



THE INNOVATION CENTER FOR BIOMEDICAL INFORMATICS



Georgetown  
University

# Genomics-based clinical informatics resources to support Precision Oncology & Evidence Curation



Subha Madhavan, PhD, FACMI  
Chief Data Scientist  
Georgetown University Medical Center  
@subhamadhavan

***Genomic Medicine XIII: Developing a Clinical Genomic Informatics Research Agenda***  
***Feb 9-Feb 10, 2021***

# Conflicts Disclosure

- Scientific Advisory Board member, NHGRI Ethical, Legal, Social Implications (ELSI) Center
- Medical and Scientific Advisory Board, Colon Cancer Alliance
- Scientific Advisory Board, Perthera Inc
- Research Funding from NIH, FDA, VA, Private organizations and Philanthropy
- Chair of Planning committee for the Georgetown annual Health Informatics & Data Science symposium since 2012 with industry sponsorship

# Outline

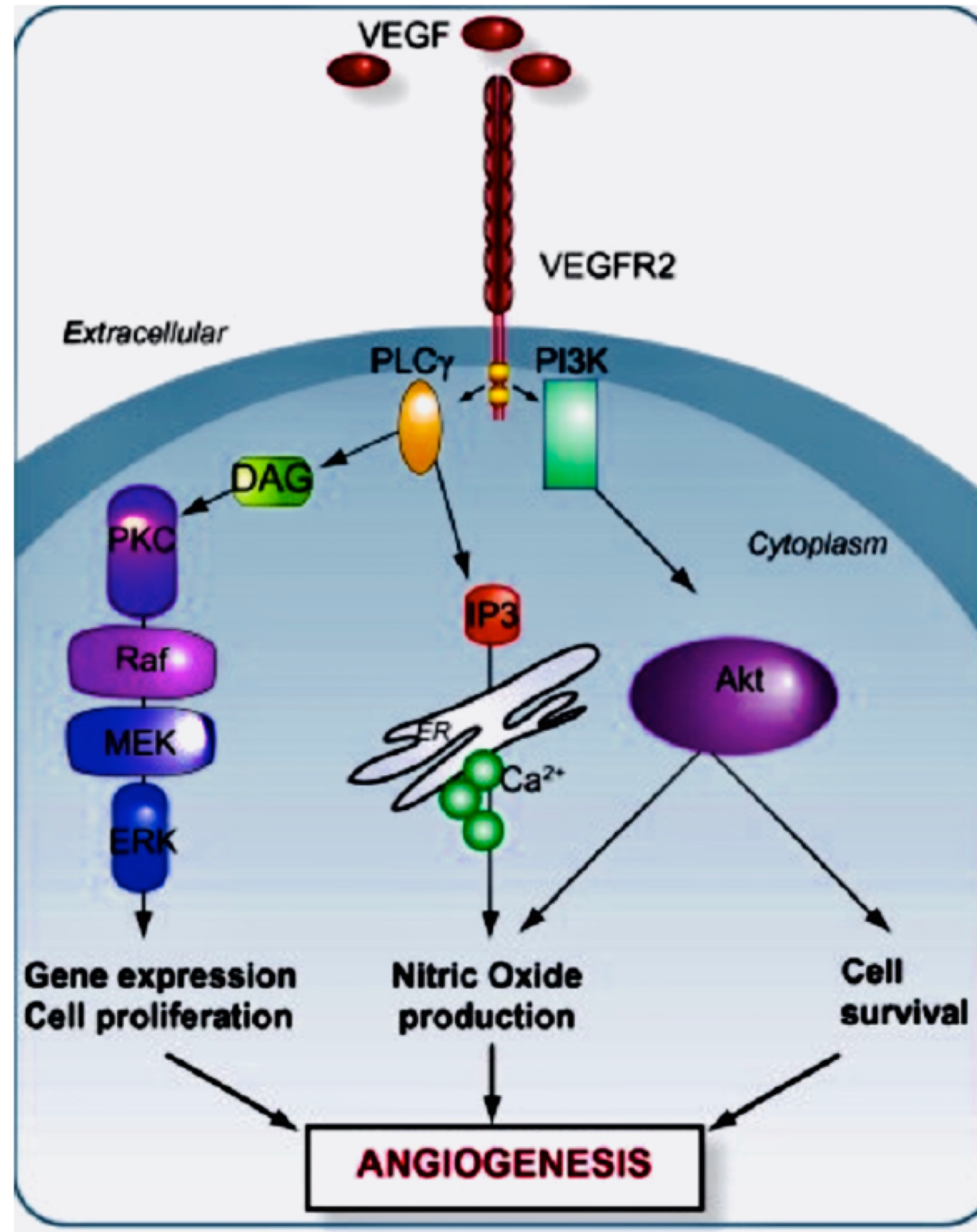
- Story 1: Collaborative, Multidisciplinary Evaluation of Cancer Variants Through Virtual Molecular Tumor Boards
  - Facilitators & Barriers for deploying genomics-based clinical informatics resources
- Story 2: Automatic extraction of Clinical genomic evidence for variant interpretation using Natural Language Processing and Machine Learning
  - Identify research needed to improve how genomic-based clinical informatics resources impact the patient and clinician decision-making processes

