

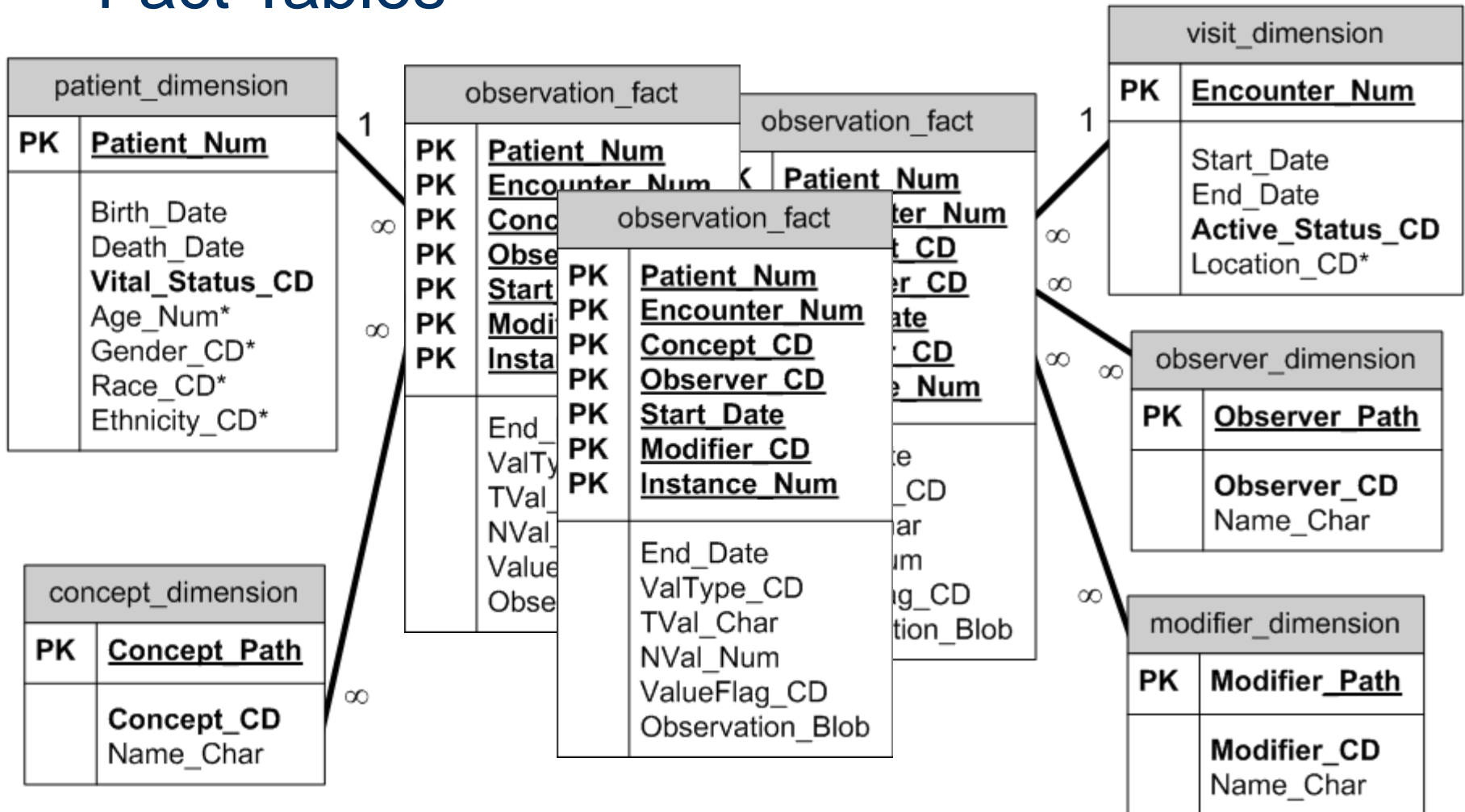
# Advanced i2b2 Topics

*Shawn Murphy MD, Ph.D.*

*Mass General Brigham and Harvard Medical School*

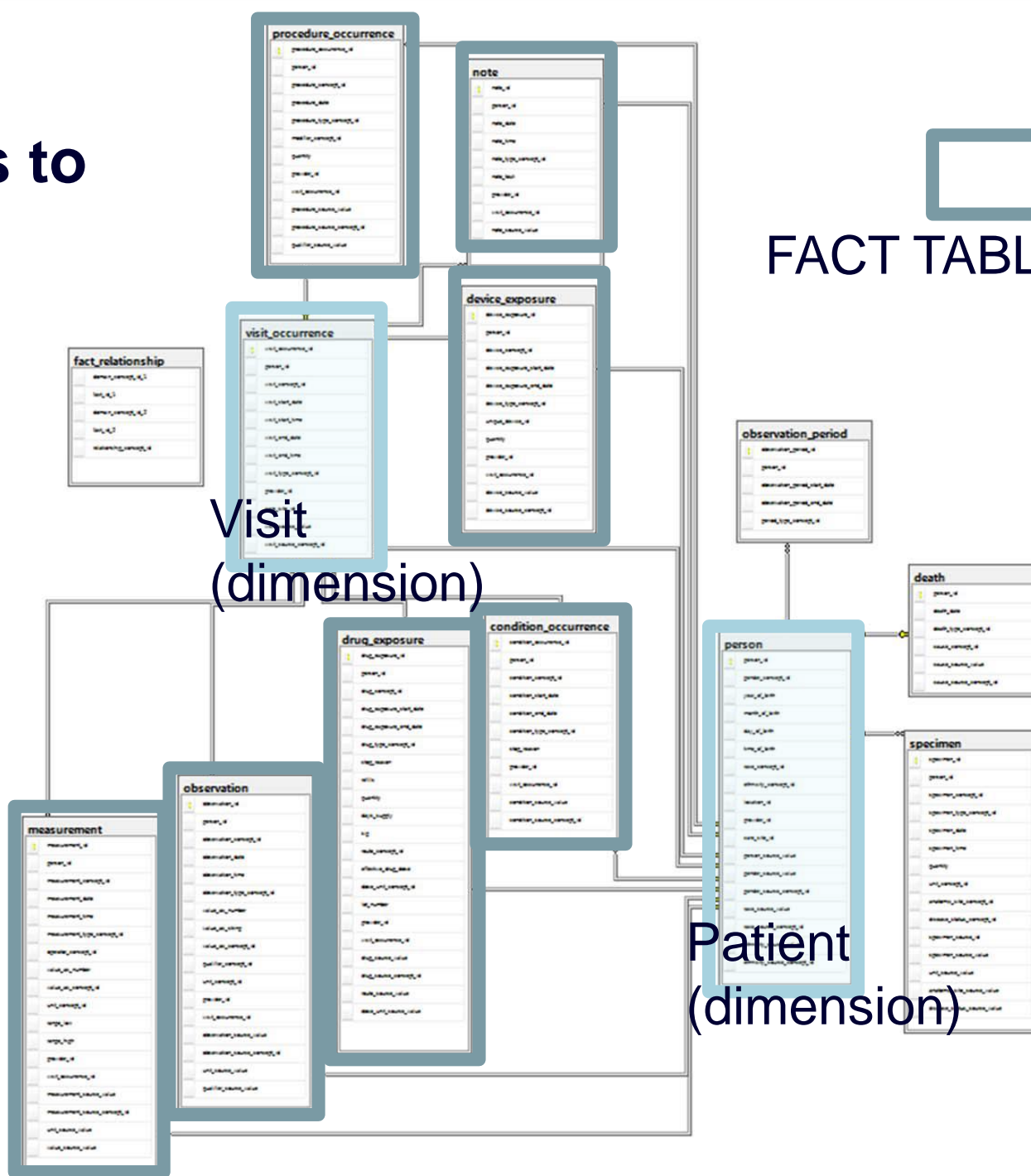
*November 14<sup>th</sup>, 2020*

# i2b2 Star Schema can adapt to Multiple Fact Tables

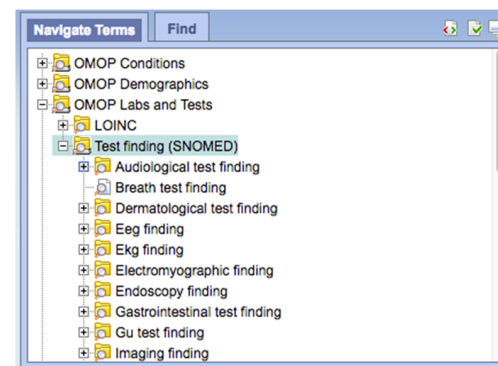
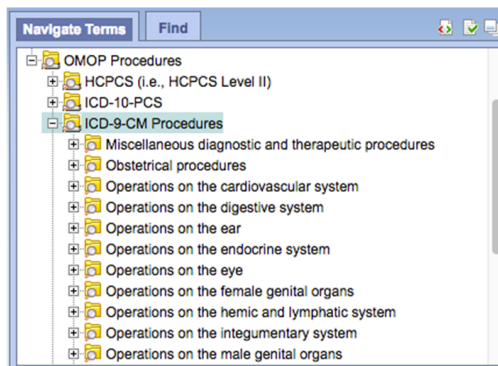
