

Storing and interpreting genomic information in widely deployed electronic health record systems

Andrew G. Ury, MD¹

Electronic health record systems are in widespread use but currently have very limited genomic capabilities. Electronic health record systems are a logical place for patient genomic information to be stored and used for decision support and improvement of patient care. Genomic data and their automated interpretation are very challenging for electronic health record software development because they are substantially different than other kinds of electronic health record

data and decision support. Those differences, the resulting challenges, and possible solutions are reviewed in this article.

Genet Med advance online publication 15 August 2013

Key Words: decision support; EHR; electronic health records; electronic medical records; genomics

In 2010, the United States reached a turning point, with more than half of office-based US physicians using an electronic health record (EHR) system. By 2012, the figure had risen to 72%, up from 29% in 2003.¹ The rapid adoption of EHRs was certainly due in part to the financial incentives offered by the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act, which reimbursed providers for the meaningful use of a certified EHR.

Individualizing the treatment of patients by taking into account individual risks and variations in treatment response has long been a goal of modern medicine.² The idea of using genomics to further this vision has become widespread,^{3,4} aided by the plummeting cost of DNA sequencing.⁵

For those physicians who are not medical geneticists, education about genomics has lagged far behind the rate of the advances in genomic knowledge and technology.⁶ Guttmacher et al.⁷ note that for biomedical researchers, this is the “genome era,” but for “most clinicians the genome era has not yet arrived.” Although using genetic counselors has been the traditional medical approach, it may not scale to meet patient and physician needs. Indeed, Belmont and McGuire⁸ have gone so far as to argue that “without the integration of a[n]...electronic health record, counseling patients on the basis of genome-wide data will be futile.”

Currently, the EHR systems that are in wide use do not offer much functionality to support genomics. A Rand Corporation study, “Are Electronic Health Records Ready for Genomic Medicine,” concluded that only 9% of those surveyed felt that EHRs currently had an impact on genetic/genomic medicine, whereas 36% thought there would be an impact over the next 5–10 years.⁹ Those surveyed included EHR specialists, primary-care clinicians, medical geneticists, and genetic counselors.

EHRs are the primary informatics and clinical decision support tools used by most physicians as well as the primary

repositories of patient data. It thus seems logical that EHRs should support the storage and interpretation of genomic data and the use of genomic data in decision support. Ideally, EHRs would help make medical treatment more precise and aid in the genomic education of physicians and patients.

Although the HITECH Act created a regulatory environment that accelerated EHR adoption, the current phases I and II of EHR certification and meaningful use do not require specific genomics functionality and thus do not provide an incentive for EHR vendors to innovate in the genomics area.^{10–12} That being said, EHRs do continue to innovate in response to market demands and technological advances. This review will cover many of the serious obstacles faced by widely deployed EHRs when implementing more complete genomic support.

GENOMIC DATA ARE CHALLENGING TO STORE IN WIDELY DEPLOYED EHRs

Normal characteristics and life cycle of EHR data

The normal characteristics and life cycle of EHR data are important to understand so that the challenges of incorporating genomic data into the EHR can be understood. Although EHR data structures vary by individual EHR,¹³ in the United States, due to certification requirements to qualify for provider reimbursement, the general method of storing core data is becoming more uniform.^{10,11} One way of thinking about EHR data is to divide it into three very general types of patient-specific data: granular data, textual data, and images.^{11,14–16}

Granular data consist of separate fields, each of which can in theory be separately accessed.¹⁷ For vital signs, one could store pulse, systolic blood pressure, diastolic blood pressure, temperature, respiratory rate, and pO₂ as separate fields. Along with date, time, and patient identity, they would form a database record. Similarly, numeric-value laboratory tests, such as serum

¹ActX, Seattle, Washington, USA. Correspondence: Andrew G. Ury (aury@actx.com)

Submitted 12 April 2013; accepted 24 June 2013; advance online publication 15 August 2013. doi:[10.1038/gim.2013.111](https://doi.org/10.1038/gim.2013.111)

sodium, blood glucose, and potassium levels, can be stored as discrete values in separate fields, along with the date, time, and patient identity.

