

Vanderbilt University Law School

**Scholarship@Vanderbilt Law**

---

Vanderbilt Law School Faculty Publications

Faculty Scholarship

---

1-6-2023

## Returning Integrated Genomic Risk and Clinical Recommendations: The eMERGE Study

Ellen W. Clayton

Jodell E. Linder

Aimee Allworth

Sara T. Bland

100 others...

Follow this and additional works at: <https://scholarship.law.vanderbilt.edu/faculty-publications>



Part of the [Computer Law Commons](#), and the [Health Law and Policy Commons](#)

---



## ARTICLE

# Returning integrated genomic risk and clinical recommendations: The eMERGE study



### ARTICLE INFO

#### Article history:

Received 26 July 2022

Received in revised form

16 December 2022

Accepted 21 December 2022

Available online 6 January 2023

#### Keywords:

Common variants

Family history

Genotyping

Monogenic risks

Polygenic risk scores

### ABSTRACT

**Purpose:** Assessing the risk of common, complex diseases requires consideration of clinical risk factors as well as monogenic and polygenic risks, which in turn may be reflected in family history. Returning risks to individuals and providers may influence preventive care or use of prophylactic therapies for those individuals at high genetic risk.

**Methods:** To enable integrated genetic risk assessment, the eMERGE (electronic MEDical Records and GENomics) network is enrolling 25,000 diverse individuals in a prospective cohort study across 10 sites. The network developed methods to return cross-ancestry polygenic risk scores, monogenic risks, family history, and clinical risk assessments via a genome-informed risk assessment (GIRA) report and will assess uptake of care recommendations after return of results.

**Results:** GIRAs include summary care recommendations for 11 conditions, education pages, and clinical laboratory reports. The return of high-risk GIRA to individuals and providers includes guidelines for care and lifestyle recommendations. Assembling the GIRA required infrastructure and workflows for ingesting and presenting content from multiple sources. Recruitment began in February 2022.

**Conclusion:** Return of a novel report for communicating monogenic, polygenic, and family history-based risk factors will inform the benefits of integrated genetic risk assessment for routine health care.

© 2023 The Authors. Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Introduction

Using genetic risk factors to identify individuals at high risk of disease promises to improve screening practices that are currently based on clinical risk factors and family history.<sup>1,2</sup> Monogenic disease risks are now incorporated into screening guidelines for several cancers (breast, ovarian, colorectal) and cardiometabolic (familial hyperlipidemia) conditions.<sup>3,4</sup> Yet, risk assessment for monogenic diseases based on family history alone is suboptimal and may miss more than three-quarters of affected

patients.<sup>5,6</sup> The introduction of polygenic risk scores (PRS), reflecting the aggregate effect of many common genetic variants of small individual effect, may enable more integrated genetic risk assessment but complicates point-of-care translation. Because the analytical validity of PRS is largely unknown for groups whose genetic ancestries are underrepresented in genome-wide association studies (GWAS) data sets, improving the generalizability of genetic risk estimation is critical to ensuring equitable implementation of genomic risk assessment across diverse populations.<sup>7-9</sup>

\*Correspondence and requests for materials should be addressed to Josh F. Peterson, Center for Precision Medicine, Department of Biomedical Informatics, Vanderbilt University Medical Center, 2525 West End Ave - Suite 1500, Nashville, TN 37203. E-mail address: [josh.peterson@vumc.org](mailto:josh.peterson@vumc.org)

A full list of authors and affiliations appears at the end of the paper.

doi: <https://doi.org/10.1016/j.gim.2023.100006>

1098-3600/© 2023 The Authors. Published by Elsevier Inc. on behalf of American College of Medical Genetics and Genomics. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

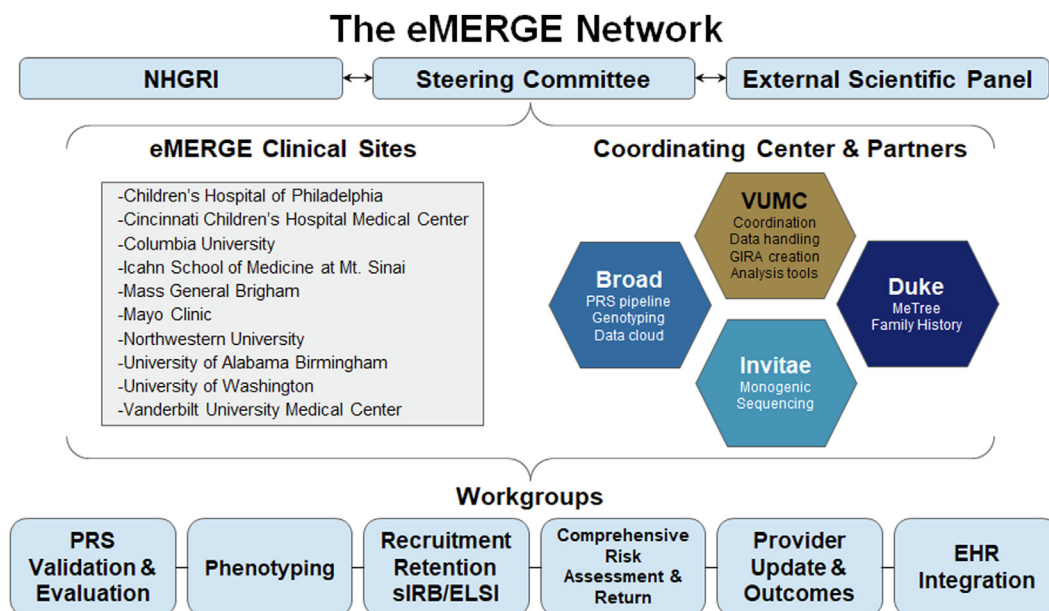
Implementing new and existing methods to apply complex genetic risk stratification to preventive care decisions requires new methods to communicate comprehensive genomic risk assessments to patients and providers. Such methods should integrate 3 components of genetic risk (family history, monogenic, and polygenic), incorporate clinical risk factors, connect genetic risk to clinical utility, and provide clear guidance on interpretation and clinical actionability. The electronic Medical Records and GENomics (eMERGE) network developed a “genome-informed risk assessment” (GIRA) report that compiles risk information from 4 sources: (1) clinical data from self-report and the electronic health record (EHR), (2) family history information provided by individuals during study enrollment, (3) clinical PRS results, and (4) clinical sequencing for a limited number of monogenic risks. The GIRA report will be incorporated into individuals’ medical records and returned to both individuals and their health care providers. Guidance is included in the report to tailor screening tests and behavioral interventions to genetic risks that could prevent or mitigate adverse health outcomes. The network will assess whether the GIRA influences downstream health care use, behavior changes, and understanding of disease risk. This manuscript describes the network’s approach to assessing integrated risk, development of the GIRA report, and the design of the prospective cohort study aimed at examining GIRA utility for 11 complex conditions.

## Materials and Methods

### Network organization

The eMERGE (electronic Medical Records and GENomics) network is a National Human Genome Research Institute–supported consortium of 10 health care institutions within the United States and a coordinating center (CC) responsible for data integration, report generation, and overall network logistics (Figure 1). In its first 3 phases, eMERGE successfully developed tools to use EHRs for discovery and to create programs to implement genomic medicine across large cohorts. For example, eMERGE developed methods for accurately extracting phenotypes from EHRs for use within large GWAS and pioneered the development of phenome-wide association studies.<sup>10-15</sup> eMERGE also studied the clinical implementation of sequencing across large cohorts, returning genomic results to individuals, placing results directly into EHRs, and deploying related clinical decision support.<sup>11,12,16-19</sup>

To complete the study objectives of the fourth phase, the consortium has engaged 3 new partners to provide essential laboratory or point-of-care data collection capabilities: the Broad Institute (clinical PRS testing and reporting), Duke University (family history collection tool, MeTree), and Invitae (Invitae Corporation; clinical monogenic sequencing). The network hosts 6 workgroups (Supplemental



**Figure 1 eMERGE network organizational structure.** Under the guidance of the NHGRI, External Scientific Panel, and Steering Committee, the 10 clinical sites, coordinating center, and network partners make up 6 workgroups charged with developing the network protocol, risk scores, and methods as well as recruiting and returning results to 25,000 individuals. Broad, Broad Institute of MIT and Harvard; Duke, Duke University; EHR, electronic health record; eMERGE, electronic Medical Records and Genomics; ELSI, ethical legal and social implications; NHGRI, National Human Genome Research Institute; PRS, polygenic risk score; sIRB, single institutional review board; VUMC, Vanderbilt University Medical Center.

Table 1) and is guided by a steering committee that meets twice monthly via teleconference and 3 times per year at hybrid meetings, as well as with a 7-member External Scientific Panel. Network decisions, rationale, and timelines are tracked centrally by the CC and displayed to the network via dashboards.

