EHRs and large scale comparative effectiveness research

September 16, 2014
Dana C. Crawford, PhD
Associate Professor
Epidemiology and Biostatistics
Institute for Computational Biology
Single Nucleotide Polymorphism (SNP)
EHRs and Research Towards Personalized Medicine

- Use of EHRs for Genomic Discovery
- Assessing Clinical Utility
- Towards Comparative Effectiveness Research

Ex: Hypothyroidism
Genomic Discovery Study Designs

Linkage Studies in Families

“Simple” Inheritance
Single Gene
Rare variants

Genetic Association Studies

Complex Inheritance
Multiple Genes
Common Variants

Controls
Cases
Study Designs

Candidate gene/pathway

Genome-wide association study (GWAS)

Published Genome-Wide Associations through 12/2012
Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories

NHGRI GWA Catalog
www.genome.gov/GWASTudies
www.ebi.ac.uk/fgpt/gwas/
Beyond GWAS

Box 6 | Clinical translation of findings from GWA studies

YOU ARE HERE

- Identification of regulatory variants
  - Novel biological insights
    - Clinical advances
      - Therapeutic targets
      - Biomarkers
      - Prevention
  - Improved measures of individual aetiological processes
    - Personalized medicine
      - Diagnostics
      - Prognostics
      - Therapeutic optimization

Genetic Association Studies Require Large Sample Sizes

But…..

• Already several existing cohorts

• Very expensive

Commentary

New Models for Large Prospective Studies: Is There a Better Way?


Biobanks Linked to EHRs

- Large
- Relatively inexpensive compared with cohorts
- Some are population-based
- Some are prospective
- Store more than DNA
- Ethical issues
  - Consent
  - Return results?
eMERGE Network

electronic medical records & genomics

emerge.mc.vanderbilt.edu
• Assess utility of DNA collections integrated with the EHRs as a resource for genome science

• Ethics (RROR)

• Clinical translation: assessment and model

• Network-wide genomic data
  ~87K samples
  Pediatric and adult
  Racial/ethnic diversity
EHRs and Challenges for Uses in Research

Variable Data Density

Mean (SD) # of clinical visits  
81.8 (107.8)

Range of # of clinical visits  
1 to 1,456

Mean (SD) # of ICD9 codes  
147.3 (230.4)

Range of # of ICD9 codes  
1 to 3,617

Ex: EAGLE BioVU (n=15,863)
Clinical versus Population-based

Ten most common codes among African Americans (>18 years):

Hypertension (401.9, 401.1)
Diabetes Mellitus (250)
End-stage renal disease (585.6)
Sequestrectomy (77)
Long-term current use of other medications (v58.69)
Other malaise/fatigue (780.79)
Chest pain (786.5)
Heart failure (428)
Kidney transplant (v42.0)
Ten most common codes among Hispanics (>18 years):

Sequestrectomy (77)
Supervision of other normal pregnancy (v22.1)
End-stage renal disease (585.6)
Hypertension (401.9, 401.1)
Outcome of delivery, single liveborn (V27.0)
Diabetes Mellitus (250)
Unknown; default (000.00)
Supervision of normal first pregnancy (V22.0)
Long-term current use of other medications (v58.69)
EHRs and Genetic Association Studies

• Challenges
  – EHR is not built for research
  – Billing codes can be unreliable in defining cases and controls

• “Demonstration Project”
  – Can the EHR be used to define cases and controls for genetic association studies?
  – Can these EHR-defined cases and controls replicate known genotype-phenotype associations?
General algorithm for determining an EHR phenotype

Definite Cases (algorithm-defined)

Possible Cases (require manual review)

Excluded (algorithm-defined)

Controls (algorithm-defined)

- Billing codes, procedure codes, labs, meds, free text
- Iterative process
- Strive for PPV>95%
Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record


The diagram shows the number of patients for various diseases:
- Atrial fibrillation
- Crohn disease
- Multiple sclerosis
- Rheumatoid arthritis
- Type II diabetes

The bars represent the number of cases (red) and controls (blue) for each disease.
Where previous OR>1.25, 8/14 replicated
Where previous OR<1.25, 0/7 replicated
GWAS for PR Interval


