AHA Science Advisory

Merging Electronic Health Record Data and Genomics for Cardiovascular Research

A Science Advisory From the American Heart Association

Jennifer L. Hall, PhD, FAHA, Chair; John J. Ryan, MD, FAHA, Co-Chair; Bruce E. Bray, MD; Candice Brown, PhD; David Lanfear, MD; L. Kristin Newby, MD, MHS; Mary V. Relling, PharmD; Neil J. Risch, PhD; Dan M. Roden, MD; Stanley Y. Shaw, MD, PhD; James E. Tcheng, MD; Jessica Tenenbaum, PhD; Thomas N. Wang, MD, PhD; William S. Weintraub, MD, FAHA; on behalf of the American Heart Association Professional and Public Education and Publications Committee of the Council on Functional Genomics and Translational Biology, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Quality of Care and Outcomes Research, and Stroke Council

Abstract—The process of scientific discovery is rapidly evolving. The funding climate has influenced a favorable shift in scientific discovery toward the use of existing resources such as the electronic health record. The electronic health record enables long-term outlooks on human health and disease, in conjunction with multidimensional phenotypes that include laboratory data, images, vital signs, and other clinical information. Initial work has confirmed the utility of the electronic health record for understanding mechanisms and patterns of variability in disease susceptibility, disease evolution, and drug responses. The addition of biobanks and genomic data to the information contained in the electronic health record has been demonstrated. The purpose of this statement is to discuss the current challenges in and the potential for merging electronic health record data and genomics for cardiovascular research. (***Circ Cardiovasc Genet*****. 2016;9:193-202. DOI: 10.1161/HCG.0000000000000029.)**

Key Words: AHA Scientific Statements ■ electronic health records ■ environment and public health ■ genetic research ■ genetics ■ genomics

Electronic health records (EHRs) have assumed a major role in medical practice in the United States.[1] EHRs also have the potential to accommodate genetic and genomic data in a manner similar to how they handle clinical laboratory data. Including genomic data with the clinical information in EHRs provides the potential to improve our understanding of the underlying mechanisms of health and disease and to improve the overall care of patients. The announcement of the Precision Medicine Initiative by the US government reaffirms the change in patient care on the horizon.[2,3] In this report, we summarize the existing landscape and current hurdles of genomic research in cardiovascular disease in the era of the EHR.

Genetic Findings in Cardiovascular Research: Historical Perspective

The goal of this advisory statement is to assess the utility of using the EHR for cardiovascular genetic research. Landmark genetic findings in cardiovascular research have come from

The American Heart Association makes every effort to avoid any actual or potential conflicts of interest that may arise as a result of an outside relationship or a personal, professional, or business interest of a member of the writing panel. Specifically, all members of the writing group are required to complete and submit a Disclosure Questionnaire showing all such relationships that might be perceived as real or potential conflicts of interest.

This Science Advisory was approved by the American Heart Association Science Advisory and Coordinating Committee on September 29, 2015, and the American Heart Association Executive Committee on October 27, 2015. A copy of the document is available at http://professional.heart.org/statements by using either "Search for Guidelines & Statements" or the "Browse by Topic" area. To purchase additional reprints, call 843-216-2533 or e-mail kelle.ramsay@wolterskluwer.com.

The Appendix is available with this article at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/HCG.0000000000000029/-/DC1.

The American Heart Association requests that this document be cited as follows: Hall JL, Ryan JJ, Bray BE, Brown C, Lanfear D, Newby LK, Relling MV, Risch NJ, Roden DM, Shaw SY, Tcheng JE, Tenenbaum J, Wang TN, Weintraub WS; on behalf of the American Heart Association Professional and Public Education and Publications Committee of the Council on Functional Genomics and Translational Biology, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Quality of Care and Outcomes Research, and Stroke Council. Merging electronic health record data and genomics for cardiovascular research: a science advisory from the American Heart Association. *Circ Cardiovasc Genet.* 2016;9:193–202. DOI: 10.1161/HCG.0000000000000029.

Expert peer review of AHA Scientific Statements is conducted by the AHA Office of Science Operations. For more on AHA statements and guidelines development, visit http://professional.heart.org/statements. Select the "Guidelines & Statements" drop-down menu, then click "Publication Development."

both single-center studies and large consortia. In the majority of these studies, DNA from blood samples in a biorepository was linked to clinical information in either a paper medical record or a study record or file. We are moving to an era of EHR-coupled biobanks.

An EHR-coupled biobank is a bank of biological specimens linked to data in the EHR. The biological samples can be DNA or other biospecimens such as serum, plasma, or tissue. The nature of the clinical information varies widely. For example, biobanks may be collections of convenience, derived by collecting samples from large numbers of patients with a single disease (heart disease, cancer, type 2 diabetes mellitus). Others are community based, and the phenotypic information is generally acquired by detailed structured questionnaires and examinations performed at enrollment and often at intervals thereafter. The development and increasing permeation of EHR systems into clinical practice environments provide another source of phenotypic information for biorepositories.[4] Accrual models vary. In some instances, patients are recruited from specific practice sites (eg, internal medicine or specialty clinics), whereas other programs have adopted an "all-comers" approach. Advantages include the potential to accrue large and relatively disease-agnostic data sets. Initial studies support the concept that because genomic and phenotypic data are in place, large data sets can be generated more rapidly and inexpensively than in conventional clinical trials.[5]

The EHR may contain more clinical information than the earlier approaches. Genetic information may be coupled to the EHR for approved studies. Improved clinical data and juxtaposed access to genetic information may lead to better studies. In addition, EHRs supply large numbers of patients with longitudinal data that may improve the ability to separate true-positive from false-positive associations. A number of studies have now identified rare variants with remarkably large effect sizes for traits such as disease susceptibility,[6–9] but large numbers are required to enable this approach. For example, EHR-derived data have helped identify single-nucleotide polymorphisms in 9p21 that are noted to be associated with early myocardial infarction and cardiovascular disease.[10]

One early concern about EHR-based genomic research was the robustness of the phenotypes that might be extractable. Multiple studies have now demonstrated the ability of electronic phenotyping to replicate known genomic signals and to identify new ones.[11] Many of these findings are ancestry specific, so future genomic research will require inclusion of multiple ancestries.

EHR-based biobanks have drawbacks. One drawback is that the subjects included are those who encounter a healthcare system and thus may not represent an entire community. Some information such as dietary or family histories may not be systematically obtained in an EHR, although recognizing this shortcoming may be a first step to expanding the collection of such data in EHR systems. The ancestry representation in an EHR will reflect local patient populations; thus, creation of networks may be a mechanism to include multiple ancestries. A need exists in the EHR to develop efficient and effective methods to collect and record family history information between patient charts and to allow the integration updates in family history into the EHR. Additionally, the absence of standardized human phenotype ontology is a drawback to deriving information from the EHR, and standardized human phenotype ontology needs to be developed in line with EHRs, especially as it pertains to hereditable diseases.[12] Finally, the development of standards for storing and displaying genomic information in an EHR is evolving, as well as the development of clinical decision support tools to advise EHR users on how to act on genomic variant data.

Consideration of the Patient

The use of the EHR for research involves a variety of stakeholders, including patients, researchers, clinicians, healthcare systems, and funders of clinical care and research. Given that the principal purpose of the EHR is to facilitate patient care and to improve patient outcomes, the interests of the patients are paramount in the consideration of the benefits, risks, and acceptability of EHR-based research. Patients are generally supportive of biomedical research involving "big data," as evidenced by their positive view of such research in national surveys. For instance, in a survey of 4659 US adults, 73% of respondents indicated that they would be willing to the release their medical records for research, and 84% supported the idea of large, national cohort studies.[13] A study conducted by the Patient Centered Outcomes Research Institute found that 66% of patient-respondents had interest in engaging in research, and 83% felt that direct involvement with investigators could lead to more valuable research.[14] Nevertheless, as the field continues to progress, understanding the role of the patient in the process is essential. It is important for participants to be aware that the personal benefits to patients for involvement in EHR-based genomic researcher are uncertain and most likely societal in nature. Use of healthcare data in research generally raises specific privacy concerns on the part of participants; thus, EHR-coupled biobanks have undergone extensive review, usually involving not only local ethical review groups, but also community consultation.

