Integration of Biobanks into Clinical Healthcare through Linkage to Medical Records—The eMERGE experience

Rex L. Chisholm, PhD
Adam and Richard T. Lind Professor of Medical Genetics
Vice Dean for Scientific Affairs
Feinberg School of Medicine
Northwestern University
Personalized Medicine
NON-RESPONDERS AND TOXIC RESPONDERS

Treat with alternative drug or dose

RESPONDERS AND PATIENTS NOT PREDISPOSED TO TOXICITY

Treat with conventional drug or dose

ALL PATIENTS WITH SAME DIAGNOSIS

PRESCRIPTION BLANK

DEAF ALD/NOID/10001
B3A0024001

PATIENT

D.O.B.

ADDRESS

DATE

INSTRUCTIONS PERMITTED DO NOT SUBSTITUTE

SIGNATURES OF PRESCRIBER

 Prescriptions is valid if more than one controlled substance prescription is written out in block.

Rebecca Heredia
Critical factors for Defining Genetic Contributions to Disease: A key to Personalized Medicine

- Methods to measure genetic variation in an individual human genome
- Large numbers of well phenotyped human genomes
- Development of standards of care / best practices
- Methods to deploy genomics based decision support
Northwestern’s Biobank: NUgene

- Launched in fall 2002
- IRB approved
- Voluntary participation via informed consent
- Collect biological specimens under a broad consent
- One-time questionnaire at enrollment
- Longitudinal medical information captured from EMR
- Secure application and database
- Enables High-throughput phenotyping
- Recontact option for additional research
- Resource facilitating genetic research institution-wide
- Genotype data returned to NUgene
Informed Consent Document

Allows:

– Blood sample to be drawn and DNA stored for length of study
– Access to participants’ EHR data
– Questionnaire information to be obtained
– Future use of DNA sample for “genetic variation” research
– Access to de-identified DNA and medical information by third parties with approved research studies, including companies
– Recontact (optional)

• Most EHR not appropriately consented (special protections for mental health and ?! genetics)
Privacy Protections

• Patient identifiers replaced with a computer generated barcode
• Single link between the DNA sample and medical information remains in a secure database, called NOTIS
• Physicians are not informed of patient’s participation or results
• Certificate of Confidentiality
  – Issued by the NIH to protect participant’s information involved in certain research projects from subpoena and other legal actions
Data Flow

Participant Enrollment Materials

Coding & Data Parsing Process

Coded Data

Patient Identifiers

MRNs SSNs

Medical Record

NOTIS

NUgene Database

Enterprise Data Warehouse

Encryption Decryption

Phenotypic Engine
NUgene today

• Over 10,600 participants are enrolled
  – Male: 42%  Female: 58%
  – Median age: 51
  – Age distribution: 18 - 85+
  – Ethnic breakdown similar to census data for 6-county area
  – Over half enrolled through &/or seen at primary care clinics
  – Average participant has 31 distinct diagnoses (ICD9 non V/E codes), & an average of 16 distinct diagnoses assigned at least twice**
  – Average patient followed over 8 years, some patients >20 years

• Overall participation rate is 28%
  – Uptake rate is ~52% if physician mentions the study

• 92% of participants agree to be contacted for future research or additional health information

**assigned the same ICD-9 code on 2 or more dates
Data Sources

• Questionnaire (self-report):
  – Completed once, at time of enrollment:
    • Demographic information
    • Environmental exposures
    • Medications
    • Self-reported family and medical history for select conditions

• Electronic billing record data
  – Retrospective and prospective diagnosis (ICD9) & procedure (CPT/ICD9CM) codes

• Electronic medical record data
  – Retrospective and prospective:
    • Medical history and diagnoses
    • Lab tests and results
    • Medications and therapies
    • Family and social history
    • Free text physician notes*
EHRs at the Northwestern Medical Center

