The eMERGE Network

eMERGE WORKGROUPS
- Clinical Annotation
- Genomics
- PGx
- Return of Results/ELSI
- EHR Integration
- Outcomes
- Phenotyping

eMERGE SUPPLEMENTS
- Geocoding
- Health Care Provider Survey
- Phenotyping – OMOP Model

eMERGE SUBGROUPS
- Familial Implications of ROR
- Infobutton
- Participant Survey
- HLA
- ROR Legal Considerations
- Phenotype Variables

eMERGE PHASE III: SEPTEMBER 2015 – MAY 2019
eMERGE

E1 2007-2011
Can EMR and biobank be used for genomic research?
• Genome-wide genotyping
• GWAS

E2 2011-2015
Can genomic findings be applied in clinical care and how?
• Clinical implementation Pilots
• GWAS

E3 2015-2019
Can sequencing technology improve genomic discovery and clinical implementations?
• Sequencing
• Clinical implementation
• GWAS

ELSI Research
eMERGE III: What do we do?

SPECIFIC AIMS of the eMERGE Network

1. Sequence and assess clinically relevant genes presumed to affect gene function in about 25,000 individuals
2. Assess the phenotypic implications of these variants
3. Integrate genetic variants into EMRs for clinical care
4. Create community resources
Impact: 110k genomic dataset

- Data on over 110,000 participants and informatics tools with which to harness the data

- eMERGE Record Counter
  - Drag and drop demographics, phenotypes, ICD codes to obtain preliminary cohort counts

- SPHINX (Sequence and PHenotype INtegration Exchange)†
  - Search catalog by genes, drugs, rsID and pathways
  - For each gene view: SNVs, pathways, drug interactions
  - For each variant view: SNPid, category, frequencies
  - European, African, & Asian ancestry allele data

### Set Name | Platform | Count
--- | --- | ---
el-ell Merged* | GWAS | 83,717
Exome chip | Exome | 12,330
Whole exome | Sequencing | 3,745
PGx | Sequencing | 9,010
Whole genome | Sequencing | 1,800
eMERGEseq | Sequencing | 25,000

**Total Current** | **111,078**

**Total Expected** | **136,078**

*el-II: 55,029 samples; and eIII: 28,688 samples

Deliverable: Imputation and merging of eI-III GWAS data

- Imputation of all eMERGE array data against the HRC reference using the Michigan Imputation Server

- HRC reference contains 39,235,157 SNPs, no indels, provides access to rare variation (low as 0.1%)

- 83,717 individuals in data set released to network
  - Principle components analysis (PCA) examined ancestry
Deliverable: Development of an eMERGEseq Platform

- Clinical reports are generated on the “Consensus Actionable List” and any specific genes or SNVs requested by individual sites
- To date: **14,077** samples sequenced and **3,716** reports issued
Indication-based returnable results
(n=2531)

<table>
<thead>
<tr>
<th>Indication</th>
<th>Total</th>
<th>Positive</th>
<th>Negative</th>
<th>Inconclusive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorectal cancer/Polyps</td>
<td>1165</td>
<td>24</td>
<td>966</td>
<td>175</td>
</tr>
<tr>
<td>Ehlers-Danlos Syndrome</td>
<td>66</td>
<td>1</td>
<td>65</td>
<td>n/a</td>
</tr>
<tr>
<td>Abnormality of pain sensation</td>
<td>545</td>
<td>0</td>
<td>545</td>
<td>n/a</td>
</tr>
<tr>
<td>Pediatric migraine</td>
<td>443</td>
<td>0</td>
<td>443</td>
<td>n/a</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>258</td>
<td>12</td>
<td>246</td>
<td>n/a</td>
</tr>
<tr>
<td>Autistic Behavior</td>
<td>54</td>
<td>0</td>
<td>54</td>
<td>n/a</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2531</td>
<td>13</td>
<td>2319</td>
<td>175</td>
</tr>
</tbody>
</table>