New Models

Initial studies support the concept that because genomic and phenotypic data are in place, large data sets can be generated more rapidly and inexpensively than in conventional clinical trials or cohort studies.[5] The recently described experimental approach of phenome-wide association study seems especially well suited to EHR-based biobanks.[15] In contrast to a conventional genome-wide association study, which interrogates in an unbiased fashion multiple genomic variants as a function of a target phenotype, phenome-wide association study interrogates multiple phenotypes as a function of a target genotype. The phenotypes interrogated to date have been based on diagnostic codes, but in the future, more sophisticated approaches may be used to diagnosis. Phenome-wide association study not only can replicate known genetic associations but also can identify multiple phenotypes associated with specific genomic variants.

In 2007, the National Human Genome Research Institute created the Electronic Medical Records and Genomics

Network, which now includes 10 EHR-based DNA repositories and >350 000 subjects.[11,16] The Electronic Medical Records and Genomics Network has demonstrated that phenotype definitions can be successfully deployed across multiple EHR systems.[4,17] As of June 2014, the Kaiser Permanente Northern California Research Program on Genes, Environment and Health biobank[18] included ≈200 000 consented subjects with saliva or blood samples linked to comprehensive longitudinal EHR data and self-reported demographic and behavioral information. A subset of 110 266 of these individuals have genome-wide genotype and telomere length data available, forming the Genetic Epidemiology Research on Adult Health and Aging cohort. Data from 78 486 of these individuals are now available through dbGaP (dbGaP Study Accession; phs000674.v1.p1).

A number of other resources have coupled DNA or other biosamples to the EHR in >100 000 subjects. One example is Vanderbilt BioVU, which has pioneered an "opt-out" model in which DNA extracted from discarded blood samples obtained during routine care is coupled to deidentified EHRs.[19] The BioVU program is accompanied by extensive publicity, including a formal opt-out opportunity provided to patients yearly.[20] The opt-out approach has the advantage of scale (>185 000 subjects as of June 2014). The Precision Medicine Initiative will likely lead to important changes in the regulatory framework and paperwork that encourage patient participation.[2]

Biobanks based in countries with EHRs and single-payer systems such as the Danish national biobank[21] also have the potential to generate large data sets. The US Million Veterans Program[22] and the Kadoorie Biobank in China[23] are population-based cohorts with 500 000 subjects that are working to couple to national EHR systems. Although the UK Biobank[24] was not originally created with a link to EHRs, health outcome data are being added to the resource.

Collecting Data for Research in the EHR

Hospital and clinical practice environments are complex systems with multiple streams of data. The EHR is a mix of structured and narrative text data. Structured data (entered into a designated data field or using a controlled vocabulary) typically consist of billing codes, laboratory tests, medication prescriptions, and certain standardized document elements (eg, height, weight, vital signs, problem lists). EHR billing codes include diagnosis-related groups to categorize hospitalizations, *International Classification of Disease* (ninth revision, soon to be replaced by the 10th revision) codes to describe diagnoses and morbidities, and Current Procedural Terminology codes to describe procedures. Narrative or text data constitute the bulk of provider notes, especially those portions entered as "free" or unstructured text; the technique of natural language processing can be used to create structured data sets from notes such as these. The EHR may also contain scanned data in analog form such as a radiographic image or scanned text document, but these forms cannot easily be searched for content.

The counterpart to EHR data is administrative data, which include data on healthcare delivery, insurance enrollment, and claims data for reimbursement. Administrative data are generally structured and often encompass a large number of patients, making them historically attractive as a source for analyses. The most widely used administrative data set is likely from the Centers for Medicare and Medicaid Services, which has been widely used for outcomes research. There is a worldwide effort to develop standardized terminologies that facilitate interoperability, unambiguous data exchange, and interpretation across various sources of healthcare data. The Appendix provides an in-depth review of interoperability.

There is a need for structured data elements in clinical documentation to ensure more robust data collection that will allow pooling of data across providers and institutions. Systematized Nomenclature of Medicine–Clinical Terms is a hierarchical data set of >311 000 concepts covering much of medicine, including symptoms, diagnoses, procedures, anatomical locations, and medications. Systematized Nomenclature of Medicine–Clinical Terms cross-maps to other terminologies such as Logical Observations Identifier Names and Codes, a universal set of codes for laboratory tests, vital signs, survey instruments, and other clinical measurements; RxNorm for pharmaceuticals; and *International Classification of Disease* billing codes. New data standards produced by the American College of Cardiology/American Heart Association Task Force on Data Standards will be linked to Systematized Nomenclature of Medicine–Clinical Terms, Logical Observations Identifier Names and Codes, and RxNorm. In the future, such data standards will be machine ready, reducing the need for coding from scratch for each application. These clinical data standards from the task force can then be used for registries, clinical trials, and structured data within EHRs. Mapping clinical data standards to Systematized Nomenclature of Medicine–Clinical Terms, Logical Observations Identifier Names and Codes, and RxNorm (and in parallel mapping administrative data to these same sets of terms) allows interoperability between clinical and administrative data to be achieved. Interoperability on this scale remains a conceptual model to be implemented over the next several years.

Adding Genomics to the EHR

Principal among the relevant limitations and risks for merging genomic data with EHR data for research purposes are computational burden and information security versus accessibility. Fully processed genome-wide genotyping data can be quite large, roughly 1.5 to 2 MB per person. DNA sequencing and RNA sequencing have storage requirements that easily jump into the terabyte range for even a small number of samples.[25] High-level genomic analysis is computationally demanding, often requiring specialized computing infrastructure even for relatively focused projects. In addition, the EHR needs to have the capacity to handle files that can be several gigabytes in size so that the clinical care, for which the EHR is designed, does not become impeded by these oversized genomic files.[4]

Data security breach is an important concern, and the need for access to information and the goal of keeping it secure are often at odds. Genetic data breaches, particularly if comingled with clinical data, might be expected to be even larger

in scale, taking into account the inherently identifiable nature of DNA.[26]

The ideal situation may be one in which genomic data are stored along with the EHR data. This would give the broadest access, enabling widespread use and collaboration (clinicians, scientists, patients) and, we hope, readying the infrastructure for a future in which genetic data routinely enhance clinical care. This approach has been used at some centers, for example, in association with the Electronic Medical Records and Genomics consortium,[27] but it requires substantial investment to handle the computational burden[28] and some accommodations in terms of information security. With widespread accessibility and given its sensitivity and vastness, one approach is to deidentify the data sets. Thus, the ethical and breach issues are somewhat mitigated because subjects are not easily identified. With this compromised format, however, identifying subjects for additional study or data collection would be difficult or impossible.

An alternative strategy is to take a more incremental approach and keep the 2 data sets separate, bringing them together in analytic subsets of data, unified for specific analyses. This allows limitation and stricter control of who has access to the genomic data and a more insular storage format (ie, these data do not need to be connected to clinical data sources or network), more easily allowing heightened security.

From a practical standpoint, it is often simpler to keep the genomic data separate from the clinical EHR and to create a devoted repository that can be directed at specific projects and questions as needed. There is growing interest in and application of unified genomic and clinical data sets for research, and it is likely that this approach will be increasingly used as some of the challenges are overcome and the current pioneering centers become templates for others. Another facet that will be addressed as more centers emerge is the development of genomic data harmonization standards to facilitate collaborations.[29] Developing standards on the genomic side will facilitate cross-validation of findings and the construction of mega data sets.

Can Clinicians Use EHR and Genomics for Improving Patient Care?