- Operates a state of the art EHR system
  - Well integrated systems
  - Mix of commercially and internally developed systems
- Electronic data capture for virtually all aspects of inpatient and outpatient care, including
  - Over 20 years of clinical data
  - Anthropometric and clinical measures, prescriptions, diagnoses, lab measures, and clinical notes
- Using high quality, widely used commercial systems
  - Cerner PowerChart – inpatient records
  - EpicCare Ambulatory EHR – outpatient records
  - PRIMES – inpatient registration and billing records
  - IDX – outpatient registration and billing records
<table>
<thead>
<tr>
<th>ICD9</th>
<th>ICD9 Description</th>
<th># Pts**</th>
<th>% Pts**</th>
</tr>
</thead>
<tbody>
<tr>
<td>401</td>
<td>Essential hypertension</td>
<td>3310</td>
<td>34%</td>
</tr>
<tr>
<td>272</td>
<td>Disorders of lipoid metabolism</td>
<td>3269</td>
<td>34%</td>
</tr>
<tr>
<td>530</td>
<td>Diseases of esophagus</td>
<td>2259</td>
<td>23%</td>
</tr>
<tr>
<td>427</td>
<td>Cardiac dysrhythmias</td>
<td>1825</td>
<td>19%</td>
</tr>
<tr>
<td>451</td>
<td>Phlebitis and thrombophlebitis</td>
<td>1688</td>
<td>17%</td>
</tr>
<tr>
<td>278</td>
<td>Overweight, obesity and other hyperalimentation</td>
<td>1566</td>
<td>16%</td>
</tr>
<tr>
<td>715</td>
<td>Osteoarthrosis and allied disorders</td>
<td>1411</td>
<td>15%</td>
</tr>
<tr>
<td>477</td>
<td>Allergic rhinitis</td>
<td>1408</td>
<td>14%</td>
</tr>
<tr>
<td>216</td>
<td>Benign neoplasm of skin</td>
<td>1380</td>
<td>14%</td>
</tr>
<tr>
<td>250</td>
<td>Diabetes mellitus</td>
<td>1367</td>
<td>14%</td>
</tr>
<tr>
<td>311</td>
<td>Depressive disorder, not elsewhere classified</td>
<td>1261</td>
<td>13%</td>
</tr>
<tr>
<td>692</td>
<td>Contact dermatitis and other eczema</td>
<td>1238</td>
<td>13%</td>
</tr>
<tr>
<td>244</td>
<td>Acquired hypothyroidism</td>
<td>1102</td>
<td>11%</td>
</tr>
<tr>
<td>493</td>
<td>Asthma</td>
<td>1043</td>
<td>11%</td>
</tr>
<tr>
<td>366</td>
<td>Cataract</td>
<td>1010</td>
<td>10%</td>
</tr>
</tbody>
</table>

*From billing, encounter, problem list, med Hx.*

**Participants w/ EMR data as of 9/24/2010.
## Top Laboratory Tests within Population

<table>
<thead>
<tr>
<th>Laboratory Tests</th>
<th># Participants*</th>
<th>% Participants*</th>
<th># of Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose</td>
<td>8,509</td>
<td>88</td>
<td>265,963</td>
</tr>
<tr>
<td>Creatinine</td>
<td>8,422</td>
<td>87</td>
<td>124,848</td>
</tr>
<tr>
<td>Calcium</td>
<td>8,392</td>
<td>86</td>
<td>258,308</td>
</tr>
<tr>
<td>Potassium</td>
<td>8,387</td>
<td>86</td>
<td>265,022</td>
</tr>
<tr>
<td>Sodium</td>
<td>8,384</td>
<td>86</td>
<td>262,436</td>
</tr>
<tr>
<td>Chloride</td>
<td>8,384</td>
<td>86</td>
<td>261,752</td>
</tr>
<tr>
<td>Mean Corpuscular Volume (MCV)</td>
<td>8,364</td>
<td>86</td>
<td>247,740</td>
</tr>
<tr>
<td>Red Cell Distribution Width (RDW)</td>
<td>8,364</td>
<td>86</td>
<td>247,733</td>
</tr>
<tr>
<td>Mean Corp. Hemoglobin Conc. (MCHC)</td>
<td>8,364</td>
<td>86</td>
<td>247,740</td>
</tr>
<tr>
<td>Mean Corpuscular Hemoglobin (MCH)</td>
<td>8,364</td>
<td>86</td>
<td>247,740</td>
</tr>
<tr>
<td>Blood Urea Nitrogen</td>
<td>8,366</td>
<td>86</td>
<td>245,748</td>
</tr>
<tr>
<td>Bicarbonate</td>
<td>8,332</td>
<td>86</td>
<td>243,499</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>8,320</td>
<td>86</td>
<td>227,208</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>8,320</td>
<td>86</td>
<td>227,201</td>
</tr>
<tr>
<td>Red Cell Count</td>
<td>8,296</td>
<td>85</td>
<td>224,380</td>
</tr>
</tbody>
</table>