Typically for each patient, in this example, there is a relatively small number of vital signs and laboratory tests, with the most recent being the most important. Decision-support rules can easily access the latest version of granular data or look back for a defined period. Visual-trend displays would look back further but still typically for a defined period. Normally, these data sets for a given patient are relatively small and, hence, can be retrieved rapidly and easily.

Another example of granular data are problem lists, orders, and medication lists.¹⁷ The list of the patient's problems is often stored as a text field and a code field (such as the International Classification of Diseases (ICD9 or ICD10), or SNOMED Clinical Terms (SNOMED CT)), along with other information. Medications similarly are stored as the medication name, dose, sig, date, prescriber, and other information. Granularity of medication information varies by EHR, but again due to certification standards and meaningful use, the trend is toward increased granularity to make interoperability, decision support, and reporting easier. For a given patient, the number of current problems and current medications is generally modest—with 20 being considered a large number. Decision support such as checking for drug interactions can easily operate on granular lists, which are easy to retrieve and write rules for.

Progress notes, nursing notes, radiology reports, pathology reports, and consult notes are all examples of textual data.¹⁷ Textual data are generally unstructured. It is easy for systems to retrieve and display but challenging to use for decision support or data mining.¹⁸ Textual data are also generally not very large; the text in a long progress note is ~5,000 characters long. The bulk of the information in the EHR for a patient consists of textual data. In some systems, even data such as numeric value laboratory reports are stored as text data.

The last type of data is images.¹⁹ For example, images can be photos of the patient or a lesion, electrocardiograms, X-rays, ultrasounds, magnetic resonance imaging, and computed tomography. Images tend to be large and, although they actually consist of individual pixels, are generally treated as large data objects to be occasionally retrieved and displayed. In many EHRs, larger types of images such as X-rays, magnetic resonance imaging, and computed tomography tend to be stored in a separate picture archiving and communication system.¹⁹ In that case, often only the radiology report, which is textual data, is stored in the EHR. Generally, only the title of the report, along with patient, date, and time, is easily available for decision support and reporting, and decision support and reporting do not use large data objects like images.

It is also important to understand the life cycle of EHR data.¹⁹ Older data, with a few exceptions, need to be accessed only rarely. For example, the vital signs, complete blood count, and chemistry panels done during an intensive care unit stay are rarely relevant again, with the exception of discharge values. In the outpatient arena, only the last note, a note on a

specific problem, or a few older progress notes are read again. An exception would be older pathology and unique radiology reports, which frequently are useful for long periods of time. Trend analysis for a flow chart does often require the retrieval of older information, but typically the set of information retrieved is small.

Typical EHR data storage systems

The type of data storage system used by an EHR is important because individual database systems have strengths and weaknesses. Most EHRs today store their data in Structured Query Language (SQL) relational databases or databases derived from the Multi-User Multi-Programming System (MUMPS), although some use a variety of other databases (EHR Association, personal communication). The SQL databases are often Oracle, SQL Server, or MySQL. Although MUMPS was originally developed in the 1960s,²⁰ derivatives of MUMPS such as MAGIC and Cache are still in wide use by some EHR vendors.

Although SQL- and MUMPS-derivative databases are well suited for many existing types of EHR data, they do not necessarily work well for all types of data. An area of relative weakness for these data storage systems is the rapid retrieval of very large granular data sets.

Genomic data characteristics and life cycle

One gene has ~3,000 base pairs, the exome has ~50 million base pairs,²¹ and the genome has ~3.2 billion base pairs.²² Apart from potentially being very large, an additional attribute of next-generation sequencing genomic data is that it has uneven depth of coverage, and thus, ideally it would be important to store information pertaining to the quality of the base-pair calls and which regions may have been missed entirely because coverage was too low.²³ Storing such attributes further complicates the storage of raw genomic data.

A way to reduce the complexity of genomic data is to process it to a set of variants or even further to a list of known pathologic variants and thus only store variants or selected variants. This substantially reduces the amount of data that must be stored.

If one stores all variants for an individual's complete genome, there are still an estimated 3–4 million variants per patient to store.²⁴ In addition, there are times when accessing the normal (reference) variant is still important, for example, for copy-number variants (e.g., three or more copies of a normal gene) and some heterozygous conditions. Therefore, even normal variants should be stored in some cases. Additional storage would also be required if one stores the variants found in abnormal tissue, such as cancers.

