Genetic data and electronic health records: a discussion of ethical, logistical and technological considerations

Kimberly Shoenbill,[1,2] Norman Fost,[3,4] Umberto Tachinardi,[2] Eneida A Mendonca[1,2,4]

OPEN ACCESS

[1]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, Wisconsin, USA
[2]Institute for Clinical and Translational Research (ICTR), University of Wisconsin, Madison, Wisconsin, USA
[3]Department of Medical History and Bioethics, University of Wisconsin, Madison, Wisconsin, USA
[4]Department of Pediatrics, University of Wisconsin, Madison, Wisconsin, USA

Correspondence to
Dr Kimberly Shoenbill, Department of Biostatistics and Medical Informatics, University of Wisconsin, 600 Highland Avenue, H6-528, Madison, WI 53711, USA; shoenbill@wisc.edu

ABSTRACT

Objective The completion of sequencing the human genome in 2003 has spurred the production and collection of genetic data at ever increasing rates. Genetic data obtained for clinical purposes, as is true for all results of clinical tests, are expected to be included in patients' medical records. With this explosion of information, questions of what, when, where and how to incorporate genetic data into electronic health records (EHRs) have reached a critical point. In order to answer these questions fully, this paper addresses the ethical, logistical and technological issues involved in incorporating these data into EHRs.

Materials and methods This paper reviews journal articles, government documents and websites relevant to the ethics, genetics and informatics domains as they pertain to EHRs.

Results and discussion The authors explore concerns and tasks facing health information technology (HIT) developers at the intersection of ethics, genetics, and technology as applied to EHR development.

Conclusions By ensuring the efficient and effective incorporation of genetic data into EHRs, HIT developers will play a key role in facilitating the delivery of personalized medicine.

INTRODUCTION

Initial sequencing of the human genome in 2003 launched a new era of genetic research that promised personalized medicine with improved diagnosis, prevention and treatment of disease (box 1). Before such promises can be fulfilled, many questions must be answered surrounding the incorporation of genetic data into electronic health records (EHRs). As noted by Lose,[1] "even before the race to integrate genomic medicine into the daily routine begins, the initial question of what to disseminate remains." A common misperception is that the field of genetics is advancing faster than ethical standards to help guide its use. Guidelines on ethically responsible genetic screening were proposed in 1975 by the Committee for the Study of Inborn Errors of Metabolism of the National Academy of Sciences, reiterated in an Institute of Medicine report in 1994, and similar guidelines promulgated by the World Health Organization and other groups.[2–5] Despite wide support of these principles, in practice they are often not followed.[6] This paper examines many of the same ethical and logistical problems anticipated by the 1975 committee/report and evaluates them in relation to technological issues faced in the current era of EHRs, genomic medicine and big data. This examination is based on a review of journal articles, government documents and websites addressing ethical, genetic, technology and infrastructure concerns as they relate to EHRs (figure 1). The rapid pace of genetic research necessitates urgent focus on integration solutions that minimize risk while maximizing benefit. Box 1 provides a glossary of relevant genetic and genomic terms.

CLASSIFYING GENETIC TESTS

To identify tests that should be included in EHRs, an evaluation of validity, utility and intended purpose is required. As outlined by the Secretary's Advisory Committee on Genetic Testing, and summarized by Lose,[1] "Three concepts are central to understanding how knowledge graduates to evidence sufficient to alter clinical practice. The first is analytical validity which is defined as the likelihood that the reported results are correct. The second, clinical validity is the degree to which the test correctly assesses risk of health or disease [eg, the test's sensitivity, specificity, incidence of false positive and false negative results, and positive and negative predictive value]. Finally, clinical utility is the degree to which the test guides medical management [including whether there is scientific evidence for interventions that are safe, effective and available to the individual being tested]."[1 18]

As genetic tests in research protocols are often still under investigation, are not required to follow clinical laboratory improvement amendments (CLIA) regulation, and might not have established clinical validity or utility, research results are often not reported to subjects or included in EHRs.[10 19–29] However, debate on this continues.[19 21 26 29] A summary of current policies from cancer genomic studies and a recent guideline from the National Heart, Lung and Blood Institute can provide a framework to determine if results should be returned to study participants.[19 26] These articles agree that data to be returned to patients should: be analytically valid; be clinically valid; and have established benefit for patients.[19 26] Clinical validity should distinguish results that identify known diseases from conditions that are merely biological variations of uncertain clinical significance.[30] If a research protocol anticipates the return of individual results, participants' informed and comprehending preferences should be documented and specify the type of information desired (expected results, incidental findings, or no results) and whether the information should be made available to personal physicians, relatives and/or included in EHRs.

Box 1 Definitions

CDS (clinical decision support)—"a process for enhancing health-related decisions and actions with pertinent, organized clinical knowledge and patient information to improve health and healthcare delivery"[7]

Chromosome—the self-replicating structure of cells made of genes[8]

CLIA (clinical laboratory improvement amendments)—federally mandated oversight of the quality of laboratory tests to ensure the accuracy, reliability and timeliness of patient test results[9]

Clinical tests—"those [tests] in which specimens are examined and results reported to the provider for medical purposes, such as diagnosis, prevention or treatment in the care of individual patients"[10]

DICOM (digital imaging and communications in medicine)—a standard method for medical images and associated information to be exchanged between devices manufactured by various vendors[11]

Direct to consumer (DTC) genetic test—"genetic tests marketed and offered directly to consumers"[10]

DNA (deoxyribonucleic acid)—the chemical that contains instructions to direct cell activity[8]

Exome sequencing—determination of the order of base sequences in all the protein coding regions of an organism's genome[8]

Gene—the basic and functional units of heredity that are made of sequences of base pairs that provide instructions on how to make proteins[8]

Genetic tests—"the analysis of human DNA, RNA, chromosomes, proteins or certain metabolites in order to detect alterations related to a heritable disorder"[12]

Genome—an organism's complete set of DNA[8]

Genome wide association studies—large research projects that evaluate thousands of genomes looking for variants (usually single nucleotide polymorphisms) that are present in greater numbers in subjects with a given disorder than in subjects without a disorder[13]

Mendelian inheritance genetic test—"[a genetic test that] focuses on one or a few genes, which are chosen because of the medical and family history and are strongly associated with disease"[14]

Microarray—a device consisting of a glass slide and a sample of many separate DNA sequences arranged in a pattern. Messenger RNA (mRNA) will bind to the DNA from which it was originally made. mRNA is placed on the slide and the amount and types that are bound indicates what DNA (genes) are expressed in the sample.[15]

PACS (picture archiving and communication system)—a medical imaging system that provides efficient storage and access to medical images from multiple sources[16]

Personalized medicine—"an emerging practice of medicine that uses an individual's genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease. Knowledge of a patient's genetic profile can help doctors select the proper medication or therapy and administer it using the proper dose or regimen."[17]

Personal genomic screening—"a rapidly evolving area of DTC genetic testing which is based on multiple statistical comparisons, or genome-wide associations"[14]

Research tests—"those [tests] in which specimens are examined for the purpose of gaining fundamental scientific knowledge, or for early stage development of a clinical test"[10]

RNA (ribonucleic acid)—a chemical found in the nucleus of cells that codes for proteins and is studied to determine which genes in a cell are expressed[15]

SNP (single nucleotide polymorphism)—DNA sequence variations that occur when a single base in the genome sequence is changed[8]

Whole genome sequencing—determination of the order of base sequences in an organism's complete set of DNA[8]

