EHR Challenges & Solutions for Genomic / Proteomics Personal Medical Data
MI 403 – Final Presentation

Ty Lee, Michael Sleep & Biljana Spasic
November 23, 2010
Personalized medicine with the advancement of genomics & proteomics (e.g., protein biomarkers) are becoming more integral to patient care.

Patient privacy and confidentiality risks do exist. Existing EHR technologies are not prepared to leverage genomic/proteomic advancements.

Secure solutions must be recommended of how to handle, store, reference, and utilize highly sensitive and confidential patient information.
Hypothesis & Aims

- **Personalized medicine via the genomic & proteomic fields are evolving**
  - Patient rights & confidentiality protection laws have not kept pace
  - FDA oversight absent / ‘who’ then educates patients on risks unclear

- **EHR technical hurdles exist**
  - Existing architectures unprepared to utilize genomic & phenotypic correlated data
  - Solutions still under development

- **Our Aims**
  - Develop legal view of how to best protect patients from discrimination / bias
  - Develop a technology view of how best to store and utilize important genomic & proteomic data
  - Develop a view of how to engage FDA to oversee the developing genomic & proteomic field
The Current State of Personalized Care with Genomics & Proteomics
What is Personalized Medicine?

- **Description**
  - Genomics medicine, enabled by advances of genomics and proteomics sciences
  - Takes into account personal variability and tailors medical treatments to the individual patient
  - Classifies individuals into subpopulations that differ in susceptibility to a particular disease or their response to a specific treatment
  - Addition to traditional methods of clinical medicine
  - Applied to prevention and treatment of disease

- **Main Targets**
  - Complex diseases, e.g. cancer, heart disease, diabetes, Parkinson’s, Alzheimer’s, AIDS
  - These diseases are results of interactions between genes as well as environmental factors, and require studies of whole genomes not individual genes only

- **Role in Improving Patient Care**
  - Take into account individual variations
  - Start or modify treatments earlier in the process
  - Improve outcomes with better targeted treatments
  - Reduce undesirable side effect
  - More proactive prevention based on genetic predisposition
An integrated use of genomic, clinical, and other data to predict clinical and biological phenotypes.

Integrated Models

Gene Expression Profiles

Proteomic Profiles
Metabolic Profiles

Genomic Data
SNPs
Epigenetic

Clinical Data
Treatments
Family history
Demographics
Environmental

Imaging

Predictions:
Risk
Individualized Prognosis & Diagnosis
Drug Response
Environment (eg Diet) Response

West M et al. Genome Res. 2006;16:559-566
Integration of Genomic Data into EMR Systems

**Potential Advantages**
- EMR can make the details of individual variations available to physicians at the point of care.
- EMR can be a starting point for integration of genomic data with clinical details, medical history, socio/environmental and other patient data, as well as with other data in systems that cover functions such as clinical decision support, diagnostic testing, and medications, etc.
- EMR can become a comprehensive source of targeted population data for genomic-based research and a tool in advancing genomic science.

**Technology Response**
- Some EHR vendors are working on incorporating genetic data, e.g. Cerner.
Some Challenges of Incorporating Genomic Data into Patient Care Practices

- The National Institute for Health (NIH) and Food and Drug Administration (FDA) are taking steps in supporting and regulating the advances of personalized medicine methods and services.

- **Scientific Research**
  - Various studies use different genetic markers; determine the most significant genetic markers and introduce some standardization.
  - Conduct clinical studies to understand correlations between genetic characteristics and effects of specific medication.

- **Diagnostic Testing**
  - Regulate genetic testing to protect patients and not hinder scientific progress.
  - Provide a single source for over 2000 known genetic tests available through clinical laboratories along with their FDA approval status.

- **Therapies**
  - Set acceptable limits for effects of new therapies.

- **Path to Implementation**
  - Define processes for review, assessment, documentation and coordinated approval of new diagnostics and therapies for clinical use. The FDA's “Critical Path Initiative aims to develop better evaluation tools, such as biomarkers and new assays.”
Insights into the Developing Ethical & Legal Issues
Patient EHR Data Security

- **Personalized diagnoses & treatments integral part of medicine**
  - Advancements in genomics & proteomics

- **Solutions must address ethical & legal issues**
  - Patient privacy & confidentiality
  - Use & access of patient genetic information
  - Patient testing & results interpretation
  - Clinical issues related to data & testing
  - Commercialization & product property rights

- **2008 U.S. Genetic Information Nondiscrimination Act (GINA) good, but not good enough**
  - Law prohibits U.S. health insurance firms and companies / employers from accessing private patient data and discriminating on the basis of the patient data and test results
  - Gaps still exist for patients
Genetic Information Nondiscrimination Act (GINA) & the FDA

- **GINA provides basic protection against discrimination**
  - Protection from health insurance/employment decisions or genomic test demands
  - Patient privacy & confidentiality ensured researchers would stay focused on developing disease therapies and treatments with genetic linkages
  - No protection from life insurance, long-term care, or disability insurance companies

- **FDA oversight not yet in place**
  - FDA has little federal regulation: test are considered “services”
  - Key unanswered issues:
    - Robustness or utility of certain tests when science that may be questionable
    - Oversight of genomic / proteomic testing continually being developed
    - Responsibility for patient education on limitations of science & privacy risks