Clinical pharmacogenetics provides information on the role of genetic testing to improve patient care. Some genetic variants have already reached the point of clinical utility for drug prescribing.[30–35] The goal of implementing pharmacogenomic testing is to improve outcomes for patients by minimizing the use of ineffective medications or ineffective doses and to minimize adverse effects. Pharmacogenomic research has identified some genes linked to efficacy or adverse effects of specific drugs. Several centers now use multigene sequencing or array-based methods to interrogate genomic variation for a variety of purposes, and such results are increasingly finding their way into the EHR. It is anticipated that data from such systems may be useful for identifying healthcare benefits. Thus, clinicians are faced with the challenge of knowing which gene variants are actionable for guiding prescribing, even gene variants that may have been identified as incidental to the primary purpose of the genetic testing.

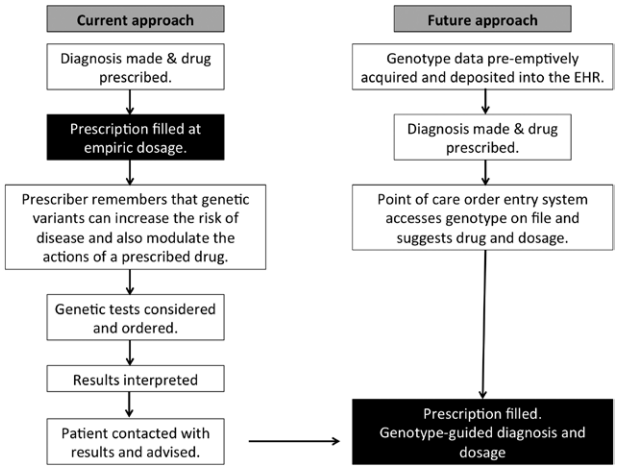

Figure. Current and future approaches to filling prescriptions. Genetic test results obtained from electronic health records (EHRs) could guide diagnosis and suggest drug dosages.

The Clinical Pharmacogenetics Implementation Consortium has prioritized genes and drugs that are actionable, maintains an updated list of these genes and drugs,[36] and has developed a number of clinical guidelines for using pharmacogenetic test results to guide prescribing,[37–43] including the prescription of several cardiovascular agents.[44–48] Clinical Pharmacogenetics Implementation Consortium guidelines do not address whether clinicians should order genetic tests but rather how to use genetic test results to guide prescribing. Thus, these guidelines are consistent with the notion of preemptive genomic testing: the generation of genetic test results before the need to use the results for specific drug-prescribing decisions. Several clinical sites are now performing some form of preemptive pharmacogenetic testing, which capitalizes on EHRs for the process of clinical implementation.[49–52] An EHR is needed to accommodate the active and passive clinical decision support that is important for interpreting and acting on genetic test results.[53,54] Such systems must be extensible to allow reinterpretation of genetic test results as new knowledge is generated. Active interruptive clinical decision support can alert the prescriber in the pretest state (a high-risk drug is being contemplated but the appropriate genetic test is not yet available in the EHR) and in the posttest state (a high-risk genetic test result is already in the EHR).[54] In the future, interoperability of EHRs could allow genetic test results to be available at all relevant points of care, including dispensing pharmacies and outpatient clinics (Figure).

Preemptive testing has the advantage of being available at the time of prescribing; there is no need to wait for the genetic test result. On the other hand, genetics will not solve all prescribing problems. Of the ≈1000 US Food and Drug Administration–approved molecular drug entities, only ≈60 to 100 are candidates for clinical actionability on the basis of germline genetic testing. Although every individual could harbor at least 1 high-risk genotype,[51] the chance that any individual has a high-risk genotype and receives a high-risk drug is low. Data from preemptive genotyping programs will help address the utility of the approach. In addition, even for pharmacogenetically high-risk drugs, other factors such as

Table. Advantages and Disadvantages of Informed Consent Versus Waiver of Consent Models for Biobanking

	Approach	Pros	Cons
Informed consent	Participant signs explicit informed consent form.	Most ethically rigorous. Ability to recontact participant.[63,64]	Requires more time, effort, and infrastructure. Generally yields lower number of participants.
Waiver of consent	IRB grants waiver of consent. Language may be inserted into permission to treat form with opt-in/opt-out check box.	Ability to accrue large numbers of specimens, particularly under opt-out model.	Individual cannot be recontacted.

IRB indicates institutional review board.

concurrent illnesses, drug-drug interactions, poor adherence to therapy, age, nutrition, and liver and renal function have critical effects on the probability of efficacy and toxicity. The cost-effectiveness of preemptive testing is not well defined, although with the declining cost of genomic testing[55] and the development of computational systems with active clinical decision support,[54] there could potentially be a role for preemptive genotyping for personalized medicine.[56] However, the ethics and medicolegal aspects surrounding this issue need to be addressed. Another issue worth considering is the rapid flux in the interpretation of genetic variants and who has responsibility for providing updated, dynamic information to patients and their providers on the identification of implicated genes.

Institutional Review Boards and Participant Consent

Institutional review boards (IRBs) are charged with ensuring the moral, legal, and ethical treatment of human research participants. This mandate includes regulation of tissue samples and genetic data. Local IRBs must ensure that data collection and distribution procedures comply with ethical standards. They must ensure that the needs of all stakeholders, including patients and their families, are considered in all aspects of the decision-making process.[57,58] One of the largest challenges in genomic medicine is developing a series of best practices that will unify local IRB policies to facilitate sharing of EHR data across national and international biorepositories. Two Electronic Medical Records and Genomics working groups have used cross-disciplinary strategies to approach this problem. The Genetics Research and Review Project consortium was designed with the goal of documenting current issues associated with genetic and genomic research among local IRBs, genetic researchers, and bioethicists and developing strategies to overcome current challenges facing IRBs.[58–61] The Consent, Education, Regulation, and Consultation Working Group was developed in response to these challenges, and it is currently developing a core set of genomic research review guidelines for local IRBs.[27,59,62]

Approaches to the creation of large biorepositories vary and have their unique advantages and disadvantages (Table).[20,65] It is important to create standards in data sharing from biorepositories, and much of these standards could be modeled from the sharing of clinical trial data, although there are unique aspects of biorepositories, especially pertaining to consent, that are not present in clinical trials.[66] Therefore, engaging and educating all stakeholders, especially community members, in the biorepository planning and implementation processes are crucial for the successful integration of genomic medicine with EHRs.[57]

Data Analytics and Visualization

For analyses of medical big data to be valid and relevant, key characteristics of the analytics environment itself must be present. The analytics environment is tightly integrated into the existing data architecture.[67] Key properties must be in place to help ensure the applicability of analytic findings to both specific individuals and populations and to extend and scale the analyses beyond singular applications[68,69] (Appendix, Table A1). Given the complexity of both clinical and genomic data, computerized clinical decision support has been proposed to assist clinicians in making decisions based on genomic data.[69] Results from analyses need to be presented in an appropriate format that is easily understood and applied. This is true whether the target audience is a genomics researcher, a genetics specialist, a (nongeneticist) clinician, or a patient. At least 3 categories of report types can be envisioned—the traditional laboratory report, embedded clinical decision support, and advanced types of reporting (external to the EHR)—that leverage visualization approaches to illustrate complex analyses and findings.

The traditional laboratory analysis approach to reporting data, text reports inclusive of interpretation, are already widely applied to genomics data. Similar to pathology reports, these are created largely as unstructured text. The logical evolution is for this style of report to evolve into partially structured, synoptic reporting that represents text information as discrete data elements.[70,71] With synoptic reporting, the free text is associated with meta-data and placed in a structured format, with the advantage that converting these into clinical document architecture documents should be relatively straightforward.

For the clinical decision support use case, the presentation of complex aggregate analyses requires that several issues be addressed.[72] The analyses must be delivered at the point of care via the EHR system as part of normal clinician workflow, not through a separate system.[73] Ideally, this would be accomplished with the use of rules-based decision support system standards currently being deployed such as the Clinical Quality Framework.[74] Furthermore, authoritative resources will be needed for reference purposes to identify whether a

variant is pathological and the significance and meaning of the variant (Appendix, Table A2).

