

# Lessons learned from the eMERGE Network: balancing genomics in discovery and practice

eMERGE Consortium\*

## Summary

The Electronic Medical Records and Genomics (eMERGE) Network, established in 2007, is a consortium of academic and integrated health systems conducting discovery and implementation research in translational genomics. Here, we outline the history of the network, highlight major impacts and lessons learned, and present the tools and resources developed for large-scale genomic analyses and translation into a clinical setting. The network developed methods to extract phenotypes from the electronic medical record to perform genome-wide and phenome-wide association studies. Recruited cohorts were clinically sequenced off a custom panel for targeted sequencing of variants and monogenic disease risks and returned to participants to investigate the impact of return of genomic results. After generating a 105,000 participant-imputed genome-wide association study (GWAS) dataset for discovery, the network enrolled and sequenced 24,998 participants. Integration of these results into the medical record and the effects of results on participants provided key lessons to the field. These learned lessons inform genetic research in diverse populations and provide insights into the clinical impact of return and implementation of genomic medicine using the electronic medical record. The lessons produced by the eMERGE Network can be utilized by other consortia as translational genomic medicine research evolves.

The utilization of electronic medical record (EMR) data for clinical research has been an integral component of translational science over the last 15 years.<sup>1,2</sup> The ability to pair clinical data with biobank samples and conduct large-scale genome-wide association studies (GWASs) has paved the way for translational research to provide insights into diabetes,<sup>3</sup> cataracts,<sup>4</sup> cardiovascular disease,<sup>5</sup> and obesity,<sup>6</sup> among many others.<sup>7,8</sup> The Electronic Medical Records and Genomics (eMERGE) Network has pioneered discovery research methods using longitudinal EMR data linked to genotyping and sequence data<sup>9–13</sup> across diverse geographical, racial, and age distributions. To date, the network has produced a merged, imputed, multi-sample genotyping file representing data from 105,000 participants recruited across three phases<sup>14–16</sup> to investigate genetic associations with disease phenotypes.<sup>16–18</sup> In later phases, the scope of the network expanded to include the clinical applications of genetics. Diverse sites, including pediatric and adult academic medical centers, integrated health systems, and community-based clinics, have sequenced clinically relevant portions of the genome and returned actionable results using the EMR. This diversity provided a natural experiment to study differences and aggregate lessons learned for delivering both sequencing and pharmacogenomic data to diverse populations. As a result, the network has contributed to research in clinical genomics,<sup>14,15</sup> pharmacogenomics,<sup>19</sup> phenotyping of clinically relevant diseases,<sup>20</sup> clinical annotation,<sup>21</sup> return of results (RoRs),<sup>22</sup> and assessment of clinical outcomes.<sup>23</sup> The network's work on establishing methods for transmitting genetic test results from laboratories into heterogeneous health care provider organizations and into clinical practice has helped the network activities span discovery and patient care.<sup>24</sup> This paper describes how the network was structured to achieve clinical implementation and ongoing discovery-based research and provides an overview of developed tools, lessons learned, and resources available to other researchers.

The initial goals of the eMERGE Network established the methods for developing and validating electronic pheno-

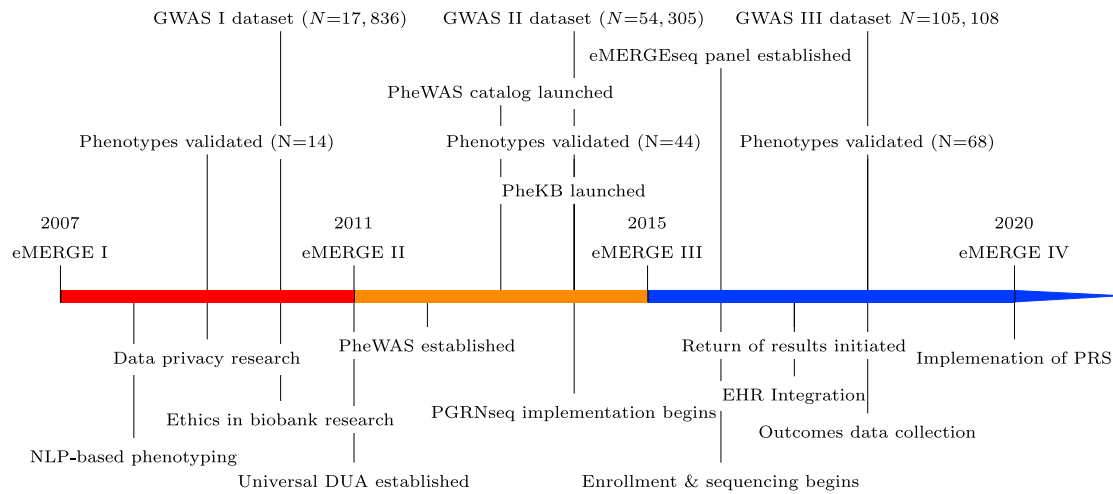
typing algorithms across multiple sites and medical record systems. The network also created a pipeline for combining GWAS data from 18,663 participants across five sites to create an imputed, multi-sample file for discovery research.<sup>13,15</sup> The PheWAS (phenome-wide association scan) method was originally developed by Denny et al.<sup>25</sup> for conducting disease-gene associations using billing codes to automatically categorize over 700 disease phenotypes. The PheWAS Catalog was launched in 2013 and cited over 300 times,<sup>26</sup> making it one of the most highly utilized tools developed by the network. In August 2011, the second phase of eMERGE began with the goal of advancing translational efforts. This was achieved by incorporating available genotyping data with electronic phenotyping and privacy protection methods into clinical research and ongoing clinical care. While eMERGE I focused on discovery in phenotyping and genomics, eMERGE II shifted to returning clinically relevant findings as sites recruited, sequenced, and returned results from the targeted Pharmacogenomics Research Network sequence platform (PGRNseq).<sup>27</sup> These data combined variant information with drug interactions, which led to the development of Sequence and Phenotype Integration Exchange (SPHINX), a web-based portal for exploring data for hypothesis generation, especially around drug response implications of genetic variation across the eMERGE Network.<sup>28</sup> This component of eMERGE II enabled the network to develop processes for the RoRs and EMR integration and to assemble a diverse set of participants, of which approximately 20% was non-European ancestry.

Phase III of the eMERGE Network had four aims: (1) sequence and assess variants in targeted clinically relevant

\*Correspondence: [david.crosslin@gmail.com](mailto:david.crosslin@gmail.com)  
<https://doi.org/10.1016/j.xhgg.2020.100018>.

© 2020 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).





