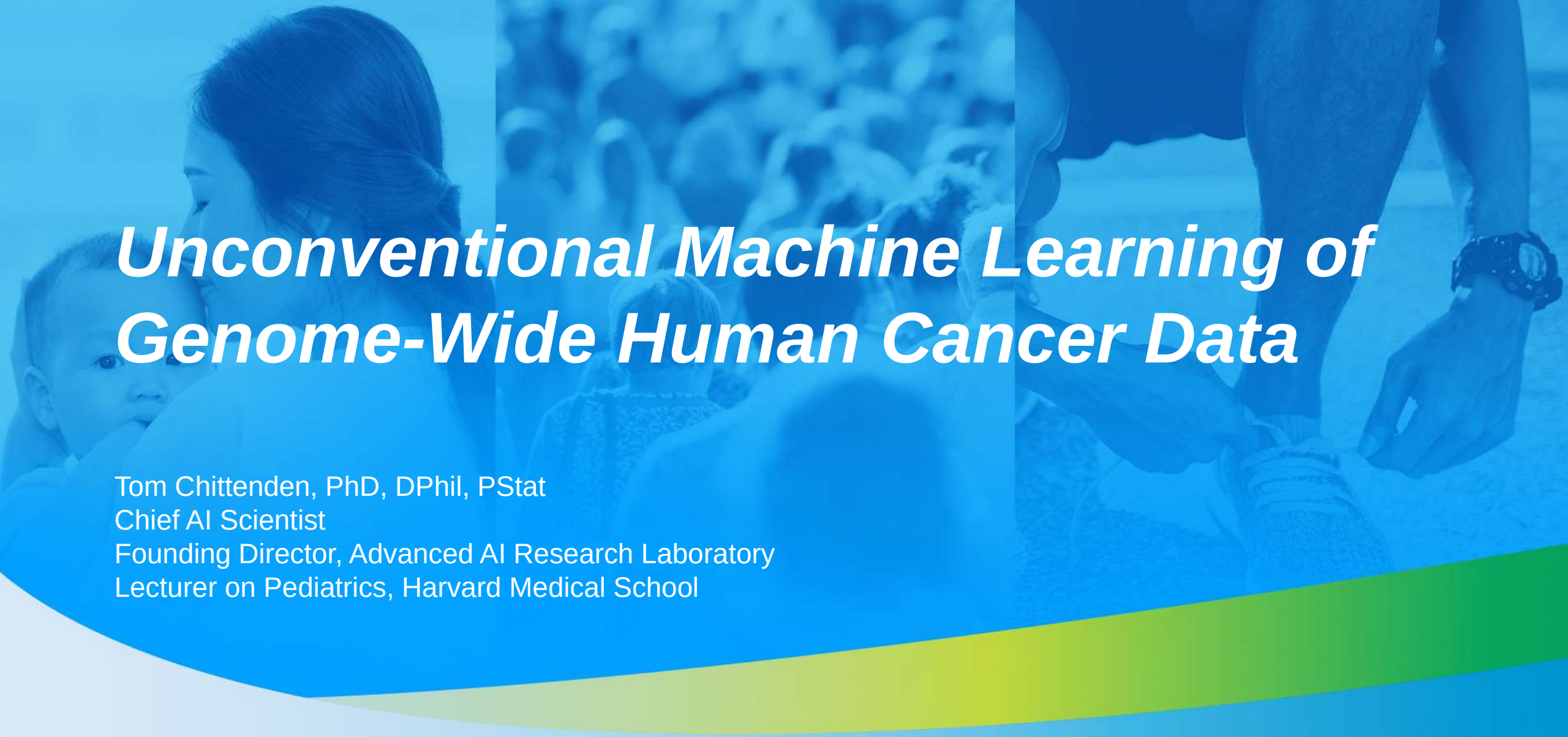


Unconventional Machine Learning of Genome-Wide Human Cancer Data

Tom Chittenden, PhD, DPhil, PStat
Chief AI Scientist
Founding Director, Advanced AI Research Laboratory
Lecturer on Pediatrics, Harvard Medical School

WuXiNextCODE

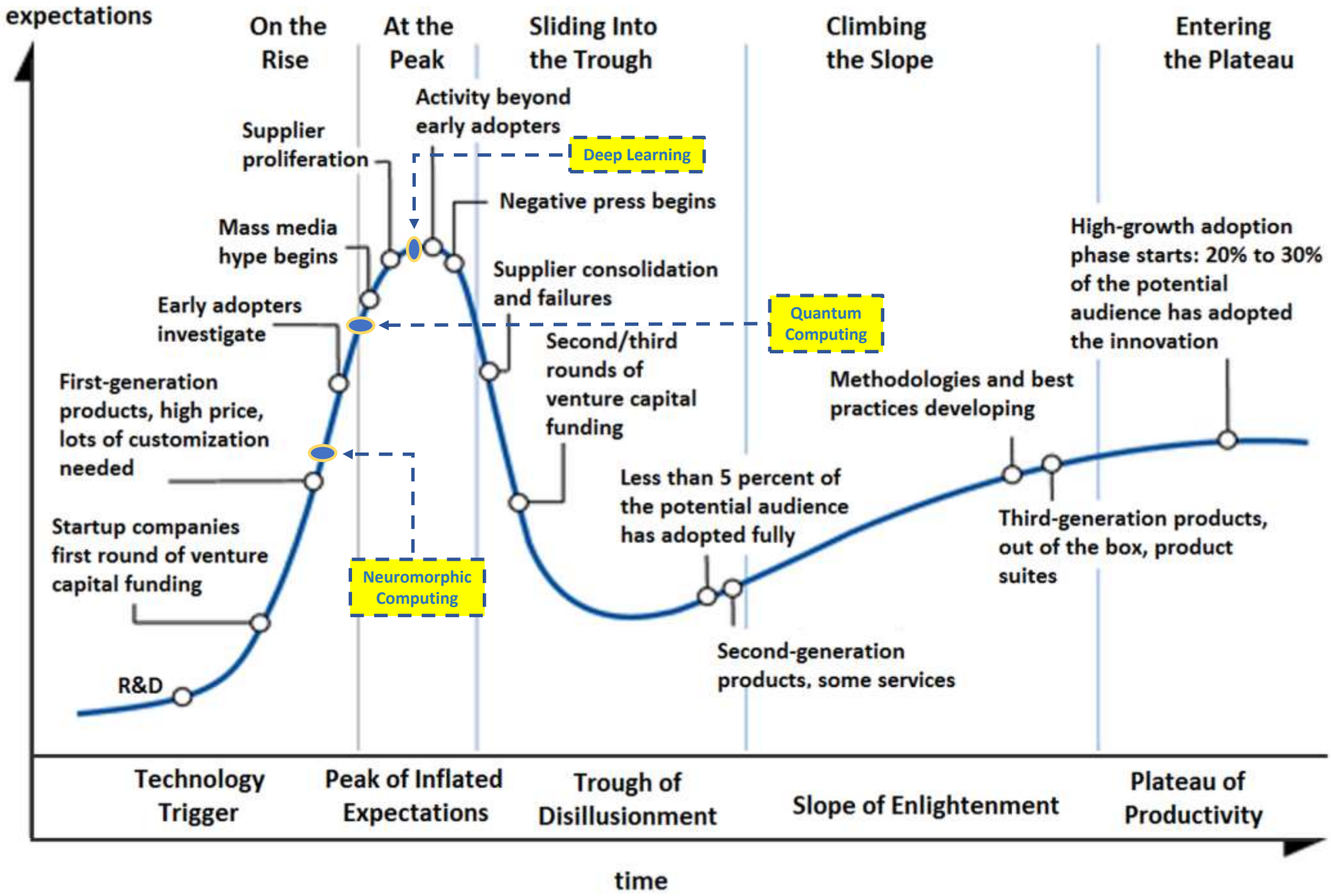
WuXi NextCODE Global Predictive Analytics Initiative

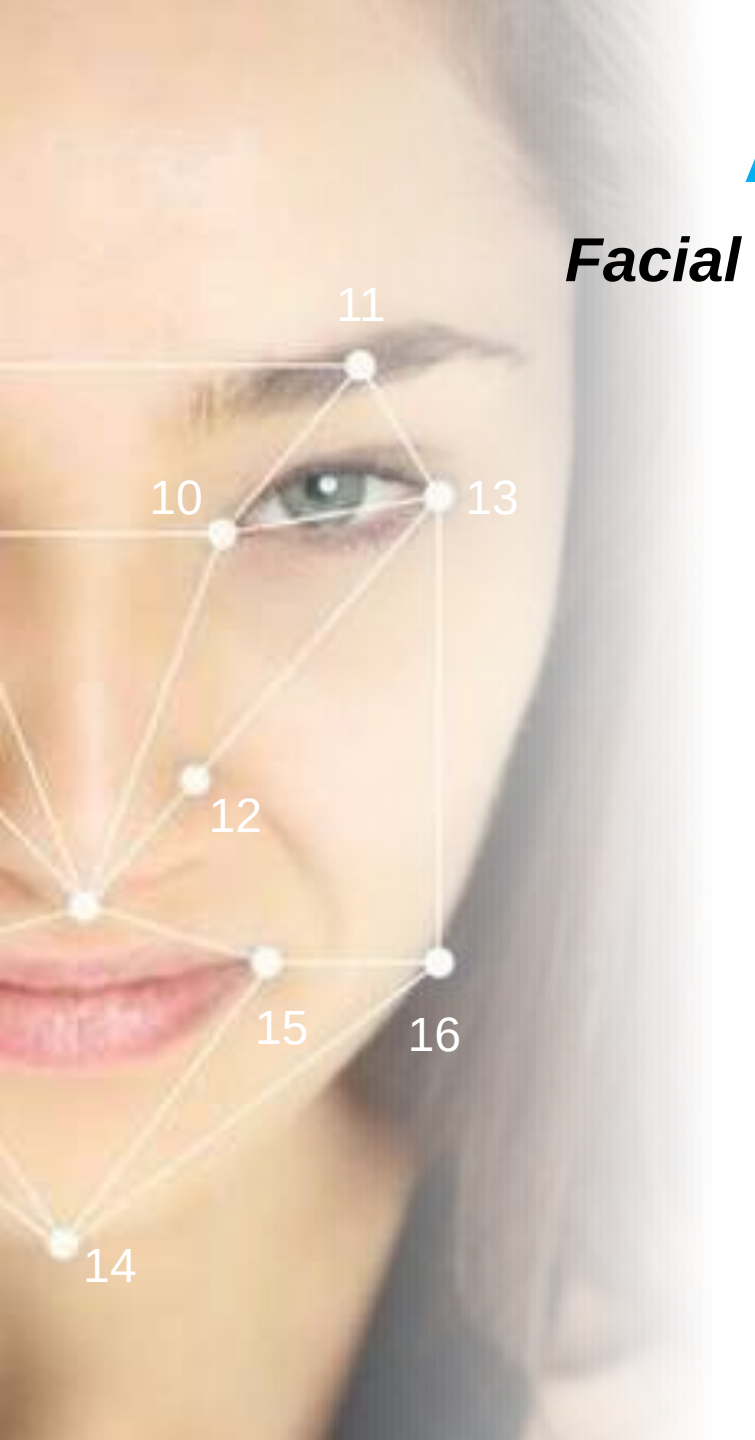
deepCODE Deep Learning and Probabilistic Programming – Faster, cost-effective drug development

Adding value to drug discovery pipelines

- **Drug target discovery and drug repurposing with novel ensemble computational intelligence strategies with integrated data platforms to identify 'causal' driver genes and molecular signal transduction networks**
  - ***Proof of concept for causal statistical learning approaches.***
  - ***Focus of Today's Talk.***
- **Discover accurate integrated 'omics' profile that defines responders and non-responders for a drug in development**
  - ***Pharma partners can use our profile to decrease cost and time of phase II or phase III trials.***
  - ***WXNC can provide sequencing/ GOR database/ analysis/ deep learning.***
  - ***Note approach may work on small sample sizes - deep learning is powerful enough to potentially find drug response profiles even in phase I clinical trials with only 40 to 60 patients on drug.***
- **Discover accurate integrated 'omics' profile that defines responders and non-responders for an approved expensive drug that is being underutilized**
  - ***Pharma can use our profile to justify use and reimbursement for their drug.***
  - ***A drug response profile could salvage the marketing of their drug.***

WuXiNextCODE





AI & Deep Learning

Facial Recognition & DeepCODE Feature Selection Analogy

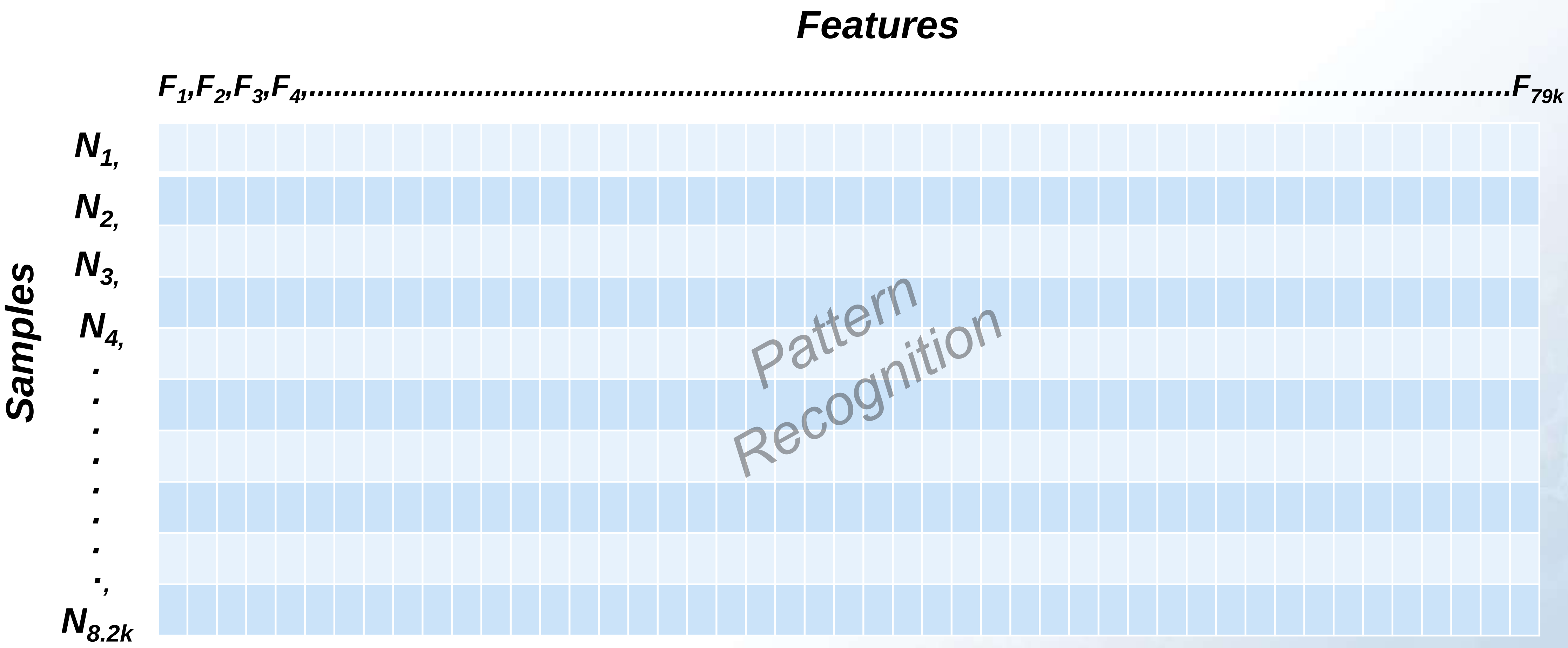
(Facebook AI team's Facial Recognition Algorithm boasts 97.25% Accuracy)

Facial Recognition

Thousands of people

Tens of Features

Our deepCODE dimensionality reduction methods enhance algorithm stability and allow us to handle tens of thousands of features without overfitting



A.I. and Precision Medicine

The computational power of modern A.I. technology is well-positioned to uncover new and actionable insights from the exponentially growing pool of biological data.

FEATURE LEARNING

The intelligent simplification of high-dimensional multi-omic data without loss of information

MACHINE & DEEP LEARNING

Intelligent algorithms capable of self-optimization to achieve incredible accuracy with complex, layered data

CAUSAL INFERENCE

Specialized statistical learning models capable of elucidating casual dependencies within biological data

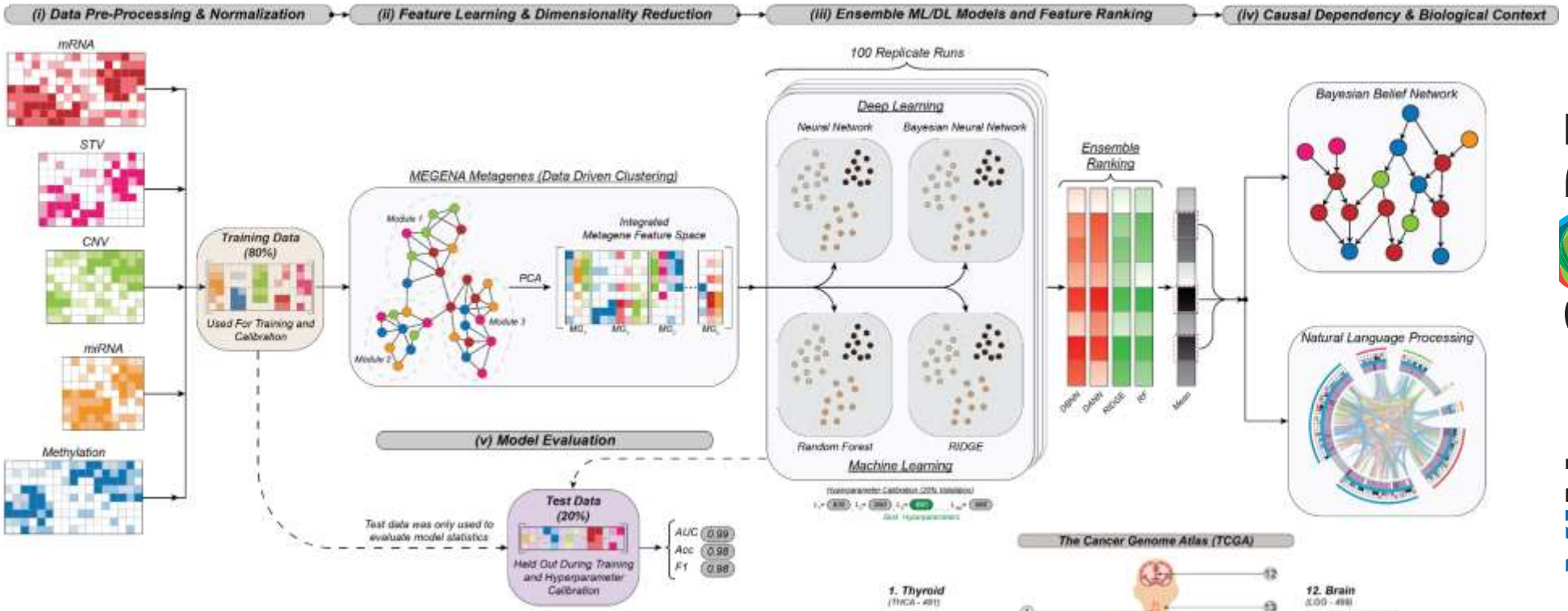
NATURAL LANGUAGE PROCESSING

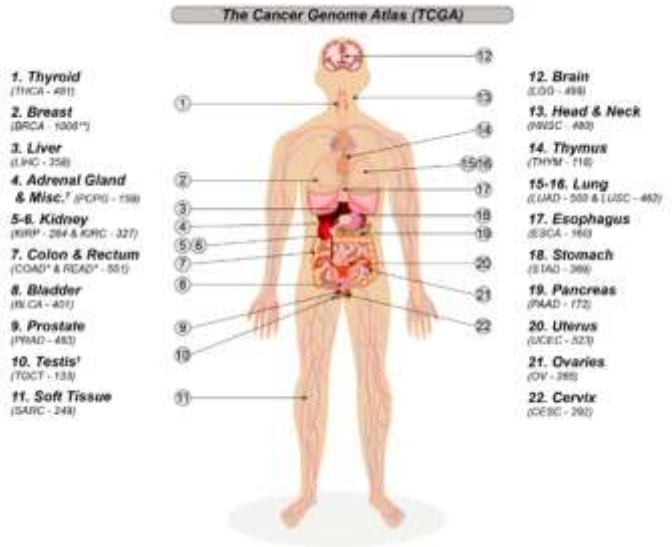
Intelligent scanning of sentence syntax to understand and validate findings in context, at scale

The combination of several A.I. methods create a proprietary ensemble A.I. strategy capable of revealing novel patterns and causal dependencies in disparate and varied biological data.

Ensemble Computational Intelligence

Enhanced Feature Reproducibility for Causal Statistical Learning



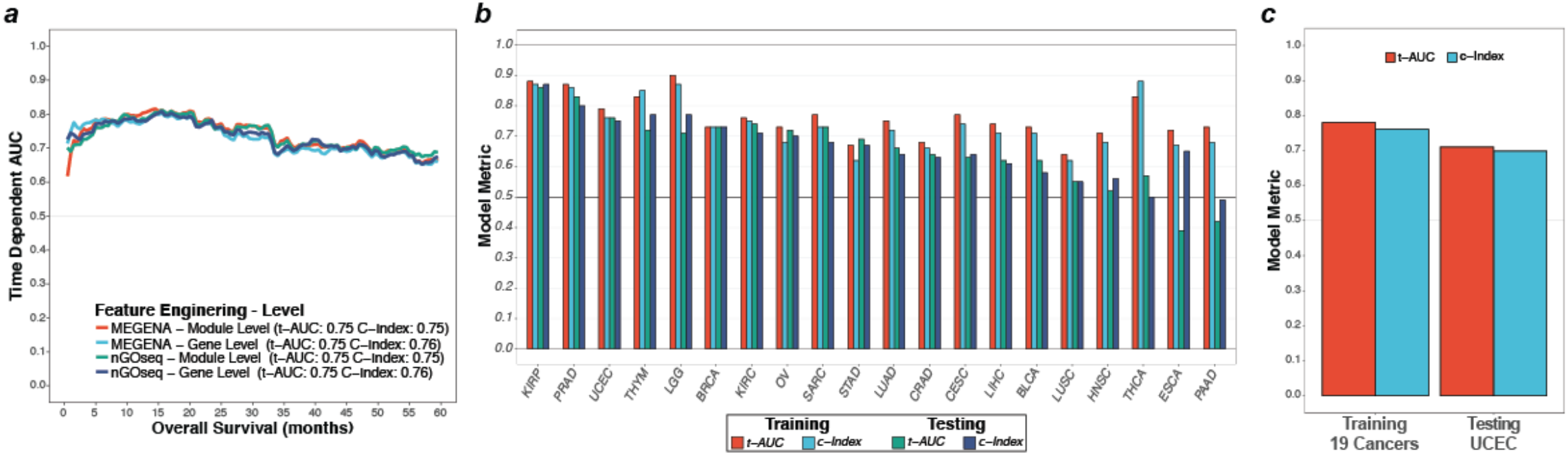


Multinomial Classification of 22 TCGA Cancer Types with Greater than 99.7% Accuracy = Disease Recognition

WuXiNextCODE

Large-scale clinical outcome study: TCGA Pan-Cancer Time-dependent Survival Analysis

Prediction of overall survival across 20 different cancers types with 75% accuracy