* Based on participants with EMR data as of 9/24/2010.
The eMERGE Network

A consortium of biorepositories linked to electronic medical records data for conducting genomic studies

An NHGRI funded consortium
eMERGE I Goals

Test the ability to leverage EMRs and biobanks for genomic research

- Evaluate validity & utility of EMR phenotypes for Genomics
- Develop & validate electronic phenotyping algorithms
- Conduct association studies of genome-wide data with EMR-derived phenotypes
<table>
<thead>
<tr>
<th>Site</th>
<th>eMERGE Phase I</th>
<th>eMERGE Phase II</th>
<th>eMERGE I &amp; II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants</td>
<td>Genotyped</td>
<td>Participants</td>
</tr>
<tr>
<td>GHC</td>
<td>2,820</td>
<td>2,789</td>
<td>5,291</td>
</tr>
<tr>
<td>Marshfield</td>
<td>20,000</td>
<td>4,210</td>
<td>20,000</td>
</tr>
<tr>
<td>Mayo</td>
<td>3,769</td>
<td>3,755</td>
<td>6,916</td>
</tr>
<tr>
<td>NU</td>
<td>10,500</td>
<td>1,907</td>
<td>12,000</td>
</tr>
<tr>
<td>VU</td>
<td>70,000</td>
<td>6,055</td>
<td>155,000</td>
</tr>
<tr>
<td>Geisinger</td>
<td>N/A</td>
<td>N/A</td>
<td>22,000</td>
</tr>
<tr>
<td>Mt. Sinai</td>
<td>N/A</td>
<td>N/A</td>
<td>25,000</td>
</tr>
<tr>
<td>CCHMC/BCH</td>
<td>N/A</td>
<td>N/A</td>
<td>11,799</td>
</tr>
<tr>
<td>CHOP</td>
<td>N/A</td>
<td>N/A</td>
<td>60,000</td>
</tr>
<tr>
<td>TOTAL</td>
<td>107,089</td>
<td>18,716</td>
<td>347,090</td>
</tr>
</tbody>
</table>
Approach to electronic phenotyping

1. Identify phenotype of interest
2. Case & control algorithm development and refinement
3. Manual review; assess precision
   - PPV < 95%
   - PPV ≥ 95%
4. Deploy at site 1
   - Validate at other sites
5. Genetic association tests; replicate
Type II Diabetes Case Algorithm

*Abnormal lab* = Random glucose > 200mg/dl, Fasting glucose > 125 mg/dl, or hemoglobin A1c ≥6.5%.
Type II Diabetes Control Algorithm

1. >=2 in person clinician visits? 
   - Yes → >=1 glucose measure? 
     - Yes → Abnormal* glucose or HbA1c? 
     - No → ICD-9 code for diabetes or related condition? 
       - No → Rx diabetes med, inc. insulin or supplies? 
         - No → Family Hx of diabetes (type 1 or 2)? 
         - Yes → T2DM control
   - No → ICD-9 code for diabetes or related condition? 
     - No → Rx diabetes med, inc. insulin or supplies? 
       - Yes → Family Hx of diabetes (type 1 or 2)? 
       - No → T2DM control

*Note: Abnormal glucose levels can be defined by specific clinical guidelines, such as fasting blood glucose levels above 126 mg/dL or HbA1c levels above 6.5%.
Type II Diabetes Chart Review

- Blinded clinician review of 100 random charts (50 cases & 50 controls)
- Case PPV = 98%
- Control PPV = 98%
Mega-Analysis (adjusted)

Imputed T2D Merged (Case/Ctrl) – 98GE SNPs, Adjusted Sex, Age, BMI, PC1, PC.

