**Electronic medical record**

The Electronic medical record and genomics (eMERGE) network review1

**NHGRI, March 2007, Phase I [eMERGE-I]-Vanderbilt University**

Five sites:

“In September 2007, grants were awarded to five sites (hereaf- ter referred to as eMERGE-I)—Group Health Cooperative/University of Washington, Marshfield Clinic, Mayo Clinic, Northwestern University, and **Vanderbilt University**, which also served as the network’s **coordinating center**.”  
eMERGE-II (9 sites):

CHOP [Children’s hospital of Philadelphia], GHC & U Wash, Mount Sinai, CCHMC & BCH [Cincinnati Children’s Hospital Medical Center with Boston Children’s Hospital], Marshfield & Essentia, Northwestern, Geiginger, Mayo, Vanderbilt

Three major aims:

1 EMR for robust electronic phenotyping;

2 conduct GWAS using the phenotypes derived in the above-mentioned first aim; PheGWAS2

3 explore the ethical, legal, and social implications associated with EMR-based GWAS and wide-scale data sharing.

The workgroups included:  
an informatics group [published a phenotyping algorithm, PheKb: Phenotype KnowledgeBase, 20http://www.PheKb.org]. Developed by the first site and deployed to the second, and then across the network.   
a genomics group [imputation, strand issue, relatedness, ],   
and a consent and community consultation group [model language for EMR-linked ].

**Lesson learned from Phase I**

1 Phenotyping across the network rather than local sites.

2 Most eMERGE Participants have consented to contributing their data to health research of any kind. Through network-wide projects, eMERGE-I was compelled to develop best practices for sharing genomic data and EMR-derived phenotypes while protecting the privacy of participants.

3 Returing of results (RoR) 3. Turner syndrome, Klinefelter syndrome, FVL, hereditary hemochromatosis. There is no clear agreement how the finding should be returned. In some case, the clinically actionable results identified in EMR was already known. “In some case, health records shed light on ambiguous findings, increasing the likelihood that some findings represented acquired rather than congenital genetic changes.”

**Phase II [eMERGE-II]**

Phenotyping workgroup

Genomics workgroup

RoR workgroup

Consent, education, regulation, and consultation workgroup: evaluating the impact of returning hemochromatosis results

EMR integration workgroup: PGx [pharmcogenomic] pilot project

Collaboration with external groups

**Lessons**

1 Portability of electronic phenotypes within and outside eMERGE

“There is currently no formal phenotyping language for the purpose of building EMR phenotyping algorithms nor is their a common approach to their implementation.

2 Approaches to EMR integration of genomic information

3 Integration of pediatric sites

4 Longitudinal cost-effective genomic medicine discovery and implementation

5 Generalized framework for the return of genomic results

6 The eMERGE network in the context of a translational framework.   
T0: early phases focusing of biologic discoveries (eMERGE-I focused large on T0)  
T1: development of candidate health applications (eMERGE-II)

T2: assessing outcomes of intervertions

T3: move genomic findings into health practice.

T4: public health surveillance

**Application case**

Balanced crystalloids与生理盐水Saline一直在临床使用上有争议，但差异可能比较小，而要组织能够区分这类差异的临床试验比较困难。之前两者的差异一直基于生理角度分析，而EMR的出现，让大规模基于病人的循证医学成为一个选项。

**Case 1:** Balanced crystalloids versus saline in the intensive care unit: The SALT randomized trial4. Data 2015/Feb/3~2015/May/31

**Case 2:** Balanced Crystalloids versus Saline in critically Ill Adults5. 2015/Jun/1~2017/4/30.

In particular, the design and the protocol of the study was published in BMC Trials6. Screenplay -> movie.

在这个Vanderbilt基于eMERGE之前的标杆研究是新西兰多中心的SPLIT RCT研究7—收集了1000：1000的buffered crystalloid 和 saline，当没有显示差异。

Vanderbilt的这个研究展示了一个比较完整的pre-study，design，到大规模实验的流程。

在prestudy中4，使用了2015/Feb/3-2015/May/31的数据，454saline，520 balanced crystalloids。也就是说，仅仅基于Vanderbilt ICD三个月的EMR数据量就赶上了新西兰的研究水平。

实验设计的具体流程则在Trials上发表6，最终发表了基于2015/jun~2017/April的数据，样本量达到15802(7942 balanced-crystalloids group, 7860 saline group)，也就是case 2。

**Case 3:** Balanced crystalloid versus saline in noncritically ill adults8.

**Geisinger**9 **[The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research]**

**1996 started to collect EMR**

**2007 GHS launched a project now known as MyCode Community Health Initiative to create a system-wide biorepository of blood serum and DNA samples for broad research used, including genomic analysis.**

**In their demo, the loss-of-function *APOC3* was tested. *APOC3* had strong effect in lipd values**10

**GERA [Genetic Epidemiology Research on Adult Health and Aging cohort]**11

About a cohort of > 100,000 subjects who are participants in the Kaiser Permanente Medical Care Plan, Northern California Region. 20% African Africa, Asian, and Latino or mixed, 80% non-Hispanic white.