# Case Study

- A 52 year old Caucasian male with locally advanced Stage IV pancreatic adenocarcinoma with possible lung metastases
- Progressed on Gemcitabine and nab-paclitaxel with a later addition of Erlotinib
- Multi-panel genomic testing lab identified 4 difficult to target mutations - KRAS, p53, CDKN2A, and SMAD4
- Curation and Informatic review identified a possible pathogenic VUS in KDR/VEGFR2



# Case Study



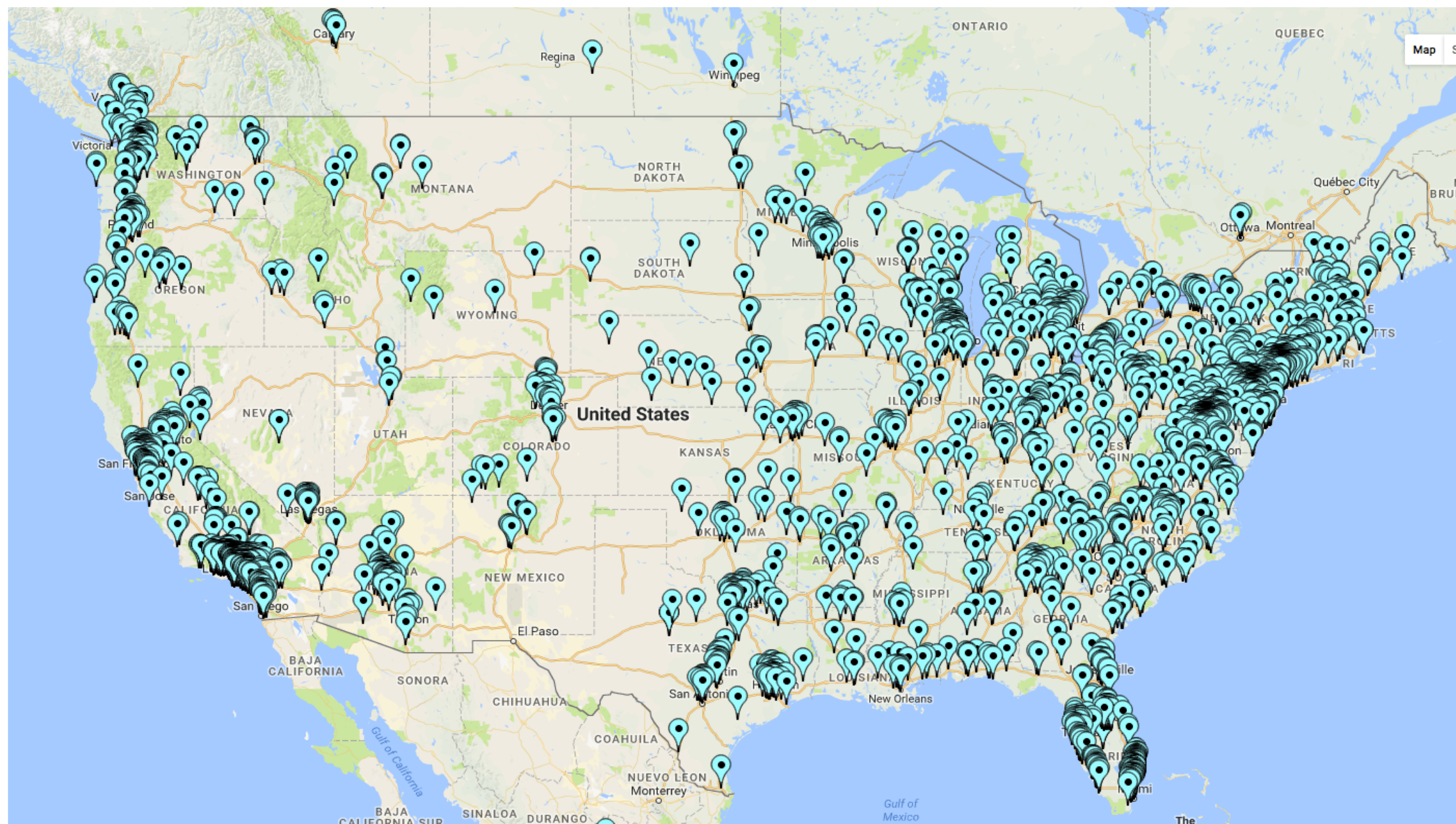
Ligand independent constitutive phosphorylation leading to excessive MAPK/PI3K/AKT/mTOR and ANGIOGENIC ACTIVITY

