Integrating Genomic, Imaging, and Clinical Data for Precision Medicine

Shawn Murphy MD, Ph.D.
Partners Healthcare and Harvard Medical School
November 3rd, 2017
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 provides samples (plasma, serum, and DNA) collected from consented patients.

- 64,000 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).

Research Discoveries

Improved Clinical Care for All Patients
Biobank Integrative Genomics Strategy

Partners BioBank Samples
(Whole Blood Extracted DNA/RNA)

Genotyping
- Illumina MEGArray:
  Multi-Ethnic GWAS/Exome SNP Array
  Array Cost: $59/sample

Transcriptome
- Whole Transcriptome Analysis:
  RNA-seq
  Array Cost: $40-50/sample

Epigenome Profiling
- Methylation Analysis:
  HumanM450K Array
  Array Cost: $150/sample

Genome/Transcriptome Analysis: ~$100/sample

Genome/Transcriptome/Epigenome Analysis: ~$260/sample
Partners Personalized Medicine Components

- **Partners Biobank**
  - DNA, Plasma & Serum from 47,000 Consented Patients
  - Sample Processing Services

- **Translational Genomics CORE**
  - Sequencing
  - Genotyping
  - Microarray
  - RNA Seq.

- **Laboratory for Molecular Medicine**
  - Sanger Sequencing
  - Targeted Next Gen Seq
  - Whole Genome Seq (WGS)
  - Medical Exome (WES)

- **Phenotype Core/Research Patient Data Repository**
  - 6.7 million Partners patients
  - Data from EMR
  - Additional research data
  - Validated phenotypes/controls
  - Genomic Data Repository (GDR)

- **PPM Information Technology /IT, Bioinformatics, Research, Administrative**
Use Phenotyping Algorithms to define cohorts of treatment-resistant and treatment-responsive depression

Need to Determine: Depressed or Well at Encounter

Must Improve Accuracy of Diagnoses from Electronic Health Record

<table>
<thead>
<tr>
<th>Clinical Status</th>
<th>Model</th>
<th>Specificity</th>
<th>Sensitivity</th>
<th>Precision</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depressed</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.09 (0.03)</td>
<td>0.57 (0.14)</td>
<td>0.54 (0.02)</td>
</tr>
<tr>
<td></td>
<td>NLP</td>
<td>0.95</td>
<td>0.42 (0.05)</td>
<td>0.78 (0.02)</td>
<td>0.88 (0.02)</td>
</tr>
<tr>
<td></td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.06)</td>
<td>0.78 (0.02)</td>
<td>0.87 (0.02)</td>
</tr>
<tr>
<td>Well</td>
<td>Billing Codes</td>
<td>0.95</td>
<td>0.06 (0.02)</td>
<td>0.26 (0.27)</td>
<td>0.55 (0.03)</td>
</tr>
<tr>
<td></td>
<td>NLP</td>
<td>0.95</td>
<td>0.37 (0.06)</td>
<td>0.86 (0.02)</td>
<td>0.85 (0.02)</td>
</tr>
<tr>
<td></td>
<td>NLP + Billing Codes</td>
<td>0.95</td>
<td>0.39 (0.07)</td>
<td>0.85 (0.02)</td>
<td>0.86 (0.02)</td>
</tr>
</tbody>
</table>

Initially: AUC = 0.54
Finally: AUC = 0.87

Initially: AUC = 0.55
Finally: AUC = 0.86
Data Integration in Biobank Portal

Electronic Medical Record (EMR) Data
- RPDR
  - Coded Data
    - Demographics
    - Diagnoses
    - Lab Results
  - Text Data (Notes/Reports)
    - Medications
    - Procedures
    - Visits
    - Physician Notes
    - Imaging Reports
    - Pathology Reports
    - Surgery Notes

Informatics Tools
- Calculated Controls (Charlson Index)
- Data Visualization
- Data Queries
- Annotation
- Extract Data
- Natural Language Processing

Additional Data
- Other Research Data
- Survey Data

Genetic Data
- GWAS

Biobank Data
- Samples
  - DNA
  - Serum
  - Plasma
- Consent
  - Recontact
  - Consent Status

Validated Phenotypes
- Type II Diabetes
- Coronary Artery Disease
- Congestive Heart Failure
- Rheumatoid Arthritis
- IBD
- Multiple Sclerosis
- Bipolar Disorder

Research
Curating a Disease Algorithm

1. **Create a gold standard training set.**

2. **Create a comprehensive list of features from patient’s electronic data that describe the disease of interest.**

3. **Develop the classification algorithm.** Using the data analysis file and the training set from step 1, assess the frequency of each variable. Remove variables with low prevalence. Apply adaptive LASSO penalized logistic regression to identify highly predictive variables for the algorithm.