TCF7L2

Chromosome

-\log_{10}(p)$
# Phase I Phenotypes

<table>
<thead>
<tr>
<th></th>
<th>GHC/UW</th>
<th>Marshfield</th>
<th>Mayo</th>
<th>Northwestern</th>
<th>Vanderbilt</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dementia</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Cataract</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>PAD</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Type 2 Diabetes</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>QRS Duration</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>Secondary</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WBC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Diabetic Retinopathy</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>RBC</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lipids</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>PheWAS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>HDL</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Resistant HTN</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

*The eMERGE Network*

electronic Medical Records & Genomics
### Primary Phenotype-Gene Associations in eMERGE I

**Associations between 19 phenotypes and 38 genes**

<table>
<thead>
<tr>
<th>Disease Phenotype</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiac Conduction</td>
<td>SCN5A, SCN10A</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>FOXE1</td>
</tr>
<tr>
<td>LDL Cholesterol</td>
<td>APOE, TRIB1, LPL, ABCA1</td>
</tr>
<tr>
<td>Platelet Count &amp; Volume</td>
<td>5 Chromosomes Associated with PLT &amp; 8 with MPV</td>
</tr>
<tr>
<td>Glaucoma, Primary Open-Angle</td>
<td>....and MORE</td>
</tr>
<tr>
<td>Glaucoma, Optic Nerve Degeneration</td>
<td></td>
</tr>
<tr>
<td>Red Blood Cell Traits, Erythroid Differentiation and Cell Cycle Regulation</td>
<td>THRBP, PTPLAD1, CDT1</td>
</tr>
<tr>
<td>RBC Traits, Erythrocyte Sedimentation Rate (ESR)</td>
<td>CR1</td>
</tr>
<tr>
<td>RBC Traits, Malaria Resistance</td>
<td>HBB, HBA1/HBA2, G6PD</td>
</tr>
<tr>
<td>RBC Traits, Peripheral Artery Disease (PAD)</td>
<td>SLC17A1, BLS1/MYB, TMPRSS6, HFE</td>
</tr>
<tr>
<td>White Blood Cell Count</td>
<td>DARC, GSDMA, MED24, PSMD3</td>
</tr>
</tbody>
</table>
Merged Genotype Dataset

• 17,046 eMERGE samples with GWAS data

• Majority of samples genotyped using 660W

• Samples collected for various studies
  GH – Dementia
  Marshfield – Cataracts and HDL-C
  Mayo – PAD
  NW – T2D
  VU – Normal ECGs

Can we use existing dataset for another experiment?
Primary Hypothyroidism

- Most common form is chronic lymphocytic hypothyroidism (Hashimoto’s thyroiditis)
- More common in females (~10x)
- Other associated factors
  - age
  - race/ethnicity
  - family history of thyroid disease
- No published* GWAS (as of 07/23/2011)
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
## Phenotype Algorithm Validation

<table>
<thead>
<tr>
<th>Site</th>
<th>EMR-based Cases/Controls</th>
<th>Chart Review Cases/Controls</th>
<th>Case PPV (%)</th>
<th>Control PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health</td>
<td>397/1,160</td>
<td>50/50</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Marshfield</td>
<td>514/1,187</td>
<td>50/50</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>233/1,884</td>
<td>100/100</td>
<td>82</td>
<td>96</td>
</tr>
<tr>
<td>Northwestern</td>
<td>92/470</td>
<td>50/50</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>81/352</td>
<td>50/50</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>All sites (weighted)</td>
<td>1,317/5,053</td>
<td>—</td>
<td>92.4</td>
<td>98.5</td>
</tr>
</tbody>
</table>

An eMERGE-wide phenotype analyzed with no extra genotyping: hypothyroidism

European Americans (1,306 cases and 5,013 controls)

Denny et al., 2011
FOXE1

• Forkhead box E1 (thyroid transcription factor 2)

• ~3.46kb intronless gene

• Thyroid transcription factor which likely plays a crucial role in thyroid morphogenesis

• Mutations associated with congenital hypothyroidism and cleft palate with thyroid dysgenesis

• The map localization of this gene suggests it may also be a candidate gene for squamous cell epithelioma and hereditary sensory neuropathy type I
The phenome-wide association study (PheWAS)

<table>
<thead>
<tr>
<th>PheWAS requirement: A large cohort of patients with genotype data and many diagnoses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GWAS:</strong></td>
</tr>
<tr>
<td><strong>PheWAS (ΦWAS):</strong> Target genotype</td>
</tr>
</tbody>
</table>

