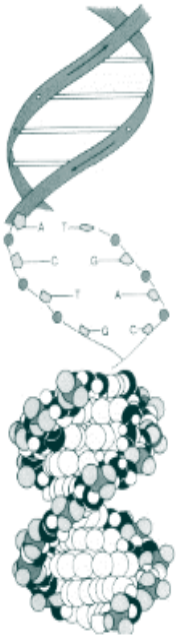
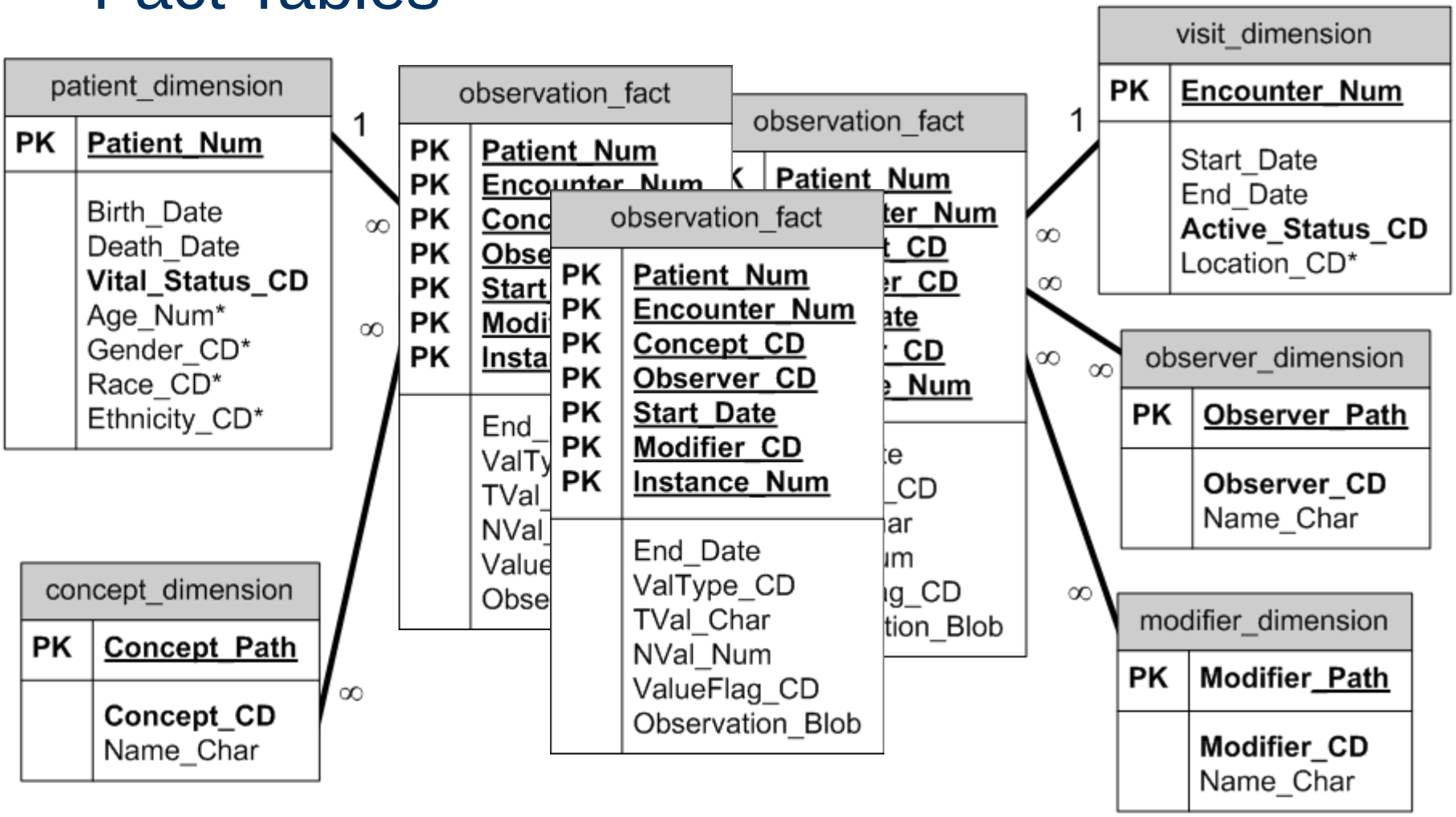


