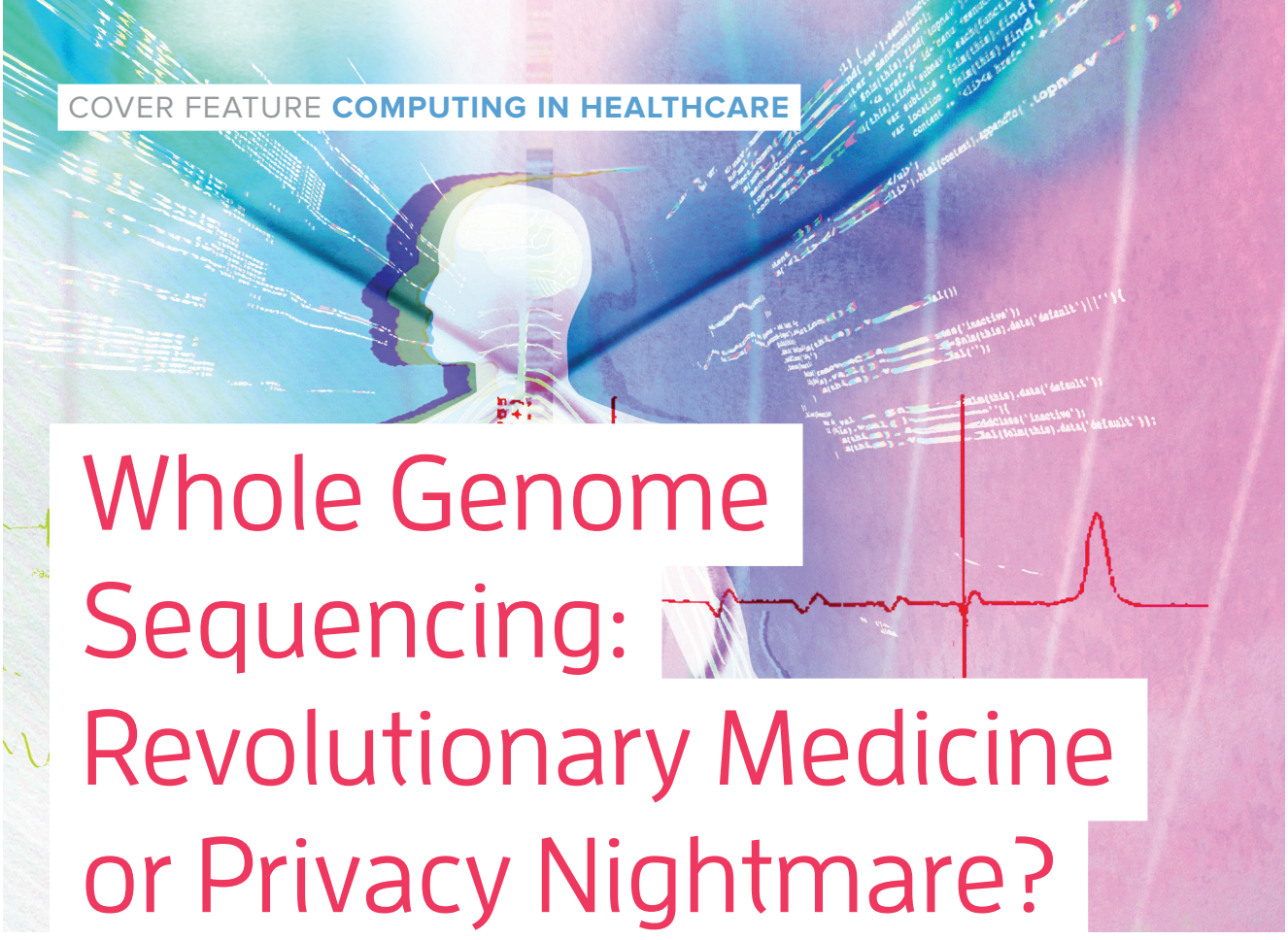

Whole Genome Sequencing: Revolutionary Medicine or Privacy Nightmare?

Erman Ayday, Bilkent University
Emiliano De Cristofaro, University College London
Jean-Pierre Hubaux, EPFL, Lausanne
Gene Tsudik, University of California, Irvine

Whole genome sequencing will soon become affordable for many individuals, but thorny privacy and ethical issues could jeopardize its popularity and thwart the large-scale adoption of genomics in healthcare and slow potential medical advances.

In the past decade, whole genome sequencing (WGS) has evolved from a futuristic concept to a realistic technology that yields an individual's complete genome. Each genomic sequence contains a vast amount of information that enables significant progress in understanding, treating, and preventing disease. As such, WGS has the potential to revolutionize healthcare.

