Urologic Oncology: Seminars and Original Investigations 32 (2014) 198–201

UROLOGIC ONCOLOGY

Seminar article

"Use it or lose it" as an alternative approach to protect genetic privacy in personalized medicine

Jennifer K. Wagner, J.D., Ph.D.[a], Jessica T. Mozersky, Ph.D.[a], Reed E. Pyeritz, M.D., Ph.D.[a,*]

[a] Center for the Integration of Genetic Healthcare Technologies (CIGHT), University of Pennsylvania, Philadelphia, PA

Keywords: Privacy; Genetic information; Whole genome sequencing; Precision medicine; ELSI

Scholars have spent considerable time discussing the challenges and importance of protecting privacy in the context of genomic medicine [see, e.g., 1]. There is a great deal of research and clinical potential to be gained from storing massive amounts of genotypic data, but this must be weighed against the possible risks to individual privacy, a quandary some have described as "privacy vs. the goldmine" [2]. DNA is a unique identifier and has the potential to reveal medical (and nonmedical) information about individuals and by consequence, their family members. For instance, genomic data could reveal undesired presymptomatic health information, nonpaternity, or even be used to frame a suspect at a crime scene [3]. Access to, or disclosure of, this information, authorized or not, could lead to various forms of misuse and discrimination (e.g., employment, insurance, and financial). The advent of whole-genome sequencing (WGS) has compounded privacy concerns because it could reveal entirely unanticipated information, particularly as our ability to (re)interpret genomic sequence data is continually improving [1; see also 4–6].

In addition, genomic information could have consequences beyond the individual from whom it was derived, including both lineal relatives (e.g., children and grandchildren) and collateral relatives (e.g., siblings, cousins, nieces, and nephews) who may be unaware that the individual is undergoing sequencing. Intrafamilial privacy issues are complex and vary from 1 family to another; however, in the clinical context, physicians typically give precedence to individual patient privacy and autonomy over the interests of a patient's relatives. For example, a patient who is found to carry a mutation in the *BRCA1* or *2* genes, which increases the risk of breast or ovarian cancer or both, may not want this information revealed to family members for a variety of reasons whereas another patient receiving the same may choose to share this with all of his/her family members, only specific family members he/she worries may be at heightened risk, or even complete strangers to advance biomedical research [see, e.g., 7]. A patient's care might be improved if the health care provider was able to ascertain genetic information of the relatives as well as that limited information on family medical history known to and shared by a specific patient; however, the idea of linking genomic records between family members or creating joint accounts containing certain genetic data to which multiple patients' records have access has not yet garnered strong support.

A frequent, if not ubiquitous, assumption in these discussions has been that the integration of WGS (In this article we refer to WGS although we acknowledge that whole-exome sequencing is currently often being used. We anticipate that as costs continue to drop whole-genome sequencing may become the norm and replace whole-exome sequencing entirely.) in health care would consist of a one-time sequencing of the patient's genome and subsequent storage of the WGS data in the patient's electronic health record (EHR) [e.g., 8–10]. These WGS data could then be interpreted and reinterpreted by health care providers over the patient's lifetime (or collaboratively "managed" by patients and clinicians [11]) as knowledge about the clinical relevance of genomic variants increases. One of the biggest clinical challenges to interpretation of an individual's genome currently is the uncertain significance of many DNA variants, although this will likely improve over time. This ongoing need for reinterpretation of an individual's genome has raised additional clinical concerns such as who is responsible for recontacting patients when and if new information emerges and how patients can be expected to give informed consent when the potential implications of WGS are unknown currently [12]. With such access to WGS data alongside scientific and technological capabilities for health care providers to mine those

* Corresponding author. Tel.: +1-215-614-0933; fax: +1-215-573-8606.
E-mail address: reed.pyeritz@uphs.upenn.edu (R.E. Pyeritz).

1078-1439/$ – see front matter
http://dx.doi.org/10.1016/j.urolonc.2013.09.016

WGS data (sometimes for information that exceeds the scope of patient expectations or immediate health concerns), policy discussions have focused on risk management of "the incidentalome" [e.g., 13-15]. Discovering multiple abnormal incidental findings, including the dreaded variants of unknown significance, could place many undue burdens on clinicians and patients alike.