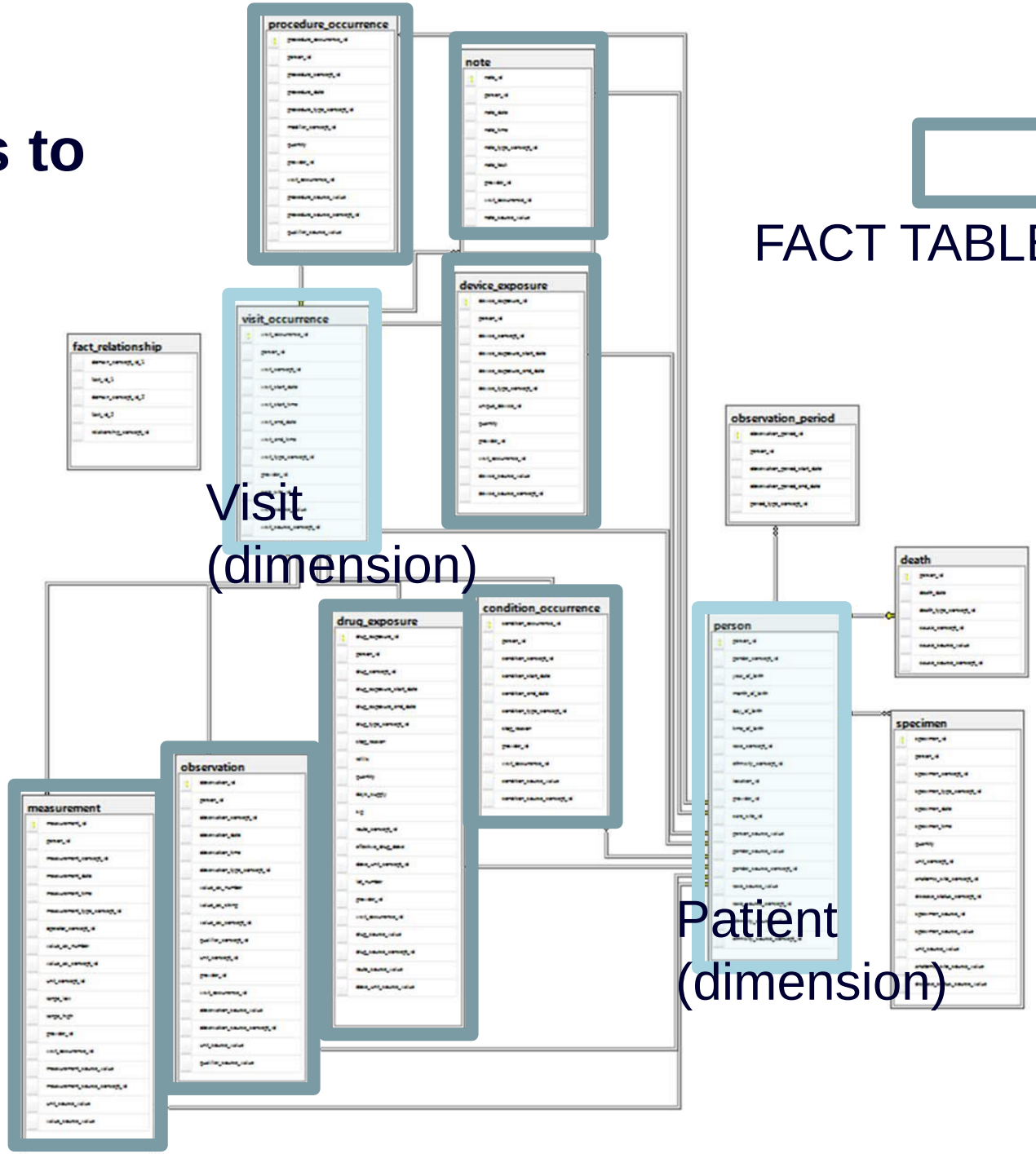
Advanced i2b2 Topics

Shawn Murphy MD, Ph.D.
Mass General Brigham and Harvard Medical School
November 14th, 2020

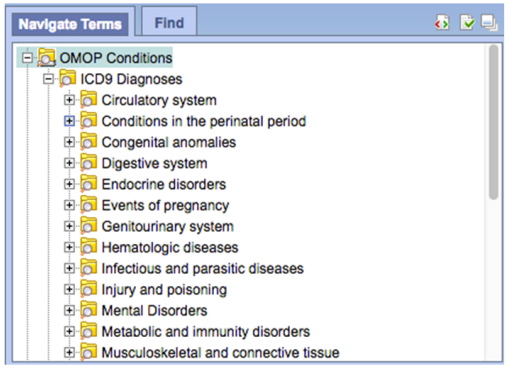
i2b2 Star Schema can adapt to Multiple Fact Tables

