2014 ACTS Meeting

Translational Informatics: From PCORI to NCI to BD2K – The Pittsburgh Roadmap for Unlocking the Value of Health Data from EHRs

Michael J. Becich, MD PhD (presenter)
Chairman, Department of Biomedical Informatics (DBMI)
University of Pittsburgh School of Medicine (UPSoM)
Rebecca Crowley, MD MSIS (CIO Personalized Medicine, DBMI)
Greg Cooper, MD PhD (Vice Chair, DBMI)

PCORI – PaTH towards a Learning Health System for the Mid-Atlantic Region
Learning Objectives

To understand the opportunity health data (from EHRs) can play in clinical and translational research

To understand how data models, deep phenotyping and causal modeling and discovery play critical roles in the effective secondary use of EHR data in clinical and translational science

To introduce innovations from the Pitt Department of Biomedical Informatics and how we are aiming to address biospecimen sharing, deep phenotyping and “causal modeling/analytics” of health care data for personalized medicine
Disclosures of COI for 2014 for MJB

- Corporate Support for API, Strategic Summit and Pathology Informatics 2014

- Corporate Sponsored Research – ZERO (1st time in 15 years!!!)

- Startup/Public Companies (Consulting, Royalties/Licensing, Stock - MJB):
  - De-ID Data Corp – de-identification software (licensing agreement) http://www.de-idata.com/
  - Empire Genomics - Scientific Advisory Board (http://www.empiregenomics.com)
  - Omnyx – Joint Venture with UPMC and GE (http://www.omnyx.com)
  - NinePoint Medical – Scientific Advisory Board (www.ninepointmedical.com)

- Consultancy (honoraria)
  - Cancer Center Consulting – MD Anderson, Karmanos Cancer Center, Moffitt Cancer Center, NFGC, Penn State CC, Roswell Park Cancer Institute, UMDNJ, U Colorado, VCU
  - CTSA Consulting – Duke, Emory, MCW, Northwestern, UAK, UC Davis, UCLA, U Chicago, U Cincinnati, U IN, U KY, UC Davis, UMich, UMN, UNC, UNM, UWI and Wash U
  - Pathology – Roswell Park Cancer Institute
What is PaTH? A consortium of clinical research sites sharing EHR data via i2b2/SHRINE and using a common data model based on common data elements (CDEs).
PaTH Partners and Funding

- PaTH = Pittsburgh (UPMC/Pitt), Penn State and Temple, Hopkins

- Partner organizations:
  - University of Pittsburgh School of Medicine and UPMC
  - Penn State College of Medicine/Hershey Medical Center
  - Temple University School of Medicine/Temple Health
  - Johns Hopkins University/Johns Hopkins Health System/Johns Hopkins Health Care

- One of eleven funded sites - $6.84M for 1/1/14 to 6/30/15

- Other funded networks include approximately $70M in funding for:
  - Harvard led SCILHS network
  - NYC Consortium
  - Greater Plains Collaborative led by KUMC
  - Vanderbilt led – Mid-South Consortium with Greenaway Health
  - UCSD led UC-Systems plus VA - pSCANNER
  - Chicago Consortium – CAPriCORN
  - Four other sites in Oregon, Kaiser CA, Louisiana and Children’s Hospitals
Technology Approach at Pitt and Across PaTH

- **Health Information System**
- **Enterprise Data Warehouse**

Extraction, Transformation, and Loading (ETL)
- **Common Data Elements**
- **De-identification**