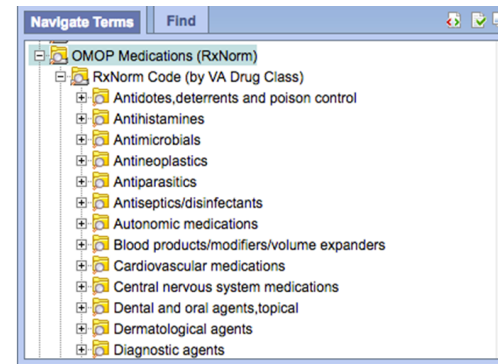
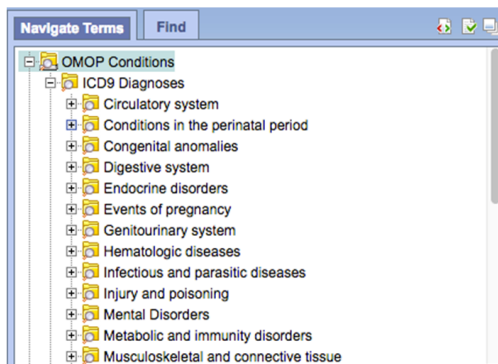


# Adapts to OMOP

FACT TABLES



# Connect to OMOP by building ontology of OMOP standard concepts

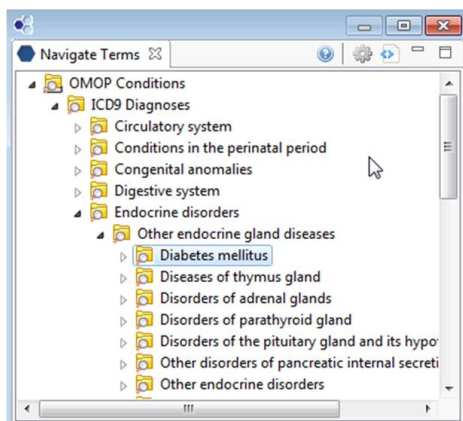


Ontologies covering the condition, procedures, drug, measurement and observation domains. All terms are mapped to standard concepts using OMOP's mapping tables

# 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

The screenshot displays the i2b2 Query & Analysis Tool interface. The browser address bar shows <https://www.i2b2.org/webclient/>. The page title is "i2b2 Query & Analysis Tool" and the user is logged in as "omop User".