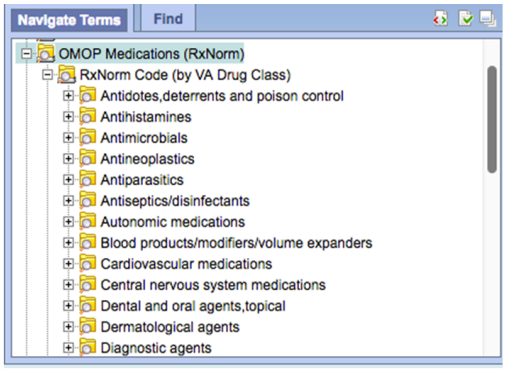


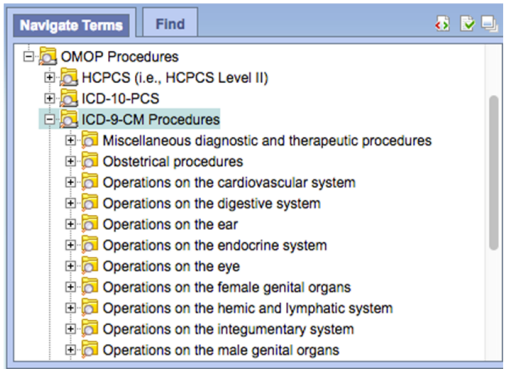
Adapts to OMOP

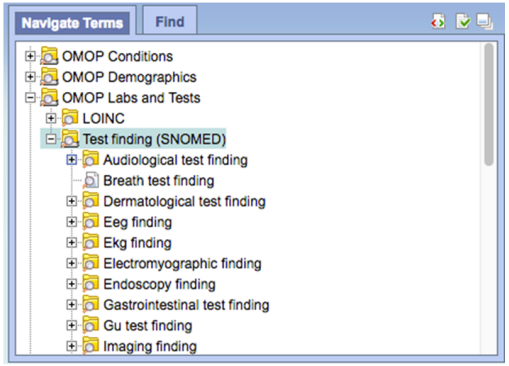


Connect to OMOP by building ontology of OMOP standard concepts







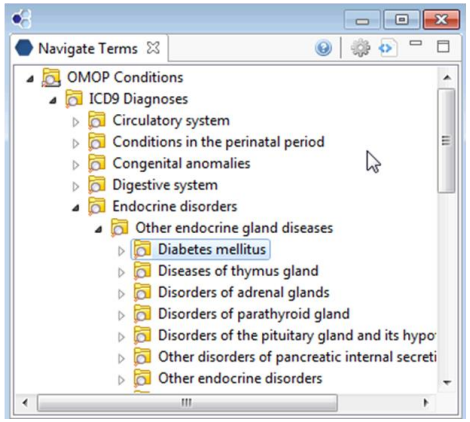


Ontologies covering the condition, procedures, drug, measurement and observation domains. All terms are mapped to standard concepts using OMOP's mapping tables

Use Ontology Tables to direct Queries to proper Fact Table view

Prepend c_facttablecolumn with OMOP domain view and modify CRC to parse into ‘domain_view’ and ‘c_facttablecolumn’

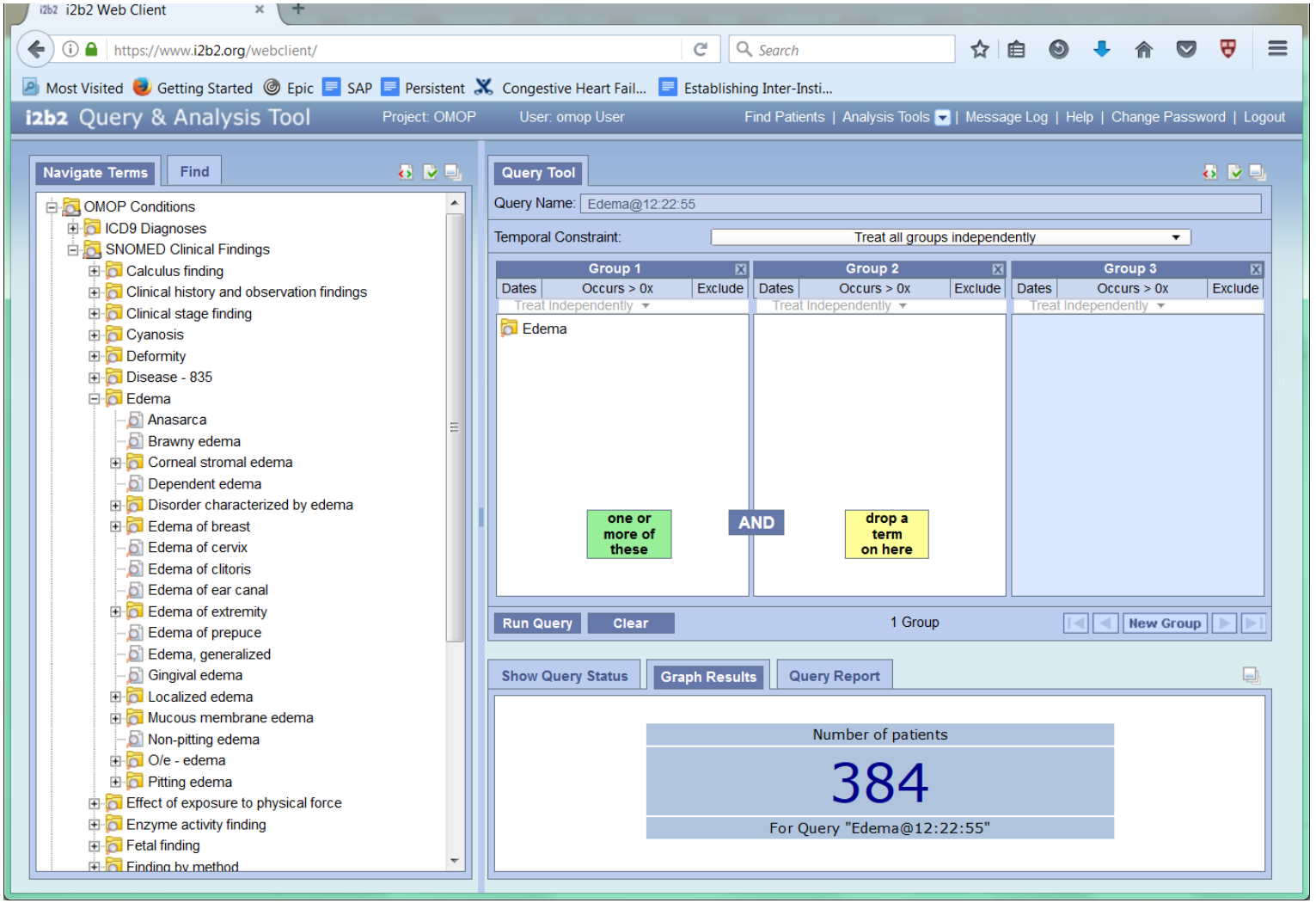
c_name	c_facttablecolumn	c_tablename	c_columnname	c_operator	c_dimcode
Diabetes mellitus	condition_view.concept_cd	concept_dimension	concept_path	LIKE	\i2b2\Diagnoses\Endocrine disorders (240-259)\Other endocrine gland