Multiple TKIs block anti-VEGFR2 activity both off label and in clinical trials including sorafenib, regorafenib, sunitinib, axitinib, cabozantinib, ponatinib

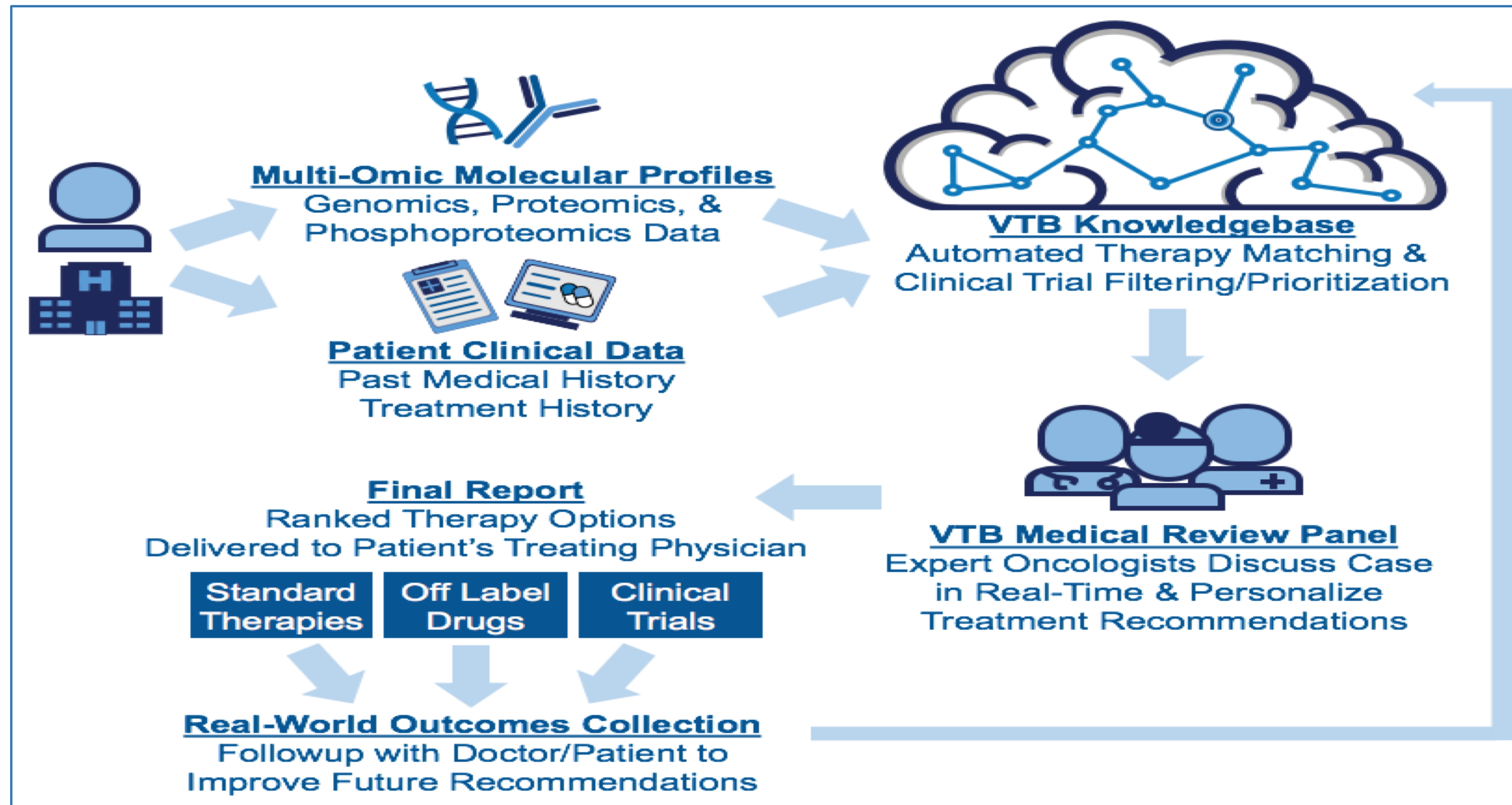
NCI-sponsored Sorafenib trial

*"I am working again; travelling; playing squash, golf, and hockey; and have gained 20lbs"*

# >1500 patients over 300 hospitals across the US



# AI-based Virtual Molecular Tumor Board



PanCAN Know Your Tumor Program

# KYT Program Information Engineering

- 4+ Years Published Data Collected, Extracted, Structured
- 4,000+ Rules for Treatment implications

