



# Democratization of genetic data: connecting government approval of clinical tests with data sharing

Theodora S. Ross

Departments of Internal Medicine and Cancer Genetics, UT Southwestern Medical Center, Dallas, Texas 75390, USA

**Abstract** When a doctor orders a genetic test, patients assume that the test will yield a useful result to guide how their physicians take care of them. That assumption is frequently correct, but not always. Until recently, a genetic test only interrogated the sequence of one or two genes. Now, DNA-sequencing technologies are so fast and cheap that they have enabled clinicians to sequence panels of genes that may or may not be relevant to the patient's condition. The technology has outpaced our ability to interpret the results. Connecting approval of clinical tests to data sharing could help close this gap.

## THE PROBLEM: MISINTERPRETATION OF GENETIC TESTS

When a genome or part of a genome is sequenced, the data are filtered to remove the millions of common gene variants; only the rare variants (those that occur at frequencies of <1% in databases such as the Exome Variant Server [<http://evs.gs.washington.edu/EVS/>]) are left for us to interpret (Fig. 1A). Unfortunately, this still leaves us with many rare variants of undetermined clinical significance (VUS) (Fig. 1B). We simply do not understand the meaning of most of the genome, therefore genetic test results reported to patients can be incorrect, confusing, or simply of little use. This applies to germline and tumor genomes. Here, I focus on germline variant test interpretation rather than tumor sequencing tests, but clearly there is a need for accurate variant interpretation and data sharing in both cases.

The case of a 33-year-old breast cancer patient from our cancer genetics clinic, who tested positive for a "deleterious" *CHEK2* mutation predisposing her to colon as well as additional breast cancers, provides a useful example. The patient's younger sister tested negative for the same *CHEK2* mutation but requested further genetic analysis because there had been many cases of cancer on both her father's and mother's sides of the family. A full panel of cancer genes, analyzed by next-generation sequencing, uncovered a different mutation in the *CHEK2* gene, a variant that is not seen frequently. The testing laboratory reported this mutation as "likely deleterious." The genetic counselor involved consulted with an independent laboratory, which instead concluded that this gene variant was not deleterious at all—just a "benign polymorphism."

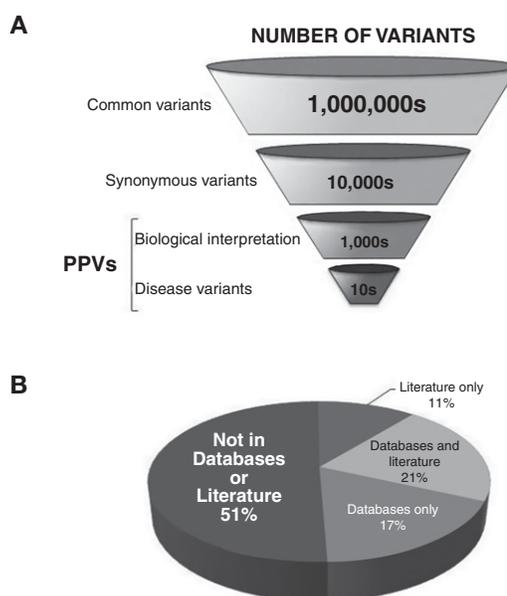
Academic clinical genetics programs frequently consult with independent commercial laboratories when faced with a laboratory's sequence interpretation that is in the gray area where the words "pathogenic" or "deleterious" are hedged with a preceding "likely." However, seeking such a second opinion is cumbersome and time consuming—it is not feasible for busy community genetics clinics. In the case above, the second laboratory claimed they had more data (e.g., large numbers of unaffected individuals who carry this variant or cancer patients who do not carry the variant even though the variant runs in their family

Corresponding author: theo.ross@utsouthwestern.edu

© 2015 Ross This article is distributed under the terms of the Creative Commons Attribution-NonCommercial License, which permits reuse and redistribution, except for commercial purposes, provided that the original author and source are credited.