**PheWAS requirement:** A large cohort of patients with genotype data and many diagnoses.
PheWAS for rs965513 (FOXE1)

Analysis of 866 phenotypes in 13,617 European Americans
Adjusted for age and sex
Hypothyroidism Conclusions

- “No genotyping” experiment yields new genetic associations
- Associations overlap with thyroid cancer GWAS
- Associations replicate in external dataset
- FOXE1 likely candidate, may be associated with other thyroid diseases
PheWAS of “all” NHGRI GWAS Catalog SNPs

3,144 SNPs with prior GWAS-discovered associations

674 SNPs with 86 phenotypes
751 SNP-phenotype associations

Test for replication of 751 associations using PheWAS

Replication Arm

3,144 SNPs

PheWAS for each SNP to discovery pleiotropy

Replication of novel associations

Discovery Arm

Denny et al, Nat Biotech 2013
Replications of NHGRI GWAS associations via PheWAS

Probability of replicating:
- All - 210/751: $2 \times 10^{-98}$
- Powered - 51/77: $3 \times 10^{-47}$

Denny et al, Nat Biotech 2013
Discovery to Practice: Integrating Genomics into the EMR

Phase 1
- Use EHR data for genome/phenome associations
- Cross institution phenotypes

Phase 2
- More sites
- Pediatrics
- Faster, better, cheaper EHR-based genomic science
- Integration of genomic information back into EHR and clinical care
# EMR-linked biobanks in eMERGE-II

<table>
<thead>
<tr>
<th>Site</th>
<th>Participants</th>
<th>(GWAS)-Genotyped Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health Seattle</td>
<td>6,381</td>
<td>3,606</td>
</tr>
<tr>
<td>Marshfield</td>
<td>20,000</td>
<td>4,225</td>
</tr>
<tr>
<td>Mayo</td>
<td>19,000</td>
<td>6,934</td>
</tr>
<tr>
<td>Northwestern</td>
<td>11,000</td>
<td>4,987</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>158,514</td>
<td>27,173</td>
</tr>
<tr>
<td>Geisinger</td>
<td>22,000</td>
<td>4,191</td>
</tr>
<tr>
<td>Mt. Sinai</td>
<td>22,000</td>
<td>2,867</td>
</tr>
<tr>
<td>CHOP</td>
<td>60,000</td>
<td>45,000</td>
</tr>
<tr>
<td>Cincinnati/Boston</td>
<td>10,000</td>
<td>3,000</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>328,895</strong></td>
<td><strong>101,983</strong></td>
</tr>
</tbody>
</table>
eMERGE II phenotyping: Lower GI Phenotypes

Colon Polyps

Diverticulosis

The eMERGE Network

electronic Medical Records & Genomics
Colon Polyp NLP Algorithm

1. **±Colon Polyps:**
   - checks EHR for colonoscopy with linked path report w. polyp mention

2. **Type + Location:**
   - NLP on path reports to extract type and location

<table>
<thead>
<tr>
<th></th>
<th>Sensitivity</th>
<th>PPV</th>
<th>Specificity</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>±Colon Polyps</td>
<td>98%</td>
<td>94%</td>
<td>94%</td>
<td>98%</td>
</tr>
<tr>
<td>Type + Location</td>
<td>96%</td>
<td>98%</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>
Colon Polyp Findings By Location

All Colon Polyps: Count per Location

- Hepatic Flexure: 1257
- Transverse: 3522
- Splenic Flexure: 574
- Descending: 2178
- Cecum: 4629
- Ileum: 259
- Sigmoid: 4906
- Rectum: 4454

Adenoma Count Per Location

- Hepatic Flexure: 904
- Transverse: 2590
- Splenic Flexure: 403
- Descending: 3423
- Cecum: 1844
- Ileum: 181
- Sigmoid: 2449
- Rectum: 1511

The eMERGE Network
electronic Medical Records & Genomics
Preliminary GWAS

Diverticulosis – Adjusted for PC1,PC2,SEX,AGE,BMI,SMOKE

Diverticulitis – Adjusted for PC1,PC2,SEX,AGE,BMI,SMOKE
Pharmacogenomics

- Deploy VIP platform developed by PGRN investigators
- Apply to participants enriched for encountering drugs for which there is a CPIC guideline
- Return appropriate genotype results through appropriate decision support tools
- Archive novel variants for further study
eMERGE PGx - Overview by Project

