



Published in final edited form as:

Genet Med. 2013 October ; 15(10): 772–778. doi:10.1038/gim.2013.131.

Practical challenges in integrating genomic data into the electronic health record

Abel N. Kho, MD, MS¹, Luke V. Rasmussen, BS¹, John J. Connolly, PhD², Peggy L. Peissig, MBA³, Justin Starren, MD, PhD¹, Hakon Hakonarson, MD, PhD^{2,4}, and M. Geoffrey Hayes, PhD¹

¹Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

²Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

³Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA

⁴Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

Abstract

Genetic testing has had limited impact on routine clinical care. Widespread adoption of electronic health records presents a promising means of disseminating genetic testing into diverse care settings. Practical challenges to integration of genomic data into electronic health records include size and complexity of genetic test results, inadequate use of standards for clinical and genetic data, and limitations in electronic health record capacity to store and analyze genetic data. Related challenges include uncertainty in the interpretation of regulatory requirements for return of results, and privacy concerns specific to genetic testing. Successful integration of genomic data may require significant redesign of existing electronic health record systems.

Keywords

clinical decision support systems; electronic health records; genetic testing; genomic medicine

Following the trend of most other industries, health care is transitioning to digital records. Over the past several years, the use of electronic health records (EHRs) has more than doubled nationwide, with little likelihood of a return to paper records.¹ With widespread use of EHRs comes the promise of “big data” and innovative applications to dramatically measure, change, and improve the way healthcare is delivered and measured.²

In particular, the advent of lower cost genetic sequencing technology has greatly increased the availability of genomic data.³ Initiatives such as the Electronic Medical Records and

Genomics (eMERGE) consortium have focused not only on when results potentially should be returned, but also on exploring how genomic data can be integrated into EHRs to influence patient care.^{4,5}

In this article, we describe the current state of data capture and storage in EHRs, and identify practical challenges to the integration of genetic and genomic data into EHRs, using examples from our eMERGE experience.

EHRs: STRUCTURE AND STANDARDS

Much of the early published experience in EHR implementations and discovery derived from so-called home-grown EHRs, tailored over the years by early adopters to fit the workflow and needs of particular institutions.^{6–8} By contrast, the rapid adoption of EHRs, encouraged by recent government incentives, necessarily involves a wide range of commercial products. Successful implementations still require the same lessons from those early sites: attention to workflow, involvement of a physician champion, and responsive support.^{9–11}

Most current EHRs share common data categories, often reflecting a common provider workflow or need to document billing, clinical observations, or clinician actions. The stage 1 Meaningful Use standard outlines both the most common data categories and the expectations around appropriate documentation to enable improvement in care processes and outcomes in later stages.³ These subsequent stages of Meaningful Use outline the widely accepted data standards that should be used to ensure the consistency of data concepts.^{12,13} Achieving the Meaningful Use standard may improve the quality of data for research applications and identification of specific phenotypes from EHRs.¹⁴

Data in EHRs can be structured or unstructured (free text). Structured data are often a requirement to enable current EHR systems to readily generate reports or trigger clinical decision support tools, such as computerized reminders, because they are in a format that may be processed unambiguously by the computer. The use of free-text clinical notes supplements structured data by using the richness of language to describe a patient's clinical condition. Up to 80% of the value in data may be locked up in free text, requiring so-called natural language processing (NLP) to derive structured data elements from clinician notes.^{2,15} Table 1 outlines the common data categories in EHRs, most common data format (structured versus unstructured versus mixed) and commonly applied data standards. Each of these concepts, structured versus unstructured data, use of widely accepted data standards, and attention to work-flow, have parallels in the storage and use of genetic data, and it is at this interface that much recent work in the eMERGE network focuses.