Unlike research genetic tests, clinical genetic tests in the USA must be performed in CLIA-certified laboratories, offering assurance of analytical validity. Clinical genetic tests are obtained to diagnose, treat, or manage a patient's medical condition and traditionally have been used to diagnose monogenic inherited diseases with high penetrance (Mendelian inherited disorders).[14] Traditional clinical genetic tests are ordered and interpreted by providers trained in genetics, and generally have high analytical validity, clinical validity, and utility.[13] In contrast, personal genomic tests can be ordered by providers or lay persons (direct to consumer (DTC)) and are often not validated.[13 31 32] Many personal genomic tests use genome-wide association studies to predict the likelihood of a person having or developing a particular disease or to predict a patient's response to a drug. These studies look at hundreds of thousands of human genomes and determine which variations (single nucleotide polymorphisms (SNPs)) are more commonly present in people with a particular condition.[31]
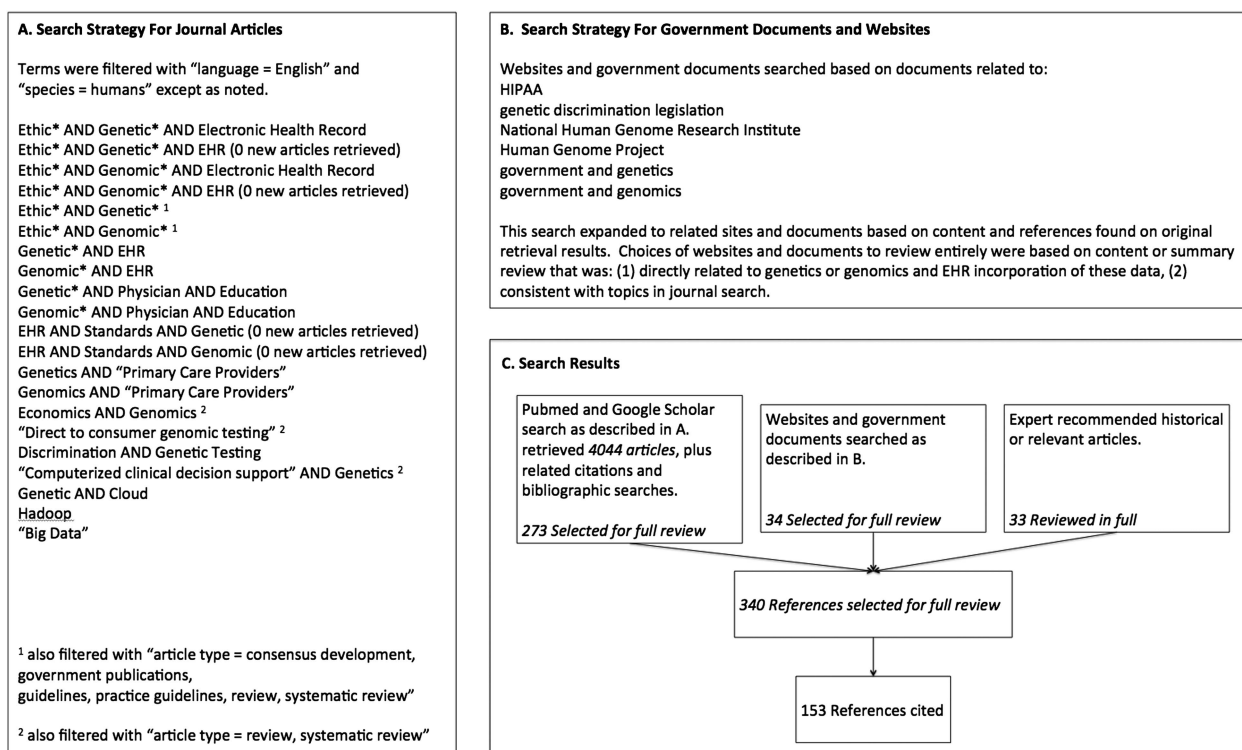
The public's growing interest in personal genomic testing coupled with limited government oversight (including limited validation and clinical guidelines for these tests) has encouraged the growth of DTC companies.[18 33–36] Despite improvements in the analytical validity of tests offered by these companies, concern regarding the clinical validity and utility of these tests remains high among scientific and medical communities.[13 14 21 31 37–43] Although DTC companies offer some monogenic tests that have shown clinical validity in specific populations, the validity may not be generalizable or applicable to different ethnic groups.[14] In addition, most DTC personal genomic tests based on SNP associations with diseases or traits have not yet shown clinical validity or utility, especially in terms of outcome evaluations.[13 14 37 44] Because of these limitations and concerns about inappropriate marketing, the Federal Trade Commission issued a consumer warning about at-home genetic tests and the US Food and Drug Administration and the Centers for Disease Control and Prevention (CDC) have advised that "at-home genetic tests are not a suitable substitute for a traditional healthcare evaluation."[45]

ETHICAL ISSUES SURROUNDING GENETIC DATA

When considering incorporation of genetic data into EHRs, one must realize that as analytical validity, clinical validity, and

A. Search Strategy For Journal Articles

Terms were filtered with "language = English" and "species = humans" except as noted.

Ethic* AND Genetic* AND Electronic Health Record
Ethic* AND Genetic* AND EHR (0 new articles retrieved)
Ethic* AND Genomic* AND Electronic Health Record
Ethic* AND Genomic* AND EHR (0 new articles retrieved)
Ethic* AND Genetic* [1]
Ethic* AND Genomic* [1]
Genetic* AND EHR
Genomic* AND EHR
Genetic* AND Physician AND Education
Genomic* AND Physician AND Education
EHR AND Standards AND Genetic (0 new articles retrieved)
EHR AND Standards AND Genomic (0 new articles retrieved)
Genetics AND "Primary Care Providers"
Genomics AND "Primary Care Providers"
Economics AND Genomics [2]
"Direct to consumer genomic testing" [2]
Discrimination AND Genetic Testing
"Computerized clinical decision support" AND Genetics [2]
Genetic AND Cloud
Hadoop
"Big Data"

[1] also filtered with "article type = consensus development, government publications, guidelines, practice guidelines, review, systematic review"

[2] also filtered with "article type = review, systematic review"

B. Search Strategy For Government Documents and Websites

Websites and government documents searched based on documents related to:
HIPAA
genetic discrimination legislation
National Human Genome Research Institute
Human Genome Project
government and genetics
government and genomics

This search expanded to related sites and documents based on content and references found on original retrieval results. Choices of websites and documents to review entirely were based on content or summary review that was: (1) directly related to genetics or genomics and EHR incorporation of these data, (2) consistent with topics in journal search.

C. Search Results

Pubmed and Google Scholar search as described in A. retrieved *4044 articles*, plus related citations and bibliographic searches.

273 Selected for full review

Websites and government documents searched as described in B.

34 Selected for full review

Expert recommended historical or relevant articles.

33 Reviewed in full

340 References selected for full review

153 References cited

Figure 1   Detailed description of search strategy and results.

utility decline, potential risks can outweigh potential benefits.[37 46] Data that are inaccurate, unreliable and/or not useful in guiding care have a low likelihood of improving patients' lives but may still cause harm (figure 2). The potential risks of these tests are outlined in box 2.[37 46–54]

Health information technology (HIT) professionals play a key role in protecting against these harms using effective data security and data governance. By ensuring high security (access management) in the use, storage and sharing of genetic data, the risk of confidentiality breach can be minimized. This can lessen other risks such as stigmatization, discrimination and family conflict. Central to this goal is the appropriate categorization and management of genetic data in relation to other data within EHRs. Proponents of 'genetic exceptionalism' recommend completely separate methods to protect, store and access genetic data due to the high potential for the harms noted below.[55 56] Others argue that genetic data are not sufficiently different from other medical data to justify different security measures.[46 55–58] In discerning how to store and secure access to different components of clinical records, Green and Botkin[58] recommend evaluating all data based on four characteristics: "(1) [the] degree to which information learned can be stigmatizing, (2) the effect of the test results on others, (3) the availability of effective interventions to alter the natural course predicted by the information, and (4) the complexity involved in interpreting test results." We agree that tests that raise problems associated with these characteristics should be treated with extra caution involving extra security and high standards for consent, similar to how HIV and mental health data are managed. However, genetic data are unique in the permanence of biomolecular observations coupled with the impermanence of interpretations. HIT professionals will play key roles designing methods to allow: ready access and use of validated genetic data; thorough documentation of when, how and where these data were obtained; and storage of genetic data that will not yet be incorporated in EHRs.