Data Matrix

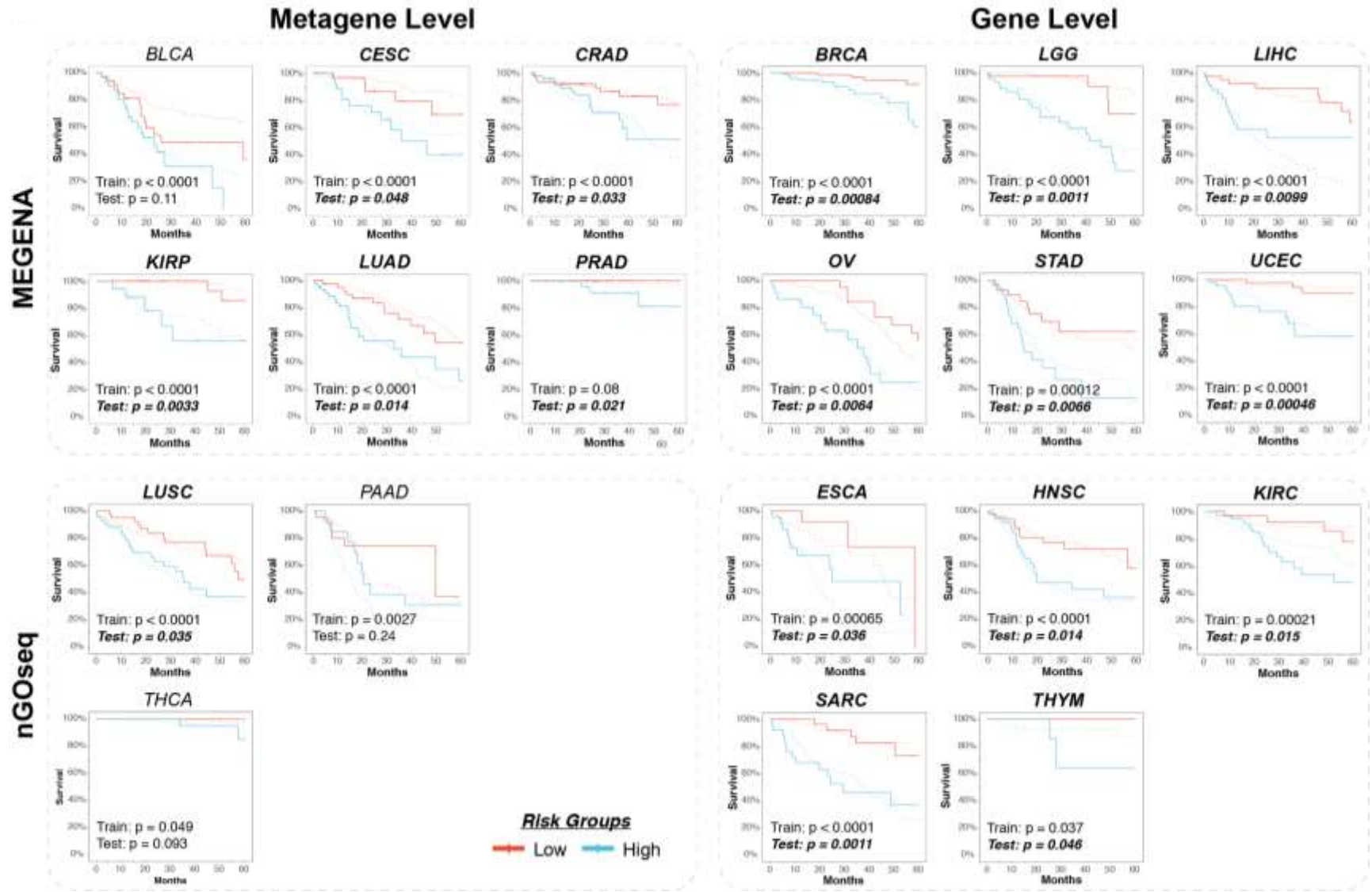
- 79k Molecular Features + 1 Clinical variable: Age
- 6,122 Training Samples
- 1,853 Testing Samples
- 20 Cancer Types

Interpretation: Compensating for overall survival instead of disease specific survival

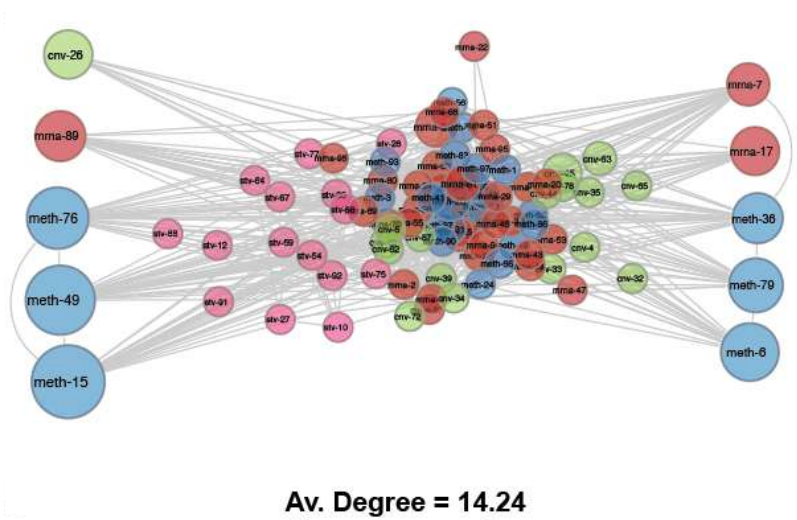
Estimator of cumulative/dynamic AUC for right-censored time-to-event data: Uno *et al. Journal of the American Statistical Association*, 2007

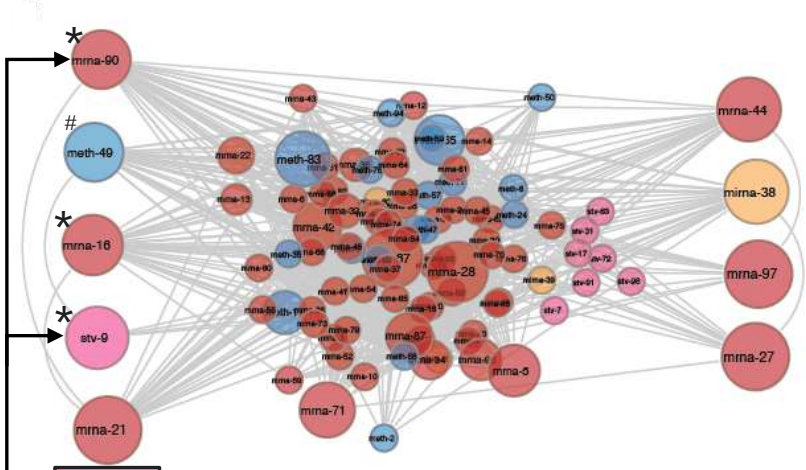
Large-scale clinical outcome study: TCGA Pan-Cancer Survival Analysis

Risk Stratification across 20 TCGA Cancers Types



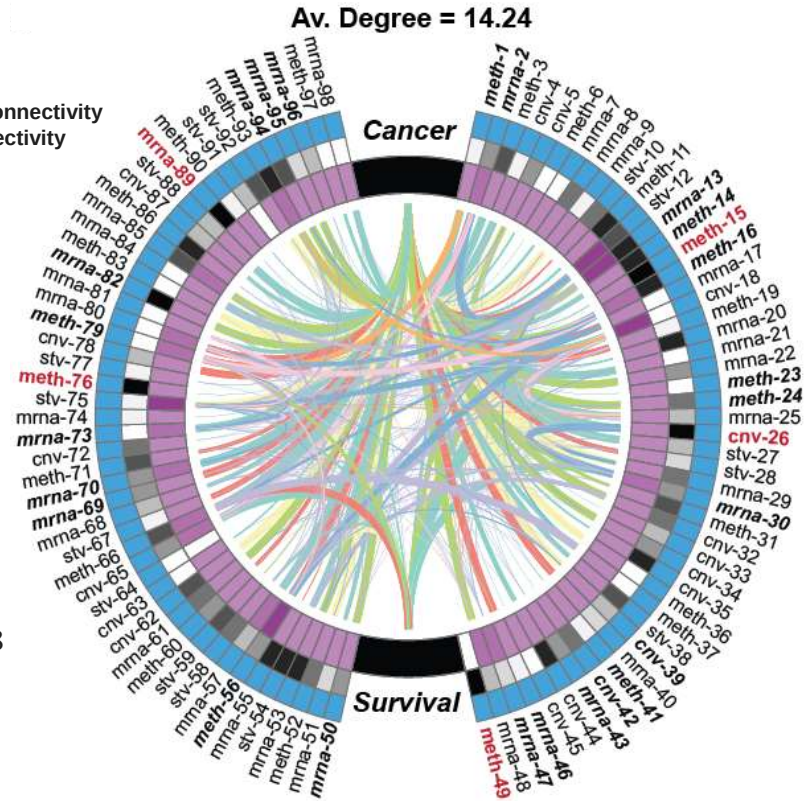
Blue = DNA Methylation
Red = mRNA
Orange = miRNA
Green = CNV
Pink = STV

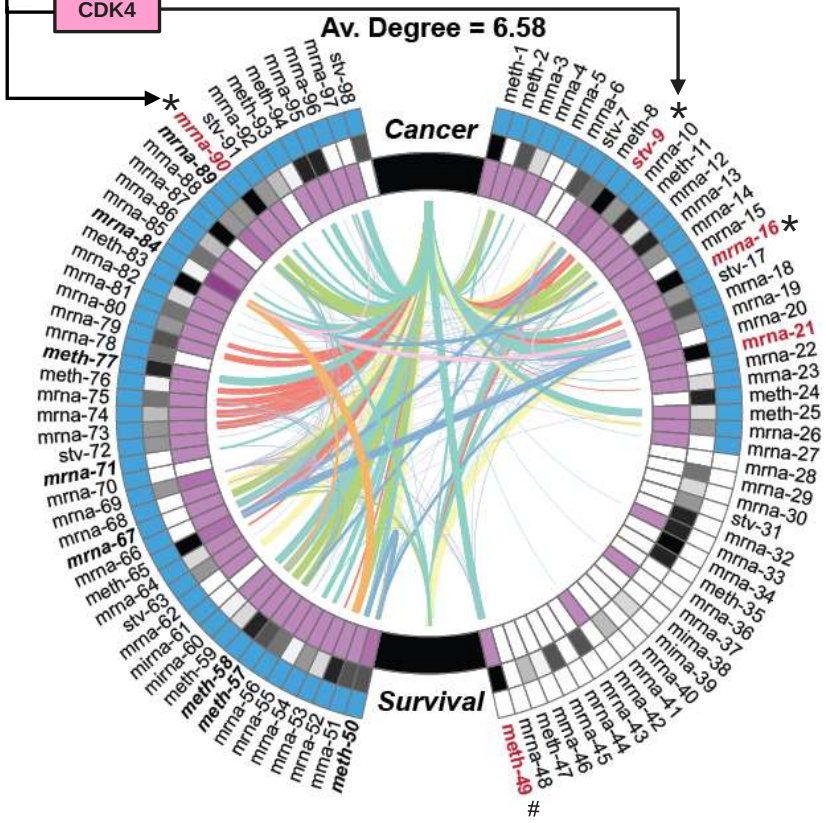




Bayesian Belief Networks

Purple Band = Degree NLP Network Connectivity
Black Band = Degree BNN Node Connectivity
Blue Band = Function Annotation
Red = BNN Driver Gene
Bold + Italic = Known Drug Target
*Driver Gene & Known Drug Target
#Driver Gene of Unknown Function





Natural Language Processing

nGOseq
DNA Methylation = 8
mRNA = 14
CNV = 2

MEGENA
DNA Methylation = 3
mRNA = 7
STV = 1

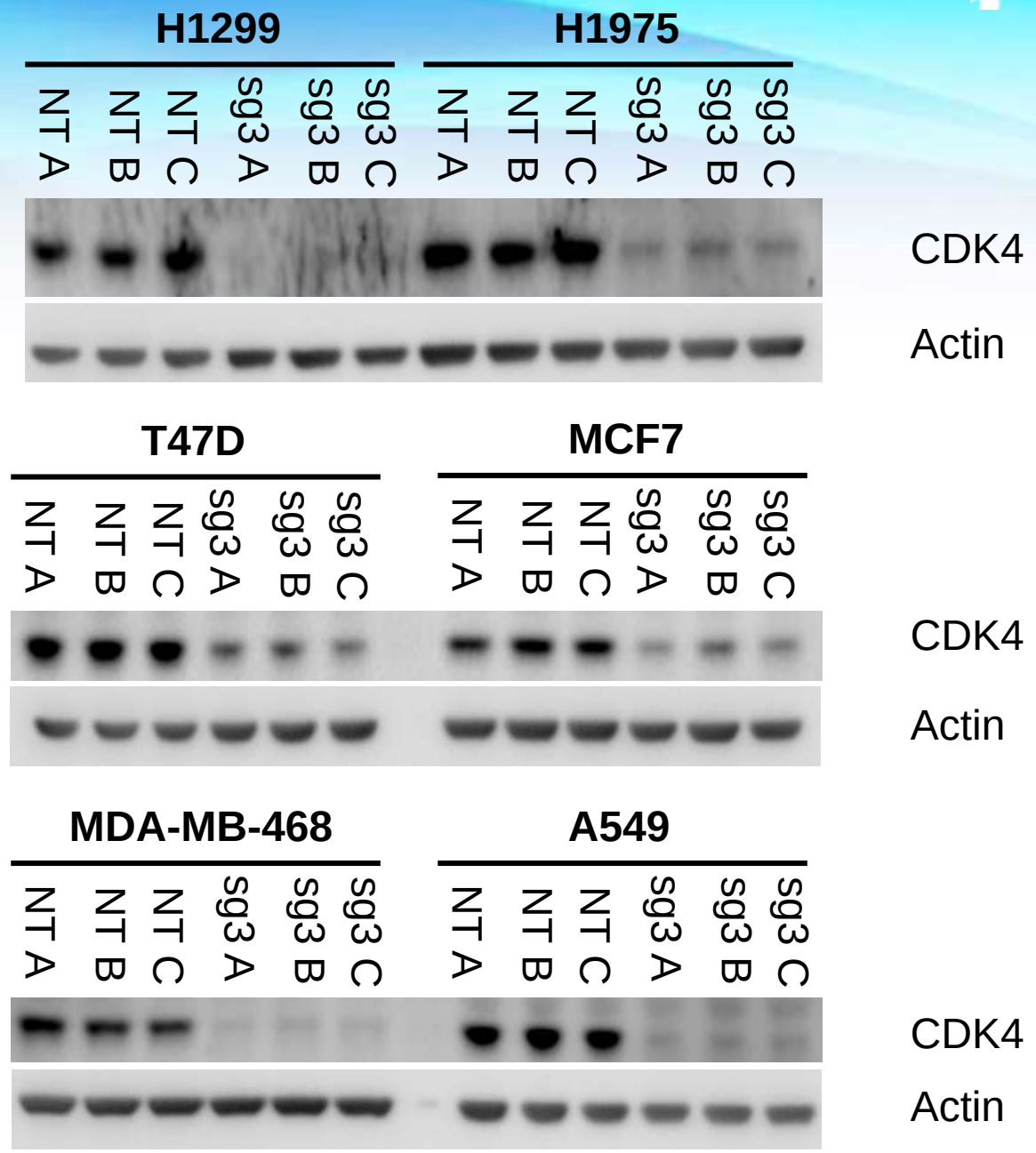
nGOseq

MEGENA

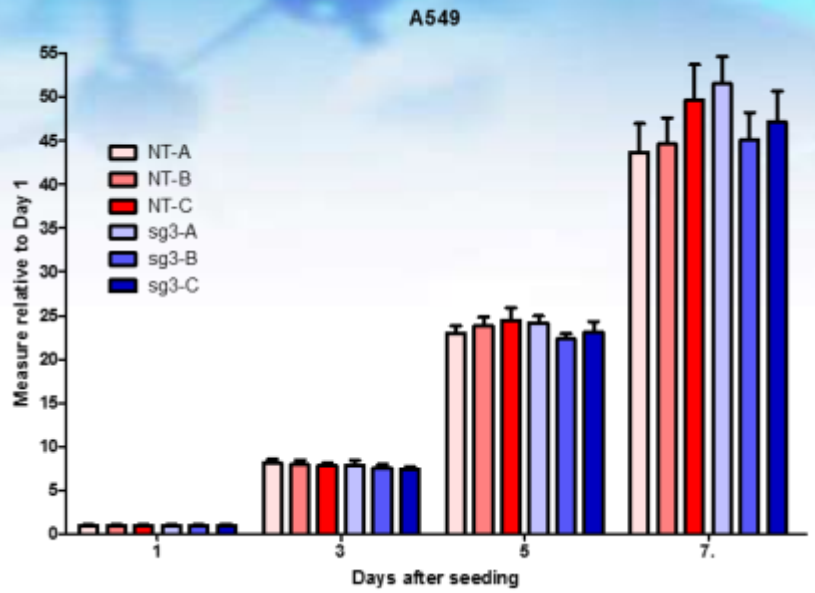
WuXiNextCODE

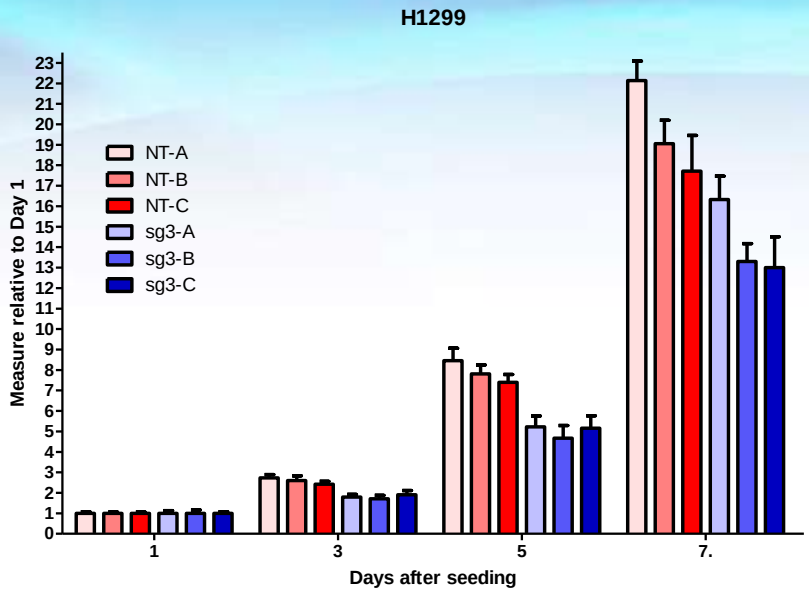
CDK4 KO confirmation by WB:

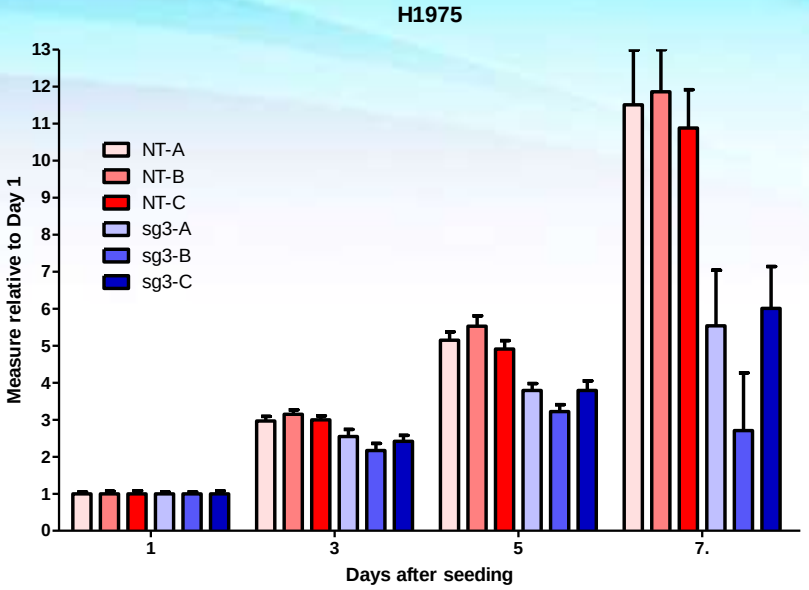
*__Approved CDK4/6 inhibitors__ for metastatic ER-positive/HER2-negative breast cancer: *Kisqali* (Norvartis), *Verzenio* (Lilly), and *Ibrance* (Pfizer).