Multifactorial, multidimensional representations of complex relationships between clinical parameters and genomics data are likely the domain of advanced visualization approaches beyond the native capabilities of conventional EHR systems. Interaction with graphically presented data provides users the opportunity to instantaneously alter perspectives and model findings but generally requires proprietary engines to accomplish this degree of interactivity.[75] The recent availability of EHR-generated rich phenotype data allows new approaches to visualization of potential disease-gene associations through phenome-wide association study.[76]

Future Directions

The document-centric storage paradigm of current-generation EHR systems does not scale to allow the storage and retrieval of raw "omic" data.[77] For the foreseeable future, it is anticipated that dedicated omic ancillary systems will be required, analogous to how picture archiving and communication systems store the raw image data for radiology and cardiology with only the interpretive report being posted to the EHR. In the case of omic data, whereas the germline genetic sequence of an individual remains largely static over time, the interpretation of the sequence is a reflection of our level of understanding. Given that our knowledge is rapidly expanding, the ability to reanalyze and adjust the representation of the source sequence data is an added requirement of omic ancillary systems. For single-nucleotide variants, big data from EHRs can be useful for looking at prevalence and distribution and, with follow-up data, for looking at the ability of single-nucleotide variants to predict events. For polygenic disorders such as coronary disease, hypertension, and type 2 diabetes mellitus, big data, perhaps with risk calculators based on genome-wide association study data and mathematical risk models connected to the EHRs, will be useful for establishing risk of disease and outcomes.[67] In both cases, EHR genetic data can help define groups for randomized clinical trials of multiple types. Additionally, it would be exciting in the future to provide patients with information as it pertains to their genomics and to enable shared decision making through patient portals in the EHR.

Summary and Conclusions

The rapid pace of advancement in the field of genomics and the growth in adaptation of EHRs and the data-handling potential of information technology offer great promise for combining these resources to increase our understanding of cardiovascular genomics and, in turn, to transform cardiovascular care. In this advisory, we address the standards as they currently exist and we have laid out a framework to guide clinicians, researchers, and patients on the potential contributions to cardiovascular health that can be made by combining genomic data with information from the EHR.

Disclosures

Writing Group Disclosures

Writing Group Member	Employment	Research Grant	Other Research Support	Speakers' Bureau/ Honoraria	Expert Witness	Ownership Interest	Consultant/ Advisory Board	Other
Jennifer L. Hall	University of Minnesota	NIH*	None	None	None	None	*JACC* (associate editor)†; *JCTR* (editor)†	None
John J. Ryan	University of Utah	None	None	None	None	None	None	None
Bruce E. Bray	University of Utah	None	None	None	None	None	None	None
Candice Brown	Wake Forest School of Medicine	NIH/NINDS*	None	None	None	None	None	None
David Lanfear	Henry Ford Hospital	None	None	None	None	None	None	None
L. Kristin Newby	Duke Clinical Research Institute	David H Murdock Institute for Business and Culture (MURDOCK Study)†; NIA*; PCORI*; Google Life Sciences†; Sanofi*	None	None	None	None	None	None
Mary V. Relling	St. Jude Children's Research Hospital, Pharmaceutical Sciences	None	None	None	None	None	None	None
Neil J. Risch	University of California, San Francisco, Institute for Human Genetics	None	None	None	None	None	None	None
Dan M. Roden	Vanderbilt University School of Medicine	None	None	None	None	None	None	None
Stanley Y. Shaw	Massachusetts General Hospital Center for Systems Biology	Merck†	None	None	None	None	None	None
James E. Tcheng	Duke University Medical Center, Duke Clinical Research Institute	None	None	None	None	None	None	None
Jessica Tenenbaum	Duke University	None	None	None	None	None	None	None
Thomas N. Wang	Vanderbilt University School of Medicine	None	None	None	None	None	None	None
William S. Weintraub	Christiana Care Health System	None	None	None	None	None	None	None

This table represents the relationships of writing group members that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all members of the writing group are required to complete and submit. A relationship is considered to be "significant" if (a) the person receives $10 000 or more during any 12-month period, or 5% or more of the person's gross income; or (b) the person owns 5% or more of the voting stock or share of the entity, or owns $10 000 or more of the fair market value of the entity. A relationship is considered to be "modest" if it is less than "significant" under the preceding definition.

*Modest.

†Significant.

Reviewer Disclosures

Reviewer	Employment	Research Grant	Other Research Support	Speakers' Bureau/ Honoraria	Expert Witness	Ownership Interest	Consultant/ Advisory Board	Other
Wendy K. Chung	Columbia University	None	None	None	None	None	BioReference Laboratories*	None
Jonathan R. Egan	The Children's Hospital at Westmead (Australia)	None	None	None	None	None	None	None
Christina Pettey	University of Arkansas for Medical Sciences	None	None	None	None	None	None	None

This table represents the relationships of reviewers that may be perceived as actual or reasonably perceived conflicts of interest as reported on the Disclosure Questionnaire, which all reviewers are required to complete and submit. A relationship is considered to be "significant" if (a) the person receives $10 000 or more during any 12-month period, or 5% or more of the person's gross income; or (b) the person owns 5% or more of the voting stock or share of the entity, or owns $10 000 or more of the fair market value of the entity. A relationship is considered to be "modest" if it is less than "significant" under the preceding definition.

*Significant.