Maximizing benefit and minimizing harm requires the provision of informed consent before genetic testing along with

Figure 2   Examples of harm. Case 1 involves personal harms of confusion and refusal of recommended care because of false reassurance from genomic test. Case 2 relates to personal harm of confusion in diagnosis. Case 3 involves personal harm of confusion prompting changes in reproductive plans and financial harm with increased use of medical consultation.

Case 1:
Eng et al. offer a hypothetical example of delayed cancer diagnosis due to falsely reassuring DTC testing that prompted the patient to forgo validated cancer screening.[13]

Case 2:
Incorrect diagnoses based on unvalidated genetic tests, led to false diagnosis of a metabolic disorder with severe neurodevelopmental symptoms in 27 newborns. If the parents followed the physicians' recommendation, the children would have been kept on a special diet throughout their lives.[30]

Case 3:
Parents of children who were carriers for the gene for cystic fibrosis changed their reproductive choices based on the false belief that the child had the disease.[63]

Box 2   Potential harms of genetic testing

Personal

Anxiety, depression, confusion, changes in life plans, changes in reproductive plans, parental guilt about passing on a deleterious mutation, survivor guilt about not having a deleterious genetic mutation when other family members do, refusal of recommended medical care because of false reassurance from an invalid genetic test

Social

Stigmatization, breach of confidentiality, identification of misattributed parentage, privacy concerns or desire not to know a genetic result that may be at odds with family members' desire to know a genetic result, misuse of genetic data (surreptitious DNA testing or transfer of genetic data to third parties after sale of direct to consumer testing company)

Professional

Employment and training concerns (see legislative discussion below)

Financial

Increased use of medical consultation and follow-up tests, employment and insurance concerns (see legislative discussion below)

Insurance

Discrimination in obtaining health, disability, life and long-term care insurance (see legislative discussion below)

Box 3   Legislative acts relevant to genetic testing

Health Information Portability and Accountability Act of 1996 (HIPAA)[73]

Sets standards on how protected health information should be controlled

Does not apply to many companies or laboratories that perform direct to consumer genetic testing and analysis

Protects against genetic discrimination in employer-sponsored group health plans

Genetic Information Nondiscrimination Act of 2008 (GINA)[74]

Extends HIPAA protections by making it illegal to use genetic information to underwrite group and individual health insurance

Prohibits employers from making employment decisions based on genetic information

Does not address life insurance, disability insurance or long-term care insurance discrimination

Does not apply to health benefits for federal employees, members of the military, veterans seeking healthcare through the Department of Veterans Affairs, or the Indian Health Service

Does not apply to athletic programs

Patient Protection and Affordable Care Act of 2010 (ACA)[75]

Prohibits health insurers from determining eligibility for coverage based on signs and symptoms of genetic disease

Changes in 2014: prohibits differences in premiums according to health status and genetic information

post-test counseling so individuals have adequate information to make their own decisions.[59][60] Predictive genetic testing, especially non-targeted, multiple-gene testing or whole genome or exome sequencing, challenges providers' abilities to complete these tasks. Due to the large volume of information produced by these tests, lack of validation of many of the results, and scarcity of trained specialists to provide genetic counseling, patients may be inadequately prepared to make informed choices regarding testing or follow-up. In addition, many healthcare providers and lay persons possess limited understanding of genetics and statistics, which may result in unfounded worry, requests for unnecessary tests or interventions, and irrational reproductive decisions.[61–64] HIT professionals can aid progress towards the goal of informed consent through research and development of interactive genetic computerized clinical decision support (CDS) tools, the provision of easy access to genetic information databases, and the development of online genetic and statistical educational resources.[65–72] Developers can also help create efficient online consent forms for genetic testing, research participation and incorporation of genetic data into EHRs.

Key legislative mandates related to genetic testing, insurance and employment are outlined in box 3. Discrimination in access to insurance and employment opportunities related to genetic testing has been discussed in many articles.[47–50][52] These US laws were enacted to minimize the risk of employment and health insurance discrimination due to genetic findings. However, discrimination in eligibility for disability, life or long-term care insurance is still not formally addressed by federal statutes. Also, despite the passage of the Genetic Information Nondiscrimination Act in 2008, concern and controversy regarding unfair treatment due to a genetic predisposition still exist.[76] HIT professionals can assist in this area by ensuring high security and Health Information Portability and Accountability Act (HIPAA) compliance in the design of patient care software, providing for easy updates should relevant laws change, and enabling effective, efficient and HIPAA-compliant research access to genetic data as authorized by research subjects.[77]

LOGISTICAL ISSUES SURROUNDING GENETIC DATA INCORPORATION INTO EHRS

A fundamental barrier to the ethical and efficient incorporation of genetic data into EHRs is the lack of central validation and oversight of genetic tests.[3][40][42][78] The USA's current fragmented system can result in confusion and increased potential for harm due to incorrect interpretations, use of genetic tests of low validity and/or utility, or the under-utilization of tests with high validity and utility.[79–81] Box 4, adapted from Fabsitz et al,[19] presents online sources of information about genetic tests and their validity and utility.

It is clear that a central validation body for genetic tests is long overdue.[2][3] The evidence-based validation efforts of the Working Group of Evaluation of Genomics Applications in Practice and Prevention (EGAPP) are exemplary; however, their ability to keep pace with the rapid development of new genetic tests is evident in the small number of tests for which they have offered guidance (six recommendations and eight evidence reports) compared to the 2687 tests currently listed on the National Institutes of Health's (NIH) Genetic Test Registry.[84] The rapid development of these tests is further documented by the GeneTest Laboratory directory growth chart, which lists just over 100 tests available in 1993 and over 2500 now.[93] Before genomics can be used in daily clinical decision-making, more research is needed to establish systematic methods to ensure the

Box 4	Online sources of genetic test information

CDC Public Health Genomics[82]

Provides information on multiple topics in genomics and links to specific topic sites (such as EGAPP and GAPP KB below). It also provides an alphabetized list of websites and topics related to genomics (genomic resources)

http://www.cdc.gov/genomics/

Evaluation of Genomic Applications in Practice and Prevention (EGAPP)[83]

Established in 2004 by the CDC's Office of Public Health Genomics

Goal: "establish a systematic, evidence based process for assessing genetic tests and the application of genetic technology in transition from research to clinical and public health practice"

http://www.egappreviews.org

GeneTests[12]

Sponsored by the University of Washington, Seattle and Bioreference Laboratories

Provides authoritative information from global genetic experts on genetic testing and its use in diagnosis, management and treatment of disease

www.genetests.org

Genetic Testing Registry (GTR)[84]

Established in 2012 by the NIH

Goal: to provide "a central location for the voluntary submission of genetic test information by providers (including) the test's purpose, methodology, validity, evidence of the test's usefulness, and laboratory contacts and credentials (in order to) advance the public health and research into the genetic basis of health and disease"

Intended audience: providers and researchers

No NIH oversight (voluntary submission of information)

GTR to replace GeneTests Laboratory directory in 2013

http://www.ncbi.nlm.nih.gov/gtr

Genomic Applications in Practice and Prevention Network (GAPPNet)[85]

Established in 2009 by the CDC's Office of Public Health Genomics, the National Cancer Institute's Division of Cancer Control and Population Sciences and other stakeholders