**Left Panel: Navigate Terms**

- OMOP Conditions
  - ICD9 Diagnoses
  - SNOMED Clinical Findings
    - Calculus finding
    - Clinical history and observation findings
    - Clinical stage finding
    - Cyanosis
    - Deformity
    - Disease - 835
    - Edema
      - Anasarca
      - Brawny edema
      - Corneal stromal edema
      - Dependent edema
      - Disorder characterized by edema
      - Edema of breast
      - Edema of cervix
      - Edema of clitoris
      - Edema of ear canal
      - Edema of extremity
      - Edema of prepuce
      - Edema, generalized
      - Gingival edema
      - Localized edema
      - Mucous membrane edema
      - Non-pitting edema
      - O/e - edema
      - Pitting edema
    - Effect of exposure to physical force
    - Enzyme activity finding
    - Fetal finding
    - Findings by method

**Right Panel: Query Tool**

Query Name: Edema@12:22:55

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
Edema								

one or more of these      AND      drop a term on here

Run Query   Clear      1 Group      New Group

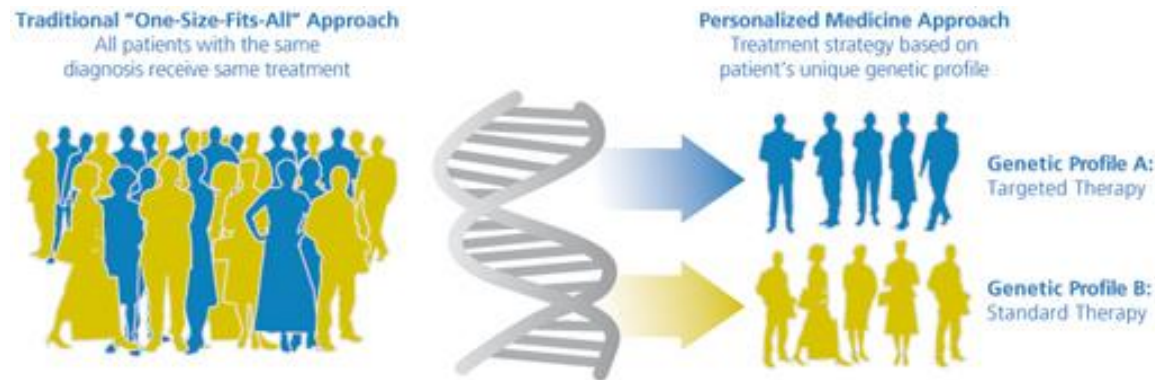
Show Query Status   Graph Results   Query Report

Number of patients

**384**

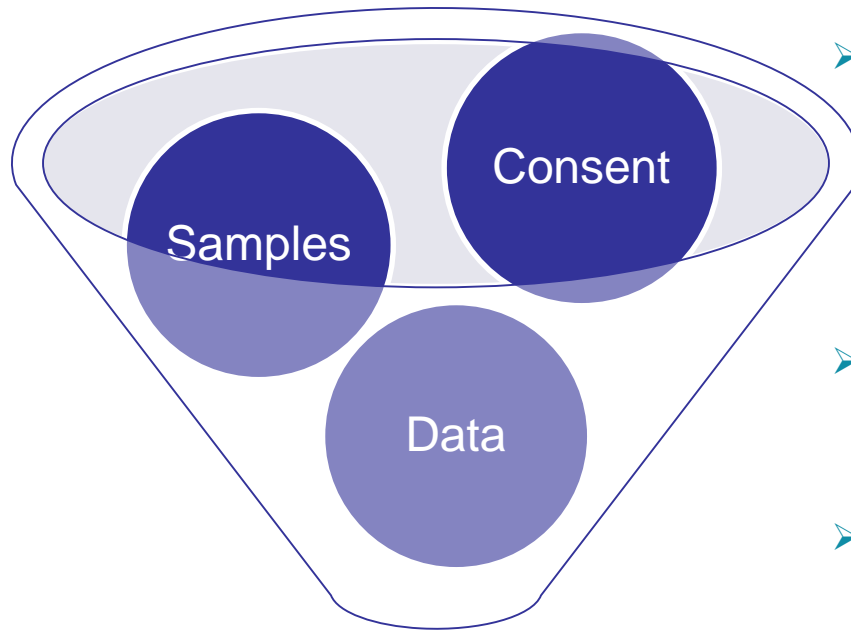
For Query "Edema@12:22:55"