PaTH i2b2
- I2b2 data warehouse
- **Analytic tools**

---

**PaTH Network Overview**

- JHU i2b2
- PSU i2b2
- Temple i2b2
- UPMC i2b2

Connected Using i2b2 and SHRINE+

- **PopMedNet Queries**
- **PCORINet**
<table>
<thead>
<tr>
<th>First Tier</th>
<th>Second Tier</th>
<th>Third Tier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Month of birth</td>
<td>Radiographic Imaging</td>
<td>Echocardiography</td>
</tr>
<tr>
<td>Year of birth</td>
<td>Oxygen Therapy</td>
<td>Six Minutes Walk Test (6MWT) or (SMW)</td>
</tr>
<tr>
<td>Gender</td>
<td>Vital Status</td>
<td>Advance Medical Directive</td>
</tr>
<tr>
<td>Race</td>
<td>Procedure</td>
<td>MRC Scale (Dyspnea evaluation)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td>Transplant Status</td>
</tr>
<tr>
<td>BMI (Body Mass Index)</td>
<td></td>
<td>Cardiac Cath</td>
</tr>
<tr>
<td>History of Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoke pack/year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pulmonary Function Test (PFT)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laboratory (LAB)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drugs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hospital Admission Status</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
NCI U24 ‘TIES Cancer Research Network’ or TCRN

- Funded in first round of NCI Informatics Program U Grants
- Builds on success of TIES natural language processing system for supporting translational research
- Extends system to develop a data sharing and tissue sharing network among four cancer centers
- Potential for developing a national collaborative cancer research network
TCRN U24 Specific Aims

Specific Aim 1. Enhance the informatics technology to support inter-institutional “trust”, paraffin registry development, tissue microarray (TMA) development, and nondestructive tissue use.

Specific Aim 2. Establish the TIES Cancer Research Network (TCRN) with four founding member institutions. Develop governance, network agreements, and policies for operating the TCRN.

Specific Aim 3. Recruit and support pilot scientific collaborations across the network, especially focused on personalized medicine.

Specific Aim 4. Disseminate the software and measure its impact.
TCRN Workflow

Institution B

1. TCRN Oversight Committee
2. Institution B Tissue Bank

4. Committee reviews documents and registers Institution B as data provider

5. Deidentification service
6. MTA process
7. Tissue Bank reviews documents
8. Determine acceptability of materials transfer

Updates from clinical system
Firewall
Clinical System
Accessed by honest brokers behind firewall

Deidentified Database
Identified Database
TIES system

Institution A

2. Account approval and creation
3. Request to access data and tissue
4. Tissues transferred
5. View data and create tissue requests
6. Deidentified Database
7. Identified Database

Institution A Tissue Bank

1. Investigator requests account

Deidentified Database
Identified Database
Clinical System
TIES system

information extraction v5
NCI U24 ‘Cancer Deep Phenotyping’ (CDP)

- Funded in second round of NCI Informatics Program U grants
- Collaboration between UPCI (Crowley) and Harvard Boston Children’s (Savova) with cross-UPCI collaboration (Co-Is Lee, Kirkwood, Day and Edwards)
- Open source software built on foundation of two mature products
- Develop new methods for extracting cancer phenotype information from electronic medical records using Natural Language Processing
- Focus on extracting variables needed for Personalized Medicine research teams and future decision support

PLUS = CDP for EHR
Personalized Medicine Requires Biomedical Informatics Science

A Research Data Warehouse (RDW) and robust Biorepository Supported by Biomedical Informatics Science are key enablers!!!

- This will require biomedical informatics expertise:
  - Expertise in **database design and query** – PCORnet and PaTH
  - **Natural language processing** of text (H&P, Consults, Discharge, Pathology Reports, etc..) – TIES/TCRN and CDP
  - Structured capture of key medical data will require controlled **vocabularies** and implementation of **ontologies** - TIES/TCRN & CDP
  - **De-identification** of text for sharing with researchers (De-ID Data Corp) – PCORnet/PaTh and TIES/TCRN & CDP
  - Implementation of **Bayesian algorithms** to make genomic data “actionable” via Causal Modeling and Discovery – BD2K – Center for Causal Modeling and Discovery – Application being reviewed today!!!