Goal: "to accelerate and streamline effective and responsible use of validated and useful genomic knowledge and applications, such as genetic tests, technologies, and family history, into clinical and public health practice"

http://www.cdc.gov/genomics/translation/GAPPNet

Genomics Applications in Practice and Prevention Knowledge Base (GAPP KB)[86]

Supported by the CDC's Office of Public Health Genomics

Goal: to provide access to information on applications of genomic research to healthcare. This site provides links to four other sites that provide information on genomic applications (GAPPFinder), genomic test evidence (the evidence aggregator), genomic research studies archive (the project locator), and the online journal of genomic research (*PLoS Currents: Evidence on Genomic Tests*)

http://www.hugenavigator.net/GAPPKB/home.do

Human Genome Epidemiology Network (HuGENet)[87]

Established by the CDC's Office of Public Health Genomics

A volunteer collaborative effort "to help translate genetic research findings into opportunities for preventive medicine and public health."

http://www.cdc.gov/genomics/hugenet

Institute of Medicine of the National Academies: Roundtable on Translating Genomic-Based Research for Health[88]

A non-profit, multidisciplinary collaborative effort of leaders from academia, industry and government to translate genomic research findings into health care, education and policy improvements

http://www.iom.edu/Activities/Research/GenomicbasedResearch.aspx

Online Mendelian Inheritance of Man—(OMIM)[89]

A catalog of human genes and genetic disorders maintained by experts in genetics at Johns Hopkins University

Intended audience: providers and researchers

http://www.omim.org

Pharmacogenomics Knowledgebase (PharmGKB)[90]

An authoritative "pharmacogenomics knowledge resource that encompasses clinical information including dosing guidelines and drug labels, potentially clinically actionable gene-drug associations and genotype–phenotype relationships."

http://www.pharmgkb.org

US DHHS Secretary's Advisory Committee on Heritable Disorders in Newborns and Children—(SACHDNC)[91 92]

A website containing information on the SACHDNC's recommendations to the secretary of the US Department of Health and Human Services on the "most appropriate application of universal newborn screening tests, technologies, policies, guidelines and standards."

http://www.hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/

analytical validity, clinical validity, and utility of these tests.[44] These investigations must evaluate morbidity and mortality outcomes along with expected healthcare cost savings from these predictive tests. Primary research studies and methodical review of peer-reviewed evidence of the benefit and harms of genetic tests, similar to the US Preventive Services Task Force methods, should be pursued.[94]

Another logistical problem with incorporating genetic data into EHRs is the need for more clinicians trained in genetics.[2 95 96] Recommendations to improve genetic education in medical schools and guidelines on genetic competency for primary care physicians have been proposed.[61 97–102] Studies of primary care physicians have shown a poor understanding of genetics and dissatisfaction with support tools at the point of care.[1 101 103–106] Computerized training tools used at providers' convenience and computerized CDS tools used at the point of care will be instrumental in facilitating clinicians' effective management of genetic issues.[68 103 105 107–110] Also, there is an urgent need to improve primary and secondary school math training in statistics, which is poorly understood by the majority of the population and is essential in understanding genetic testing.[62 64 111] Here too, interactive computerized tools may facilitate faster, more accurate comprehension.[112]

TECHNOLOGICAL ISSUES SURROUNDING GENETIC DATA INCORPORATION INTO EHRS

The US government is promoting the evaluation of genomics incorporation into clinical workflows.[113] Currently, EHR systems handle genetic data like any other laboratory test. Those solutions are not optimal, requiring HIT developers to collaborate with scientists and health professionals to redesign EHR systems that allow efficient, secure incorporation of genetic/genomic data.[114] Workflows to be considered include genetic/genomic knowledge access at the point of care, CDS tool development, order entry, test tracking, result interpretation, result retrieval, result re-interpretation and tracking, and patient notification. Green and Guyer[69] explained, "Existing clinical informatics architectures are largely incapable of storing genome sequence data in a way that allows the information to be searched, annotated and shared across healthcare systems over an individual's lifespan." HIT developers working on genetic data incorporation must overcome many of the same difficulties that challenged developers of digital imaging and communications in medicine and picture archiving systems.[16] Radiological images and their interpretations are similar to genetic sequence data in that both require standardized methods to communicate unstructured and structured data, contain large volumes of data requiring compression, must interface with EHRs, contain identifiable patient information, and cannot be queried in their raw (unstructured) form. Masys et al[115] have addressed many of these challenges in their paper describing seven essential requirements for the incorporation of genomic data into EHRs: separation of genetic sequence and genetic interpretation data, lossless compression of the sequence data, documentation of laboratory method details (when, where, how data were obtained), use of short codes to represent large sequence data to allow faster retrieval and analysis, design for human and machine-readable forms of the data, design for future changes to the data, and design for clinical and research use of the data.

To implement these features, HIT developers must agree on standardized genetic terminology and methods of data transfer. Standard structures and language are required for genetic data to be exchanged between clinics/EHR systems and to drive clinical decision support tools (figure 3). Health Level 7 (HL7), the
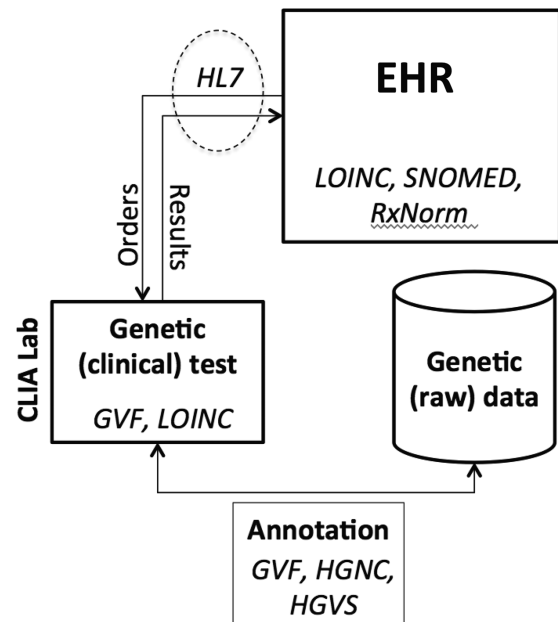
Figure 3 Diagram showing integration of genetic data to the electronic health record (EHR) and a clinical laboratory improvement amendments (CLIA) certified laboratory for genetic tests. Also shown are examples of standards available for: data messaging (HL7), genetic data annotation (genomic variation format (GVF), Human Gene Nomenclature Committee's terminology (HGNC), Human Genome Variation Society's (HGVS)) and genetic data representation for clinical use (Logical Observation Identifiers Names and Codes (LOINC), SNOMED, RxNorm, GVF).

widely used standard for EHR data transfer, has a clinical genomics work group working on areas such as the genetic testing report form and the family history model.[116–118] Currently, the main standards used for genetic tests include: the Logical Observation Identifiers Names and Codes (LOINC), the Human Genome Variation Society's (HGVS) nomenclature, the Human Gene Nomenclature Committee's (HGNC) terminology, Reference Sequences NCBI (RefSeq), the database of single nucleotide polymorphisms (dbSNP) and the International System for Human Cytogenetic Nomenclature (ISCN). The Genomic Variation Format also facilitates genetic data incorporation in EHRs through simplified formatting.[119] Standards for associations of genetic mutations and disease states or drugs are LOINC, SNOMED and RxNorm, but many terms used to describe genetic diseases that are listed in the Online Mendelian Inheritance in Man database are not easily mapped to SNOMED or integrated into existing EHR systems.[120] Further collaborative research between HIT developers and genetic research and clinical experts is needed to develop and test standardized models to ascertain the best representation of all genetic data.