## EAGLE BioVU

Phenotypes developed in BioVU/eMERGE

<table>
<thead>
<tr>
<th>Condition</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atrial fibrillation</td>
<td>QT interval</td>
</tr>
<tr>
<td>BMI</td>
<td>Red blood cell indices</td>
</tr>
<tr>
<td>Cataract</td>
<td>Resistant hypertension</td>
</tr>
<tr>
<td>Crohn’s disease</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td>Dementia</td>
<td>Thyroid stimulating hormone levels</td>
</tr>
<tr>
<td>Diabetic retinopathy</td>
<td>Type 2 diabetes</td>
</tr>
<tr>
<td>Height</td>
<td>White blood cell indices</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td></td>
</tr>
<tr>
<td>Lipids (HDL-C, LDL-C, TG)</td>
<td>Cancers (EAGLE/PAGE)*</td>
</tr>
<tr>
<td>Multiple sclerosis</td>
<td>Menopause/Menarche (EAGLE/PAGE)**</td>
</tr>
<tr>
<td>Peripheral artery disease</td>
<td>Cardiac structure and function***</td>
</tr>
<tr>
<td>PR interval</td>
<td></td>
</tr>
<tr>
<td>QRS duration</td>
<td></td>
</tr>
</tbody>
</table>

*Bush et al *Pac Symp Biocomput* 2013:373-384  
**Malinowski et al *Pac Symp Biocomput* 2014:376-387  
***Wells et al *Journal of Clinical Bioinformatics* (in press)
Genetic Discovery in eMERGE

- Five biobanks linked to EHRs with GWAS data on ~18K samples

- Can we collaborate to
  Use additional clinical data to define “new” trait?
  Implement algorithm across all sites?
  Perform discovery GWAS?

Variants Near FOX1E Are Associated with Hypothyroidism and Other Thyroid Conditions: Using Electronic Medical Records for Genome- and Phenome-wide Studies

Joshua C. Denny,1,2,17,* Dana C. Crawford,3,4,17 Marylyn D. Ritchie,1,3,4 Suzette J. Bielinski,5 Melissa A. Basford,6 Yuki Bradford,4 High Seng Chai,7 Lisa Bastarache,1 Rebecca Zuvich,3,4 Peggy Peissig,8 David Carrell,9 Andrea H. Ramirez,2 Jyotishman Pathak,7 Russell A. Wilke,2 Luke Rasmussen,8 Xiaoming Wang,6 Jennifer A. Pacheco,14 Abel N. Kho,10 M. Geoffrey Hayes,10 Noah Weston,9 Martha Matsumoto,7 Peter A. Kopp,10,14 Katherine M. Newton,8 Gail P. Jarvik,11 Rongling Li,12 Teri A. Manolio,12 Iftikhar J. Kullo,13 Christopher G. Chute,7 Rex L. Chisholm,14 Eric B. Larson,9 Catherine A. McCarty,15 Daniel R. Masys,1 Dan M. Roden,2,16 and Mariza de Andrade7

Hypothyroidism

• Symptoms
  – Fatigue
  – Weight gain

• Risk Factors
  – Female sex
  – Increased age
  – Family history

• Complications
  – Heart problems, goiter, depression, birth defects
Domain experts define phenotype (VU)

Create initial EMR-based algorithm (VU)

Evaluate & refine (VU)

Share algorithm

GH

Marsh

Mayo

NW
Case/Control Algorithm

No thyroid-altering medications (e.g., Phenytoin, Lithium)

ICD-9s for Hypothyroidism
Abnormal TSH/FT4
Thyroid replacement medication
No secondary causes (e.g., pregnancy, ablation)
Case

2+ non-acute visits

No ICD-9s for Hypothyroidism
Normal TSH
No thyroid replace. meds
No hx of myasthenia gravis
Control
eMERGE GWAS Sample Size, by Study

Total = 13,617 European Americans
eMERGE Hypothyroidism Case Sample Size, by Study

Total = 1,317
eMERGE GWAS for Hypothyroidism

1,317 cases
5,053 controls
## Results

### Unmatched Analysis

<table>
<thead>
<tr>
<th>SNP</th>
<th>Odds ratio</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7850258</td>
<td>0.74</td>
<td>0.67-0.82</td>
<td>3.93x10^{-9}</td>
</tr>
<tr>
<td>rs965513</td>
<td>0.74</td>
<td>0.67-0.82</td>
<td>4.15x10^{-9}</td>
</tr>
<tr>
<td>rs925489</td>
<td>0.74</td>
<td>0.67-0.82</td>
<td>4.64x10^{-9}</td>
</tr>
<tr>
<td>rs10759944</td>
<td>0.75</td>
<td>0.68-0.83</td>
<td>8.13x10^{-9}</td>
</tr>
</tbody>
</table>

1,317 cases; 5,053 controls; adjusted for decade of birth, sex, site (±PC)
Personalized medicine

Scientific Discovery → Validation → Implementation
Personalized medicine

Scientific Discovery → Validation → Implementation
Importance of Validation

Living well starts with knowing your DNA.

