National Advisory Council for Human Genome Research

May 19, 2014

Concept Clearance for RFAs

eMERGE (Electronic Medical Records and Genomics) Phase III

Purpose

NHGRI proposes three RFAs to continue genomic medicine discovery and implementation research using large biorepositories linked to electronic medical records (EMRs) in eMERGE Phase III. This effort will build upon the genome sequencing, phenotyping, and implementation methods and knowledge generated in eMERGE II and related programs to pursue a broad-based program sufficiently large to define health outcomes associated with rare (0.5-1.0% MAF) variants in ~100 clinically relevant genes. Specifically, this initiative proposes to: 1) detect rare variants presumed to affect gene function; 2) assess the phenotypic implications of these variants, singly or in combination, by leveraging well-validated EMR data; 3) report actionable variants to patients and clinicians to improve clinical care and ultimately health outcomes; and 4) assess health impact, cost-effectiveness, and ethical, legal and social (ELSI) implications of reporting these variants on a broader population scale for patients, clinicians, and institutions. It will also continue efforts to improve electronic phenotyping, provide electronic clinical decision support, and enable integration of genomic information into EMRs for clinical research and care.

Background

eMERGE was initiated in 2007 to develop methods and best practices for conducting genomic research in 5 biorerepositories linked to EMRs. Having demonstrated the robustness of EMR phenotyping for genome-wide studies, defined approaches for enhancing privacy of shared EMR data, and engaging patients and communities in consent and data sharing, eMERGE expanded to include 7 clinical sites in 2011 and 2 pediatric sites in 2012. Its research scope also expanded to include NHGRI's first collaborative clinical implementation efforts, returning genotypic information to patients, their EMRs, and their clinicians to assess feasibility and impact on clinical care. Initial findings focusing on 5 drug-gene interactions have identified one or more actionable variants in 91% of genotyped patients and have demonstrated the cost-effectiveness of pre-emptive vs. reactive testing.

The substantial infrastructure, datasets, and tools developed by eMERGE to date, combined with its rare (if not unique) experience in EMR data-mining and clinical implementation, position eMERGE well to continue its three-pronged emphasis on discovering clinically relevant variants, assessing impact of large-scale implementation on cost and quality of care, and enabling both discovery and implementation research in other biorepositories. In 2012 eMERGE began a collaboration with the Pharmacogenetics Research Network (PGRN) to perform targeted sequencing of 84 PGx genes in 9,000 patients. The large scale of this project has been especially illuminating, not only in the processes for consent, clinical workflow, and approval at 10 diverse institutions, but also in the number of potentially actionable variants found. Over 2% of the first 2,022 patients studied, for example, carry rare "known or expected pathogenic variants" in two arrhythmia genes, *SCN5A* and *KCNH2*; yet we know that familial arrhythmia syndromes are much less prevalent than this. eMERGE is uniquely positioned to define the phenotypic correlates of such variants and assess their penetrance through EMR data-mining in large clinical samples. This information will be critical in moving genome sequencing into wide clinical use, as institutions are increasingly concerned about their obligations for following up such results. Without accurate estimates of penetrance and pathogenicity to target feedback

only to patients truly at high risk, institutional responsibilities for curating, counseling, and following up these variants will not be sustainable.

Currently, eMERGE II has 328,895 participants with biorepository samples linked to EMRs; 105,524 of these samples have been genotyped (Table). eMERGE II has expanded its e-phenotyping library to 41 validated phenotypes and has continued GWAS and phenome-wide association studies (PheWAS, which it helped to originate) for discovery research using common variants. Widely available tools developed by eMERGE include PheKB, a phenotype algorithm repository; the eMERGE Record Counter for feasibility assessment; Natural Language Processing (NLP) tools such as cTakes and MedEx; the eMERGE InfoButton pointing physicians to clinical decision support (CDS) resources; the PGx variant and phenotype data repository (SPHINX); and MyResults.org, an online educational resource on genetic testing for patients. In 2013 eMERGE was selected to survey ~16,000 patients' attitudes toward proposed modifications to the Common Rule on broad consent for genomics research, information that is urgently needed NIH-wide for policy development. Through April 2014, eMERGE has published 60 network and 192 site-specific projects and has been cited over 2,100 times.