Certain categories of data already collected within the EHR have particular relevance to the integration of genetic data. Demographics, such as self-reported race and ethnicity, may be important factors in selecting an appropriate genetic testing platform, or may modify risk probabilities based on genetic test results. Family history captures the clinical or phenotypic presentation of disease within related family members and, as an indirect measure of related factors such as environmental exposures and disease penetrance, may provide a

complementary or even superior risk measure to genetic tests.^{16,17} However, family history may be infrequently or incompletely captured in EHRs.^{18,19}

Medication lists and orders document one of the more common clinical interventions. Increasing understanding of how genetic variation predicts response to medications has led some institutions to initiate genomic medicine pilots with pharmacogenetic applications.^{20,21}

Incorporating genetic data into an EHR and using them in clinical practice reflect processes most similar to the handling and use of laboratory data. For many laboratory tests, including DNA tests, a healthcare provider places an order and a sample of blood or biological matter is collected and sent to the laboratory for testing. All laboratories performing tests used to guide clinical treatment require compliance with federal Clinical Laboratory Improvements Amendments (CLIA) certification as administered by the Centers for Medicare and Medicaid Services. CLIA creates a quality standard to ensure the accuracy, reliability and timeliness of patient test results no matter where the test is performed.²² Genetic tests conducted in a non-CLIA-certified laboratory may be used for research purposes but should not be used to guide clinical care unless results are subsequently confirmed through a CLIA-certified laboratory. This has important implications for whether and how genetic test results are recorded within an EHR. It is worth noting that many research-driven laboratory tests, such as those pertaining to electrolytes, are routinely performed in CLIA-certified laboratories and can be included in EHRs without special handling. Many of the newer genetic tests are performed by research laboratories and are not routinely available from CLIA-certified sources. Genetic test results performed through a non-CLIA-certified laboratory should either be stored in a separate research database, or, if stored within the EHR, clearly flagged as for research or nonclinical decision purposes.

More traditional laboratory results are currently represented in a variety of data formats, which may or may not need further interpretation prior to use in clinical practice. In most cases, data from the laboratory instrument are captured in the laboratory information system, and then transferred to the EHR through an electronic interface. Results with more direct numeric or binary indicators—such as total, low-density, and high-density cholesterol values—may be returned immediately. Other results, such as microbiology tests and cultures, require further processing or interpretation prior to use in the clinical environment. To further describe and interpret the laboratory result, other reference information—such as the reference populations, instrumentation and value high and low ranges (similar to the reference information required to interpret genetic results)—is associated with the result. The clinically relevant laboratory result may be coded using standardized terminologies such as Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) so it can be easily retrieved and made available to clinicians via a variety of interfaces including graphical and tabular displays, accessed through decision support systems (e.g. medication prescribing alerts), or integrated with application-specific information for viewing. Other tests may require human interpretation, for example, a pathologist's microscopic review of a blood smear, or a genetic laboratory interpretation of karyotype. These results may be presented as a text report for review by the ordering clinician; however, these text reports may not be readily used to generate trend reports or to trigger decision support rules.

Understanding whether genetic test results can be integrated as structured data within the EHR requires a parallel understanding of the scope and nature of genomic data.

GENETIC DATA: ORIGINS, STRUCTURE, AND STANDARDS

Genetic data range in both the purpose for which they were collected and in the size and scope of the data themselves. The scope of genetic testing data can range from testing for a single base pair change in a single individual to the entire genomic sequence of multiple if not hundreds or even thousands of individuals. Genetic data can be collected as part of a research study, ordered directly by clinicians for interpreting the underlying cause of the patient's phenotypic presentation or to determine a patient's disease risk or response to medications, and even conducted as a matter of personal interest through direct-to-consumer genetic testing options such as 23andMe. Most DTC options are specifically noted as not CLIA certified and therefore should not be used directly for clinical decision making (although they may indicate that a follow-up test is warranted).

Research-derived genetic data are typically generated in multiple individuals at a time depending on the purpose of the study and its design. The two primary population-based approaches to genetic studies are linkage and association. Another potential source for genetic data from research studies are smaller diagnostic tests that focus on a limited number of individuals, often related to one another. Linkage studies genotype genetic markers intermittently spaced throughout the genome in sets of family members to assess the cosegregation of alleles at any of these polymorphic markers with the disease or phenotype of interest. By contrast, association studies genotype markers in sets of cases and controls for a given disease or phenotype and assess the association between allele counts and case/control status—typically through allelic χ^2 tests or logistic regression if adjusting for covariates is important. Alternatively, the phenotype may be a quantitative trait, and linear regression is used to assess the evidence for association between counts of a given allele and the quantitative trait. Association studies may focus on a small region of the genome (i.e., a candidate gene) in which the case/control cohorts are on the order of a few hundred of individuals. The number of markers examined depends on the size of the region examined and the patterns of linkage disequilibrium (correlation between the alleles of the polymorphisms examined), and typically number in the tens. Alternatively, association studies can be performed genome-wide (genomewide association studies), and the number of genetic markers examined can be in the hundreds of thousands; such studies typically involve cohorts of multiple thousands of individuals in order to accommodate the need to correct for multiple testing across such a large number of markers. Similarly, sequence data can be used for linkage and association studies and can span a few hundred nucleotides in a candidate gene region to all of the exons in the genome, the “exome,” which encodes all of the proteins found in the genome sequence, to the near complete (>99%) genome sequence. Genetic data (Table 2) generated for research purposes are often stored in a pedigree and map file format (see http://watson.hgen.pitt.edu/docs/mega2_html/mega2.html for examples). The pedigree specifies the relationships of the individuals in the study as well as the genotype of each individual, along with a map file that identifies the polymorphisms included in the genotype file as well as their location in the genome.