Our genes make us who we are, so naturally they impact our health. By knowing your DNA, you can take steps toward living a healthier life.

Plan for the future.
Find out if your children are at risk for inherited conditions, so you can plan for the health of your family. about carrier status

Stay one step ahead.
Understand your genetic health risks. Change what you can, manage what you can't. about health risks

Talk to your doctor.
Arm your doctor with information on how you might respond to certain medications. about drug response

www.23andme.com/health (10/17/2013)
Hypothyroidism

Hypothyroidism is a condition in which the thyroid gland does not produce enough thyroid hormone. Left untreated, hypothyroidism can lead to obesity, low energy levels, infertility, heart disease, and other serious medical problems. Primary hypothyroidism, the most common thyroid disorder, affects 1%-5% of the population. Although hypothyroidism can affect anyone, women, especially those over the age of 50, are at highest risk for developing the condition.

The following results are based on Preliminary Research for 5 reported markers, updated April 12th, 2012.

Primary hypothyroidism

<table>
<thead>
<tr>
<th>Journal</th>
<th>PLoS ONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Size</td>
<td></td>
</tr>
<tr>
<td>Replications</td>
<td>1</td>
</tr>
<tr>
<td>Contrary Studies</td>
<td>None</td>
</tr>
<tr>
<td>Applicable Ethnicities</td>
<td>European</td>
</tr>
<tr>
<td>Marker</td>
<td>rs7850258</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Who</th>
<th>Genotype</th>
<th>Genetic Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dana Crawford</td>
<td>AG</td>
<td>Typical odds of developing primary hypothyroidism.</td>
</tr>
<tr>
<td></td>
<td>GG</td>
<td>Slightly higher odds of developing primary hypothyroidism.</td>
</tr>
</tbody>
</table>
Components of Utility & Validity

- 44 questions
- Identify knowledge gaps to aid scope of future research

Image: http://www.cdc.gov/genomics/gtesting/ACCE
EGAPP Initiative 2004-present

- Develop methodology for evaluating evidence on genomic tests and translation of those tests into recommendations for use in clinical practice
- Systematic review process
Analytic framework

1. Clinical population
2. Genetic screening
3. Identify at-risk individuals from genetic risk profile
4. Early intervention
5. Improved outcomes
6. Harms caused by screening
7. Harms caused by intervention
Rapid Review for Clinical Validity

• Q: Does the genotyping of the 5 specific variants, previously associated with hypothyroidism, in adult asymptomatic women of reproductive age lead to improved health outcomes?

<table>
<thead>
<tr>
<th>Variant</th>
<th>Gene</th>
<th>GWAS Study</th>
<th>OR</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7850258</td>
<td>FOXE1</td>
<td>Denny, JC; Crawford DC, 2011</td>
<td>0.74</td>
<td>3.96x10^{-9}</td>
</tr>
<tr>
<td>rs2476601</td>
<td>PTPN22</td>
<td>Eriksson, N, 2012</td>
<td>1.36</td>
<td>3.9x10^{-13}</td>
</tr>
<tr>
<td>rs3184504</td>
<td>SH2B3</td>
<td>Eriksson, N, 2012</td>
<td>0.84</td>
<td>2.6x10^{-12}</td>
</tr>
<tr>
<td>rs4915077</td>
<td>VAV3</td>
<td>Eriksson, N, 2012</td>
<td>1.30</td>
<td>7.5x10^{-10}</td>
</tr>
<tr>
<td>rs2517532</td>
<td>HLA region</td>
<td>Eriksson, N, 2012</td>
<td>0.86</td>
<td>1.3x10^{-8}</td>
</tr>
</tbody>
</table>
Rapid vs. Systematic Review

Rapid evidence review
• Formulate overarching question
• Formulate key questions
• Create search query for literature for PubMed
• Abstract review - 2 reviewers
• Full text review - 2 reviewers*
• Analysis
• Write-up
• Time: ~7 mos, working part-time; minimal cost

Systematic evidence review
• Formulate overarching question
• Formulate key questions
• Create search query, identify databases to use
• Abstract review - 2+ reviewers
• Full text review - 2+ reviewers
• Analysis
• Write-up
• Time: ~1-2+ yrs, costly