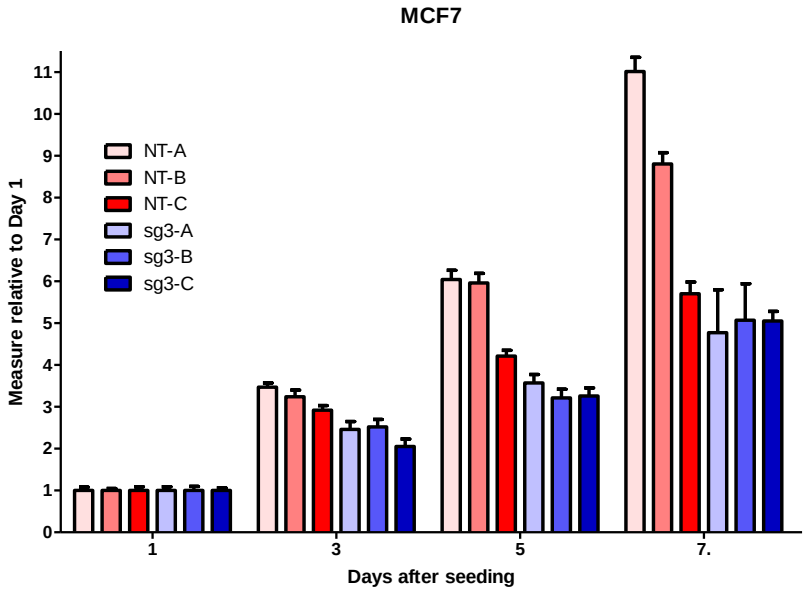


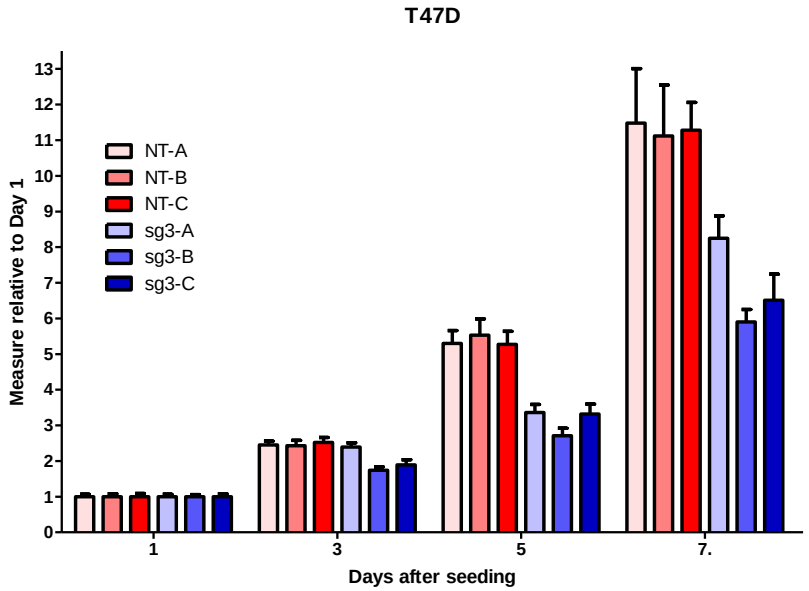
CDK4 KO vs NT Growth curves

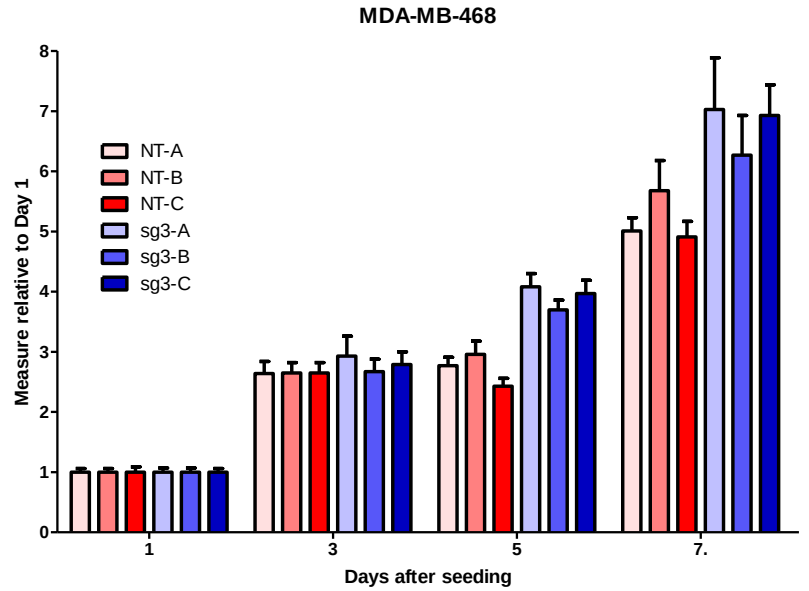




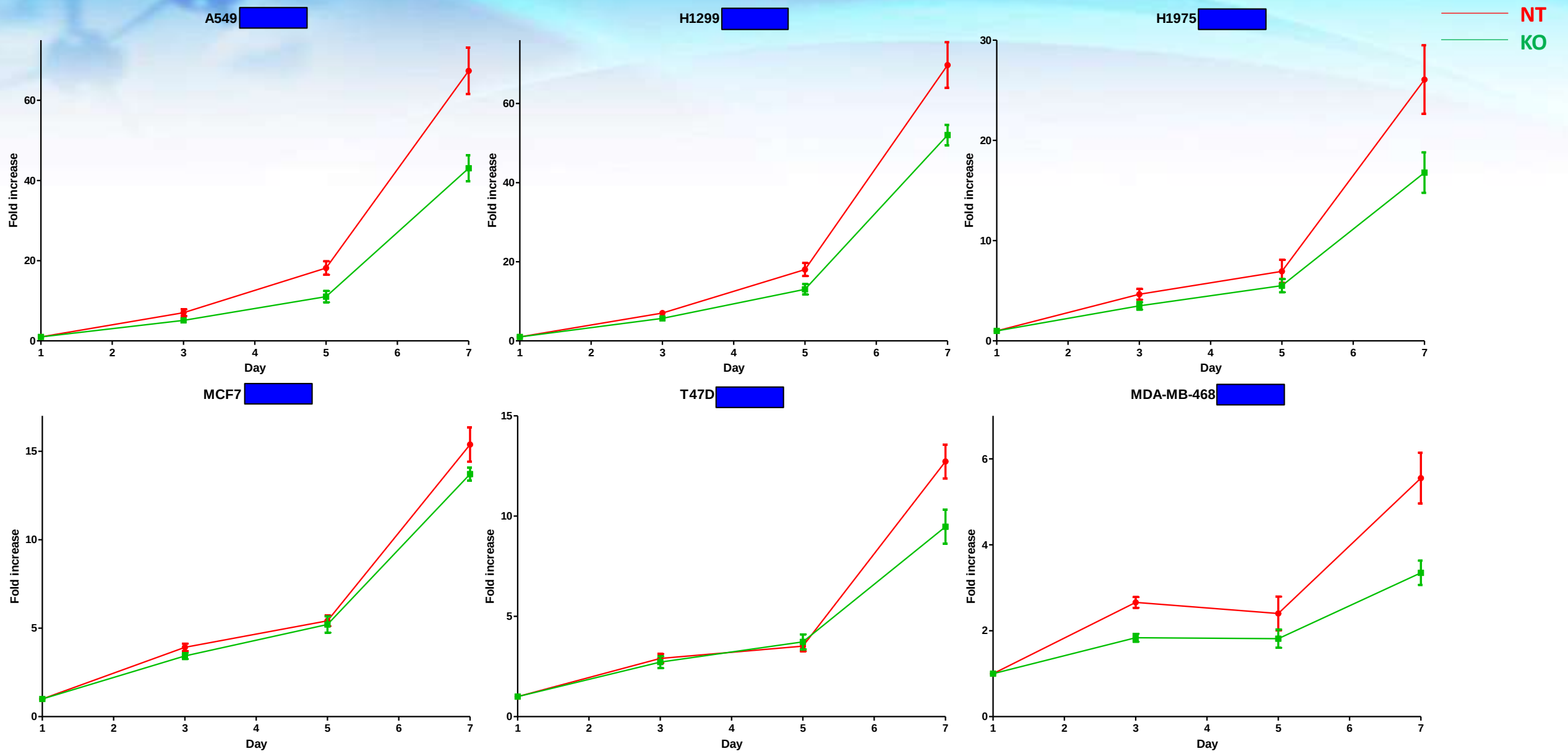






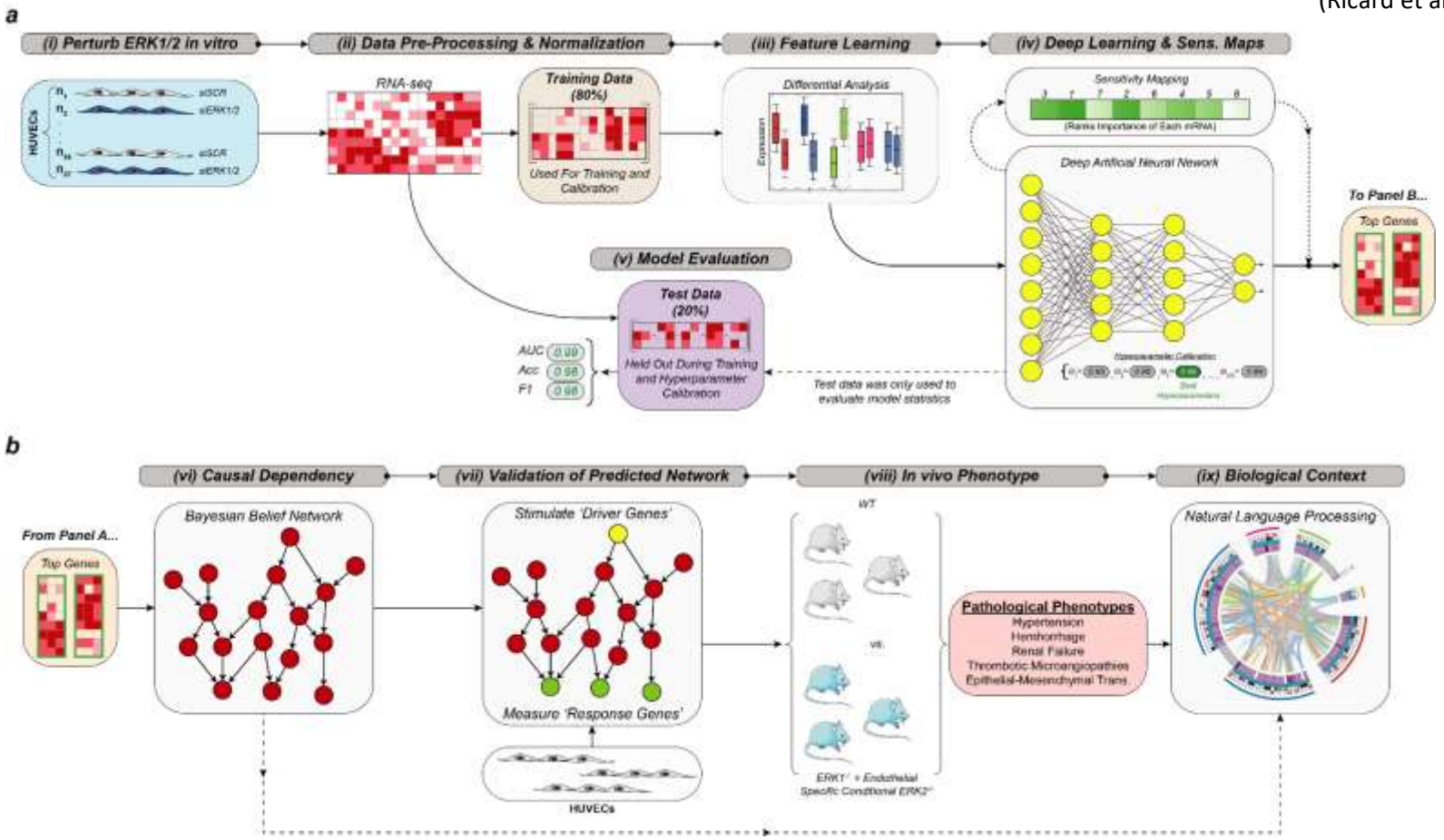


mRNA 90 KO vs NT Growth curves



Phenotype Projection: Identifying Causal Drivers of Cardiovascular Disease (Hypertension, Vascular Hemorrhage, and Renal Failure)

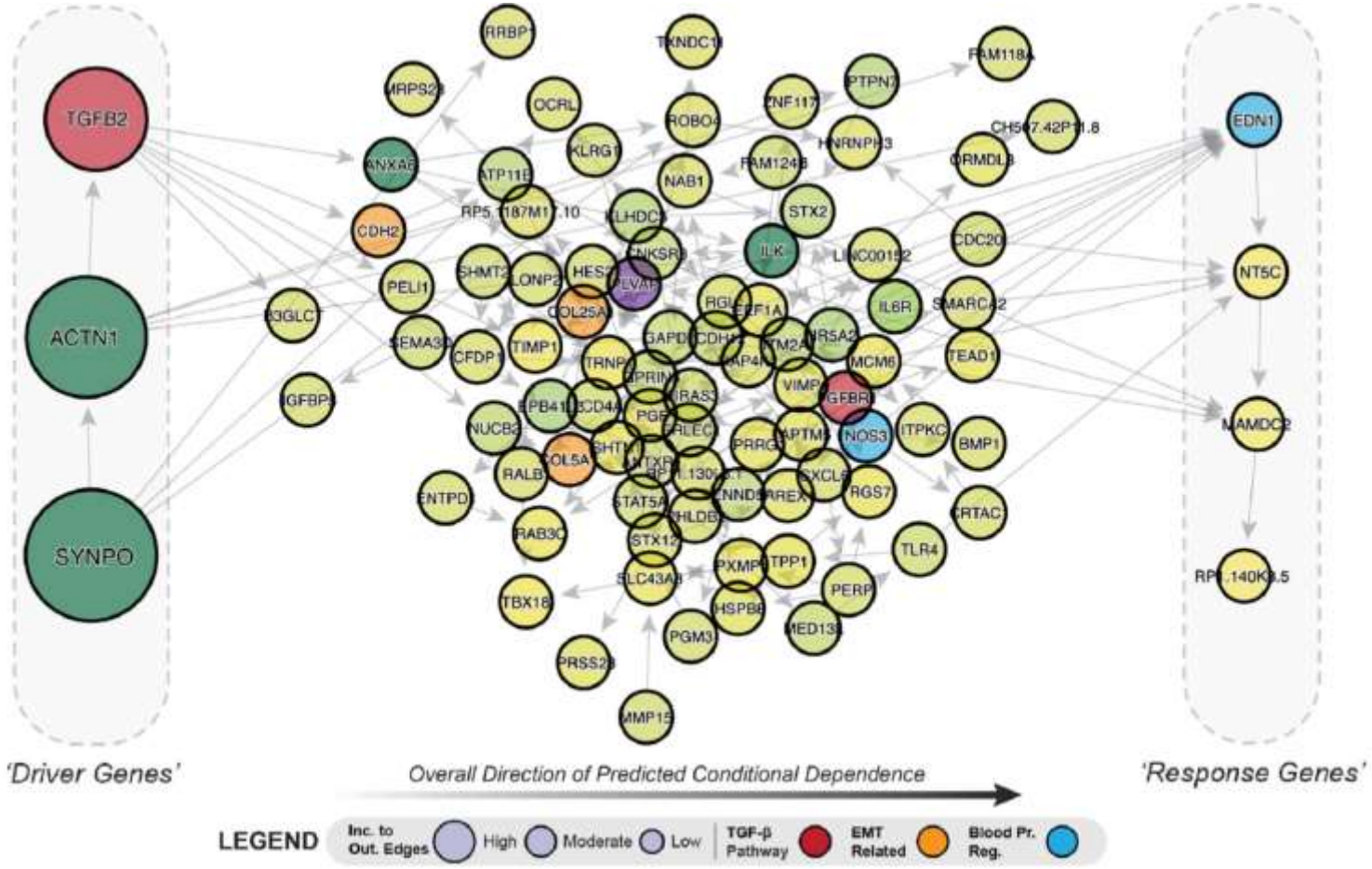
(Ricard et al., *JEM* 2019)

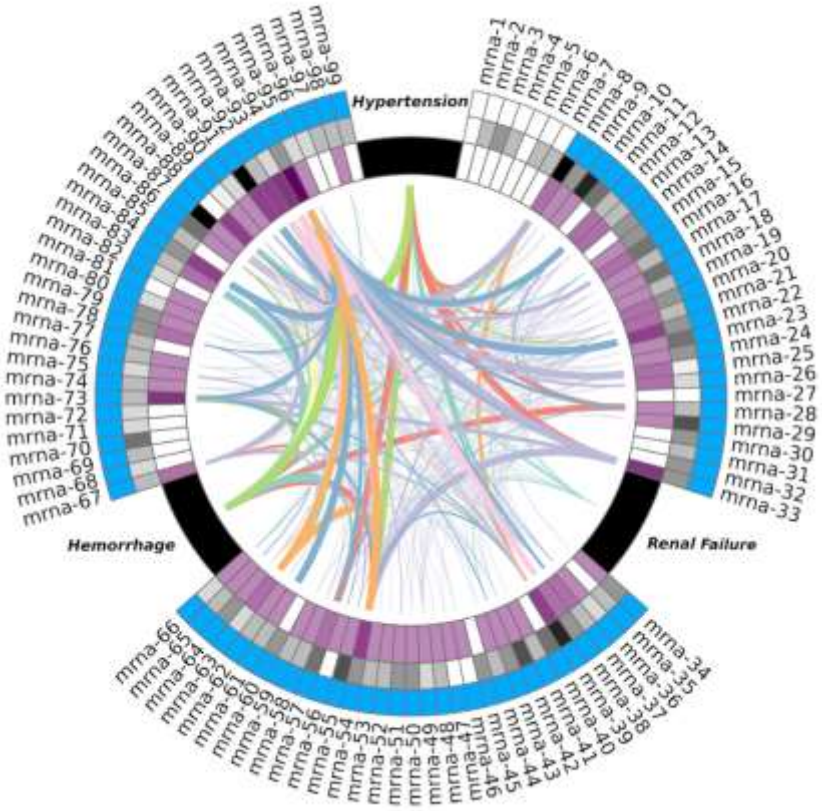


Research Collaboration with Yale Cardiovascular Research Center
Deep Learning, BBN Analysis, and NLP of Single Cell RNA-seq Data

Phenotype Projection: Identifying Causal Drivers of Cardiovascular Disease (Hypertension, Vascular Hemorrhage, and Renal Failure)

(Ricard et al., *JEM* 2019)

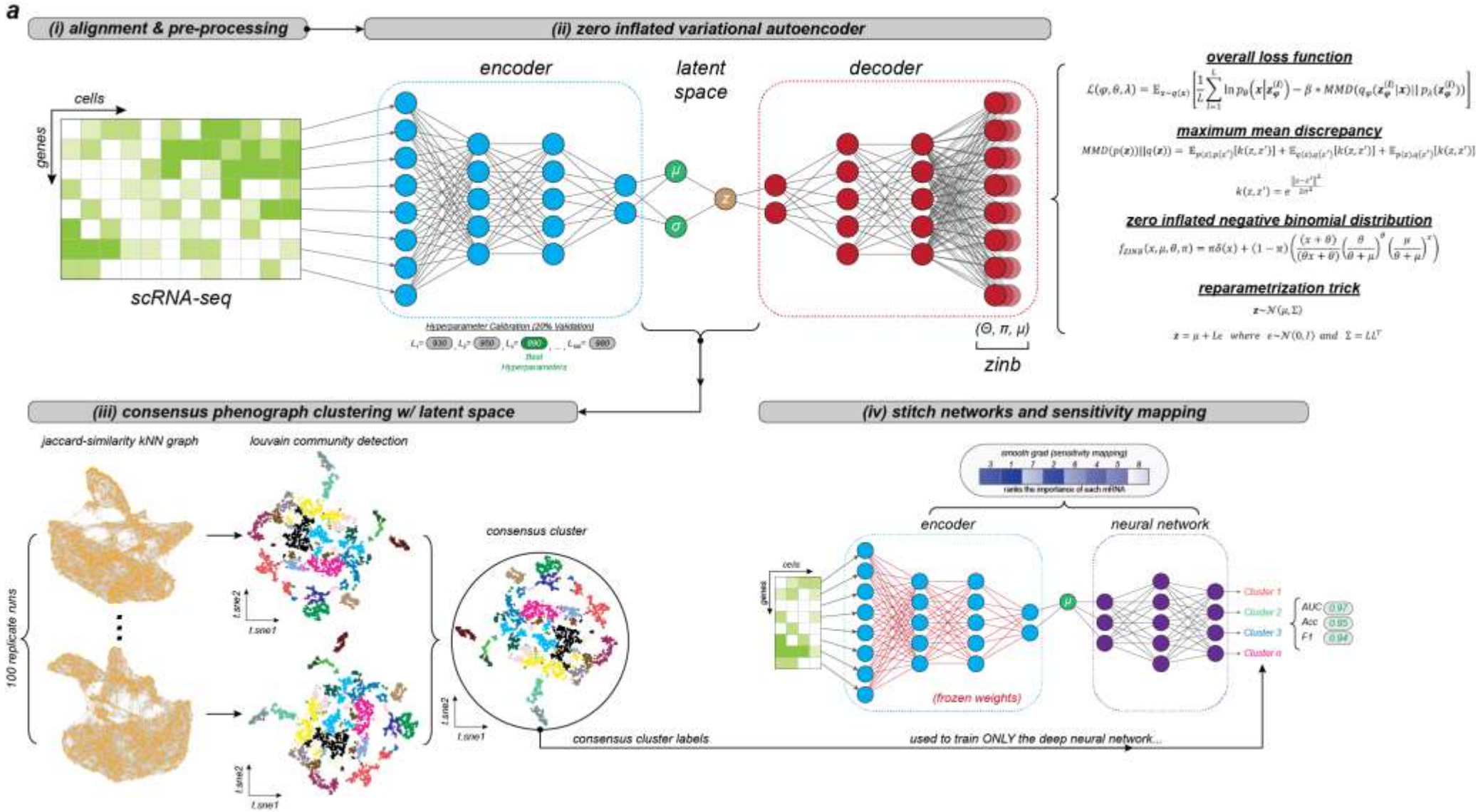




Research Collaboration with Yale Cardiovascular Research Center
Deep Learning, BBN Analysis, and NLP of Single Cell RNA-seq Data

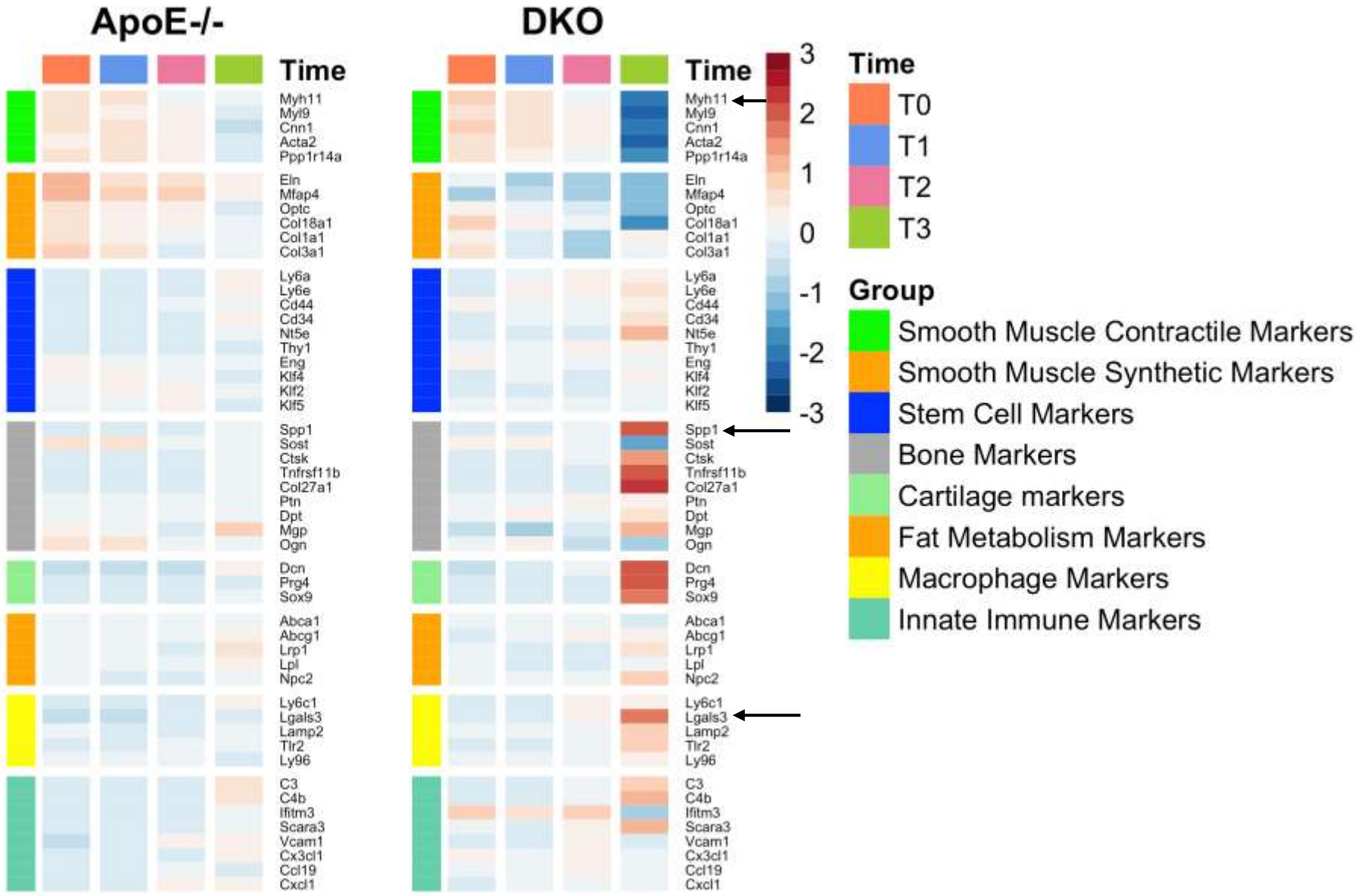
Identifying Causal Drivers of Cardiovascular Disease: Aortic Aneurysm and Atherosclerosis

(*Chen et al., Nature Metabolism 2019; Li et al., JCI In press*)

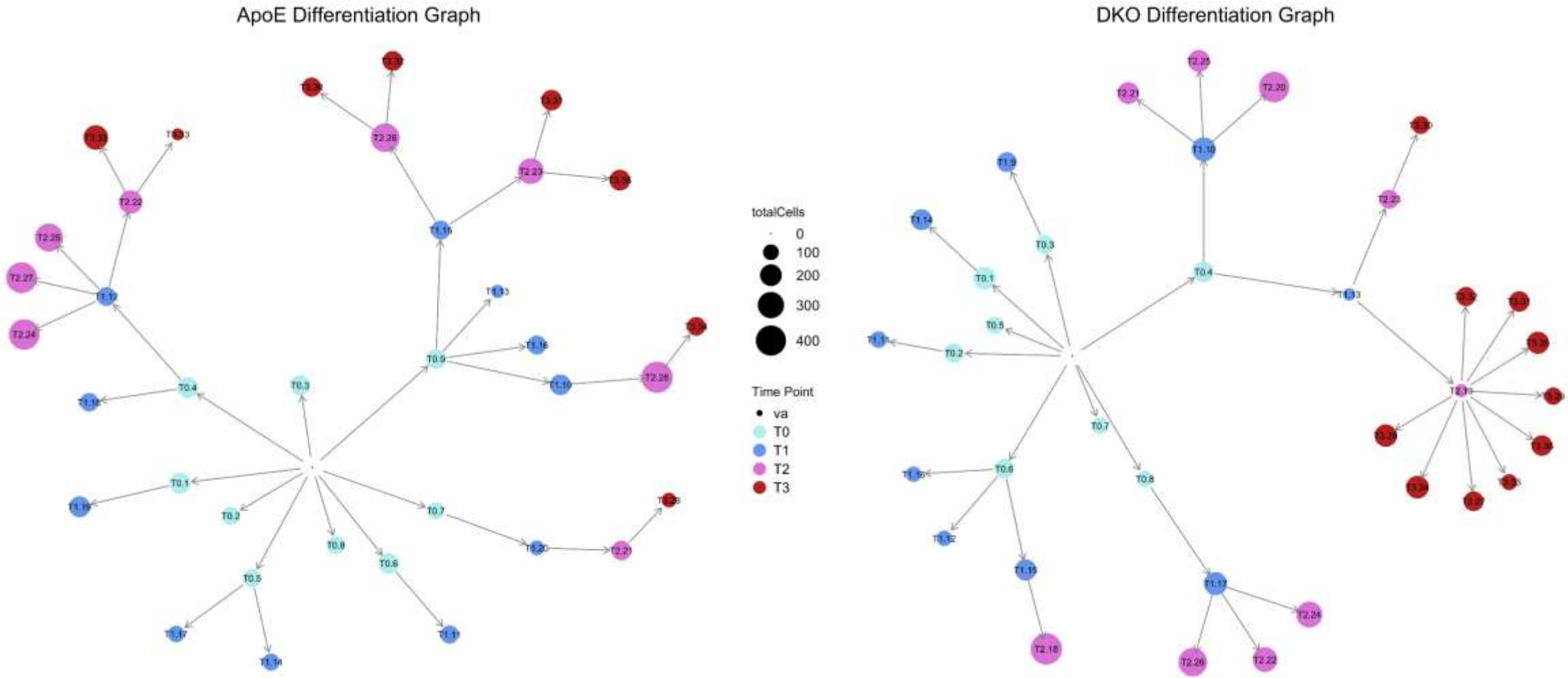


Research Collaboration with Yale Cardiovascular Research Center
Deep Learning, BBN Analysis, and NLP of Single Cell RNA-seq Data

Identifying Causal Drivers of Cardiovascular Disease: Analysis of Cellular Differentiation in Aortic Aneurysm