## Patient Database

- Patient Multi-omic Profile
- Patient Demographics
- Patient Treatment History

## Clinical Trials Database

- 1,300+ Clinical Trials

## Published Literature Database

- 1,400+ publications collected, extracted, structured

## Guidelines Database

## Patient Outcomes Database

## Drugs Database

- 700+ Drugs  
Approved  
Off Label  
NCE's in Clinical Trials

## Variants Database

- 1,700+ with Actionability

## Published Multi-omic Database

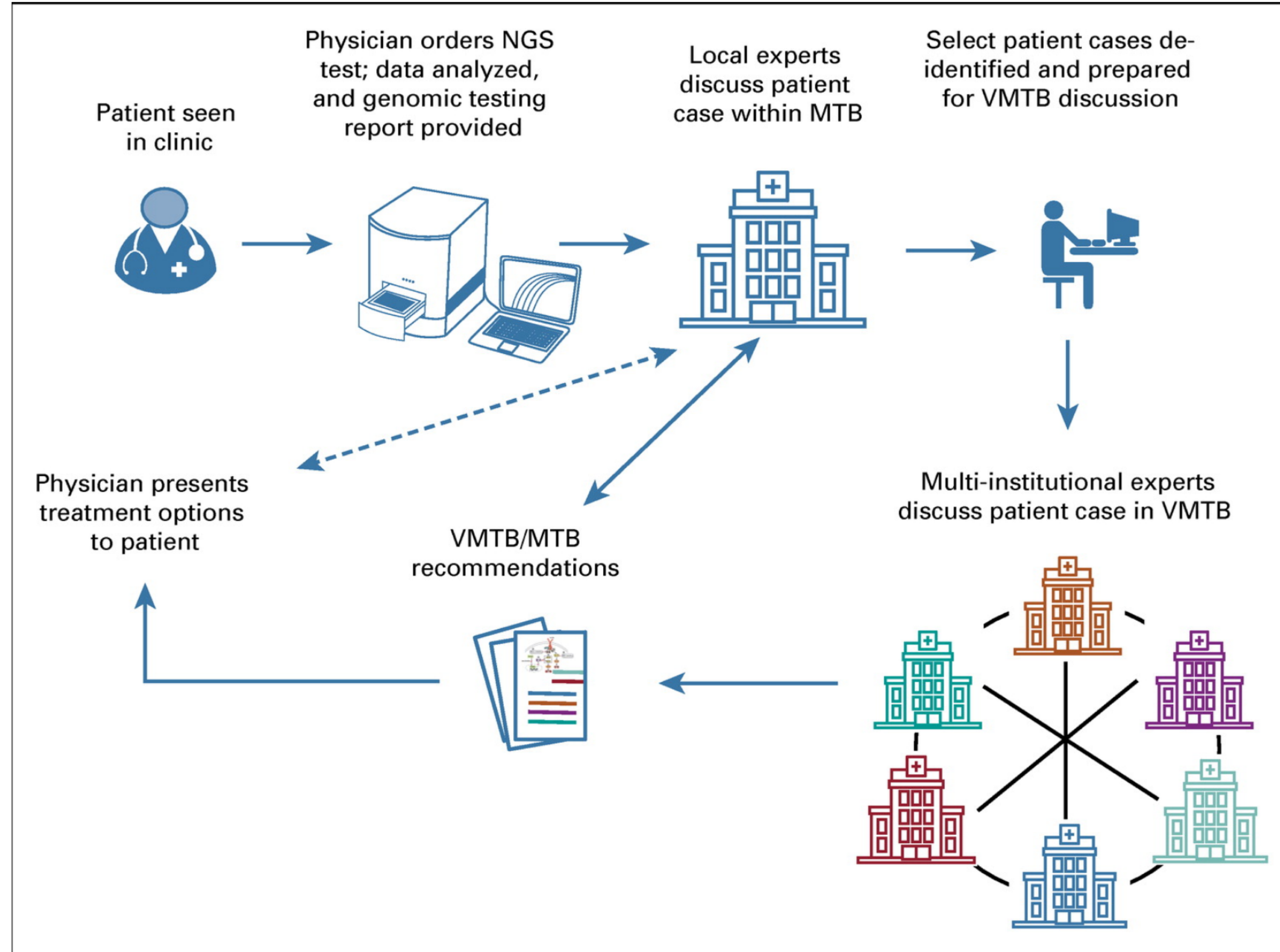
- NGS DNA
- NGS RNA
- Proteomic IHC
- ISH
- Proteomic Phosphorylation
- Germline DNA Sequencing

## Published Outcomes Database



**Continuously Monitored  
Real-time Updating**

# Collaborative, Multidisciplinary Evaluation of Cancer Variants Through Virtual Molecular Tumor Boards Informs Local Clinical Practices



# Barriers and Opportunities

- Lot of manual curation and data integration
- Need for consensus in variant interpretation
- Too many knowledgebases – onerous for clinicians and Phase 1 clinical trialists
- Lack of resources for curators creating one-off solutions
- Need more uptake by clinical researchers (training)