## Study design

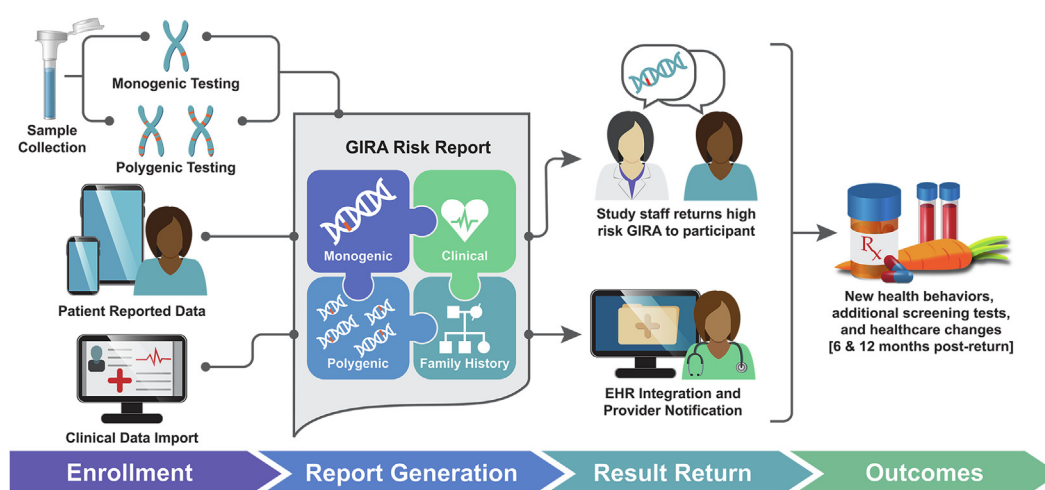
The eMERGE network aims to increase applicability of genomic risk prediction across populations by validating PRS in multiple ancestral groups and enrolling a prospective cohort that includes individuals who are currently under-represented in clinical-genomic research. Six sites are committed to recruiting an “enhanced diversity cohort” with a target of 75% of individuals belonging to a racial or ethnic minority or medically underserved population, whereas the remainder of clinical sites will target 35%. Enrollment is not targeted to individuals with specific conditions, although individuals with prevalent conditions can be included. The network focuses on 3 major aims: (1) recruit 25,000 individuals (ages 3-75 years) from general health care system populations, (2) generate cross-ancestry and ancestry-adjusted PRS as the basis for reports to return risk alongside family health history, clinical, and monogenic risk, and (3) measure individual outcomes and provider behaviors and comprehension in response to receiving this information.

The Vanderbilt University Medical Center CC is the institutional review board of record (#211043) for the network’s single institutional review board, approved in July 2021. Exclusion criteria include inability to consent or receive results in English or Spanish, history of solid organ or bone marrow transplant, lack of a health care provider at

the parent institution, and study personnel. The network uses a prescreening survey before consent to ascertain eligibility. Individuals are consented electronically and then given 2 surveys to complete before providing biological samples (saliva or blood). Samples are sent to the Broad Institute for generation of polygenic risk reports and Invitae (adults only) for monogenic reports. Individuals also provide family health history data using MeTree software. Data are received at the CC to generate GIRA reports (Figure 2).

## Condition selection

Conditions were selected after considering the analytical and clinical validity of the associated PRS, projected clinical actionability, and applicability of genomic risks to populations of diverse genetic ancestries. During validation studies, the network compared individuals above a selected threshold to all those below the threshold; scores with nonsignificant odds ratios (OR; those with 95% CIs that overlapped 1.0) in multiple genetic ancestries were deprioritized for implementation. Definitions for each selection criterion and categories of genomic risk are provided in the Supplemental Glossary. The network reviewed 23 proposed conditions and selected 11 conditions for implementation (Table 1). The initial assessments included comprehensive literature reviews focused on PRS performance; the network then conducted independent secondary validations in additional clinical-genomic data sets across 4 genetic ancestry groups (European, Asian, African, and Hispanic). Because the PRS for colorectal cancer did not validate in either Hispanic or African ancestry sample data, the colorectal cancer PRS risk component was excluded from the GIRA. Of the 11 conditions, 4 were determined to



**Figure 2** The GIRA report integrates multiple risk elements and is returned to individuals, providers, and the EHR. Individuals in the electronic Medical Records and Genomics (eMERGE) study are enrolled and provide both self-reported (clinical and family history) data and biological samples. Polygenic risk (Broad report), monogenic risk (Invitae report), family health history (MeTree pedigree), and clinical data (from EHRs and participant surveys) are collected and combined to create the GIRA report. Results are returned along with care recommendations that can be used to guide screening recommendations and next steps for primary care providers. Outcomes are assessed to determine changes in health care behavior after return. EHR, electronic health record; GIRA, genome-informed risk assessment.

**Table 1** High-risk criteria for genome-informed risk assessment conditions

Condition	Monogenic Risk		Polygenic Risk	Family History Risk		Expected Number of High Risk <sup>a</sup>
	Genes Considered	P/LP Frequency, %	High Risk Threshold, %	Relatives	% Increased Risk	
Pediatric only conditions (3-17 y at enrollment)						
Asthma	N/A	N/A	5	N/A	N/A	250
Type 1 diabetes	N/A	N/A	3	N/A	N/A	150
Pediatric and adult conditions (3-75 y at enrollment)						
Obesity/BMI	N/A	N/A	3	N/A	N/A	750
Type 2 diabetes	N/A	N/A	2	N/A	N/A	500
Adult only conditions (18-75 y at enrollment)						
Atrial fibrillation	<i>LMNA</i>	0.05	3	Parents <75 y	5	610
Breast cancer <sup>b,c</sup>	<i>BRCA1, BRCA2, PALB2, PTEN, TP53, STK11</i>	1.48	5	First and second degree	11	348
Chronic kidney disease	N/A	N/A	2	First degree	9	400
Colorectal cancer	<i>EPCAM, MLH1, MSH2, MSH6, PMS2, STK11, PTEN, TP53</i>	0.54	N/A	N/A	N/A	110
Coronary heart disease	<i>APOB, LDLR, LDLRAP1, PCSK9</i>	0.89	5	First degree	9.8	1178
Hypercholesterolemia	<i>APOB, LDLR, LDLRAP1, PCSK9</i>	0.89	3	N/A	N/A	778
Prostate cancer <sup>c</sup>	<i>BRCA1, BRCA2, EPCAM, MLH1, MSH2, MSH6, PMS2</i>	1.69	10	First degree males	9.4	1169

*BMI*, body mass index; *BOADICEA*, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; *N/A*, not applicable; *P/LP*, pathogenic/likely pathogenic.

<sup>a</sup>Predicted in-person return for combined polygenic risk and monogenic risk scores does not account for overlap. Trigger hierarchy: monogenic > polygenic > family history risk.

<sup>b</sup>Breast cancer triggers from *BOADICEA* integrated score, 25% lifetime risk.

<sup>c</sup>Breast and prostate cancer risk returned to self-reported sex at birth of female and male, respectively. *N* ~ 5000 pediatric and *N* ~ 20,000 adult individuals are expected.

be pertinent clinically for individuals aged 3 to 17 years: asthma, obesity/body mass index (BMI), type 1 diabetes (T1D), and type 2 diabetes (T2D). Asthma and T1D are only returned to those aged <18 years, whereas obesity/BMI and T2D are returned to adults as well. Considering the network ethical, legal, and social implication (ELSI) input and uncertain penetrance of monogenic genes on these 4 conditions, the network elected not to offer monogenic testing to individuals aged <18 years.

## Polygenic risk assessment

The network selected a genotyping panel for PRS generation based on the Global Diversity Array (Illumina); the Global Diversity Array balances cost and content and includes a large number of single-nucleotide variations (formerly single-nucleotide polymorphisms) optimized for imputation performance across diverse populations. Standard imputation and analytical pipelines were implemented to support clinical reporting of the PRS, and the network developed clinical report language and a logic structure for the PRS components of the GIRA.

For each of the 10 PRSs selected for clinical reporting, the threshold for high-risk status was selected by one of

several methods that varied by phenotype. Thresholds were chosen to be equivalent to a clinically meaningful risk factor such as family history or a corresponding OR of  $\geq 2$  (Table 1). The Broad Institute developed a phasing and imputation method based on the tools used by the Michigan Imputation Server (University of Michigan).<sup>20</sup> The final PRS pipeline used 2 distinct population reference panels, the 1000 Genome project panel for imputation and parameters derived from the *All of Us Research Program* cohort for ancestry calibration. Imputation and PRS analytical validity were determined through a validation study in the Broad Institute clinical laboratory that leveraged 42 samples with matched polymerase chain reaction-free whole-genome sequences in 20 specimens with matched blood and saliva sample collection. The PRS determination pipelines were developed in the Terra cloud platform ([www.terra.bio](http://www.terra.bio)) using the Workflow Description Language, allowing the methods to be made available for research purposes in the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space (AnVIL)<sup>21</sup> platform (Supplemental Table 2).

Each condition selected underwent validation on relevant data sets based on the targeted PRS populations. The PRS pipeline performance was additionally verified on the

eMERGE I-III cohorts<sup>12,22,23</sup> by the Broad Institute clinical laboratory to ensure that the ORs, magnitudes, and significance could be independently verified. To accurately and equitably return scores across ancestry groups, an ancestry adjustment step was applied to the raw PRS scores, which was a derivative of the method first proposed by Khera et al<sup>24</sup> and includes variance correction.

