The electronic MEdical Records and GEnomics (eMERGE) network
The eMERGE Network

**Electronic Medical Records & Genomics**

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

---

**2007-2011: Phase I**

**eMERGE-I Goal**: to assess utility of DNA collections integrated with electronic medical records (EMRs) as resources for genome science

- Each site identified a phenotype of interest in ~3,000 subjects and conducted a genome-wide association study (GWAS)
- To what extent can identifiers be stripped from EMRs and research utility retained?
- Assess consent for genomic technologies & data sharing
A common general approach to study design in BioVU and eMERGE

- Application of electronic phenotype selection logic
- Definite cases
- Possible Cases
- Excluded (insufficient evidence)
- Controls
  - Manual review
  - Case-control genotypic analysis
Phenotype definitions are portable across EMRs

Table 1. Evaluation of Primary Hypothyroidism Algorithm at the Five eMERGE Sites

<table>
<thead>
<tr>
<th>Site</th>
<th>Primary Phenotype</th>
<th>Total Genotyped Subjects</th>
<th>Primary Hypothyroidism</th>
<th>Cases</th>
<th>Controls</th>
<th>Case PPV (%)</th>
<th>Control PPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Health</td>
<td>dementia</td>
<td>2532</td>
<td></td>
<td>397</td>
<td>1,160</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Marshfield</td>
<td>cataracts</td>
<td>4113</td>
<td></td>
<td>514</td>
<td>1,187</td>
<td>91</td>
<td>100</td>
</tr>
<tr>
<td>Mayo Clinic</td>
<td>peripheral arterial disease</td>
<td>3043</td>
<td></td>
<td>233</td>
<td>1,884</td>
<td>82</td>
<td>96</td>
</tr>
<tr>
<td>Northwestern</td>
<td>type 2 diabetes</td>
<td>1217</td>
<td></td>
<td>92</td>
<td>470</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>normal cardiac conduction</td>
<td>2712</td>
<td></td>
<td>81</td>
<td>352</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>All sites</td>
<td></td>
<td>13,617</td>
<td></td>
<td>1317</td>
<td>5053</td>
<td>92.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>98.5&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Genotype counts represent all subjects who were found by the hypothyroidism algorithms at each site and who were genotyped. Counts are limited to those classified as “white” in the electronic medical record of each site. PPV = positive predictive value.

<sup>a</sup> Average weighted for number of samples contributed to the total.

Denny et al., 2011
An eMERGE-wide phenotype analyzed with no extra genotyping: hypothyroidism

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

Denny et al., 2011
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 useable 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
GWAS: Target phenotype → association P value

Chromosomal location

The phenome-wide association study

PheWAS (ΦWAS): Target genotype → association P value

Diagnosis code

PheWAS requirement: A large cohort of patients with genotype data and many diagnoses
PheWAS for rs10759944 near FOXE1

\[ \text{OR}_{\text{GWAS}} = 0.74 \]
\[ \text{OR}_{\text{PheWAS}} = 0.76 \]

N = 13617 subjects

Denny et al., 2011
GWAS catalog: 1751 publications, 11,912 SNPs (Dec. 2013)

~2/3 of all evaluable and well-powered GWAS results replicated by PheWAS using ICD9 codes only in 13,835 eMERGE subjects

All PheWAS results for the GWAS catalog publically available at emrphewas.org

**PheWAS applications**

- Replicating genotype-phenotype associations
- Discovering pleiotropic gene effects
- Disease subsetting
- Drug repurposing
- Expanding our understanding of gene regulation
- Engaging basic scientists
eMERGE Network
electronic medical records & genomics

2011-2015: Phase II

Coordinating
Center

GEISINGER

MAYO CLINIC

GROUP HEALTH

VANDERBILT UNIVERSITY

MEDICAL CENTER
eMERGE-II goals

- Expand the electronic phenotyping library and apply to genotyped samples
- Initiate implementation of actionable variants into the EMR
  - Site-specific projects
  - Cross network initiatives

2011-2015: Phase II
+ pediatric sites
eMERGE-PGRN Partnership

PGx capabilities:
- CPIC guidelines
- Resequencing platform for 84 Very Important Pharmacogenes
- CLIA & QC standards

EMR-informatics capabilities
- Privacy
- Electronic phenotyping
- Large populations
- Decision support
eMERGE-PGx Pharmacogene sequencing project

- identify patients (n=9,000)
- identify “actionable” variants
- sequence (82 key pharmacogenes)

- implement actionable variants
- create a repository of all variants
eMERGE-III

• Expand the electronic phenotyping library and apply to genotyped samples
• Targeted sequencing of 100 selected genes in 25,000 subjects:
  • Return “Known or Expected Pathogenic” variants
  • Create a repository of other variants: assess EMR phenotypes, penetrance