4. **Apply the algorithm to all subjects** in the superset and assign each subject a probability of having the phenotype.
White matter abnormalities associated with treatment-resistant depression

- Scans collected as part of routine clinical care
- Diffusion tensor imaging in 150 pts
- Age-related decline in white matter integrity increases with treatment resistant depression

Medial fornix shows strongest effect

## Biobank Portal | Curated Diseases

<table>
<thead>
<tr>
<th>Validated Phenotype</th>
<th>Count*</th>
<th>Predictive Positive Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bipolar Disease</td>
<td>71</td>
<td>89%</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>387</td>
<td>90%</td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>2,420</td>
<td>97%</td>
</tr>
<tr>
<td>Crohn's Disease</td>
<td>453</td>
<td>90%</td>
</tr>
<tr>
<td>Multiple Sclerosis</td>
<td>94</td>
<td>90%</td>
</tr>
<tr>
<td>Rheumatoid Arthritis</td>
<td>550</td>
<td>90%</td>
</tr>
<tr>
<td>Type 2 Diabetes Mellitus</td>
<td>1,887</td>
<td>97%</td>
</tr>
<tr>
<td>Ulcerative Colitis</td>
<td>330</td>
<td>90%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Healthy Controls based on Charlson Index</th>
<th>Count**</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 10-year survival probability is &gt;98.3%</td>
<td>2,206</td>
</tr>
<tr>
<td>1 – 10-year survival probability is &gt;95.87%</td>
<td>4,343</td>
</tr>
<tr>
<td>2 – 10-year survival probability is &gt;90.15%</td>
<td>6,545</td>
</tr>
</tbody>
</table>

* Based on 15,880 patients
** Based on 21,300 patients
High Quality Phenotypes available for Genetic Studies
Genotype Data

- 3349 SNP or indels
- 1680 Homozygous
- 1336 subjects with protein altering (frameshift, missense, nonsense, start loss, stop loss) variant

TTN
Partners Biobank Portal – Request Genetic Data

Biobank Portal Genomics Request

The Biobank contains subjects who have consented to make their genomic data available for research. To request genomic data, please fill out the form below.

Once you have submitted this form, you will be contacted by the Biobank team to complete your request. Please note that no genomic data will be sent out without a discussion with the Biobank team.

For more assistance on making a genomic data request, please refer to the Biobank Wiki section “How to Make Genomic Requests Using the Biobank Portal” or contact us.

Step 1. Patient Selection:

Choose Your Patients: Enter a list of Biobank Subject IDs OR Select a Previous Query

Your Patient List:
Type or paste a list of Biobank Subject IDs, one per line. You can find Subject IDs by downloading the de-identified data set. Do not enter MRNs.

10012040
10022384
10016324
10016243
10016229
10025693

Step 2. Details of Request:

Request Type: De-Identified OR Identified

Contact Name: Wattanasin, Nich

Contact E-mail: nwattanasin@partners.org
Common Sequence for Research Data

• 1) A researcher creates a registry of patients
• 2) Data is collected on the patients
  • Abstracted from clinical chart as summary data and imaging
  • Questionnaires are given and/or Interviews with patients are performed
• 3) Data is analyzed and published
• 4) Data is Archived

• Opportunity is lost – new research may wish to combine the data with new clinical data and data from other registries
Valuable Data

- Registry Data collected at great time and expense by departments often for reporting purposes
  - ~30 Registries of this type
- Imaging Data collected from a specialized cohort of patients
  - ~20 Collections of this type
- Genomic Data collected from a specialized cohort of patients
  - ~10 Collections of this type (besides the Biobank)
- Studies that collected data through REDCap
  - ~500 of these studies
Wish to “Publish” these data in “Portals”

• Easily assessable to researchers at Partners
• Researchers can instantly get a “feel” for the data
• Researchers can do querys similar to the RPDR
• Data is linked to clinical data and data throughout the Big Data Commons
• Researchers can look at data in tables and on timelines
• Researchers can get detailed data
• Adheres to HIPAA and Common Rule regulation
• There are a few of these appearing, but why not more, why is this so hard?
  • MIT: http://www.cerebrovascularportal.org/
Wish to overcome difficulty linking study data to Clinical Data