# 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

# The Partners Biobank



Research Discoveries

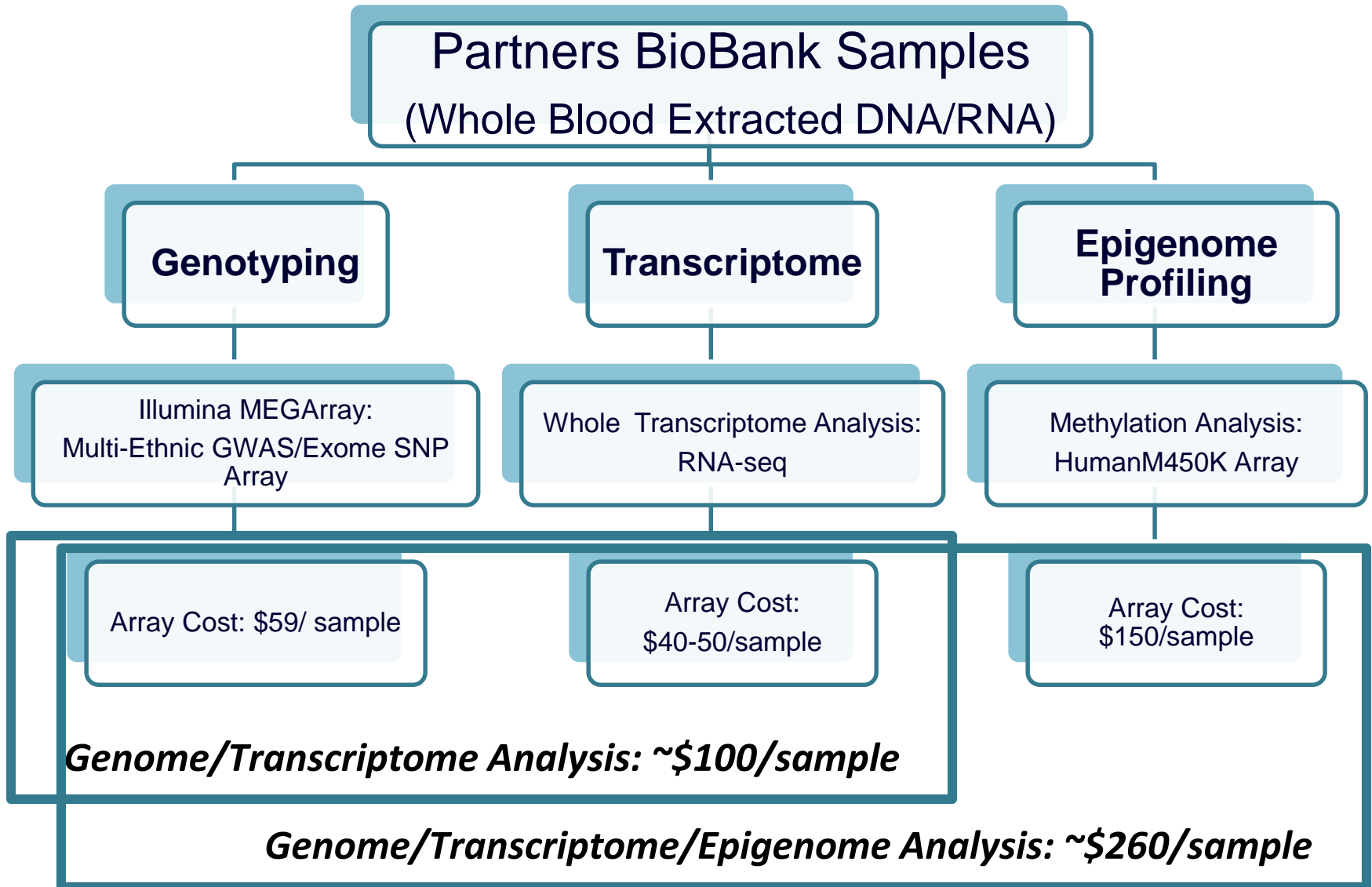
**Improved Clinical Care for All Patients**