- **Solution:** begin by amending GINA to protect patients from other insurers and (industry motivated) patient care providers
Utilization of the EHR for Genotypic and Phenotypic Association Studies
Common Technology Architecture for Research

- **Electronic Health Record (EHR)**
  - Stores patient demographic data
  - Stores primarily phenotypic (observation) data
  - Stores results from ancillary systems
    - Laboratory
    - Diagnostic Imaging
    - Genomic studies

- **DNA Biobank**
  - Stores extensive genomic data
  - Large number of discrete, codified data fields
  - Excludes demographic information
  - Ability to associate Genomic information to EHR data through unique identifier
Data Collection Methods for DNA Biobanks

- **DNA Harvested from Discarded Blood Samples**
  - Samples left over from routine clinical testing
  - Blood has been retained for 3 days post test, and scheduled for disposal

- **De-Identification**
  - In accordance with Title 45 Code of Federal Regulations part 46
  - Considered under criteria for research on non-human subjects

- **Exclusions**
  - Patients under the age of 18
  - Insufficient DNA samples
  - Samples absent a consent form
  - Samples from patients who have opted out
  - Duplicate samples
  - A randomly selected 2% of eligible samples
## Electronic Medical Records and Genomics (eMERGE) Network

<table>
<thead>
<tr>
<th>Institution</th>
<th>Biorepository Overview</th>
<th>Recruitment Model</th>
<th>Repository Size</th>
<th>EMR Summary</th>
<th>Phenotype</th>
<th>Phenotyping Methods*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health (Seattle, WA)</td>
<td>Alzheimer's Disease Patient Registry and Adult Changes in Thought Study</td>
<td>Disease specific</td>
<td>~2,800</td>
<td>20+ years pharmacy data, 15+ years radiology and pathology reports, 15+ years ICD9 data, Comprehensive EMR since 2004</td>
<td>1°: Alzheimer's Disease &amp; Dementia, 2°: White Blood Cell Counts</td>
<td>Coded data extraction, NLP, Manual chart review, Computer assisted chart abstraction</td>
</tr>
<tr>
<td>Marshfield Clinic (Marshfield, WI)</td>
<td>Personalized Medicine Research Project Marshfield Clinic, an integrated regional health system</td>
<td>Geographic</td>
<td>~21,000</td>
<td>Comprehensive EMR since 1985, 75% participants have 20+ years medical history</td>
<td>1°: Cataracts &amp; Low HDL, 2°: Diabetic Retinopathy</td>
<td>Coded data extraction, NLP, Intelligent Character Recognition</td>
</tr>
<tr>
<td>Mayo Clinic (Rochester, MN)</td>
<td>Mayo Clinic Non-Invasive Vascular Laboratory &amp; Exercise Stress Testing Lab</td>
<td>Disease specific</td>
<td>~3,300</td>
<td>Comprehensive EMR since 1995, 40 years of history of data extraction</td>
<td>1°: Peripheral Arterial Disease (PAD), 2°: Red Blood Cell Counts</td>
<td>Coded data extraction, NLP</td>
</tr>
<tr>
<td>Northwestern University (Chicago, IL)</td>
<td>Nugene Project: Northwestern affiliated hospitals and outpatient clinics</td>
<td>Clinic &amp; Hospital</td>
<td>~10,000</td>
<td>20+ years ICD9 data, Comprehensive EMR since 2000</td>
<td>1°: Type 2 Diabetes, 2°: Lipids &amp; Height</td>
<td>Coded data extraction, NLP, Mining text using regular</td>
</tr>
<tr>
<td>Vanderbilt University (Nashville, TN)</td>
<td>BioVU: Vanderbilt Clinic, diverse outpatient clinics</td>
<td>Outpatient lab draws</td>
<td>~30,000/200,000</td>
<td>35+ years medical history, Comprehensive EMR since 2000</td>
<td>1°: QRS &amp; PR Duration, Other: Phenotype</td>
<td>Coded data extraction, NLP</td>
</tr>
</tbody>
</table>
Genomic Association Studies

- **Approach**
  - Leverage linked genotype & phenotype data from EHR and Biobank
  - Genotype - > Phenotype Association
    - Can a known gene marker predict likelihood of disease state?
  - Phenotype -> Genotype Association
    - Can a known disease state be linked to specific gene markers?

- **Technology Architecture Consideration**
  - Should complete genomic data be stored in EHR?
  - Should complete genomic data be stored for potential re-test?
  - Can patients be positively identified by their genomic data?

- **Value in Clinical Research**
  - Rapid identification of research subsets
  - Low cost to produce research data set
  - Assistive technologies improve data accuracy
    - Natural Language Processing
  - Electronic codification of data fields and mapping to clinical terminologies
Recommendations

- **Personalized medicine via the genomic & proteomic fields are evolving**
  - Patient rights & confidentiality protection laws have not kept pace
  - FDA oversight absent / ‘who’ then educates patients on risks unclear

- **EHR technical hurdles exist**
  - Existing architectures unprepared to utilize genomic & phenotypic correlated data
  - Solutions still under development

- **Recommendations**
  - Amend GINA law to protect patients from industry (life / disability insurers, LT care)
  - Create and enhance bio-repositories
  - Patient / public health lobby groups need to engage FDA on oversight / patient education
Q&A

Thank you!!

MI 403 – Final Presentation

Ty Lee
Mike Sleep
Biljana Spasic
References