• Current Process to create a set of patients from the registry, import into RPDR, and then have data returned as detailed data sets, analyze data set with SAS, Excel, and other local tools
Biobank Portal is a good model for these Portals, but need more:

Perform Queries  Obtain Summary Tables  Link to Detailed Data
Two Strategies to Link Clinical Data to Registries Enabled by Big Data Commons

• Enterprise centric –
  • When regulation over data is such that a data use agreement is not required.
  • When registry data is not frequently updated.
  • When there is no “special” data like images or waveform data to be presented.

• Registry centric
  • When regulation over data requires a data use agreement
  • When data is frequently updated
  • When there are special displays and features of the data requiring specialized viewers.
An Enterprise Centered Data Network

**Genomic Data**
Genomic data collected through the Biobank lives in a separate repository, but is made available for connecting with clinical data. All patients within the Biobank are accessible.

**Research Repository**
Broad repository of clinical data made available for research is the center point for all querying. Contains the entire Partners patient population. This is the RPDR.

**Imaging Repository**
DICOM Metadata is extracted from images downloaded from mi2b2. This may be supplemented by a limited amount of tags on all images given to us by Radiology group. Contains references to all patients from who we have imaging data.

**Notes & Reports**
Notes and reports on all patients are collected and put into a separate data repository that can be full text indexed. Specific security precautions are used to limit the PHI that can be queried directly.

**Project Registry**
Individual research groups may contribute their data or findings back to the Partners enterprise for querying and use by all researchers across the organization. Data is used for the greater good.
A Registry Centered Data Network

**Genomic Data**
Genomic data collected through the Biobank lives in a separate repository, but is made available for connecting with clinical data. Only patients contained in the project registry can be queried within this network.

**Genomic Data**

**Project Registry**
The central access point for this type of data network is the project specific registry. All queries will be limited to the patients that are part of this project.

**Imaging Repository**
DICOM Metadata is extracted from images downloaded from mi2b2. This may be supplemented by a limited amount of tags on all images given to us by Radiology group. Contains references to all patients from who we have imaging data. Only patients contained in the project registry can be queried within this network.

**Clinical Data**
Repository that contains most clinical data from legacy systems as well as Epic for all patients across the enterprise. Only patients contained in the project registry can be queried within this network.

**Notes & Reports**
Notes and reports on all patients are collected and put into a separate data repository that can be full text indexed. Specific security precautions are used to limit the PHI that can be queried directly. Only patients contained in the project registry can be queried within this network.
From Registry

From Biobank
Partners Big Data Commons launches the Clinical Image Bank

The Partners Big Data Commons has launched the Clinical Image Bank portal, which allows researchers to access expertly curated, phenotypically characterized medical images, along with clinical data extracted from Partners electronic health records.

The first registry to be made available is a repository of pediatric brain magnetic resonance images comprised of a cohort of neonates with clinically confirmed hypoxic ischemic encephalopathy (HIE), along with a cohort of “normative” children imaged between the ages of 0-6 years of age with no known CNS pathology.
Register

Please enter your Partners user ID and password to proceed to the registration form.

Username

Password

Continue
Sign the DUA
Login

Please enter your Partners user ID and password:

Username

Password

Launch
Run a Query
Create a Table

[Image of a screenshot from a clinical image bank software, showing a Patient Set Viewer interface with options for selecting patients and aggregating data.]
Detailed Review of Patients

<table>
<thead>
<tr>
<th>Patient Number</th>
<th>Gender</th>
<th>Age</th>
<th>Race</th>
<th>4. Birth head circumference [List of All Values]</th>
<th>5. Cohort [All Concepts [Names/Text]]</th>
<th>Acquisition [Images [Yes/No]]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>13</td>
<td>White</td>
<td>[37.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>13</td>
<td>White</td>
<td>[30]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>13</td>
<td>White</td>
<td>[32.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>13</td>
<td>White</td>
<td></td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>13</td>
<td>Unknown</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>12</td>
<td>Black</td>
<td>[36.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>13</td>
<td>White</td>
<td>[33]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>12</td>
<td>White</td>
<td>[31]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>12</td>
<td>White</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>12</td>
<td>White</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>12</td>
<td>Black</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>10</td>
<td>Asian</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>9</td>
<td>White</td>
<td>[36]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
<td>View</td>
</tr>
</tbody>
</table>
View Image Timeline
View Images on Timeline
Download Images

The Clinical Image Bank offers downloadable image packages depending on your need. Identified images are served in DICOM format and limited data sets are served in NIfTI format.