Once genetic data standardization is agreed upon, multiple technological issues must still be addressed. As noted above, the unique nature of genetic data in its simultaneous permanence and impermanence mandates that the biomolecular (permanent) findings be available in machine and human-readable forms that allow re-interpretation as new understanding of these data develops.[121] In deciding how to incorporate genetic data into EHRs, a distinction between validated test results and unvalidated test results must be made. More research is needed to determine how best to ensure that test results that are not yet clinically valid or useful do not influence patient care inappropriately. The GeneInsight Suite platform uses a reporting system

within the EHR that designates a variant as benign, likely benign, unknown significance, likely pathogenic or pathogenic.[122] This allows providers to see all variants, however, this amount of data may be difficult for providers and EHR systems to manage. Also, research needs to evaluate whether a variant labeled as benign or of unknown significance may result in the 'vulnerable child syndrome' or a population of 'worried well' due to persistent concern about such 'abnormal' results.[123–125]

One possible mechanism to address these concerns is to use a separate warehouse (a virtual shadow chart) to store all test results and interpretations that are not yet validated or useful, only uploading them into the EHR when they are determined to meet high standards for clinical validity and utility. Ideally this determination would be made by a central body overseen by a local specialist who would approve transfer into the EHR and confirm that appropriate patient consent and counseling occurred. This system would mimic the current research protocol of the Pharmacogenomics of Anticancer Agents Research in children (PAAR4Kids).[126] Problems with such a system include the need for a separate, secure warehouse, the need for robust warehouse/EHR interoperability, the current shortage of genetic specialists, and the current lack of a central validating body to determine a test's readiness for EHR integration. Also, legal or ethical risks of excluding unvalidated information would need to be determined through further scholarly work.[127]

Given the large volume of data that comprises the human genome and its interpretation, genomic data storage requirements must be decreased and search and retrieval capabilities improved. As discussed by Masys et al,[115] general compression formats are inadequate for genetic data due to the need for lossless reduction of the data and the extremely large datasets that are involved. They proposed lossless compression of genetic data by "representing personal nucleotide and/or protein sequences as the difference between the individual and what [they] propose calling a 'Clinical Standard Reference Genome'" resulting in a 100-fold data size reduction.[115] Drawbacks of this method include the current lack of a universal standard reference genome, the large storage space needed for the reference genome, and the significant time needed for decompression of the data.[128] Qiao et al[128] proposed a compression algorithm called "SpeedGene" that chooses from among three algorithms based on disk space and claims to overcome problems of large files and long loading times of genetic data. Vey[129] proposed the "Differential Direct Coding algorithm" that is able to distinguish expected sequence data containing the usual nucleotide bases from other data types and codes these data differently. This algorithm accurately codes (rather than removes and stores) genetic data that may contain wildcards (other nucleotide bases), annotation data or special repeats. This improves time to compress and decompress the data and eliminates the need for separate storage of wildcard data.[129] These methods appear promising, but further research is needed to determine the best method of genetic data compression.

In addition to data compression, innovations are needed in data storage and computation capability. One solution to both of these problems could be cloud-based computing in which customers pay only for the amount of storage and computation time used.[130–132] The flexibility of a system that expands or shrinks as storage and computation needs change may provide a good economic option.[133 134] Disadvantages of cloud computing include data transfer time to and from the cloud and legal/ethical issues regarding patient data security and privacy.[130 135 136] Heterogeneous computational environments offer another option to increase computational speed through

the use of computers containing specialized accelerators (graphical processing units) that improve arithmetic processing by 10 to 100-fold.[130] Although this technology will not alleviate genetic data storage concerns, it could be coupled with cloud computing to maximize both computational and storage capability. A difficulty with this faster technology is that it requires specialized programming languages and informatics expertise to develop or modify applications.[130] Another solution to problems with storage and processing of genetic big data may be found in the open source software project Hadoop and its programming model, MapReduce, which partitions and distributes structured and unstructured data and their analysis in parallel, over multiple computers. This can be accomplished in the cloud or using a hybrid cloud-cluster architecture, allowing big data to be stored, queried and processed quickly and cost effectively.[137 138] Hadoop may be well suited to the volume and heterogeneity of genetic data, but security concerns over data storage location (cloud vs clusters of computers) and associated software usage will need to be addressed to ensure HIPAA compliance. A solution to security and privacy problems in genome interpretation has been proposed by Knome's stand-alone system with enough storage and power to process one genome per day.[139]

Data security and privacy concerns are not unique to genomic data but critical to address given the ethical issues discussed above, the pressing need for data sharing, and the vanishing concept of data de-identification as it relates to genomic data.[20 57 140–146] The US Presidential Commission for the Study of Bioethical Issues recently released recommendations for ensuring patient privacy and data security while promoting scientific advancement through the use of whole genome sequencing. The five recommendations discuss the need for: "strong baseline protections while promoting data access and sharing; data security and access to databases; consent; facilitating progress in whole genome sequencing; [and] public benefit."[147] These recommendations discuss many of the same issues addressed by the Working Conference of the American Medical Association in 2007, which proposed a model of data stewardship for health data collection, storage, use and disclosure. The resultant paper by Bloomrosen and Detmer[148] defined principles of data stewardship to include: accountability, transparency, need for consent, appropriate notice to patients regarding use of their data (consistent with permitted uses/disclosures), attention to technical issues involving security and de-identification of data, and enforcement of these principles. Models that allow both clinical and research use of data have been proposed including caBIG, I2B2, STRIDE and the Information Warehouse.[149–152] An excellent evaluation of different integrated data storage and use structures is provided by Louie et al.[153]

CONCLUSION AND FUTURE STEPS

Ethical issues surrounding the incorporation of genetic data into EHRs can be summarized as a need to maximize good and minimize harm while respecting people as autonomous beings with a right to make their own decisions. These principles have been applied to genetic data integration to identify the following key challenges HIT professionals face.

1. Standardization of genetic content (raw data, ontology) to allow accurate and efficient interpretation of genetic data from EHRs, facilitate development of computerized CDS tools, and secure transmission of data between EHRs. As stabilization of such standards can take decades, warehousing of raw genetic data may be a good practice for the time being.

2. Development of information infrastructure to facilitate genetic test validation and computerized CDS tool creation to assist providers with interpretation, follow-up and re-interpretation of these data.

3. Development of efficient storage of genetic data using compression and archiving methods uniquely suited to the size, structure and security requirements of these data.

4. Development of workflow models that allow secure clinical, research and auditable use of genetic data.

Other areas of need include computerized training programs in genetics and statistics, better patient consent tools and enhanced educational focus on statistics. This paper is offered as a review of current challenges and avenues for continuing exploration as we journey towards the ethical and efficient incorporation of genetic data in EHRs and, ultimately, the delivery of personalized medicine.