**Figure 1. Timeline and impact of the eMERGE Network**

The network has produced 68 clinical phenotypes validated across multiple electronic medical record systems, launched tools focused on the reuse of genomic data, created multiple iterations of a large GWAS imputed dataset culminating with 105,108 participants, and sequenced and returned results off a PGRNseq and eMERGEseq custom sequencing platform.

genes in approximately 25,000 participants; (2) assess the phenotypic associations of these variants; (3) integrate clinical reports of actionable genetic variants into EMRs for clinical care; and (4) create community resources. Two sequencing centers joined the network of ten clinical sites (Figure S1), which developed, implemented, and returned results from a network-specific sequencing panel, eMERGEseq, for use in clinical care.<sup>21</sup> As the network transitioned from discovery-based research to implementation, protocols focused on sequencing and return of clinically relevant findings were developed.<sup>29</sup> The accomplishments of eMERGE over the past decade (Figure 1) along with the transition to RoRs over the last several years has allowed the network, which has focused on genomic research and pragmatic interventions, to facilitate the integration of research findings into real-world clinical settings (Figure 2).

### Approach and methodology

#### *Achieving network goals through milestone-driven workgroups*

During eMERGE phase III, seven workgroups with specific goals and milestones were developed.

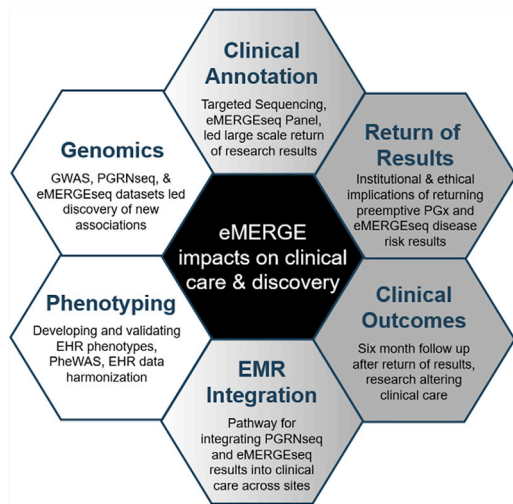
1. Clinical annotation: create consistent approaches for gene and variant interpretation across the eMERGE sequencing centers and clinical sites and support contributions to public knowledge bases;
2. EMR integration: integrate the clinical and genomic results data produced during phase III into electronic records at the clinical sites;
3. Genomics: identify best practices and facilitate analyses of common and rare variant data in previous and current phases of eMERGE;
4. Pharmacogenomics: coordinate and promote pharmacogenomic discovery and implementation work

across eMERGE utilizing previous and current pharmacogenetic (PGx) data;

5. Phenotyping: advance the science of phenotype development, including case, control, and cohort definition, and the development and implementation of 25 new disease phenotypes and incorporate natural language processing (NLP);
6. Outcomes: develop cross-site outcomes assessment to track implementation and impact of eMERGE III sequencing and assess the impact on health care utilization and outcomes of importance for participants and their families; and
7. RoRs and ethical, legal, and social implications: develop best practices for the return of actionable variants and measure the impact this return has on participants, families, health care providers, and health care systems.

#### *Data sharing, utilization, and the evolution of data security*

The notions of privacy and security are critical to supporting genome-based research and clinical care. When eMERGE began, it established protocols for collecting and sharing de-identified data within the network to facilitate investigations and ensure equity of access. These guidelines have become a critical element of success of the eMERGE Network. While many of the participants consented to participate in the study and community consultation was performed,<sup>9</sup> the process was refined as participants shifted from passive contributors of specimens and data in phase I to recipients of clinically actionable information in phases II and III.<sup>22</sup> At the same time, eMERGE aimed to support the NIH's goal of making data from genome-based investigations publicly available. It was understood that the uniqueness of certain types of data utilized by eMERGE made records vulnerable to



**Figure 2. eMERGE impacts on clinical care and discovery**

The eMERGE Network began by focusing on discovery-based research (white hexagons) before moving into clinical utility-based research (gray hexagons) in phase III. Discovery-based research remains a foundation of the network, contributing to the broad knowledgebase of clinical genomics, which, in turn, can be utilized to inform standards of clinical care and precision medicine in non-research settings. The image describes the main workgroup topics focused on during the third phase and how the network approached both clinical- and discovery-based science.

privacy intrusions, particularly through “re-identification” attacks, including demographics, and DNA. As such, organizations making genomic data available began following the practice of sharing aggregate summary statistics about the underlying records, a policy adopted by the Database of Genotypes and Phenotypes (dbGaP), where eMERGE data would eventually be made accessible. However, just as eMERGE was beginning, it was shown that data shared in such a manner could also be subject to probabilistic inference attacks that would reveal the presence of a participant in the cohort and, thus, expose phenotypic information associated with the cohort (e.g., if the cohort was composed solely of participants diagnosed with a cardiovascular disorder).<sup>30</sup> Thus, it was initially decided that all participant-level, as well as summary, data to be shared outside of the network would only be done so in a controlled fashion. All data sharing requests would be subject to review and all recipients would be required to enter into a data use agreement (DUA; see [Web resources](#)), with institutional signoff, which included a prohibition on re-identification attempts and secure data storage. The DUA allowed data sharing between parties while allowing for specific site’s institutional review board (IRB)-approved protocol to inform the transfer. This DUA has allowed the rapid onboarding of 9 affiliates, 5 collaborators, and 18 sites over the history of the network, including the Clinical Sequencing Evidence-Generating Research (CSER), Implementing Genomics in Practice (IGNITE), Population Architecture Using Genomics and Epidemiology (PAGE) Consortium, Clinical Pharmacogenetics

Implementation Consortium (CPIC), All of Us Research Program (AoURP), the Analysis, Visualization, and Informatics Lab-space (AnVIL), the US Food and Drug Administration, and the Michigan Imputation Server group. Still, given the need to make data accessible to the public more broadly, the program developed various privacy protection methodologies. These included approaches for de-identification in natural language clinical notes,<sup>31</sup> aggregation of clinical codes to support discovery,<sup>32</sup> and game theory-based risk assessment strategies to determine when genomic and phenotype data could be made public through resources like SPHINX.<sup>33</sup>

Although the datasets were centralized, the analyses, research ideas, and discovery were independently led by the eMERGE sites. An essential component of the network’s success was the ability to share data and collaboratively work on projects across multiple institutions. Well-defined policies regarding data utilization, transparent communications across the steering committee and the entire network, along with the DUA, led to the success of eMERGE over the years. The network’s “Publication Policy,” to which all members requesting access to network data adhere, requires members collaborating with investigators from other sites or utilizing data from other sites, to submit a “manuscript concept sheet” (MCS; see [Web resources](#)). These sheets capture the scope, data requested, timeline, and goals of the proposed project. The MCS is reviewed by the steering committee and participants from sites can request involvement in the analysis and publication. Discussions regarding overlap with previously proposed projects and clarification of scope occur during this time period, and modifications are made as necessary. The publications across the network are tracked centrally at the Coordinating Center. Citation counts for publications are tracked quarterly using Zotero reference software via Google Scholar.