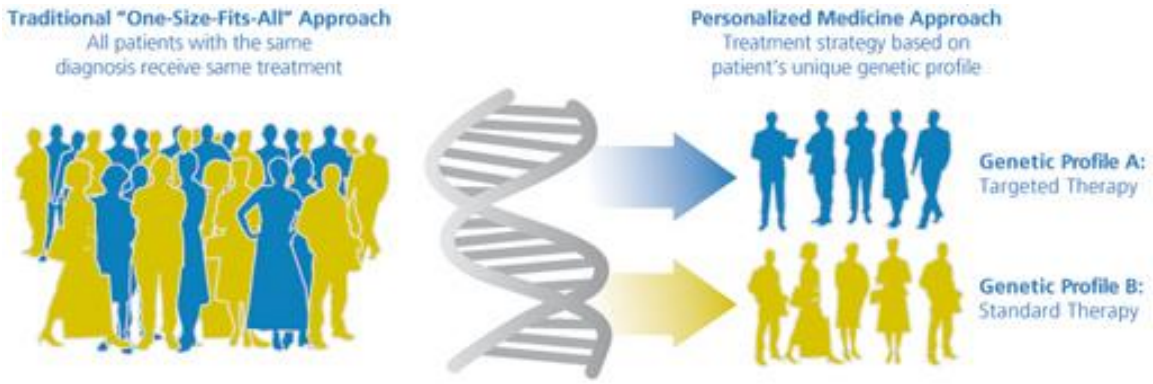

```
select patient_num from
condition_view where concept_cd
IN
(select
concept_cd from concept_dimension where
concept_path like '\i2b2\Diagnoses\Endocrine
disorders (240-259)\Other endocrine gland
diseases (250-259)\(250) Diabetes mellitus\%‘)
```

https://community.i2b2.org/wiki/display/OMOP

OMOP Linked into i2b2 software



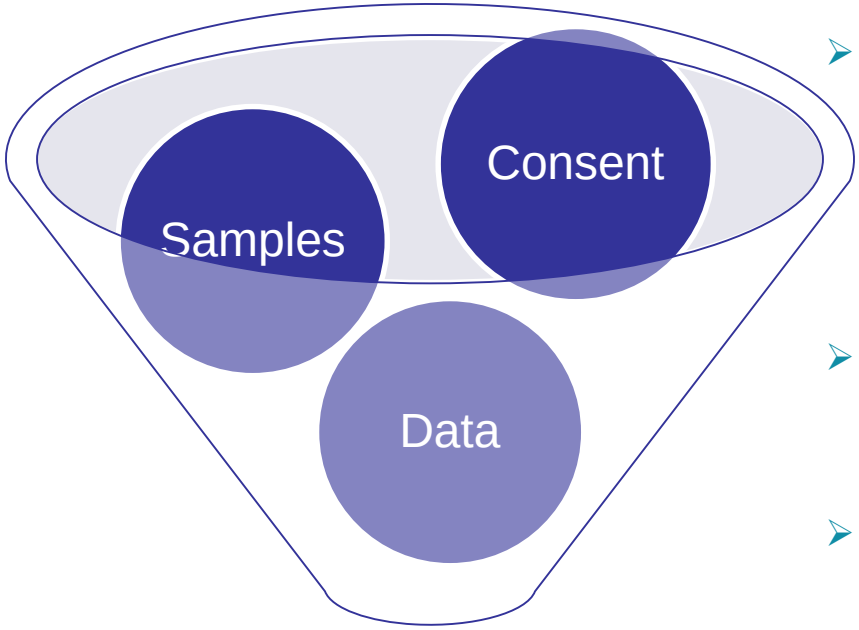
Personalized Medicine and Genomic technology are critical to managing populations



- Managing a population involves improving health outcomes of the group as a whole by identifying, monitoring and addressing health needs of individuals through:
  - Subpopulation stratification
  - Targeted, evidence-based treatment protocols
  - Predictive analytics

