First signal:
• 1 year after Celecoxib
• 8 months after Rofecoxib

What is going on in healthcare? Past and now?

Oral Hypoglycemic Agents
Challenge: Efficiently Reach Large $N$ for Population studies

- High throughput genotyping
- High throughput phenotyping
- High throughput sample acquisition

DHHS Secretary’s Advisory Committee on Genetics, Health, and Society (SACGHS) argues for the health value of a 500,000 to 1M subject study. Estimated cost: $3,000,000,000

Free and Open Source Translational Toolkit: Implementations

Major Modes
EDGR Advantages

- Timeliness
- Clinical Relevance
- Underserved populations
- Controls
- Co-morbidity recognition (e.g. PheWAS)

Who?
Health Care Utilization
(Hospitalization, ED Visits)

NLP (and comedy) is not pretty
Crimson: Core Functions

- Mined Phenotypes
- Matched Anonymous ID
- Clinical discard

Richly annotated biospecimens

Murphy et al. Genome Research, 2009

Murphy et al. Genome Research, 2009

But it works...

Kurreeman, AJHG 2011

SNP
“Crowd sourcing” physician expertise

Predicting length of stay according to “normality” of labs in first day: > 75% accuracy

Timeline

EDGR Challenges

- Consent (None/Opt-in/Opt-Out)
- Cost of EHRs
- Quality of EHR data
- Lack of Family History codification
- Lack of EHR standardization
- Cultural gulf between clinical informatics and bioinformatics.
  - Translational Bioinformatics
How might we scale nationally?

What does SHRINE allow you to do?

• Search routine clinical records from 5 major hospitals for:
  – Demographics
  – Diagnosis
  – Medications
  – Lab Results
• Reach N
  – Rare Dx
  – Small Effects

Shared Health Research Informatics Network (SHRINE)

How does it work?

• Export data from hospital systems
• Import data into i2b2 (clinical data warehouse) for each hospital
• Connect hospital i2b2 systems with a secure network
• Enforce agreed upon usage rules
• Extensive registration/authentication/auditing
SHRINE Today

• HMS: 5 major hospitals (general, pediatric & cancer)
• > 6 patients
• > 2 billion facts
• Launched January 2011
• Software released as open source
• Steadily increasing usage
• Also, NW hosp x 3, pedi IBD and CARRAnet

Application to a common pediatric disease

• With an understudied epidemiology

Bowel Disorders

- Lower gastrointestinal disease
- Gastritis and duodenitis
- Esophageal disorders
- Gastric ulcer
- Gastrojejunal ulcer
- Peptic ulcer, site unspecified
- Duodenitis
Now: Virtual Cohort Studies

- i2b2 Clinical Research Cell + Full Genome Sequence
- Nominal-Quantitative Variables
- Bioinformatics Tools
- Automating Controls

CORE 1
- Study Cohort Selection
- i2b2 Virtual Cohort Study
- Patient Characterizations
- Outcomes

CORE 2
- DBP1: Autoimmunity and CVD
- DBP2 Diabetes and CVD
- Systems Medicine / Genomics

CORE 3
- Visualization Tools
- Clinical Trial Tools
- i2b2 Analytics
- Prediction Tools
- Multilingual NLP
- Autonomous Self-Supporting Open Source Community