#### **Development of the eMERGE aggregate genomic datasets**

To compile the GWAS Haplotype Reference Consortium (HRC)-imputed array ( $N \approx 105,108$ ), the eMERGE Network included 83 genotype array batches comprising 105,108 participants from 12 contributing medical center sites in the final version of this phase III effort. The consented participants have indexed medical records linked to these genotype results for GWAS and PheWAS.<sup>14</sup> The PGRNseq targeted capture sequencing panel ( $N = 9,010$ ) was developed by the Pharmacogenomics Research Network (PGRN) as a tool to enable PGx discovery and implementation in large cohorts and includes 84 genes, representing all SNPs present on commercial PGx genotyping platforms at the time (Affy DMET+, Illumina ADME) and additional genes (exons and untranslated regions [UTRs]) nominated by PGRN for their associations to PGx phenotypes.<sup>25,26,34</sup> Development of the eMERGEseq panel targeted capture sequencing panels across  $N = 24,956$  participants has also been described.<sup>21</sup> The variants from these two datasets (eMERGEseq and PGRNseq) are available in

summary format by frequency in SPHINX. All three datasets are also available in dbGaP (GWAS: phs001584.v1.p1; PGRNseq: phs000906; and eMERGEseq: phs001616).

To facilitate use of the eMERGEseq and PGRNseq datasets, the network upgraded the SPHINX tool<sup>28</sup> during phase III. The eMERGEseq data (N = 24,956) were added to SPHINX, which previously housed the PGRNseq (N = 9,010) dataset. The tool was reconfigured to include dynamic data updates from DrugBank, the GWAS Catalog, linkage to dbSNP, and PharmGKB annotation resources. The importation of DrugBank annotations into SPHINX allowed for updates to drug-gene-pathway interactions to be maintained in a dynamic fashion. This ensured relevant updates to drug compound name information were captured and interactions to the variants detected in genes were displayed, providing a critical element for the study of PGx interactions. The addition of the GWAS Catalog annotations linked variants that were shown to have phenotypic association to the published peer reviewed literature, allow investigators to quickly reference previous findings for a given variant. The linkage to dbSNP provided an indexing service to document and standardize the naming, genomic mapping, population frequency, and discovery history of genetic variation. Similarly, PharmGKB linkage added literature references to the PGx variant evidence. These additions allow SPHINX to maintain current information tied to variants in the eMERGEseq and PGRNseq datasets, creating a dynamic resource for pharmacogenomic and genetic discovery.

### Discovery research in the eMERGE Network

As eMERGE transitioned from genotyping to sequencing, investigations utilizing eMERGE datasets began to incorporate rare variation along with common variation in their analyses. The network compiled 157,480 samples across 13 sites and 6 datasets (Table S1). The GWAS, PGRNseq, and eMERGEseq datasets were harmonized at the Coordinating Center to produce research files for network utilization. The array data were imputed and merged at each phase, and the latest effort utilized the Michigan Imputation Server and HRC version 1.1, thus allowing the aggregation of data from 105,108 eMERGE participants for ~40 million variants, both common and rare.<sup>14</sup> Figure S2 and Table 1 illustrate the evolution of enrollment and ancestry enrichment in eMERGE through all three phases. The main group consisted of European ancestry, African ancestry was the second largest ancestral group, and there was an enrichment of Hispanic eMERGE participants in phase II. This dataset served as the backbone of discovery efforts in eMERGE. These genetic data were combined with phenotypic data across the sites. Over the three phases, the network developed and validated 68 clinical phenotyping algorithms (Table S2). Case control status was centrally collected at the Coordinating Center in addition to annually refreshed commonly used variables (BMI, ICD, CPT, Phecodes, statin medications, and lipid and autoimmune labs; Table S3) linked to age at the event. These data were

also linked to the genomic datasets. The Columbia University site led the effort in leveraging the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) to implement shared phenotypes.<sup>20</sup> This effort demonstrated the effectiveness of using data standards to minimize the work needed to implement phenotypes across the network and promote scalability.<sup>20</sup> The harmonization and centralization of both EMR and genetic data at the Coordinating Center allowed for the rapid transfer of data to investigators across the network, freeing the sites from the burden of re-pulling and harmonizing the same data for every new research proposal. This dataset also qualifies eMERGE for the International HundredK+ Cohorts Consortium (IHCC), which is a global effort to aggregate data for translational research.

These large datasets enabled 685 publications focused on discovery and clinical utility (Figure S3) to date. In the latter half of phase II, the network moved into sequencing for the clinical return of actionable genetic variation with the PGRNseq dataset. With these changes, the data footprint became larger and decisions on when to disseminate research datasets to the network and investigators influenced the timing of discovery. The rich datasets produced required significant time and effort to compile and maintain. A major lesson learned by the eMERGE Network was the need for outlining the analyses and product goals prior to compiling a large dataset, including phenotypic requirements and hard cut-offs for data processing. Time and resources were required when cohorts were altered and large research datasets were recompiled, therefore groups must clearly lay out objectives and realistic timelines for data generation, quality control, and downstream investigations. Analysis and computation costs are significant in large datasets, and this should be considered when hosting and analyzing data, especially as the footprint of genomic data grow with new technological and software advances.

### The evolution of clinical return in eMERGE III

The network's focus on implementing genetic testing in a clinical setting produced successes but also identified challenges. One challenge that arose was that each site had a unique study design, cohort population, IRB regulations, and protocols, including how, when, and which results were returned. These differences were driven by study design as well as local IRB regulations. Described in detail by Fossey et al.,<sup>22</sup> IRBs vary in requirements, processes, and views toward RoRs from genomic sequencing. Sites encountered many real-world issues during the RoR process, including:

1. Delays in expected timelines for RoRs;
2. Changes in participant contact information;
3. Engagement of participants;
4. Death of participants prior to RoRs;
5. Previous genetic testing influencing whether participants were interested in the eMERGE results; and

**Table 1. Breakdown of eMERGE dataset diversity**

	eMERGEseqN = 24,956	GWAS N = 105,108	PGRNseq N = 9,010
<b>Self-identified racial category</b>			
African American	3,914 (16%)	15,836 (15%)	1,209 (13%)
Hawaiian/Pacific Islander	54 (0.2%)	23 (0.02%)	6 (0.1%)
Am Indian/Alaskan	79 (0.3%)	170 (0.2%)	26 (0.3%)
Asian American	1,578 (6%)	1,246 (1%)	135 (2%)
European	17,691 (71%)	79,764 (76%)	6,065 (67%)
Missing/unknown	1640 (7%)	8069 (8%)	1,569 (17%)
<b>Self-identified ethnicity category</b>			
Hispanic/Latino	1,506 (6%)	5,217 (5%)	413 (5%)
Not Hispanic/Latino	22,551 (90%)	93,425 (89%)	7,313 (81%)
Missing/unknown	899 (4%)	6,466 (6%)	1,284 (14%)

Self-reported ancestry counts and percent of participants in the network-wide genomic datasets. Race and ethnicity are captured independently and represented separately.