The reported result for a clinical genetic test typically presents the summary of the findings rather than the raw data themselves, although the raw data may be included as well depending on the specific test. As with research-generated data, this can range from a single marker to whole-genome sequencing. Clinical tests can determine the presence or absence of a single mutation (e.g., the G6V mutation changing HbA to HbS, resulting in sickle cell anemia), multiple mutations (e.g., C282Y and H63D in the *HFE* gene, leading to hereditary hemochromatosis), or panels of mutations (e.g., the tens to hundreds of known mutations found in the *CFTR* gene resulting in cystic fibrosis or at a minimum impacting the cystic fibrosis phenotype and its presentation) that may be clinically relevant. These tests could result from sequencing the entire gene rather than genotyping specific polymorphic sites, yet only the specific variants might be presented in structured data fields. Potentially, the raw sequence could be determined from elsewhere in the report. This is further compounded if the test examines multiple genes, exome sequencing, or whole-genome sequencing. The single-nucleotide polymorphism arrays commonly used for genome-wide association studies in research studies can be used on single individuals in a clinical test setting to examine single polymorphisms if the mutation of interest is on the array, but more importantly they can be used to detect insertions, deletions, and copy-number variants of various sizes and copy number. Finally, karyotypes are another form of commonly ordered genetic test, reporting the presence of large-scale genomic rearrangements affecting partial chromosomes (e.g., translocations, large deletions) to full chromosomes (e.g., unisomies and trisomies). Given the methods and detection limits of karyotyping, there is no way to directly convert the report to raw nucleotide-level data and it therefore can better be thought of as qualitative data in narrative form, similar to pathology reports. The American College of Medical Genetics and Genomics has previously published recommended standards and guidelines for reporting genetic tests in a clinical setting (http://www.acmg.net/Pages/ACMG_Activities/stds-2002/stdsmenu-n.htm).

USE AND LIMITATIONS IN THE CLINICAL SETTING

The diversity of potential genetic test results creates novel opportunities to influence clinical care, but application may be limited by current EHR designs. Currently, genetic tests from the prenatal setting through a patient's life offer insight into the risk or potential for acquiring a condition, or serve as an explanation for the presence of a disease (and inform risk for the heritability to offspring). In some settings, genetic test results may be accompanied by a consultation with a genetic counselor that focuses on patient education and the collection of additional information (such as a detailed family history) to aid in assessing risk. With the wider availability of genetic testing outstripping the availability of trained genetic counselors, and limited understanding by many clinicians of the appropriate interpretation of genetic results, the potential to leverage technology such as clinical decision support systems within EHRs may be more relevant than ever.²³

Results returned to physicians are most often in a textual format, regardless of whether they are returned directly to the physician by a diagnostic laboratory or are the result of an interpretation by a genetic counselor. The modality of the result may vary—from a scan of a faxed report to the integration of a textual narrative incorporated with other laboratory results in an EHR. A common limitation is not only a lack of structured data that could be

used for clinical decision support but also the focus on reporting of variants as opposed to all observations made.²⁴