Specific Aim 1
- Recruit / Collect Samples
- PGRN-Seq Sequencing
- Clinical Variant Validation

Specific Aim 2
- Return Results: EHRIntegration and CDS
- Patient & Clinician Education

Specific Aim 3
- Populate Variant and Phenotype Data Repository (SPHINX)
## PGx candidate drug-gene pairs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Drug</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP2C19</td>
<td>clopidogrel</td>
<td>Best evidence in patients with coronary stents</td>
</tr>
<tr>
<td>CYP2C9</td>
<td>Warfarin</td>
<td>Algorithms to predict starting dose available. Vary by ancestry</td>
</tr>
<tr>
<td>VKORC1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP4F2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA B*5701</td>
<td>Abacavir</td>
<td></td>
</tr>
<tr>
<td>HLA B*1502</td>
<td>carbamazepine</td>
<td>Higher frequency variant in Asian subjects</td>
</tr>
<tr>
<td>SLCO1B1</td>
<td>Simvastatin</td>
<td>Especially at higher dosages or with interacting drugs</td>
</tr>
<tr>
<td>TPMT</td>
<td>6-MP, azathioprine</td>
<td></td>
</tr>
<tr>
<td>CYP2D6</td>
<td>Codeine</td>
<td>PM status predicts non-response</td>
</tr>
<tr>
<td>CYP2D6</td>
<td>some SSRIs</td>
<td>PM status predicts intolerance of effective dosages.</td>
</tr>
</tbody>
</table>
eMERGE Infobutton Project

Develop a new information resource based on the HL7 Infobutton standard

1. Required by Meaningful Use Stage 2
2. Multiple sites contribute content, all sites can access.

- Representation from many eMERGE sites
  - Crosses EHRI and CERC workgroups
- External collaborators
  - University of Utah – Dr. Guilherme del Fiol
  - Intermountain Healthcare – Dr. Nathan Hulse
  - NIH Clinical Center – Dr. Jim Cimino
eMERGE Infobutton Project

Implement infobuttons within EHRs at eMERGE institutions

- Identify supporting platforms
- Develop systems that don’t exist, or contribute features to existing OSS projects
- Assist with implementation and configuration
- Evaluate usage

Genetic Test Results

About These Results

Many things can explain why a person has a medical condition, or why different people respond to the same medication in different ways. Genetic testing, which looks for changes (also called polymorphisms, or mutations) in your DNA, can help.

It’s important to know that being at risk for a condition doesn’t mean you will necessarily get it. The results of these tests should be used with other pieces of evidence collected by your doctor to make medical decisions.

WARFARIN DOSING

What Does this Mean?
Nam nisi fels, pulvinar egest nibh in, placerat al quam dam. Integer placerat dolor elit, at b hendu m neque luctus interdict. Phasellus at blandit justo, egestas vehicula sem. Praesent vulputate aculis risus, vitae feugiat leo sollicitudin eget. In imperdiet velit non sapien convalis placer.

Results
CYP2C15: CC
Tested On 8/19/2013

SIMVASTATIN METABOLISM

What Does this Mean?
You have no detectable mutations that should affect metabolism of simvastatin. This does not preclude you from any other potential adverse events from taking simvastatin.

Results
Normal Activity (Predicted)
Tested On 8/19/2013

CLOPIDOGREL METABOLISM

What Does this Mean?
You had a genetic test looking for genetic differences to help predict how your body might respond to using the drug clopidogrel (Plavix®). Your test result showed that clopidogrel may not work as well for you as for most people.