#### 6. Pediatric sites' participants turning 18 years of age before RoRs requiring re-consent as an adult.

Several sites had to amend their IRB protocols during the study to facilitate RoRs when participants were unable to be contacted. The major lesson learned from this work was that the RoR process is dependent on institutional culture, priorities, and local IRB regulations. However, in the absence of prior data on RoRs on a large network scale, this flexibility in the study design allowed sites more freedom to explore different approaches for RoRs across different medical settings and enable the network to learn from the complexities of return across multiple institutional settings. The diversity of approaches resulted in a natural experiment that is being explored using implementation science frameworks. This will result in publications that should inform others interested in implementing genomics in clinical care. These lessons will strengthen future research RoR efforts and, in future studies, a single central IRB could standardize the processes and allow the network to create larger, more robust standard guidelines for practice.

Another focus was measuring the impact of the RoRs on clinical outcomes. The network determined which outcome measures to collect, deployed data collection forms, and collected data 6- and 12-months post return for network-wide analyses. Due to the short time frame of the project, the network focused on "process outcomes" that examined the specific steps in a process that had evidence supporting associations with a particular health outcome (e.g., a lipid test ordered after the return of a result in a gene associated with familial hypercholesterolemia) as well as "intermediate outcomes," which are biomarkers associated with a given outcome (e.g., LDL cholesterol level). The outcomes forms were compared to the independently developed outcomes from the ClinGen Actionability working group and showed concordance, suggesting

that common outcomes could be identified for genetic diseases and interventions.<sup>23</sup> If common outcomes are used across studies, evidence can accumulate more quickly. During development of the forms, it became clear that implementation guides for a given disease were necessary as there were multiple study personnel abstracting data elements from the EMR across the 10 sites. Eleven guides were developed for the abstractors, ensuring consistent data entry for downstream analyses. Final analyses of the clinical outcomes are still pending as sites finalize data collection across their cohorts prior to the completion of phase III.

Integrating results into the EMR was a critical component of eMERGE. Several challenges were encountered as the network harmonized integration and sites placed results into the EMR. Some medical centers transitioned to new medical record systems during the project, including large infrastructure changes during the Group Health Cooperative transition to Kaiser Permanente in 2017. Compliance regulations differed across states regarding data usage and return. Creating a standard for data flow between sequencing centers and sites is key to integrating genomic test results into the EMR.<sup>24</sup> This pipeline differed depending on site regulations, study design, and requirements. The network continued to inform the genetic Health IT standards by developing a Fast Healthcare Interoperability Resources (FHIR) profile that codifies all of the combined experience learned during the return and integration of genetic information into the EMR (see [Web resources](#)).

The ability to utilize and integrate these data for both research and in clinical settings and to learn from the heterogeneity of the network is critical. Overarching lessons learned during phase III ([Table 2](#)) showcase an adaptable, cohesive network that has moved beyond genomic discovery to leading research into clinical implementation and the return of genomic results. As eMERGE continues to

**Table 2. Overarching lessons learned from the eMERGE Network**

Focus	Lessons Learned	Tools
Genomic discovery <sup>14,15,26,28</sup>	<ul style="list-style-type: none"> <li>● defined data freezes with specifications regarding diversity, phenotypic data, discovery, and implementation goals are critical to maximize resources</li> <li>● centrally collected and aligned genetic and EMR data maximize output and relieve repeated site efforts</li> </ul>	SPHINX; PheWAS Catalog
Electronic clinical phenotyping <sup>20</sup>	<ul style="list-style-type: none"> <li>● implementation and review of complex clinical algorithms using local experts maximizes phenotype accuracy</li> <li>● clearly define processes for phenotype creation, validation and implementation, using standard vocabularies and common data models increase portability</li> </ul>	PheKB
Pharmacogenomics <sup>19,28</sup>	<ul style="list-style-type: none"> <li>● shared variant knowledge base with access to structured data, and knowledge is necessary for implementation</li> <li>● provider education, customization of clinical reports, and ongoing education is critical for provider utilization</li> <li>● technical requirements and approaches for PGx integration differs from highly penetrant genetic results</li> <li>● SNP coverage in targeted panels should be sufficient for interpreting the full range and types of variants for clinical return of a given gene (e.g. <i>CYP2D6</i>)</li> </ul>	SPHINX; CDS_KB
Clinical sequencing <sup>21</sup>	<ul style="list-style-type: none"> <li>● centralized sequencing allows for harmonization across large networks</li> <li>● with multiple sequencing centers, consistency in panel validation, variant interpretations, reclassifications, and discrepancies is essential</li> </ul>	
Return of results (RoR) <sup>22</sup>	<ul style="list-style-type: none"> <li>● flexibility in study design allows exploration of different approaches for RoR</li> <li>● institutional culture and IRB regulations influence the RoR process</li> <li>● a single IRB and centralized protocol may enhance consistency and data harmonization</li> </ul>	MyResults
Integration into EMR <sup>24</sup>	<ul style="list-style-type: none"> <li>● a standard for data flow is essential for returning genomic test results across sites</li> <li>● eMERGE-informed national FHIR profiles are needed to support genomic return</li> <li>● transitions to new EMR systems can delay and alter site integration plans</li> </ul>	CDS_KB; DocUBuild

**Table 2. Continued**

Focus	Lessons Learned	Tools
Clinical outcomes <sup>23</sup>	<ul style="list-style-type: none"> <li>● implementation guides for EHR abstraction ensure consistency across personnel and sites</li> <li>● focusing on process and intermediate outcomes allows for outcomes analysis in studies with short term follow up</li> </ul>	