As noted by Hoffman,²⁵ even if structured information were available, there remains uncertainty in how best to represent genetic tests and results within existing clinical vocabularies. For example, although Logical Observation Identifiers Names and Codes represents a widely accepted vocabulary to represent specific tests and reporting styles, proliferation of genetic tests presents a logistical challenge. Should a new Logical Observation Identifiers Names and Codes code be generated for each test of a specific genetic variant? And should each result be assigned a Systematized Nomenclature of Medicine code? Although unresolved, work to address these challenges is underway within the standards community (D. J. Vreeman, personal communication, 11 July 2013). Furthermore, the availability of these results is a double-edged sword—the amount of information may be overwhelming, and/or not immediately clinically actionable in its raw form. Clinicians may not feel adequately informed or may not have the time to adequately counsel patients on genetic test results.²⁶ It is a combination of structured data that can be synthesized by the EHR to actionable results via clinical decision support that provides the greatest benefit, including links to more interpretable reports and reference sources.²⁷ Current EHR design may require significant modification to incorporate genetic test results in a meaningful and actionable format.²

TECHNICAL CHALLENGES IN INTEGRATING GENETIC DATA WITHIN CURRENT EHRs

The current eMERGE consortium consists of 10 centers, including 3 children's hospitals, and is focused on expanding the original eMERGE consortium's library of EHR-based phenotypes, and developing the means to integrate genotypic results back into the EHR. While exploring approaches to integrating genomic results into the EHR, it became apparent that solutions are needed not only to integrate genomic data into current-generation EHRs but also to plan for how next-generation EHRs could be better equipped for this type of information.

As described by Starren et al.,⁴ the differences in genetic and genomic data do not fit within current EHRs, given the increased data requirements to store such data, and the need to reprocess the data as new knowledge is made available.² In addition, the existing types of data typically stored in an electronic health record do not directly translate to a genetic counterpart. Considering laboratory results, as previously described a cholesterol measurement contains not only the actual measurement, but also additional supporting information such as the reference range for that measurement. For genomic data, there is a challenge as to how best represent several hundred or thousand measurements (summarized in Table 3).

One approach is to place each measured single-nucleotide polymorphism into the EHR as a laboratory result. This is a straightforward approach, but it introduces significant storage requirements for multiple tests looking at a wide range of polymorphisms, and may cause the EHR to run more slowly. In addition, the significant amount of information introduced

into the laboratory results by this approach may make finding other laboratory results increasingly difficult for physicians. Furthermore, to act on this information to provide decision support, the EHR would need to wade through a large amount of information each time to make a final determination, and would also require the maintenance of more complex rules as new clinical recommendations are made.²⁸

Another approach is to look at the interpretation of the genomic information at a single point in time and store a higher-level determination. For example, instead of storing several polymorphisms, their values would be converted to a single observation that is then stored as a laboratory result. This approach accepts a degree of information loss, such that when new knowledge is made available regarding how these genetic variants are interpreted, the polymorphisms would need to be measured again to create the new interpretation.

Yet another approach follows the lead of picture archiving and communication systems, in which the original data are stored external to the EHR, but a link is established.² Warehousing genetic data external to the EHR enables great flexibility for accommodating large volumes of data, or storage with loss-less data compression methods; given current limitations in EHR technology, this may be a particularly promising approach.²⁹ However, this approach requires not only attention to technical details but also development and maintenance of a robust interface back to the EHR to support data interpretation and decision support. If not well integrated with the EHR and the workflow of the clinician, it may find little use by a busy clinician.

For second-generation sequencing, these issues are magnified further. Although sequencing offers superior resolution of rare genetic variants as compared with single-nucleotide polymorphism arrays used for genome-wide association studies, this comes at the (literal) expense of a heavy burden on data management. The current cost of sequencing a human genome is ~\$3,500; the time required is several days (for a reasonable coverage of more than $\times 50$); and the outlay required in terms of data storage and analysis is considerable. Whole-exome sequencing (WES), which can be accomplished for less than \$1,000 and can yield results in <24 hours, is currently more widespread. For WES to achieve $\times 70$ to $\times 100$ coverage, several terabytes of raw sequence data are typically produced. However, such data are rarely stored, and 700–800 GB of stored data may be generated per whole-exome sequencing run. A run can be reliably accomplished by pooling ~80 samples, which further reduces the total amount of data generated to ~10 GB per sample. Such strategies can also be employed for whole-genome sequencing runs, which yield on average 12 times more data (i.e., ~120 GB of data for pooled samples). Data-management solutions similar to those discussed above can similarly be adopted to facilitate second-generation sequencing integration but would require significantly increased data storage and processing capacity. Further discussion of the challenges of dealing with sequence data in EHRs is presented in the article by Chute et al. on big data in this same issue.³⁰