- 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).*



# Biobank Integrative Genomics Strategy



# Genotype Data

**Navigate Terms** Find

- Biobank Consent Information
- Biobank Demographics
- Biobank Genomics
  - dbSNP rs Identifier - 10097
  - Gene - 10097**
  - People with genomic data - 10097
- Biobank Health Information Survey
- Biobank Sample Types
- Curated Disease Populations
- Healthcare Data
- Healthy Populations (Controls)

**Query Tool**

Query Name:

Temporal Constraint

Group 1

Dates	Occurs > 0x	Exclude	Da
Treat Independently			
Gene - 10097			

drop a term on here

**Search by Gene**

Use the gene name box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first characters.

Gene Name\*: APO

Please note the zygosity to query for patients without a particular variant (nucleotide on the right)

Zygosity\*: APOB

Consequence: APOB

APOBEC1  
APOBEC2  
APOBEC3A  
APOBEC3A\_B  
APOBEC3B

OK Cancel

**Navigate Terms** Find

- Biobank Consent Information
- Biobank Demographics
- Biobank Genomics
  - dbSNP rs Identifier - 10097**
  - Gene - 10097
  - People with genomic data - 10097
- Biobank Health Information Survey
- Biobank Sample Types
- Curated Disease Populations
- Healthcare Data
- Healthy Populations (Controls)

**Query Tool**

Query Name:

Temporal Constraint

Group 1

Dates	Occurs > 0x	Exclude	Da
Treat Independently			
dbSNP rs Identifier - 10097			

drop a term on here

**Search by dbSNP rs Identifier**

Use the rs identifier box to specify the variant for which to search. When you begin typing in the search box below, a selection list will appear after you type the first three numbers.

rs identifier\*: rs1234

Please note the zygosity to query for patients without a particular variant (nucleotide on the right)

Zygosity\*: rs12340033 | C to G

rs12340061 | G to A  
rs12340067 | C to T  
rs12340088 | T to G  
rs12340105 | A to C  
rs12340107 | G to T  
rs12340117 | G to A  
rs12340120 | G to A  
rs12340129 | A to G  
rs12340149 | G to A  
rs12340158 | T to A

OK Cancel

drop a term on here

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

Partners Biobank Portal

File Edit View History Bookmarks Tools Help

https://biobankportal.partners.org/4-0/?user=snm0

Partners HEALTHCARE 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) - 36024
- 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
Exclude	Exclude
Dates	Dates
Occurs > 0x	Occurs > 0x
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	AND
one or more of these	one or more of these