* 1 report had an additional secondary finding

Non indication-based consensus returnable results**
(n=5268)

<table>
<thead>
<tr>
<th>Indication</th>
<th>Positive</th>
<th>Negative</th>
<th>Inconclusive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer/Tumor disease#</td>
<td>516</td>
<td>319</td>
<td>n/a</td>
</tr>
<tr>
<td>Cardiac</td>
<td>68</td>
<td>4949</td>
<td>n/a</td>
</tr>
<tr>
<td>Familial Hypercholesterolemia</td>
<td>66</td>
<td>545</td>
<td>n/a</td>
</tr>
<tr>
<td>Connective Tissue</td>
<td>n/a</td>
<td>545</td>
<td>n/a</td>
</tr>
<tr>
<td>Hemochromatosis</td>
<td>n/a</td>
<td>443</td>
<td>n/a</td>
</tr>
<tr>
<td>Malignant Hyperthermia</td>
<td>n/a</td>
<td>54</td>
<td>n/a</td>
</tr>
<tr>
<td>Clotting Disorder</td>
<td>n/a</td>
<td>54</td>
<td>n/a</td>
</tr>
<tr>
<td>Endocrine (Diabetes)</td>
<td>n/a</td>
<td>258</td>
<td>n/a</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2531</td>
<td>13</td>
<td>2319</td>
</tr>
</tbody>
</table>

Path/Likely Path Variants in CCHMC-adolescent cohort specific genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHEK2</td>
<td>1/10</td>
</tr>
</tbody>
</table>

** BMPR1A excluded in 1251 cases; CACNA1A, COL5A1, HNF1B, PALB2, POLE1, POLE excluded in 2653 cases; HNF1A, KCNE1, OT, BRCA1, BRCA2, MLH1, MLH2, MSH6, PMS2 excluded in 1402 cases as per site reporting requirements.

*1 report had an additional secondary finding

Partners-Broad Interpretation and Reporting: Review of 5268 cases

- Indication-based returnable results
  - Negative: 91.62% (n=2319)
  - Positive: 1.46% (n=37)
  - Inconclusive: 6.91% (n=175)

- Non indication-based consensus returnable results
  - Negative: 93.94% (n=4949)
  - Positive: 6.06% (n=319)

- Non indication-based site-specific returnable results
  - Negative: 90% (n=9)
  - Positive (n=1): 10%

*4 reports had two pathogenic variants. Skewed positive rate due to one site with sample selection based on suspicious genotype (16% positive).
## Interpretation & reporting: Baylor

### Indication based returnable results

<table>
<thead>
<tr>
<th>Indications</th>
<th>Total</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardiomyopathy</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Cardiac Arrhythmia</td>
<td>31</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>Hyperlipidemia (^a, b)</td>
<td>808</td>
<td>22</td>
<td>786</td>
</tr>
<tr>
<td>Colorectal Cancer</td>
<td>595</td>
<td>3</td>
<td>592</td>
</tr>
<tr>
<td>Breast/Ovarian Cancer(^c)</td>
<td>72</td>
<td>16</td>
<td>56</td>
</tr>
</tbody>
</table>

\(^a\) 298 patients had colorectal cancer and hyperlipidemia

\(^b\) Hyperlipidemia includes FH, hypertriglyceridemia, hyperlipidemia and coronary artery disease indications.

\(^c\) All returned genes belong to the 68 consensus except for CHEK2 in a breast cancer patient.