Regardless of the approach, consideration of how to integrate family history with genetic data in the EHR will be important given the role of family history in guiding the choice of genetic testing, and in interpreting the results.³¹ In most systems, family history is captured as a combination of free text, diagnoses codes, and semistructured definitions of family

relationships/roles (e.g., mother, father), and it may be time-consuming.³² A number of initiatives are underway to standardize and streamline the capture of family history in EHRs; these will complement proposed approaches for inclusion of genetic data.^{33–35}

Currently within eMERGE, sites are exploring alternatives and combinations to the three potential solutions described. One early consortium pilot involves several eMERGE sites and will develop decision support rules to identify patients with increased probability of exposure to certain medications for focused pharmacogenetic testing on a limited number of single-nucleotide polymorphisms associated with drug metabolism. Although many of the eMERGE sites use the same EHR system, they may be exploring different approaches, given technical, organizational, or regulatory restrictions at each local institution.

IMPLICATIONS OF GENETIC RESULTS STORED IN THE EHR

Once genetic data are stored or available through the EHR, additional considerations are raised related to management of result interpretation for both patients and providers. Novel genetic discoveries may change the clinical implications for previously stored test results. The ability to identify patients with specific genetic results within the EHR will be important in order to recontact potentially affected patients, similar to identifying patients prescribed a recalled drug. Clear documentation of pre- and posttest counseling stored in the EHR should accompany genetic test results as recommended by the American College of Medical Genetics and Genomics statement that it is the responsibility of clinician/team to “provide comprehensive pre- and posttest counseling to the patient.” Obviously, such resources must be in place before initiating a genomics–EHR project. Initiatives such as Pharmacogenomics Knowledgebase and the related Clinical Pharmacogenetics Implementation Consortium provide guidance to clinicians interpreting genetic test results, in this instance, related to use of particular drugs.³⁶

For patients, excellent educational resources are available through, for example, the National Human Genome Research Institute (<http://www.genome.gov/Education/>), the National Library of Medicine (<http://ghr.nlm.nih.gov/>), and the Genetic Alliance (<http://www.geneticalliance.org/understanding.genetics>) on genomics in general. However, significant gaps in coverage of tests and conditions remain however. As an example, within eMERGE, we noted few public resources that specifically addressed the projects specific to eMERGE. To address this deficit, the eMERGE network recently announced plans to build an educational resource (www.myresults.org) that will focus on pharmacogenomics, relevant results, and practical issues for patients. An informal survey of 15 eMERGE representatives found that only 13% (2 representatives) agreed that “there are already relevant resources for communicating with patients.” Institutions intent on returning results to patients will need to understand the educational needs and requirements of their patients and identify available resources, where they exist.

PRIVACY CONCERNS WITH GENETIC DATA

Incorporating genetic data within the clinical EHR may present additional privacy concerns. Clinical data releases are currently regulated under the Health Insurance Privacy and Accountability Act. Genetic data constitute protected health information under the Health

Insurance Privacy and Accountability Act, although they are not explicitly mentioned as one of the 18 current identifiers specified in the act. Recent work reidentifying patients based on genetic data linked with publicly available genetic genealogy databases underscores the risk of re-identification presented by this loophole.³⁷ State laws may further restrict genetic data release above and beyond the federal Health Insurance Privacy and Accountability Act regulations; for example, in Illinois, the Genetic Information Privacy Act (410 ILCS 513) requires specific patient consent before any genetic data may be disclosed. Genetic data then, if stored in the EHR, may require segregation from other clinical data to prevent inadvertent or illegal disclosures. Although some research has explored methods of providing access controls around genetic results,³⁸ a technical framework to store and manage patient consent at this level of granularity is not widely available. The Genetic Information Nondiscrimination ACT defines legal protections preventing the misuse of genetic results but does not obviate the need for EHR design to accommodate current state and federal privacy regulations.

The association between family history and genetic testing results raises additional concerns. Genetic test results interpretation may be enhanced by testing of relatives identified through a robust and accurate family history, raising additional issues around consent to approach relatives and in whose EHR should genetic test results be stored. Because genetic test results may have significant implications for disease risk for family members, ethical concerns may be raised regarding duty to disclose to at-risk family members, which may conflict with duty to ensure privacy.