However, a genome also contains highly sensitive information that uniquely identifies an individual. When technology advances eventually make WGS affordable for the general population, individuals will need assurances about access to their genomic information. For example, who will store the digitized genome and where? How will access be controlled such that no one can inadvertently or deliberately leak genomic information to third parties? What will keep a healthcare provider's service partners from using genomic information in ways other than medical research or personalized medical treatment?

With DNA sequencing cost dropping below $1,000 per genome, these questions have become pressing. Both throughput gains and the cost reductions of new-generation sequencing platforms have defied Moore's

See **www.computer.org/computer-multimedia** for multimedia content related to this article.

0018-9162/15/$31.00 © 2015 IEEE

ONGOING WORK TO PROTECT GENOMIC DATA

Over the past few years, research in genomic privacy has accelerated and now falls into four main categories:

- » string searching and comparison,
- » release of aggregate data,
- » alignment of raw genomic data, and
- » clinical use of genomic data, such as for personalized medicine.

Work in the first category is experimenting with the use of medical tools and private string comparison for privacy-preserving paternity tests, personalized medicine, and genetic compatibility tests.[1] More recently, researchers have extended that work to implement the GenoDroid toolkit.[2] which provides paternity and ancestry testing via a smartphone.

In the second category, researchers are focusing on privacy risks of releasing aggregate genomic data.[3] Others have explored the application of *differential privacy* to the publication of aggregate genomic trial statistics.[4,5] Their work aims to ensure that two genomic databases, which differ only by one individual's data, have indistinguishable statistical features. Hence, the published result from a genomic dataset does not reveal the existence of a particular individual in that dataset.

Research in the third category is looking at secure and efficient algorithms for read mapping (aligning millions of short sequences to a reference DNA sequence). One recent attempt on this direction works in a hybrid (public and private) cloud environment.[6] In this work, authors outsource the computationally intensive steps of the operation to a public (untrusted or commercial) cloud; they propose doing sensitive and lightweight computations on a private (trusted) cloud to protect the privacy of sensitive DNA information.

In the last category is work to preserve the patient's privacy in medical tests and personalized medicine. One approach uses homomorphic encryption and secure multiparty computation to protect patients' genomic data in this context.[7,8]

Some of these efforts have already materialized into practical genomic testing. However, it is hard to foresee the range and complexity of future genetic operations: some tests might be too computationally intricate to be performed on a personal device, or genetic tests might involve multiple genomes. Consequently, we expect the scope and nature of genomic data protection work to change as researchers make new discoveries and shift their focus to address a new set of needs. At the same time, the efforts already in progress are important stepping stones to solutions that address the multifaceted challenge of protecting genomic data.

References

1. P. Baldi et al., "Countering GATTACA: Efficient and Secure Testing of Fully-Sequenced Human Genomes," *Proc. 18th ACM Conf. Computer and Communications Security* (CCS 11), 2011, pp. 691–702.
2. E. De Cristofaro et al., "Genodroid: Are Privacy-Preserving Genomic Tests Ready for Prime Time?" *Proc. ACM Workshop Privacy in the Electronic Society* (WPES 12), 2012, pp. 97–108.
3. X. Zhou et al., "To Release or Not to Release: Evaluating Information Leaks in Aggregate Human-Genome Data," *Proc. 16th European Conf. Research in Computer Security* (ESORICS 11), 2011, pp. 607–627.
4. F. Yu et al., "Scalable Privacy-Preserving Data Sharing Methodology for Genome-Wide Association Studies," *J. Biomedical Informatics*, Feb. 2014, pp. 133–141.
5. A. Johnson and V. Shmatikov, "Privacy-Preserving Data Exploration in Genome-Wide Association Studies," *Proc. 19th ACM Int'l Conf. Knowledge Discovery and Data Mining*, 2013, pp. 1079–1087.
6. Y. Chen et al., "Large-Scale Privacy-Preserving Mapping of Human Genomic Sequences on Hybrid Clouds," *Proc. 19th Network and Distributed System Security Symp.* (NDSS 12), 2012; www.informatics.indiana.edu/xw7/papers/ndss2012.pdf.
7. E. Ayday et al., "Privacy-Preserving Computation of Disease Risk by Using Genomic, Clinical, and Environmental Data," *Proc. Usenix Security Workshop Health Information Technologies* (HealthTech 13), 2013; www.usenix.org/conference/healthtech13/workshop-program/presentation/ayday.
8. E. Ayday et al., "Protecting and Evaluating Genomic Privacy in Medical Tests and Personalized Medicine," *Proc. ACM Workshop Privacy in the Electronic Society* (WPES 13), 2013, pp. 95–106.