Source: Personalized Medicine Coalition and innovation.org ; Oliver Wyman

The Partners Biobank



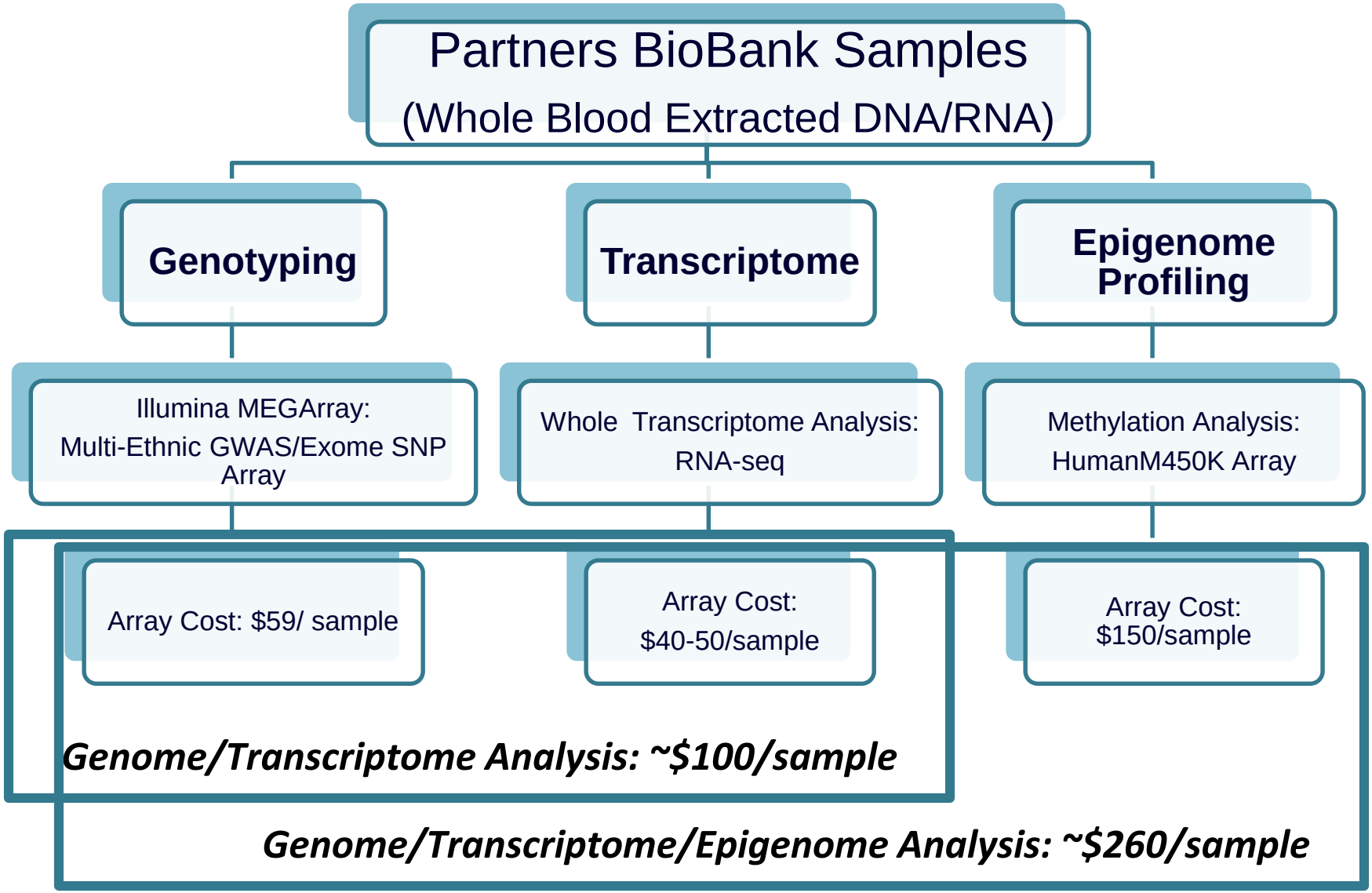
Research Discoveries

- The Partners Biobank provides samples (plasma, serum, and DNA) collected from consented patients.
- 84,000 patients have consented to date
- Samples are available for distribution to Partners investigators* to help identify novel Personalized Medicine opportunities that reduce cost and provide better care

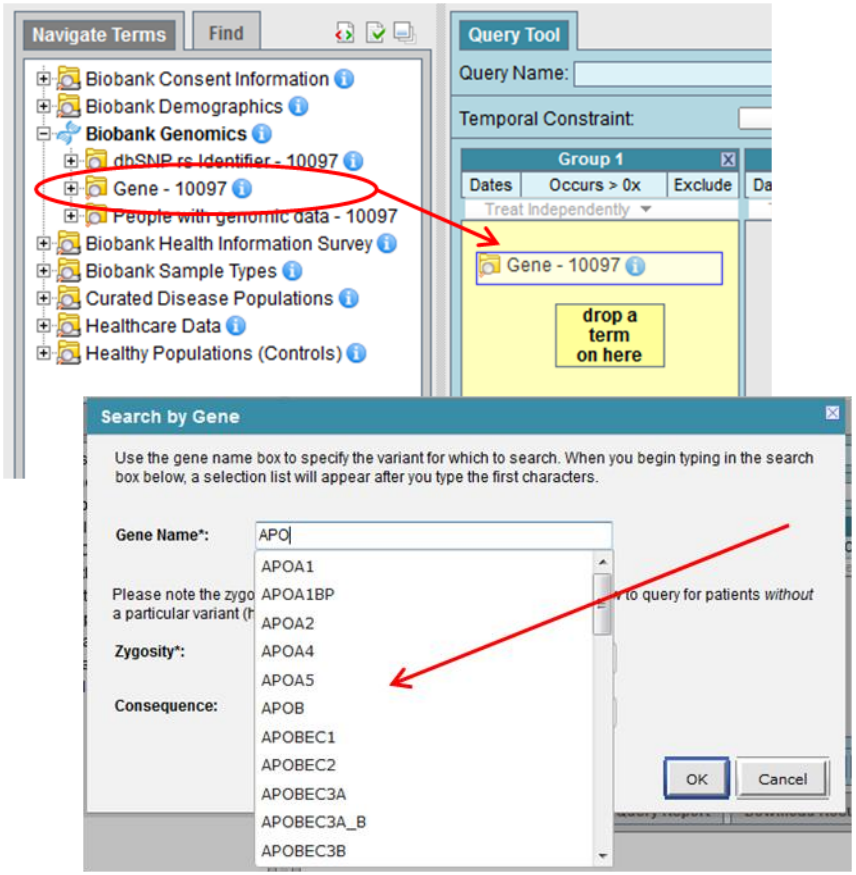
**with required approval from the Partners Institutional Review Board (IRB).*