References

1. Blumenthal D. Launching HITECH. *N Engl J Med*. 2010;362:382–385. doi: 10.1056/NEJMp0912825.
2. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372:793–795. doi: 10.1056/NEJMp1500523.
3. Ryan JJ, Butrous G, Maron BA. The heterogeneity of clinical practice patterns among an international cohort of pulmonary arterial hypertension experts. *Pulm Circ*. 2014;4:441–451. doi: 10.1086/677357.
4. Krishnamoorthy P, Gupta D, Chatterjee S, Huston J, Ryan JJ. A review of the role of electronic health record in genomic research. *J Cardiovasc Transl Res*. 2014;7:692–700. doi: 10.1007/s12265-014-9586-0.
5. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, Delaney JT, Cowan J, Weeke P, Mosley JD, Wells QS, Karnes JH, Shaffer C, Peterson JF, Denny JC, Roden DM, Pulley JM. Biobanks and electronic medical records: enabling cost-effective research. *Sci Transl Med*. 2014;6:234cm3. doi: 10.1126/scitranslmed.3008604.
6. TG and HDL Working Group of the Exome Sequencing Project, National Heart, Lung, and Blood Institute, Crosby J, Peloso GM, Auer PL, Crosslin DR, Stitziel NO, Lange LA, Lu Y, Tang ZZ, Zhang H, Hindy G, Masca N, Stirrups K, Kanoni S, Do R, Jun G, Hu Y, Kang HM, Xue C, Goel A, Farrall M, Duga S, Merlini PA, Asselta R, Girelli D, Olivieri O, Martinelli N, Yin W, Reilly D, Speliotes E, Fox CS, Hveem K, Holmen OL, Nikpay M, Farlow DN, Assimes TL, Franceschini N, Robinson J, North KE, Martin LW, DePristo M, Gupta N, Escher SA, Jansson JH, Van Zuydam N, Palmer CN, Wareham N, Koch W, Meitinger T, Peters A, Lieb W, Erbel R, Konig IR, Kruppa J, Degenhardt F, Gottesman O, Bottinger EP, O'Donnell CJ, Psaty BM, Ballantyne CM, Abecasis G, Ordovas JM, Melander O, Watkins H, Orho-Melander M, Ardissino D, Loos RJ, McPherson R, Willer CJ, Erdmann J, Hall AS, Samani NJ, Deloukas P, Schunkert H, Wilson JG, Kooperberg C, Rich SS, Tracy RP, Lin DY, Altshuler D, Gabriel S, Nickerson DA, Jarvik GP, Cupples LA, Reiner AP, Boerwinkle E, Kathiresan S. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22–31. doi: 10.1056/NEJMoa1307095.
7. Jørgensen AB, Frikke-Schmidt R, Nordestgaard BG, Tybjærg-Hansen A. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med*. 2014;371:32–41. doi: 10.1056/NEJMoa1308027.
8. Flannick J, Thorleifsson G, Beer NL, Jacobs SB, Grarup N, Burtt NP, Mahajan A, Fuchsberger C, Atzmon G, Benediktsson R, Blangero J, Bowden DW, Brandslund I, Brosnan J, Burslem F, Chambers J, Cho YS, Christensen C, Douglas DA, Duggirala R, Dymek Z, Farjoun Y, Fennell T, Fontanillas P, Forsén T, Gabriel S, Glaser B, Gudbjartsson DF, Hanis C, Hansen T, Hreidarsson AB, Hveem K, Ingelsson E, Isomaa B, Johansson S, Jørgensen T, Jørgensen ME, Kathiresan S, Kong A, Kooner J, Kravic J, Laakso M, Lee JY, Lind L, Lindgren CM, Linneberg A, Masson G, Meitinger T, Mohlke KL, Molven A, Morris AP, Potluri S, Rauramaa R, Ribel-Madsen R, Richard AM, Rolph T, Salomaa V, Segrè AV, Skärstrand H, Steinthorsdottir V, Stringham HM, Sulem P, Tai ES, Teo YY, Teslovich T, Thorsteinsdottir U, Trimmer JK, Tuomi T, Tuomilehto J, Vaziri-Sani F, Voight BF, Wilson JG, Boehnke M, McCarthy MI, Njølstad PR, Pedersen O; Go-T2D Consortium; T2D-GENES Consortium, Groop L, Cox DR, Stefansson K, Altshuler D. Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet*. 2014;46:357–363. doi: 10.1038/ng.2915.
9. Holm H, Gudbjartsson DF, Sulem P, Masson G, Helgadottir HT, Zanon C, Magnusson OT, Helgason A, Saemundsdottir J, Gylfason A, Stefansdottir H, Gretarsdottir S, Matthiasson SE, Thorgeirsson GM, Jonasdottir A, Sigurdsson A, Stefansson H, Werge T, Rafnar T, Kiemeney LA, Parvez B, Muhammad R, Roden DM, Darbar D, Thorleifsson G, Walters GB, Kong A, Thorsteinsdottir U, Arnar DO, Stefansson K. A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat Genet*. 2011;43:316–320. doi: 10.1038/ng.781.
10. Wood GC, Still CD, Chu X, Susek M, Erdman R, Hartman C, Yeager S, Blosky MA, Krum W, Carey DJ, Skelding KA, Benotti P, Stewart WF, Gerhard GS. Association of chromosome 9p21 SNPs with cardiovascular phenotypes in morbid obesity using electronic health record data. *Genomic Med*. 2008;2:33–43. doi: 10.1007/s11568-008-9023-z.
11. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, Sanderson SC, Kannry J, Zinberg R, Basford MA, Brilliant M, Carey DJ, Chisholm RL, Chute CG, Connolly JJ, Crosslin D, Denny JC, Gallego CJ, Haines JL, Hakonarson H, Harley J, Jarvik GP, Kohane I, Kullo IJ, Larson EB, McCarty C, Ritchie MD, Roden DM, Smith ME, Böttinger EP, Williams MS; eMERGE Network. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med*. 2013;15:761–771. doi: 10.1038/gim.2013.72.
12. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. 2008;83:610–615. doi: 10.1016/j.ajhg.2008.09.017.
13. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. *Am J Hum Genet*. 2009;85:643–654. doi: 10.1016/j.ajhg.2009.10.002.
14. Macchia A, Marchioli R, Marfisi R, Scarano M, Levantesi G, Tavazzi L, Tognoni G. A meta-analysis of trials of pulmonary hypertension: a clinical condition looking for drugs and research methodology. *Am Heart J*. 2007;153:1037–1047. doi: 10.1016/j.ahj.2007.02.037.
15. Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, Field JR, Pulley JM, Ramirez AH, Bowton E, Basford MA, Carrell DS, Peissig PL, Kho AN, Pacheco JA, Rasmussen LV, Crosslin DR, Crane PK, Pathak J, Bielinski SJ, Pendergrass SA, Xu H, Hindorff LA, Li R, Manolio TA, Chute CG, Chisholm RL, Larson EB, Jarvik GP, Brilliant MH, McCarty CA, Kullo IJ, Haines JL, Crawford DC, Masys DR, Roden DM. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31:1102–1110. doi: 10.1038/nbt.2749.
16. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, Crane PK, Pathak J, Chute CG, Bielinski SJ, Kullo IJ, Li R, Manolio TA, Chisholm RL, Denny JC. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med*. 2011;3:79re1. doi: 10.1126/scitranslmed.3001807.
17. Rasmussen LV. The electronic health record for translational research. *J Cardiovasc Transl Res*. 2014;7:607–614. doi: 10.1007/s12265-014-9579-z.
18. Schaefer C; RPGEH GO Project Collaboration. C-A3-04: the Kaiser Permanente Research Program on Genes, Environment and Health: a resource for genetic epidemiology in adult health and aging. *Clin Med Res*. 2011;9:177–178. doi: 10.3121/cmr.2011.1020.c-a3-04.
19. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, Masys DR. Development of a large-scale de-identified DNA biobank