Discussions that center on incorporating all genomic data into the EHR have the effect of medicalizing the genome [see also 16] by assuming all genomic information is relevant for determining medical risk when some portions of the genome have no known medical relevance whatsoever despite being useful for non–health care purposes (such as ancestral or forensic information). Here, in contemplating patient privacy in personalized medicine, we question the medicalization of genomes in a broad sense, not only questioning the restriction of an individual's access to genomic information by requiring such data to be obtained only through a health care provider but also questioning the *a priori* medical relevance of all genomic sequence data. Although genomic data can, and do, provide clinically relevant information, the entire genomic sequence would not be relevant or necessary in most contexts. Each locus in the genome has its own evolutionary story and an anthropological (not just medical) genetic perspective is necessary. Some specific loci in the genome may be medically relevant for some individuals, in some contexts, and during some stages of development whereas they might be irrelevant for other individuals, in other contexts, or during other stages of development. Assessing the medical relevance of genomic data is limited by our present understanding of normal human genomic variation, reporting biases of positive results in the literature, and the underrepresentation of genomic research involving individuals from racial and ethnic minorities. Thus, the determination of medical relevance of genomic data is appropriately a locus-by-locus, patient-by-patient, visit-by-visit, case-by-case decision. We further question the assumption that WGS data should be incorporated into the EHR without deliberate consideration given to privacy and standard-of-care problems that may accompany the "hoarding" of genomic data in clinical systems. We emphasize that our suggestion relates to the storage of clinical data as opposed to research data where storage of the entire genome may indeed be necessary.

Proposing a "use it or lose it" strategy for WGS data in health care

The cost to generate a human WGS has fallen rapidly in the last few years [17], and the coveted $1,000 genome may soon be within arm's reach. Clinical applications of WGS are no longer cost prohibitive [18]. Moreover, WGS can be done in roughly a day, with the speed of sequencing now 50,000 times faster than it was in 2000 [19]. WGS costs are descending to levels comparable to diagnostic genetic tests [see, e.g., 20]. Clinical uptake of WGS is expected to intensify as the cost of generating a patient's WGS falls, while the speed of generating and analyzing the WGS increases. Yet privacy concerns are unresolved; that WGS data are themselves identifiers and can be associated with publicly available information relatively easily is gaining broader recognition [21]. At what point does the cost of storage of WGS data locally or on the cloud—and the associated costs (economical, logistical, structural, and otherwise) of properly maintaining data security and protecting the privacy of health information in compliance with the Health Insurance Portability and Accountability Act (HIPAA), Health Information Technology for Economic and Clinical Health (HITECH), and Genetic Information Nondiscrimination Act (GINA) final rules [22; see also 23–24]—itself become an avoidable act of "raiding the medical commons" by straining already limited health care resources [25]?

Quite simply, the mass default storage of WGS data generated in clinical settings may expose patients, and their family members, to unnecessary privacy risks. Various solutions have been proposed ranging from closed or controlled-access databases [e.g., 26] to entirely publicly accessible databases that provide no privacy protection (such as the Personal Genome Project) [2]. An alternative approach may be to treat WGS as any other diagnostic test, not ordered once per lifetime of a patient (as is often envisioned today) but, rather, ordered as needed for specific medical purposes and to guide clinical intervention. Upon generation of the WGS data ordered for a specific purpose, the raw data from the WGS used as a basis for the medical decisions and course of treatment could be incorporated into the EHR along with the record of the diagnosis and course of treatment. The rest of the WGS data would be discarded. This proposed "use it or lose it" policy has several advantages that could help mitigate the currently contentious issues regarding "privacy vs. the gold-mine" [2]. Our "use it or lose it" strategy would limit the genomic privacy risks of the patient by retaining only the genomic information necessary for the patient's current management. Moreover, such a policy would reduce the "incidentalome" and associated findings such as variants of unknown significance or unwanted knowledge about presymptomatic disease or reproductive risk. This would also avoid the potential duty to recontact patients about an unlimited number of conditions with which their genomic data may become associated in the future. In trying to grapple with these issues, the American College of Medical Genetics and Genomics recently published policy recommendations that 57 particular variants, genes, and conditions be reported as incidental findings regardless of the purpose for which clinical sequencing was initially carried out [14]. This American College of Medical Genetics and Genomics recommendation has generated controversy, leading some to suggest that such a policy violates patient autonomy by denying individuals the right to opt out of receiving such information [27]. In contrast, our suggested "use it or lose it" policy