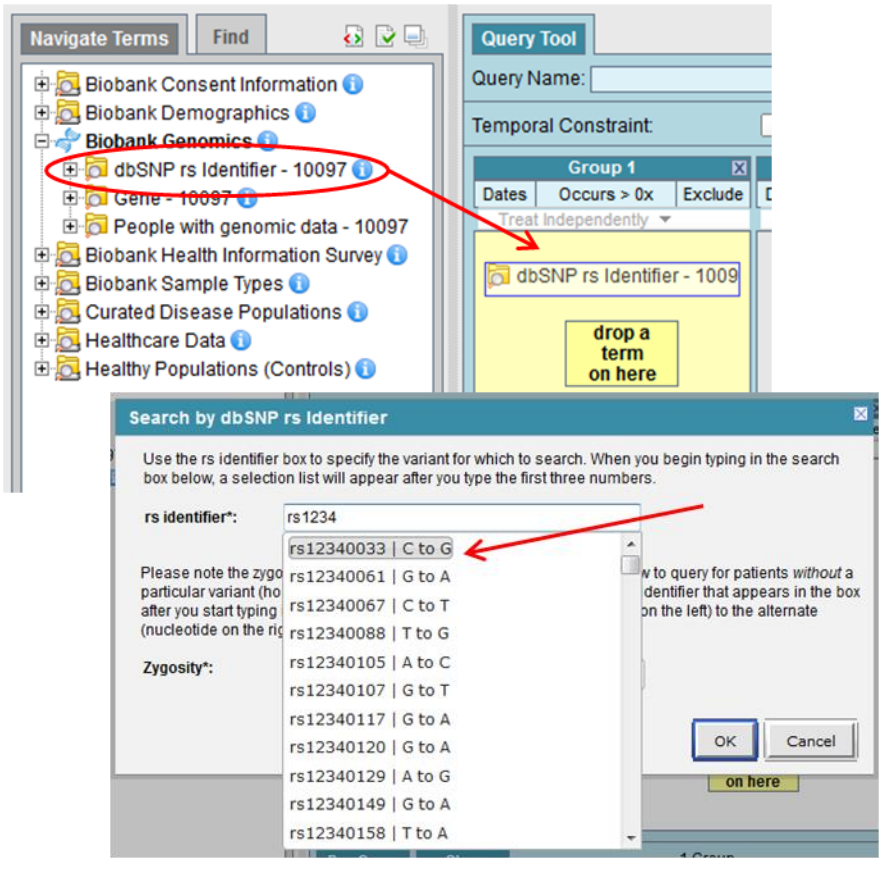
Improved Clinical Care for All Patients

Biobank Integrative Genomics Strategy

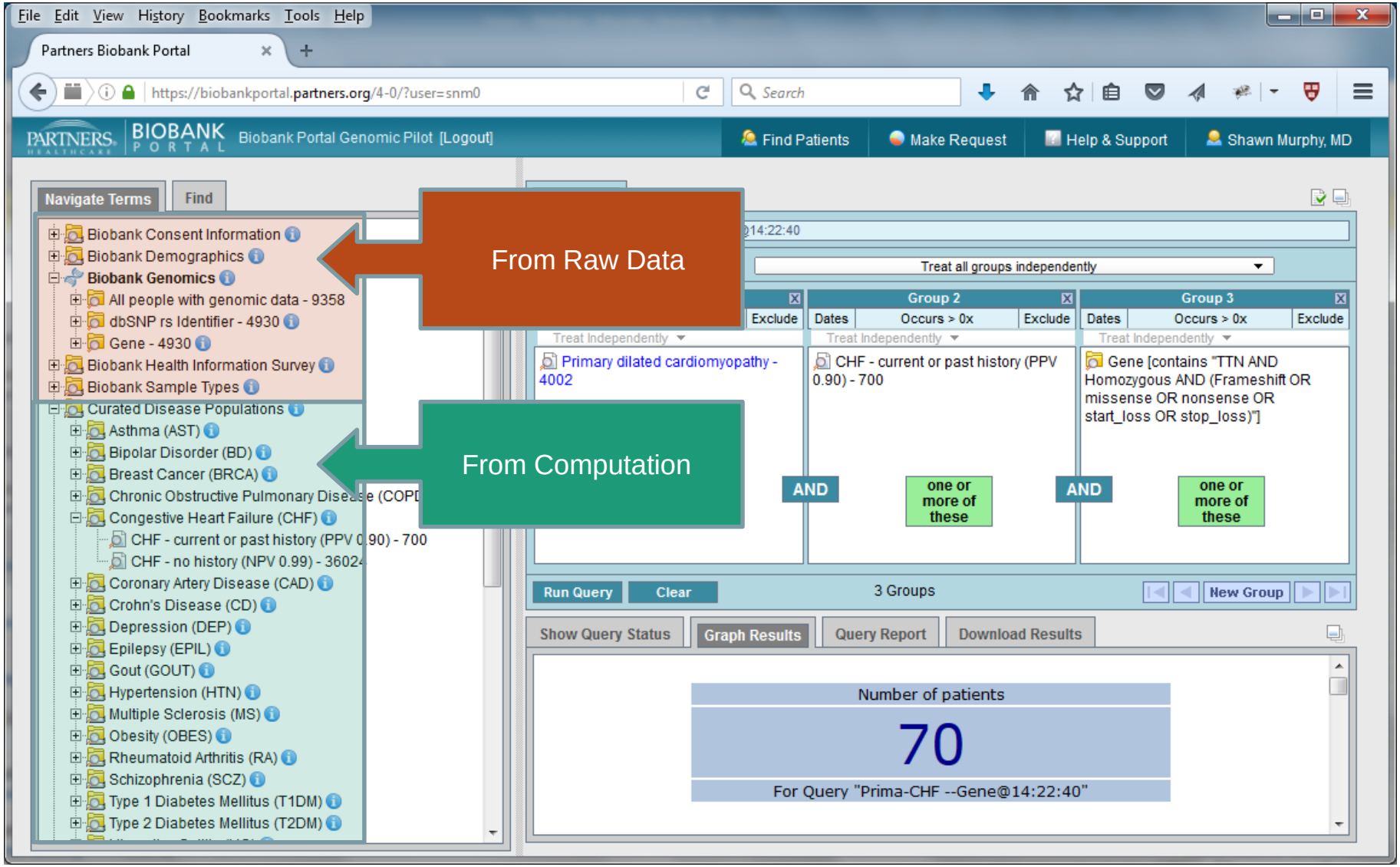


Genotype Data





https://community.i2b2.org/wiki/display/IGD/Loading+Genomic+VCF+Files+into+i2b2


File Edit View History Bookmarks Tools Help
Partners Biobank Portal
https://biobankportal.partners.org/4-0/?user=snm0
Search
PARTNERS BIOBANK PORTAL
Biobank Portal Genomic Pilot [Logout]
Find Patients
Make Request
Help & Support
Shawn Murphy, MD
Navigate Terms
Find
Biobank Consent Information
Biobank Demographics
Biobank Genomics
All people with genomic data - 9358
dbSNP rs Identifier - 4930
Gene - 4930
Biobank Health Information Survey
Biobank Sample Types
Curated Disease Populations
Asthma (AST)
Bipolar Disorder (BD)
Breast Cancer (BRCA)
Chronic Obstructive Pulmonary Disease (COPD
Congestive Heart Failure (CHF)
CHF - current or past history (PPV 0.90) - 700
CHF - no history (NPV 0.99) - 3602
