PARP1 rs1805407 Increases Sensitivity to PARP1 Inhibitors in Cancer Cells Suggesting an Improved Therapeutic Strategy”, 2019, ***Scientific Reports***