# Members : Academia, Industry, ACMG, AMP, CAP, GA4GH, NIH, FDA

*AMP, Austin TX, October 2015*



*AACR, New Orleans LA, April 2016*



*Hackathon, Scripps Research Institute, Oct, 2019*



*AACR, Washington DC, April 2017*




*AACR, Chicago, April 2018*



**ClinGen Somatic WG**

# Minimum Variant Level Data

## Somatic cancer variant curation and harmonization through consensus minimum variant level data

Deborah I. Ritter<sup>†</sup>, Sameek Roychowdhury<sup>†</sup>, Angshumoy Roy, Shruti Rao, Melissa J. Landrum, Dmitriy Sonkin, Mamatha Shekar, Caleb F. Davis, Reece K. Hart, Christine Micheel, Meredith Weaver, Eliezer M. Van Allen, Donald W. Parsons, Howard L. McLeod, Michael S. Watson, Sharon E. Plon, Shashikant Kulkarni, Subha Madhavan  and  
on behalf of the ClinGen Somatic Cancer Working Group

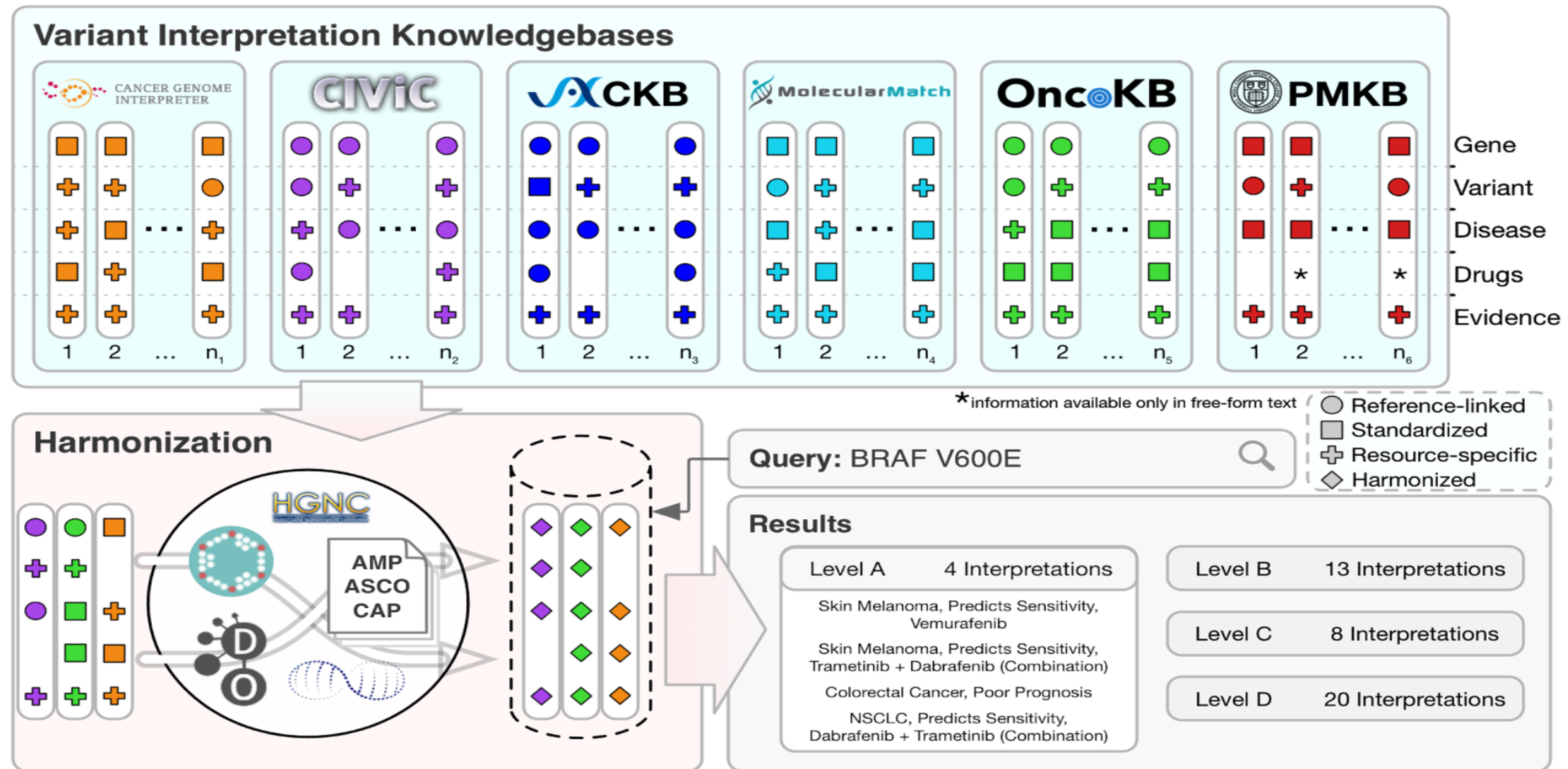
<sup>†</sup> Contributed equally

*Genome Medicine* 2016 8:117 | DOI: 10.1186/s13073-016-0367-z | © The Author(s). 2016