After your request is processed, your images will be made available on either the ERISOne Linux Computing Cluster or the HPCWIN3 Windows Analysis Server.

Please complete the form below:

- **Request Type**: NIfTI Images (Limited Data Set) / DICOM Images (Identified)

- **Distribution Type**: I Don’t Know / Linux / Windows

- **Image Package**: All available NIfTI images (for 210 patients)

- **Contact Name**:

- **Contact E-mail**:

- **Contact Phone**: 123-456-7890

- **Principal Investigator Name**:

- **Name of Study**:

- **Description**: Please enter any other relevant details about your request.
<table>
<thead>
<tr>
<th>Gender</th>
<th>Age</th>
<th>Race</th>
<th>Birth head circumference</th>
<th>Cohort [All Concepts (Names/Text)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>14</td>
<td>White</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)] [8. Other cohort type]</td>
</tr>
<tr>
<td>M</td>
<td>13</td>
<td>White</td>
<td>[30]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>White</td>
<td>[32.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>Unknown</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>M</td>
<td>12</td>
<td>Black</td>
<td>[36.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>13</td>
<td>White</td>
<td>[33]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>12</td>
<td>White</td>
<td>[31]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>M</td>
<td>12</td>
<td>Black</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>Asian</td>
<td>[34]</td>
<td>[3. Stroke]</td>
</tr>
<tr>
<td>M</td>
<td>9</td>
<td>White</td>
<td>[36]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>White</td>
<td>[31]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>White</td>
<td>[34.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>Unknown</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>M</td>
<td>7</td>
<td>Black</td>
<td>[35.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>White</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>M</td>
<td>10</td>
<td>Unknown</td>
<td>[37]</td>
<td>[8. Other cohort type]</td>
</tr>
<tr>
<td>M</td>
<td>6</td>
<td>Unknown</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>Other</td>
<td>[35]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>White</td>
<td>[34.5]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>6</td>
<td>Unknown</td>
<td>[34]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>F</td>
<td>5</td>
<td>Black</td>
<td>[32]</td>
<td>[2. Hypoxic Ischemic Encephalopathy (HIE)]</td>
</tr>
<tr>
<td>M</td>
<td>5</td>
<td>Black</td>
<td>[36]</td>
<td>[8. Other cohort type]</td>
</tr>
</tbody>
</table>
Find Normal MRI’s at All Ages 0-6 y/o

Number of patients who had a brain MRI scan at a particular age in months from 0 to 6 years (A) and in weeks from 0 to 4 months (B)
Determining a Normal Child’s MRI
Atlases provide a visual guide for Radiology Decision Support, such as determining Perinatal Hypoxic Ischemic Encephalopathy

ADC map from 4 infants: Each statistically compared to age matched atlas yields visual guide to pathology

Quantitative analysis tools + large data sets = Great insights for practicing doctors
Integration strategy from “Informatics for Integrating Biology and the Bedside (i2b2)” sponsored by the National Institutes of Health, what is it?

- Software for explicitly organizing and transforming person-oriented clinical data to a way that is optimized for clinical genomics research
  - Allows integration of clinical data, trials data, and genotypic data
- A portable and extensible application framework
  - Software is built in a modular pattern that allows additions without disturbing core parts
  - Available as open source at https://www.i2b2.org
Interrogation can occur through i2b2 web client.
I2b2 Community Software Modules contributed as “Cells”

New i2b2 Community Projects

I2b2 FHIR Cell
Built by Kavishwar Waghohlar, this i2b2 addition allows SMART cells to communicate with the i2b2 core using the Fast Healthcare Interoperability Resources.

C3-PRO FHIR Uploading Cell
C3-PRO Research Framework is an iOS framework written in Swift. Combining FHIR and ResearchKit, usually for data storage into i2b2, this framework allows you to use FHIR Questionnaire resources directly with ResearchKit and will return FHIR QuestionnaireResponse that you can send to your server. In addition, a FHIR Contract resource can be used to carry trial eligibility requirements and define content to be shown during consenting. Subsequently, the contract can be “signed” with a FHIR Patient resource and returned to your server, indicating consent.