CONCLUSION

The promise of personalized medicine has not yet been realized, despite tremendous advances in both genomic technology, and EHR adoption. Significant work remains to reconcile the proliferation of genetic data with an improved understanding of how to present succinct and actionable distillates for the busy clinician. Gaps in the coverage and application of clinical and regulatory standards, as well as current EHR design, limit successful integration of genotypic data into clinical workflow. The potentially large volume of genetic data, coupled with changing understanding of the implications and interpretation of genetic results may further challenge integration with existing EHRs.

Early pilots, including some from the eMERGE consortium, point to the need for novel EHR tools and methods. Promising and near-term goals are focused on the integration of discrete genetic results, in structured formats, to deliver actionable recommendations through existing clinical decision support systems. Longer-term goals will require systems that can potentially store whole-genome scale results and support complex rules to guide interpretation of large data. Despite these challenges, the sharing of early lessons learned, and best practices between early adopters suggests that momentum is building toward more widespread adoption of genetic testing in clinical care.³⁹

Acknowledgments

The authors were funded in part by National Human Genome Research Institute eMERGE consortium grants U01HG006388-02 (A.N.K., L.V.R., J.S., and M.G.H.), U01HG006389-02 (P.L.P.), and 1U01HG006828-01 (J.J.C., H.H.).

References

1. Hsiao CJ, Hing E. Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2012. *NCHS Data Brief*. 2012; 111:1–8. [PubMed: 23384787]
2. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA*. 2013; 309:1351–1352. [PubMed: 23549579]
3. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *N Engl J Med*. 2010; 363:501–504. [PubMed: 20647183]
4. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: a time for omic ancillary systems. *JAMA*. 2013; 309:1237–1238. [PubMed: 23494000]
5. McCarty CA, Chisholm RL, Chute CG, et al. 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. [PubMed: 21269473]
6. Pryor TA, Gardner RM, Clayton PD, Warner HR. The HELP system. *J Med Syst*. 1983; 7:87–102. [PubMed: 6688267]
7. McDonald CJ, Overhage JM, Tierney WM, et al. The Regenstrief Medical Record System: a quarter century experience. *Int J Med Inform*. 1999; 54:225–253. [PubMed: 10405881]
8. Stead WW, Brame RG, Hammond WE, Jelovsek FR, Estes EH, Parker RT. A computerized obstetric medical record. *Obstet Gynecol*. 1977; 49:502–509. [PubMed: 854253]
9. Lorenzi NM, Kouroubali A, Detmer DE, Bloomrosen M. How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings. *BMC Med Inform Decis Mak*. 2009; 9:15. [PubMed: 19236705]
10. Ludwick DA, Doucette J. Adopting electronic medical records in primary care: lessons learned from health information systems implementation experience in seven countries. *Int J Med Inform*. 2009; 78:22–31. [PubMed: 18644745]
11. Goetz Goldberg D, Kuzel AJ, Feng LB, DeShazo JP, Love LE. EHRs in primary care practices: benefits, challenges, and successful strategies. *Am J Manag Care*. 2012; 18:e48–e54. [PubMed: 22435884]
12. Office of the National Coordinator for Health IT: Health IT Policy Committee. Meaningful Use Workgroup Request for Comments Regarding Meaningful Use Stage 2. 2012
13. Office of the National Coordinator for Health IT: Health IT Policy Committee. [Accessed 27 March 2013] Meaningful Use Workgroup Request for Comments Regarding Meaningful Use Stage 3. 2012. http://www.healthit.gov/sites/default/files/draft_stage3_rfc_07_nov_12.pdf.
14. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Trans Med*. 2011; 3:79re71.
15. Voorham J, Denig P. Computerized extraction of information on the quality of diabetes care from free text in electronic patient records of general practitioners. *J Am Med Inform Assoc*. 2007; 14:349–354. [PubMed: 17329733]
16. Paynter NP, Chasman DI, Paré G, et al. Association between a literature-based genetic risk score and cardiovascular events in women. *JAMA*. 2010; 303:631–637. [PubMed: 20159871]
17. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N Engl J Med*. 2008; 359:2208–2219. [PubMed: 19020323]
18. Acheson LS. Recording, interpreting, and updating the family history of cancer: implications for cancer prevention. *JAMA*. 2011; 306:208–210. [PubMed: 21750302]
19. Murff HJ, Greevy RA, Syngal S. The comprehensiveness of family cancer history assessments in primary care. *Community Genet*. 2007; 10:174–180. [PubMed: 17575462]