Lessons broken down by phase III workgroup, with relevant resources available. SPHINX (Sequence & Phenotype Integration Exchange) links sequencing data to drug associations, GWAS variants, and ancestry. PheWAS Catalog functions as a platform for analysis of phenotypes against single gene variants. PheKB (Phenotype KnowledgeBase. Collaborative) is an environment to build and validate electronic phenotypic algorithms. CDS\_KB (Clinical Decision Support KnowledgeBase) catalogs and shares clinical decision support implementation artifacts. MyResults provides education targeted to the public and information about genetic test results and disease risks. DocUBuild is a web application for creating and sharing information resources for electronic medical record systems.

evolve, the network will utilize its experience with large datasets, innovative research methods, communications, and integration into clinical settings to push forward genomic discovery and clinical implementation

### Advancing translational genomic research

The eMERGE Network has developed several methods on the secondary use of EMR data for the discovery of genotype-phenotype associations.<sup>1,2,9-13</sup> The network's success in enabling discovery methods and findings was dependent on a coordinated workgroup infrastructure, a large and diverse genomic dataset, centralized data use agreements, adherence to policies that governed data security and privacy, and informed investigator collaboration on network publications. The lessons learned from the network (Table 2) span discovery-based research and implementation of the clinical return of genetic variations associated with clinical disease phenotypes. Clearly defined data freezes and centrally collected commonly used EMR variables maximized data availability and minimized duplicative efforts across study teams. Variability in site protocols allowed for the ability to determine how different approaches alter participant experiences and to integrate genomic data into their clinical outcomes. Harmonization of protocols and data flows helped improve sample sizes, data integration, and collection across the network. The network's transition to clinical implementation<sup>27,34</sup> leveraged the variety of strategies devised by sites for returning results to participants and integrating results into diverse EMR systems<sup>24</sup> using accepted standards (Rasmussen et al., 2017, AMIA, conference). As genomic research moves from independent studies to large networks and consortiums, such as CSER, PAGE, AoURP, and IGNITE, these lessons maximize genomic and EMR data utilization from large biobanks to push forward research in clinical and translational genomics. Collaborations between these networks are key as research moves forward.

The overarching goal of eMERGE has been to advance genomic medicine. To this end, the network has established multiple tools and resources available to external researchers, spanning the fields of genomic discovery, EMR integration, participant engagement, and education (Table 2). To date, eMERGE publications have been referenced over 40,000 times (Figure S3), impacting the field of translational genomic medicine. Collaborations with external groups have helped shape the practice of other networks, clinical research, and genetic medicine.

Though the research conducted over the last several years has led to many lessons learned, there are several limitations associated with the approach the network undertook. Based on the network's experience with developing clinical algorithms, even with restrictions imposed, false positives and low positive predictive values can be an issue, especially when incorporating complex variables and techniques like NLP. Maximizing predictive values for electronic phenotyping of diseases and outcomes will be critical as the genomic research evolves. The requirements to house, compute, and make large genomic datasets available to the network have increased over time—the data footprint of over 80 TBs presents continued challenges to store and make data accessible to the research community. Calculations on datasets of ~100,000 participants are computationally intensive and require computational capacity beyond that available to many researchers. As the field moves from local servers to cloud-based platforms for data storage, egress fees and computational costs should be considered to make these data accessible and usable by all investigators. Finally, as sites implemented different protocols, ascertainment bias during enrollment at certain sites restricted downstream analyses on penetrance and clinical outcomes. Sample size was reduced for certain populations and disease phenotypes due to differences in site protocols, and utilizing a single protocol in future work may increase power and allow for the investigation of clinical diseases with low prevalence in general populations.

Moving into the next round of eMERGE, the network is focusing on genomic risk assessments and PRS. Overall risk for many common diseases is complex, and elements of family history, a person's age, and prior health data need to be assessed in parallel with genomic data to provide accurate risk scores.<sup>35</sup> Additionally, the impact of genomic risk can be modified over the lifespan. Age, medications, and lifestyle changes have been shown to interact with the genetic risk of a disease phenotype.<sup>36</sup> The population from which a risk algorithm is determined should be considered when disseminating risk to participants.<sup>37</sup> To achieve a European ancestry level of performance of PRS in non-European ancestry groups, ancestry-specific GWAS and PRS derived from large and diverse cohorts and statistical modeling are crucial. Recent work by the network demonstrated lower hazard ratios in genome-wide PRS for coronary heart disease in African ancestry compared to European ancestry or LatinX ethnicity.<sup>38</sup> The network datasets contain approximately 15% ances-

trally diverse participants (Table 1; Figure S2), which may limit the applicability of genetic findings in more diverse populations. In the next phase of the network, the goal is to increase the diversity of underrepresented populations, with targeted recruitment aimed at over 50% non-European ancestry. The lessons from enrollment and RoRs to diverse populations, even limited, will inform our next phase as we continue to strive for a more representative population. As genomic medicine evolves, increasing this diversity is critical, and the next steps for the network include enrollment of populations underrepresented in research.

Balancing collaborative discovery-based research with implementing genomic technology and results into clinical settings across multiple sites and phases has been one of the greatest achievements of the network. The rapid access to data and collaborators affords network investigators the power to test relevant hypotheses and observe rare associations on a large scale. Network data are accessible to external researchers through dbGaP as well as through direct collaboration with the network. As the field of genetic medicine evolves, the network will continue to adapt to new techniques and standards, leading the research in prediction of risk of disease, methods to implement risk prediction on a larger scale, including integration of findings into medical records, and evaluating the impact of RoR information to participants and providers.

### Data code availability

The datasets generated and referred to during this study are available at dbGap under the following accession numbers: GWAS: phs001584.v1.p1, eMERGEseq: phs001616, and PGRNseq: phs000906.

### Supplemental information

Supplemental information can be found online at <https://doi.org/10.1016/j.xhgg.2020.100018>.

### Consortia

The members of eMERGE are Jodell E. Linder, Alanna K. Rahm, Ian B. Stanaway, Hana Zouk, Elisabeth A. Rosenthal, Ozan Dikilitas, Adam S. Gordon, Megan J. Puckelwartz, Krzysztof Kiryluk, Bahram Namjou, Ingrid A. Holm, Cynthia A. Prows, Iftikhar J. Kullo, Marc S. Williams, Heidi L. Rehm, Patrick M.A. Sleiman, Chunhua Weng, Wendy K. Chung, Maureen E. Smith, Bradley A. Malin, Samuel E. Adunyah, Christopher G. Chute, Josh C. Denny, Ali Gharavi, Richard Gibbs, Hakon Hakonarson, John Harley, George Hripcsak, Elizabeth W. Karlson, Eric B. Larson, Niall Lennon, Shawn Murphy, Dan M. Roden, Richard R. Sharp, Jordan W. Smoller, Wei-Qi Wei, Scott T. Weiss, Digna R. Velez Edwards, Melissa A. Basford, Brittany B. City, Alanna J.