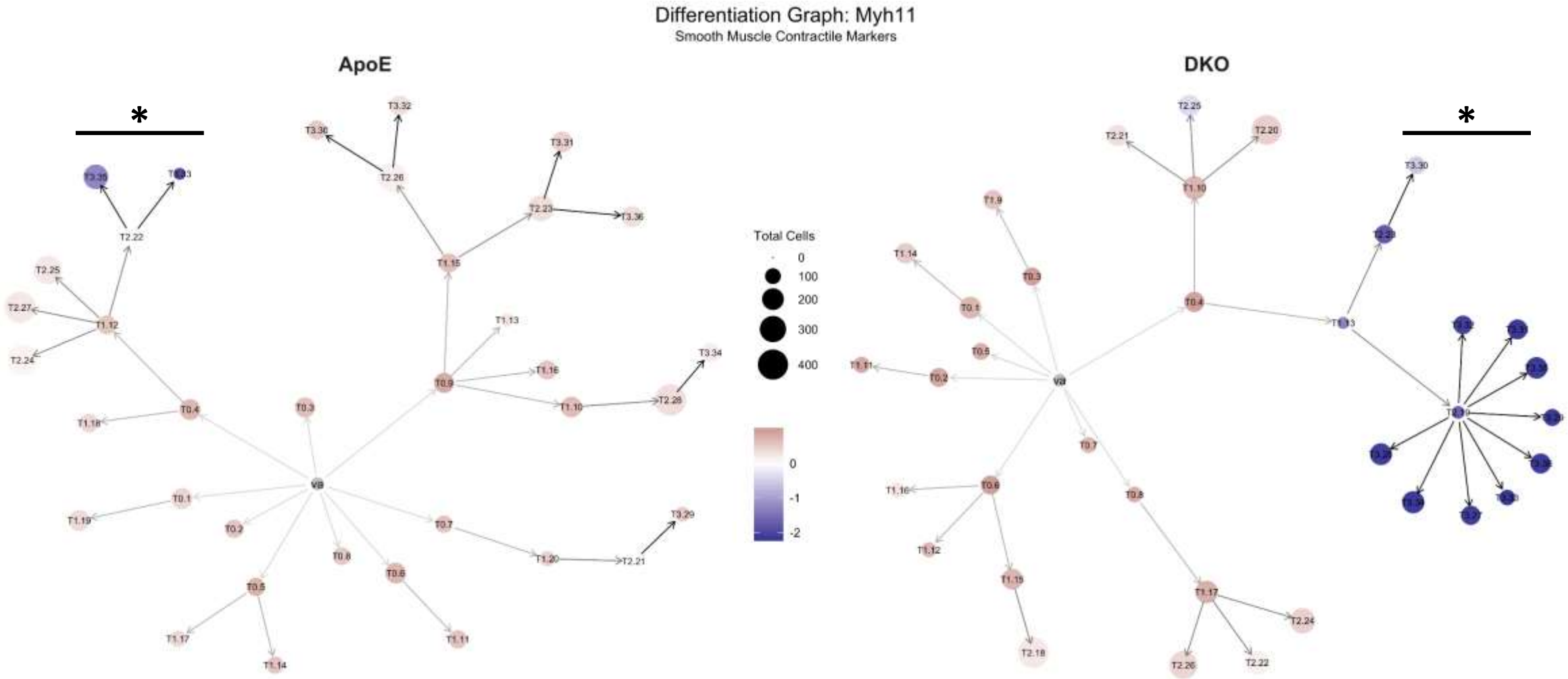


Identifying Causal Drivers of Cardiovascular Disease: Analysis of Cellular Differentiation in Aortic Aneurysm



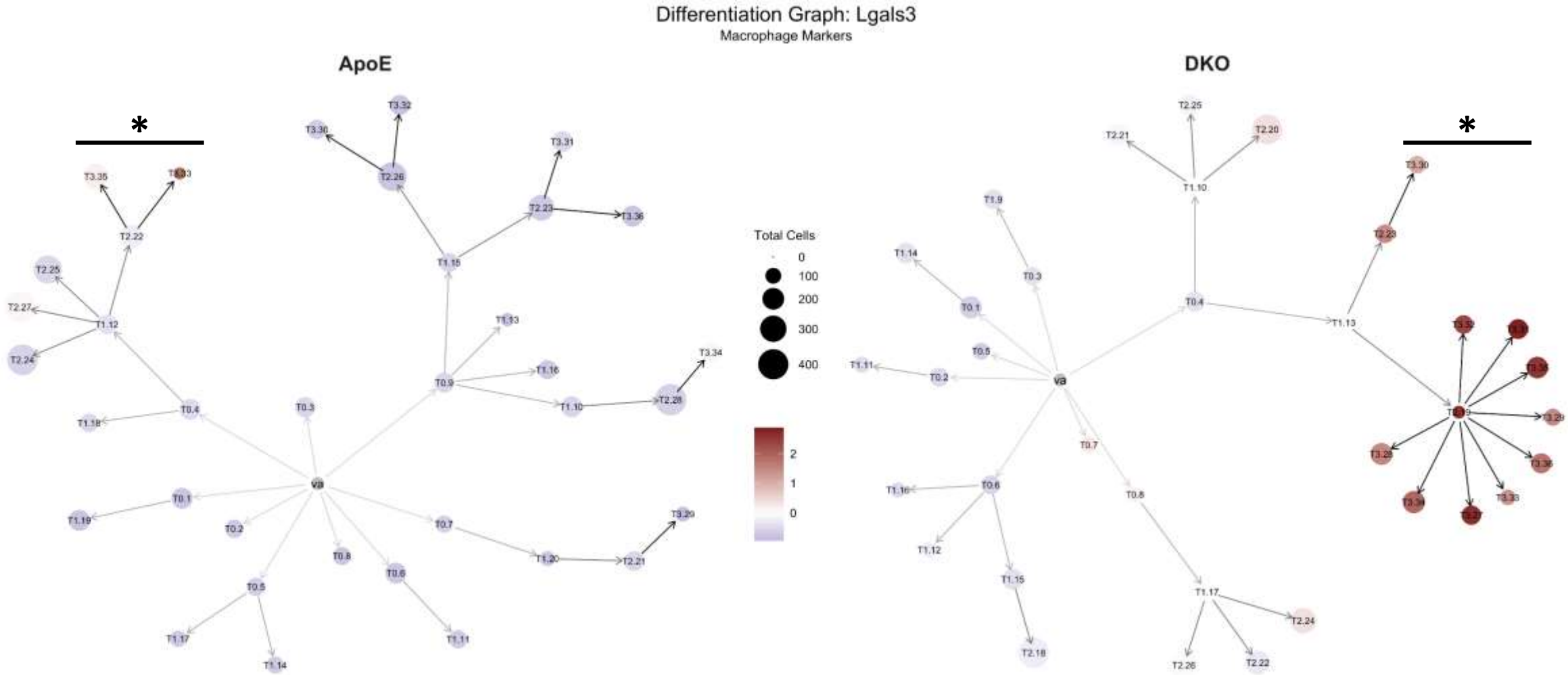
Novel means of tracking single cell differentiation across time. Holds significant commercial application in early phase clinical trials and drug efficacy studies.

Identifying Causal Drivers of Cardiovascular Disease: Analysis of Cellular Differentiation in Aortic Aneurysm



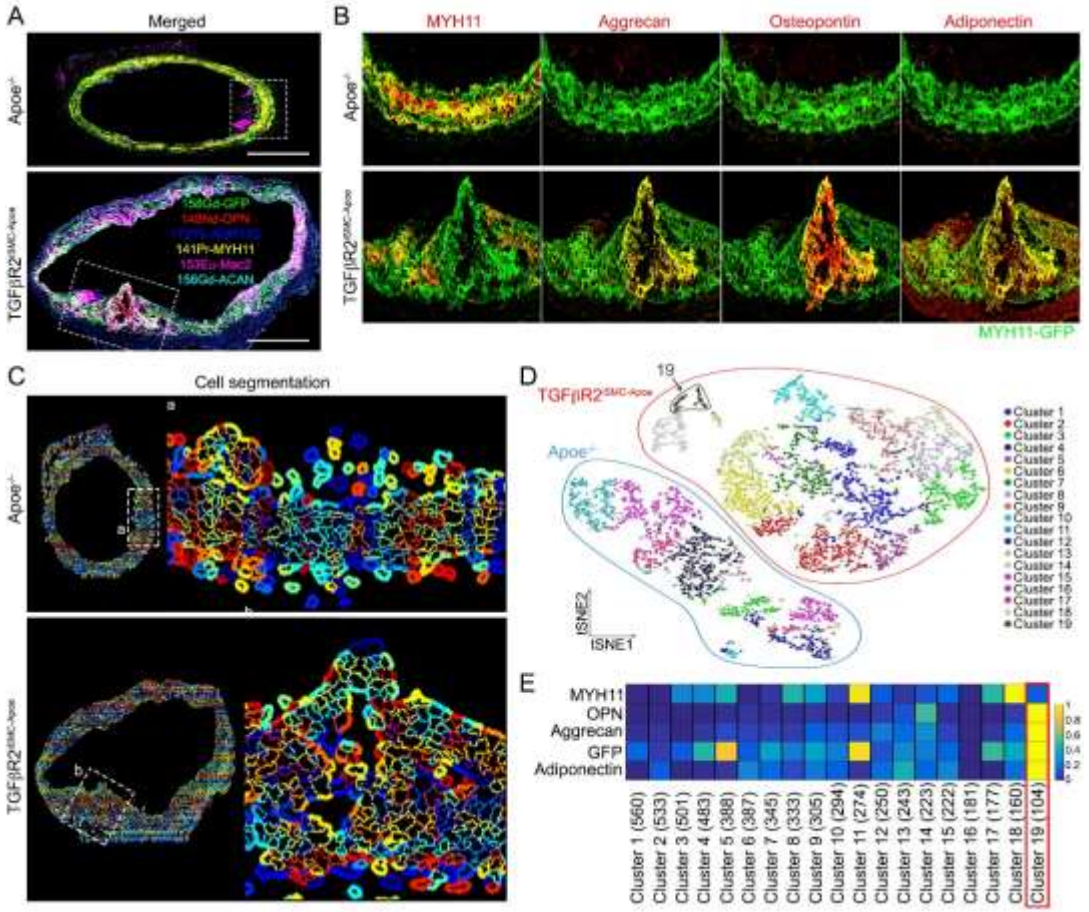
Novel means of tracking single cell differentiation across time. Holds significant commercial application in early phase clinical trials and drug efficacy studies. *Note degree of loss of Myh11 expression in distinct cell populations relative to two different experimental perturbation strategies.

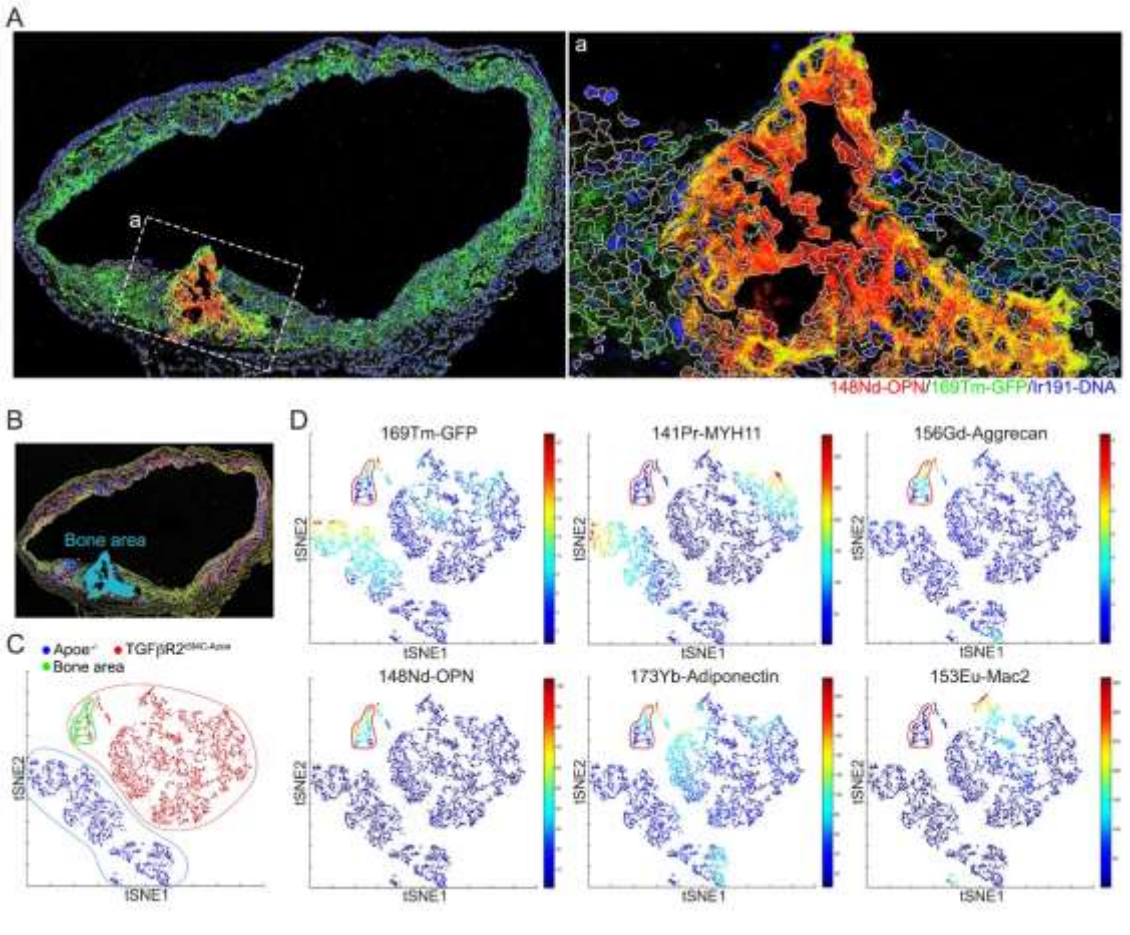
Identifying Causal Drivers of Cardiovascular Disease: Analysis of Cellular Differentiation in Aortic Aneurysm



Novel means of tracking single cell differentiation across time. Holds significant commercial application in early phase clinical trials and drug efficacy studies. *Note degree of gain in Lgals3 expression in same cell populations as in pervious Myh11 slide relative to two distinct experimental perturbation strategies.

Experimental Validation of ZI-VAE AI With Imaging Mass Cytometry





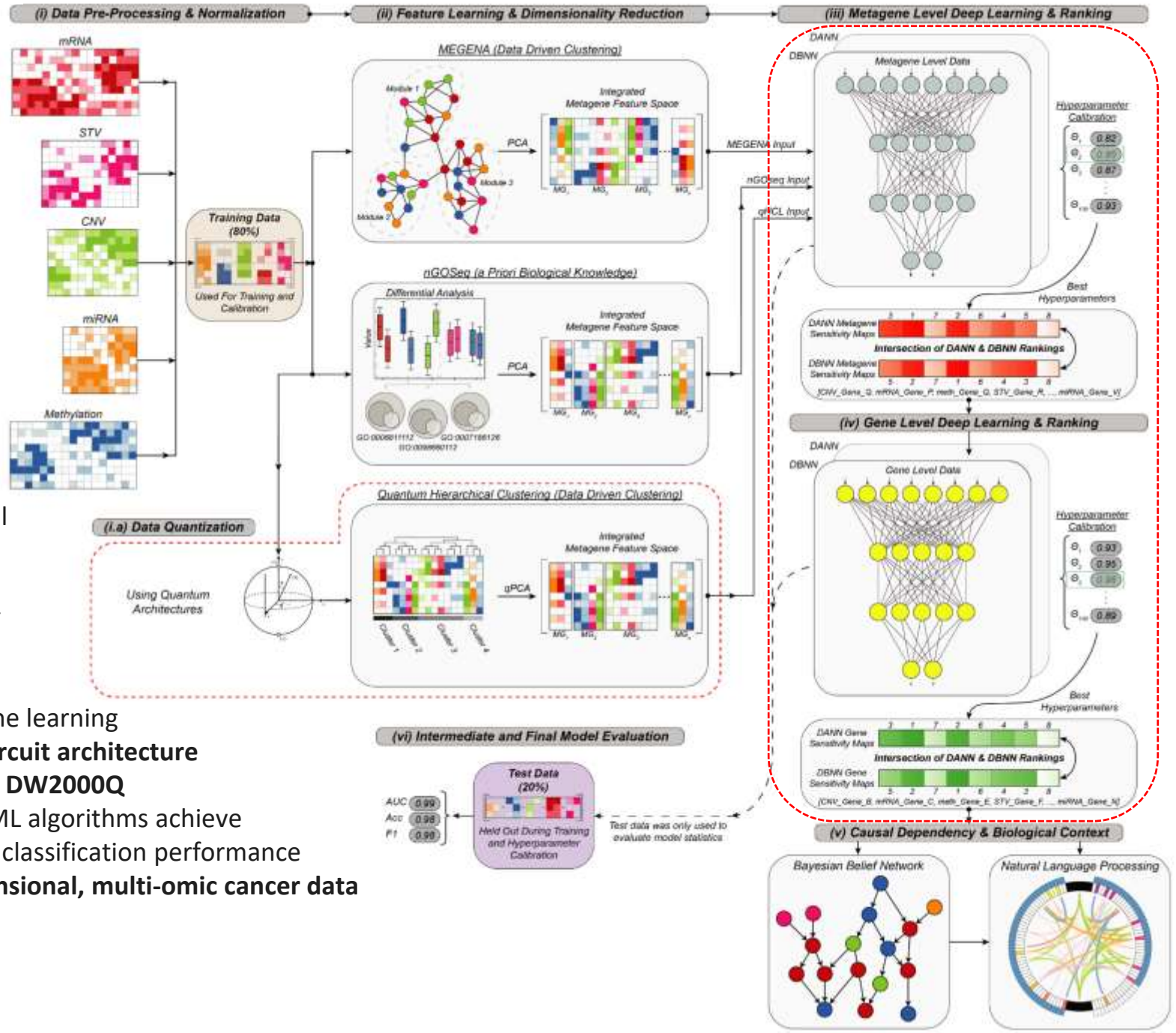
Quantum Machine Learning

- Quantum computing promises enhanced performance for many classes of problems associated with large datasets.

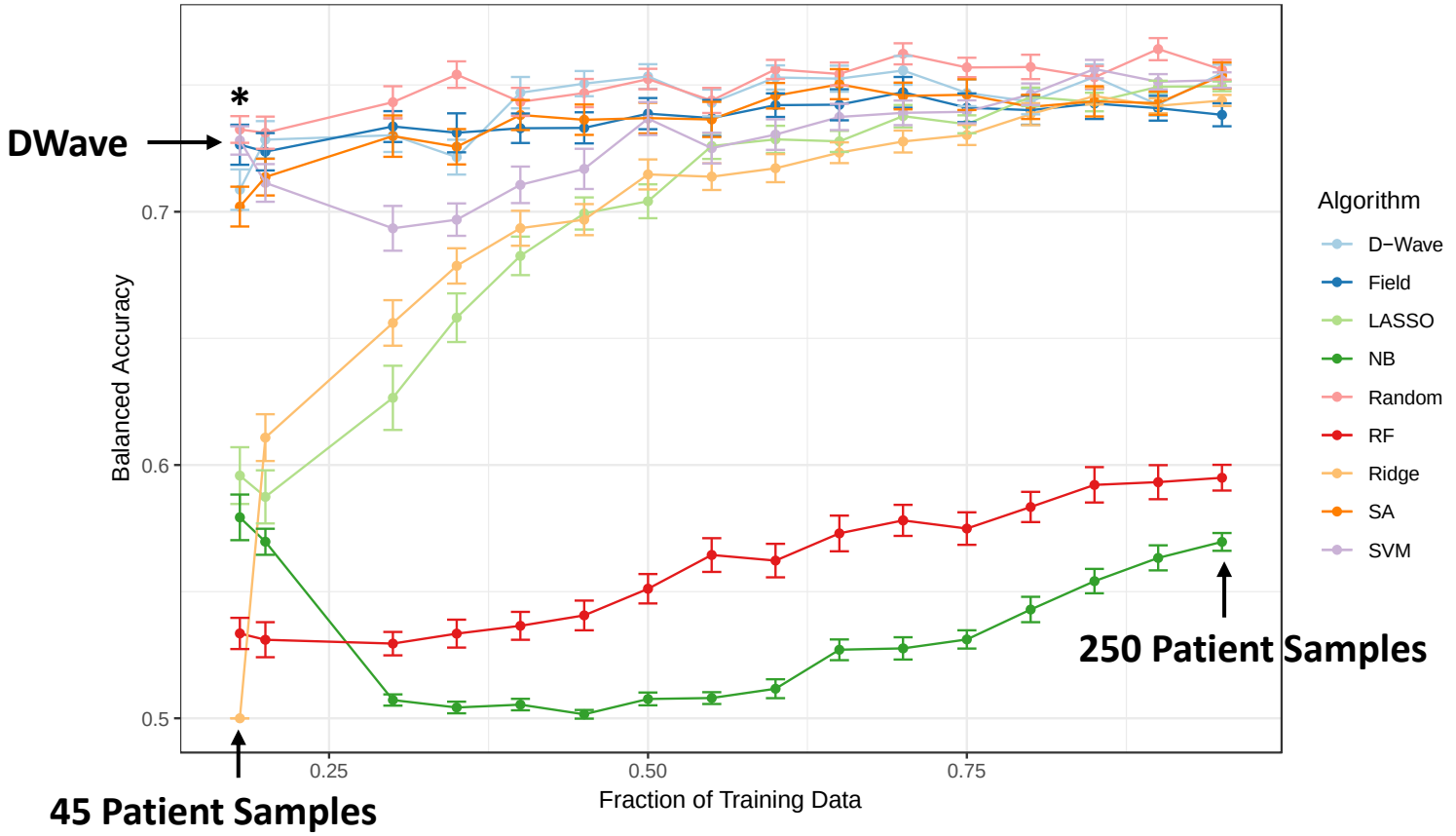
- We are in the process of replacing algorithmic components of our **Ensemble Computational Intelligence Strategy** with their respective quantum counterparts.

- Our first algorithm was a quantum hierarchical clustering (qHCl), based on a modified **Grover's algorithm**, a quantum **search algorithm** that runs quadratically faster than any equivalent classical algorithm.

- We have now built statistical quantum machine learning classifiers on both **IBM's universal quantum circuit architecture** and the **D-Wave Two X (DW2X) processor and DW2000Q Adiabatic quantum computer**. Our D-Wave qML algorithms achieve comparable, and in some cases slightly better, classification performance than their classical counterparts on **high-dimensional, multi-omic cancer data from the Cancer Genome Atlas (TCGA)**.