Published by Cold Spring Harbor Laboratory Press

doi: 10.1101/mcs.a000414

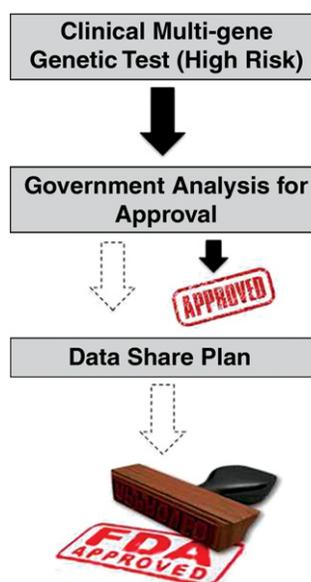


**Figure 1.** Variant classification and distribution. (A) Variant annotation. The sequenced genome is filtered by comparison to the reference genome to remove “common variants” (occurs in >5% of the general population) and “synonymous variants” that have no obvious effect on protein sequence (rarely these could be deleterious by affecting RNA stability or translational efficiency or start sites). The remaining rare “potentially pathogenic variants” (PPVs) are evaluated for biological relevance (e.g., associations with diseases and disease genes, signaling pathways, cellular processes, biomarkers, and clinical phenotypes). For variants to be considered disease variants they need experimental validation. (B) Variant distribution. This chart illustrates the frequency of VUSs (calculated from interrogation of 695 unique PPVs) from analysis of 163 well-known disease genes in 176 patients’ whole-genome sequencing data (Foley et al. 2015). A variant not found in the literature or in databases (Exome Variant Server [ESP6500] or ClinVar) was by default a VUS. Most of the remaining variants found in the literature or databases were VUSs as well.

tree) and were therefore able to reach a more informed conclusion. Even then, we had to trust their interpretation and could not verify that their interpretation was based on additional supporting cases, as their data were not in the public domain. If the two laboratories had pooled their data, both they and the patient would have benefited. They did not share and the question therefore remains unanswered: Whose interpretation is correct?

### A POTENTIAL SOLUTION: GOVERNMENT REGULATION

The U.S. Food and Drug Administration (FDA), in a move to address such challenges of complex clinical test interpretation, is now choosing to enforce its authority to regulate commercial clinical laboratory-developed tests. Laboratory-developed tests are defined as diagnostic tests developed and used within a single laboratory. Some have nicknamed these tests “home brews.” Examples include simple tests such as a measurement of blood potassium levels and more complex tests such as identification of DNA variants to diagnose a genetic disease. Increased test oversight on the part of the FDA is a healthy development because it will improve patient care and increase innovation. This will be especially true if this enhanced regulation also includes rules for data sharing, rather than competition (Fig. 2). With this new enforcement, the expectations for progress in the area of genetic testing will certainly outpace the hype. This will mean that expectations for most genetic tests will



**Figure 2.** U.S. government approval of genetic tests: now and the future. Laboratory-developed genetic tests are currently performed in a CLIA-approved laboratory. If classified as high risk, the newly enforced requirement for FDA approval will ensure their safety and effectiveness. To improve the interpretation of genetic tests in the future, government approval could require sharing of sequence data with other clinical and academic laboratories without restrictions, such as through public databases (dashed arrows).

be realistically lowered, with the hype following suit. In addition, with proper regulation we can comfortably elevate expectations that some of our genetic testing will match the hype. Overpromises are especially harmful in oncology, where some hospitals put out misleading advertisements that exaggerate how their next-generation sequencing tests lead to better cancer care.

In 1996 when Myriad Genetics started testing for breast and ovarian cancer susceptibility mutations, there were just a few genes on the genetic testing menu. Now, in 2015, our ability to generate genetic data with next-generation sequencing is outpacing our ability to understand the data. The proliferation of tests is the reason we need a governing agency to gather data to protect patients from inaccurate interpretations that lead to unnecessary action. Individual commercial laboratories cannot be expected to do this alone. We currently test for hundreds of genes that, when mutated, can increase the risk of cancer and other diseases (Stenson et al. 2014). More and more of these genes will be identified over time with our increasing ability to sequence entire genomes. Our ability to understand the significance of changes in the ~19,000 human genes is also improving (Ezkurdia et al. 2014). Yet, even the most established clinical laboratories can struggle with how to interpret the clinical significance of many genetic variants.

## WHAT THE GOVERNMENT PROPOSES TO DO

There are two main regulatory issues to consider when it comes to laboratory-developed test standards: technical accuracy of the test and the correct interpretation of the results. Until now, the government has chosen to enforce regulation of the technical standards through CLIA approval (Clinical Laboratory Improvement Amendments of 1988; see below) while leaving the interpretation open to, well, interpretation.