Current i2b2 Community Projects

Loyalty Cohorts
Because electronic health records are often missing information about patients, we developed and validated a tunable computer algorithm to identify subsets of patients whose data are relatively complete and therefore better suited for clinical research studies.

Workplace Items Sharing Enhancement
A collection of webclient plugins that enhance the sharing of items within the Workplace pane

Ontology Tools
Tools to extract and organize ontologies. The tools are organized by Lori Phillips. Recent additions is a library of ontologies which can be downloaded using the i2b2 workbench.
i2b2 Cell: The Canonical Software Module

HTTP XML
(minimum: RESTful)
An i2b2 Environment (the Hive) is built from i2b2 Cells

“Hive” of software services provided by i2b2 cells
I2b2 Software components are distributed as open source
Implementations

CTSA’s

- Boston University
- Case Western Reserve University (including Cleveland Clinic)
- Children's National Medical Center (GWU), Washington D.C.
- Duke University
- Emory University (including Morehouse School of Medicine and Georgia Tech)
- Harvard University (including Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Dana Farber Cancer Center, Joslin Diabetes Center, Massachusetts General Hospital)
- Medical University of South Carolina
- Medical College of Wisconsin
- Oregon Health & Science University
- Penn State Milton S. Hershey Medical Center
- Tufts University
- University of Alabama at Birmingham
- University of Arkansas for Medical Sciences
- University of California Davis
- University of California, Irvine
- University of California, Los Angeles*
- University of California, San Diego*
- University of Chicago
- University of Cincinnati (including Cincinnati Children's Hospital Medical Center)
- University of Colorado Denver (including Children's Hospital Colorado)
- University of Florida
- University of Kansas Medical Center
- University of Kentucky Research Foundation
- University of Massachusetts Medical School, Worcester
- University of Michigan
- University of Pennsylvania (including Children's Hospital of Philadelphia)
- University of Pittsburgh (including their Cancer Institute)
- University of Rochester School of Medicine and Dentistry
- University of Texas Health Sciences Center at Houston
- University of Texas Health Sciences Center at San Antonio
- University of Texas Medical Branch (Galveston)
- University of Texas Southwestern Medical Center at Dallas
- University of Utah
- University of Washington
- University of Wisconsin - Madison (including Marshfield Clinic)
- Virginia Commonwealth University
- Weill Cornell Medical College

Academic Health Centers (does not include AHCs that are part of a CTSA):

- Arizona State University
- City of Hope, Los Angeles
- Georgia Health Sciences University, Augusta
- Hartford Hospital, CN
- HealthShare Montana
- Massachusetts Veterans Epidemiology Research and Information Center (MAVERICK), Boston
- Nemours
- Phoenix Children's Hospital
- Regenstrief Institute
- Thomas Jefferson University
- University of Connecticut Health Center
- University of Missouri School of Medicine
- University of Tennessee Health Sciences Center
- Wake Forest University Baptist Medical Center

HMOs:

- Group Health Cooperative
- Kaiser Permanente

International:

- Georges Pompidou Hospital, Paris, France
- Hospital of the Free University of Brussels, Belgium
- Inserm U936, Rennes, France
- Institute for Data Technology and Informatics (IDI), NTNU, Norway
- Institute for Molecular Medicine Finland (FIMM)
- Karolinska Institute, Sweden
- Landspitali University Hospital, Reykjavik, Iceland
- Tokyo Medical and Dental University, Japan
- University of Bordeaux Segalen, France
- University of Erlangen-Nuremberg, Germany
- University of Goettingen, Goettingen, Germany
- University of Pavia, Pavia, Italy
- University of Seoul, Seoul, Korea

Companies:

- Johnson and Johnson (TransMART)
- GE Healthcare Clinical Data Services
Federated Queries in PCORNet

- Partners HealthCare System
- Boston Children’s Hospital
- BIDMC
- Boston Health Net (BMC and Community Health Centers)
- Columbia U. Medical Center and New York Presbyterian Hospital
- Wake Forest Baptist Medical Center
- Morehouse/Grady/RCMI
- University of California, Davis
- Washington University in St. Louis
- U Texas Health Science Center/Houston
Upcoming Project Releases