to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84:362–369. doi: 10.1038/clpt.2008.89.
20. Pulley J, Clayton E, Bernard GR, Roden DM, Masys DR. Principles of human subjects protections applied in an opt-out, de-identified biobank. *Clin Transl Sci*. 2010;3:42–48. doi: 10.1111/j.1752-8062.2010.00175.x.
21. Christensen H, Nielsen JS, Sørensen KM, Melbye M, Brandslund I. New national biobank of the Danish Center for Strategic Research on Type 2 Diabetes (DD2). *Clin Epidemiol*. 2012;4:37–42. doi: 10.2147/CLEP.S33042.
22. Austin ED, Cogan JD, West JD, Hedges LK, Hamid R, Dawson EP, Wheeler LA, Parl FF, Loyd JE, Phillips JA 3rd. Alterations in oestrogen metabolism: implications for higher penetrance of familial pulmonary arterial hypertension in females. *Eur Respir J*. 2009;34:1093–1099. doi: 10.1183/09031936.00010409.
23. Chen Z, Chen J, Collins R, Guo Y, Peto R, Wu F, Li L; China Kadoorie Biobank (CKB) Collaborative Group. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol*. 2011;40:1652–1666. doi: 10.1093/ije/dyr120.
24. Peakman TC, Elliott P. The UK Biobank sample handling and storage validation studies. *Int J Epidemiol*. 2008;37(suppl 1):i2–i6. doi: 10.1093/ije/dyn019.
25. Puckelwartz MJ, Pesce LL, Nelakuditi V, Dellefave-Castillo L, Golbus JR, Day SM, Cappola TP, Dorn GW 2nd, Foster IT, McNally EM. Supercomputing for the parallelization of whole genome analysis. *Bioinformatics*. 2014;30:1508–1513. doi: 10.1093/bioinformatics/btu071.
26. Hayden EC. Privacy protections: the genome hacker. *Nature*. 2013;497:172–174. doi: 10.1038/497172a.
27. McCarty CA, Chisholm RL, Chute CG, Kullo IJ, Jarvik GP, Larson EB, Li R, Masys DR, Ritchie MD, Roden DM, Struewing JP, Wolf WA; eMERGE Team. The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med Genomics*. 2011;4:13. doi: 10.1186/1755-8794-4-13.
28. Lin YC, Yu CS, Lin YJ. Enabling large-scale biomedical analysis in the cloud. *Biomed Res Int*. 2013;2013:185679. doi: 10.1155/2013/185679.
29. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. *Genet Med*. 2013;15:802–809. doi: 10.1038/gim.2013.121.
30. Meyer UA. Pharmacogenetics: five decades of therapeutic lessons from genetic diversity. *Nat Rev Genet*. 2004;5:669–676. doi: 10.1038/nrg1428.
31. Veenstra DL, Roth JA, Garrison LP Jr, Ramsey SD, Burke W. A formal risk-benefit framework for genomic tests: facilitating the appropriate translation of genomics into clinical practice. *Genet Med*. 2010;12:686–693. doi: 10.1097/GIM.0b013e3181eff533.
32. Green ED, Guyer MS; National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470:204–213. doi: 10.1038/nature09764.
33. Manolio TA, Green ED. Genomics reaches the clinic: from basic discoveries to clinical impact. *Cell*. 2011;147:14–16. doi: 10.1016/j.cell.2011.09.012.
34. Voora D, Ginsburg GS. A hub for bench-to-bedside pharmacogenomic-based research. *Pharmacogenomics*. 2011;12:1095–1098. doi: 10.2217/pgs.11.62.
35. Wang L, McLeod HL, Weinshilboum RM. Genomics and drug response. *N Engl J Med*. 2011;364:1144–1153. doi: 10.1056/NEJMra1010600.
36. Galiè N, Olschewski H, Oudiz RJ, Torres F, Frost A, Ghofrani HA, Badesch DB, McGoon MD, McLaughlin VV, Roecker EB, Gerber MJ, Dufton C, Wiens BL, Rubin LJ; Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy Studies (ARIES) Group. Ambrisentan for the treatment of pulmonary arterial hypertension: results of the Ambrisentan in Pulmonary Arterial Hypertension, Randomized, Double-Blind, Placebo-Controlled, Multicenter, Efficacy (ARIES) Study 1 and 2. *Circulation*. 2008;117:3010–3019. doi: 10.1161/CIRCULATIONAHA.107.742510.
37. Muir AJ, Gong L, Johnson SG, Lee MT, Williams MS, Klein TE, Caudle KE, Nelson DR; Clinical Pharmacogenetics Implementation Consortium (CPIC). Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for IFNL3 (IL28B) genotype and PEG interferon-α-based regimens. *Clin Pharmacol Ther*. 2014;95:141–146. doi: 10.1038/clpt.2013.203.
38. Caudle KE, Thorn CF, Klein TE, Swen JJ, McLeod HL, Diasio RB, Schwab M. Clinical Pharmacogenetics Implementation Consortium guidelines for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing. *Clin Pharmacol Ther*. 2013;94:640–645. doi: 10.1038/clpt.2013.172.
39. Relling MV, Gardner EE, Sandborn WJ, Schmiegelow K, Pui CH, Yee SW, Stein CM, Carrillo M, Evans WE, Hicks JK, Schwab M, Klein TE; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium guidelines for thiopurine methyltransferase genotype and thiopurine dosing: 2013 update. *Clin Pharmacol Ther*. 2013;93:324–325. doi: 10.1038/clpt.2013.4.
40. Leckband SG, Kelsoe JR, Dunnenberger HM, George AL Jr, Tran E, Berger R, Müller DJ, Whirl-Carrillo M, Caudle KE, Pirmohamed M; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium guidelines for HLA-B genotype and carbamazepine dosing. *Clin Pharmacol Ther*. 2013;94:324–328. doi: 10.1038/clpt.2013.103.
41. Hicks JK, Swen JJ, Thorn CF, Sangkuhl K, Kharasch ED, Ellingrod VL, Skaar TC, Müller DJ, Gaedigk A, Stingl JC; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. *Clin Pharmacol Ther*. 2013;93:402–408. doi: 10.1038/clpt.2013.2.
42. Hershfield MS, Callaghan JT, Tassaneeyakul W, Mushiroda T, Thorn CF, Klein TE, Lee MT. Clinical Pharmacogenetics Implementation Consortium guidelines for human leukocyte antigen-B genotype and allopurinol dosing. *Clin Pharmacol Ther*. 2013;93:153–158. doi: 10.1038/clpt.2012.209.
43. Martin MA, Klein TE, Dong BJ, Pirmohamed M, Haas DW, Kroetz DL; Clinical Pharmacogenetics Implementation Consortium. Clinical pharmacogenetics implementation consortium guidelines for HLA-B genotype and abacavir dosing. *Clin Pharmacol Ther*. 2012;91:734–738. doi: 10.1038/clpt.2011.355.
44. Scott SA, Sangkuhl K, Stein CM, Hulot JS, Mega JL, Roden DM, Klein TE, Sabatine MS, Johnson JA, Shuldiner AR; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium guidelines for CYP2C19 genotype and clopidogrel therapy: 2013 update. *Clin Pharmacol Ther*. 2013;94:317–323. doi: 10.1038/clpt.2013.105.
45. Scott SA, Sangkuhl K, Gardner EE, Stein CM, Hulot JS, Johnson JA, Roden DM, Klein TE, Shuldiner AR; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin Pharmacol Ther*. 2011;90:328–332. doi: 10.1038/clpt.2011.132.
46. Johnson JA, Gong L, Whirl-Carrillo M, Gage BF, Scott SA, Stein CM, Anderson JL, Kimmel SE, Lee MT, Pirmohamed M, Wadelius M, Klein TE, Altman RB; Clinical Pharmacogenetics Implementation Consortium. Clinical Pharmacogenetics Implementation Consortium Guidelines for CYP2C9 and VKORC1 genotypes and warfarin dosing. *Clin Pharmacol Ther*. 2011;90:625–629. doi: 10.1038/clpt.2011.185.
47. Wilke RA, Ramsey LB, Johnson SG, Maxwell WD, McLeod HL, Voora D, Krauss RM, Roden DM, Feng Q, Cooper-Dehoff RM, Gong L, Klein TE, Wadelius M, Niemi M; Clinical Pharmacogenomics Implementation Consortium (CPIC). The Clinical Pharmacogenomics Implementation Consortium: CPIC guideline for SLCO1B1 and simvastatin-induced myopathy. *Clin Pharmacol Ther*. 2012;92:112–117. doi: 10.1038/clpt.2012.57.
48. Ramsey LB, Johnson SG, Caudle KE, Haidar CE, Voora D, Wilke RA, Maxwell WD, McLeod HL, Krauss RM, Roden DM, Feng Q, Cooper-DeHoff RM, Gong L, Klein TE, Wadelius M, Niemi M. The Clinical Pharmacogenetics Implementation Consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin Pharmacol Ther*. 2014;96:423–428. doi: 10.1038/clpt.2014.125.
49. Shuldiner AR, Relling MV, Peterson JF, Hicks JK, Freimuth RR, Sadee W, Pereira NL, Roden DM, Johnson JA, Klein TE; Pharmacogenomics Research Network Translational Pharmacogenetics Program Group, Shuldiner AR, Vesely M, Robinson SW, Ambulos N Jr, Stass SA, Kelemen MD, Brown LA, Pollin TI, Beitelshees AL, Zhao RY, Pakyz RE, Palmer K, Alestock T, O'Neill C, Maloney K, Branham A, Sewell D, Relling MV, Crews K, Hoffman J, Cross S, Haidar C, Baker D, Hicks JK, Bell G, Greeson F, Gaur A, Reiss U, Huettel A, Cheng C, Gajjar A, Pappo A, Howard S, Hudson M, Pui CH, Jeha S, Evans WE, Broeckel U, Altman RB, Gong L, Whirl-Carrillo M, Klein TE, Sadee W, Manickam K, Sweet KM, Embi PJ, Roden D, Peterson J, Denny J, Schildcrout J, Bowton E, Pulley J, Beller M, Mitchell J, Danciu I, Price L, Pereira NL, Weinshilboum R, Wang L, Johnson JA, Nelson D, Clare-Salzler M, Elsey A, Burkley B, Langaee T, Liu F, Nessl D, Dong HJ, Lesko L, Freimuth RR, Chute CG. The Pharmacogenomics Research Network Translational Pharmacogenetics Program: overcoming challenges of real-world implementation. *Clin Pharmacol Ther*. 2013;94:207–210. doi: 10.1038/clpt.2013.59.
50. Pulley JM, Denny JC, Peterson JF, Bernard GR, Vnencak-Jones CL, Ramirez AH, Delaney JT, Bowton E, Brothers K, Johnson K, Crawford

DC, Schildcrout J, Masys DR, Dilks HH, Wilke RA, Clayton EW, Shultz E, Laposata M, McPherson J, Jirjis JN, Roden DM. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. *Clin Pharmacol Ther*. 2012;92:87–95. doi: 10.1038/clpt.2011.371.