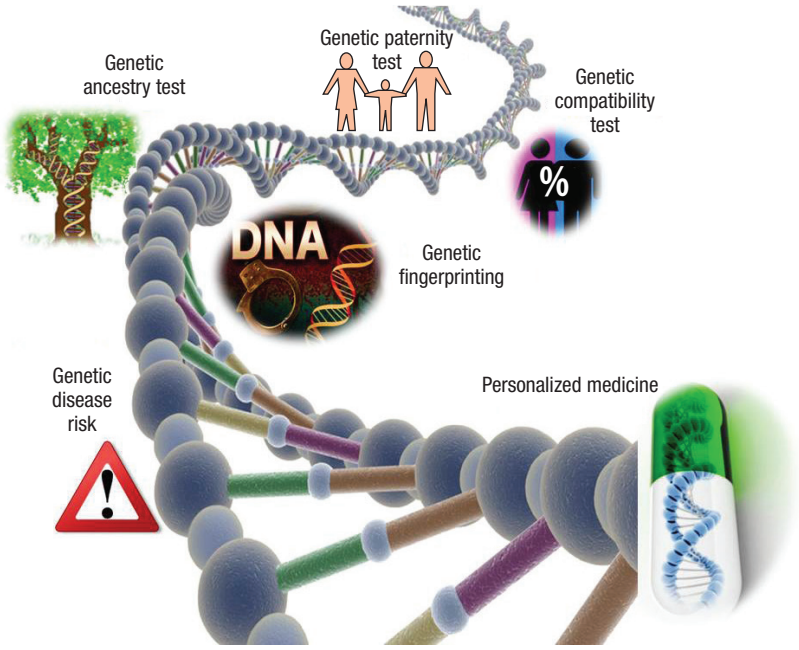

FIGURE 1. Genomics applications. Whole genome sequencing will enable personalized genomic medicine and facilitate testing for genetic disease risk and ancestry.

law. Thus, it is safe to assume that, in a few years, most individuals in developed countries will be able to obtain their digitized genomes for any number of purposes—from personalized medicine to paternity testing. Commercial entities, such as Knome and Illumina, already offer services that create reports from raw genomic data, which doctors use to guide treatment.

However, without a deeper understanding of the complex interplay between genomes and healthcare, WGS applications will be limited. Achieving progress in this research will require patients (or volunteers) who are willing to share their genetic data—an agreement that raises privacy protection, ethical use, and legal rights concerns. For example, in the Personal Genome Project (www.personalgenomes.org), participants agree to make their genomic data and other personal information publicly available on the Internet. Such pilot projects offer a glimpse into the future concerns of handling large-scale genomic data.

DNA sequencing greatly exacerbates data exposure and exploitation issues that social media and personal health records (PHRs) have already brought to the forefront. The genome represents an individual's biological identity and thus contains rich information about that person's ancestry. By combining the genomic data with data on the person's environment or lifestyle, a third party can infer the individual's phenotype, including predisposition to physical and mental health conditions (such as Alzheimer's disease, cancer, or schizophrenia).

If a genomic information leak occurs, revoking or replacing an individual's DNA sequence is impossible, which has serious implications for applications that depend on accurate genomic information. The use of DNA analysis in law enforcement and healthcare, for example, is already prompting ethical questions, such as how to guarantee the genomic information's integrity.

Until researchers address these open problems, the much anticipated benefits of personalized medicine could remain on hold.

GENOMICS 101

The human genome is encoded in double-stranded DNA molecules that consist of two complementary polymer chains. Each chain is a series of nucleotides, represented as the letters A, C, G, and T. Technicians collect DNA samples from a person's saliva, hair, skin, or blood, among other sources, and extract genetic material for sequencing. The resulting genome is a unique string of approximately 3.2 billion letter pairs (an arrangement of A, C, G, and T).

The reference genome, which scientists have assembled as a representation of the human genome, makes up 99.5 percent of a human's DNA sequence. The remaining 0.5 percent represents the individual's genetic variation. Although it might seem insignificant relative to the reference genome, this minuscule 0.5 percent corresponds to several million nucleotides.

The genetic variation can take several forms, the most common being single nucleotide polymorphism (SNP, pronounced "snip"). In simplest terms, a SNP is a position in the genome sequence with a nucleotide that varies between individuals. For example, in two sequenced DNA fragments from different individuals, AAGCCTA and AAGCTTA, the fifth nucleotide is C in one and T in the other.

Researchers have confirmed that humans have approximately 50 million unique SNPs,[1] a number that becomes more exact as more individuals consent to sequencing.