While challenges remain in integrating genomics and CDS into EMRs, and in navigating diverse approaches to implementing genomic medicine across institutions, eMERGE as a whole may be sufficiently prominent to begin to engage major EMR vendors in providing effective tools for incorporating genomic data and facilitating their use in clinical care and research. Critical issues in genomics and EMRs that eMERGE is well-positioned to address, and to encourage vendors to implement, include facilitating confidential, standardized, and efficient genomic data sharing across providers; developing and assessing effectiveness of genomic CDS; harnessing genomic EMR data for quality improvement research around use of genomic tests; and enhancing the usefulness of genomic EMR data for patient education, self-management, and identification of at-risk family members. Such issues must be addressed if genomic medicine is to be widely implemented and successful in improving outcomes and reducing healthcare costs.

Proposed Scope and Objectives

This initiative will support 8-12 study sites, a coordinating center, and 1-2 genome sequencing and genotyping facilities. eMERGE Phase III will conduct research in both genomic discovery and clinical implementation; share expertise and experience within and outside eMERGE; and disseminate association findings, tools and best practices to the scientific community.

In eMERGE Phase III, 2,000-3,000 DNA samples per site (total n~25,000) will be sequenced at the CLIA-certified eMERGE central sequencing and genotyping facility(-ies). Within the first year, the eMERGE Steering Committee will agree upon genes to be sequenced and types of sequencing to be done (targeted genes, exome or genome). Sequence variants with presumed detrimental impact on gene function will be analyzed separately and collectively (by gene, pathway, or other logical method of clustering) for associations with clinical and psychosocial outcomes, health behaviors following testing, and attitudes toward testing. Exploratory functional studies capitalizing on the unique phenotyping strengths of eMERGE, including the potential for re-contact and re-examination, may be supported as funds permit, especially those leveraging consortia such as NHGRI's ENCODE and GTEx.

The phenotyping library will be expanded from 41 phenotypes in Phase II to up to 80 in Phase III, focusing on phenotypes likely to be related to the genes to be sequenced. Continued enhancements to phenotyping tools, including "modular" phenotypes combinable into multiple complex phenotypes, are expected to increase phenotyping efficiency and transportability.

eMERGE III will develop and implement institution- or system-wide methods for re-annotation of variants as knowledge accrues, as well as methods for re-contact of clinicians and patients with

updated information on actionability. These challenging situations will permit investigation of the impact of re-annotation on attitudes of clinicians and patients toward genetic testing and their behaviors in response. Integration of genomic findings into EMRs will facilitate system-wide tracking of and feedback on clinician errors such as incorrect or redundant ordering or inadequate follow-up, as well as other quality improvement efforts, thus bringing learning healthcare systems to genomics and vice versa. Patient portals and related tools will facilitate patient education and self-management and explore challenges involved in identifying at-risk family members. eMERGE III will continue to engage stakeholders such as IRBs, health systems' clinical leaders, administrators, and EMR vendors on improving integration of genomic data for clinical care and genomic research. eMERGE III sites will also be expected to explore system-wide legal and economic implications of genomic medicine implementation, such as genomic sequencing-based treatment decisions and institutional impact of returning results.