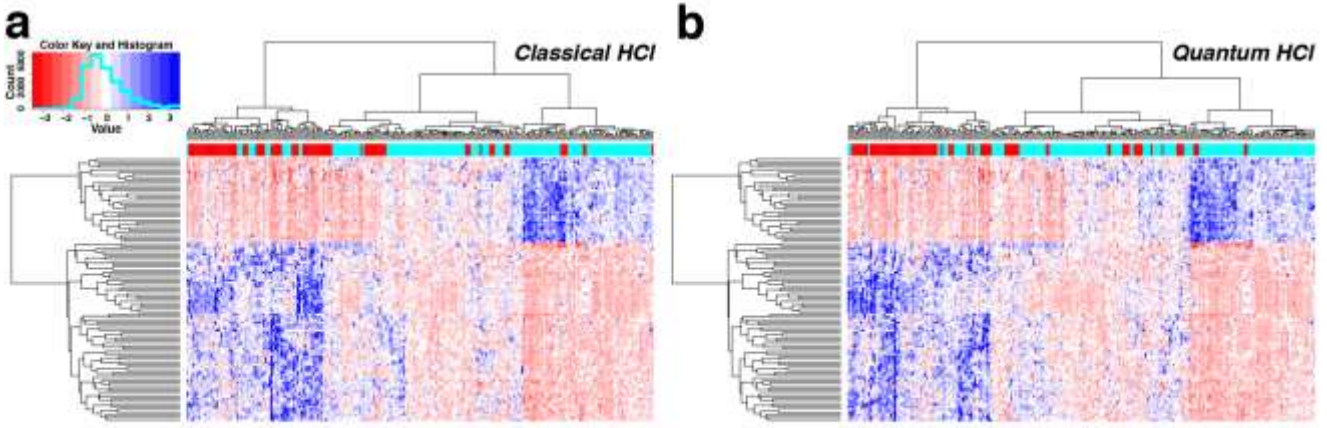
Binomial Classification of Tumor Molecular Subtypes
Luminal A vs. Luminal B Human Breast Cancers



LumA vs. LumB Status	
Tumor Samples	311
Luminal A	199
Luminal B	112
Train	250
Test	61

***We have developed a novel solution for the Ising problem and statistical optimization. Significant commercial application in early phase clinical trials and drug efficacy studies. High-profile research manuscript in preparation.**

Binomial Classification of Tumor Molecular Subtypes: Luminal A vs. Luminal B Human Breast Cancers

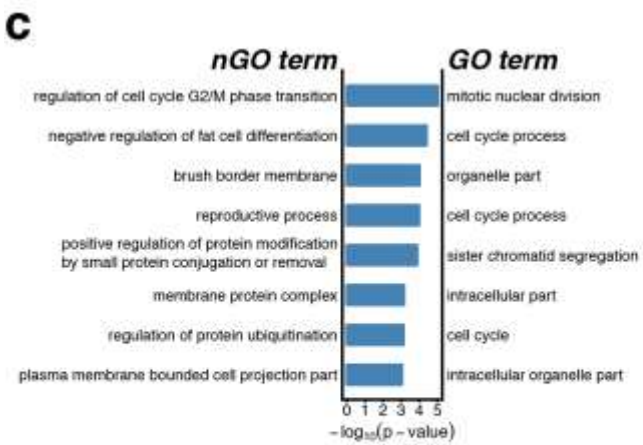


LumA vs. LumB Status	
Tumor Samples	311
Luminal A	199
Luminal B	112
Train	250
Test	61

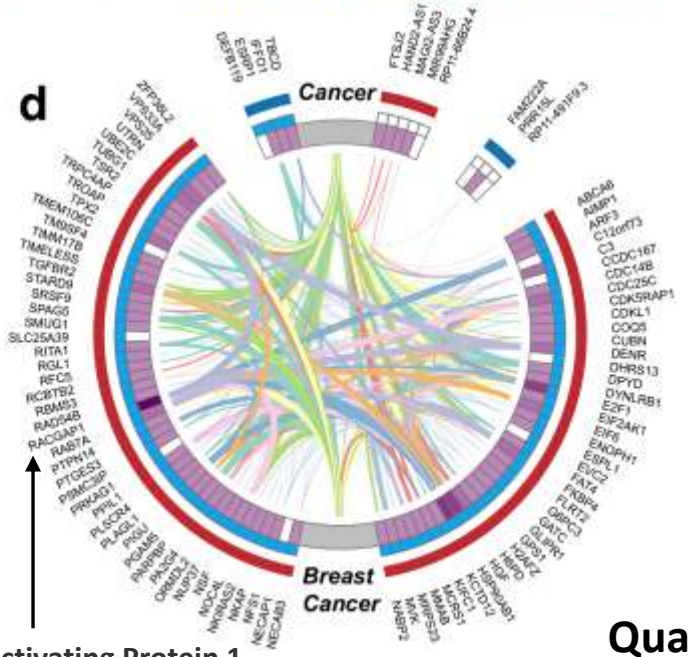
*Quantum and classical trees are 88% concordant based on the standard Robinson–Foulds metric

*qHCL - Durr-Hoyer method based on a modified Grover's search algorithm with Euclidean distance and Ward linkage

*qHCL ran on a IBM quantum simulator using 19 qubits



Functional Enrichment



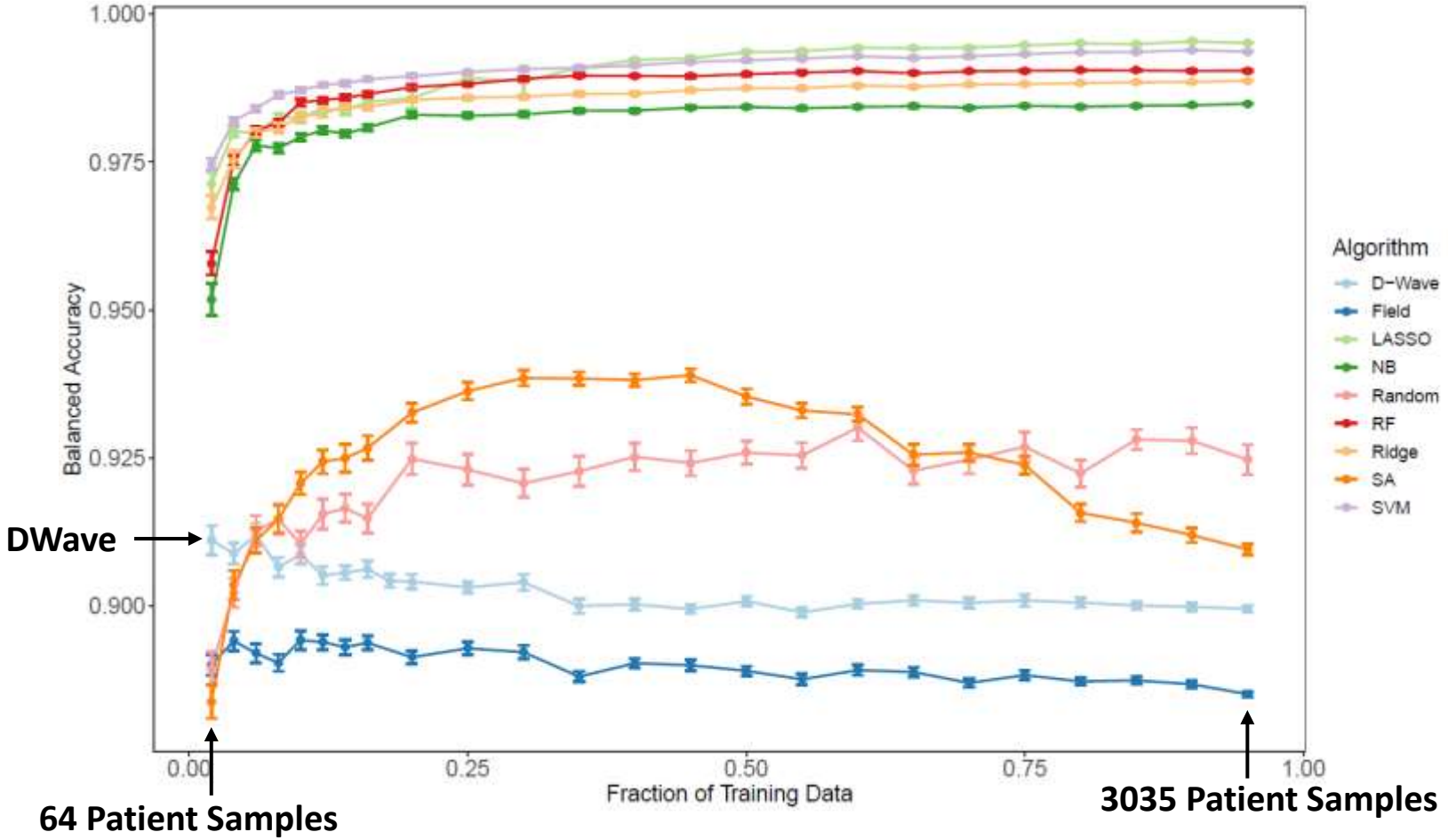
Rac GTPase Activating Protein 1
(proposed oncogene)

Natural Language Processing

*Outer red band: mrna data
*Outer blue band: methylation data
*Inner blue band: genes of known function

Quantum machine learning identifies same top 100 genes as classical machine learning

Multinomial Classification of Human Cancer Types
Quantum Machine Learning



Human Cancer Types	Sample Number
Liver Hepatocellular Carcinoma	358
Breast cancer	1006
Brain Lower Grade Glioma	499
Colon Adenocarcinoma/ Rectum Adenocarcinoma	551
Kidney Cancer	611
Lung Cancer	962
Total	3987
Train	3190
Test	797

***We have developed a novel solution for the Ising problem and statistical optimization. Significant commercial application in early phase clinical trials and drug efficacy studies. High-profile research manuscript in preparation.**

Advanced Artificial Intelligence Research Laboratory

Academic and Industry Research Collaborations

Harvard Medical School
Professor Chris Walsh
Chief of Genetics and Genomics

University of Oxford
Professor Chris Holmes
Computational Statistics and Machine Learning

University of Southern California
Professor Daniel Lidar
Quantum Computing and Quantum Machine Learning

University of Toronto
Professor Alán Aspuru-Guzik
Quantum Chemistry and Chemical Biology

WuXi AppTec Oncology
Fabrice Alphonse

Yale University School of Medicine
Professor Michael Simons
Director of Yale Cardiovascular Research Center

Professor Karen Hirschi
Yale Cardiovascular Research Center

WuXiNextCODE

Acknowledgements

WuXi NextCODE
- Jeff Gulcher
- Richard Williams
- Rob Brainin
- Hákon Guðbjartsson
- Hongye Sun
- Jim Lund
- Shannon Bailey
- Irene Blat

WuXi NextCODE Advanced A.I. Research Laboratory
- Sharvari Gujja
- Joe White
- Sweta Bajaj
- Kris Holton
- Jose Malagon Lopez
- Richard Li

Boston Children's Hospital Harvard Medical School
- Chris Walsh
- Mike Lodato

Cardiovascular Research Center Yale University Medical School
- Mike Simons
- Pengchun Yu

University of Oxford
- Chris Holmes

University of Tennessee
- Hao Chen

WuXiNextCODE

APPENDIX

WuXiNextCODE

14 DECEMBER | TWO THOUSAND EIGHTEEN

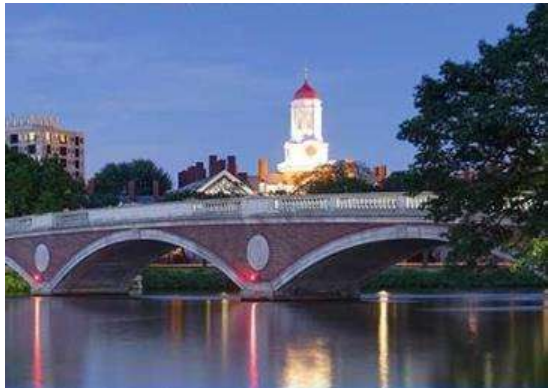
WuxiNextCODE: A Global Contract Genomics Organization

Natively Global, rapid expansion – 700+ employees, raised $260 million (Oct 2017)
Nov 2018: GMI Acquisition and $200 million investment

ICELAND

- Birthplace of population genomics
- Database, Clinical Interpretation, Sequence Analysis development

deCODE genetics
a subsidiary of AMGEN



US

- Global capital of life sciences
- World-leading clinical, deep learning capabilities

NEXTCODE HEALTH

CHINA

- WuXi – the quality leader in Life Sciences with pharma
- CLIA, CAP certified laboratory in China

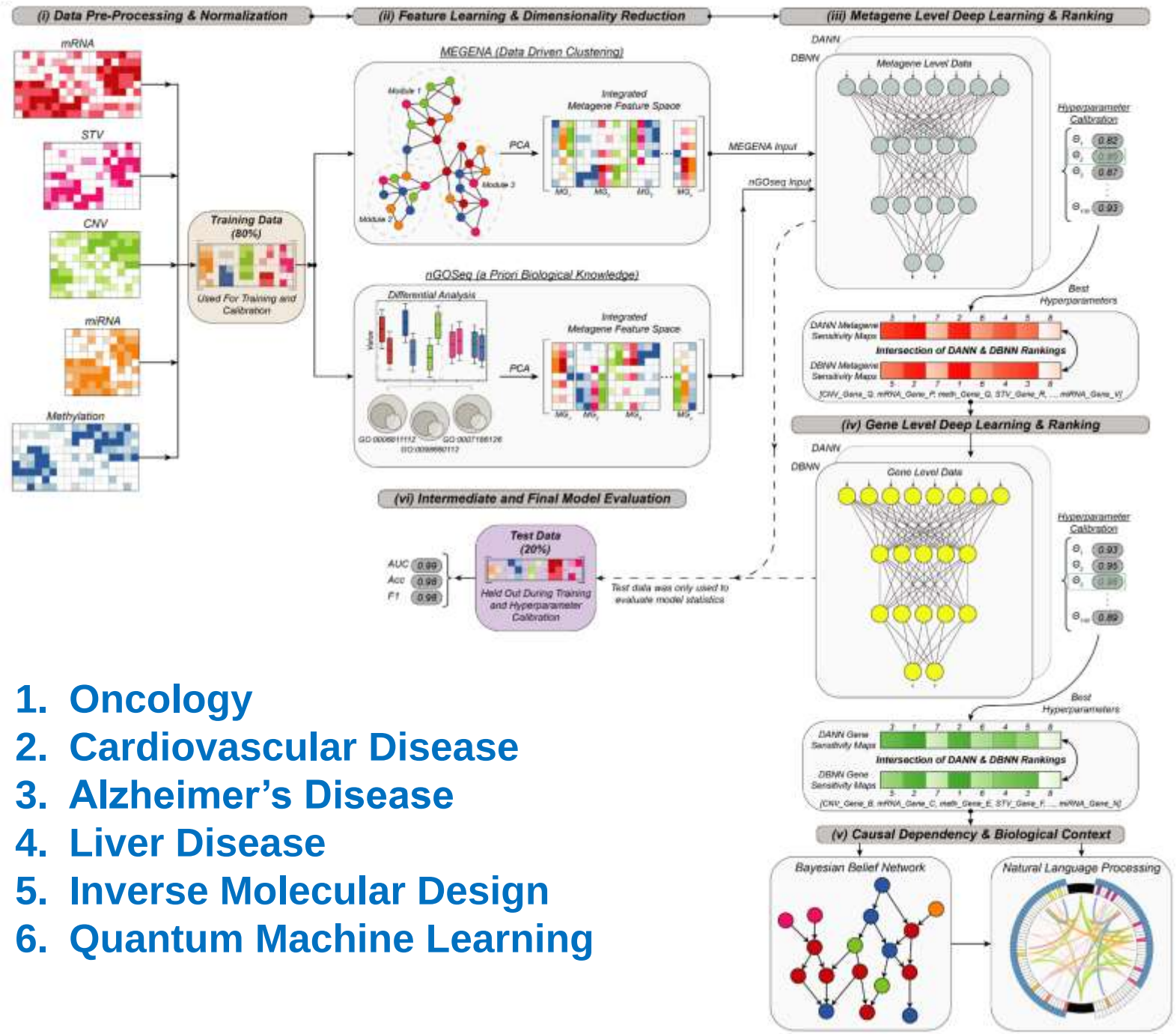
药明康德
WuXi AppTec

IRELAND

- GMI – now a wholly-owned subsidiary of WXNC
- Recruit & Whole genome sequence (WGS) 400,000 of the Irish population

gmi genomics medicine ireland

Ensemble Computational Intelligence



1. Oncology
2. Cardiovascular Disease
3. Alzheimer's Disease
4. Liver Disease
5. Inverse Molecular Design
6. Quantum Machine Learning

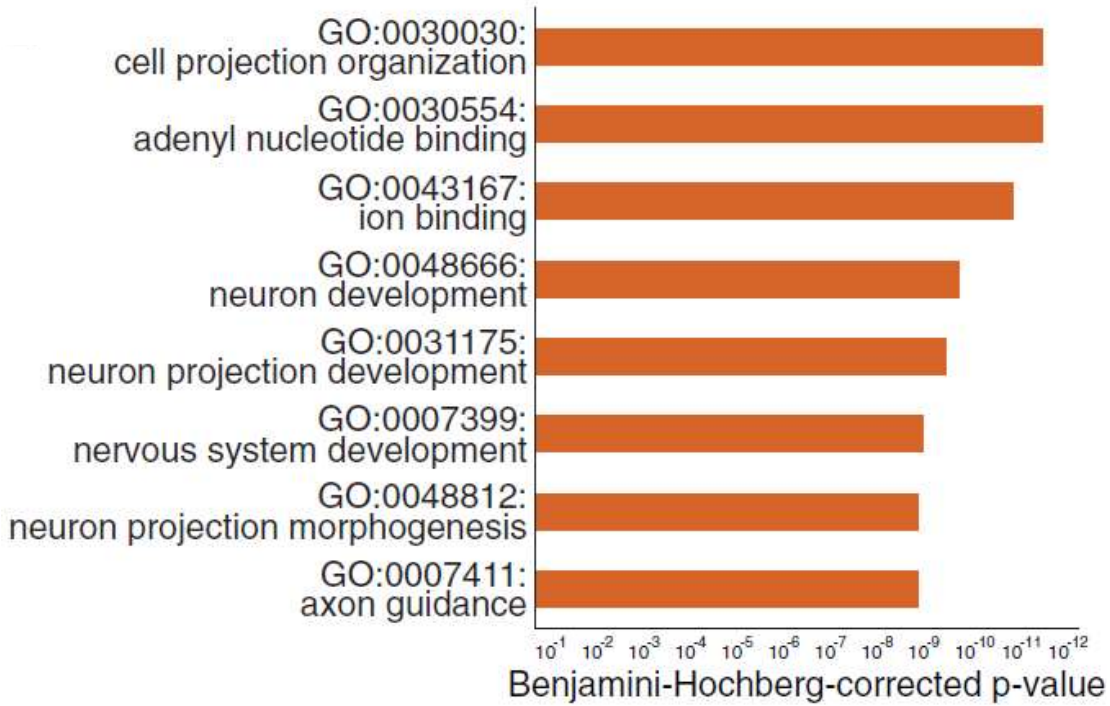
WuXiNextCODE

Lodato *et al.*, *Science* 2015
Lodato *et al.*, *Science* 2017

NEURODEVELOPMENT

Somatic mutation in single human neurons tracks developmental and transcriptional history

Michael A. Lodato,[1*] Mollie B. Woodworth,[1*] Semin Lee,[2*] Gilad D. Evrony,[1] Bhaven K. Mehta,[1] Amir Karger,[3] Soohyun Lee,[2] Thomas W. Chittenden,[3,4]† Alissa M. D'Gama,[1] Xuyu Cai,[1]‡ Lovelace J. Luquette,[2] Eunjung Lee,[2,5] Peter J. Park,[2,5]§ Christopher A. Walsh[1]§



WuXiNextCODE

LETTER

doi:10.1038/nature22322

FGF–dependent metabolic control of vascular development

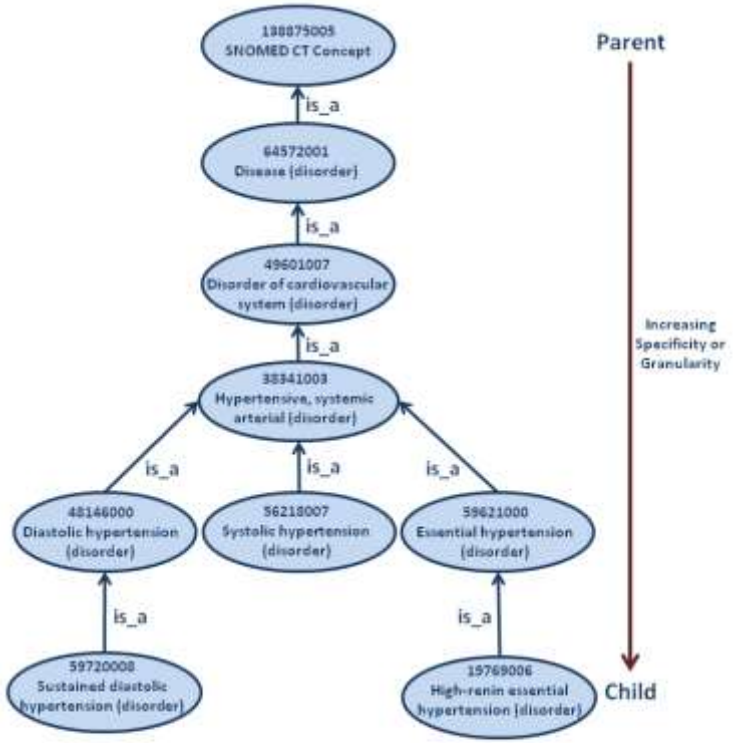
Pengchun Yu[1], Kerstin Wilhelm[2]*, Alexandre Dubrac[1]*, Joe K. Tung[1]*, Tiago C. Alves[3], Jennifer S. Fang[1], Yi Xie[1], Jie Zhu[4], Zehua Chen[5], Frederik De Smet[6,7], Jiasheng Zhang[1], Suk-Won Jin[1,8], Lele Sun[9], Hongye Sun[9], Richard G. Kibbey[3], Karen K. Hirschi[1], Nissim Hay[10], Peter Carmeliet[11,12], Thomas W. Chittenden[5], Anne Eichmann[1,13], Michael Potente[2] & Michael Simons[1,14]

GO Class	Accession Number	nGOSeq Term	List Hits	List Size	Pop Hits	Pop Size	Fisher's Exact	Gene Enrich	%Gene Enrich	Pvalue LogDiff	nGOseq Gene Enrich	GOseq Accession	GOSeq Term
BP	1900744	regulation of p38MAPK cascade	2	63	4	889	0.027	1.72	42.91	0.65	0.33	0007155	cell adhesion
BP	0060055	angiogenesis involved in wound healing	3	20	4	199	0.003	2.60	64.95	0.88	0.25	0001666	response to hypoxia
BP	0001935	endothelial cell proliferation	22	1127	72	7178	0.001	10.66	14.86	0.26	0.25	0044237	cellular metabolic process
BP	0043114	regulation of vascular permeability	2	85	4	868	0.050	1.61	40.21	0.68	0.54	0006629	lipid metabolic process
BP	0010573	vascular endothelial growth factor production	3	41	6	488	0.013	2.50	41.60	0.70	0.43	0033993	response to lipid
BP	0071604	transforming growth factor beta production	3	37	8	441	0.022	2.33	29.11	0.21	0.06	2000145	regulation of cell motility
BP	0006006	glucose metabolic process	19	576	73	3432	0.028	6.75	9.244	0.64	1.53	0044767	single-organism developmental process

Chittenden *et al.*, *Bioinformatics* 2012
Fang *et al.*, *Nature Communications* 2017
Yu *et al.*, *Nature* 2017

a priori Biomedical Knowledge-based Feature selection for deepCODE deep learning models

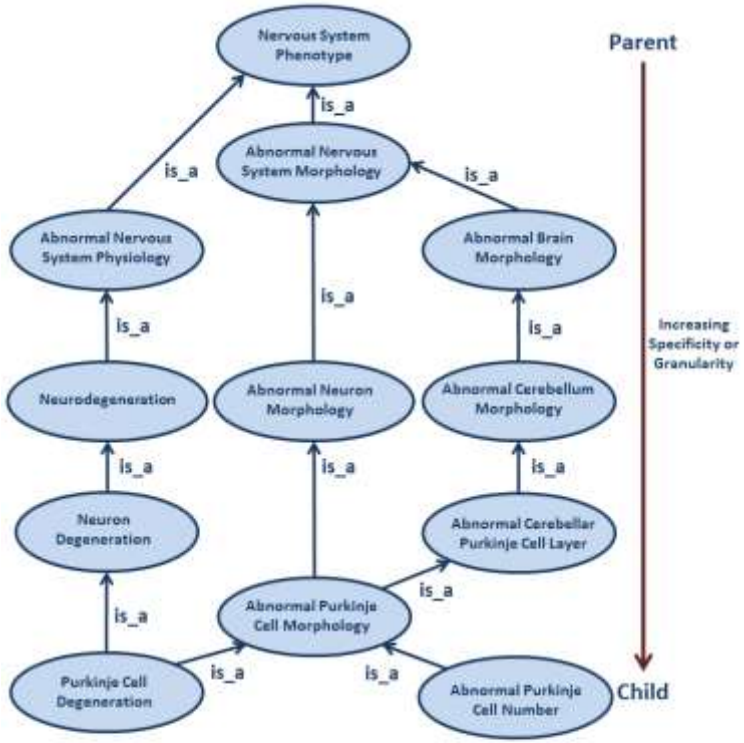
A. Clinical Ontology



SNOMED CT Clinical Ontology - Directed Acyclic Graph (DAG)

- Terms may have multiple parents on the tree.
- All attributes of a selected term must hold true for all its parents.
- Governed by "is_a" relationships.

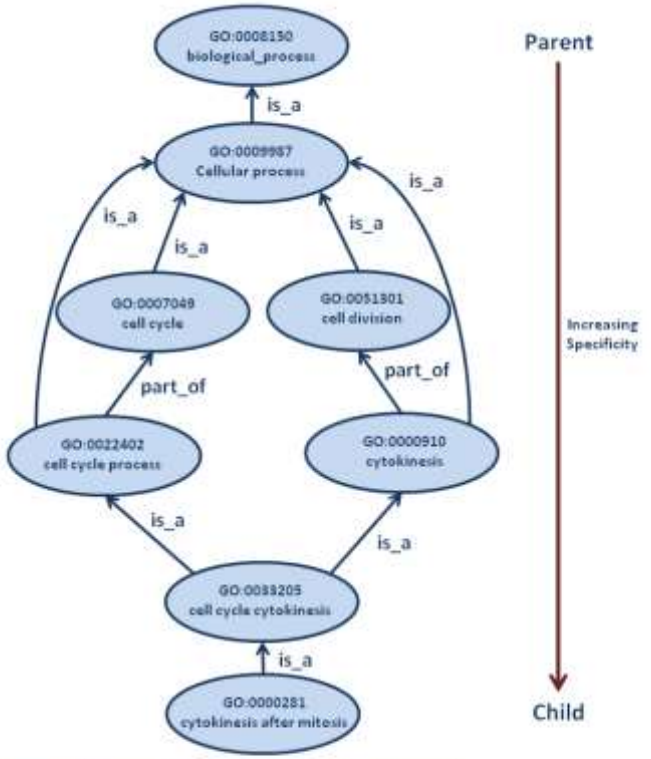
B. Phenotype Ontology



Human Phenotype Ontology - Directed Acyclic Graph (DAG)

- Terms may have multiple parents on the tree.
- All attributes of a selected term must hold true for all its parents.
- Governed by "is_a" relationships.