If one stores only clearly pathologic variants, the number stored will be significantly smaller. However, there is a substantial risk of false positives and false negatives with this approach as our understanding of the genome evolves. Although false positives can originate from an error in the original testing, a likely significant future source of false positives is the incorrect original classification of a variant as pathologic.²⁵ False negatives are also a big concern. Many known pathologic variants

remain unpublished and are stored only in the databases of individual laboratories.²⁶ Variants of unknown significance often outnumber pathologic variants. Moreover, we are only at the beginning of the research on genomics, and consequently the pathology of many variants remains undiscovered.

Standardization of terminology for the storage of variants is an issue.²⁷ Although rs numbers (reference SNP ID numbers) are becoming the standard way of identifying single-nucleotide polymorphism variants, many variants have not yet been assigned rs numbers. For variants that are haplotypes, there are often multiple ways of describing the variant.

The life cycle of genomic data is different than the life cycle of most other EHR data. The patient's genome in general remains the same throughout his or her lifetime. This is to be contrasted with vital signs, laboratory tests, and much of the other data gathered on the patient, which are less relevant as the data become older. On the other hand, because current sequencing techniques are usually partial and are not fully accurate, it is expected that until sequencing matures, a patient's DNA will be sequenced more than once in their lifetime.

The genomic data challenge

The current approach to storing genomic data in an EHR is as a textual laboratory report, usually a long, fairly complex document, with no granularity.²⁸ Because this conforms to the usual EHR textual data type, it is easy for EHRs to do. However, such a report does not make the data available for decision support, future reinterpretation of the data, use of the data for purposes other than the original diagnostic purpose, or for detailed population reporting. The physician must remember to access it to learn its contents.

To fully support genomic data, EHRs must store either raw genomic data and/or the variants derived from it. Both represent serious challenges for current EHR data schemas and storage systems. The problem is that the data are very large (like images) and very granular (like laboratory data). This combination is problematic if information needs to be rapidly retrieved and analyzed. Although raw genomic data can be heavily compressed,²⁹ and thus will take less space to store, it still must be uncompressed to be used for analysis, so its problematic large nature remains.

If the decision is made to only store variants, then there could still be millions of variants to store for each patient once whole-genome sequencing is widespread.²⁴ That would still be a very large and granular data set for an EHR to store and analyze on an individual patient. The rapid retrieval and analysis of such a data set can be challenging for the typical database systems used by most EHRs,³⁰ which could lead to a substantial performance issue.

If the decision is made to only store highly selected variants in order to improve performance, then as noted above, given the current early, and imprecise, state of our knowledge on which variants are truly pathologic, it will lead to missing many important variants as our knowledge evolves. Future clinical decision support will be hampered. To further complicate

matters, there currently is no consensus on which variants to select.³¹ Finally, there is the problem of knowing how meaningful a negative result is. As noted above, current sequencing methods do not cover all genes equally reliably, and there can be areas that were not covered.

GENOMIC DATA ARE CHALLENGING TO INTERPRET IN WIDELY DEPLOYED EHRs

Clinical result interpretation and display in typical EHRs

EHRs display and interpret clinical results in a variety of ways. If the result is received as a textual report, such as a radiology or pathology report, then it usually contains the interpretation and conclusion within it. Displaying the report is sufficient to provide an interpretation.

If the result is received as granular data, then the EHR will often need to provide an interpreted context.³² For example, for granular laboratory data, a chemistry panel, whether a value is high or low, is often shown through the use of color or a nearby marker. The range of normal values comes from the EHR or the laboratory. Similarly, for vital signs, EHRs often will provide an interpreted context. For example, the height, weight, and head circumference of young children can be plotted on growth charts with reference percentile curves.¹⁰

Clinical decision support in typical EHRs

Almost all modern EHRs offer some form of clinical decision support. US certification^{10,11} requires some clinical decision support elements, and some EHRs offer more than what is required. For example, when a drug is prescribed, an allergy and drug-interaction check is performed, and warnings are displayed if necessary.^{19,33} Another example is when, on the basis of the patient's age and sex, a mammogram is suggested for the patient.