SNPs can help determine an individual's predisposition to certain disorders or diseases. For example, recent genome-wide association studies show that the presence of three genes with 10 particular SNPs can indicate susceptibility to Alzheimer's disease.[2,3]

Interdependent SNPs sometimes result in linkage disequilibrium (LD)[4]—the nonrandom association of alleles at two or more loci. The alleles descend

from single, ancestral chromosomes, so LD makes it possible to infer the nucleotide of a SNP from the contents of other SNPs. This relationship obviously complicates privacy protection.

PERSONALIZED MEDICINE AND BEYOND

WGS has the potential to bring about a new era of predictive, preventive, participatory, and personalized (P4) medicine[5] and enable applications such as those in Figure 1. P4 represents a significant healthcare paradigm shift[6] from the current trial-and-error treatment because it enables medication tailored to a patient's precise genetic makeup. P4 applications include assessments of disease and treatment risk, and paternity and ancestry testing, and the evaluation of genetic compatibility between potential partners to reduce the possibility of passing genetic diseases to their offspring.

Pharmacogenomics

Experiments have shown that certain genetic mutations alter drug metabolism and that genomic tests can help predict a patient's response to particular drugs. This experimentation and testing is part of *pharmacogenomics*—the study of how genetic variations affect an individual's response to medications. Examples of pharmacogenomics include testing for SNP mutations in the tpmt gene of children with leukemia and pretreatment testing for the correlation of the BRCA1/BRCA2 genes to familial breast and ovarian cancer syndromes.

Genomic tests to determine drug response are expected to become more widespread in the near future. Experts estimate that about a third of the 900 cancer drugs now in clinical trials could soon come to market with an enclosed recommendation for a DNA or another molecular test.[7]

Programs are underway to support pharmacogenomics. For example, Vanderbilt University's Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment (Predict) program[8] evaluates patients' genetic characteristics to help physicians determine which drugs are most likely to work, thus avoiding the long trial-and-error period characteristic of traditional drug evaluation. In one case,[9] Predict program researchers used the genetic profile of a patient with coronary artery disease to help doctors select a specific cholesterol-lowering drug and successfully treat the patient in a fraction of the time with a conventional approach.

Testing for genetic disease risk

Low-cost WGS will give individuals direct access to their genomic information, which they could share with sites that test for genetic disease risks. One such site, 23andMe, already provides relatively low-cost genetic ancestry and disease risk tests for 960,000 specific SNPs, although it does not yet offer WGS. Since November 2013, the US authorities have suspended the health-related 23andMe tests, pending FDA investigation; however, such tests are still offered in the UK.

In parallel to direct-to-consumer services, national and regional efforts are attempting to introduce genomics into the clinical setting. Examples include the UK's 100,000 Genomes Project (www.genomicsengland.co.uk) and University Hospital Lausanne's biobank (www.chuv.ch/biobanque/bil_home/bil-patients-famille/bil-la_bil.htm).

Although researchers are enthusiastically exploring the relationship of genetics and personalized medicine, biomedical experts have expressed doubts about the extent to which gene mapping can predict the likelihood of developing a disease.[10] They argue that, although scientists have a list of genetic features that correlate to certain diseases,[2] they do not know whether (and to what extent) environmental factors also come into play.

Paternity and ancestry testing

The availability of a patient's fully sequenced genome will enable clinicians, doctors, and testing facilities to run complex, correlated genetic tests in a matter of seconds. Compared with the more expensive in vitro tests, these specialized computational algorithms enable faster and more accurate testing while preserving legal acceptance.

Commercial entities already offer ancestry and genealogical testing in which software compares an individual's genomic information with publicly available genomic data from a particular ethnic group to determine how the individual relates to the group. Online services also offer genetic compatibility tests that assess the risk of Mendelian inheritance[11]—the chance of transmitting genetic diseases to any offspring—in the couple being tested.

THREATS TO GENOMIC DATA PRIVACY

Many view genomic privacy with skepticism, since every individual constantly leaves behind biological material, such as hair, skin, or saliva—evidence that a third party can collect even days later and use to construct a DNA sequence. However, this threat is credible only for a targeted individual or a small group, not for a large