Network discussions determined clinical report content. A pipeline for automated report creation was built at the Broad Institute and includes a step for clinical laboratory staff review of results before release. Clinical reports are created as both PDF documents and structured data in JSON format, both of which are returned to the CC for inclusion in the GIRA and delivery to the clinical sites.

### Monogenic risk assessment

For a subset of conditions (Table 1), the network assessed monogenic risk with a limited panel of genes. Although funding for monogenic testing was not included in the initial eMERGE project, the network partnered with Invitae to support this testing with an in-kind contribution.

Consideration of which genes to include in monogenic testing occurred in parallel with selection of the final conditions for which PRS would be reported. The network finalized the list of genes, including the Centers for Disease Control and Prevention tier 1 conditions<sup>25</sup> (hereditary breast and ovarian cancer syndrome, Lynch syndrome, and familial hypercholesterolemia) as well as 5 additional genes: *TP53*, *STK11*, *PTEN*, *PALB2*,<sup>26</sup> which are important risk factors for colorectal and breast cancer, and *LMNA*, which is associated with arrhythmias, including atrial fibrillation (Table 1). Variants of uncertain significance will not be returned. Based on Invitae data, the network estimated a pathogenic or likely pathogenic result rate for this expanded panel, excluding heterozygote results from *LDLRAP1*, *LMNA*, and *EPCAM* of approximately 2.5%.<sup>27</sup>

### Family history risk assessment

Family history is an important component of risk assessment for most conditions in this study and was therefore included as a high-risk criterion in the GIRA. MeTree, a family history application,<sup>28</sup> is being used to collect variables on first and second degree relatives directly from the individuals for the conditions of interest. Family history text is displayed on all GIRA reports as contextual factors along with a pedigree. Family history is used as a high-risk return trigger for 5 conditions: atrial fibrillation, breast cancer (integrated into the Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm [BOADICEA]<sup>29</sup> model), chronic kidney disease, coronary artery heart disease (CHD), and prostate cancer (Table 1). The Duke University team collaborated with network workgroups to customize MeTree to display only relevant eMERGE conditions and allow for a manual and

automated data transfer to the CC 8 weeks after individual account creation.

### Clinical variables

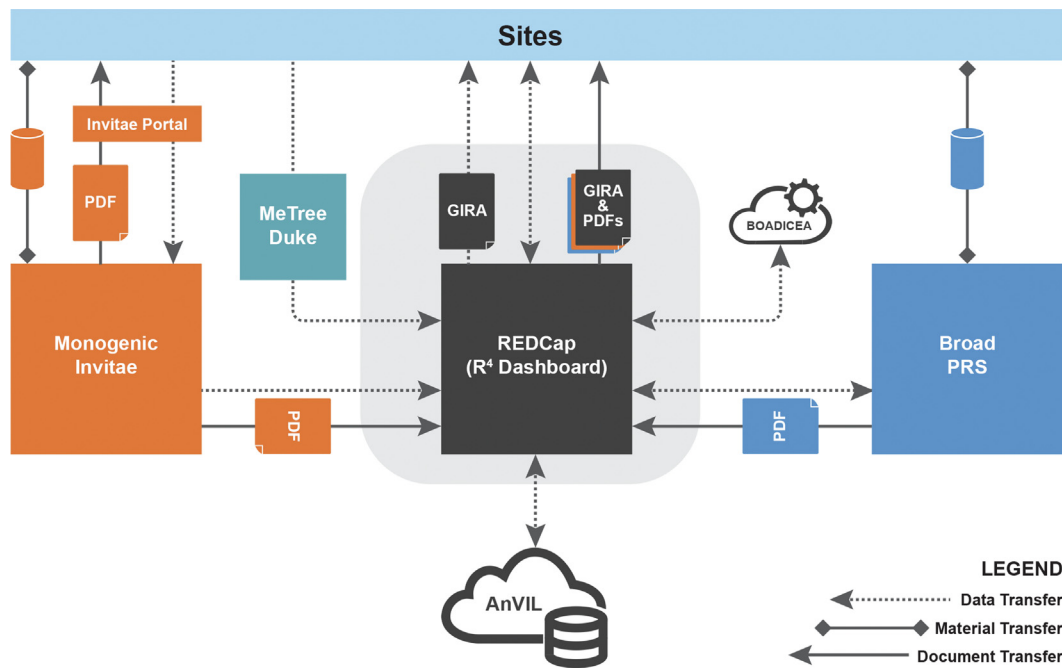
Because the study was focused on genetic risk, clinical risk factors alone do not trigger a high-risk notification. However, they are incorporated into the GIRA as contextual information for 5 conditions (asthma, chronic kidney disease, hypercholesterolemia, obesity, and T2D); clinical data that is displayed include laboratory values, blood pressure, and International Classification of Diseases codes representing known diagnoses. Clinical variables are used to calculate a comprehensive risk score for 2 conditions (BOADICEA for breast cancer and the pooled cohort equation [PCE]/integrated risk score for CHD).

### Data transfer and risk integration

Data use and transfer agreements for the eMERGE project were complex because they needed to encompass multiple sites, participant entered surveys, individual's EHR data, and data generated by network partners. The CC executed data use and transfer agreements to authorize the intake and storage of identifiable data in a centralized repository and web application based on Research Electronic Data Capture (REDCap; VUMC).<sup>30,31</sup> The custom application was named R<sup>4</sup> after the primary functions supported for the study: Recruitment, Results reporting, and Risk Reduction. This REDCap project uses data access groups to allow site-specific access to identifiable information and customized programming to generate individualized GIRA reports based on upstream data variables and associated standardized display text. Figure 3 describes the data flow across the network. The CC established application programming interfaces with the 3 network partners (Broad Institute, Duke University, and Invitae) to receive structured and PDF reports. GIRA also uses data elements from the participant surveys to generate calculated fields. BMI and the PCE that predicts 10-year risk for a first atherosclerotic cardiovascular disease event<sup>32</sup> were integrated directly into REDCap, and an application programming interface was established with CanRisk (University of Cambridge)<sup>33</sup> to calculate and send back BOADICEA<sup>29</sup> scores for breast cancer. Once all data import instruments are completed, the REDCap record is locked and the GIRA is generated for a given individual.

### GIRA return and outcome assessment plan

Because care recommendations based on polygenic risk are not currently driven by guidelines, the network adapted existing clinical recommendations associated with risks of similar magnitude to provide evidence-based guidance to individuals and providers. Recommendations were drafted and reviewed iteratively over an 8-month period in light of



**Figure 3** Research infrastructure for assembling risk information from site enrollment data collection, site EHRs, partner laboratories, and point-of-care tools for collecting family history. Data are transferred through application programming interfaces from partners and sites to the central VUMC housed REDCap system R<sup>4</sup>. Custom REDCap programming was developed to generate the overall GIRA report, which is made available to sites along with component data and PRS and monogenic laboratory reports. AnVIL, Analysis, Visualization, and Informatics Lab-space; BOADICEA, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; GIRA, genome-informed risk assessment; PRS, polygenic risk score; REDCap, Research Electronic Data Capture; VUMC, Vanderbilt University Medical Center.

feedback from study site physicians, findings from ELSI-led focus groups and interviews with physicians,<sup>34</sup> and the steering committee. The goal was to supplement the current standard of care for conditions of interest and provide individuals and providers with well-reasoned next steps. Family history and monogenic recommendations were taken directly from existing guidelines.

Individuals receiving high-risk GIRA will have a high-touch “one-on-one” return, during which a trained study staff member, study physician, or site genetic counselor will meet with the individual to explain the results and next steps. For scenarios when only family history risk triggers a “high-risk” GIRA and there are no identified genomic risk factors, individuals will still receive a “high-risk” GIRA; however, results will be returned by mail or electronically to individuals. GIRA reports are placed into the EHR and communicated to the individual’s study provider, which encompasses traditional primary care and specialists providing longitudinal care to individuals. The return methods at each site are customized based on current legal and institutional regulations. GIRA results without risk factors are deemed “not high-risk”; the network elected to avoid the terms “low risk” in any study communications to avoid inappropriate reassurance. “Not high-risk” results are returned via mail or patient portal and placed in the EHR. Educational material for patients and clinicians, as well as training of study staff about the study and its processes, were harmonized across sites.

The effect of return of results on the process and clinical outcomes will be assessed 6 months after the return of results through EHR data and participant survey and 12 months using EHR data alone. The participant survey will examine individuals’ understanding, clarity of results, how individuals felt after receiving results, and lifestyle changes. The network will examine the following outcomes among “high-risk” individuals compared with “not high-risk” individuals (Supplemental Figure 1): adoption of recommendations, uptake of risk-reducing interventions, and new diagnoses. Condition-specific analyses will be conducted that analyze prespecified process and clinical outcomes and will assess potential confounders and effect modifiers. For example, we will conduct a stratified analysis of high-risk individuals based on whether they have prevalent disease. In addition, the propensity for high-risk individuals without disease to have risk factors that could prompt testing will be investigated with a regression discontinuity design that controls for disease related confounders. A control analysis will evaluate the contribution of an in-person, high-touch return to changes in process outcomes independent of the risk information, by assessing condition-specific outcomes among individuals who receive high-touch return for an unrelated condition.