Another form of decision support is rule-based decision support. The rules can be provided by the EHR, or in some cases, by the users of the EHR. The rules involve reading some portion of EHR data and then using predefined logic to display an alert or message to the EHR user.^{34–36} The EHR data to be accessed must be granular and well defined or validated. An example would be an alert reminding a user to order a renal or liver function test because the patient is on a medication that requires monitoring of renal or liver function.³³

Practicalities and challenges of genomic data interpretation

When genomic data are presented as a textual laboratory report, the report includes the interpretation, which is done outside the EHR. However, if an EHR stores the actual genomic data, whether in raw base-pair form or as lists of variants found, then the EHR will need to aid the clinician in interpreting the genomic data. It is unreasonable to expect clinicians (and especially nongeneticists) to understand the meaning of each genomic variant stored for the patient. For the EHR to interpret the patient's genomic data in an automated manner, the EHR faces several challenges.

First, there is a knowledge challenge, which is the lack of availability of an accessible, clinically reliable source for

knowing whether a variant is pathologic and what its meaning is.^{31,37} Although there are multiple public and paid websites that contain variant information and information on pathogenicity, they are for use by sophisticated users who can understand differences in terminology and often have to reference the original literature to make a final decision. The available amalgamated data is simply not reliable enough to be used as is.^{31,37,38} Individual molecular diagnostic laboratories use a combination of proprietary databases and human geneticists to overcome this interpretation problem. However, because their approach is only partly automated and usually focused on a narrow set of genes, it is not well suited for an EHR. It is likely that most, if not all, EHR vendors do not have the resources or expertise to develop or maintain their own genomic knowledge bases.

Second, there is a computing challenge. Much of decision support is delivered at the point of care, and clinicians are very sensitive to small delays in receiving information.^{39,40} If the amount of genomic data for a single patient is large, analysis is needed, and if time is limited, then unacceptable delays in interpretation could result.

Third is a usability challenge.⁴⁰ The genomic interpretation needs to be delivered in context and as part of the normal clinician workflow. If alerts are too frequent, users may start ignoring them (alert fatigue), so it is important to decide what data and alerts are truly worthwhile.⁴¹

Finally, there is the issue of whether existing EHR rules-based decision support systems can be adapted for widespread genomic use. There are examples in which for a few, carefully selected genomic conditions, with a limited number of pathologic variants, EHR rules-based decision support systems can be made to work.⁴² However, there are many genomic conditions with dozens or hundreds of pathologic variants and others with gene relationship complexity. Given that reality, it is unlikely that current EHR rules-based decision support systems can scale to an environment containing tens or hundreds of thousands of meaningful variants and rapidly evolving knowledge.

OTHER GENOMICS CHALLENGES FOR EHRs

Family history is an important part of the genomic information for a patient.⁴³ It can help to assess a patient's true genomic risks. Family history data can be granular and relatively easy to store in the data schema of a typical EHR. Many EHRs support the collection of family history.⁴⁴ The challenge for EHRs is to collect family history in a way that is accurate and also supports automated clinical decision support.⁴⁵⁻⁴⁹

Patient privacy is a major concern for genomic data.^{50,51} Security is very important but can be the same as for other patient data. However, the sharing and use of patient genomic data may need a different set of rules for genomic data than other patient data.⁵¹⁻⁵⁴ EHRs may need to provide a different mechanism for handling the sharing and use of genomic data.

Genomic data pose medical ethics dilemmas, which can be a challenge for automated interpretation in EHRs. Although this topic is reviewed elsewhere in this issue, it is worth noting that when a patient is asymptomatic, deciding which conditions and

level of evidence that will trigger an alert or risk profile is an issue that must be considered carefully.^{53,55,56}

The lack of physician education on genomics is another challenge for EHRs.⁶ When presenting genomic data and its interpretation, the overall knowledge of the receiving physician should be considered.

POSSIBLE SOLUTIONS

A number of groups have proposed possible solutions for the major EHR challenges of storing, interpreting, and providing decision support for genomic data (Table 1).