*access to 3rd reviewer as necessary
Key Questions for Rapid Review

• What is the clinical validity of these SNPs?
  – odds ratios/effect sizes
  – positive predictive values
Key Questions for Rapid Review

• What is the clinical validity of these SNPs?
  – odds ratios/effect sizes
  – positive predictive values

• Does the genetic testing of these SNPs lead to improved health outcomes?
  – increased screening
  – reduction in time between symptom and diagnosis
  – increased treatment (including for subclinical disease)
  – harms associated with testing
Search Query for Rapid Review

- PubMed database, MeSH terms
Rapid Evidence Review Results

- 631 abstracts reviewed
- 346 full-text articles reviewed

<table>
<thead>
<tr>
<th>Reason for removal</th>
<th>Count (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full text unavailable</td>
<td>25</td>
</tr>
<tr>
<td>Study included pediatric samples (age ≤17), age not adjusted for or pediatric results reported separately</td>
<td>22</td>
</tr>
<tr>
<td>Study included males, sex not adjusted for or male results reported separately</td>
<td>47</td>
</tr>
<tr>
<td>Study excluded due to small sample size (n&lt;10)</td>
<td>3</td>
</tr>
<tr>
<td>Study excluded due to race/ethnicity not adjusted for or results stratified by race/ethnicity</td>
<td>5</td>
</tr>
<tr>
<td>Study excluded because sample included individuals with thyroid cancer, results not stratified by cancer status</td>
<td>1</td>
</tr>
<tr>
<td>Study excluded because no effect sizes/odds ratios given</td>
<td>3</td>
</tr>
<tr>
<td>Study excluded because no positive predictive values/AUC given</td>
<td>15</td>
</tr>
<tr>
<td>Study excluded because study did not address any of the key questions of the rapid evidence review</td>
<td>225</td>
</tr>
</tbody>
</table>
Summary: Rapid Evidence Review

• Streamlined the evidence review process
Summary: Rapid Evidence Review

• Streamlined the evidence review process

• Found inadequate evidence to recommend for or against genetic testing in asymptomatic women of childbearing age for hypothyroidism
Summary: Rapid Evidence Review

- Streamlined the evidence review process
- Found inadequate evidence to recommend for or against genetic testing in asymptomatic women of childbearing age for hypothyroidism
- Results consistent with others
  - lack of clinical validity/utility data
Components of Utility & Validity

- 44 questions
- Identify knowledge gaps to aid scope of future research

Image: [http://www.cdc.gov/genomics/gtesting/ACCE](http://www.cdc.gov/genomics/gtesting/ACCE)
Comparative Effectiveness Research

CER aims to provide evidence towards consequences of treatment options:

• Effectiveness
• Benefits
• Harms

http://effectivehealthcare.ahrq.gov
Comparative Effectiveness Research

CER research is conducted as

• Systematic reviews of existing datasets
  Clinical trials
  Clinical reviews
  Other research

• Studies that generate new evidence

http://effectivehealthcare.ahrq.gov
Personalized Medicine
Ex. Hypothyroidism

Scientific Discovery

Validation

Clinical validity
Clinical utility (CER)

Implementation

FoxE1
PTPN22
SH2B3
VAV3
HLA region

23andMe
CER and EHR Opportunities

• Large and “inexpensive”
• Prospective (with regular clinic visits)
• Can be harmonized* with epidemiologic studies

• Variable data density
• May or may not be representative of population
• Patients lost to follow-up
• Effort required to define phenotypes
• Exposure data difficult to extract

*Depends on phenotype/trait
Personalized Medicine and Role of EHRs for Research

Scientific Discovery → Validation → Implementation
Population Architecture Using Genomics and Epidemiology

PAGE

EAGLE

Epidemiologic Architecture for Genes Linked to Environment

WHI

FRED HUTCHINSON CANCER RESEARCH CENTER

WOMEN'S HEALTH INITIATIVE

MEC

The Multiethnic Cohort Study
Follow-Up Health Survey

University of Hawai‘i Cancer Research Center of Hawai‘i
(808) 586-2996

University of Southern California Keck School of Medicine
1-800-786-3538

CALiCo

CARDIA
Coronary artery risk development in young adults

CHS
SOL Study of osteoporotic fracture study

ARIC
Atherosclerosis Risk in Communities Study

Coordinating Center

NHGRI