20. McClure JB, Swan GE, St John J, et al. Pharmacogenetic smoking cessation intervention in a health care setting: a pilot feasibility study. *Nicotine Tob Res.* 2013; 15:518–526. [PubMed: 22949583]
21. Price MJ, Berger PB, Teirstein PS, et al. GRAVITAS Investigators. Standard-vs high-dose clopidogrel based on platelet function testing after percutaneous coronary intervention: the GRAVITAS randomized trial. *JAMA.* 2011; 305:1097–1105. [PubMed: 21406646]
22. Cimino JJ. Improving the electronic health record—are clinicians getting what they wished for? *JAMA.* 2013; 309:991–992. [PubMed: 23483171]
23. Overby CL. Opportunities for genomic clinical decision support interventions. *Genet Med.* 2013; 15:817–823. [PubMed: 24051479]
24. Mitchell DR, Mitchell JA. Status of clinical gene sequencing data reporting and associated risks for information loss. *J Biomed Inform.* 2007; 40:47–54. [PubMed: 16617035]
25. Hoffman MA. The genome-enabled electronic medical record. *J Biomed Inform.* 2007; 40:44–46. [PubMed: 16616698]
26. Nippert I, Harris HJ, Julian-Reynier C, et al. Confidence of primary care physicians in their ability to carry out basic medical genetic tasks—a European survey in five countries—Part 1. *J Community Genet.* 2011; 2:1–11. [PubMed: 22109718]
27. Lubin IM, McGovern MM, Gibson Z, et al. Clinician perspectives about molecular genetic testing for heritable conditions and development of a clinician-friendly laboratory report. *J Mol Diagn.* 2009; 11:162–171. [PubMed: 19197001]
28. Deshmukh VG, Hoffman MA, Arnoldi C, Bray BE, Mitchell JA. Efficiency of CYP2C9 genetic test representation for automated pharmacogenetic decision support. *Methods Inf Med.* 2009; 48:282–290. [PubMed: 19387508]
29. Masys DR, Jarvik GP, Abernethy NF, et al. Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform.* 2012; 45:419–422. [PubMed: 22223081]
30. Chute CG, Ullman-Cullere M, Wood G, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. *Genet Med.* 2013; 15:802–809. [PubMed: 24008998]
31. Guttmacher AE, Collins FS, Carmona RH. The family history—more important than ever. *N Engl J Med.* 2004; 351:2333–2336. [PubMed: 15564550]
32. Acheson LS, Wiesner GL, Zyzanski SJ, Goodwin MA, Stange KC. Family history-taking in community family practice: implications for genetic screening. *Genet Med.* 2000; 2:180–185. [PubMed: 11256663]
33. Cohn WF, Ropka ME, Pelletier SL, et al. Health Heritage© a web-based tool for the collection and assessment of family health history: initial user experience and analytic validity. *Public Health Genomics.* 2010; 13:477–491. [PubMed: 20424421]
34. Feero WG, Bigley MB, Brinner KM. Family Health History Multi-Stakeholder Workgroup of the American Health Information Community. New standards and enhanced utility for family health history information in the electronic health record: an update from the American Health Information Community’s Family Health History Multi-Stakeholder Workgroup. *J Am Med Inform Assoc.* 2008; 15:723–728. [PubMed: 18755994]
35. Carmona RH, Wattendorf DJ. Personalizing prevention: the U.S. Surgeon General’s Family History Initiative. *Am Fam Physician.* 2005; 71:36–39. [PubMed: 15663022]
36. Whirl-Carrillo M, McDonagh EM, Hebert JM, et al. Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther.* 2012; 92:414–417. [PubMed: 22992668]
37. Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. *Science.* 2013; 339:321–324. [PubMed: 23329047]
38. Darcy DC, Lewis ET, Ormond KE, Clark DJ, Trafton JA. Practical considerations to guide development of access controls and decision support for genetic information in electronic medical records. *BMC Health Serv Res.* 2011; 11:294. [PubMed: 22047175]
39. Manolio TA, Chisholm RL, Ozenberger B, et al. Implementing genomic medicine in the clinic: the future is here. *Genet Med.* 2013; 15:258–267. [PubMed: 23306799]
40. Register F. Executive Office of the President Office of Management and Budget. Revisions to the standards for the classification of federal data on race and ethnicity. 621997:58781–58790.