C. Genomic Ontology



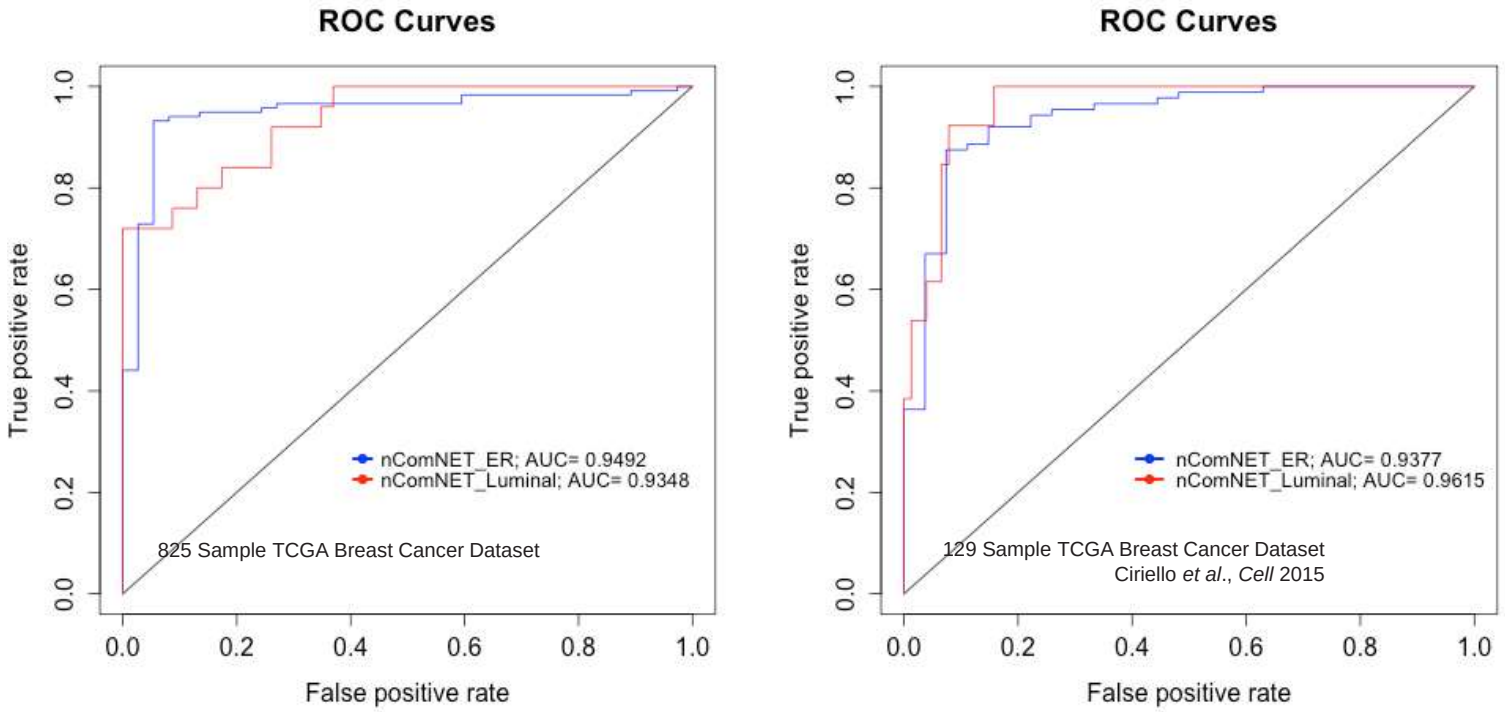
Gene Ontology - Directed Acyclic Graph (DAG)

- Terms may have multiple parents on the tree.
- All attributes of a selected term must hold true for all its parents.
- Governed by "is_a" and "part-of" relationships.

Modeling Human Breast Cancer – High Generalizability

Molecular Subtypes using Somatic Tumor Variants (STVs) and mRNA

Novel deepCODE pathway-based integration approach classifies tumor subtypes and tumor vs. normal at high accuracy
This classification reveals key mutated and expressed genes/pathways.

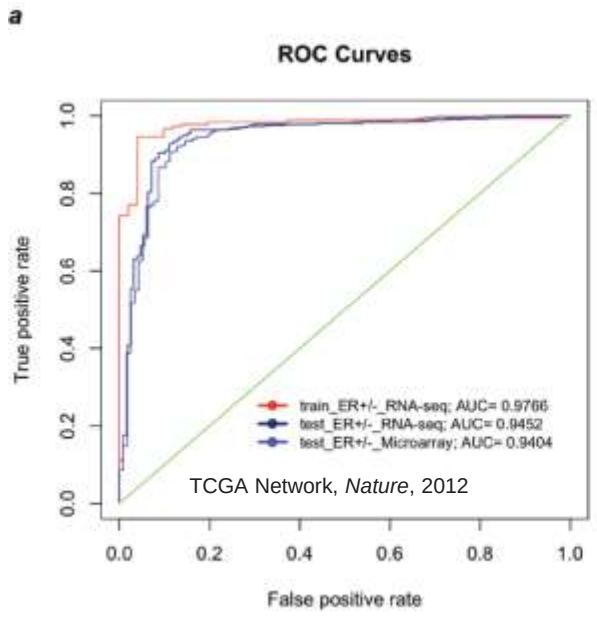


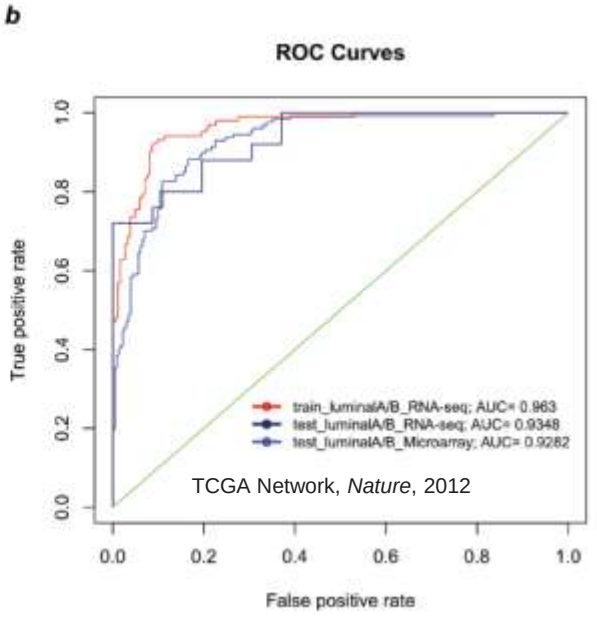
ER- vs. ER+ Breast Tumor Classification with 0.95 accuracy
2 Mutated Pathways (10 genes); 5 Aberrant Expression Pathways (146 genes)

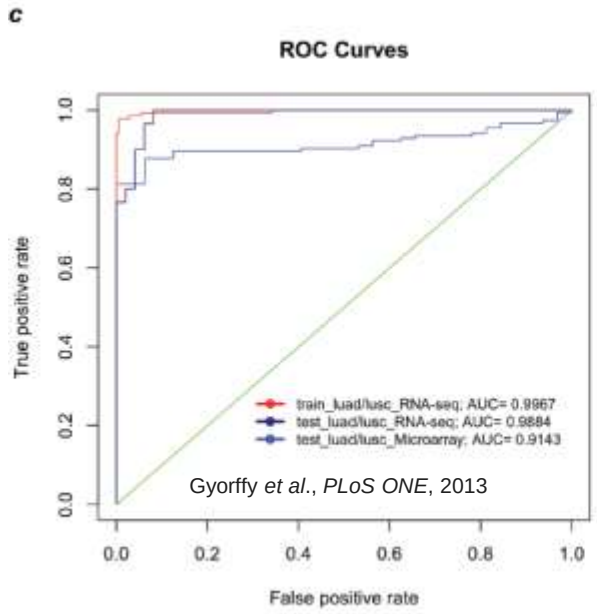
Luminal A vs. B Breast Tumor Classification with 0.94 accuracy
4 Mutated Pathways (172 genes); 8 Aberrant Expression Pathways (72 genes)

WuXiNextCODE

Cross-Platform Analysis: RNA-seq to DNA Microarray – High Generalizability







a. ER- vs. ER+ Classification with AUC = 0.95
5 Aberrant Expression Pathways (146 genes)

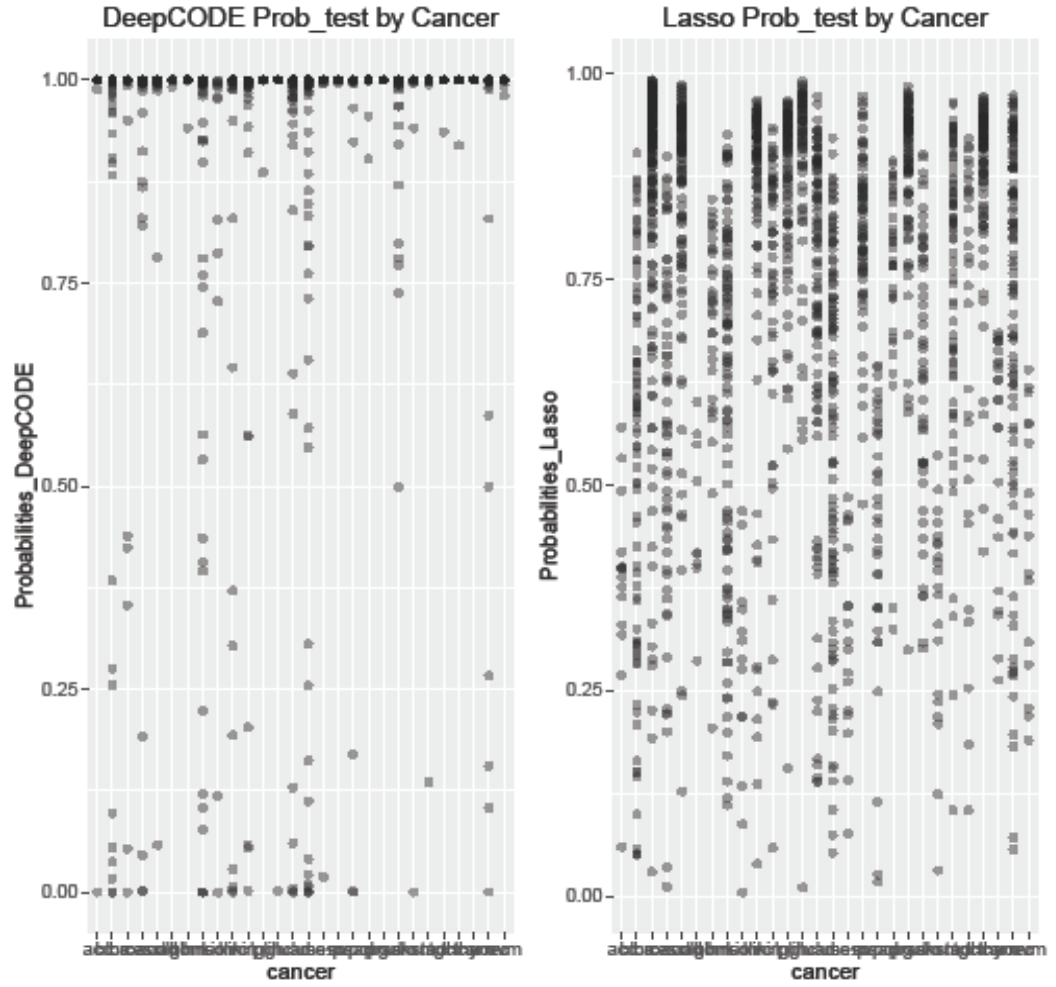
b. Luminal A vs. B Classification with AUC = 0.94
8 Aberrant Expression Pathways (72 genes)

c. LUAD vs. LUSC Classification with AUC = 0.98
9 Aberrant Expression Pathways (60 genes)

WuXiNextCODE

Our deep learning approach to classification of TCGA tumor Types is far superior to traditional machine learning methods (LASSO)

DeepCODE Deep Learning vs. LASSO Machine Learning Multinomial Regression Models on 28 TCGA cancer types



True Positive Probability Distributions per Cancer type

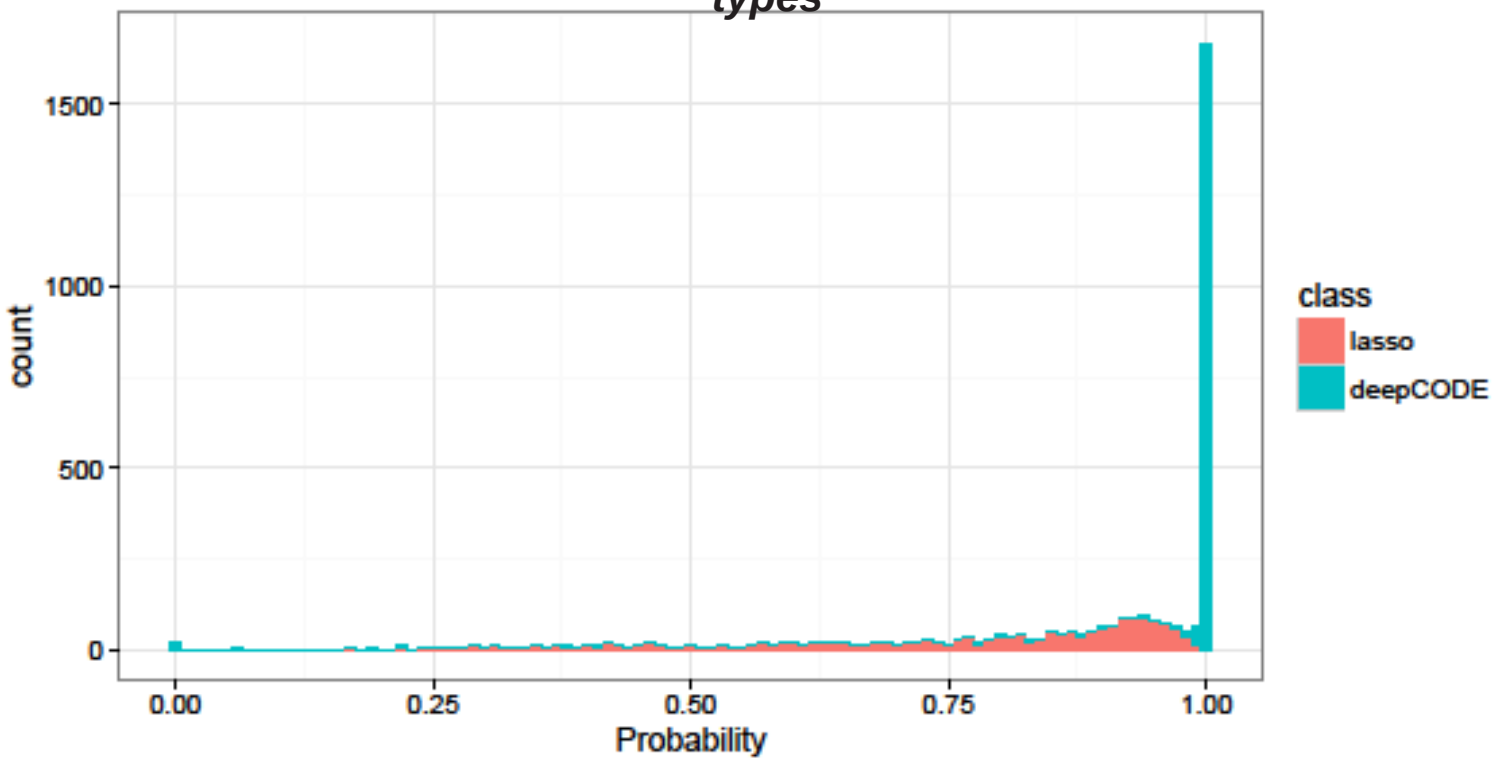
Note: deepCODE Model Calls True Positives with far greater confidence

Multinomial Human Cancer Classification:
Trained: 7,618 RNA-seq samples; Tested:1,889 RNA-seq samples

WuXiNextCODE

Our deep learning approach to classification of TCGA tumor Types is far superior to traditional machine learning methods (LASSO)

DeepCODE Deep Learning vs. LASSO Machine Learning Multinomial Regression Models on 28 TCGA cancer types



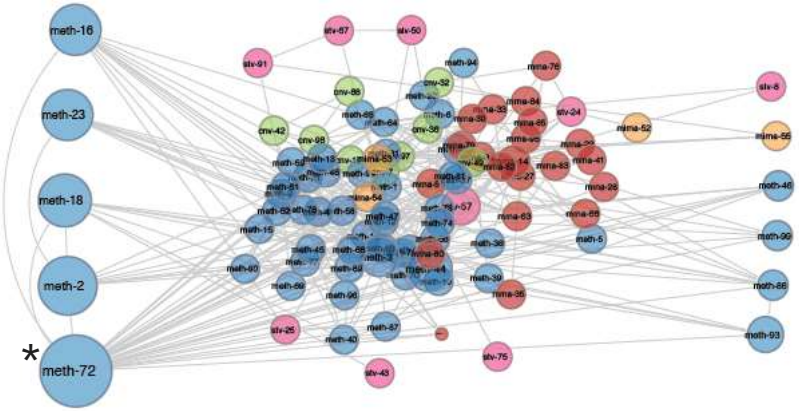
Total True Positive Probability Counts for Test Data Across All 28 Cancer Types

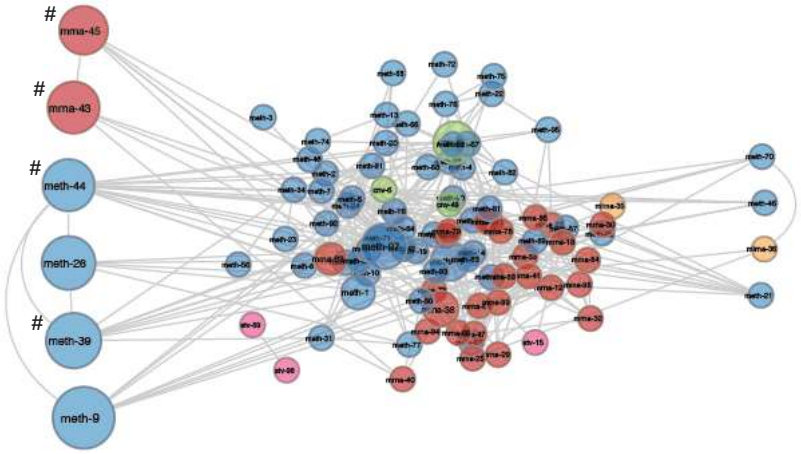
Note: deepCODE Model Calls True Positives with far greater confidence

Multinomial Human Cancer Classification:

Trained: 7,618 RNA-seq samples; Tested:1,889 RNA-seq samples

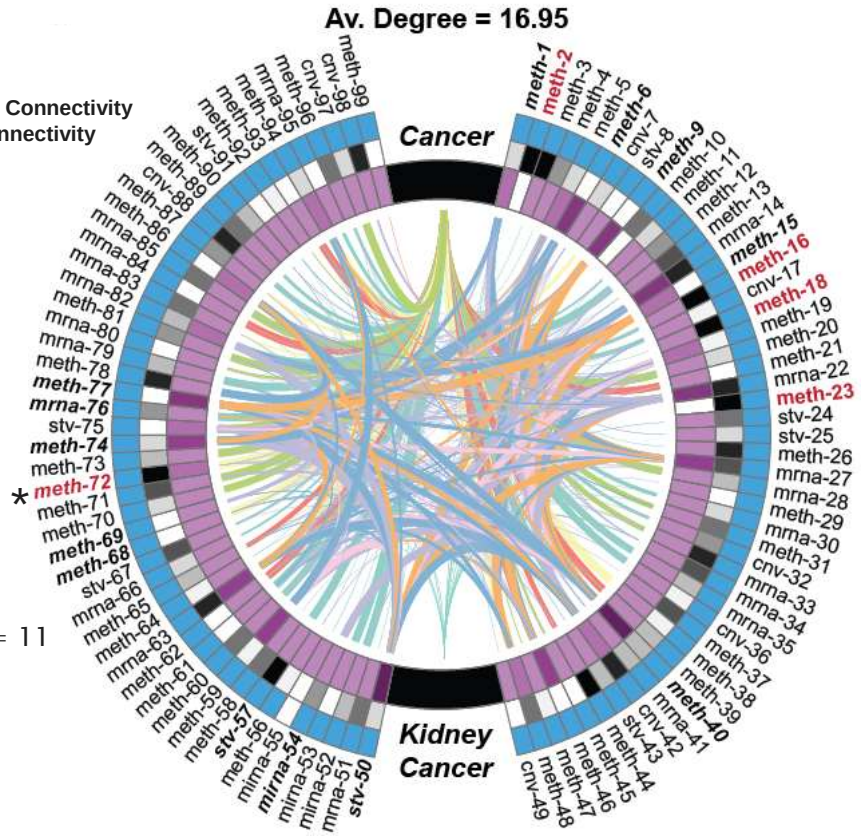
Blue = DNA Methylation
Red = mRNA
Orange = miRNA
Green = CNV
Pink = STV

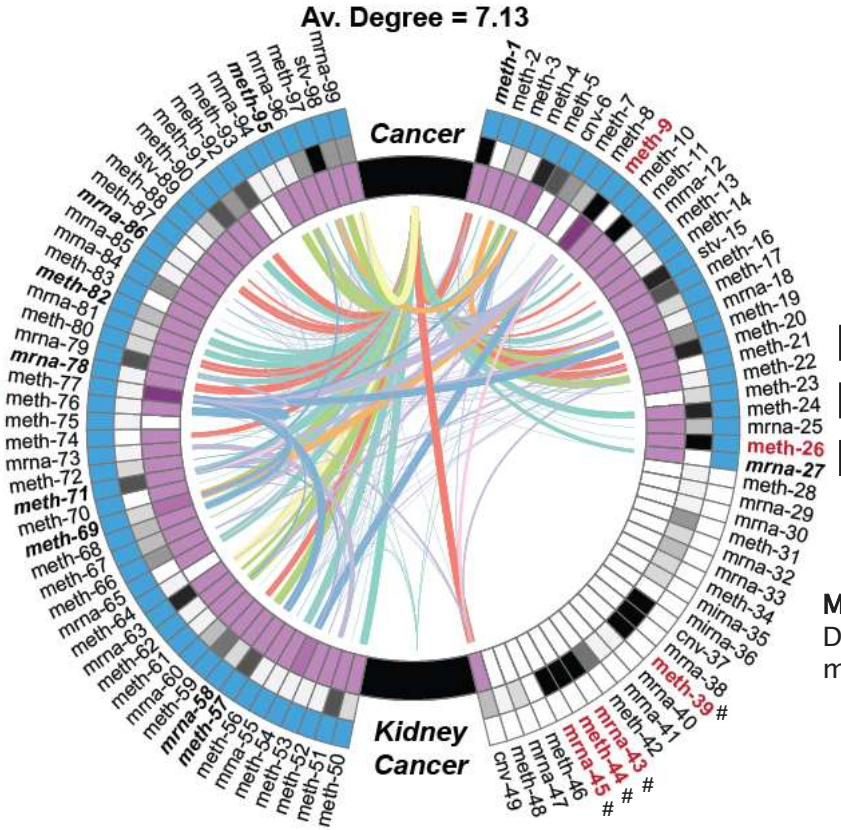




Bayesian Belief Networks

Purple Band = Degree NLP Network Connectivity
Black Band = Degree BNN Node Connectivity
Blue Band = Function Annotation
Red = BNN Driver Gene
Bold + Italic = Known Drug Target
*Driver Gene & Known Drug Target
#Driver Gene of Unknown Function