DiVietro, Brandy M. Mapes, Timoethia M. Stone, Laura Allison B. Woods, Jyoti G. Dayal, Robb K. Rowley, Baergen I. Schultz, Ken L. Wiley, Jr., Sheethal Jose, Christine Eng, Jianhong Hu, David Murdock, Donna Muzny, Steven Scherer, Eric Venner, Kimberly Walker, Mullai Murugan, Viktoriya Korchina, Christie Kovar, Kevin R. Dufendach, Kenneth M. Kaufman, Todd Lingren, John Lynch, Keith Marsolo, Erin M. Miller, Melanie F. Myers, Yizhao Ni, Beth L. Cobb, Debra J. Abrams, Berta Almoguera, Meckenzie Behr, Elizabeth J. Bhoj, John J. Connolly, Joe T. Glessner, Margaret Harr, Heather S. Hain, Frank D. Mentch, Addie I. Nesbitt, Renata P. da Silva, Avni Santani, Lifeng Tian, Lyam Vazquez, Paul S. Appelbaum, Wendy Chung, Katherine D. Crew, David Fasel, Alex Fedotov, Alexander L. Hsieh, Atlas Khan, Cong Liu, Maddalena Marasa, Hila Milo Rasouly, Jordan Nestor, Lynn Petukhova, Soumitra Sengupta, Yufeng Shen, Ning Shang, Miguel Verbitsky, Julia Wynn, Kenneth M. Borthwick, Adam H. Buchanan, David J. Carey, Jessica M. Goehringer, Darren K. Johnson, Laney K. Jones, Navya Shilpa Josyula, Anne E. Justice, H. Lester Kirchner, Benjamin R. Kuhn, Ming Ta Michael Lee, Yanfei Zhang, Melissa A. Kelly, Casey Overby Taylor, Thomas N. Person, Cassandra Pisieczko, Amy C. Sturm, Agnes S. Sundaresan, Nephi Walton, Janet Williams, Juliann M. Savatt, Barbara Benoit, Andrew Cagan, Victor M. Castro, Vivian S. Gainer, Shawn N. Murphy, Ladia H. Albertson-Junkans, Deborah J. Bowen, David S. Carrell, Paul K. Crane, Stephanie M. Fullerton, Andrea L. Hartzler, Nora B. Henrikson, Dustin L. Key, Kathleen A. Leppig, James D. Ralston, Arvind Ramaprasan, Aaron Scrol, Peter Tarczy-Hornoch, David L. Veenstra, Hana Bangash, Pedro J. Caraballo, Mariza De Andrade, David C. Kochan, Noralane M. Lindor, Daniel J. Schaid, Joel E. Pacyna, Maya S. Safarova, Gabriel Q. Shaibi, Janet E. Olson, Philip Lammers, Siddharth Pratap, Rajbir Singh, Duane T. Smoot, Sharon Aufox, Christin Hoell, Yoonjung Y. Joo, Yuan Luo, Elizabeth McNally, Jennifer A. Pacheco, Luke V. Rasmussen, Laura J. Rasmussen-Torvik, Justin Starren, Theresa Walunas, Firas H. Wehbe, Samuel Aronson, Lawrence J. Babb, Mark Bowser, Birgit Funke, Stacey Gabriel, Chet Graham, Maegan V. Harden, Elizabeth D. Hynes, Barbara J. Klanderma, Emily Kudalkar, Matthew S. Lebo, Chiao-Feng Lin, Alyssa Macbeth, Lisa Mahanta, Himanshu Sharma, Matthew Varugheese, Leora Witkowski, Sarah T. Bland, Ellen Wright Clayton, Todd L. Edwards, Jacklyn N. Hellwege, Jacob M. Keaton, Sara L. Van Driest, Quinn S. Wells, Murray Brilliant, Scott Hebring, Terrie Kitchner, Erwin P. Bottinger, Eimear E. Kenny, Aniwaa Owusu Obeng, Rex L. Chisholm, Gail P. Jarvik, Josh F. Peterson, David R. Crosslin.

Author affiliations can be found in [Table S4](#).

## Acknowledgments

The network would like to thank all members throughout phases I, II, and III who contributed and made this work possible. The following grants funded phase III: U01HG8657 (Kaiser Permanente Washington/University of Washington), U01HG8685 (Brigham

and Women's Hospital), U01HG8672 (Vanderbilt University Medical Center), U01HG8666 (Cincinnati Children's Hospital Medical Center), U01HG6379 (Mayo Clinic), U01HG8679 (Geisinger Clinic), U01HG8680 (Columbia University Health Sciences), U01HG8684 (Children's Hospital of Philadelphia), U01HG8673 (Northwestern University), U01HG8701 (Vanderbilt University Medical Center serving as the Coordinating Center), U01HG8676 (Partners Healthcare/Broad Institute), and U01HG8664 (Baylor College of Medicine) The following grants funded phase II: U01HG006828 (Cincinnati Children's Hospital Medical Center/Boston Children's Hospital), U01HG006830 (Children's Hospital of Philadelphia), U01HG006389 (Essentia Institute of Rural Health, Marshfield Clinic Research Foundation and Pennsylvania State University), U01HG006382 (Geisinger Clinic), U01HG006375 (Group Health Cooperative/University of Washington), U01HG006379 (Mayo Clinic), U01HG006380 (Icahn School of Medicine at Mount Sinai), U01HG006388 (Northwestern University), U01HG006378 (Vanderbilt University Medical Center), and U01HG006385 (Vanderbilt University Medical Center serving as the Coordinating Center). The following grants funded eMERGE imputed merged phase I and phase II dataset: U01HG004438 (CIDR) and U01HG004424 (the Broad Institute) serving as Genotyping Centers. The PGRNSeq dataset (eMERGE PGx) was funded by grant U01HG004438 (CIDR) serving as a Sequencing Center. The following grants funded phase I: U01-HG-004610 (Group Health Cooperative/University of Washington), U01-HG-004608 (Marshfield Clinic Research Foundation and Vanderbilt University Medical Center), U01-HG-04599 (Mayo Clinic), U01HG004609 (Northwestern University), U01-HG-04603 (Vanderbilt University Medical Center, also serving as the Administrative Coordinating Center); U01HG004438 (CIDR), and U01HG004424 (the Broad Institute) serving as Genotyping Centers.