Coronary Artery Disease (CAD)
Crohn's Disease (CD)
Depression (DEP)
Epilepsy (EPIL)
Gout (GOUT)
Hypertension (HTN)
Multiple Sclerosis (MS)
Obesity (OBES)
Rheumatoid Arthritis (RA)
Schizophrenia (SCZ)
Type 1 Diabetes Mellitus (T1DM)
Type 2 Diabetes Mellitus (T2DM)
From Raw Data
From Computation
14:22:40
Treat all groups independently
Group 2
Group 3
Dates
Occurs > 0x
Exclude
Treat Independently
Primary dilated cardiomyopathy - 4002
CHF - current or past history (PPV 0.90) - 700
Gene [contains "TTN AND Homozygous AND (Frameshift OR missense OR nonsense OR start_loss OR stop_loss)"]
AND
one or more of these
AND
one or more of these
Run Query
Clear
3 Groups
New Group
Show Query Status
Graph Results
Query Report
Download Results
Number of patients
70
For Query "Prima-CHF --Gene@14:22:40"

In Digitally Driven Healthcare, Disease Labels determine Algorithms for Managing Patient

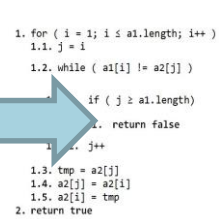
Mary

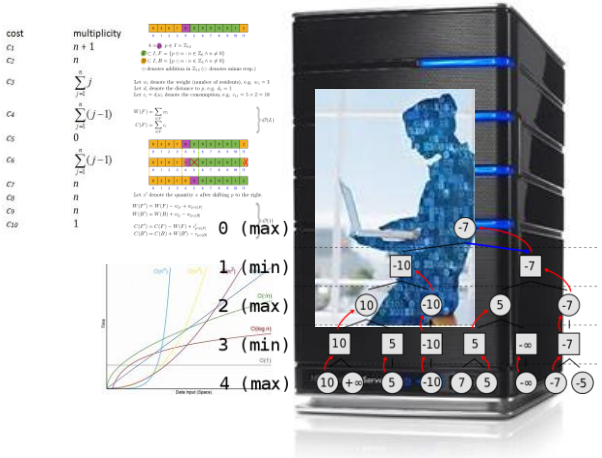
Followed for:

- Prevention
- **Diabetes Type II**
- Hypertension
- Heart Failure

Innovations Enabling Digital Care

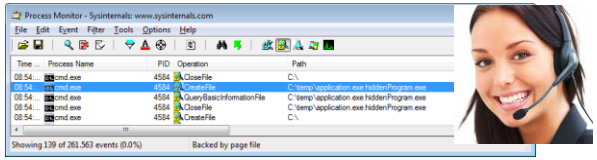
Digital and IoT devices continuously output Patient Data



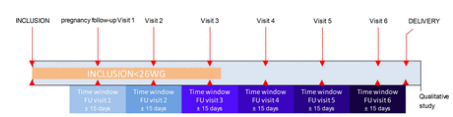


PATIENT

Digital Twin of patient enables continuous assessment of patient with Real Time Algorithms

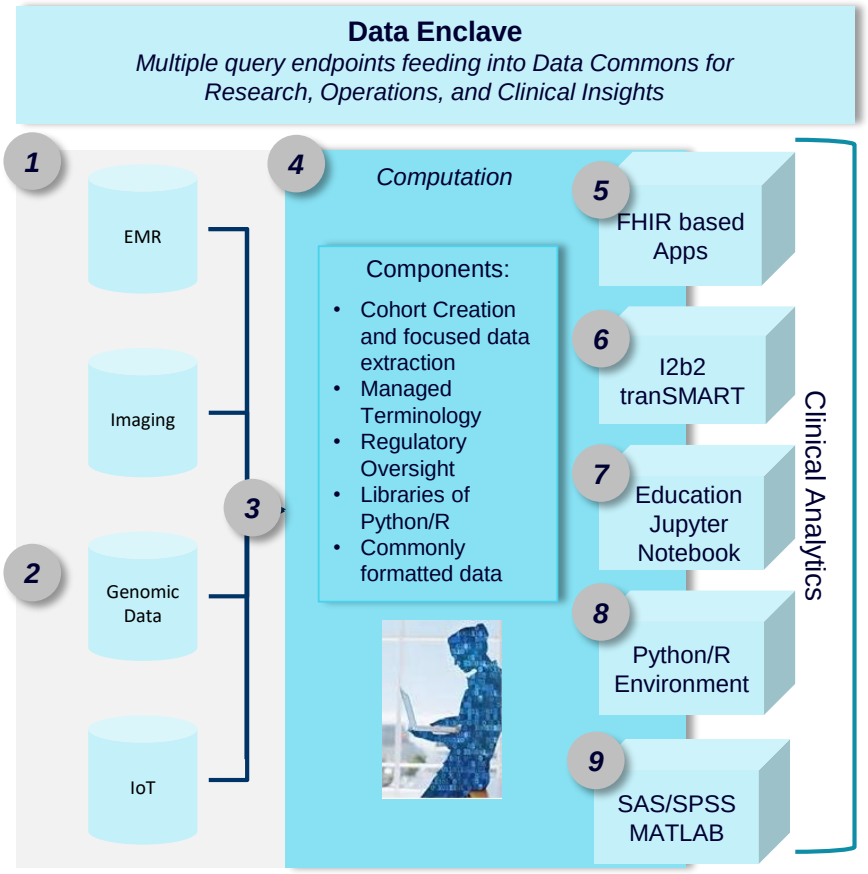


Navigator Model dramatically increases Frequency and Convenience for Patient Communication



System drives Pragmatic Clinical Trials Leading to Continuous Process Improvement

Digital Twin for Continuously Assessing Patient



Data Enclave	
1	Enable Data Extracts, perhaps some are Federated
2	Combine and Link data, put in common OMOP/i2b2 format
3	Conduit to Data Enclave
Technical Solution Development	
4	Healthcare Ready Bundle
Research and Clinical Application Projects	
5	FHIR based SMART Apps
6	I2b2 tranSMART with Fractalis plugin (next version of SmartR plugin)
7	Jupyter Notebook with AI Visualizations – code can advance to production
8	Python/R Environment full interactive development in Data Lake

Many, many contributions …

- Management
  - Isaac Kohane
  - Susanne Churchill
  - Griffin Weber
  - Lee Nadler
  - Douglas MacFadden
  - Christopher Herrick
  - Victor Castro
  - Roy Perlis
  - Beth Karlson
  - Paul Avillach
  - Jeff Klann
  - Michael Mendis
- Biobank/Phenotyping
  - Scott Weiss
  - Beth Karlson
  - Jordon Smoller
  - Natalie Bouton
  - Vivian Gainer
  - Nich Wattanasin
- Healthcare Innovation Team
  - Kavi Wagholikar
  - Randy Gollub
  - Calum Macrea
  - Sandy Aronson

Special thanks to…

Core i2b2 Developers over past 14 years:

- Michael Mendis
- Lori Phillips
- Raj Kuttan
- Wensong Pan
- Griffin Weber
- Andrew McMurry
- Janice Donahoe
- Nich Wattanasin
- William Simons
- Christopher Herrick
- Bill Wang
- Kavi Wagholikar
- Wayne Chan
- Bhaswati Ghosh
- Reeta Metta
- Barbara Benoit
- David Wang
- Jeff Klann

THANK YOU