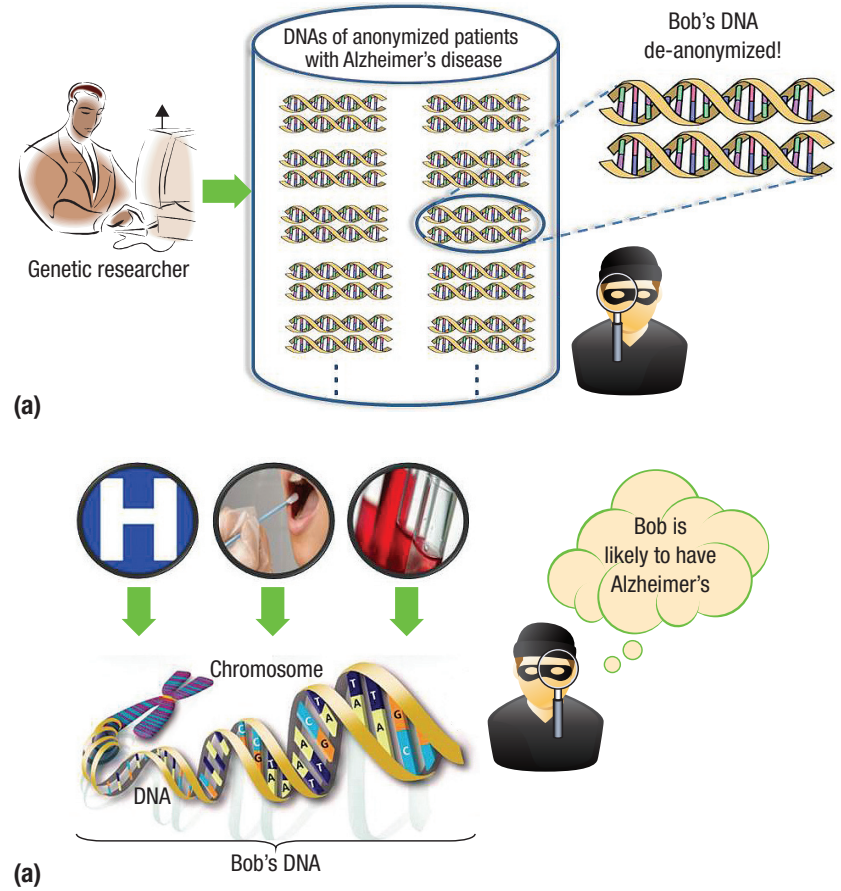

FIGURE 2. Two main threats to human genomic data privacy. (a) DNA donors in a public research database lose anonymity (de-anonymization), and (b) partial genomic data leakage allows outsiders to infer sensitive information. Figure used with permission from the US Department of Energy Genomic Science program (https://public.ornl.gov/site/gallery/detail.cfm?id=398&topic=&citation=&general=dna&restsection=all).

number of digitized genomes, such as in a research database.

Genomes in the latter setting face two main threats, as Figure 2 illustrates. Although existing laws protect data privacy in general, genomic data has certain characteristics that require more restrictive provisions to address unique privacy threats.[12]

Loss of donor anonymity

The primary traditional approaches to privacy protection are data de-identification or aggregation. Common de-identification strategies, which include deleting or masking identifiers, such as names and Social Security numbers, are ineffective for genomic data because the genome is the ultimate identifier.[13]

Aggregation—a strategy that combines data for a population—is also ineffective because enough published information is available to identify the individual from a case study and, in some instances, to recover parts of the genome sequence. For example, a 2009 study[14] shows that even the test statistics (such as *p*-values, *r*-squares) calculated from allele frequencies and published papers give away enough information to identify genetic trial participants. A 2013 study[15] demonstrated that third parties can use information from popular genealogy websites along with other available personal data to re-identify (counter de-identification of) DNA donors from a public research database.

Data leaks

Because the genomes of two closely related individuals are highly similar, the disclosure of a person's genome can possibly leak significant genomic information about that person's close relatives. This disclosure is a problem regardless of whether it was voluntary, accidental, or malicious.

The possibility of revealing others' identities makes genomic data privacy a unique issue, since, in most other sensitive scenarios, only the individual's data is at stake. Depending on the number of siblings and children, disclosure can affect a large group.[16] Failing to consider this possibility can have severe consequences, as the recent controversy about Henrietta Lacks' genome sequence attests. In researching Lacks' disease nearly five decades ago, scientists discovered cell properties in her cancerous tissue that made the cells highly suitable for biogenetic research. They harvested more cells without the family's knowledge and began using the HeLa cell (in honor of Lacks' first and last names) in studies. It eventually became so popular in genetics research that Lacks' surviving family members began receiving requests for tissue and blood samples. After several court cases to address privacy violations, in 2013, the National Institutes of Health (NIH) agreed to give the family some control over the HeLa cells' use.

Exacerbating the data leak problem is the genome's immutability and longevity. An individual can change passwords, account numbers, and even public key certificates. The same is not true of a genome. Moreover, future generations will inherit most of their ancestor's DNA, so genomic information disclosure can become an endless curse.

PRIVACY PROTECTION LAWS

Clearly, privacy concerns represent a formidable obstacle to assembling large human genomic databases and can delay (or derail) genome-wide association studies, which in turn could thwart advances in medicine and subsequent healthcare improvements. In law enforcement, which increasingly uses DNA-based identification, the need for genomic data security and reliability is also evident.