A 2012 NHLBI workshop on integrating genetic results into electronic medical records resulted in the publication of a list of technical desiderata for EHR genomic data.²³ For the storage of patient genetic information, they included the following desiderata: (i) maintain separation of primary molecular observations from the clinical interpretations of those data; (ii) support lossless data compression from primary molecular observations to clinically manageable subsets; (iii) maintain linkage of molecular observations to the laboratory methods used to generate them; (iv) support compact representation of clinically actionable subsets for optimal performance; and (v) simultaneously support human-viewable formats and machine-readable formats to facilitate implementation of decision support rules.

Berg et al.^{24,37} have proposed a methodology to deal with the large number of variants that can be found in a single patient

Table 1 Challenges for fully integrating genomic data into the EHR

Challenge	Possible solutions
Size and complexity of genomic data	2012 NHLBI technical desiderata (ref. 23) Berg et al. (refs. 24,37) Starren et al. (ref. 64)
Genomic data interpretation	Welch and Kawamoto review (ref. 58) Vanderbilt PREDICT (ref. 42) ACMG 2013 recommendations (ref. 56) Berg et al. (refs. 24,37) Overby (ref. 60)
Processing and storage demands	Proprietary efforts
Ethics and privacy	Darcy et al. (ref. 53) eMERGE (ref. 61)
Using family history	Doerr and Teng (ref. 43)
Research use	Kohane (ref. 65) Roden et al. (ref. 66) Pathak et al. (ref. 67) Ritchie et al. (ref. 68) eMERGE (ref. 61)

ACMG, American College of Medical Genetics and Genomics; EHR, electronic health record; eMERGE, Electronic Medical Records and Genomics Network; NHLBI, National Heart, Lung, and Blood Institute; PREDICT, Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment.

during whole-genome or -exome sequencing. They propose sorting them into bins based on clinical utility/actionability, clinical validity, and the potential to cause harm. Although their work is intended to aid molecular diagnosis laboratories struggling with the “incidentalome,” it is applicable to the interpretation issue EHRs will face handling genomic data.

The American College of Medical Genetics and Genomics recently published its recommendations for reporting incidental sequencing findings, specifically listing 24 disease groups.⁵⁶ The American Society of Clinical Oncology has also published its recommendations on genetic testing for cancer susceptibility.⁵⁷

Welch and Kawamoto,⁵⁸ in a systematic review of the literature, found 38 primary research articles focused on clinical decision support for genetically guided personalized medicine. The lack of automatic provision of clinical decision support in routine clinical workflow was strongly associated with a negative outcome. Of the 38 primary research articles, 9 were randomized trials, and 7 of the 9 randomized trials reported positive results. The key factor for a positive trial appeared to be incorporation into routine clinical workflow. Welch and Kawamoto note in their conclusion that research on these kinds of clinical decision support systems is still in its infancy.

The incorporation of pharmacogenomics into EHRs has been proposed by several groups.^{59,60} The Vanderbilt PREDICT group has published their operational design for a project that incorporated CYP2C19 variants into EHR decision support for prescribing clopidogrel.⁴²

Darcy et al.⁵³ have published a guide to practical considerations for access controls and decision support for genetic information. Best practices for merging genomic data repositories with EHRs are being developed by the eMERGE group,⁶¹ whose work is described in an article in this issue.⁶² Martin-Sanchez et al.⁶³ have called for a synthesis of medical informatics and bioinformatics to create an integration of genomic data and the EHR.

As an interim solution, Starren et al.⁶⁴ have proposed creating separate, ancillary genomic systems. Finally, family history remains important in the genomic era. Doerr and Teng⁴³ recently reviewed the validity and usefulness of family history tools.

CONCLUSION

Given the storage, interpretation, and processing challenges, along with the press of other priorities such as certification and meaningful use, it is not surprising that up to now EHRs have made very little progress in their use of genomic data. Because of the data's large but granular nature and interpretation complexity, genomic data and their interpretation are substantially different than other kinds of EHR data and decision support and thus do not easily lend themselves to extensions of existing EHR technologies.

EHRs do support the only kind of genomic data that fits into their current schemas, which is a textual laboratory

report from a molecular diagnosis laboratory. Those reports generally address a very specific diagnostic issue and contain a full interpretation. However, the reports are not well suited for clinical decision support, generally contain limited genomic information, and can be rendered obsolete by advances in medical knowledge.