- Personalized Medicine critically requires **biorepositories***

*(U Pitt is #1 contributor to The Cancer Genome Atlas – see subsequent slides)
27 centers contribute to TCGA Working Groups as Tissue Providers and Clinical Experts

~50% of all qualified cases in TCGA come from NCI Comprehensive Cancer Centers

University of Pittsburgh:
# 1 contributor in breast & prostate
# 2 contributor in H/N & renal
# 3 contributor in melanoma

Co-authors on reports on breast, bladder, colorectal, endometrial, lung, ovarian and renal cancers in Nature (2011-2014)
Leveraging TCGA to Support Personalized Medicine at UPCI/UPMC

Pittsburgh Genome Resource Repository

**Regulatory Foundation**

- **dbGAP Data Use Certificate** provides access for broad collaboration by ~52 investigators at Pitt and UPMC.
- **Contracted Services Agreement** with PSC to host the TCGA Data.
- **Data Use Agreement** with UPMC to store and provide access to TCGA Data.

Pittsburgh Genome Resource Repository **IRB Protocol** approves mechanism for UPMC to link TCGA data to clinical data while providing only deidentified data to requesting researchers.

**The Cancer Genome Atlas (TCGA)**

- **TCGA data files from all institutions**
  - N = 7603 patients
  - 853 Tb Data as of 11/24/13
  - Uses TCGA provided phenotype data only

- **TCGA data files from UPMC patients**
  - N = 681 patients
  - 66 Tb Data as of 11/24/13
  - Merges omics data with rich UPMC clinical phenotype data

**Modified TCGA Roadmap**

- Used to update constantly changing TCGA files on a weekly basis.
Pitt Causal Modeling and Discovery Center

Led by Greg Cooper, MD PhD (Biomedical Informatics), Ivet Bahar, PhD (Computational and Systems Biology) and Jeremy Berg, PhD (Director of Institute for Personalized Medicine)

Theme: Modeling and discovery of causal networks from genome and phenome (patient EHR) biomedical datasets including those from clinical trials and other controlled biospecimen based efforts

Aims

- **Represent** causal knowledge within a unified, formal framework
- **Discover** causal knowledge from biomedical data (both observational and experimental) and background knowledge (e.g., from the literature) using efficient algorithms
- **Apply** causal knowledge to support browsing, answering causal queries, simulating causal processes, and designing experiments to resolve causal uncertainties

Driving Biological Problem areas: Signalling Pathways in Cancer/TCGA, Idiopathic Pulmonary Fibrosis (same as PaTH) and "Connectome" via fMRI brain imaging
Personalized Medicine and Causal Modeling

• Causal modeling can be applied via computational pathology to predict therapeutic response from genomics and EHR data
• Determining whether a treatment is working
• Producing safer drugs by predicting potential for adverse effects earlier
• Targeting groups of people most likely to benefit from a drug, while keeping its use from those who may be harmed by it
• Producing better medical outcomes
• Decreasing health care costs

Modified from: From Schwartz, CAP, 2009
Emerging Developments and Your Future in Pathology
PaTH - Towards a Learning Health System in the Mid-Atlantic Region

Waqas Amin¹, Fuchiang (Rich) Tsui¹, Charles Borromeo¹, Cynthia H. Chuang³, Jeremy U. Espino¹, Daniel Ford⁴, Wenke Hwang⁷, Wishwa Kapoor,² Harold Lehmann⁴, G. Daniel Martich⁵, Sally Morton⁸, Anuradha Paranjape⁶, William Shirey¹, Aaron Sorensen⁶, Michael J. Becich¹, Rachel Hess² and the PaTH network team*  

¹Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, PA  
²Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA  
³Department of Medicine and Public Health Sciences, Penn State College of Medicine, Hershey, PA  
⁴Department of Medicine, Division of Health Science Informatics John Hopkins School of Medicine, Baltimore, MD  
⁵Chief Medical Information Officer, UPMC, Department of Critical Care Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA  
⁶Department of Medicine, Temple University School of Medicine, Philadelphia, PA  
⁷Department of Public Health Sciences, Division of Health Services Research, Penn State College of Medicine, Hershey, PA  
⁸Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA
THANK YOU FOR ACTS INVITE

Questions?

NOTE: e-mail me at becich@pitt.edu for:
- Additional questions/clarifications
- PDFs of articles mentioned
- Copy of PPT

DBMI Web Site = http://www.dbmi.pitt.edu
Video Profile: https://www.youtube.com/watch?v=BZjexgkGK8