The network estimated the overall effect size (difference in adoption of recommendations among high-risk vs not high-risk) under conservative assumptions that the uptake of recommendations would be 20% in the high-risk group, and

background rates of these recommendations in standard health care (not high-risk group) would be approximately 10%. Provider surveys will focus on perceived utility and understanding of the GIRA as well as assess if providers communicated with individuals about results. Targeted provider interviews are also planned to gain a more in-depth understanding of the effect of the GIRA on care.

## Results

Enrollment began in February of 2022. As of December 2022, the 10 sites have enrolled 5671 individuals. Of those, 3688 (65%) are female at birth; 1978 (35%) are male at birth, and <1% did not indicate sex assigned at birth. Nearly half (47%) of the individuals self-reported that they are members of a racial or ethnic minority group, and an additional 4% indicated more than one race. Hispanic ethnicity accounted for approximately 18% of enrolled individuals ([Supplemental Figure 2](#)). Samples have begun being processed by the Broad Institute, Invitae, and family history data is currently being collected through MeTree. GIRA generation is targeted to begin in December 2022 with the first return to individuals and providers shortly thereafter.

## GIRA

The network created the genome-informed risk assessment (GIRA) report to concisely summarize the risk (polygenic, monogenic, family history, and clinical factors) of developing 11 common, complex conditions and to display clinical recommendations to providers and individuals. For the breast cancer and coronary heart disease conditions integrated risk scores are displayed on the GIRA (BOADICEA and PCE, respectively); for the remaining conditions, the independent risk factors are displayed along with relevant clinical factors for context). The GIRA report ([Supplemental Document 1](#)) consists of a cover page, 1-page summary of high-risk conditions, a breakdown of risk type (monogenic, polygenic, family history), education pages, frequently asked questions, methods, and a family history pedigree. In addition, laboratory reports from both Invitae (for adults) and the Broad Institute are attached to the GIRA (laboratory reports not shown in [supplemental information](#)). The example GIRA includes high-risk findings for atrial fibrillation (pathogenic *LMNA* variant), breast cancer (BOADICEA score result > 25%), and T2D (PRS above threshold) with accompanying clinical risk factors. The frequently asked questions and education materials were included to clarify each risk type and increase overall understanding. The network gathered multiple sources of expertise including ELSI studies, physician, and potential participant feedback to develop the report.<sup>34</sup> Each GIRA report is dichotomized into high-risk and not high-risk to more clearly communicate actionability because interviews with potential participants indicated a dichotomized risk was

much more understandable than a quantitative risk. Details on the ORs and CIs for the high-risk cutoff are reported stratified by ancestry in the Broad Institute laboratory report to provide details for providers (or individuals) interested in the specifics. A table of GIRA PRS text including ORs can be found in [Supplemental Table 3](#). For the 2 conditions using integrated risk scores (breast cancer and CHD), calculated absolute risks are displayed on the GIRA itself because they are commonly used in health care for clinical decision making.

The network focused on PRS as the core component of genomic risk assessment. ELSI studies of study candidates were interested in receiving PRS even if they could not be fully validated in all ancestral populations, demonstrating an urgent need for developing risk assessment across populations. Monogenic and family history risk could also trigger high-risk status for many of the conditions. [Table 1](#) displays how the different components of the GIRA trigger high-risk status. Although not all conditions use all 3 risk types, overall risk was generally designated using the following hierarchy: monogenic risk > polygenic risk > family history. Although the overall risk on the first page of the GIRA is triggered by the aforementioned hierarchy, all risk types (monogenic, polygenic, family history) are displayed on the second page of the GIRA along with risk specific text to provide the individual and provider with the most comprehensive information. There are 2 exceptions in which integrated scores are available. Breast cancer for adult females at birth used a BOADICEA score; individuals with greater than or equal to 25% lifetime breast cancer risk conveyed a high-risk status. For individuals at high-risk for CHD, the PCE both with and without the PRS incorporated is displayed on reports and is intended to assist physicians with next steps while providing standard of care information. Unlike for BOADICEA, the PCE is not used to determine risk, it is displayed as context for the provider and individual.

The network modeled the expected number of high-risk return of conditions ([Table 1](#)) to determine maximum effect and sample sizes; approximately 25% (~6200) of the 25,000 individuals are expected to receive at least one high-risk return. Estimates were primarily based on frequency of high monogenic and polygenic risk because those elements were the main components to generate a one-on-one (high-touch) return for most conditions. These calculations do not account for potential correlations between conditions in same individual or the small overlap between monogenic and high PRS risk.

## Care recommendations

The GIRA report includes clinical recommendations for individuals receiving high-risk results. The network conducted multiple physician-led discussion groups, compiled recommendations from experts across 10 clinical sites, and went through multiple iterations of the care



recommendations to supply the providers with artifacts that accurately reflected clinical care based on other risk factors (clinical, family history, monogenic) in this novel risk space. The network aimed to supplement current guidelines and to provide recommendations for those with a high PRS. Table 2 summarizes the projected clinical utility and next steps recommended for each of the 11 conditions, and Supplemental Table 4 provides the detailed recommendations for high-risk PRS results along with references for rationale on recommendations. Recommended actions included counseling and education as well as additional screening to consider involving laboratory tests, imaging, and referrals. Monogenic risk recommendations and family history-based recommendations were taken directly from clinical guidelines by the National Comprehensive Cancer Network<sup>1,2</sup> and cardiology societies<sup>3</sup>; these recommendations are not displayed in the supplemental materials. Outcomes assessments (Table 2) are derived directly from the care recommendations.

## Data harmonization and EHR integration

Multiple types of data representing the 3 types of genomic risk will be generated, collated, and used to produce the GIRA report for each individual (Figure 3). The study aims to track how providers make use of the GIRA to inform medical decision making and care once it is incorporated into the EHR and into existing workflows for the review of medical information. Educational materials are being developed for linkage through info buttons and different versions of clinical decision support to assist clinicians receiving GIRA results. eMERGE sites primarily use the Epic EHR, with one site using Cerner. The network attempted to harmonize EHR GIRA integration methodology in advance, but discussions during monthly EHR Integration Workgroup meetings indicated that some heterogeneity across sites was unavoidable. In addition to the configuration differences of each EHR, sites considered the GIRA a “research report,” in that it is generated on

**Table 2** Clinical utility and actionability of eMERGE return

Condition	Clinical Utility	Actionability	
		Counseling and Education	Laboratory Tests and Procedures
Asthma	Early intervention and therapy, 4× more common in AA	Reduce environmental triggers	Provider encounter (education)
Atrial fibrillation	Cost effective screening, early detection	Emphasize healthy lifestyle, assess symptoms	Test ordered (ECG, rhythm monitor)
Breast cancer	Enhanced screening and therapy, higher mortality in AA, BOADICEA score returned	Emphasize healthy lifestyle, screening	Imaging order (breast MRI, mammogram)
Chronic kidney disease	Early noninvasive screening, high prevalence >10% adults, 3.7× more common in AA	Emphasize healthy lifestyle, screening	Laboratory order (serum creatinine, urine albumin to creatinine ratio)
Colorectal cancer	Early noninvasive screening, risk varies by race and ethnicity	Early screening	Procedure order (stool test, flex sigmoidoscopy, colonoscopy, CT colonography)
Coronary heart disease	Noninvasive screening and intervention, highly prevalent disease, pooled cohort equation returned	Emphasize healthy lifestyle, screening	Laboratory order (lipid profile), imaging order (coronary CT, carotid ultrasound)
Hypercholesterolemia	Noninvasive, accessible screenings, disproportionate burden in AA	Emphasize healthy lifestyle, screening, medication	Laboratory order (lipid profile)
Obesity/BMI	Early intervention to modify behavior, weight in childhood is predictive of adult obesity	Emphasize healthy lifestyle, referral	Referral (weight loss or nutrition consultation)
Prostate cancer	Noninvasive, accessible screenings, population disparities in incidence and mortality	Shared decision making about early screening	Provider encounter, examination
Type 1 diabetes	Treatment dependent on distinguishing between diabetes types	Education regarding symptoms, screening	Laboratory order (autoantibodies)
Type 2 diabetes	Noninvasive screening, highly prevalent disease, racial disparities in care	Emphasize healthy lifestyle, screening	Laboratory order (fasting blood glucose, HgA1c)

*HgA1c*, glycated hemoglobin; *AA*, African American; *BMI*, body mass index; *BOADICEA*, Breast and Ovarian Analysis of Disease Incidence and Carrier Estimation Algorithm; *CT*, computed tomography; *ECG*, electrocardiogram; *eMERGE*, electronic Medical Records and Genomics; *MRI*, magnetic resonance imaging.

behalf of a research study. Although all sites are able to make a PDF of the GIRA accessible to providers within their EHR, site-specific policies resulted in variation about where in the EHR the report was displayed or localized. In addition to the GIRA PDF, sites also had to develop integration plans for the 2 component genetic reports (monogenic and PRS) and had to consider whether structured data representations of any of the 3 reports would also be incorporated into the EHR, particularly for use in providing context-aware decision support for clinicians and individuals.<sup>35</sup>

## Discussion

This phase of the eMERGE network was established to formulate scalable methods for returning integrated genomic risk to individuals and providers, including PRS, monogenic, and family history-based risks in the context of traditional clinical risks. Using a prospective cohort design, the study will determine whether returning the GIRA with recommendations and counseling for high-risk individuals is effective at increasing appropriate clinical actions to mitigate the risk of future disease or detect unrecognized disease.