Received: 16 July 2016 | Accepted: 13 October 2016 | Published: 4 November 2016

27364 downloads on Genome Medicine  
Top 5% of all articles by Altametrics

# GA4GH Driver Project: A harmonized meta-knowledgebase of somatic clinical interpretations of cancer genomic variants



# N-of-1 to all patients with advanced cancers

- Need to scale for all patients with cancer
- Infrastructure must allow for multi-omic analysis to assess trial eligibility
- Harmonized knowledgebases of curated evidence must be readily available and updated to keep up with evolving science



# Story 2

Automatic extraction of Clinical genomic evidence for  
variant interpretation using Natural Language Processing  
and Machine Learning

# Limitations of Bio- and Health- NLP Tools

- Mostly Rule-based
- Domain-specific
- Abstract-based

# Goal: Improve and Automate Relationship Extraction

---

- Machine learning approach
  - Learning from “less” annotated data
- Extend relation extraction to cover germline mutations
- Extract information from the literature to support variant pathogenicity curation in ClinGen based on ACMG guidelines for sequence variant interpretation.
- Extend processing to full-text, tables, figures, supplementary information.

# Learning from “Less Data”

Use annotations in existing knowledge bases (CIVIC, ClinGen, Hypotheses, Clinvar) as gold/training set for ML models.

Problem: Small size of training set -- Less Data

Solution: **Distant Supervision, Semantic Abstraction**

# Semantic Abstraction

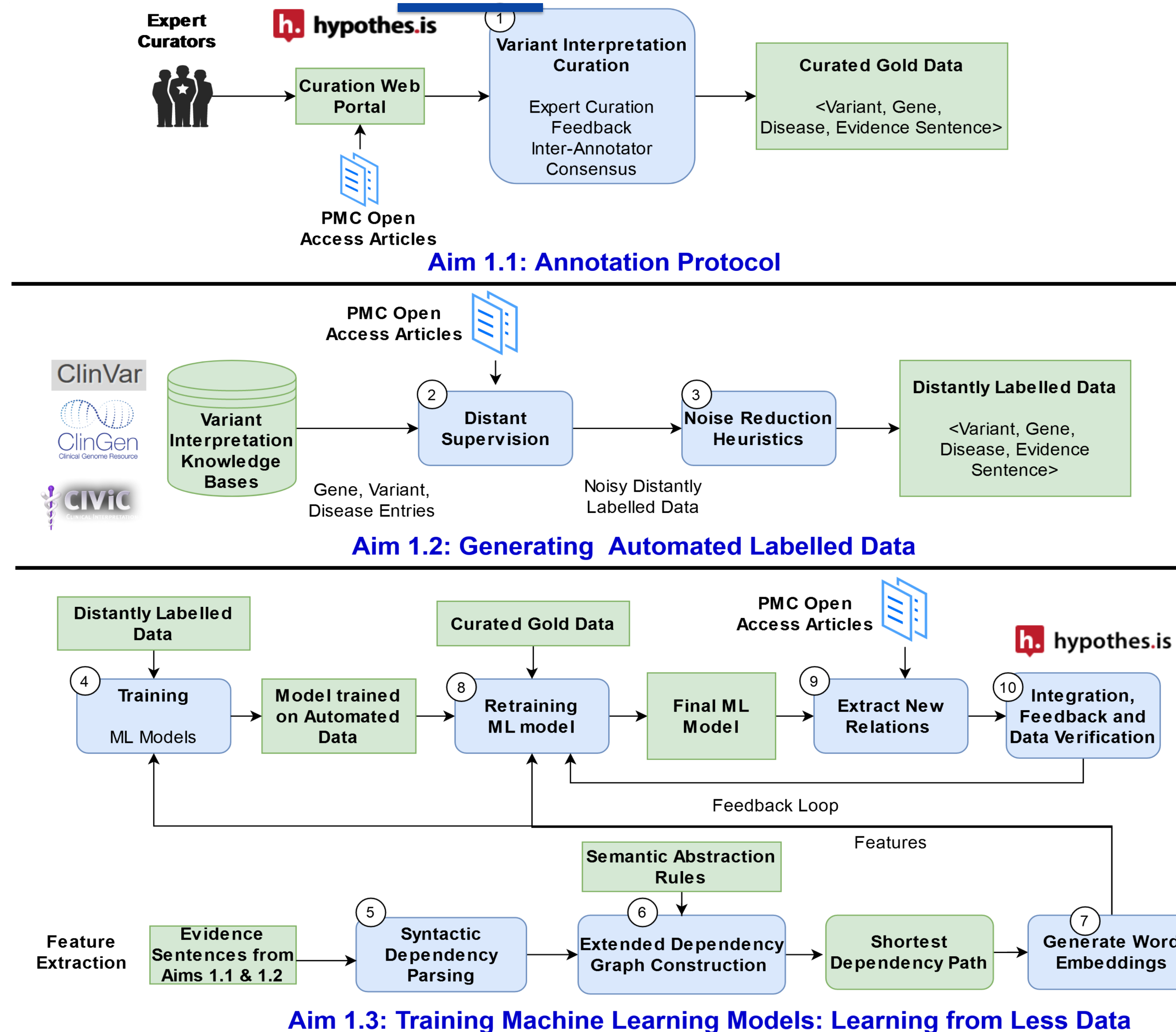
- Abstract away from textual and syntactic variations.
- Extended Dependency Graph: unified representation (BioNLP 2015)
- Explore Extended Dependency Graph (EDG) with the shortest dependency path as feature representation.