Although existing EHR rules-based decision support systems can be used for a few, selected genomic decision support conditions, they are unlikely to scale to support a rapidly evolving environment containing tens of thousands of meaningful variants. Many EHRs support the gathering of family history, and family history is an important part of understanding the meaning of genomic findings. However, there is a need to technically improve the manner in which family history is gathered, to make it more useful for clinicians.

To truly integrate genomic decision support into the EHR involves solving several difficult challenges, and substantial research and development is still needed, along with the collaboration of the medical genetics community. In addition to the technical challenges of fully incorporating genomic data into EHRs, important patient privacy and bioethical issues need to be addressed. The research and development effort is very worthwhile because integrating genomic decision support into the EHR is a key to beginning a new era of precision medicine.

Incorporating genomic information into the EHR will have the additional benefit of enabling medical research, allowing much more to be learned about the relationship between genotype and phenotype.^{18,61,65–68}

ACKNOWLEDGMENTS

The author thanks James Mategko for his help in reviewing the literature, and Isabelle Lucas, Marie Lee, Steve Ornstein, and Susan Croft for their feedback.

DISCLOSURE

A.G.U. is the chairman of ActX, a privately held company. ActX is working in the area of genomics and informatics. Moreover, A.G.U. was an unpaid ex-officio member of the Executive Committee of the Electronic Health Association (EHRA) in 2010 and 2011. For the past 3 years, A.G.U. has had no financial interest in any EHR company. A.G.U.'s past involvement with EHRs is as follows: (i) 6 years ago, A.G.U. was CEO of Practice Partner, an EHR company, which was sold to McKesson; (ii) until 4 years ago, A.G.U. was chief medical officer at McKesson Provider Technologies; and (iii) A.G.U. has in the past twice been vice chair of the EHRA.

REFERENCES

- Centers for Disease Control and Prevention. Products, Data Briefs, Number 111, December 2012. <http://www.cdc.gov/nchs/data/databriefs/db111.htm>. Accessed 5 February 2013.
- Conti R, Veenstra DL, Armstrong K, Lesko LJ, Grosse SD. Personalized medicine and genomics: challenges and opportunities in assessing effectiveness, cost-effectiveness, and future research priorities. *Med Decis Making* 2010;30:328–340.