### Non indication based consensus returnable results

- Positive 3.2% (n=39)
- Negative 96.8% (n=1170)

### Non indication based site-specific returnable results

- Positive 3.1% (n=90) \(^a\)
- Negative 96.9% (n=2,803)

\(^a\) 1 patient had 2 variants

### Path and Lpath variants in NU and Vanderbilt specific returned

<table>
<thead>
<tr>
<th>Path and Lpath variants in NU and Vanderbilt specific returned</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHEK2</td>
<td>24</td>
</tr>
<tr>
<td>ATM</td>
<td>7</td>
</tr>
<tr>
<td>SERPINA1</td>
<td>3</td>
</tr>
<tr>
<td>MC4R</td>
<td>3</td>
</tr>
<tr>
<td>F11, FLG, KCNE2 (x1)</td>
<td>3</td>
</tr>
</tbody>
</table>

Others include MEFV, HNF1A, CACNA1A, OTC, LDLR
Impact: Electronic phenotyping & PheKB

• PheKB (Phenotype Knowledge Base)
  • Collaborative environment to building and validating electronic algorithms
  • Computational algorithm library
    • 37 finalized, public phenotypes

• Demonstrated feasibility of use in Genomic Medicine

• Tools and process allowed for computational and algorithm development cross collaboration around the world

September 11, 2017 Progress

Phenotype Development Workflow

Create
- Phenotype algorithm and data dictionary are in development
  - Share algorithm with project team
  - Standardize Phenotype Development
  - Standardize data collection

Validate
- Algorithm and Data Dictionary in review by validation site(s)
  - Share algorithm with validation team
  - Validate algorithm
  - Validate Data Dictionary

Share
- Share and implement algorithm and data dictionary for multi-site data collection
  - Validate Dataset against Data dictionary

Publish
- Phenotype published and Algorithm is sharable to public

Tool Support
- KNIME
- eMERGE RecordCounter
- PheKB
- Dictionary/Dataset Validation
- eMERGE RecordCounter
- PheKB
- PheKB
- PheKB
Phenotypes

PHENOTYPES: Development & Implementation

Phase 1
TOTAL PHENOTYPES

14

Phase II
TOTAL PHENOTYPES

29

Phase III
TOTAL PHENOTYPES*

27

70
Network Phenotypes Developed by Summer 2018

*Phenotypes in development
Impact: eMERGE PheWAS

• Developed methods for large scale genotype/phenotype analyses and implemented them across an entire collaborative Network

• Phenome-wide association studies (PheWAS)
  • 3144 SNPS present in NHGRI catalog (2012) in 13,835 individuals across 5 sites.
    • 1358 phenotypes analyzed for each SNP
  • Addition of Neanderthal PheWAS catalogue
  • Creation of Phecode mappings from ICD codes

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

Discovery Arm

Replication of novel associations

PheWAS for each SNP to discovery pleiotropy

3,144 SNPs

Denny et al, Nat Biotech 2013
Impact: eMERGE Pharmacogenomics (PGx)

• Multi-site test of the concept that genetic sequence information can be coupled to electronic medical records (EMRs) for use in healthcare

• Genetic sequencing on a 9010 participant data set
  • Sequencing and phenotype data available on SPHINX

• 82 pharmacogenetic genes investigated

• Many more opportunities for research on these data
  • PGx SNVs on the eMERGE-Seq panel

• Sites continue to collect utilization and outcomes data

Deliverable: PGRNseq multi-sample calling

- Original PGRNseq aligned to multiple references used by the original five sequencing centers

- All 9010 BAMs re-aligned to the same genome reference hs37d5.fa

- 9010 individuals in data set provided to the network for analysis

- Principle components analysis (PCA) examined ancestry
Impact: Return of genomic data via EMR

• Infrastructure and tools, in particular decision support tools, to enable genomic medicine

• InfoButton*
  • Explored use of infobuttons as a decision support tool to provide context specific links within the electronic health record (EHR) to relevant genomic medicine content
  • Assessed the coverage of content topics among information resources developed

• CDS_KB (Clinical Decision Support KnowledgeBase)
  • Partnership with IGNITE network
  • Goal is to catalog and share CDS implementation artifacts and design considerations for genomic medicine programs from a broad community of institutions

**Impact: Network wide analyses ‘DNAnexus’**

- Utilization for sharing and managing genetic data in a cloud-based system

- Network seminar series demonstrating utility of the analysis pipeline and development of apps
  - Large scale analyses possible for all investigators, regardless of local computing power