Run Query Clear 3 Groups

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

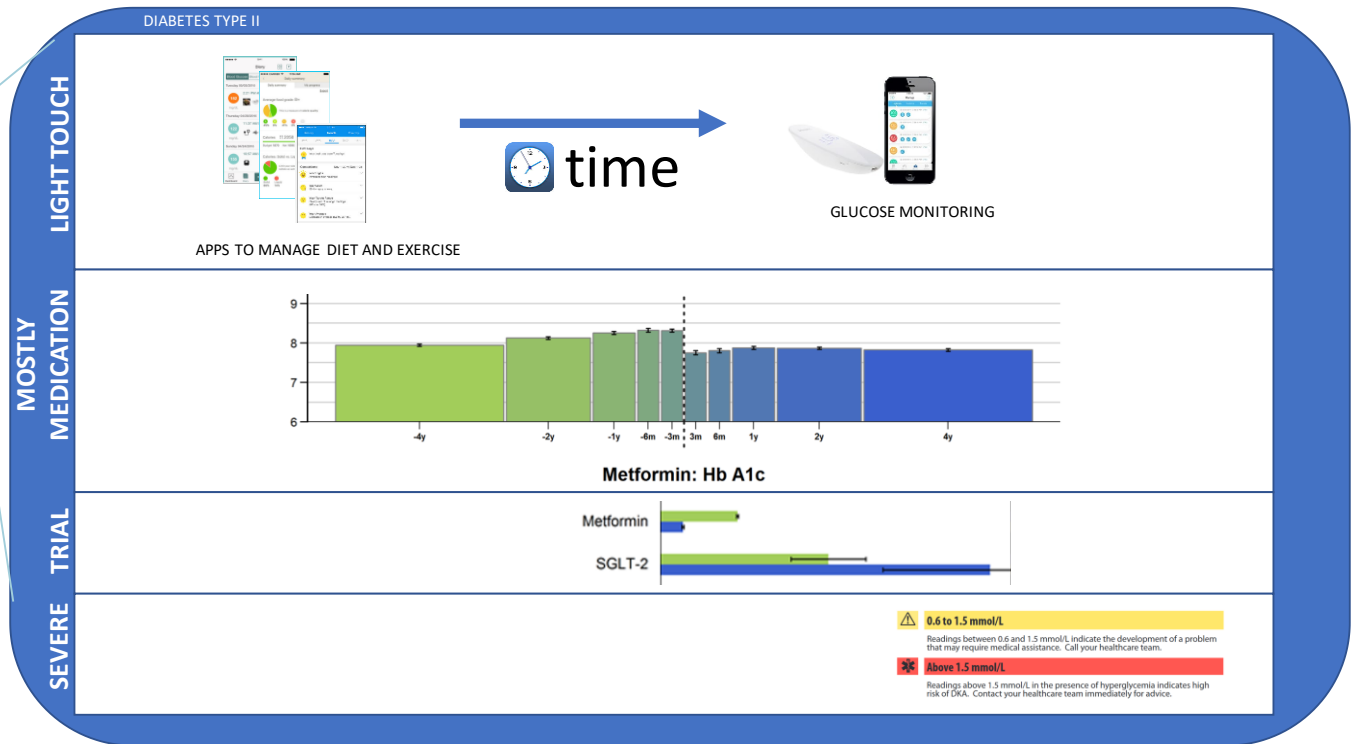
Followed for:

Prevention

Diabetes Type II

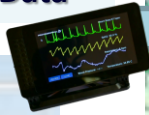
Hypertension

Heart Failure



# Innovations Enabling Digital Care

Digital and IoT devices continuously output Patient Data

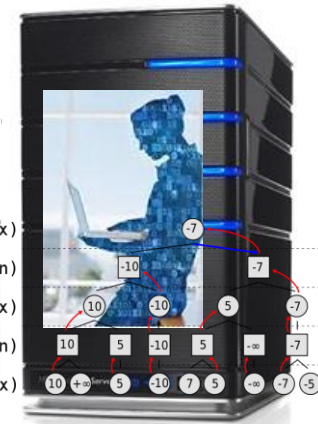


```

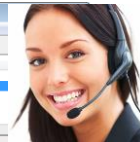
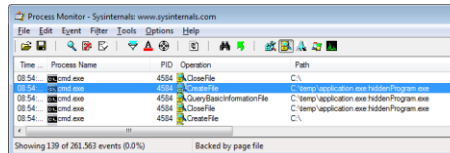
1. for ( i = 1; i ≤ a1.length; i++ )
  1.1. j = 1
  1.2. while ( a1[i] != a2[j] )
    if ( j ≥ a1.length )
      .. return false
    .. j++
  1.3. tmp = a2[j]
  1.4. a2[j] = a2[i]
  1.5. a2[i] = tmp
2. return true
    
```

```

cost
C1 n + 1
C2 n
C3 ∑_{j=1}^i (j-1)
C4 ∑_{j=1}^i (j-1)
C5 n
C6 n
C7 n
C8 n
C9 n
C10 1
    
```