Existing laws protect genomic data privacy to some degree. In 1990, the National Human Genome Research Institute established the Ethical, Legal, and Social Implications Research Program to explore the repercussions of advances in genetic and genomic research on individuals, families, and communities. In 2008, the US government established the Genetic Information Nondiscrimination Act (GINA), which prohibits health insurance and employment discrimination on the basis of genetic information. Also, the Health Insurance Portability and Accountability Act (HIPAA) provides a general framework for protecting and sharing health information, and the State of California has begun to consider DNA privacy laws.[17] Meanwhile, in Europe, legislators are taking similar precautions.[18]

Discrimination through genetic data is not a new idea. As far back as 1997, *Gattaca*, a popular science fiction movie, touched on the notion of genism—the theory that genes determine distinctive human characteristics and abilities—and explored the idea that genetic discrimination could be as pernicious as overt racism.

THE CASE FOR STRICTER POLICY

Although current legislation provides guidelines for genomic data use, it does not contain enough technical information about safe and secure ways to store and process digitized genomes. One reason is that security and privacy issues for genomic data—both individual genomes and the genome collections in genomic databases—are not well understood.

Privacy practitioners and consumer organizations are strongly advocating the need for more restrictive legislation to close current policy gaps. A recent report from the US Presidential Commission for the Study of Bioethical Issues[19] analyzed WGS advances, highlighted growing privacy and security concerns, and made a few privacy and security recommendations.

We believe these recommendations reflect a general lack of understanding about the associated open technical problems. For example, one recommendation was to use de-identification, which is clearly unsuitable. The recommendations also fail to address several important points. For example, to guard against surreptitious DNA testing, any genomic data protection policy must recognize the need for informed consent. The policy should set forth procedures for authorities and companies to obtain written permission from an individual before collecting, analyzing, storing, or sharing that person's genetic information, such as hair or saliva samples—thus ensuring that no individual will be a victim of unauthorized sequencing.

A measure such as this will not be popular with those who view privacy-friendly measures as hindrances to genomic research. Scientists typically sequence DNA from large groups to determine genes associated with particular diseases. The informed consent restriction would mean that they cannot reuse large genomic datasets to study a different disease. Rather, they would have to destroy the data after each study or track down all previously enrolled study participants and secure a new authorization from each for the next study. Also, because related individuals have similar genomes, the participant's relatives might have to give consent as well.

GUIDELINES FOR GENOMIC DATA PROTECTION AND USE

The individual who requests and likely pays for genome sequencing should own the result, as is already the case for any other personal medical information. However, genomes are a new kind of personal health information, which raises numerous issues that technical approaches alone cannot address. Rather, technology must work with legal and professional guidelines that govern how to transmit, store, process, and eventually dispose of genomic information.

Storage and long-term protection

Storing and protecting the genome raises several important questions:

- Should the genome be stored on the individual's personal device? What special hardware security features are needed to prevent tampering?
- Should genome storage be outsourced to a cloud provider?
- Should the genome be encrypted? If so, what organization will generate and store the encryption keys?

Although encryption might seem the ideal answer to many of these questions, it has drawbacks. Encryption schemes that many consider strong at present might gradually weaken,

but the genome's sensitivity will not. Thus, a third party that cannot decrypt an encrypted genome might be able to do so years later. The Advanced Encryption Standard (AES) scheme supports key lengths up to 256 bits. Although several standardization bodies and intelligence agencies believe it will be secure for several decades,[20] computational breakthroughs or unforeseen weaknesses might allow early decryption.

One option is to periodically reencrypt the genome, assuming it cannot be copied. Another option is to use secret-sharing techniques to split the genome and partition it among several providers. However, efficient reassembly is problematic, as is the guarantee that providers do not collude in genome reconstruction. Moreover, the providers themselves must have sufficient longevity.

Finally, encryption will not prevent leaks of a long-deceased individual's genomic data, which can affect the privacy of that person's living progeny.

> ENCRYPTION SCHEMES THAT MANY CONSIDER STRONG AT PRESENT MIGHT GRADUALLY WEAKEN, BUT THE GENOME'S SENSITIVITY WILL NOT.

Accessibility

Given the genome's sensitivity, an individual should never disclose any genomic information, which would certainly prevent access to any genomic application except within the individual's secured personal device. Although it sounds ideal, such a restriction might be possible if operations were represented in some standardized form that some trusted agency has certified. For example, if testing for a genetic disease requires matching a well-known pattern in some approximate location in the genome, the US Food and Drug Administration (FDA) might certify that pattern and its parameters. Individuals would then be assured that the operation is a legitimate test for a specific genetic disease and that they will receive the results, which they then can opt to keep private.