Predicting the onset of common diseases or other health outcomes is essential to tailoring preventive care to individual risk. Established models focused on clinical variables for cardiovascular disease (PCE) and breast cancer (eg, Breast Cancer Risk Assessment Tool—BRCA1/2) lack simultaneous consideration of monogenic and polygenic risk, and family history-based risk is inconsistently measured and applied.<sup>36,37</sup> Developing more comprehensive tools, similar to the BOADICEA equation used in the eMERGE GIRA, will improve risk prediction and promises to guide clinical management more effectively, especially for individuals who do not know their family history of disease. In addition, adding PRS to existing models, as has been done for the PCE, could identify an expanded group of individuals for whom primary prevention or intensified surveillance is appropriate.<sup>38</sup> To learn how best to maximize the utility of the GIRA, the eMERGE study will both identify and address barriers to implementing health recommendations as well as assess psychosocial harms incurred from learning high-risk or overinterpreting not high-risk results.

A major obstacle the network had to overcome while developing PRS for the chosen conditions was ensuring that they were applicable to the diverse range of individuals that the network aims to recruit. Sufficiently diverse GWAS data are lacking to support cross-ancestry scores for many common medical conditions, even though it has been demonstrated how genetically diverse population data are critical to the accuracy of the method.<sup>9,39,40</sup> By limiting clinically implemented scores to those validated in multiple ancestral populations, the network aimed to increase the applicability of the risk scores.

The network's decisions regarding conditions included in the GIRA implementation, selection of PRS thresholds or

integrated scores, and clinical recommendations were based on review of evidence, expert opinion, and the application of clinical guidelines written to address risks of comparable magnitude. However, there are currently no clinical standards for PRS implementation, and the study may reveal the strengths and weaknesses of these early decisions. We expect significant refinements will be needed within PRS development, reporting, and integration with other genetic findings for future implementations. The study will also be challenged to ensure provider and individuals understanding of the limitations and clinical significance of the GIRA results and appropriately translate the report recommendations to changes to screening practices or lifestyle. By measuring the uptake of follow-up testing and recommendations in parallel with provider and individuals understanding, the study will shed light on how integrated genomic risk will be received in clinical practice.

As currently designed, the GIRA report is limited to information that is available within electronic records and laboratory reporting. Tailored preventive care plans recommended by providers at the point of care may need to consider many patient factors not included in the report, such as medical history, personal preferences, access to care, influence of social and physical environments, and the cost of care. In addition, by limiting the recruited population to those already receiving longitudinal care at established health care institutions, the study will not reflect the full set of challenges experienced by the general population in obtaining individualized preventive care. Finally, because the report summarizes evidence within a rapidly evolving scientific field, it is likely that report redesign and updates to the underlying PRS, integrated score algorithms, and clinical recommendations will be necessary if it is proven valuable to adopt in clinical practice.

This eMERGE study aims to improve understanding of risk stratification in the field of genomics by implementing a novel integrated genomic risk with increased applicability to diverse populations. Demonstrating clinical uptake and understanding will help to establish clinical utility and facilitate additional comparative or clinical implementation studies.

## Data Availability

The de-identified individual data that underlie the results reported in this article (including text, tables, figures, and appendices) will be made available for noncommercial, academic purposes upon request. De-identified data derived from individuals enrolled in the study will be made available on the AnVIL platform (<https://anvil.terra.bio/>) at periodic intervals over the course of the study as it is generated.

## Acknowledgments

The network thanks the patient and physician advisory groups that helped shape study design, report content, and

education materials. Owing to the structure of the U01 cooperative agreement, the recruitment sites, coordinating center, and funding agency collaborated on the study design as well as collection, management, analysis, and interpretation of the data. This trial was registered with [clinicaltrials.gov](https://clinicaltrials.gov) under identifier NCT05277116.

## Funding

The eMERGE Genomic Risk Assessment Network is funded by the National Human Genome Research Institute (NHGRI) through the following grants: U01HG011172 (Cincinnati Children's Hospital Medical Center), U01HG011175 (Children's Hospital of Philadelphia), U01HG008680 (Columbia University), U01HG011176 (Icahn School of Medicine at Mount Sinai), U01HG008685 (Mass General Brigham), U01HG006379 (Mayo Clinic), U01HG011169 (Northwestern University), U01HG011167 (University of Alabama at Birmingham), U01HG008657 (University of Washington), U01HG011181 (Vanderbilt University Medical Center), and U01HG011166 (Vanderbilt University Medical Center serving as the Coordinating Center).

## Author Information

Conceptualization: J.E.L., R.L.C., D.R.C., G.P.J., E.W.K., N.J.L., N.A.L., S.N.M., C.A.P., L.R.-T., R.R., D.R.V.E., N.S.A.-H., W.K.C., J.J.C., Ha.H., G.H., E.E.K., L.K., K.K., I.J.K., T.A.M., E.M.M., D.M.R., R.S., J.W.S., W.-Q.W., C.W., K.W., J.F.P.; Methodology: J.E.L., A.A., S.T.B., P.J.C., R.L.C., E.W.C., D.R.C., O.D., A.D., E.D.E., S.F., R.R.F., A.S.G., R.G., M.V.H., I.A.H., G.P.J., E.W.K., S.L., N.J.L., N.A.L., K.F.M., S.N.M., L.O., C.A.P., L.V.R., L.R.-T., R.R., K.T.S., T.S., S.T., D.V., D.R.V.E., J.F.P., D.A., N.S.A.-H., J.A., H.B., M.B., J.E.B., E.S.B., J.B., W.K.C., J.J.C., J.C., P.D., B.D., S.M.F., C.G., M.L.H., Ha.H., He.H., M.H., E.H., V.H., C.H., M.H.-P., G.H., M.R.I., C.K., D.K., E.E.K., A.K., K.K., B.K., L.K., I.J.K., K.L., C.L., E.M., T.A.M., T.M., E.M.M., F.M., A.M., S.D.M., P.M., B.M., N.M., B.N., E.F.P., M.J.P., T.R.-B., D.M.R., E.A.R., S.S., M.S., D.J.S., Ba.S., L.S., G.Q.S., R.R.S., Br.S., M.E.S., J.W.S., R.S., S.A.S., J.T., H.K.T., S.B.T., T.W., W.-Q.W., Q.S.W., C.W., G.L.W., K.W.; Project Administration: J.E.L., S.F., S.L., L.S., A.D., J.C., M.H., A.M., T.R.-B., E.F.P., B.M., S.T.B., J.F.P.; Writing-original draft: J.E.L., A.A., S.T.B., P.J.C., R.L.C., E.W.C., D.R.C., O.D., A.D., E.D.E., S.F., R.R.F., A.S.G., R.G., M.V.H., I.A.H., G.P.J., E.W.K., S.L., N.J.L., N.A.L., K.F.M., S.N.M., L.O., C.A.P., L.V.R., L.R.-T., R.R., K.T.S., T.S., S.T., D.V., D.R.V.E., J.F.P.; Writing-review and editing: J.E.L., A.A., S.T.B., P.J.C., R.L.C., E.W.C., D.R.C., O.D., A.D., E.D.E., S.F., R.R.F., A.S.G., R.G., M.V.H., I.A.H., G.P.J., E.W.K., S.L.,

N.J.L., N.A.L., K.F.M., S.N.M., L.O., C.A.P., L.V.R., L.R.-T., R.R., K.T.S., T.S., S.T., D.V., D.R.V.E., D.A., N.S.A.-H., J.A., H.B., M.B., J.E.B., E.S.B., J.B., W.K.C., J.J.C., J.C., P.D., B.D., S.M.F., C.G., M.L.H., Ha.H., He.H., M.H., E.H., V.H., C.H., M.H.-P., G.H., M.R.I., C.K., D.K., E.E.K., A.K., K.K., B.K., L.K., I.J.K., K.L., C.L., E.M., T.A.M., T.M., E.M.M., F.M., A.M., S.D.M., P.M., B.M., N.M., B.N., E.F.P., M.J.P., T.R.-B., D.M.R., E.A.R., S.S., M.S., D.J.S., Ba.S., L.S., G.Q.S., R.R.S., Br.S., M.E.S., J.W.S., R.S., S.A.S., J.T., H.K.T., S.B.T., T.W., W.-Q.W., Q.S.W., C.W., G.L.W., K.W., J.F.P.

## Ethics Declaration

The eMERGE Genomic Risk Assessment study was approved by a central institutional review board at Vanderbilt University Medical Center (#211043) and acknowledged by the local institutional review boards under reliance agreements. Informed consent is required from all enrolled individuals.