# PATIENT



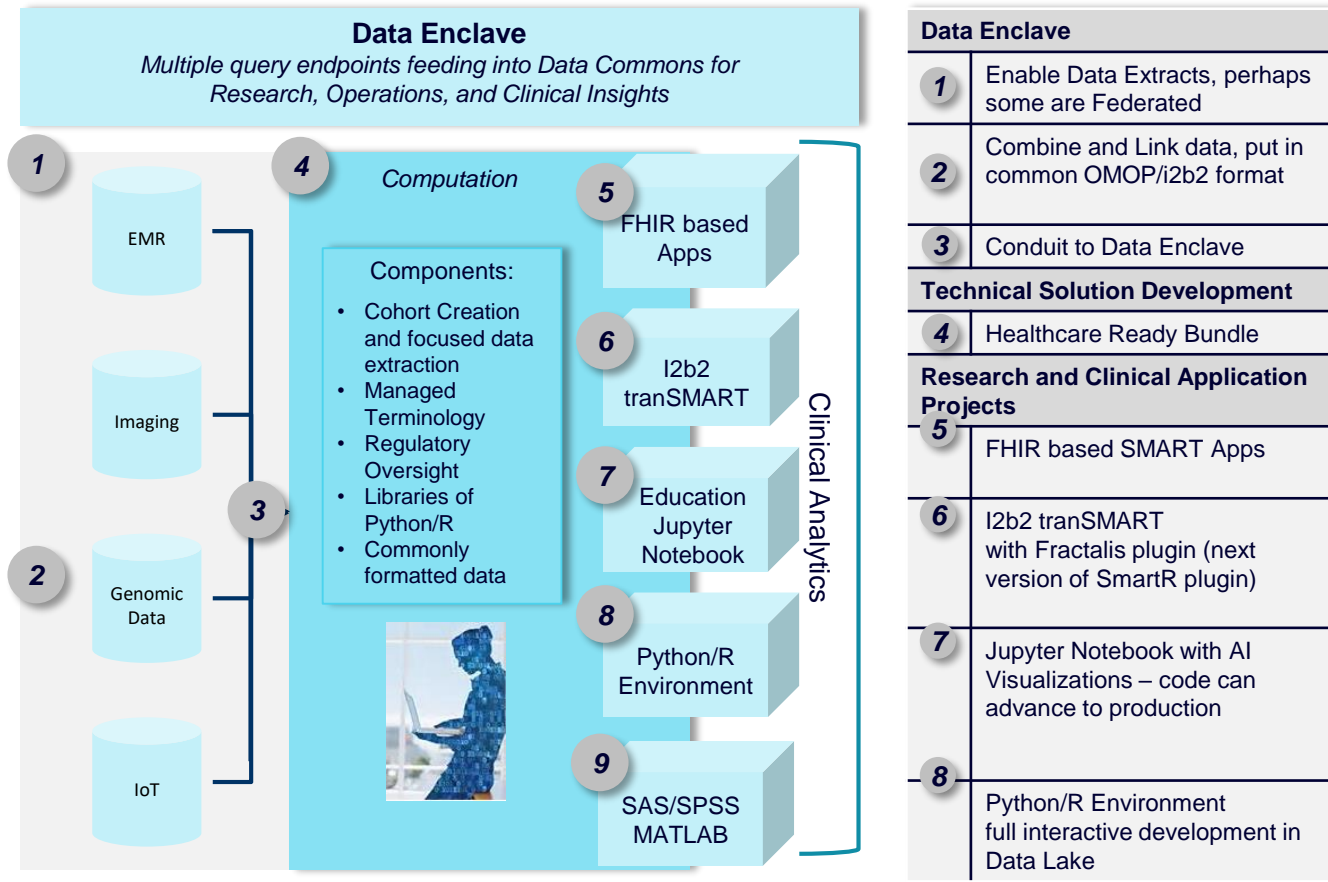
Navigator Model dramatically increases Frequency and Convenience for Patient Communication

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



System drives Pragmatic Clinical Trials Leading to Continuous Process Improvement

# Digital Twin for Continuously Assessing Patient



# 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 Waghlikar
  - 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 Waghlikar
- Wayne Chan
- Bhaswati Ghosh
- Reeta Metta
- Barbara Benoit
- David Wang
- Jeff Klann



THANK YOU