Natural Language Processing

nGOseq
DNA Methylation = 11
mRNA = 1
miRNA = 1
STV = 1

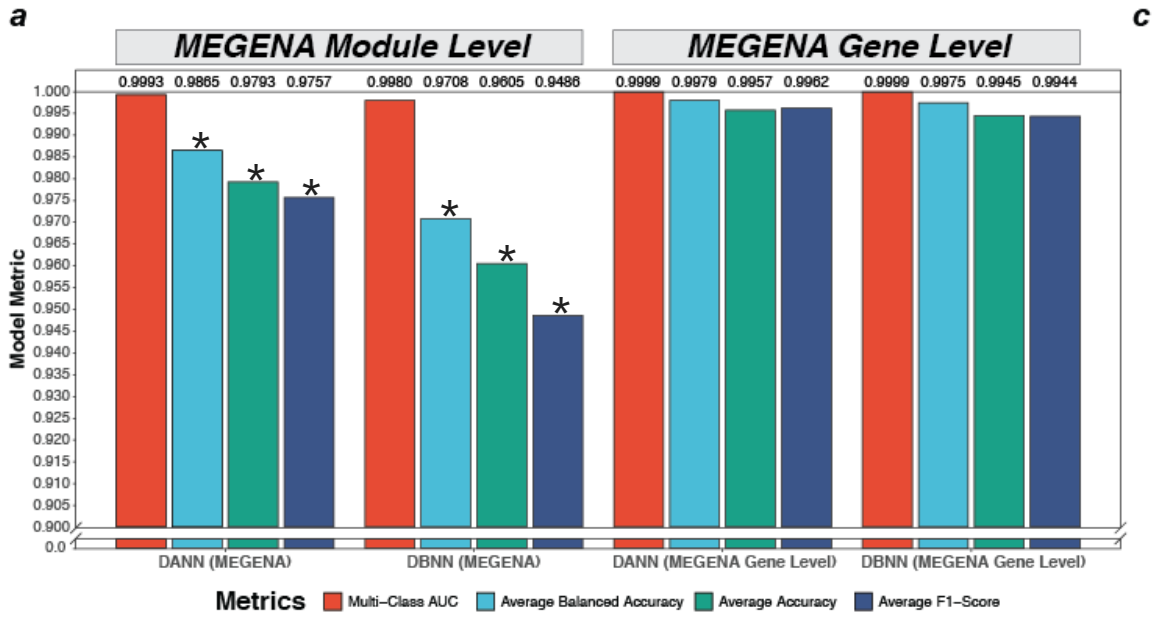
MEGENA
DNA Methylation = 6
mRNA = 4

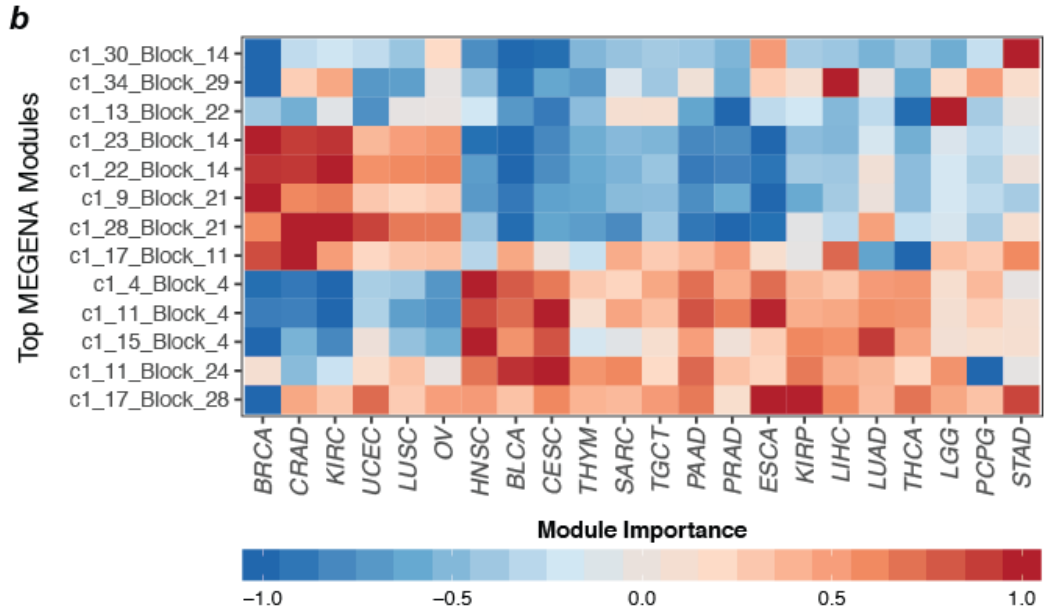
nGOseq

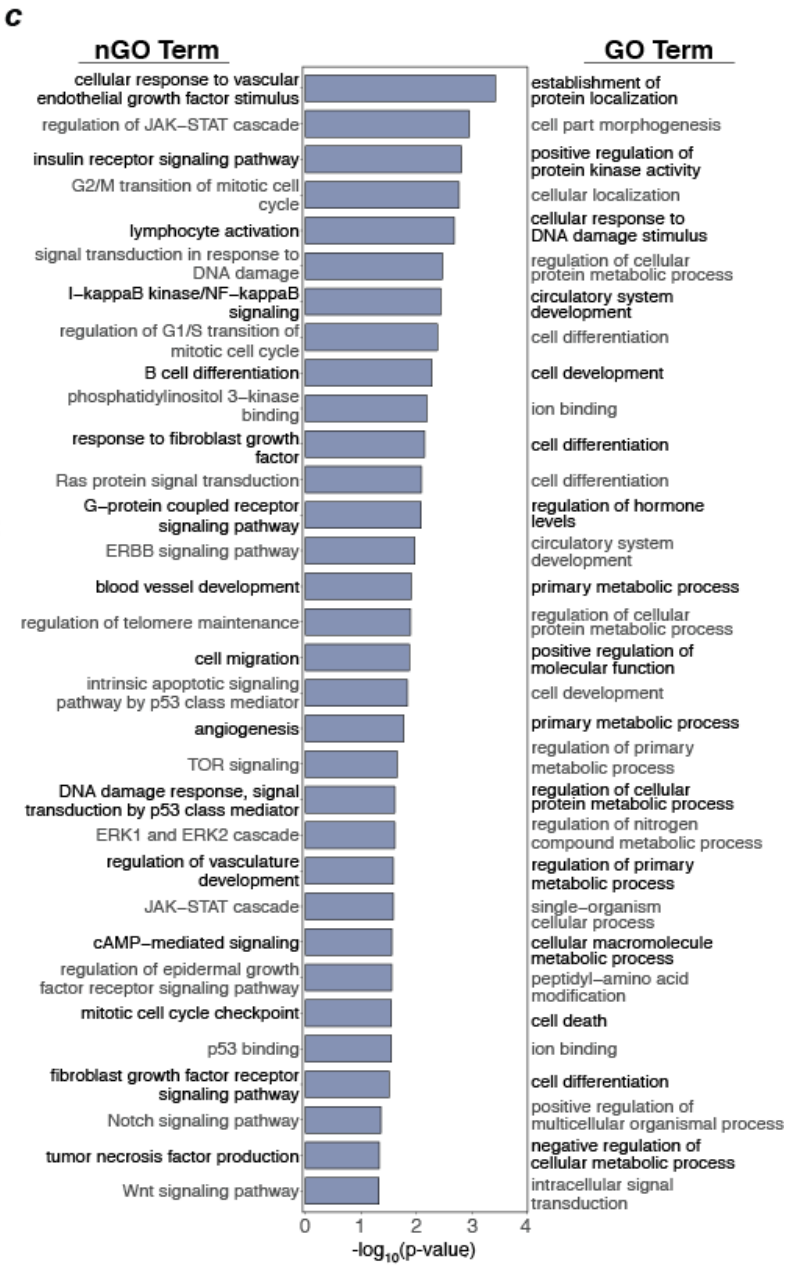
MEGENA

WuXiNextCODE

Multinomial Classification of 22 TCGA Cancer Types with Greater than 99.6% Accuracy

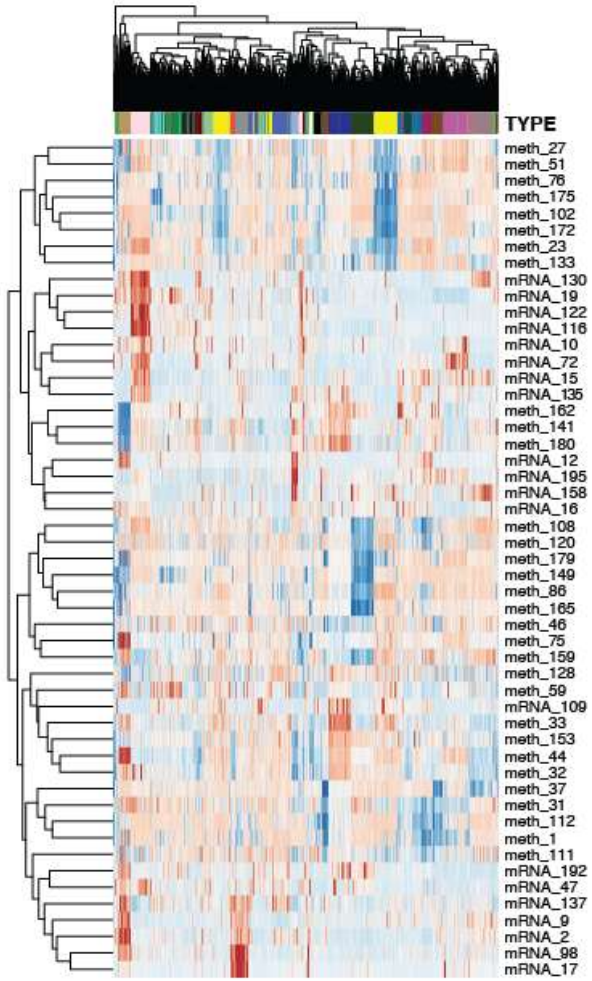


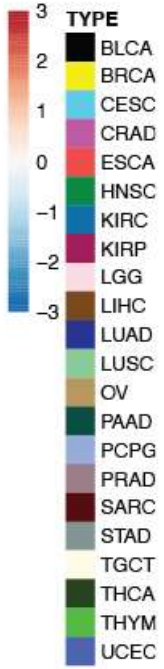


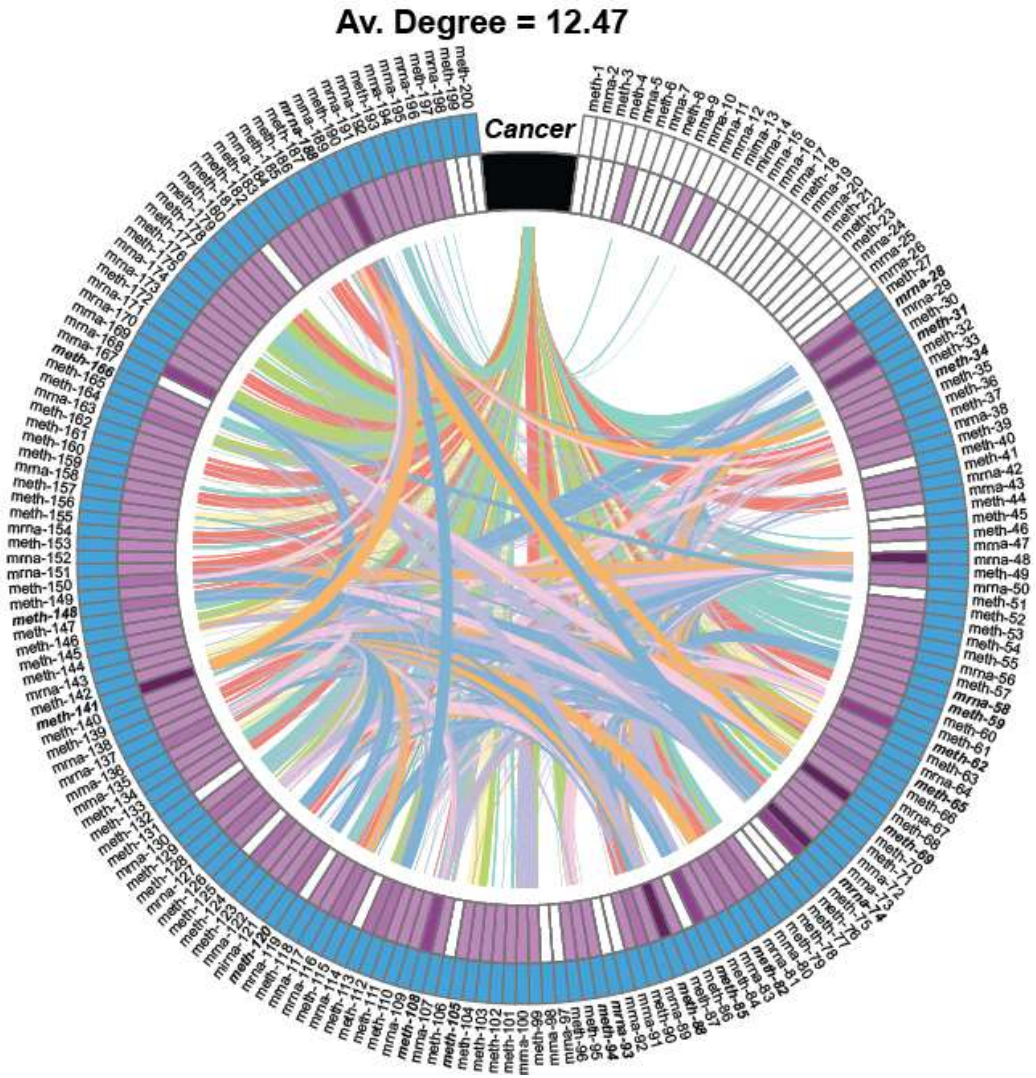


WuXiNextCODE

Multinomial Classification of 22 TGCA Cancer Types with Greater than 99.6% Accuracy





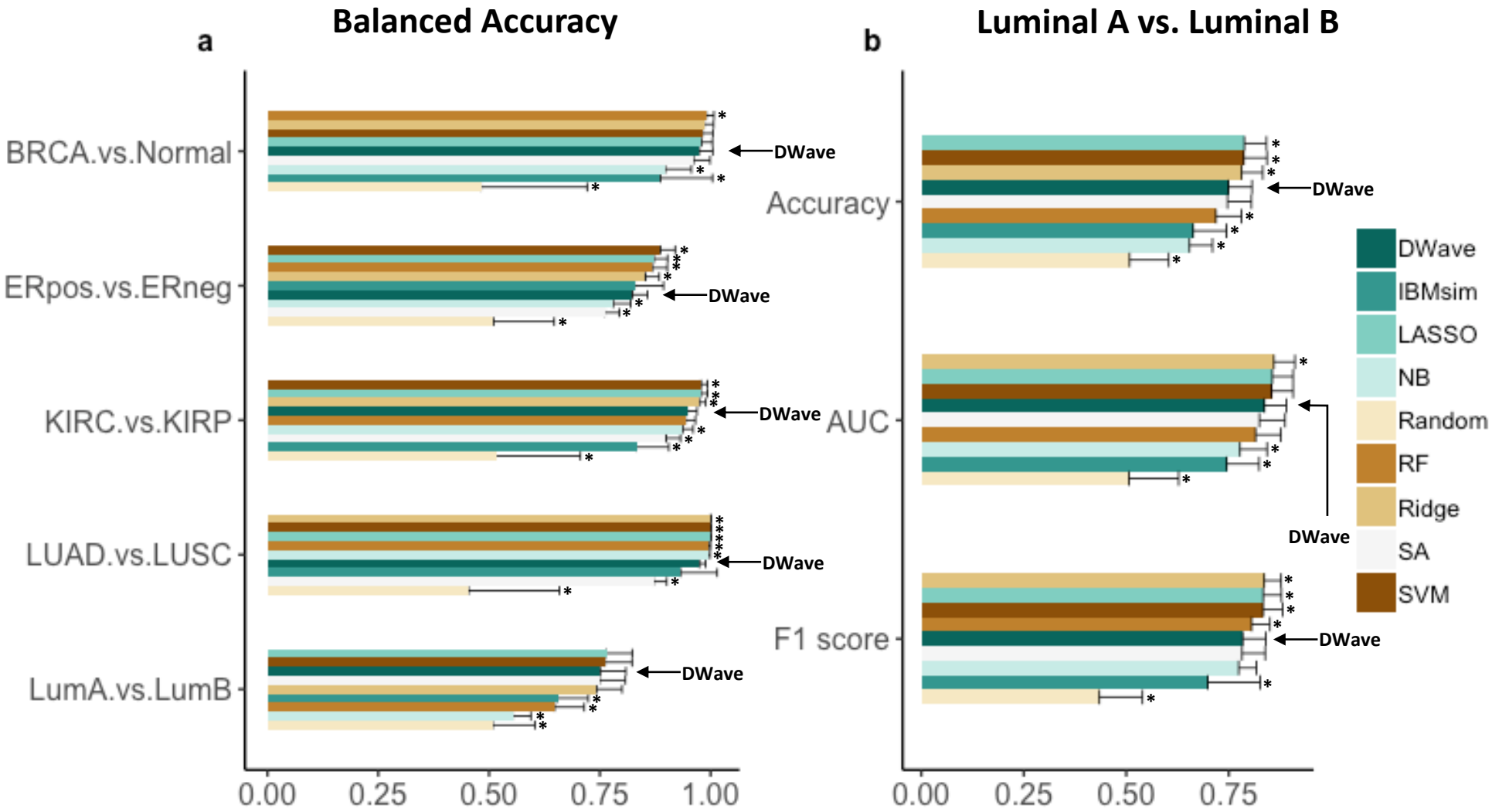


Natural Language Processing

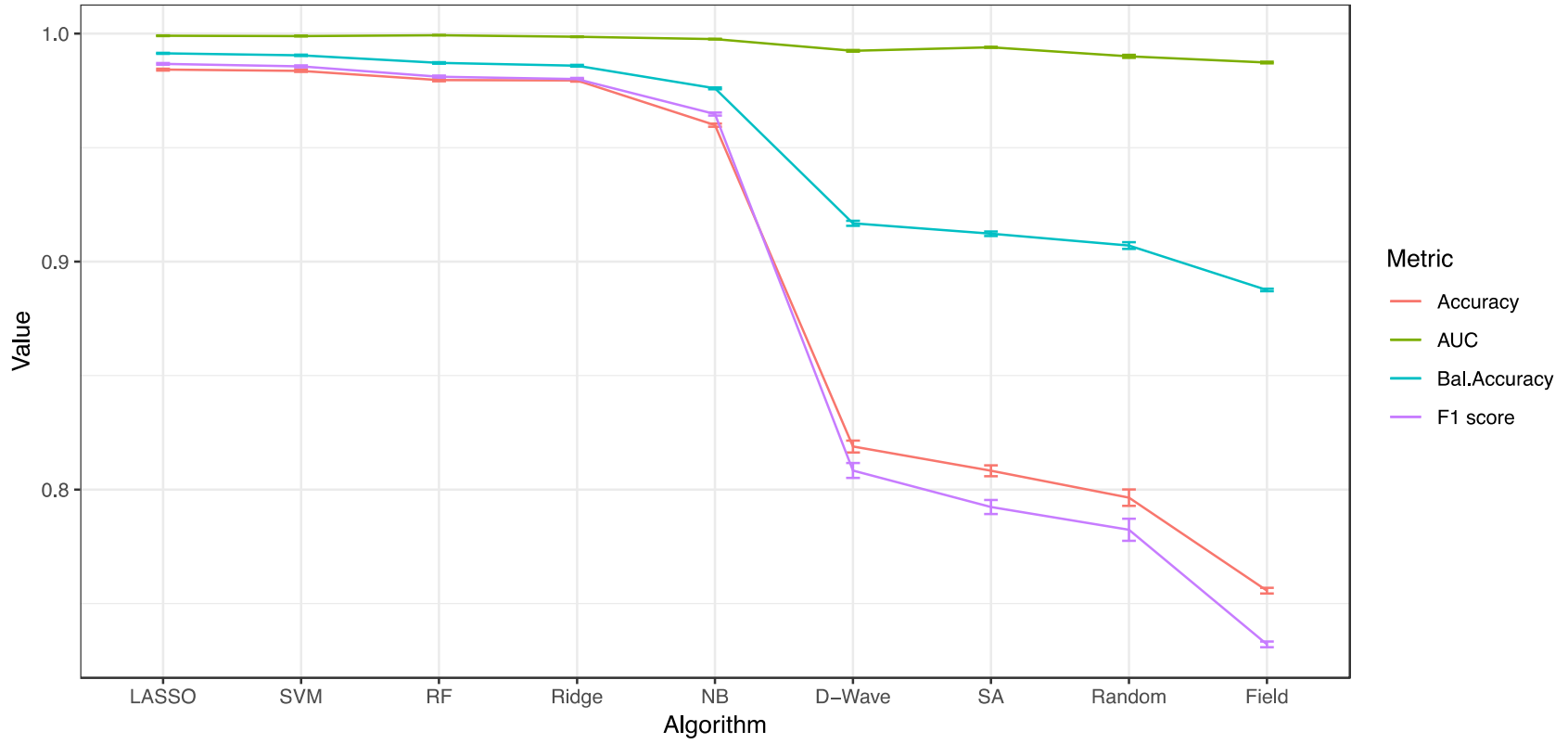
Purple Band = Degree NLP Network Connectivity
Blue Band = Function Annotation
Bold + Italic = Known Drug Target (12 DNA Methylation; 4 mRNA)

WuXiNextCODE

Binomial Classification of Tumor Molecular Subtypes Quantum Machine Learning



Multinomial Classification of Human Cancer Types Quantum Machine Learning



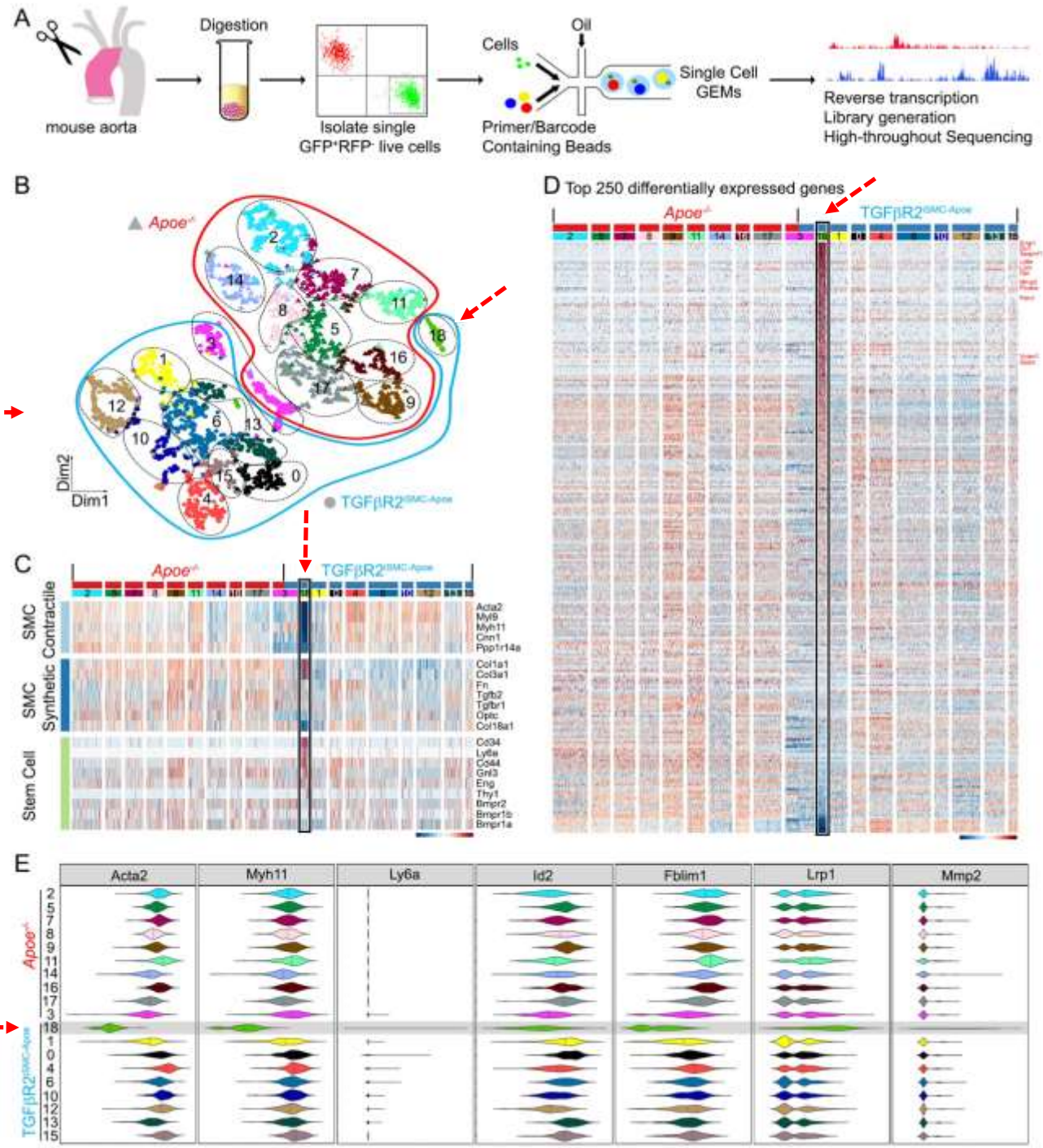
Human Cancer Types	Sample Number
Liver Hepatocellular Carcinoma	358
Breast cancer	1006
Brain Lower Grade Glioma	499
Colon Adenocarcinoma/ Rectum Adenocarcinoma	551
Kidney Cancer	611
Lung Cancer	962
Total	3987
Train	3190
Test	797

Identifying Causal Drivers of Cardiovascular Disease: Aortic Aneurysm

CyTOF (Cytometry by Time of Flight)

Vs.