REFERENCES

1. Lose EJ. The emerging role of primary care in genetics. *Curr Opin Pediatr* 2008;20:634–8.
2. National Research Council (US). Committee for the Study of Inborn Errors of Metabolism, Genetic Screening: programs, principles and research/Committee for the Study of Inborn Errors of Metabolism, Division of Medical Sciences, Assembly of Life Sciences, National Research Council. Washington, DC: National Academy of Sciences, 1975.
3. Simopoulos AP. Genetic screening: programs, principles, and research–thirty years later. Reviewing the recommendations of the Committee for the Study of Inborn Errors of Metabolism (SIEM). *Public Health Genomics* 2009;12:105–11.
4. Andermann A, Blancquaert I, Beauchamp S, et al. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull WHO* 2008;86:317–19.
5. Andrews LB. *Assessing genetic risks: implications for health and social policy*. National Academies Press, 1994.
6. Botkin JR, Clayton EW, Fost NC, et al. Newborn screening technology: proceed with caution. *Pediatrics* 2006;117:1793–9.
7. HIMSS. [1 August 2012]. http://www.himss.org/ASP/topics clinicalDecision.asp.
8. Human Genome Project Information, Genome Glossary. [1 August, 2012]. http://genomics.energy.gov.
9. US Food and Drug Administration. [23 July 2012]. http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm124105.htm.
10. FDA Executive Summary, Molecular and Clinical Genetics Panel, 8 & 9 March, 2011. 2011 [23 July 23 2012]. http://www.fda.gov/downloads/AdvisoryCommittees?CommitteesMeetingMaterials/MedicalDevices/MedicalDevicesAdvisoryCommittee/MolecularandClinicalGeneticsPanel/UCM245660.pdf.
11. Digital Imaging and Communications in Medicine. 2012 [24 August 2012]. medical.nema.org/.
12. University of Washington, Seattle, GeneTests Medical Genetics Information Resource (database online)]. cited 1 August 2012]. http://www.ncbi.nlm.nih.gov/sites/GeneTests. Note: new website under construction June 2013 at www.genetests.org
13. Eng C, Sharp RR. Bioethical and clinical dilemmas of direct-to-consumer personal genomic testing: the problem of misattributed equivalence. *Sci Transl Med* 2010;2:17cm5.
14. Edelman E, Eng C. A practical guide to interpretation and clinical application of personal genomic screening. *BMJ (Clinical Research edn)* 2009;339:b4253.
15. National Center for Biotechnology Information (NCBI). [22 July 2012]. http://www.ncbi.nlm.nih.gov.
16. Napoli M, Nanni M, Cimarra S, et al. Picture archiving and communication in radiology. *Rays* 2003;28:73–81.
17. NLM. Genetics Home Reference. [9 April 2013]. http://ghr.nlm.nih.gov/glossary=personalizedmedicine.
18. NIH. Department of Health and Human Services Secretary's Advisory Committee on Genetic Testing. 2008 [7 July 2012]. http://oba.od.nih.gov/SACGHS/sacghs_documents.html.
19. Fabsitz RR, McGuire A, Sharp RR, et al. Ethical and practical guidelines for reporting genetic research results to study participants: updated guidelines from a National Heart, Lung, and Blood Institute working group. *Circ Cardiovasc Genet* 2010;3:574–80.
20. Haga SB, Beskow LM. Ethical, legal, and social implications of biobanks for genetics research. *Adv Genet* 2008;60:505–44.
21. Kohane IS, Masys DR, Altman RB. The incidentalome: a threat to genomic medicine. *JAMA* 2006;296:212–15.
22. Wolf SM. Introduction: the challenge of incidental findings. *J Law Med Ethics* 2008;36:216–18.
23. Wolf SM, Paradise J, Caga-anan C. The law of incidental findings in human subjects research: establishing researchers' duties. *J Law Med Ethics* 2008;36:361–83, 214.
24. Wolf SM, Lawrenz FP, Nelson CA, et al. Managing incidental findings in human subjects research: analysis and recommendations. *J Law Med Ethics* 2008;36:219–48, 1.
25. Cho MK. Understanding incidental findings in the context of genetics and genomics. *J Law Med Ethics* 2008;36:280–5; 12.
26. Dressler LG. Disclosure of research results from cancer genomic studies: state of the science. *Clin Cancer Res* 2009;15:4270–6.
27. Van Ness B. Genomic research and incidental findings. *J Law Med Ethics* 2008;36:292–7; 12.
28. Clayton EW. Informed consent and biobanks. *J Law Med Ethics* 2005;33:15–21.
29. Hudson KL. Genomics, health care, and society. *N Engl J Med* 2011;365:1033–41.
30. van Calcar SC, Gleason LA, Lindh H, et al. 2-Methylbutyryl-CoA dehydrogenase deficiency in Hmong infants identified by expanded newborn screen. *WMJ* 2007;106:12–15.
31. Samuel GN, Jordens CF, Kerridge I. Direct-to-consumer personal genome testing: ethical and regulatory issues that arise from wanting to 'know' your DNA. *Intern Med J* 2010;40:220–4.
32. Wright CF, Hall A, Zimmern RL. Regulating direct-to-consumer genetic tests: what is all the fuss about? *Genet Med* 2011;13:295–300.
33. Direct-to-Consumer Genetic Testing. Report of the Secretary's Advisory Committee on Genetics Health and Society, 2010.
34. 23andMe. 2012 [15 July 2012]. http://www.23andme.com.
35. deCODEme. 2012 [cited 2012 July 15 2012]. http://www.decodeme.com.
36. Navigenics. 2012 [cited 2012 July 15 2012]. http://www.navigenics.com.
37. Bunnik EM, Schermer MH, Janssens AC. Personal genome testing: test characteristics to clarify the discourse on ethical, legal and societal issues. *BMC Med Ethics* 2011;12:11.
38. Udesky L. The ethics of direct-to-consumer genetic testing. *Lancet* 2010;376:1377–8.
39. US Government Accountability Office. Direct to Consumer Genetic Tests: Misleading Test Results Are Further Complicated by Deceptive Marketing and Other Questionable Practices. 2010 [August 01, 2012]. http://www.gao.gov/products/GAO.
40. Javitt GH, Stanley E, Hudson K. Direct-to-consumer genetic tests, government oversight, and the First Amendment: what the government can (and can't) do to protect the public's health. *Okla Law Rev* 2004;57:251–302.
41. Javitt G. Which way for genetic-test regulation? Assign regulation appropriate to the level of risk. *Nature* 2010;466:817–18.
42. Javitt GH. In search of a coherent framework: options for FDA oversight of genetic tests. *Food Drug Law J* 2007;62:617–52.
43. Skirton H, Goldsmith L, Jackson L, et al. Direct to consumer genetic testing: a systematic review of position statements, policies and recommendations. *Clin Genet* 2012;82:210–18.
44. Khoury MJ, McBride CM, Schully SD, et al. The Scientific Foundation for Personal Genomics: recommendations from a National Institutes of Health–Centers for Disease Control and Prevention multidisciplinary workshop. *Genet Med* 2009;11:559–67.
45. Federal Trade Commission. At-Home Genetic Tests: A Healthy Dose of Skepticism May Be the Best Prescription. 2007. 3 August 2012. http://www.ftc.gov/bcp/edu/pubs/consumer/health/hea02.shtm.

46 Burke W, Pinsky LE, Press NA. Categorizing genetic tests to identify their ethical, legal, and social implications. *Am J Med Genet* 2001;106:233–40.

47 Robertson JA. The $1000 genome: ethical and legal issues in whole genome sequencing of individuals. *Am J Bioeth* 2003;3:W–IF1.

48 Fost N. Ethical implications of screening asymptomatic individuals. *FASEB J* 1992;6:2813–17.

49 Fost N. Ethical issues in genetics. *Pediatr Clin North Am* 1992;39:79–89.

50 Wilcken B. Ethical issues in genetics. *J Paediatr Child Health* 2011; 47:668–71.

51 Almond B. Genetic profiling of newborns: ethical and social issues. *Nat Rev Genet* 2006;7:67–71.

52 Fost N. Genetic diagnosis and treatment. Ethical considerations. *Am J Dis Child (1960)* 1993;147:1190–5.

53 Rogowski WH, Grosse SD, John J, *et al*. Points to consider in assessing and appraising predictive genetic tests. *J Commun Genet* 2010;1:185–94.

54 Heshka JT, Palleschi C, Howley H, *et al*. A systematic review of perceived risks, psychological and behavioral impacts of genetic testing. *Genet Med* 2008;10:19–32.

55 Rothstein MA. Genetic exceptionalism and legislative pragmatism. *J Law Med Ethics* 2007;35(2 Suppl):59–65.

56 Ross LF. Genetic exceptionalism vs. paradigm shift: lessons from HIV. *J Law Med Ethics* 2001;29:141–8.

57 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;10:495–9.