Other questions about accessibility are more complicated:

- Should the sequencing facility keep an escrowed copy of the genome?
- Should the individual entrust a genome copy to his personal physician or health insurance provider?
- Is it possible to guarantee the digitized genome's integrity and authenticity? If so, how?
- If backups are made, how often and where should they be kept?
- Is it possible to securely erase a genome?
- Should individuals periodically request a new genome sequence to keep pace with more accurate technology?

Testing guidelines

To effectively replace their in vitro counterparts, computational genomic tests must be accurate, efficient, and usable for individuals who are not geneticists.

Accuracy. A computational genomic test should guarantee accuracy that is at least equivalent to the in vitro test. For example, a computational paternity test should provide the same confidence as the in vitro test, which is currently admissible in a court of law. Computational tests should also strive for accountability by furnishing guarantees of correctness for both execution and input information.

Efficiency. Computational genomic tests should incur minimal communication and computing costs. Patients might be used to waiting several days to obtain genetic test results. However, in a computational setting, long runtimes on personal devices might hinder the test's practicality.

Usability. Computational genomic tests are likely to involve the general population, which raises several usability questions:

- How much should the user know about genomic test aspects?
- What information about the test and results is appropriate, and at what granularity should it be presented?
- Do individual's privacy perceptions and concerns match the scientific community's expectations?

The last question is particularly complex. Some users might be willing

to forego their genomic privacy. For example, the expectation is that patients will reveal their genomes to their doctors so that they can benefit from tests that can possibly save them from a life-threatening disease, such as cancer. However, the same individual might not wish to reveal that information to an online service or pharmaceutical company.

These considerations are for the most part educated guesses, since few efforts have focused on users' concerns. Therefore, one research focus should be on exploratory user studies[21] to elicit insights into this issue and address the open problem of how to effectively communicate the potential privacy risks associated with genomic information and its disclosure.

Affordable, readily available WGS will stimulate thrilling opportunities, but it will also raise privacy concerns; addressing both sides of WGS will require long-term collaboration among geneticists, other healthcare providers, ethicists, lawmakers, and computer scientists. To this end, we helped organize the first multidisciplinary Dagstuhl seminar on genomic privacy, which took place in 2013[22] and will be held again in October 2015. We also helped launch an international workshop on genomic privacy, which took place in 2014 and will be held again in conjunction with the 2015 IEEE Symposium on Security and Privacy (www.genopri.org). Finally, we have set up www.genomeprivacy.org, a site that offers computer scientists tutorials and links to genome privacy research groups.

Long-term collaboration will require targeted funding support. In the US, genomic privacy has fallen into funding gap between agencies. The NIH funding, for example, solidly covers both bioinformatics and WGS ethical issues, but only sparsely supports research on genomic data privacy. The National Science Foundation's (NSF's) Smart and Connected Health program includes integrative projects that require collaboration among computer and health sciences, but the program may or may not engender long-range genomic privacy research.

Other US funding agencies have not, thus far, explicitly addressed genomic privacy. In Europe, numerous EU and nationally funded projects are focusing on e-health, and some consider data protection, but they largely overlook genomic data privacy. In addition, although most officials in charge of data protection typically have a strong legal background, they lack computer science expertise. Consequently and not surprisingly, they tend to rely on legislation more than on technology.

Our work is thus a call for research collaboration to specifically and vigorously address the privacy issues we have identified. Overcoming these obstacles will free WGS to reach its full potential to revolutionize medicine and allow individuals and society overall to reap the considerable benefit.

REFERENCES

1. Nat'l Center for Biotechnology Information, "dbSNP," Dec. 2014; www.ncbi.nlm.nih.gov/projects/SNP.
2. Eupedia, "Genetically Inherited Traits, Conditions, and Diseases," 2014; www.eupedia.com/genetics/medical_dna_test.shtml
3. S. Seshadri et al., "Genome-Wide Analysis of Genetic Loci Associated with Alzheimer Disease," *J. Am. Medical Assoc.*, vol. 303, no. 18, 2010, pp. 1832–1840.
4. D.S. Falconer and T.F. Mackay, *Introduction to Quantitative Genetics*, 4th ed., Addison Wesley, 1996.
5. L. Hood and D. Galas, "P4 Medicine: Personalized, Predictive, Preventive, Participatory: A Change of View That Changes Everything," 2009; www.cra.org/ccc/files/docs/init/P4_Medicine.pdf.
6. A. Weston and L. Hood, "Systems Biology, Proteomics, and the Future of Healthcare: Toward Predictive, Preventive, and Personalized Medicine," *J. Proteome Research*, vol. 3, no. 2, 2004, pp. 179-196.
7. A. Burke, "Foundation Medicine: Personalizing Cancer Drugs," 2012; www.technologyreview.com/featuredstory/426987/foundation-medicine-personalizing-cancer-drugs/.
8. My Drug Genome, "Using Genetics to Personalize Medication Treatment," 2014; www.mydruggenome.org/overview.php.
9. K. Whitney, "PREDICT Helps Pinpoint Right Statin for Patient," *Vanderbilt Univ. Medical Center Report*, 4 Oct. 2012; http://news.vanderbilt.edu/2012/10/predict-helps-pinpoint.
10. G. Naik, "Gene Maps Are No Cure-All," *Wall Street J.*, 3 Apr. 2012; www.wsj.com/articles/SB10001424052702304023504577319604245325644.
11. V. McKusick and S. Antonarakis, *Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders*, John Hopkins Univ. Press, 1994.
12. Y. Erlich and A. Narayanan, "Routes for Breaching and Protecting Genetic Privacy," *Nature Reviews Genetics*, vol. 15, no. 6, 2014, pp. 409–421.
13. N. Homer et al., "Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping