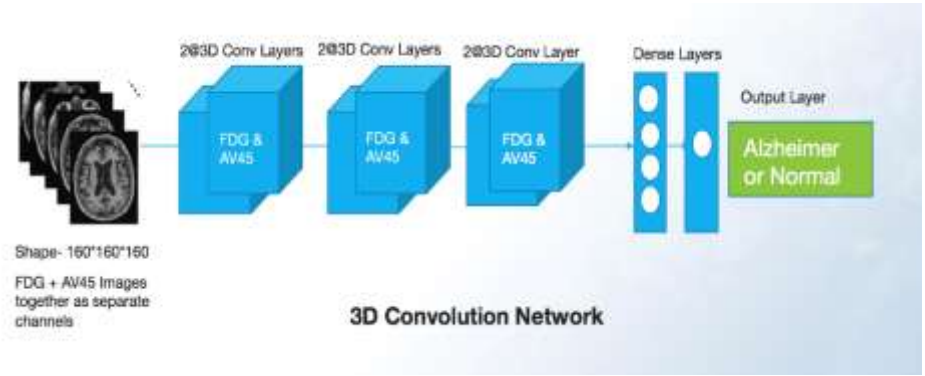
Zero Inflated Variational Autoencoder (VAE) of Single Cell RNA-seq Data



Deep Learning, Machine Learning and Alzheimer Disease (ADNI)

Image Data

PET Scans (AV45 + FDG)



	Count
Samples	326
Alzheimer	144
Controls	182

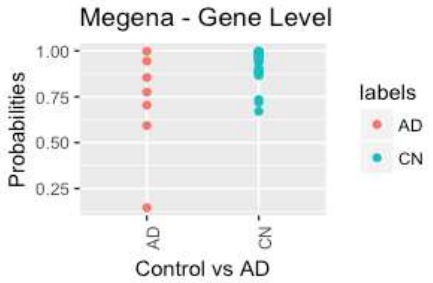
DCNN (Test set)	
AUC	0.99
Accuracy	0.94

Molecular Signature

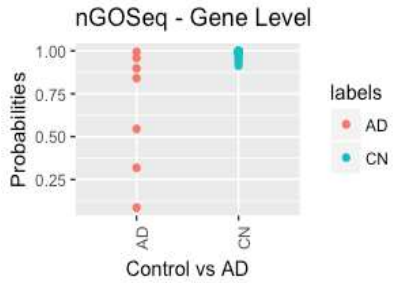
Integrated Datatypes: Methylation , Expression, Variant Data

	Count
Samples	152
Alzheimer	36 (29/7)
Controls	116 (93/23)

	LASSO (Test set)			
	MEGENA	nGOSeq	MEGENA	nGOSeq
	MetaGene	MetaGene	Gene	Gene
AUC	0.87	0.94	0.98	1
Accuracy	0.83	0.93	0.97	0.93



Non-zero Genes
Methylation: 31
Expression: 17
STV: 7



Non-zero Genes
Methylation: 29
Expression: 5
STV: 10

WuXiNextCODE

Deep Learning, Machine Learning and Alzheimer Disease (ADNI)

Molecular Signature

Single Datatypes: Methylation , Expression, Variant Data

	Count
Samples	152
Alzheimer	36 (29/7)
Controls	116 (93/23)

	Methylation Genes	Expression Genes	Variant Genes
AUC	0.93	0.77	0.80
Accuracy	0.90	0.76	0.77
AD Test Acc	0.57	0.54	0.54

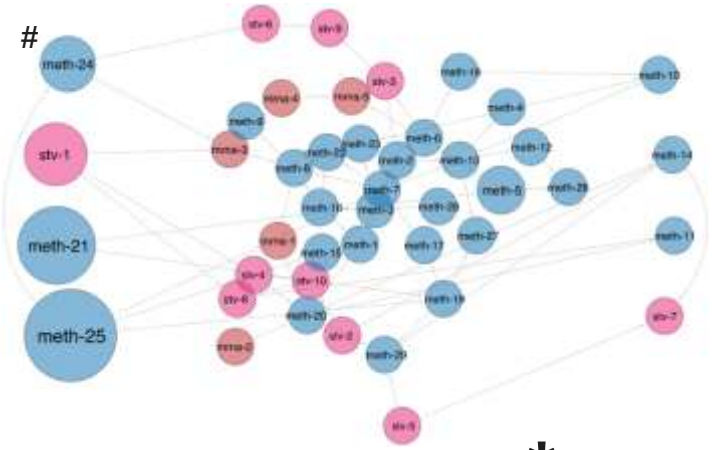
No Feature Selection

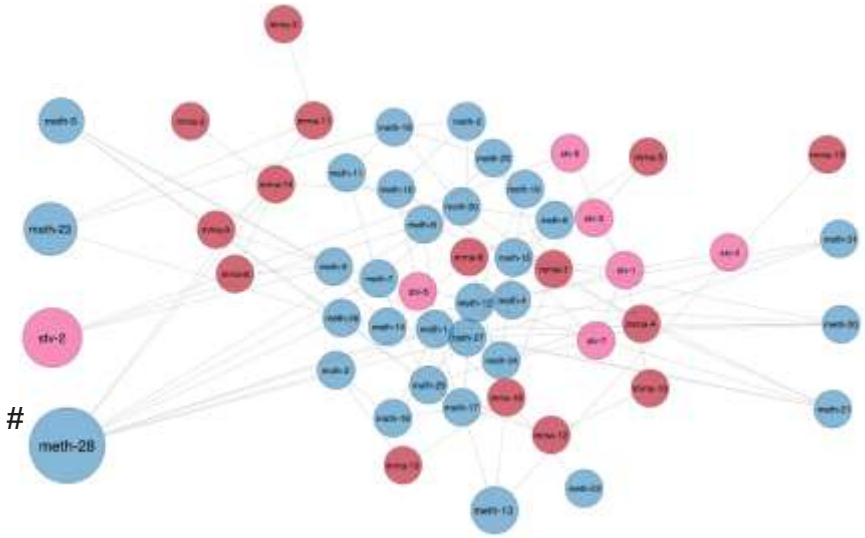
	MEGENA Feature Selection		
	Methylation Genes	Expression Genes	Variant Genes
AUC	0.99	0.75	0.75
Accuracy	0.93	0.73	0.73
AD Test Acc	0.86	0.64	0.64

	nGOSeq Feature Selection		
	Methylation Genes	Expression Genes	Variant Genes
AUC	0.98	0.81	0.75
Accuracy	0.93	0.73	0.73
AD Test Acc	0.71	0.64	0.58

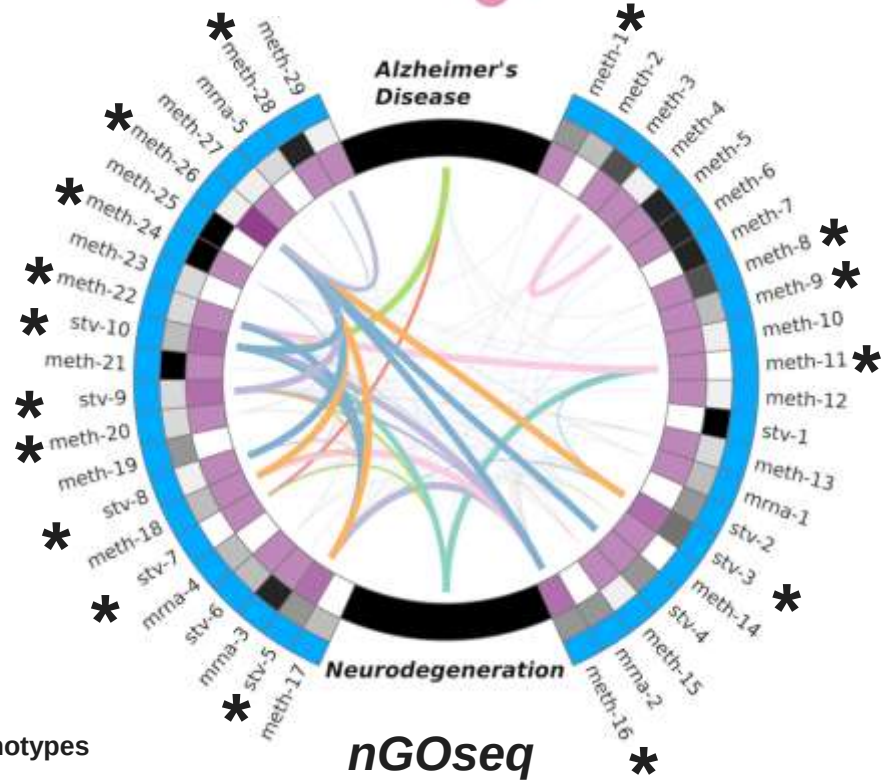
Blue = DNA Methylation
Red = mRNA
Pink = STV

#Same Gene



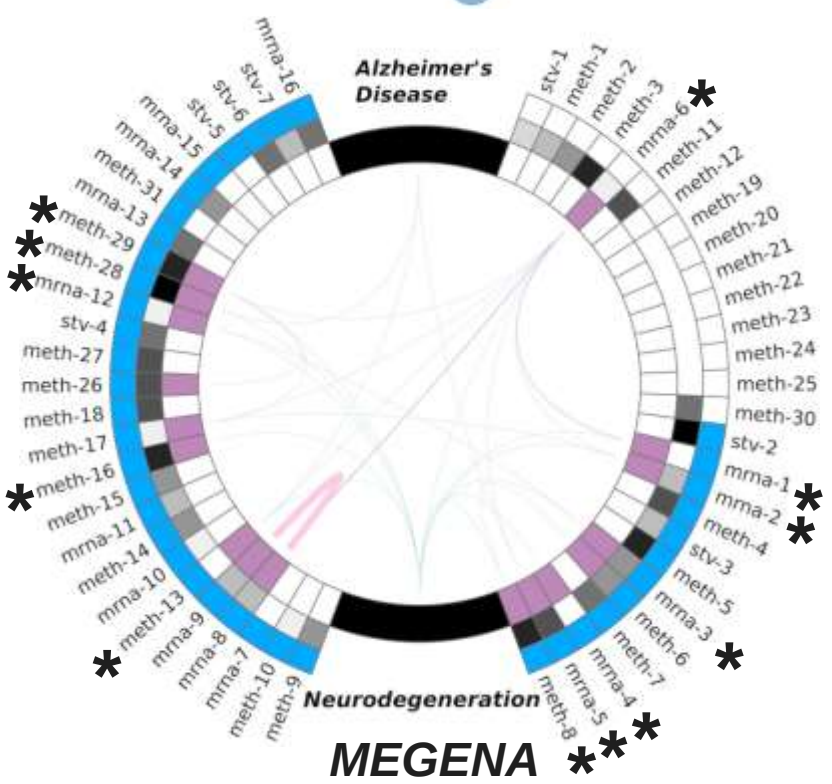


Bayesian Belief Networks



nGOseq

Av. Degree = 12.00



MEGENA

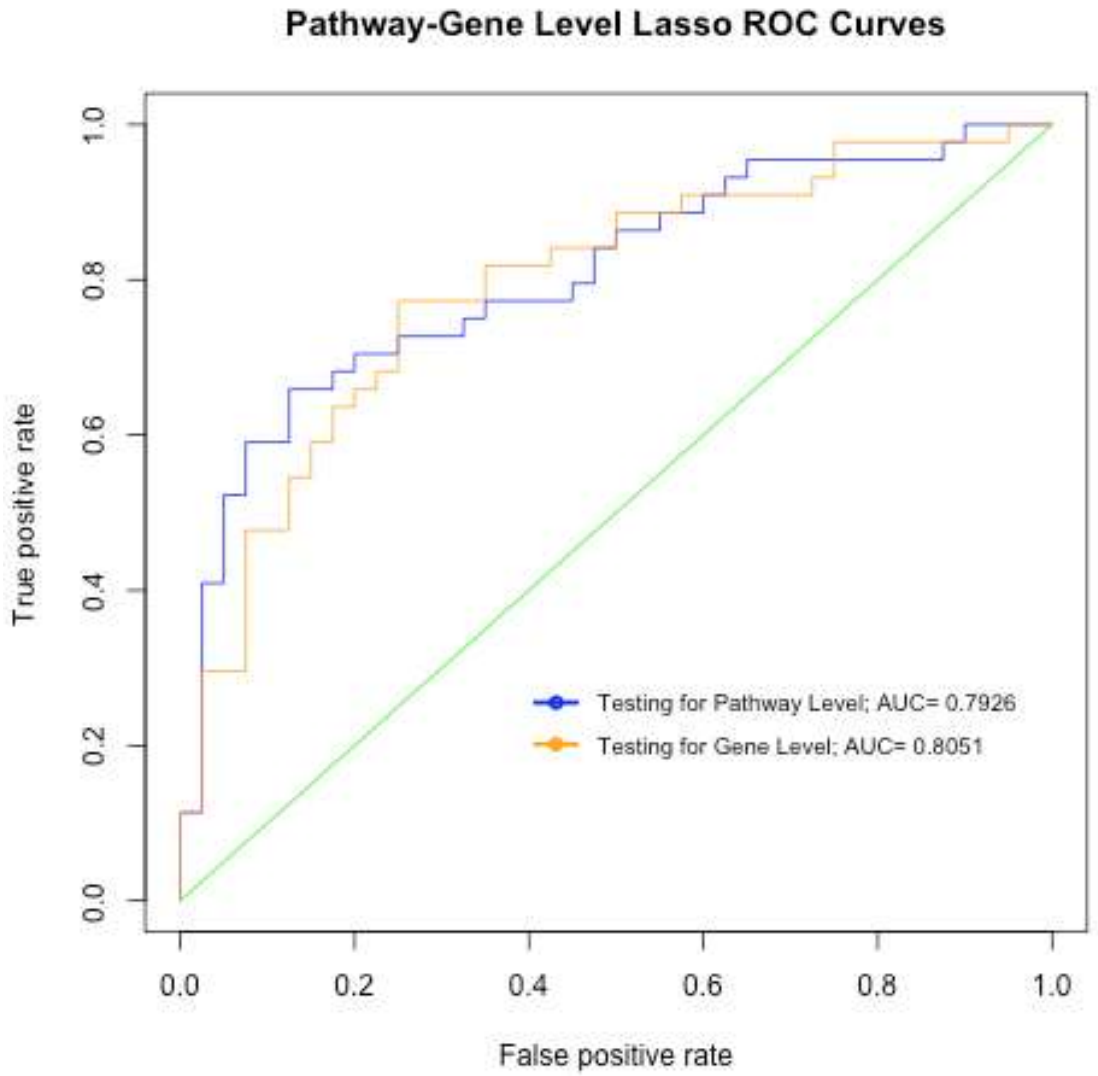
Av. Degree = 4.58

Natural Language Processing

*Implicated in Phenotypes

WuXiNextCODE

Deep Learning, Machine Learning and Alzheimer Disease (RosMap)



RNA extracted from dorsolateral prefrontal cortex of 724 subjects

Sample set:
AD: 222 [Train: 178, Test: 44]
CN: 201 [Train: 161, Test: 40]

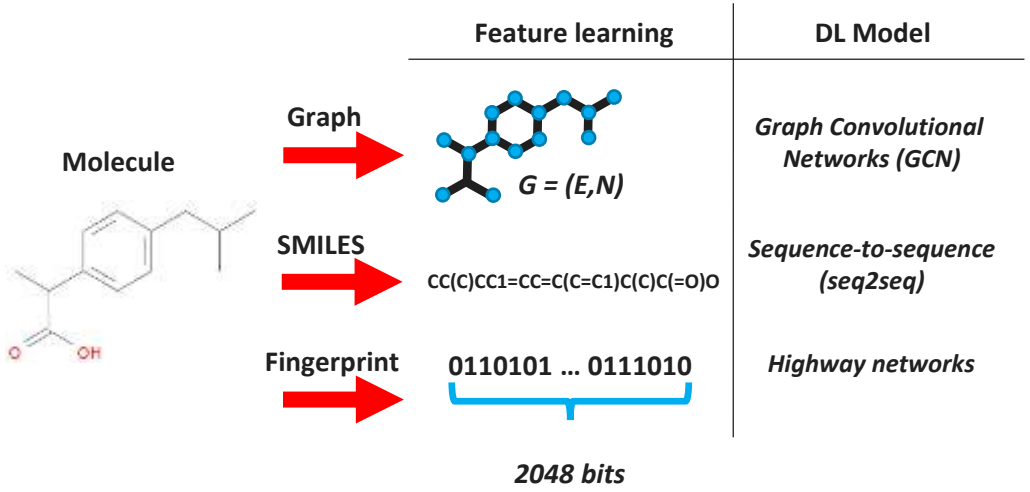
Pathway Level Analysis:
Number of Pathways: 3340
Test Accuracy: 72.61
Test AUC: 79.26
Number of Non-Zero Pathways: 76

Gene Level Analysis:
Number of Genes: 342 Genes from 76 Non-zero Pathways
Test Accuracy: 72.61
Test AUC: 80.51
Number of Non-Zero Genes: 45

Deep Learning for Chemical Reactions

Modeling Chemical Data

DL models based on different representations of molecules:



Retrosynthesis

Learning how molecules are produced using chemical reaction datasets (~1.1 M chemical reactions from U.S. patents)

	Count
Product molecules	431485
Chemical reactions for classification	462

Multinomial classification with Highway networks (20% - Test set)	
Accuracy	0.79 (0.12)*

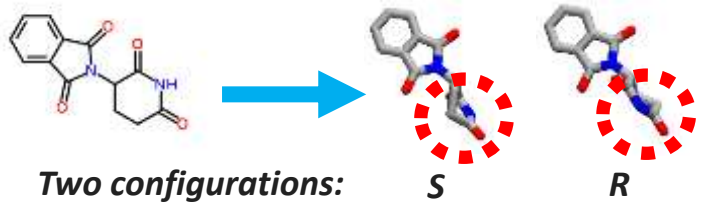
Multinomial classification with Multiscale approach (20% - Test set)	
Accuracy	0.90 (0.08)*

*s.d. in parentheses

Taking stereochemistry into account

Learning about molecular 3D shape for chemical reaction prediction

Atoms can be arranged differently for same molecule:

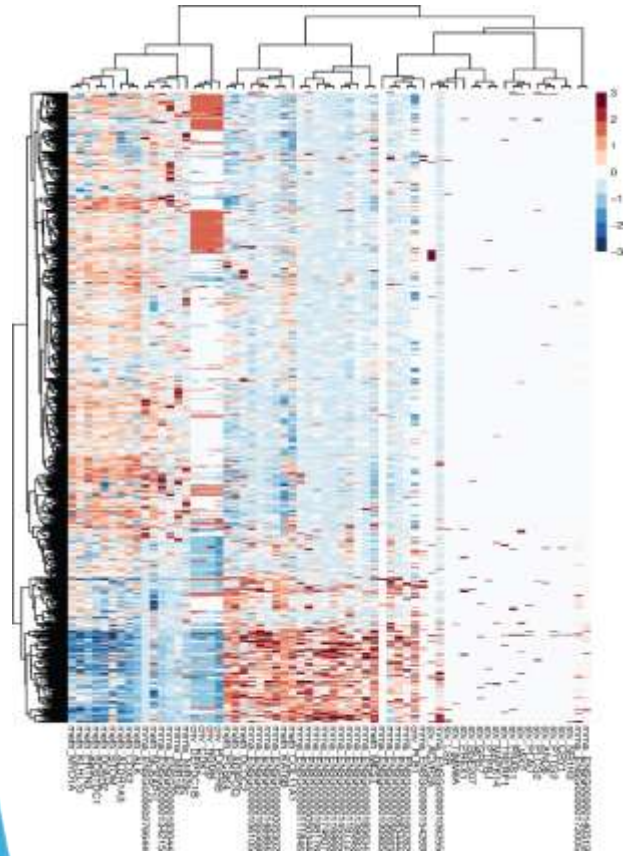


	Count
Molecules with single chiral center	2762

Binomial classification based on chirality (20% - Test set)	
Accuracy	0.89

Modeling Human Breast Cancers Quantum Machine Learning

Classical HCL



Estrogen Receptor Status	
Tumor Samples	959
ER Negative	740
ER Positive	219
Train	768
Test	191

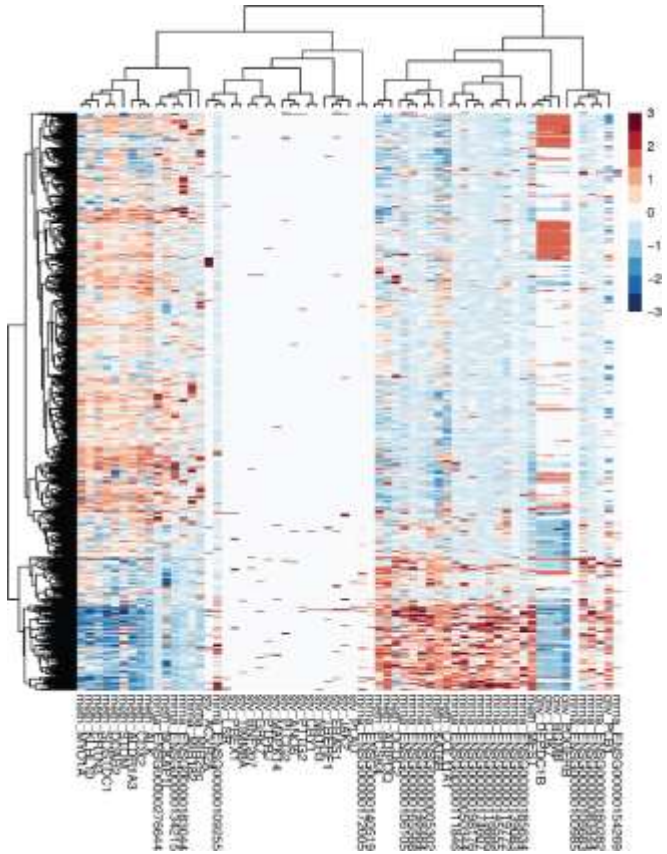
	Performance	
Algorithm	HCL	qHCL
Clustering (genes)	64	64
Clustering (sec)*	0.02	10078.30 (2h 48m)
Cluster Number	8	9
LASSO Classification Accuracy	0.9215	0.9267
LASSO ROC AUC	0.945	0.944
DANN Classification Accuracy	0.9267	0.9267
DANN ROC AUC	0.943	0.944

*Quantum and classical trees are 88% concordant based on the standard Robinson–Foulds metric

*qHCL - Durr-Hoyer method based on a modified Grover's search algorithm with Euclidean distance and Ward linkage

*qHCL ran on a IBM quantum simulator using 19 qubits

quantum HCL



WuXiNextCODE

WUXI NEXTCODE ANALYSIS PLATFORM

Clinical interpretation and research in **one, scalable platform built for the genome from the ground up**

GOR (Genomically Ordered Relational) Database Infrastructure

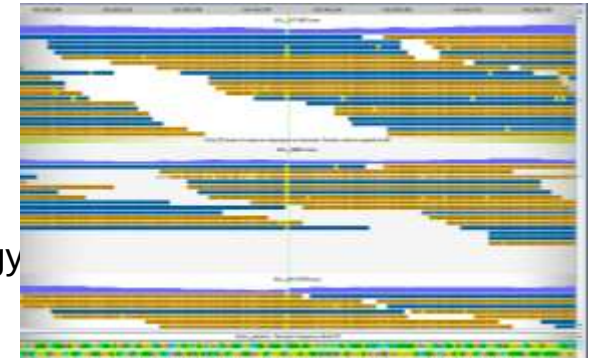
- For efficient storage and queries for whole genome and whole exome data using the tools listed below

Clinical Sequence Analyzer (CSA)

- Clinical geneticist-friendly tools for germline analysis of large or small families
- Automatic gene carrier analysis for confirmation
- Generate candidate genes from a standard list or with phenotype tools and stratify by variant annotations

Sequence Miner (SM)

- Advanced tool for case-control disease gene discovery or responder non-responder companion diagnostic discovery
- Additional algorithms for covariate adjustment and pathway enrichment
- Perform phenotype scans and carrier analysis



Tumor Mutation Analyzer (TMA)

- Somatic variant analysis for defining tumor-specific variations and oncology annotations including actionable databases