58 Green MJ, Botkin JR. "Genetic exceptionalism" in medicine: clarifying the differences between genetic and nongenetic tests. *Ann Intern Med* 2003;138:571–5.

59 Botkin JR. Genetic counseling: making room for beneficence. *J Clin Ethics* 1995;6:182–4.

60 Moulton B, King JS. Aligning ethics with medical decision-making: the quest for informed patient choice. *J Law Med Ethics* 2010;38:85–97.

61 Guttmacher AE, Porteous ME, McInerney JD. Educating health-care professionals about genetics and genomics. *Nat Rev Genet* 2007;8:151–7.

62 Gaissmaier W, Gigerenzer G. Statistical illiteracy undermines informed shared decision making. *Zeitschrift fur Evidenz, Fortbildung und Qualitat im Gesundheitswesen* 2008;102:411–13.

63 Mischler EH, Wilfond BS, Fost N, *et al*. Cystic fibrosis newborn screening: impact on reproductive behavior and implications for genetic counseling. *Pediatrics* 1998;102:44–52.

64 Gigerenzer G, Gaissmaier W, Kurz-Milcke E, *et al*. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8:53–96.

65 Jing X, Kay S, Marley T, *et al*. Incorporating personalized gene sequence variants, molecular genetics knowledge, and health knowledge into an EHR prototype based on the Continuity of Care Record standard. *J Biomed Inform* 2012;45:82–92.

66 Green MJ, McInerney AM, Biesecker BB, *et al*. Education about genetic testing for breast cancer susceptibility: patient preferences for a computer program or genetic counselor. *Am J Med Genet* 2001;103:24–31.

67 Green MJ, Fost N. An interactive computer program for educating and counseling patients about genetic susceptibility to breast cancer. *J Cancer Educ* 1997;12:204–8.

68 Green MJ, Biesecker BB, McInerney AM, *et al*. An interactive computer program can effectively educate patients about genetic testing for breast cancer susceptibility. *Am J Med Genet* 2001;103:16–23.

69 Green ED, Guyer MS. Charting a course for genomic medicine from base pairs to bedside. *Nature* 2011;470:204–13.

70 Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. *J Am Med Inform Assoc* 2013;20:388–400.

71 Green MJ, Fost N. Who should provide genetic education prior to gene testing? Computers and other methods for improving patient understanding. *Genet Test* 1997;1:131–6.

72 Dabney MK, Huelsman K. Counseling by computer: breast cancer risk and genetic testing. Developed by the University of Wisconsin-Madison Department of Medicine and the Program in Medical Ethics. *Genet Test* 2000;4:43–4.

73 US Department of Health and Human Services. Office for Civil Rights, Health Information Privacy, Understanding HIPAA Privacy]. 30 July 2012]. http://www.hhs. gov/ocr/privacy/.

74 US Equal Employment Opportunity Commission. [cited 1 August 2012]. http:// www.eeoc.gov/laws/statutes/gina.cfm.

75 US Department of Health and Human Services, Healthcare.gov. [1 August 2012]. http://www.healthcare.gov/law/index.html.

76 Klitzman R. Views of discrimination among individuals confronting genetic disease. *J Genet Couns* 2010;19:68–83.

77 Weng C, Appelbaum P, Hripcsak G, *et al*. Using EHRs to integrate research with patient care: promises and challenges. *J Am Med Inform Assoc* 2012; 19:684–7.

78 Wilfond BS, Nolan K. National policy development for the clinical application of genetic diagnostic technologies. Lessons from cystic fibrosis. *JAMA* 1993;270:2948–54.

79 Wilfond BS, Fost N. The cystic fibrosis gene: medical and social implications for heterozygote detection. *JAMA* 1990;263:2777–83.

80 Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. *Transl Res* 2009;154:277–87.

81 Malinowski MJ. Separating predictive genetic testing from snake oil: regulation, liabilities, and lost opportunities. *Jurimetrics* 2000;41:23–52.

82 CDC. Office of Public Health Genomics]. August 2012]. http://www.cdc.gov/ genomics/

83 EGAPP Work Group. Evaluation of Genomic Applications in Practice and Prevention. 2012 [20 July 2012]. http://www.egappreviews.org.

84 National Center for Biotechnology Information. Genetic Testing Registry]. 22 July 2012]. http://www.ncbi.nlm.nih.gov/gtr/.

85 Centers for Disease Control and Prevention. Genomic Applications in Practice and Prevention Network. 2012 [21 July 2012]. http://www.cdc.gov/genomics/ translation/GAPPNet.

86 Genomics in Application, Practice, and Prevention Knowledge Base (GAPP KB). [1 August 2012]. http://www.hugenavigator.net/GAPPKB/home.do.

87 Centers for Disease Control and Prevention, the Office of Public Health Genomics, Human Genome Epidemiology Network. [21 July 2012]. http://www.cdc.gov/ genomics/hugenet.

88 Institute of Medicine of the National Academies: Roundtable on Translating Genomic-Based Research for Health. 2012 [cited 2012, 21 July 2012]. http:// www.iom.edu/Activities/Research/GenomicBasedResearch.aspx.

89 Johns Hopkins University, Online Mendelian Inheritance In Man. 2012 [21 July 2012]. http://www.omim.org.

90 PharmGKB. The Pharmacogenomics Knowledgebase. 2012 [cited 2012, 21 July 2012]. http://www.pharmgkb.org.

91 US Department of Health and Human Services Hg. Secretary's Advisory Committee on Heritable Disorders in Newborns and Children [3 January 2013]. http://www. hrsa.gov/advisorycommittees/mchbadvisory/heritabledisorders/.

92 Trotter TL, Fleischman AR, Howell RR, *et al*. Secretary's Advisory Committee on Heritable Disorders in Newborns and Children response to the President's Council on Bioethics report: the changing moral focus of newborn screening. *Genet Med* 2011;13:301–4.

93 GeneTests: Laboratory Directory. 2012 [cited 21 July 2012]. http://www.ncbi.nlm. nih.gov/sites/GeneTests/lab?db=GeneTests.

94 Force USPST. US Preventive Services Task Force, Methods and Processes]. 3 January 2013]. http://www.uspreventiveservicestaskforce.org/methods.htm.

95 American Board of Medical Geneticists. 2012 [16 July 2012]. http://www.abmg.org.

96 Genetics Education Training, Report of the Secretary's Advisory Committee on Genetics, Health, and Society. In: Services DoHaH, editor, 2011.

97 American Academy of Family Physicians Core Educational Guidelines: Medical Genetics (2011). 2010, 26 February 2012.

98 Thurston VC, Wales PS, Bell MA, *et al*. The current status of medical genetics instruction in US and Canadian medical schools. *Acad Med* 2007; 82:441–5.

99 Burke S, Martyn M, Thomas H, *et al*. The development of core learning outcomes relevant to clinical practice: identifying priority areas for genetics education for non-genetics specialist registrars. *Clin Med* 2009;9:49–52.

100 Trinidad SB, Fryer-Edwards K, Crest A, *et al*. Educational needs in genetic medicine: primary care perspectives. *Community Genet* 2008;11:160–5.

101 Burke W, Emery J. Genetics education for primary-care providers. *Nat Rev Genet* 2002;3:561–6.

102 Laberge AM, Fryer-Edwards K, Kyler P, *et al*. Long-term outcomes of the "Genetics in Primary Care" faculty development initiative. *Fam Med* 2009;41:266–70.

103 Carroll JC, Rideout AL, Wilson BJ, *et al*. Genetic education for primary care providers: improving attitudes, knowledge, and confidence. *Can Fam Physician* 2009;55:e92–9.