Criteria for selecting eMERGE III sites, whether new or existing, will include population diversity, particularly for ancestral groups under-represented in genomic studies to date; documented availability at the time of award of high-quality GWAS data in at least 3,000 patients with accessible EMRs and at least 2,000 DNA samples for CLIA sequencing and return of results; consent for sharing individual-level data through dbGaP and for returning results; completeness and usability of EMR data; and ability to implement existing eMERGE phenotypes and protocols reliably. Applicants will be expected to include expertise in sequencing, EMR phenotyping and integration, informed consent and genetic counseling, assessment of clinical and psychosocial outcomes, health administration, legal implications, and health economics. Applicants new to eMERGE with unusual strengths in population diversity or in these key areas of expertise will be encouraged to apply; those with smaller biorepositories will be encouraged to consider partnering with other sites. Existing sites will also be assessed on ongoing productivity and collaborative performance in eMERGE II.

Relationship to Ongoing Activities

eMERGE III's large size and focus on EMR phenotyping positions it well to assess clinical manifestations of sequence variants in the large numbers of patients needed for reliable estimates of association, and to provide that evidence to ClinGen to support consensus decisions on actionability. In contrast to the very deep phenotyping and/or genome sequencing possible in smaller, more patient-focused programs such as UDN, NSIGHT, and CSER, eMERGE will combine less extensive sequencing with a breadth of EMR phenotypes and institution- or system-wide implementation of return of results and CDS. eMERGE III implementation models and EMR integration methods will provide further best practice approaches testable for real-world dissemination in routine clinical settings within programs like the Implementing Genomics in Practice (IGNITE) consortium. All six projects will participate in NHGRI-wide working groups related to reporting results, EMR integration, and CDS, and in a program-wide review of NHGRI genomic medicine programs being planned for mid-2015 by NHGRI's Genomic Medicine Working Group.

Mechanism of Support

The NIH U01 (Cooperative Agreement) mechanism will continue to be used. Three RFAs will be issued for 8-12 study sites, 1 coordinating center, and 1-2 central genome sequencing and genotyping facilities.

Funds Anticipated:

NHGRI will commit roughly $60M over four years based on current consortium costs and targeted sequencing charges: $44M for the study sites and coordinating center, and $16M for the genome sequencing and genotyping facilities.

Table. eMERGE Phase II biorepositories, EMR characteristics, and study samples.

Institution	Repository size; Ancestry	GWA study size (% Female)	Selected Network Phenotyping led by site
Children's Hospital of Philadelphia	60,000; 38% AA	42,290 (50% female)	Asthma, atopic dermatitis, ADHD, lipids
Cincinnati Children's Hospital Medical Center; Boston Children's Hospital	Combined 10,000; 9.8% AA, 3.3% HL	5,360 (41.9% female)	Autism, appendicitis, childhood obesity
Geisinger Health System	22,000; 99.4% EA	4,191 (47.5% female)	AAA, extreme obesity and related conditions, remission of diabetes after bariatric surgery
Group Health, University of Washington	6,381; 3.7% AA	3,606 (57% female)	Dementia, C difficile diarrhea, Herpes zoster, carotid artery atherosclerosis disease
Marshfield Clinic	20,000; 99% EA	4,693 (58.4% female)	Cataracts, glaucoma, ocular HTN, age related macular degeneration, dry eye
Mayo Clinic	19,000; 93.5% EA	6,934 (38% female)	Peripheral arterial disease, RBC indices, CHD, VTE, cardio respiratory fitness, heart failure
Mount Sinai School of Medicine	22,000; 24.1% EA, 30.6% AA, 43.8% HL	6,290 (52.4% female)	Diabetic hypertensive chronic kidney disease, rapid declines in renal function, CHD, drug-induced liver injury
Northwestern University	11,000; 67.7% EA, 19.7% AA, 4.8% Asian	4,987 (83% female)	Type 2 diabetes, diverticulosis, lower GI non-syndromic polyp, Methicillin-resistant Staphylococcus aureus (MRSA)
Vanderbilt University	158,514; 56% EA, 34% AA	27,173 (58.9% female)	QRS duration, hypothyroidism, resistant hypertension, ACE-inhibitor cough, statins for MI prevention

Key: EA: European Americans **AA:** African Americans **HL:** Hispanic/Latino