51. Van Driest SL, Shi Y, Bowton EA, Schildcrout JS, Peterson JF, Pulley J, Denny JC, Roden DM. Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. *Clin Pharmacol Ther*. 2014;95:423–431. doi: 10.1038/clpt.2013.229.
52. Johnson JA, Elsey AR, Clare-Salzler MJ, Nessl D, Conlon M, Nelson DR. Institutional profile: University of Florida and Shands Hospital Personalized Medicine Program: clinical implementation of pharmacogenetics. *Pharmacogenomics*. 2013;14:723–726. doi: 10.2217/pgs.13.59.
53. Hicks JK, Crews KR, Hoffman JM, Kornegay NM, Wilkinson MR, Lorier R, Stoddard A, Yang W, Smith C, Fernandez CA, Cross SJ, Haidar C, Baker DK, Howard SC, Evans WE, Broeckel U, Relling MV. A clinician-driven automated system for integration of pharmacogenetic interpretations into an electronic medical record. *Clin Pharmacol Ther*. 2012;92:563–566. doi: 10.1038/clpt.2012.140.
54. Bell GC, Crews KR, Wilkinson MR, Haidar CE, Hicks JK, Baker DK, Kornegay NM, Yang W, Cross SJ, Howard SC, Freimuth RR, Evans WE, Broeckel U, Relling MV, Hoffman JM. Development and use of active clinical decision support for preemptive pharmacogenomics. *J Am Med Inform Assoc*. 2014;21:e93–e99. doi: 10.1136/amiajnl-2013-001993.
55. Schildcrout JS, Denny JC, Bowton E, Gregg W, Pulley JM, Basford MA, Cowan JD, Xu H, Ramirez AH, Crawford DC, Ritchie MD, Peterson JF, Masys DR, Wilke RA, Roden DM. Optimizing drug outcomes through pharmacogenetics: a case for preemptive genotyping. *Clin Pharmacol Ther*. 2012;92:235–242. doi: 10.1038/clpt.2012.66.
56. Bielinski SJ, Olson JE, Pathak J, Weinshilboum RM, Wang L, Lyke KJ, Ryu E, Targonski PV, Van Norstrand MD, Hathcock MA, Takahashi PY, McCormick JB, Johnson KJ, Maschke KJ, Rohrer Vitek CR, Ellingson MS, Wieben ED, Farrugia G, Morrisette JA, Kruckeberg KJ, Bruflat JK, Peterson LM, Blommel JH, Skierka JM, Ferber MJ, Black JL, Baudhuin LM, Klee EW, Ross JL, Veldhuizen TL, Schultz CG, Caraballo PJ, Freimuth RR, Chute CG, Kullo IJ. Preemptive genotyping for personalized medicine: design of the right drug, right dose, right time-using genomic data to individualize treatment protocol. *Mayo Clin Proc*. 2014;89:25–33. doi: 10.1016/j.mayocp.2013.10.021.
57. Hartzler A, McCarty CA, Rasmussen LV, Williams MS, Brilliant M, Bowton EA, Clayton EW, Faucett WA, Ferryman K, Field JR, Fullerton SM, Horowitz CR, Koenig BA, McCormick JB, Ralston JD, Sanderson SC, Smith ME, Trinidad SB. Stakeholder engagement: a key component of integrating genomic information into electronic health records. *Genet Med*. 2013;15:792–801. doi: 10.1038/gim.2013.127.
58. Edwards KL, Lemke AA, Trinidad SB, Lewis SM, Starks H, Snapinn KW, Griffin MQ, Wiesner GL, Burke W; GRRIP Consortium. Genetics researchers' and IRB professionals' attitudes toward genetic research review: a comparative analysis. *Genet Med*. 2012;14:236–242. doi: 10.1038/gim.2011.57.
59. Clayton EW, Smith M, Fullerton SM, Burke W, McCarty CA, Koenig BA, McGuire AL, Beskow LM, Dressler L, Lemke AA, Ramos EM, Rodriguez LL; Consent and Community Consultation Working Group of the eMERGE Consortium. Confronting real time ethical, legal, and social issues in the Electronic Medical Records and Genomics (eMERGE) Consortium. *Genet Med*. 2010;12:616–620. doi: 10.1097/GIM.0b013e3181efdbd0.
60. Dressler LG, Smolek S, Ponsaran R, Markey JM, Starks H, Gerson N, Lewis S, Press N, Juengst E, Wiesner GL; GRRIP Consortium. IRB perspectives on the return of individual results from genomic research. *Genet Med*. 2012;14:215–222. doi: 10.1038/gim.2011.10.
61. Lemke AA, Wu JT, Waudby C, Pulley J, Somkin CP, Trinidad SB. Community engagement in biobanking: experiences from the eMERGE Network. *Genomics Soc Policy*. 2010;6:35–52.
62. Jarvik GP, Amendola LM, Berg JS, Brothers K, Clayton EW, Chung W, Evans BJ, Evans JP, Fullerton SM, Gallego CJ, Garrison NA, Gray SW, Holm IA, Kullo IJ, Lehmann LS, McCarty C, Prows CA, Rehm HL, Sharp RR, Salama J, Sanderson S, Van Driest SL, Williams MS, Wolf SM, Wolf WA; eMERGE Act-ROR Committee and CERC Committee; CSER Act-ROR Working Group, Burke W. Return of genomic results to research participants: the floor, the ceiling, and the choices in between. *Am J Hum Genet*. 2014;94:818–826. doi: 10.1016/j.ajhg.2014.04.009.
63. McCarty CA, Chapman-Stone D, Derfus T, Giampietro PF, Fost N; Marshfield Clinic PMRP Community Advisory Group. Community consultation and communication for a population-based DNA biobank: the Marshfield Clinic Personalized Medicine Research Project. *Am J Med Genet A*. 2008;146A:3026–3033. doi: 10.1002/ajmg.a.32559.
64. Ollier W, Sprosen T, Peakman T. UK Biobank: from concept to reality. *Pharmacogenomics*. 2005;6:639–646. doi: 10.2217/14622416.6.6.639.
65. Hazin R, Brothers KB, Malin BA, Koenig BA, Sanderson SC, Rothstein MA, Williams MS, Clayton EW, Kullo IJ. Ethical, legal, and social implications of incorporating genomic information into electronic health records. *Genet Med*. 2013;15:810–816. doi: 10.1038/gim.2013.117.
66. The Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. 2015. http://iom.nationalacademies.org/reports/2015/sharing-clinical-trial-data. Accessed February 13, 2016.
67. Antman EM, Benjamin EJ, Harrington RA, Houser SR, Peterson ED, Bauman MA, Brown N, Bufalino V, Califf RM, Creager MA, Daugherty A, Demets DL, Dennis BP, Ebadollahi S, Jessup M, Lauer MS, Lo B, MacRae CA, McConnell MV, McCray AT, Mello MM, Mueller E, Newburger JW, Okun S, Packer M, Philippakis A, Ping P, Prasoon P, Roger VL, Singer S, Temple R, Turner MB, Vigilante K, Warner J, Wayte P; on behalf of the American Heart Association Data Sharing Summit Attendees. Acquisition, analysis, and sharing of data in 2015 and beyond: a survey of the landscape: a conference report from the American Heart Association Data Summit 2015. *J Am Heart Assoc*. 2015;4. doi: 10.1161/JAHA.115.002810.
68. Schneeweiss S. Learning from big health care data. *N Engl J Med*. 2014;370:2161–2163. doi: 10.1056/NEJMp1401111.
69. Welch BM, Eilbeck K, Del Fiol G, Meyer LJ, Kawamoto K. Technical desiderata for the integration of genomic data with clinical decision support. *J Biomed Inform*. 2014;51:3–7. doi: 10.1016/j.jbi.2014.05.014.
70. Hassell L, Aldinger W, Moody C, Winters S, Gerlach K, Schwenn M, Perriello D. Electronic capture and communication of synoptic cancer data elements from pathology reports: results of the Reporting Pathology Protocols 2 (RPP2) project. *J Registry Manag*. 2009;36:117–124; quiz 163.
71. Baskovich BW, Allan RW. Web-based synoptic reporting for cancer checklists. *J Pathol Inform*. 2011;2:16. doi: 10.4103/2153-3539.78039.
72. Ury AG. Storing and interpreting genomic information in widely deployed electronic health record systems. *Genet Med*. 2013;15:779–785. doi: 10.1038/gim.2013.111.
73. Manolio TA, Chisholm RL, Ozenberger B, Roden DM, Williams MS, Wilson R, Bick D, Bottinger EP, Brilliant MH, Eng C, Frazer KA, Korf B, Ledbetter DH, Lupski JR, Marsh C, Mrazek D, Murray MF, O'Donnell PH, Rader DJ, Relling MV, Shuldiner AR, Valle D, Weinshilboum R, Green ED, Ginsburg GS. Implementing genomic medicine in the clinic: the future is here. *Genet Med*. 2013;15:258–267. doi: 10.1038/gim.2012.157.
74. Ryerson CJ, Nayar S, Swiston JR, Sin DD. Pharmacotherapy in pulmonary arterial hypertension: a systematic review and meta-analysis. *Respir Res*. 2010;11:12. doi: 10.1186/1465-9921-11-12.
75. Keim D, Qu H, Ma KL. Big-data visualization. *IEEE Comput Graph Appl*. 2013;33:20–21.
76. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, Wang D, Masys DR, Roden DM, Crawford DC. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics*. 2010;26:1205–1210. doi: 10.1093/bioinformatics/btq126.
77. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: a time for omic ancillary systems. *JAMA*. 2013;309:1237–1238. doi: 10.1001/jama.2013.1579.