3. Burton H, Cole T, Lucassen AM. Genomic medicine: challenges and opportunities for physicians. *Clin Med* 2012;12:416–419.
4. Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007;9:665–674.
5. Wetterstrand KA. DNA sequencing costs: data from the NHGRI genome sequencing program (GSP). 2013. <http://www.genome.gov/sequencingcosts>. Accessed 20 March 2013.
6. Salari K. The dawning era of personalized medicine exposes a gap in medical education. *PLoS Med* 2009;6:e1000138.
7. Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet* 2007;8:151–157.
8. Belmont J, McGuire AL. The futility of genomic counseling: essential role of electronic health records. *Genome Med* 2009;1:48.
9. Scheuner MT, de Vries H, Kim B, Meili RC, Olmstead SH, Teleki S. Are electronic health records ready for genomic medicine? *Genet Med* 2009;11:510–517.
10. HealthIT.gov. Testing and certification. 2011. <http://www.healthit.gov/policy-researchers-implementers/2011-edition-approved-test-methods>. Accessed 20 March 2013.
11. HealthIT.gov. Testing and certification. 2014. <http://www.healthit.gov/policy-researchers-implementers/2014-edition-final-test-method> Accessed 20 March 2013.
12. Centers for Medicare & Medicaid Services. CMS meaningful use. 2013. http://www.cms.gov/Regulations-and-guidance/Legislation/EHRIncentivePrograms/Meaningful_Use.html Accessed 22 March 2013.
13. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform* 2008;77:291–304.
14. Mitre – NIH National Center for Research Resources: electronic health records overview. 2006. <http://www.mendeley.com/catalog/electronic-health-records-overview/>. Accessed 25 March 2013.
15. HL7/ASTM implementation guide for CDA release 2 –continuity of care document (CCD) release 1, 2005. <http://www.hl7.org/implement/standards>. Accessed 21 March 2013.
16. HL7 implementation guide for CDA release 2: IHE health story consolidation, release 1.1 - US Realm, 2012. <http://www.hl7.org/implement/standards/hhsifr.cfm>. Accessed 21 March 2013.
17. ASTM. Standard guide for content and structure of the electronic health record (EHR), designation: E 1384–99, 1999. <http://www.astm.org/Standards/E1384.htm>. Accessed 4 April 2013.
18. Denny JC. Chapter 13: mining electronic health records in the genomics era. *PLoS Comput Biol* 2012;8:e1002823.
19. Carter, JH. *Electronic Medical Records: A Guide for Clinicians and Administrators*. ACP-ASIM: Philadelphia, PA, 2001.
20. O’Kane KC. *The MUMPS Programming Language*, 2nd ed. CreateSpace: Seattle, 2010.
21. Coffey AJ, Kokocinski F, Calafato MS, et al. The GENCODE exome: sequencing the complete human exome. *Eur J Hum Genet* 2011;19:827–831.
22. Human genome project information - functional and comparative genomics fact sheet. 2011. http://www.ornl.gov/sci/techresources/Human_Genome/faq/compngen.shtml#genomesize. Accessed 4 April 2013.
23. 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.
24. Berg JS, Khoury MJ, Evans JP. Deploying whole genome sequencing in clinical practice and public health: meeting the challenge one bin at a time. *Genet Med* 2011;13:499–504.
25. Kohane IS, Hsing M, Kong SW. Taxonomizing, sizing, and overcoming the incidentalome. *Genet Med* 2012;14:399–404.
26. Campbell EG, Clarridge BR, Gokhale M, et al. Data withholding in academic genetics: evidence from a national survey. *JAMA* 2002;287:473–480.
27. Yu W, Ned R, Wulf A, Liu T, Khoury MJ, Gwinn M. The need for genetic variant naming standards in published abstracts of human genetic association studies. *BMC Res Notes* 2009;2:56.
28. Gulley ML, Brazier RM, Halling KC, et al.; Molecular Pathology Resource Committee, College of American Pathologists. Clinical laboratory reports in molecular pathology. *Arch Pathol Lab Med* 2007;131:852–863.
29. Giancarlo R, Scaturro D, Utro F. Textual data compression in computational biology: a synopsis. *Bioinformatics* 2009;25:1575–1586.
30. Pavlo A, Paulson E, Rasin A, et al. A comparison of approaches to large-scale data analysis. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*. ACM: New York, 2009:165–178.
31. Jarvik G, Turner E, Robertson P, et al. Actionable, known pathogenic incidental findings in 1000 subjects’ exomes. Paper presented at ACMG Annual Clinical Genetics Meeting (Abstract 12). Phoenix, 18–23 March 2013.
32. Powsner SM, Tuft E. Graphical summary of patient status. *Lancet* 1994;344:386–389.
33. Wessell AM, Ornstein SM, Jenkins RG, Nemeth LS, Litvin CB, Nietert PJ. Medication safety in primary care practice: results from a PPRNet quality improvement intervention. *Am J Med Qual* 2013;28:16–24.
34. Test procedure for §170.306 (c) clinical decision support. 2010. http://healthcare.nist.gov/docs/170.306.c_ClinicalDecisionSupportIL_v1.1.pdf. Accessed 20 March 2013.
35. Sittig DF, Wright A, Osheroff JA, et al. Grand challenges in clinical decision support. *J Biomed Inform* 2008;41:387–392.
36. Nemeth LS, Ornstein SM, Jenkins RG, Wessell AM, Nietert PJ. Implementing and evaluating electronic standing orders in primary care practice: a PPRNet study. *J Am Board Fam Med* 2012;25:594–604.
37. Berg JS, Adams M, Nassar N, et al. An informatics approach to analyzing the incidentalome. *Genet Med* 2013;15:36–44.
38. Moorthie S, Hall A, Wright CF. Informatics and clinical genome sequencing: opening the black box. *Genet Med* 2013;15:165–171.
39. Shneiderman B. Response time and display rate in human performance with computers. *Computing Surveys* 1984;16:265–285.
40. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10:523–530.
41. Hoffman MA, Williams MS. Electronic medical records and personalized medicine. *Hum Genet* 2011;130:33–39.
42. Pulley JM, Denny JC, Peterson JF, et al. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. *Clin Pharmacol Ther* 2012;92:87–95.
43. Doerr M, Teng K. Family history: still relevant in the genomics era. *Cleve Clin J Med* 2012;79:331–336.
44. CCHIT final test scripts ambulatory rev. 6.5.07, 2007. <https://www.cchit.org>. Accessed 20 March 2013.
45. Orlando LA, Hauser ER, Christianson C, et al. Protocol for implementation of family health history collection and decision support into primary care using a computerized family health history system. *BMC Health Serv Res* 2011;11:264.
46. Yoon PW, Scheuner MT, Jorgensen C, Khoury MJ. Developing family healthware, a family history screening tool to prevent common chronic diseases. *Prev Chronic Dis* 2009;6:A33.
47. Acheson LS, Wang C, Zyzanski SJ, et al.; Family Healthware Impact Trial (FHIT) Group. Family history and perceptions about risk and prevention for chronic diseases in primary care: a report from the family healthware impact trial. *Genet Med* 2010;12:212–218.
48. Melton GB, Raman N, Chen ES, Sarkar IN, Pakhomov S, Madoff RD. Evaluation of family history information within clinical documents and adequacy of HL7 clinical statement and clinical genomics family history models for its representation: a case report. *J Am Med Inform Assoc* 2010;17:337–340.
49. Melton GB, Raman N, Chen ES, Sarkar IN, Pakhomov S, Madoff RD. Evaluation of family history information within clinical documents and adequacy of HL7 clinical statement and clinical genomics family history models for its representation: a case report. *J Am Med Inform Assoc* 2010;17:337–340.
50. McGuire AL, Basford M, Dressler LG, et al. Ethical and practical challenges of sharing data from genome-wide association studies: the eMERGE Consortium experience. *Genome Res* 2011;21:1001–1007.
51. McGuire AL, Fisher R, Cusenza P, et al. Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genet Med* 2008;10:495–499.
52. Clayton EW, Smith M, Fullerton SM, et al.; 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.
53. 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.
54. McGuire AL, Fisher R, Cusenza P, et al. Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genet Med* 2008;10:495–499.