Results
Poor Metabolizer (Predicted)
Tested On 8/19/2013

Back to the Home Page
## PGRNSeq-Incidental Findings (IFs)

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Gene Name</th>
<th>Phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACNA1S</td>
<td>Calcium channel, voltage-dependent, L type, alpha 1S subunit</td>
<td>malignant hyperthermia</td>
</tr>
<tr>
<td>KCNH2</td>
<td>Potassium voltage-gated channel, subfamily H (eag-related), member 2</td>
<td>long QT arrhythmia</td>
</tr>
<tr>
<td>LDLR</td>
<td>Low density lipoprotein receptor</td>
<td>hyperlipidemia</td>
</tr>
<tr>
<td>RYR1</td>
<td>Ryanodine receptor 1 (skeletal)</td>
<td>malignant hyperthermia</td>
</tr>
<tr>
<td>RYR2</td>
<td>Ryanodine receptor 2 (cardiac)</td>
<td>catecholaminergic polymorphic ventricular tachycardia</td>
</tr>
<tr>
<td>SCN5A</td>
<td>Sodium channel, voltage-gated, type V, alpha subunit</td>
<td>long QT arrhythmia</td>
</tr>
</tbody>
</table>
What is the Phenotype KnowledgeBase?

The reuse of data from electronic medical records (EMRs) and other clinical data systems holds tremendous promise for improving the efficiency and effectiveness of health research. Clinical data in the EMR is a potential source of rich longitudinal data for research, and the recent government efforts to promote the use of EMRs in the clinical setting may further promote the use of such systems in the US healthcare system. As the use of EMRs expands, the demand for usable data from these systems for research has also expanded.

One such effort by the Electronic Medical Records and Genomics Network (eMERGE) has investigated whether data captured through routine clinical care using EMRs can identify disease phenotypes with sufficient positive and negative predictive values for use in genome-wide association studies (GWAS). Most EMRs captured key information (diagnoses, medications, laboratory tests) used to define phenotypes in a structured format; in addition, natural language processing has also been shown to improve case identification rates.*

PheKB is an outgrowth of that validation effort and provides a collaborative environment of building and sharing validated phenotype algorithms.
PheKB.org
a knowledgebase for discovering phenotypes

• Public Side of the Website
  – Final phenotypes algorithms and implementation results

• Private Side of the Website
  – Ability to create phenotype workspaces and collaborate on development, testing, and validation on algorithms
  – Version history
  – Access controls
    • Only viewable by authors
    • Shared with a collaborative groups(s)
    • Public (“Final” algorithms only)
  – Comment on Phenotypes shared with you
  – Receive alerts on Phenotypes you are following
Patient survey: biobanking consent issues

Questions:
• Do participants view specific consent to be a requirement for sharing biosamples and data for future research?
• Which biospecimen and biobanking-related research practices are likely to have the greatest impact on willingness to participate under broad consent?

Plan
• Survey 100,000 participants and patients across the eMERGE institutions to elicit a wide cross-section of patient perspectives.

Outcome
• Recommendations to inform future policy for the ethical conduct of human subject research
Summary

• Biobanks and EHRs are increasingly playing a critical role in identifying associations between genetic variation, disease risk, drug efficacy and clinical outcomes
• Longitudinal mining of electronic medical records can be used to provide the most up to date phenotype associated with human biospecimens
• Research use can be an important driver of EHR quality
• Networks of EHR-linked biobanks that share samples and data have the potential to increase statistical power to detect genetic associations, population diversity in these studies, and overall research efficiency
• Methods to store genomic variation in EHR will enable personalized medicine
Acknowledgements

• eMERGE PIs and teams
  – Geisinger
  – Group Health/UW
  – Mayo
  – Marshfield
  – Mt. Sinai
  – Vanderbilt
  – CHOP
  – Boston Childrens Hospital
  – Cincinnati Childrens Hospital

• NU eMERGE team:
  – Rex Chisholm
  – Abel Kho
  – Bill Lowe
  – Wendy Wolf
  – Maureen Smith
  – Geoff Hayes
  – Loren Armstrong
  – Laura Rasmussen-Torvik
  – Jen A. Pacheco
  – Will Thompson

• Funding for eMERGE studies at NU is provided by NHGRI (U01HG004609)

• NUgene Governance Committee & Community Advisory Committee

• NUgene team:
  – Rex Chisholm, PhD (PI)
  – Warren Kibbe, PhD (Co-founder)
  – William Lowe, MD (Medical Dir.)
  – Wendy Wolf, PhD
  – Maureen Smith, MS, CGC
  – Jennifer A. Pacheco
  – Tony Miqueli
  – Sharon Aufox, MS, CGC
  – Oana Popescu
  – Nicole Sheehan
  – Noah Goss
  – Maribeth Miceli

• More information about NUgene: http://www.nugene.org