## Conflict of Interest

N.S.A.-H. is an employee and equity holder of 23andMe; serves as a scientific advisory board member for Allelica, Inc; received personal fees from Genentech Inc, Allelica Inc, and 23andMe; received research funding from Akcea Therapeutics; and was previously employed by Regeneron Pharmaceuticals. T.W. has grant funding from Gilead Sciences, Inc. L.O. and T.R.-B are founders of a company developing MeTree. T.S., E.D.E., and E.H. are employees and stockholders of Invitae Corporation. E.M.M. has been a consultant for Avidity Bioscience, Amgen Inc, AstraZeneca, Cytokinetics, Invitae Corporation, Janssen Pharmaceuticals, Pfizer Inc, PepGen Inc, Tenaya Therapeutics, and Stealth BioTherapeutics Inc; she is also the founder of Ikaika Therapeutics. E.E.K. received personal fees from Illumina Inc, 23andMe, and Regeneron Pharmaceuticals and serves as a scientific advisory board member for Encompass Bioscience, Foresite Labs, and Galateo Bio. B.K. is an advisory board member and stockholder of Genome Medical. M.S. is a member of the Institutional Review Board of the All of Us Research Program. E.F.P. is a paid consultant for Allecia Inc. J.F.P. is a paid consultant for Natera Inc. All other authors declare no conflicts of interest.

## Additional Information

The online version of this article (<https://doi.org/10.1016/j.gim.2023.100006>) contains supplementary material, which is available to authorized users.

## Authors

Jodell E. Linder<sup>1</sup> , Aimee Allworth<sup>2</sup>, Sarah T. Bland<sup>3</sup>, Pedro J. Caraballo<sup>4</sup>, Rex L. Chisholm<sup>5</sup>, Ellen Wright Clayton<sup>6</sup>, David R. Crosslin<sup>7</sup>, Ozan Dikilitas<sup>8</sup>, Alanna DiVietro<sup>1</sup>, Edward D. Esplin<sup>9</sup>, Sophie Forman<sup>1</sup>, Robert R. Freimuth<sup>10</sup>, Adam S. Gordon<sup>11</sup>, Richard Green<sup>12</sup>, Maegan V. Harden<sup>13</sup>, Ingrid A. Holm<sup>14</sup>, Gail P. Jarvik<sup>15</sup>, Elizabeth W. Karlson<sup>16</sup>, Sofia Labrecque<sup>1</sup>, Niall J. Lennon<sup>13</sup>, Nita A. Limdi<sup>17</sup>, Kathleen F. Mittendorf<sup>18</sup>, Shawn N. Murphy<sup>19</sup>, Lori Orlando<sup>20</sup>, Cynthia A. Prows<sup>21</sup>, Luke V. Rasmussen<sup>22</sup>, Laura Rasmussen-Torvik<sup>22</sup>, Robb Rowley<sup>23</sup>, Konrad Teodor Sawicki<sup>24</sup>, Tara Schmidlen<sup>9</sup>, Shannon Terek<sup>25</sup>, David Veenstra<sup>26</sup>, Digna R. Velez Edwards<sup>27</sup>, Devin Absher<sup>28</sup>, Noura S. Abul-Husn<sup>29</sup>, Jorge Alsip<sup>30</sup>, Hana Bangash<sup>31</sup>, Mark Beasley<sup>32</sup>, Jennifer E. Below<sup>33</sup>, Eta S. Berner<sup>34</sup>, James Booth<sup>35</sup>, Wendy K. Chung<sup>36</sup>, James J. Cimino<sup>37</sup>, John Connolly<sup>25</sup>, Patrick Davis<sup>12</sup>, Beth Devine<sup>26</sup>, Stephanie M. Fullerton<sup>38</sup>, Candace Guiducci<sup>13</sup>, Melissa L. Habrat<sup>12</sup>, Heather Hain<sup>25</sup>, Hakon Hakonarson<sup>25</sup>, Margaret Harr<sup>25</sup>, Eden Haverfield<sup>9</sup>, Valentina Hernandez<sup>39</sup>, Christin Hoell<sup>40</sup>, Martha Horike-Pyne<sup>2</sup>, George Hripcsak<sup>41</sup>, Marguerite R. Irvin<sup>42</sup>, Christopher Kachulis<sup>13</sup>, Dean Karavite<sup>43</sup>, Eimear E. Kenny<sup>29</sup>, Atlas Khan<sup>44</sup>, Krzysztof Kiryluk<sup>44</sup>, Bruce Korf<sup>45</sup>, Leah Kottyan<sup>46</sup>, Iftikhar J. Kullo<sup>31</sup>, Katie Larkin<sup>13</sup>, Cong Liu<sup>41</sup>, Edyta Malolepsza<sup>13</sup>, Teri A. Manolio<sup>23</sup>, Thomas May<sup>47</sup>, Elizabeth M. McNally<sup>5</sup>, Frank Mentch<sup>25</sup>, Alexandra Miller<sup>31</sup>, Sean D. Mooney<sup>12</sup>, Priyanka Murali<sup>2</sup>, Brenda Mutai<sup>2</sup>, Naveen Muthu<sup>43</sup>, Bahram Namjou<sup>46</sup>, Emma F. Perez<sup>48</sup>, Megan J. Puckelwartz<sup>11</sup>, Tejinder Rakhra-Burris<sup>20</sup>, Dan M. Roden<sup>49</sup>, Elisabeth A. Rosenthal<sup>2</sup>, Seyedmohammad Saadatagah<sup>31</sup>, Maya Sabatello<sup>50</sup>, Dan J. Schaid<sup>51</sup>, Baergen Schultz<sup>23</sup>, Lynn Seabolt<sup>1</sup>, Gabriel Q. Shaibi<sup>52</sup>, Richard R. Sharp<sup>53</sup>, Brian Shirts<sup>54</sup>, Maureen E. Smith<sup>24</sup>, Jordan W. Smoller<sup>55</sup>, Rene Sterling<sup>56</sup>, Sabrina A. Suckiel<sup>57</sup>, Jeritt Thayer<sup>43</sup>, Hemant K. Tiwari<sup>32</sup>, Susan B. Trinidad<sup>38</sup>, Theresa Walunas<sup>58</sup>, Wei-Qi Wei<sup>3</sup>, Quinn S. Wells<sup>59</sup>, Chunhua Weng<sup>41</sup>, Georgia L. Wiesner<sup>33</sup>, Ken Wiley<sup>23</sup>, eMERGE Consortium, Josh F. Peterson<sup>60,\*</sup> 

## Affiliations

<sup>1</sup>Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN; <sup>2</sup>Division of Medical Genetics, Department of Medicine, University of Washington Medical Center, Seattle, WA; <sup>3</sup>Department of Biomedical Informatics and Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; <sup>4</sup>Department of Internal Medicine and Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN; <sup>5</sup>Center for Genetic Medicine,

Northwestern University, Chicago, IL; <sup>6</sup>Center for Biomedical Ethics and Society, Vanderbilt University Medical Center, Nashville, TN; <sup>7</sup>Division of Biomedical Informatics and Genomics, John W. Deming Department of Medicine, Tulane University School of Medicine, New Orleans, LA; <sup>8</sup>Mayo Clinician Investigator Training Program, Department of Internal Medicine and Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN; <sup>9</sup>Invitae, San Francisco, CA; <sup>10</sup>Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN; <sup>11</sup>Department of Pharmacology, Feinberg School of Medicine, and Center for Genetic Medicine, Northwestern University, Chicago, IL; <sup>12</sup>Department of Biomedical Informatics and Medical Education, University of Washington Medical Center, Seattle, WA; <sup>13</sup>Broad Institute of MIT and Harvard, Cambridge, MA; <sup>14</sup>Division of Genetics and Genomics and Manton Center for Orphan Diseases Research, Boston Children's Hospital, Department of Pediatrics, Harvard Medical School, Boston, MA; <sup>15</sup>Division of Medical Genetics, Department of Medicine and Department of Genome Science, University of Washington Medical Center, Seattle, WA; <sup>16</sup>Division of Rheumatology, Inflammation and Immunity, Department of Medicine, Brigham and Women's Hospital, Boston, MA; <sup>17</sup>Department of Neurology, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL; <sup>18</sup>Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN; <sup>19</sup>Department of Neurology, Massachusetts General Hospital, Boston, MA; <sup>20</sup>Center for Applied Genomics and Precision Medicine, Duke University, Durham, NC; <sup>21</sup>Divisions of Human Genetics and Patient Services, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; <sup>22</sup>Department of Preventive Medicine, Northwestern University, Chicago, IL; <sup>23</sup>Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD; <sup>24</sup>Department of Cardiology and Center for Genetic Medicine, Northwestern University, Chicago, IL; <sup>25</sup>Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; <sup>26</sup>School of Pharmacy, University of Washington, Seattle, WA; <sup>27</sup>Division of Quantitative Science, Department of Obstetrics and Gynecology, Department of Biomedical Sciences, Vanderbilt University Medical Center, Nashville, TN; <sup>28</sup>Kaiser Permanente Research Bank, Oakland, CA; <sup>29</sup>Institute for Genomic Health, Department of Medicine, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; <sup>30</sup>UAB Medicine Enterprise, Birmingham, AL; <sup>31</sup>Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN; <sup>32</sup>Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, Birmingham, AL; <sup>33</sup>Division of Genetic Medicine, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; <sup>34</sup>Department of Health Services Administration, University of Alabama at Birmingham, Birmingham, AL; <sup>35</sup>Department of Emergency Medicine, Heersink School of Medicine, University of Alabama at Birmingham,