ABOUT THE AUTHORS

ERMAN AYDAY is an assistant professor of computer science at Bilkent University, Ankara, Turkey. While conducting the research reported in this article, he was a postdoctoral researcher in the School of Computer and Communication Sciences at EPFL, Switzerland. His research interests include privacy, genomics, trust and reputation systems, and network security. Ayday received a PhD in electrical and computer engineering from Georgia Institute of Technology. He is a member of IEEE and ACM. Contact him at erman@cs.bilknet.edu.tr.

EMILIANO DE Cristofaro is a senior lecturer (associate professor) at University College London (UCL). While conducting the work reported in this article, he was a research scientist at Xerox's Palo Alto Research Center (PARC). His main research interests are privacy-enhancing technologies and applied cryptography. De Cristofaro received a PhD in networked systems from University of California, Irvine. Contact him at me@emilianodc.com.

JEAN-PIERRE HUBAUX is a professor in the School of Computer and Communication Sciences at EPFL, Switzerland. His research interests include privacy protection, notably in mobile networks and genomics. Hubaux received a DrEng in electrical engineering from Politecnico di Milano. He is a Fellow of IEEE and ACM. Contact him at jean-pierre.hubaux@epfl.ch.

GENE TSUDIK is a Chancellor's Professor of Computer Science at University of California, Irvine. His research interests include security, privacy, and applied cryptography. Tsudik received a PhD in computer science from University of Southern California. He is a Fellow of IEEE and ACM. Contact him at gts@ics.uci.edu.

Microarrays," *PLoS Genetics*, vol. 4, no. 8, 2008, pp. 1–9.
14. R. Wang et al., "Learning Your Identity and Disease from Research Papers: Information Leaks in Genome-Wide Association Study," *Proc. 15th ACM Conf. Computer and Communications Security* (CCS 09), 2009, pp. 534–544.
15. M. Gymrek et al., "Identifying Personal Genomes by Surname Inference," *Science*, vol. 339, no. 6117, 2013, pp. 321–324.
16. M. Humbert et al., "Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy," *Proc. 20th ACM Conf. Computer and Communications Security* (CCS 13), 2013, pp. 1141–1152.
17. H. Shen, "California Considers DNA Privacy Law—Academic Researchers Fear Measures Would Prohibit Work with Genetic Databases," *Nature*, 18 May 2012; www.nature.com/news/california-considers-dna-privacy-law-1.10677.
18. Council of Europe, "Additional Protocol to the Convention on Human Rights and Biomedicine, Concerning Genetic Testing for Health Purposes," 2008; http://conventions.coe.int/Treaty/EN/Treaties/html/203.htm.
19. Presidential Commission for the Study of Bioethical Issues, "Privacy and Progress in Whole Genome Sequencing," 2012; www.bioethics.gov/cms/sites/default/files/PrivacyProgress508.pdf.
20. Nat'l Inst. Standards and Tech., "Cryptographic Key Length Recommendation," 2014; www.keylength.com/en/4.
21. E. De Cristofaro, "An Exploratory Ethnographic Study of Issues and Concerns with Whole Genome Sequencing," *Proc. 8th Network and Distributed System Security Symp. (NDSS) Workshop Usable Security* (USEC 2014), 2014; http://arxiv.org/abs/1306.4962.
22. K. Hamacher, J.-P. Hubaux, and G. Tsudik, "Dagstuhl Seminar on Genomic Privacy," Oct. 2013; www.dagstuhl.de/en/program/calendar/semhp/?semnr=13412.

Selected CS articles and columns are also available for free at **http://ComputingNow.computer.org**.