- DNAnexus houses Network wide genetic datasets
  - GWAS
    - Including a subset of geocoded samples
  - PGRNseq
  - eMERGEseq

Number of Published Projects Through August 2017

<table>
<thead>
<tr>
<th>Year</th>
<th>Site Projects</th>
<th>Network Projects</th>
<th>Total Projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-2011</td>
<td>13</td>
<td>67</td>
<td>633</td>
</tr>
<tr>
<td>2013</td>
<td>26</td>
<td>88</td>
<td>114</td>
</tr>
<tr>
<td>2015</td>
<td>34</td>
<td>91</td>
<td>125</td>
</tr>
<tr>
<td>2017</td>
<td>17</td>
<td>73</td>
<td>90</td>
</tr>
<tr>
<td>2017</td>
<td>10</td>
<td>9</td>
<td>119</td>
</tr>
</tbody>
</table>

633 Total Projects
Citations of eMERGE Publications by Category

Cumulative Citation Counts: 17,115 (2007-March 2017)
**Impact: dbGaP & Website analytics 2007-2017**

Data Reuse: # Downloads of e MERGE dbGaP Submissions as of August 2017

- **eMERGE Website**
  - Average usage past 6 months
  - 63.1% new visitors
  - 1596 sessions/month
  - 1043 users/month
  - Views from 96 countries

- **PheKB Website**
  - Average usage past 6 months
  - 56.2% new visitors
  - 1171 sessions/month
  - 540 users/month
  - Views from 76 countries

> **1100 external downloads as of August 2017**
eMERGE Tools

- **PheKB**: A knowledgebase for discovering phenotypes from electronic medical records.
- **MyResults.org**: An informational tool for educating patients about genetic test results.
- **SPHINX**: A data exploration tool for genetics-related drug response hypothesis generation.
- **Infobutton Project template**
- **eMerged Model Consent Language**
- **PheWAS catalog**

**Additional Tools**

- **GENOTYPING tools**
- **PHENOTYPING tools**
- **CDS tools**
- **Natural Language Processing (NLP) Tools**
eMERGE III: Future Deliverables

- dbGaP submissions
  - GWAS eI-III imputed set (*ready to submit*)
  - Interim (Fall 2017) and final eMERGEseq data

- Return of clinical results and EHR integration at all sites
  - Establishment of IT support for return of results processes based on data delivered through the network
  - Analysis of solutions, challenges and lessons learned
    - Manuscripts and methods documentation of Network-wide efforts
    - Sharing with CDSKB and standard bodies as appropriate

- Outcomes analysis for effect of return of results on patients and providers across sites
  - Compare differences in health outcomes and provider behaviors for return of negative and positive results
  - EHR and survey based methods for examining patient impacts and changes in care or awareness by providers

- Creation and deployment of 27 phenotypes, 8 deployed to date
  - 25 eI-III imputed GWAS
  - 13 PGRNseq
  - 24 eMERGEseq
Questions??
## eMERGE Geocoding