Birmingham, AL; <sup>36</sup>Departments of Pediatrics and Medicine, Columbia University Irving Medical Center, Columbia University, New York, NY; <sup>37</sup>Division of General Internal Medicine and the Informatics Institute, Department of Medicine, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL; <sup>38</sup>Department of Bioethics and Humanities, University of Washington School of Medicine, Seattle, WA; <sup>39</sup>Mountain Park Health Center, Phoenix, AZ; <sup>40</sup>Department of Obstetrics & Gynecology and Center for Genetic Medicine, Northwestern University, Chicago, IL; <sup>41</sup>Department of Biomedical Informatics, Columbia University Irving Medical Center, Columbia University, New York, NY; <sup>42</sup>Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL; <sup>43</sup>Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA; <sup>44</sup>Division of Nephrology, Department of Medicine, Columbia University Irving Medical Center, New York, NY; <sup>45</sup>Department of Genetics, Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL; <sup>46</sup>The Center for Autoimmune Genomics and Etiology, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; <sup>47</sup>Elson S. Floyd College of Medicine, Washington State University, Vancouver, WA; <sup>48</sup>Department of Medicine, Brigham and Women's Hospital, Mass General Brigham Personalized Medicine, Boston, MA; <sup>49</sup>Departments of Medicine, Pharmacology, and Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; <sup>50</sup>Division of Nephrology, Department of Medicine & Division of Ethics, Department of Medical Humanities and Ethics, Columbia University Irving Medical Center, New York, NY; <sup>51</sup>Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN; <sup>52</sup>Center for Health Promotion and Disease Prevention, Arizona State University, Phoenix, AZ; <sup>53</sup>Biomedical Ethics Program, Department of Quantitative Health Science, Mayo Clinic, Rochester, MN; <sup>54</sup>Department of Laboratory Medicine & Pathology, University of Washington Medical Center, Seattle, WA; <sup>55</sup>Department of Psychiatry and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; <sup>56</sup>Division of Genomics and Society, National Human Genome Research Institute, Bethesda, MD; <sup>57</sup>The Institute for Genomic Health, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; <sup>58</sup>Department of Medicine and Center for Health Information Partnerships, Northwestern University, Chicago, IL; <sup>59</sup>Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; <sup>60</sup>Center for Precision Medicine, Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

## The eMERGE Consortium

Adam Gordon, Agboade Sobowale, Aimee Allworth, Akshar Patel, Alanna DiVietro, Alanna Strong, Alborz Sherafati,

Alborz Sherfati, Alex Bick, Alexandra Miller, Alka Chandel, Alyssa Rosenthal, Amit Khera, Amy Kontorovich, Andrew Beck, Andy Beck, Angelica Espinoza, Anna Lewis, Anya Prince, Atlas Khan, Ayuko Iverson, Bahram Namjou Khales, Barbara Benoit, Becca Hernan, Ben Kallman, Ben Kerman, Ben Shoemaker, Benjamin Satterfield, Beth Devine, Bethany Etheridge, Blake Goff, Bob Freimuth, Bob Grundmeier, Brenae Collier, Brenda Mutai, Brett Harnett, Brian Chang, Brian Piening, Brittney Davis, Bruce Korf, Candace Patterson, Carmen Demetriou, Casey Ta, Catherine Hammack, Catrina Nelson, Caytie Gascoigne, Chad Dorn, Chad Moretz, Chris Kachulis, Christie Hoell, Christine Cowles, Christoph Lange, Chunhua Weng, Cindy Prows, Cole Brokamp, Cong Liu, Courtney Scherr, Crystal Gonzalez, Cynthia Ramirez, Daichi Shimbo, Dan Roden, Daniel Schaid, Dave Kaufman, David Crosslin, David Kochan, David Veenstra, Davinder Singh, Dean Karavite, Debbie Abrams, Devin Absher, Digna Velez Edwards, Eden Haverfield, Eduardo Morales, Edward Esplin, Edyta Malolepsza, Ehsan Alipour, Eimear Kenny, Elisabeth Rosenthal, Eliza Duvall, Elizabeth McNally, Elizabeth Bhoj, Elizabeth Cohn, Elizabeth Hibler, Elizabeth Karlson, Elizabeth McNally, Ellen Clayton, Emily Chesnut, Emily DeFranco, Emily Gallagher, Emily Soper, Emma Perez, Erin Cash, Eta Berner, Fei Wang, Firas Wehbe, Francisco Ricci, Frank Mentch, Gabriel Shaibi, Gail Jarvik, George Hahn, George Hripsak, Georgia Wiesner, Gillian Belbin, Gio Davogustto, Girish Nadkarni, Haijun Qiu, Hakon Hakonarson, Hana Bangash, Hannah Beasley, Hao Liu, Heide Aungst, Hemant Tiwari, Hillary Duckham, Hope Thomas, Iftikhar Kullo, Ingrid Holm, Isabelle Allen, Iuliana Ionita-Laza, Jacklyn Hellwege, Jacob Petzelka, Jacqueline Odgis, Jahnvi Narula, Jake Petzelka, Jalpa Patel, James Cimino, James Meigs, James Snyder, Janet Olson, Janet Zahner, Jeff Pennington, Jen Pacheco, Jennifer Allen Pacheco, Jennifer Morse, Jeremy Corsmo, Jeritt Thayer, Jim Cimino, Jingheng Chen, Jocelyn Fournier, Jodell Jackson, Joe Glessner, Joel Pacyna, Johanna Smith, John Connolly, John Lynch, John Shelley, Jonathan Mosley, Jordan Nestor, Jordan Smoller, Jorge Alsip, Joseph Kannry, Joseph Sutton, Josh Peterson, Joshua Smith, Julia Galasso, Julia Smith, Julia Wynn, Justin Gundelach, Justin Starren, Karmel Choi, Kate Mittendorf, Katherine Anderson, Katherine Bonini, Kathleen Leppig, Kathleen Muenzen, Katie Larkin, Kelsey Stuttgart, Ken Wiley, Kenny Nguyen, Kevin Dufendach, Kiley Atkins, Konrad Sawicki, Kristjan Norland, Krzysztof Kiryluk, Laura Beskow, Laura Rasmussen-Torvik, Leah Kottyan, Li Hsu, Lifeng Tian, Lisa Mahanta, Lisa Martin, Lisa Wang, Lizbeth Gomez, Lorenzo Thompson, Lori Orlando, Lucas Richter, Luke Rasmussen, Lynn Petukhova, Lynn Seabolt, Madison O'Brien, Maegan Harden, Malia Fullerton, Margaret Harr, Mark Beasley, Marta Guindo, Martha Horike, Martha Horike-Pyne, Marwah Abdalla, Marwan Hamed, Mary Beth Terry, Mary Maradik, Matt Wyatt, Matthew Davis, Matthew Lebo, Maureen Smith, Maya del Rosario, Maya Sabatello, Meckenzie Behr, Meg Roy-Puckelwartz, Mel Habrat, Melanie Myers, Meliha Yetisgen, Merve Iris, Michael DaSilva, Michael Preuss, Michelle McGowan, Mingjian Shi, Minoli

Perera, Minta Thomas, Mitch Elkind, Mohammad Abbass, Mohammad Saadatagah, Molly Hess, Molly Maradik, Nataraja “RJ” Vaitinadin, Nataraja Vaitinadin, Naveen Muthu, Neil Netherly, Niall Lennon, Ning Shang, Nita Limdi, Noah Forrest, Noheli Romero, Nora Robinson, Noura Abul-Husn, Omar Elsekaily, Ozan Dikilitas, Patricia Kovatch, Patrick Davis, Paul Appelbaum, Paul Francaviglia, Paul O’Reilly, Paulette Chandler, Pedro Caraballo, Peter Tarczy-Hornoch, Pierre Shum, Priya Marathe, Priyanka Murali, Qiping Feng, Quinn Wells, Rachel Atchley, Radhika Narla, Rene Barton, Rene Sterling, Rex Chisholm, Richard Green, Richard Sharp, Riki Peters, Rita Kukafka, Robb Rowley, Robert Freimuth, Robert Green, Robert Winter, Roger Mueller, Ruth Loos, Ryan Irvin, Sabrina Suckiel, Sajjad Hussain, Samer Sharba, Sandy Aronson, Sarah Jones, Sarah Knerr, Scott Nigbur, Scott Weiss, Sean Mooney, Shannon Terek, Sharon Aufox, Sharon Nirenberg, Shawn Murphy, Sheila O’Byrne, Shing Wang (Sam) Choi, Sienna Aguilar, ST Bland, Stefanie Rodrigues, Stephanie Ledbetter, Stephanie Rutledge, Stuart James Booth, Su Xian, Susan Brown Trinidad, Suzanne Bakken, Tara Schmidlen, Tejinder Rakhra-Burris, Teri Manolio, Tesfaye Mersha, Theresa Walunas, Thevaa Chandereeng, Thomas May, Tian Ge, Todd Edwards, Tom Kaszemacher, Valentina Hernandez, Valerie Willis, Vemi Desai, Vimi Desai, Virginia Lorenzi, Vivian Gainer, Wei-Qi Wei, Wendy Chung, Wu-Chen Su, Xiao Chang, Yiqing Zhao, Yuan Luo, Yufeng Shen.