Circulation
Cardiovascular Genetics

Merging Electronic Health Record Data and Genomics for Cardiovascular Research: A Science Advisory From the American Heart Association

Jennifer L. Hall, John J. Ryan, Bruce E. Bray, Candice Brown, David Lanfear, L. Kristin Newby, Mary V. Relling, Neil J. Risch, Dan M. Roden, Stanley Y. Shaw, James E. Tcheng, Jessica Tenenbaum, Thomas N. Wang and William S. Weintraub
on behalf of the American Heart Association Professional and Public Education and Publications Committee of the Council on Functional Genomics and Translational Biology, Council on Clinical Cardiology, Council on Epidemiology and Prevention, Council on Quality of Care and Outcomes Research, and Stroke Council

Circ Cardiovasc Genet. 2016;9:193-202; originally published online March 14, 2016;
doi: 10.1161/HCG.0000000000000029
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

Appendix: Merging Electronic Health Record Data and Genomics for Cardiovascular Research

Appendix

Linkage of data between different systems remains a challenge in health care environments. Typically this is programed one interface at a time using specialized software. This can be improved by using a process called Integrating the Health Care Enterprise (IHE). IHE has been developed in several areas of medicine, in particular radiology and cardiology. IHE profiles permit interoperability, potentially at least in some areas at the level of "plug and play". New equipment or services should be labeled as IHE compliant for appropriate profiles.

As Electronic Health Records (EHR) datasets include larger numbers of patients and data types, they become more attractive for population-scale analyses. However, both EHR and administrative datasets primarily serve non-research purposes (medical documentation, billing, claims, insurance enrollment), which means that their adaptation for research and genomics comes with limitations. Administrative datasets such as Centers for Medicare and Medicaid Services include data on large numbers of patients spanning multiple hospital systems. But because of the inherent limitations of the available data, it remains challenging to reconstruct a patient's full clinical picture based on claims and other types of administrative data. While EHRs also experience missing data (eg, care at multiple institutions may be better captured by administrative rather than EHR data), important advantages include the availability of diagnostic test results (laboratory, radiographic, pathologic), growing numbers of institutional specimen biorepositories linked to EHR data,[1, 2] and the potentially rich content of narrative text, which can encapsulate clinician impressions. Approaches ranging from straightforward searches for pre-specified text strings to computer science approaches such as Natural

Language Processing (NLP) can convert these text sources into structured data for analysis. NLP is an area of active investigation; multiple approaches can scan narrative notes and map identified concepts to controlled vocabularies such as SNOMED-CT, identify relations between identified concepts (eg, connecting a symptom with an anatomic location), and distinguish positive *vs.* negative mentions (eg, "no history of angina").[3-5] Machine-learning algorithms that use both structured and NLP-extracted narrative EHR data to identify classes of patients have been recently shown to work robustly across multiple institutions.[6]

References

1. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR and Masys DR. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clinical pharmacology and therapeutics*. 2008;84:362-9.
2. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nature reviews Genetics*. 2011;12:417-28.
3. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC and Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. *Journal of the American Medical Informatics Association : JAMIA*. 2010;17:507-13.
4. Lemke AA, Wu JT, Waudby C, Pulley J, Somkin CP and Trinidad SB. Community engagement in biobanking: Experiences from the eMERGE Network. *Genomics, society, and policy / ESRC Genomics Network*. 2010;6:35-52.
5. Dligach D, Bethard S, Becker L, Miller T and Savova GK. Discovering body site and severity modifiers in clinical texts. *Journal of the American Medical Informatics Association : JAMIA*. 2014;21:448-54.
6. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, Pacheco JA, Boomershine CS, Lasko TA, Xu H, Karlson EW, Perez RG, Gainer VS, Murphy SN, Ruderman EM, Pope RM, Plenge RM, Kho AN, Liao KP and Denny JC. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. *Journal of the American Medical Informatics Association : JAMIA*. 2012;19:e162-9.
7. Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, Hoffman MA, Kohane IS and Levy HP. Technical desiderata for the integration of

genomic data into Electronic Health Records. *Journal of biomedical informatics*. 2012;45:419-22.
8. Welch BM, Eilbeck K, Fiol GD, Meyer LJ and Kawamoto K. Technical desiderata for the integration of genomic data with clinical decision support. *Journal of biomedical informatics*. 2014.

Appendix: Merging Electronic Health Record Data and Genomics for Cardiovascular Research

Appendix Table A1. Requirements of the analytics environments for medical big data

Maintain separation of primary molecular observations from the clinical interpretations of those data.
Support lossless data compression from primary molecular observations to clinically manageable subsets.
Maintain linkage of molecular observations to the laboratory methods used to generate them.
Support compact representation of clinically actionable subsets for optimal performance.
Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules.
Anticipate fundamental changes in the understanding of human molecular variation.
Support both individual clinical care and discovery science.

Reproduced with permission from Masys et al.[7]

Appendix: Merging Electronic Health Record Data and Genomics for Cardiovascular Research

Appendix Table A2. Desiderata for the integration of whole genome studies into clinical decision support (CDS)

CDS knowledge must have the potential to incorporate multiple genes and clinical information.
Keep CDS knowledge separate from variant classification.
CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification.
Support a large number of gene variants while simplifying the CDS knowledge to the extent possible.
Leverage current and developing CDS and genomics standards.
Support a CDS knowledge base deployed at and developed by multiple independent organizations.
Access and transmit only the genomic information necessary for CDS

Reproduced with permission from Welch et al.[8]