55. Khoury MJ, Coates RJ, Evans JP. Evidence-based classification of recommendations on use of genomic tests in clinical practice: dealing with insufficient evidence. *Genet Med* 2010;12:680–683.
56. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med* 2013;15:565–574.
57. Robson ME, Storm CD, Weitzel J, Wollins DS, Offit K; American Society of Clinical Oncology. American Society of Clinical Oncology policy statement update: genetic and genomic testing for cancer susceptibility. *J Clin Oncol* 2010;28:893–901.
58. Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc* 2013;20:388–400.
59. Roden DM, Xu H, Denny JC, Wilke RA. Electronic medical records as a tool in clinical pharmacology: opportunities and challenges. *Clin Pharmacol Ther* 2012;91:1083–1086.
60. Overby CL, Tarczy-Hornoch P, Hoath JI, Kalet IJ, Veenstra DL. Feasibility of incorporating genomic knowledge into electronic medical records for pharmacogenomic clinical decision support. *BMC Bioinformatics* 2010;11(suppl 9):S10.
61. 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.
62. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. *Genet Med* 2013; e-pub ahead of print.
63. Martin-Sanchez F, Iakovidis I, Nørager S, et al. Synergy between medical informatics and bioinformatics: facilitating genomic medicine for future health care. *J Biomed Inform* 2004;37:30–42.
64. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: a time for omic ancillary systems. *JAMA* 2013;309:1237–1238.
65. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 2011;12:417–428.
66. Roden DM, Xu H, Denny JC, Wilke RA. Electronic medical records as a tool in clinical pharmacology: opportunities and challenges. *Clin Pharmacol Ther* 2012;91:1083–1086.
67. Pathak J, Pan H, Wang J, et al. Evaluating phenotypic data elements for genetics and epidemiological research: experiences from the eMERGE and PhenX Network Projects. *AMIA Summits Transl Sci Proc* 2011;2011:41–45.
68. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet* 2010;86:560–572.
69. Casey J. Clinical decision support integrated with the EHR for genomic medicine. *Genet Med* 2013; e-pub ahead of print.