## References

- Daly MB, Pilarski R, Yurgelun MB, et al. NCCN guidelines insights: genetic/familial high-risk assessment: breast, ovarian, and pancreatic, version 1.2020. *J Natl Compr Canc Netw*. 2020;18(4):380-391. <http://doi.org/10.6004/jnccn.2020.0017>
- Gupta S, Provenzale D, Regenbogen SE, et al. NCCN guidelines insights: genetic/familial high-risk assessment: colorectal, version 3. 2017. *J Natl Compr Canc Netw*. 2017;15(12):1465-1475. <http://doi.org/10.6004/jnccn.2017.0176>
- Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/AphA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. *Circulation*. 2019;139(25):e1082-e1143. Published correction appears in *Circulation*. 2019;139(25):e1182-e1186. <https://doi.org/10.1161/CIR.0000000000000625>
- Hattersley AT, Greeley SAW, Polak M, et al. ISPAD clinical practice consensus guidelines 2018: the diagnosis and management of monogenic diabetes in children and adolescents. *Pediatr Diabetes*. 2018;19(suppl 27):47-63. <http://doi.org/10.1111/pedi.12772>
- Murray MF, Khoury MJ, Abul-Husn NS. Addressing the routine failure to clinically identify monogenic cases of common disease. *Genome Med*. 2022;14(1):60. <http://doi.org/10.1186/s13073-022-01062-6>
- Grzymalski JJ, Elhanan G, Morales Rosado JA, et al. Population genetic screening efficiently identifies carriers of autosomal dominant diseases. *Nat Med*. 2020;26(8):1235-1239. <http://doi.org/10.1038/s41591-020-0982-5>
- Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet*. 2009;25(11):489-494. <http://doi.org/10.1016/j.tig.2009.09.012>
- Landry LG, Ali N, Williams DR, Rehm HL, Bonham VL. Lack of diversity in genomic databases is a barrier to translating precision medicine research into practice. *Health Aff (Millwood)*. 2018;37(5):780-785. <http://doi.org/10.1377/hlthaff.2017.1595>
- Khera AV, Chaffin M, Zekavat SM, et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation*. 2019;139(13):1593-1602. <http://doi.org/10.1161/CIRCULATIONAHA.118.035658>
- McCarty CA, Chisholm RL, Chute CG, et al. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13. <http://doi.org/10.1186/1755-8794-4-13>
- Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. 2013;15(10):761-771. <http://doi.org/10.1038/gim.2013.72>
- eMERGE consortium. Lessons from the eMERGE Network: balancing genomics in discovery and in practice. *HGG Adv*. 2020;2(1):100018. <http://doi.org/10.1016/j.xhgg.2020.100018>
- Denny JC, Ritchie MD, Basford MA, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26(9):1205-1210. <http://doi.org/10.1093/bioinformatics/btq126>
- Ritchie MD, Verma SS, Hall MA, et al. Electronic medical records and genomics (eMERGE) network exploration in cataract: several new potential susceptibility loci. *Mol Vis*. 2014;20:1281-1295.
- Dumitrescu L, Ritchie MD, Denny JC, et al. Genome-wide study of resistant hypertension identified from electronic health records. *PLoS One*. 2017;12(2):e0171745. <http://doi.org/10.1371/journal.pone.0171745>
- Crawford DC, Crosslin DR, Tromp G, et al. eMERGEing progress in genomics-the first seven years. *Front Genet*. 2014;5:184. <http://doi.org/10.3389/fgene.2014.00184>
- Crosslin DR, Robertson PD, Carrell DS, et al. Prospective participant selection and ranking to maximize actionable pharmacogenetic variants and discovery in the eMERGE Network. *Genome Med*. 2015;7(1):67. <http://doi.org/10.1186/s13073-015-0181-z>
- Fossey R, Kochan D, Winkler E, et al. Ethical considerations related to return of results from genomic medicine projects: the eMERGE network (phase III) experience. *J Pers Med*. 2018;8(1):2. <http://doi.org/10.3390/jpm8010002>
- Williams JL, Chung WK, Fedotov A, et al. Harmonizing outcomes for genomic medicine: comparison of eMERGE outcomes to ClinGen outcome/intervention pairs. *Healthcare (Basel)*. 2018;6(3):83. <http://doi.org/10.3390/healthcare6030083>
- Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284-1287. <http://doi.org/10.1038/ng.3656>
- Schatz MC, Philippakis AA, Afgan E, et al. Inverting the model of genomics data sharing with the NHGRI Genomic Data Science Analysis, Visualization, and Informatics Lab-space. *Cell Genom*. 2022;2(1):100085. <http://doi.org/10.1016/j.xgen.2021.100085>
- eMERGE Consortium. Harmonizing clinical sequencing and interpretation for the eMERGE III network. *Am J Hum Genet*. 2019;105(3):588-605. <http://doi.org/10.1016/j.ajhg.2019.07.018>
- Stanaway IB, Hall TO, Rosenthal EA, et al. 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*. 2019;43(1):63-81. <http://doi.org/10.1002/gepi.22167>
- Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. <http://doi.org/10.1038/s41588-018-0183-z>
- Centers for Disease Control and Prevention. Tier 1 genomics applications and their importance to public health. Centers for Disease Control and Prevention. Published March 16, 2021. Accessed June 15, 2022. <https://www.cdc.gov/genomics/implementation/toolkit/tier1.htm>

26. Breast Cancer Association Consortium, Dorling L, Carvalho S, et al. Breast cancer risk genes – association analysis in more than 113,000 women. *N Engl J Med*. 2021;384(5):428-439. <http://doi.org/10.1056/NEJMoa1913948>
27. Haverfield EV, Esplin ED, Aguilar SJ, et al. Physician-directed genetic screening to evaluate personal risk for medically actionable disorders: a large multi-center cohort study. *BMC Med*. 2021;19(1):199. Published correction appears in *BMC Med*. 2021;19(1):288. <https://doi.org/10.1186/s12916-021-01999-2>
28. Orlando LA, Buchanan AH, Hahn SE, et al. Development and validation of a primary care-based family health history and decision support program (MeTree). *N C Med J*. 2013;74(4):287-296.
29. Lee A, Mavaddat N, Wilcox AN, et al. BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. *Genet Med*. 2019;21(8):1708-1718. Published correction appears in *Genet Med*. 2019;21(6):1462. <http://doi.org/10.1038/s41436-018-0406-9>
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377-381. <http://doi.org/10.1016/j.jbi.2008.08.010>
31. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform*. 2019;95:103208. <http://doi.org/10.1016/j.jbi.2019.103208>
32. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129(25 suppl 2):S49-S73. Published correction appears in *Circulation*. 2014;129(25 Suppl 2):S74-S75. <https://doi.org/10.1161/01.cir.0000437741.48606.98>
33. Lee A, Mavaddat N, Cunningham A, et al. Enhancing the BOADICEA cancer risk prediction model to incorporate new data on *RAD51C*, *RAD51D*, *BARD1* updates to tumour pathology and cancer incidence. *J Med Genet*. 2022;59(12):1206-1218. <http://doi.org/10.1136/jmedgenet-2022-108471>
34. Lewis ACF, Perez EF, Prince AER, et al. Patient and provider perspectives on polygenic risk scores: implications for clinical reporting and utilization. *Genome Med*. 2022;14(1):114. <http://doi.org/10.1186/s13073-022-01117-8>
35. Cook DA, Teixeira MT, Heale BS, Cimino JJ, Del Fiol G. Context-sensitive decision support (infobuttons) in electronic health records: a systematic review. *J Am Med Inform Assoc*. 2017;24(2):460-468. <http://doi.org/10.1093/jamia/ocw104>
36. Ginsburg GS, Wu RR, Orlando LA. Family health history: underused for actionable risk assessment. *Lancet*. 2019;394(10198):596-603. [http://doi.org/10.1016/S0140-6736\(19\)31275-9](http://doi.org/10.1016/S0140-6736(19)31275-9)
37. Polubriaginof F, Tatonetti NP, Vawdrey DK. An assessment of family history information captured in an electronic health record. *AMIA Annu Symp Proc*. 2015;2015:2035-2042.
38. Aragam KG, Dobbyn A, Judy R, et al. Limitations of contemporary guidelines for managing patients at high genetic risk of coronary artery disease. *J Am Coll Cardiol*. 2020;75(22):2769-2780. <http://doi.org/10.1016/j.jacc.2020.04.027>
39. Wojcik GL, Graff M, Nishimura KK, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*. 2019;570(7762):514-518. <http://doi.org/10.1038/s41586-019-1310-4>
40. Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021;600(7890):675-679. <http://doi.org/10.1038/s41586-021-04064-3>