## Declaration of interests

C.E., R.G., J. Hu, D. Murdock, D. Munzy, S.S., E.V., and K.W. list Baylor Genetics Laboratories (BGL) as a competing financial interest, and BGL is co-owned by Baylor College of Medicine. A.H.B. declares MeTree and You, Inc. as a competing financial interest and has equity stakes in the company. D.R.C. is a consultant for UnitedHealth Group R&D. E.E.K. received speaker honorariums from Regeneron, Illumina, and KPMG. J.F.P. is a consultant for Color Genomics. E.M. is a consultant to Invitae. E.V. is cofounder of Codified Genomics, which provided variant interpretation services for a portion of the work conducted. All other authors declare no competing interests.

## Web resources

CDSKB, <https://cdskb.org/>  
dbGaP, <https://www.ncbi.nlm.nih.gov/gap/>  
DocUBuild, <https://docubuild.fsm.northwestern.edu/>  
eMERGE data use agreement, <https://emerge-network.org/collaborate>  
eMERGE manuscript concept sheet, <https://emerge-network.org/publications/>  
eMERGE FHIR reference, <https://emerge-fhir-spec.readthedocs.io/en/latest/>  
MyResults, <https://myresults.org/>  
PheKB, <https://phekb.org/>  
PheWAS, <https://phewascatalog.org/>  
SPHINX, <https://www.emergesphinx.org/>



## References

1. Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., and Shekelle, P.G. (2006). Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann. Intern. Med.* *144*, 742–752.
2. Baron, R.J. (2007). Quality improvement with an electronic health record: achievable, but not automatic. *Ann. Intern. Med.* *147*, 549–552.
3. Ng, M.C.Y., Shriner, D., Chen, B.H., Li, J., Chen, W.-M., Guo, X., Liu, J., Bielinski, S.J., Yanek, L.R., Nalls, M.A., et al.; FIND Consortium; eMERGE Consortium; DIAGRAM Consortium; MuTHER Consortium; and MEta-analysis of type 2 Diabetes in African Americans Consortium (2014). Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* *10*, e1004517.
4. Ritchie, M.D., Verma, S.S., Hall, M.A., Goodloe, R.J., Berg, R.L., Carrell, D.S., Carlson, C.S., Chen, L., Crosslin, D.R., Denny, J.C., et al. (2014). Electronic medical records and genomics (eMERGE) network exploration in cataract: Several new potential susceptibility loci. *Mol Vis.* *20*, 1281–1295.
5. Dumitrescu, L., Ritchie, M.D., Denny, J.C., El Rouby, N.M., McDonough, C.W., Bradford, Y., Ramirez, A.H., Bielinski, S.J., Basford, M.A., Chai, H.S., et al. (2017). Genome-wide study of resistant hypertension identified from electronic health records. *PLoS ONE* *12*, e0171745.
6. Cronin, R.M., Field, J.R., Bradford, Y., Shaffer, C.M., Carroll, R.J., Mosley, J.D., Bastarache, L., Edwards, T.L., Hebring, S.J., Lin, S., et al. (2014). Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index. *Front. Genet.* *5*, 250.
7. Edwards, T.L., Giri, A., Hellwege, J.N., Hartmann, K.E., Stewart, E.A., Jeff, J.M., Bray, M.J., Pendergrass, S.A., Torstenson, E.S., Keaton, J.M., et al. (2019). A *Trans*-Ethnic Genome-Wide Association Study of Uterine Fibroids. *Front. Genet.* *10*, 511.
8. Hellwege, J.N., Stallings, S., Torstenson, E.S., Carroll, R., Borthwick, K.M., Brilliant, M.H., Crosslin, D., Gordon, A., Hripscak, G., Jarvik, G.P., et al. (2019). Heritability and genome-wide association study of benign prostatic hyperplasia (BPH) in the eMERGE network. *Sci. Rep.* *9*, 6077.
9. McCarty, C.A., Chisholm, R.L., Chute, C.G., Kullo, I.J., Jarvik, G.P., Larson, E.B., Li, R., Masys, D.R., Ritchie, M.D., Roden, D.M., et al.; eMERGE Team (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* *4*, 13.
10. Pathak, J., Wang, J., Kashyap, S., Basford, M., Li, R., Masys, D.R., and Chute, C.G. (2011). Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. *J. Am. Med. Inform. Assoc.* *18*, 376–386.
11. Kho, A.N., Pacheco, J.A., Peissig, P.L., Rasmussen, L., Newton, K.M., Weston, N., Crane, P.K., Pathak, J., Chute, C.G., Bielinski, S.J., et al. (2011). Electronic medical records for genetic research: results of the eMERGE consortium. *Sci. Transl. Med.* *3*, 79re1.
12. Gottesman, O., Kuivaniemi, H., Tromp, G., Faucett, W.A., Li, R., Manolio, T.A., Sanderson, S.C., Kannry, J., Zinberg, R., Basford, M.A., et al. (2013). The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet. Med.* *15*, 761–771.
13. Crawford, D.C., Crosslin, D.R., Tromp, G., Kullo, I.J., Kuivaniemi, H., Hayes, M.G., Denny, J.C., Bush, W.S., Haines, J.L., Roden, D.M., et al. (2014). eMERGEing progress in genomics-the first seven years. *Front. Genet.* *5*, 184.
14. Stanaway, I.B., Hall, T.O., Rosenthal, E.A., Palmer, M., Naranbhai, V., Knevel, R., Namjou-Khales, B., Carroll, R.J., Kiryluk, K., Gordon, A.S., et al. (2018). The eMERGE genotype set of 83,717 subjects imputed to ~40 million variants genome wide and association with the herpes zoster medical record phenotype. *Genet. Epidemiol.* *43*, 63–81.
15. Zuvich, R.L., Armstrong, L.L., Bielinski, S.J., Bradford, Y., Carlson, C.S., Crawford, D.C., Crenshaw, A.T., de Andrade, M., Doheny, K.F., Haines, J.L., et al. (2011). Pitfalls of merging GWAS data: lessons learned in the eMERGE network and quality control procedures to maintain high data quality. *Genet. Epidemiol.* *35*, 887–898.
16. Verma, S.S., de Andrade, M., Tromp, G., Kuivaniemi, H., Pugh, E., Namjou-Khales, B., Mukherjee, S., Jarvik, G.P., Kottyan, L.C., Burt, A., et al. (2014). Imputation and quality control steps for combining multiple genome-wide datasets. *Front. Genet.* *5*, 370.
17. Rasmussen-Torvik, L.J., Pacheco, J.A., Wilke, R.A., Thompson, W.K., Ritchie, M.D., Kho, A.N., Muthalagu, A., Hayes, M.G., Armstrong, L.L., Scheftner, D.A., et al. (2012). High density GWAS for LDL cholesterol in African Americans using electronic medical records reveals a strong protective variant in APOE. *Clin. Transl. Sci.* *5*, 394–399.
18. Namjou, B., Lingren, T., Huang, Y., Parameswaran, S., Cobb, B.L., Stanaway, I.B., Connolly, J.J., Mentch, F.D., Benoit, B., Niu, X., et al.; eMERGE Network (2019). GWAS and enrichment analyses of non-alcoholic fatty liver disease identify new trait-associated genes and pathways across eMERGE Network. *BMC Med.* *17*, 135.
19. Rohrer Vitek, C.R., Abul-Husn, N.S., Connolly, J.J., Hartzler, A.L., Kitchner, T., Peterson, J.F., Rasmussen, L.V., Smith, M.E., Stallings, S., Williams, M.S., et al. (2017). Healthcare provider education to support integration of pharmacogenomics in practice: the eMERGE Network experience. *Pharmacogenomics* *18*, 1013–1025.
20. Hripscak, G., Shang, N., Peissig, P.L., Rasmussen, L.V., Liu, C., Benoit, B., Carroll, R.J., Carrell, D.S., Denny, J.C., Dikilitas, O., et al. (2019). Facilitating phenotype transfer using a common data model. *J. Biomed. Inform.* *96*, 103253.
21. eMERGE Consortium (2019). Harmonizing Clinical Sequencing and Interpretation for the eMERGE III Network. *Am J Hum Genet.* *105*, 588–605.
22. Fossey, R., Kochan, D., Winkler, E., Pacyna, J.E., Olson, J., Thibodeau, S., Connolly, J.J., Harr, M., Behr, M.A., Prows, C.A., et al. (2018). Ethical Considerations Related to Return of Results from Genomic Medicine Projects: The eMERGE Network (Phase III) Experience. *J Pers Med.* *8*, 2.
23. Williams, J.L., Chung, W.K., Fedotov, A., Kiryluk, K., Weng, C., Connolly, J.J., Harr, M., Hakonarson, H., Leppig, K.A., Larson, E.B., et al. (2018). Harmonizing Outcomes for Genomic Medicine: Comparison of eMERGE Outcomes to ClinGen Outcome/Intervention Pairs. *Healthcare (Basel)* *6*, 83.
24. Aronson, S., Babb, L., Ames, D., Gibbs, R.A., Venner, E., Connelly, J.J., Marsolo, K., Weng, C., Williams, M.S., Hartzler, A.L., et al. (2018). Empowering genomic medicine by establishing critical sequencing result data flows: the eMERGE example. *J Am Med Inform Assoc.* *25*, 1375–1381.