- Incorporation of FHIR into i2b2
  - FHIR Cell to allow single patient’s data to be returned in FHIR
  - FHIR Ontology and ETL to allow direct import of FHIR into i2b2
  - FHIR Cell to allow i2b2 to extend to FHIR query endpoints

- i2b2-based system for Accruing Patients for Clinical Trials
  - Extensions to SHRINE for management of clinical trials
  - Web Client Plug-ins to extend SHRINE queries to local i2b2-based patient recruitment

- i2b2-based system to query and return data from Observation-Fact tables contained in multiple different i2b2 Hives
No Small Change for the Health Information Economy

Kenneth D. Mandell, M.D., M.P.H., and Isaac S. Kohane, M.D., Ph.D.

The economic stimulus package signed by President Barack Obama on February 17 included a $19 billion investment in health information technology. How can we best take advantage of this unprecedented opportunity to improve patient care and stimulate the health information economy while also stimulating the U.S. economy? A health care system adapting to the effects of an aging population, growing expenditures, and diminishing primary care workforce needs the support of a flexible information infrastructure that facilitates innovation in wellness, health care, and public health.

Health care is critical, since the system will have to function under new policies and in the service of new health care delivery mechanisms, and it will need to incorporate emerging information technologies on an ongoing basis. As we seek to design a system that will constantly evolve and encourage innovation, we can glean lessons from large-scale information-technology successes in other fields. An essential first lesson is that ideally, system components should be not only interoperable but also substitutable.

The Apple iPhone, for example, uses a software platform with a published interface that allows software developers outside Apple to create applications. There are now nearly 10,000 applications that consumers can download and use with the common phone interface. The platform separates the system from the functional-
1 SMART App in 3 SMART Systems
What Big Data can do for the Everyday Clinician
Finding Similar Patients

• Looking at similar patients can help predict:
  • Future outcomes and responses to therapy
  • Course of disease
  • Penetrance of genetic variants
  • Likelihood that a diagnostic pathway might be fruitful

• Big Data Commons is an opportunity for combining data from the Electronic Health Record, Specialized Health Databases, Analytics from Big Data Queries, and presentation in SMART Apps

• Presentation of results can be greatly enhanced with engaging visualizations for the provider making difficult, complex decisions
Bringing Big Data into Clinical Care with Open App Development

Clinician

Core Integration Database

FHIR interface for real-time updates

Epic Data Repository

DATA
- GenelInsight
- mHealth
- ePath
- Medical Images
- 25 years of Legacy electronic data
- Other External Systems

SMART App embedded in Epic

Analytic Calculation Engine

Analytics have direct access to repository
FULL SUPPLY CHAIN CONTROL

INTEGRATION ENVIRONMENT

SMART Apps

CLINICAL ENVIRONMENT

Clinical Data

RESEARCH ENVIRONMENT

AI Applications

STRUCTURED DATA

COMPUTATIONAL ENVIRONMENT
Tribute to…

- I2b2 Core Team
  - Isaac Kohane
  - Susanne Churchill
  - Michael Mendis
  - Christopher Herrick
  - Griffin Weber
  - Paul Avillach
  - Lori Phillips
  - Nich Wattanasin
  - David Wang
  - Vivian Gainer
  - Victor Castro
  - Andrew Cagan
  - Wayne Chan

- SMART Apps Team
  - Calum MacRae
  - Sandy Aronson
  - Mike Oats
  - Layne Ainsworth
  - Kenneth Mandl
  - Joshua Mandel

- Radiology (mi2b2) Team
  - Randy Gollub
  - P Ellen Grant
  - Kathy Andriole
  - Kallirroi Retzepi
  - Rudolph Pienaar
  - Lilla Zollei
  - Yangming Ou
I2b2, SHRINE, and SMART
Information and Software
on the Web

i2b2 Homepage (https://www.i2b2.org)
i2b2 Software (https://www.i2b2.org/software)
i2b2 Community Site (https://community.i2b2.org)
SMART Platforms Homepage (http://smarthealthit.org)

Partners Healthcare, NIH/NCBC/BD2K; /NIMH; /NCATS; /NIBIB; /NHGRI

NIH R01 EB014947          NIH U54 HG007963
NIH U54 LM008748          NIH R01 AT006364
NIH U01 HG008685          NIH R01 AT005280
PCORI 282364.5077585.0007 NIH P01 AT006663
THANK YOU