41. World Health Organization. [Accessed 23 April 2013] International Classification of Disease (ICD). 2013. <http://www.who.int/classifications/icd/en/>.
42. International Health Terminology Standards Development Organisation. SNOMED CT. 2013 <http://www.ihtsdo.org/snomed-ct/>.
43. American Medical Association. CPT - Current Procedural Terminology. 2013 <http://www.ama-assn.org/ama/pub/physician-resources/solutionsmanaging-your-practice/coding-billing-insurance/cpt.page>.
44. U.S. National Library of Medicine. RxNorm Overview. 2013. <http://www.nlm.nih.gov/research/umls/rxnorm/overview.html#>.
45. Forrey AW, McDonald CJ, DeMoor G, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem*. 1996; 42:81–90. [PubMed: 8565239]
46. National Electrical Manufacturers Association. DICOM Homepage. 2013 <http://medical.nema.org/>.
47. Radiological Society of North America. <http://www.rsna.org/RadLex.aspx>.

Table 1

Common data categories and how they are represented in electronic health records (EHRs)

EHR data category	Structured/unstructured/mixed	Common data standard^a
Demographics (age, gender)	Structured	
Demographics (race/ethnicity)	Mixed	Office of management and budget categories (ref. 40)
Diagnoses	Structured	ICD9/10 (ref. 41) or SNOMED-CT (ref. 42)
Procedures	Structured	Common Procedural Terminology (CPT) (ref. 43)
Medications	Mixed	RxNorm (ref. 44)
Vital signs (e.g., blood pressure, heart rate)	Structured (numeric)	SNOMED-CT
Laboratory tests	Structured	LOINC (ref. 45)
Laboratory results/reports	Mixed/unstructured	SNOMED-CT
Family history	Mixed	Under development
Radiology images	Structured	DICOM (ref. 46)
Radiology reports	Semistructured (unstructured text organized into consistent sections)	SNOMED-CT or RadLex (ref. 47)

^aThese are not exclusive, and there can be overlap of available standards across data categories (e.g., LOINC codes can represent radiology report titles).

DICOM, Digital Imaging and Communications in Medicine; ICD-9/ICD-10, International Classification of Diseases, Ninth Revision/Tenth Revision; LOINC, Logical Observation Identifiers Names and Codes; SNOMED-CT, Systematized Nomenclature of Medicine–Clinical Terms

Table 2

Representations of genetic and genomic data

Variant type	Description	How coded
SNPs	Single-nucleotide polymorphisms	ACGT or 0/1 (presence/absence for each allele copy relative to the alleles found in the map file)
Indels	Insertions/deletions	0/1 (presence/absence) or copy number (e.g., 0/1/2, etc. of the repeat polymorphism)
CNVs	Copy-number variants	Number of copies of the copy-number variants and map position boundaries of the variants
STRs	Short tandem repeats	Number of copies of the repeat (integer)
WGS/WES	Whole-genome/whole-exome sequences	ACGT. Can be full sequence or limited to just the variant positions relative to the reference sequence

ACGT: nucleotide base sequence such as adenine, cytosine, guanine, thymine.

Table 3

Strategies for storing genetic and genomic test results in electronic health records (EHRs)

Storage strategy	Pros	Cons
Detailed results stored in EHR as a laboratory result (e.g., each SNP, insertion/deletion)	All information about the test result is available directly from the EHR Changes to how genetic variants are interpreted are immediately actionable through decision support	Significant storage needed for EHRs Introduces more results that must be sorted/filtered when reviewing laboratory results Complex decision support logic and reasoning is required within the EHR
Point-in-time interpretation	Reduced storage requirements in the EHR Computable result may be integrated into decision support	Loss of detailed information regarding genetic results Requires retesting and reinterpretation if knowledge about affected variants changes over time
External data warehouse linked back to EHR	Same benefits as the point-in-time interpretation Preserves original data, allowing review and reinterpretation of the results as needed Interpretation of genomic results may take advantage of more advanced computational approaches not available within EHRs Allows time for best practice evaluation and integration into the clinical workflow	Requires development and maintenance of a robust interface to integrate back into the EHR Investment in creation and maintenance of the ancillary storage system

SNP, single-nucleotide polymorphism.