104 Bellcross CA, Kolor K, Goddard KA, *et al*. Awareness and utilization of BRCA1/2 testing among US primary care physicians. *Am J Prev Med* 2011;40:61–6.

105 Houwink EJ, van Luijk SJ, Henneman L, *et al*. Genetic educational needs and the role of genetics in primary care: a focus group study with multiple perspectives. *BMC Fam Pract* 2011;12:5.

106 Wideroff L, Vadaparampil ST, Greene MH, *et al*. Hereditary breast/ovarian and colorectal cancer genetics knowledge in a national sample of US physicians. *Am J Med Genet* 2005;42:749–55.

107 Fiks AG. Designing computerized decision support that works for clinicians and families. *Curr Probl Pediatr Adolesc Health Care* 2011;41:60–88.

108 Garg AX, Adhikari NK, McDonald H, *et al*. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005;293:1223–38.

109 Emery J, Morris H, Goodchild R, *et al*. The GRAIDS Trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. *Br J Cancer* 2007;97:486–93.

110 Acheson LS, Stange KC, Zyzanski S. Clinical genetics issues encountered by family physicians. *Genet Med* 2005;7:501–8.

111 Gigerenzer G. Collective statistical illiteracy: a cross-cultural comparison with probabilistic national samples: comment on "Statistical Numeracy for Health". *Arch Intern Med* 2010;170:468.

112 Garfield J, Ben-Zvi D. How students learn statistics revisited: a current review of research on teaching and learning statistics. *Int Stat Rev* 2007;75:372–96.

113 Kuehn BM. NIH shifts focus from sequencing genes to fostering clinical applications. *JAMA* 2012;307:132.

114 Schrijver I, Aziz N, Farkas DH, *et al*. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the Association for Molecular Pathology. *J Mol Diagn* 2012;14:525–40.

115 Masys DR, Jarvik GP, Abernethy NF, *et al*. Technical desiderata for the integration of genomic data into Electronic Health Records. *J Biomed Inform* 2012;45:419–22.

116 Health Level 7 International—Clinical Genomics. [28 March 2013]. http://www.hl7.org/Special/committees/clingenomics/overview.cfm.

117 HL7 Version 2 Implementation Guide: Clinical Genomics; Fully LOINC-Qualified Genetic Variation Model (US Realm). [1 September 2012].

118 Melton GB, Raman N, Chen ES, *et al*. 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;17:337–40.

119 Reese MG, Moore B, Batchelor C, *et al*. A standard variation file format for human genome sequences. *Genome Biol* 2010;11:R88.

120 Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the Electronic Health Record for personalized medicine. *Hum Mutat* 2011;32:512–16.

121 Shirts BH, Parker LS. Changing interpretations, stable genes: responsibilities of patients, professionals, and policy makers in the clinical interpretation of complex genetic information. *Genet Med* 2008;10:778–83.

122 Aronson SJ, Clark EH, Babb LJ, *et al*. The GeneInsight Suite: a platform to support laboratory and provider use of DNA-based genetic testing. *Hum Mutat* 2011;32:532–6.

123 Green M, Solnit AJ. Reactions to the threatened loss of a child: a vulnerable child syndrome. Pediatric management of the dying child, Part III. *Pediatrics* 1964;34:58–66.

124 McGuire AL, Cho MK, McGuire SE, *et al*. Medicine. The future of personal genomics. *Science (New York, NY)* 2007;317:1687.

125 Jonsen AR, Durfy SJ, Burke W, *et al*. The advent of the 'unpatients'. *Nat Med* 1996;2:622–4.

126 Pharmacogenomics of Anticancer Agents Research in Children (PAAR4Kids). [10 August 2010]. http://pharmacogenetics.org/paar4kids/research.html.

127 Klitzman R. Exclusion of genetic information from the medical record: ethical and medical dilemmas. *JAMA* 2010;304:1120–1.

128 Qiao D, Yip WK, Lange C. Handling the data management needs of high-throughput sequencing data: SpeedGene, a compression algorithm for the efficient storage of genetic data. *BMC bioinformatics* 2012;13:100.

129 Vey G. Differential direct coding: a compression algorithm for nucleotide sequence data. *Database: J Biol databases curation* 2009;2009:bap013.

130 Schadt EE, Linderman MD, Sorenson J, *et al*. Computational solutions to large-scale data management and analysis. *Nat Rev Genet* 2010;11:647–57.

131 Stein LD. The case for cloud computing in genome informatics. *Genome Biol* 2010;11:207.

132 Baker M. Next-generation sequencing: adjusting to data overload. *Nat Methods* 2010;7:495–9.

133 Giniat EJ. Cloud computing: innovating the business of health care. *Healthc Financ Manage* 2011;65:130–1.

134 Dudley JT, Pouliot Y, Chen R, *et al*. Translational bioinformatics in the cloud: an affordable alternative. *Genome Med* 2010;2:51.

135 Klein CA. Cloudy confidentiality: clinical and legal implications of cloud computing in health care. *J Am Acad Psychiatry Law* 2011;39:571–8.

136 Schweitzer EJ. Reconciliation of the cloud computing model with US federal electronic health record regulations. *J Am Med Inform Assoc* 2012;19:161–5.

137 Taylor RC. An overview of the Hadoop/MapReduce/HBase framework and its current applications in bioinformatics. *BMC Bioinformatics* 2010;11(Suppl 12):S1.

138 Qiu J, Ekanayake J, Gunarathne T, *et al*. Hybrid cloud and cluster computing paradigms for life science applications. *BMC Bioinformatics* 2010;11:S3.

139 Baker M. Genome interpreter vies for place in clinical market. *Nature* 2012;490:157.

140 Arnason V. Coding and consent: moral challenges of the database project in Iceland. *Bioethics* 2004;18:27–49.

141 Clayton EW, Smith M, Fullerton SM, *et al*. Confronting real time ethical, legal, and social issues in the Electronic Medical Records and Genomics (eMERGE) Consortium. *Genet Med* 2010;12:616–20.

142 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;21:1001–7.

143 Malin BA. An evaluation of the current state of genomic data privacy protection technology and a roadmap for the future. *J Am Med Inform Assoc* 2005;12:28–34.

144 Angrist M. Genetic privacy needs a more nuanced approach. *Nature* 2013;494:7.

145 Lowrance WW, Collins FS. Ethics, Identifiability in genomic research. *Science (New York, NY)* 2007;317:600–2.

146 Gymrek M, McGuire AL, Golan D, *et al*. Identifying Personal Genomes by Surname Inference. *Science (New York, NY)* 2013;339:321–4.

147 Presidential Commission for the Study of Bioethical: Privacy and Progress in Whole Genome Sequencing. 2012 [14 October 2012]. http://www.bioethics,gov/cms/sites/default/files/PrivacyProgress508.pdf.

148 Bloomrosen M, Detmer D. Advancing the framework: use of health data – a report of a working conference of the American Medical Informatics Association. *J Am Med Inform Assoc* 2008;15:715–22.

149 Saltz J, Oster S, Hastings S, *et al*. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics* 2006;22:1910–16.

150 Murphy SN, Mendis M, Hackett K, *et al*. Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside. *AMIA Annu Symp Proc* 2007:548–52.

151 Lowe HJ, Ferris TA, Hernandez PM, *et al*. STRIDE–An integrated standards-based translational research informatics platform. *AMIA Annu Symp Proc* 2009;2009:391–5.

152 Kamal J, Liu J, Ostrander M, *et al*. Information warehouse—a comprehensive informatics platform for business, clinical, and research applications. *AMIA Annu Symp Proc* 2010;2010:452–6.

153 Louie B, Mork P, Martin-Sanchez F, *et al*. Data integration and genomic medicine. *J Biomed Inform* 2007;40:5–16.