Affymetrix Axiom Genotyping chip designed by GERA12,13.

Saliva-based genotyping experiments, Oragene OG-250 disc format saliva kit.

Population genetics study analysis for GERA found about 3400 Chinese samples14.

GERA also measure Telomere Length15. It was pretty petty because mDNA provided a much accurate measure for aging compared with Telomere length16.

**Partners Healthcare System (**<https://www.partners.org/About/Default.aspx>**)**

**Founding members**

Brigham and Women’s Hospital, Massachusetts General Hospital

International code of disease [<https://www.who.int/classifications/icd/en/>]

Natural Language Processing and EMR

Open source NLP cTAKES <http://ctakes.apache.org/>

Unified Medical Language System (UMLs)17 <https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/>

Compared with using ICD code in billing data alone for treatment resistant depression as a model18.

Using NLP in EMR to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem19.

**Reference**

1. Gottesman, O. *et al.* The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet. Med.* **15**, 761–71 (2013).

2. Denny, J. C. *et al.* PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* **26**, 1205–1210 (2010).

3. Fullerton, S. M. *et al.* Return of individual research results from genome-wide association studies: Experience of the Electronic Medical Records and Genomics (eMERGE) Network. *Genet. Med.* **14**, 424–431 (2012).

4. Semler, M. W. *et al.* Balanced crystalloids versus saline in the intensive care unit: The SALT randomized trial. *Am. J. Respir. Crit. Care Med.* **195**, 1362–1372 (2017).

5. Semler, M. W. *et al.* Balanced Crystalloids versus Saline in Critically Ill Adults. *N. Engl. J. Med.* **378**, 819–828 (2018).

6. Semler, M. W. *et al.* Balanced crystalloids versus saline in the intensive care unit: Study protocol for a cluster-randomized, multiple-crossover trial. *Trials* **18**, 129 (2017).

7. Young, P. *et al.* Effect of a buffered crystalloid solution vs saline on acute kidney injury among patients in the intensive care unit: The SPLIT randomized clinical trial. *JAMA* **314**, 1701–1710 (2015).

8. Self, W. H. *et al.* Balanced Crystalloids versus Saline in Noncritically Ill Adults. *N. Engl. J. Med.* **378**, 819–28 (2018).

9. Carey, D. J. *et al.* The Geisinger MyCode community health initiative: An electronic health record-linked biobank for precision medicine research. *Genet. Med.* **18**, 906–913 (2016).

10. The TG and HDL Working Group of the Exome Sequencing Project. Loss-of-Function Mutations in APOC3, Triglycerides, and Coronary Disease. *N. Engl. J. Med.* **371**, 22–31 (2014).

11. Kvale, M. N. *et al.* Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1051–1060 (2015).

12. Hoffmann, T. J. *et al.* Next generation genome-wide association tool: Design and coverage of a high-throughput European-optimized SNP array. *Genomics* **98**, 79–89 (2011).

13. Hoffmann, T. J. *et al.* Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics* **98**, 422–430 (2011).

14. Banda, Y. *et al.* Characterizing Race/Ethnicity and Genetic Ancestry for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1285–1295 (2015).

15. Lapham, K. *et al.* Automated Assay of Telomere Length Measurement and Informatics for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *Genetics* **200**, 1061–1072 (2015).

16. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).

17. Bodenreider, O. The Unified Medical Language System (UMLS): Integrating biomedical terminology. *Nucleic Acids Res.* **32**, 267–270 (2004).

18. Perlis, R. H. *et al.* Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. *Psychol. Med.* **42**, 41–50 (2012).

19. Zhong, Q.-Y. *et al.* Use of natural language processing in electronic medical records to identify pregnant women with suicidal behavior: towards a solution to the complex classification problem. *Eur. J. Epidemiol.* **34**, 153–162 (2019).