Association of Desmin **mutation** with idiopathic dilated cardiomyopathy ...

Desmin **mutation** [association] idiopathic dilated  
cardiomyopathy

Desmin **mutation** is associated with idiopathic dilated cardiomyopathy.

# MACE2K Informatics Infrastructure (Proof of Concept)



# One-year Masters Program Health Informatics & Data Science



**GEORGETOWN UNIVERSITY**  
Biomedical Graduate Education



THE INSTITUTE FOR COMPUTATIONAL BIOMEDICAL INFORMATICS  
OFFICE OF HEALTH INFORMATICS AND DATA SCIENCE

## Master's Degree Health Informatics & Data Science

The Master's in Health Informatics & Data Science (HIDS) at Georgetown University is an industry-driven, career-ready program focused on current and emerging technologies that will inform healthcare. The program offers students access to Georgetown's distinguished faculty and unique opportunities in the nation's capital. Students will gain core competency in data science, big data analytics, digital ethics, artificial intelligence and machine learning applications to inform health outcomes and reduce costs. Our goal is to develop responsible leaders in health data science and medical informatics with rewarding careers in industry, government or academia.

For more information, visit  
<https://healthinformatics.georgetown.edu/>

**PROGRAM LEADERSHIP**  
Subha Madhavan, Ph.D., FACMI  
Director, M.S HIDS  
  
Adil Alaoui, M.S, MBA.  
Director, Industry Capstone Program  
  
Yuriy Gusev, Ph.D.  
Curriculum Director  
  
Shruti Rao, M.S, MBA  
Program Coordinator



**PROGRAM FEATURES**



12 month  
accelerated program



32 credits  
Full-time or part-time



Final capstone project with  
industry partners or government agencies

**WE ARE NOW ACCEPTING APPLICATIONS  
FOR FALL 2019 ADMISSION**

Application Deadline: July 1st, 2019

For further information visit:  
Website: <https://healthinformatics.georgetown.edu/>  
Email: [hids@georgetown.edu](mailto:hids@georgetown.edu)  
Phone: +1 (202) 687 1093



**GEORGETOWN UNIVERSITY**  
Biomedical Graduate Education



THE INSTITUTE FOR COMPUTATIONAL BIOMEDICAL INFORMATICS  
OFFICE OF HEALTH INFORMATICS AND DATA SCIENCE

## Master's Degree Health Informatics & Data Science



<https://healthinformatics.georgetown.edu/>

The program includes comprehensive coursework, hands-on projects and a mandatory Capstone Project that is the culmination of the student's coursework and experience at Georgetown University. Students will work with a host organization and an academic mentor to critically analyze a real world problem, design, conceptualize and implement a solution that applies the knowledge and skills acquired during the Masters program.

FALL	SPRING	SUMMER
<ul style="list-style-type: none"> <li>• Introduction to Health Data Science &amp; Analytics (HIDS-501)</li> <li>• Health Informatics fundamentals (HIDS-502)</li> <li>• Utilizing Data in Electronic Medical Records (HIDS-503)</li> <li>• Evidence based data analysis in population health (HIDS-504)</li> </ul>	<ul style="list-style-type: none"> <li>• Massive Data Fundamentals (ANLY-502)</li> <li>• Precision Medicine Informatics: MolDx, high throughput omics and analytics (HIDS-505)</li> <li>• AI for health applications (HIDS-506)</li> <li>• Digital Health Applications (HIDS-507)</li> </ul>	<ul style="list-style-type: none"> <li>• Human Factor Engineering, Usability and Safety (HIDS-508)</li> <li>• Imaging Informatics (HIDS-509)</li> <li>• Mandatory Capstone (HIDS-510)</li> </ul>

**INDUSTRY & GOVERNMENT ADVISORY PANEL**

<ul style="list-style-type: none"> <li>• Sage Bionetworks</li> <li>• ESAC</li> <li>• FDA</li> <li>• Deloitte</li> <li>• Flatiron Health</li> <li>• COTA</li> <li>• Syapse</li> <li>• TrinetX</li> <li>• INOVA</li> <li>• MedStar Institute for Innovation (MI2)</li> <li>• Cerner Corporation</li> </ul>	<ul style="list-style-type: none"> <li>• Indivumed Group</li> <li>• Argentys</li> <li>• CVS Health</li> <li>• PwC</li> <li>• Open Source Electronic Health Record Alliance (OSEHRA)</li> <li>• RadAmerica</li> <li>• LexisNexis</li> <li>• Assymetrik</li> </ul>
--	--