<table>
<thead>
<tr>
<th>Factors</th>
<th>Source</th>
<th>Resolution</th>
<th>National/Local</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td>Coordinating Center/Site EDW</td>
<td>Patient Level</td>
<td>National</td>
</tr>
<tr>
<td>SES</td>
<td>Census/ACS</td>
<td>Block Group Level</td>
<td>National</td>
</tr>
<tr>
<td>Built Environment</td>
<td>RUCA (rural-urban-commuting-area-codes)</td>
<td>Tract Level</td>
<td>National</td>
</tr>
<tr>
<td>Traffic Volume</td>
<td>Google?</td>
<td>Block Group Level</td>
<td>National</td>
</tr>
<tr>
<td>Road Density</td>
<td>ArcGIS shapefiles</td>
<td>Block Group Level</td>
<td>National</td>
</tr>
<tr>
<td>Food Accessibility</td>
<td>Food Environment Atlas (USDA Economic Research Service)</td>
<td>County Level</td>
<td>National</td>
</tr>
<tr>
<td>Water Quality</td>
<td>NURE-HSSR database; Enviromapper?</td>
<td>Various</td>
<td></td>
</tr>
<tr>
<td>Density of Parks</td>
<td>ArcGIS shapefiles</td>
<td>Block Group Level</td>
<td>National</td>
</tr>
<tr>
<td>Walkability</td>
<td>Walk Score Professional</td>
<td>Zip Code</td>
<td>National</td>
</tr>
<tr>
<td>Entropy Index</td>
<td>Census/ACS</td>
<td>Block Group Level</td>
<td>National</td>
</tr>
<tr>
<td>Crime</td>
<td></td>
<td></td>
<td>Local</td>
</tr>
<tr>
<td>Hospital Utilization</td>
<td>AHRF, HHS, HRSA</td>
<td>County Level</td>
<td>National</td>
</tr>
</tbody>
</table>
CSER, eMERGE, and IGNITE

- Clinical Sequencing Evidence Generating Research (FY2017-2020)
- Electronic Medical Records and Genomics (FY2015-2018)
- Implementing Genomics in Practice (FY2018-2022)
Commonalities and Complementarity of CSER and eMERGE

**CSER (FY2017-2020)**
- ~4,600 pts, 6 sites
- Community clinical scenarios
- Focus: clinical encounter
- Increased ethnic and socioeconomic diversity
- Evidence generation for clinical utility of genomic sequencing
- Real-world barriers to integrating genomic data for healthcare utilization

**eMERGE (FY2015-2018)**
- 25K pts, 9 sites
- Electronic phenotyping
- Focus: system-wide
- Health outcomes of rare variants in ~100 clinically relevant genes
- System-wide impact of reporting actionable variants
- Improved e-phenotyping
- Novel variant discovery
- Electronic CDS

- EMR integration
- Clinical impact of RoR
- Data sharing concerns
Commonalities and Complementarity of eMERGE and IGNITE

**eMERGE (FY2015-2018)**
- 25K pts, 9 sites
- Electronic phenotyping
- Focus: system-wide
- Health outcomes of rare variants in ~100 clinically relevant genes
- System-wide impact of reporting actionable variants
- Improved e-phenotyping
- Novel variant discovery
- Electronic CDS

**IGNITE (FY2018-2022)**
- ~15K pts, 4-6 sites
- Diverse, real-world clinical settings
- Focus: pragmatic trials
- Clinical utility of established genomic medicine interventions
- Increased ethnic and socioeconomic diversity
- Generalizable knowledge on use of trials in genomic medicine interventions

**Commonalities**
- EMR integration
- Cost-effectiveness
- Patient/clinician education

**Complementarity**
Timeline of NHGRI Genomic Medicine Programs

Programs: eMERGE, CSER, IGNITE, UDN (CF), UDN (HD), ClinGen, NSIGHT

Fiscal Year

- Renewal Phases
- Projected Phases

You are here: Sept 2017
Resources

emerge network
ELECTRONIC MEDICAL RECORDS & GENOMICS

www.gwas.org

Manuscripts (to date)
https://emerge.mc.vanderbilt.edu/publications/

dbGaP (published to date)
https://emerge.mc.vanderbilt.edu/dbgap/

GWAS sequencing platforms (eI-III)
https://emerge.mc.vanderbilt.edu/wp-content/uploads/2015/02/Platform-Information-eMERGE.docx

TOOLS

PheWAS Resources
https://phewascatalog.org/

PheKB
a knowledgebase for discovering phenotypes from electronic medical records
https://phekb.org/

CDS_KB
CLINICAL DECISION SUPPORT KNOWLEDGEBASE
https://cdskb.org/

SPHINX
A resource of the eMERGE Network
https://www.emergesphinx.org/