25. Denny, J.C., Ritchie, M.D., Basford, M.A., Pulley, J.M., Bastarache, L., Brown-Gentry, K., Wang, D., Masys, D.R., Roden, D.M., and Crawford, D.C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26, 1205–1210.
26. Denny, J.C., Bastarache, L., Ritchie, M.D., Carroll, R.J., Zink, R., Mosley, J.D., Field, J.R., Pulley, J.M., Ramirez, A.H., Bowton, E., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* 31, 1102–1110.
27. Rasmussen-Torvik, L.J., Almqguera, B., Doheny, K.F., Freimuth, R.R., Gordon, A.S., Hakonarson, H., Hawkins, J.B., Husami, A., Ivacic, L.C., Kullo, I.J., et al. (2017). Concordance between Research Sequencing and Clinical Pharmacogenetic Genotyping in the eMERGE-PGx Study. *J. Mol. Diagn.* 19, 561–566.
28. Rasmussen-Torvik, L.J., Stallings, S.C., Gordon, A.S., Almqguera, B., Basford, M.A., Bielinski, S.J., Brautbar, A., Brilliant, M.H., Carrell, D.S., Connolly, J.J., et al. (2014). Design and anticipated outcomes of the eMERGE-PGx project: a multi-center pilot for preemptive pharmacogenomics in electronic health record systems. *Clin. Pharmacol. Ther.* 96, 482–489.
29. Kullo, I.J., Olson, J., Fan, X., Jose, M., Safarova, M., Radecki Breitkopf, C., Winkler, E., Kochan, D.C., Snipes, S., Pacyna, J.E., et al. (2018). The Return of Actionable Variants Empirical (RAVE) Study, a Mayo Clinic Genomic Medicine Implementation Study: Design and Initial Results. *Mayo Clin. Proc.* 93, 1600–1610.
30. Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J.V., Stephan, D.A., Nelson, S.F., and Craig, D.W. (2008). Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet.* 4, e1000167.
31. Carrell, D., Malin, B., Aberdeen, J., Bayer, S., Clark, C., Wellner, B., and Hirschman, L. (2013). Hiding in plain sight: use of realistic surrogates to reduce exposure of protected health information in clinical text. *J. Am. Med. Inform. Assoc.* 20, 342–348.
32. Heatherly, R., Rasmussen, L.V., Peissig, P.L., Pacheco, J.A., Harris, P., Denny, J.C., and Malin, B.A. (2016). A multi-institution evaluation of clinical profile anonymization. *J. Am. Med. Inform. Assoc.* 23 (e1), e131–e137.
33. Wan, Z., Vorobeychik, Y., Xia, W., Clayton, E.W., Kantarcioglu, M., and Malin, B. (2017). Expanding Access to Large-Scale Genomic Data While Promoting Privacy: A Game Theoretic Approach. *Am. J. Hum. Genet.* 100, 316–322.
34. Gordon, A.S., Fulton, R.S., Qin, X., Mardis, E.R., Nickerson, D.A., and Scherer, S. (2016). PGRNseq: a targeted capture sequencing panel for pharmacogenetic research and implementation. *Pharmacogenet. Genomics* 26, 161–168.
35. Weigl, K., Chang-Claude, J., Knebel, P., Hsu, L., Hoffmeister, M., and Brenner, H. (2018). Strongly enhanced colorectal cancer risk stratification by combining family history and genetic risk score. *Clin. Epidemiol.* 10, 143–152.
36. Mega, J.L., Stitzel, N.O., Smith, J.G., Chasman, D.I., Caulfield, M., Devlin, J.J., Nordio, F., Hyde, C., Cannon, C.P., Sacks, F., et al. (2015). Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* 385, 2264–2271.
37. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* 100, 635–649.
38. Dikilitas, O., Schaid, D.J., Kosel, M.L., Carroll, R.J., Chute, C.G., Denny, J.A., Fedotov, A., Feng, Q., Hakonarson, H., Jarvik, G.P., et al. (2020). Predictive Utility of Polygenic Risk Scores for Coronary Heart Disease in Three Major Racial and Ethnic Groups. *Am. J. Hum. Genet.* 106, 707–716.