**APPLICATION CHECKLIST**

- 4-year Bachelor's degree
- Grade point average of 3.0 or higher
- Official transcripts
- 2 letters of recommendation
- Academic Statement of Purpose
- Interview with faculty
- Non-refundable application fee
- TOEFL/IELTS Score (if applicable)

# VTB Collaborators

## Georgetown Lombardi Comprehensive Cancer Center

Subha Madhavan, PhD  
Michael Pishvaian, MD, PhD  
Paula Pohlmann, MD, MSc, PhD  
Mike Atkins, MD

## George Mason University

Emanuel Petricoin III, PhD

## City of Hope Cancer Center

Vincent Chung, MD

## Cedars-Sinai Medical Center

Andrew Hendifar, MD

## Arizona Oncology

Bradley Monk, MD

## Perthera

David Halverson, MD  
Edik Blais, PhD  
R Joseph Bender, PhD

## Memorial Sloan Kettering Cancer Center / Weill Cornell Medical College

Maeve Lowery, MD  
Neil Shah, MD

## The Ohio State University Comprehensive Cancer Center

Sam Mikhail, MD  
Kai He, MD

## Case Comprehensive Cancer Center, University Hospitals Seidman Cancer Center and Cleveland Clinic Taussig Cancer Institute

Davendra Sohal, MD, MPH

## University of Cincinnati Cancer Institute

Thomas Herzog, MD

## Stephenson Oklahoma Cancer Center at the University of Oklahoma

Kathleen Moore, MD

## MD Anderson Cancer Center

Robert Coleman, MD

**Funding: Pancreatic Cancer Action Network RAN grant, NLM U01**

# ClinGen & MACE2K Collaborators

**Funding: NIH U01 HG007437**  
*ClinGen: A Knowledge Base for  
Clinically Relevant Genes and  
Variants*

Samantha	Baxter	Broad Institute
Rong	Chen	Mt. Sinai
Barbara	Conley	ASCO
Helio	Costa	Stanford University
Amy	Dodson	WUSTL
Sheryl	Elkin	N of One
Mezbah	Faruque	Howard University
Kelly	Filipski	NIH
Andrew	Freedman	NIH
Laurie	Gay	Foundation Medicine
Stacy	Gray	Dana Farber
Malachi	Griffith	WUSTL
Obi	Griffith	WUSTL
Michael	Harris	Georgetown University
Steven	Harrison	Harvard/Partners
Reece	Hart	Invitae
Carolyn	Hutter	NHGRI
Anthony John	Iafrate	Harvard/Partners
Michael	Italia	Syapse
Katherine	Janeway	Dana Farber
Sapna	Joshi	Moffitt Cancer Center
Michael	Kelley	Duke University
Warren	Kibbe	NIH
Shashi	Kulkarni	Baylor College of Medicine
Melissa	Landrum	ClinVar
Shirely	Li	Molecular Match
Marilyn	Li	CHOP
Rachel	Liao	Genomics & Health
Subha	Madhavan	Georgetown University
Peter	McGarvey	Georgetown University

Howard	McLeod	Moffitt Cancer Center
Annette	Meredith	Color
Funda	Meric-Bernstam	MD Anderson
Jason	Merker	Stanford University
Christine	Micheel	Vanderbilt Medical Center
Vincent	Miller	Foundation Medicine
Susan	Mockus	Jackson Labs
Will	Parsons	Baylor College of Medicine
Nirali	Patel	University of North Carolina
Anand	Pathak	FDA
Snehit	Prabu	Stanford University
Erin	Ramos	NHGRI
Shruti	Rao	Georgetown University
Sandeep	Reddy	Caris Life Sciences
Deborah	Ritter	Baylor College of Medicine
Angshumoy	Roy	Baylor College of Medicine
Sameek	Roychowdhury	Ohio State Cancer Center
Richard	Schilsky	ASCO
Safaraz	Shakih	Dana Farber
Mamatha	Shekar	Illumina
James	Shima	Syapse
Nikoletta	Sidiropoulos	University of Vermont
Dmitriy	Sonkin	NIH
Eli	Van Allen	Dana Farber
Christine	Walko	Moffitt Cancer Center
Kai	Wang	Foundation Medicine
Mike	Watson	ACMG
Meredith	Weaver	ACMG
Firas	Wehbe	Northwestern University
Nick	Xiao	Caris Life Sciences

# ICBI Center Core Team Members

---

- <http://icbi.georgetown.edu/>
- [icbi@georgetown.edu](mailto:icbi@georgetown.edu)
- <https://icbi.github.io/>
- @ICBI\_Georgetown
- <https://healthdatascience.blogspot.com/>