“High-risk” clinical tests are the tests the FDA proposes to regulate. Laboratory-developed genetic tests are considered “high risk” if the result and subsequent interpretation is the basis for expensive or life-changing medical recommendations and decisions. They are “high risk” because there is harm to the patient if the interpretations are incorrect. This risk-based regulatory approach is necessary. The FDA’s resources cannot be exhausted on low-risk tests in lieu of higher risk complex tests, such as those that interpret multiple gene sequences and risk misdiagnoses, inappropriate treatments or avoidable adverse events. Although the precise number of current laboratory-developed tests is not known, the FDA estimates that there are thousands. With such a large number, risk stratification is required.

Opponents of FDA regulation of high-risk tests claim that other mechanisms already exist to ensure that tests are clinically valid (Evans and Watson 2015). They say there is no need to regulate laboratory tests beyond the mechanism of an enhanced CLIA program. Based on our experience, this is not correct.

Another case illuminates the current capabilities of the CLIA program: a patient who had a technically inaccurate test result. The patient was a young woman with a family history of breast cancer who was found to have a “normal” *BRCA* gene. Shortly after she received the normal test result, she was diagnosed with breast cancer. We retested her through a second laboratory and found that she, in fact, has a *BRCA* mutation that is known to predispose to breast cancer. The first laboratory’s *BRCA* test result was a “false negative”—a major laboratory mistake.

Because the CLIA program monitors these types of mistakes to ensure laboratory quality, we reported this mistake to the CLIA program. When a laboratory repeatedly makes technical mistakes, it loses “CLIA approval.” For this particular patient’s problem, they were the right people to report such a mistake to. And this is where CLIA stops.

A CLIA representative will tell you they are limited to maintaining technical standards of clinical laboratories and that the CLIA program has no way to assess or regulate whether laboratories are correct in their interpretation of what tests results mean for the patient. Simply put, even when test data are technically sound, as assured by the CLIA regulatory program, the data can still be interpreted differently by different laboratories. This is where the FDA sees a vacuum it must fill by enforcing regulation of test interpretation. With laboratory tests becoming more complex, the technical accuracy of the test result does not guarantee we will know how to interpret the test or what the result means for our patients.

How a testing laboratory reads and then translates the results for physicians to use for care of our patients, like the woman possessing the *CHEK2* mutation described above, is not regulated by CLIA. Currently, the system relies on doctors to be judicious about which genetic tests they order and how to use the information. The assumption is that doctors will stop ordering tests that do not return clinically useful information. However, simply because a laboratory claims to have a test to tell patients and their doctors what their risk for a disease is does not mean the laboratory interpretation is correct. This also places an unreasonable burden on doctors because it requires them to stay on top of a large, constantly changing and diverse body of scientific literature.

You might argue the patient with the *CHEK2* variant and the different clinical interpretations between commercial laboratories provides merely anecdotal evidence. Unfortunately, my colleagues and I see this more often than we would like. To address this systematically, we asked one clinical laboratory to reinterpret the clinical significance of genetic variants that had been found in 105 of our cancer genetics patients by a second company (Yorczyk et al. 2014). More than half of the 32 gene variants seen in these patients were interpreted differently by the second laboratory. Some of the genetic variants were called normal, clinically significant, or of unknown significance by one laboratory but not by the other, suggesting that physician care would differ based on different interpretations of different companies.

The FDA's decision to enforce regulation of these high-risk tests should help prevent different interpretations of the same clinical data.

Of course the need for regulation of complex genetic tests goes beyond germline cancer variants. We recently found that a large number of our cancer genetics patients' whole-genome sequences contained a mutation in *PRSS1* (Foley et al. 2015). This mutation supposedly causes pancreatitis in many of its carriers, but none of the *PRSS1*-mutant patients were diagnosed with pancreatitis or reported a history of symptoms that could reflect pancreatitis. It is possible that this "mutation" has been given an undeserved bad name. Several other studies have found the same thing—for example, 27% of "disease" mutations have been found to be incorrect, incomplete, or just common polymorphisms (Bell et al. 2011). Knowledge of the many correct spellings of our genes will be achieved only after large numbers of genomes are sequenced (MacArthur et al. 2014). In the past, we simply did not have the tools to do this. Now we do. We need more research, more subjects, and more sequence data published and put in public databases. This has crucial implications for people who make life-changing decisions based on what is known about sequence variants.

### ADDING TO THE SOLUTION: DATA SHARING

---

The FDA's goal (and that of other governments around the world) to standardize the interpretation of "high-risk" test results that affect patient care will be made possible with more data sharing. Many laboratories are now sequencing panels of genes in the germline DNA to look for mutations that will influence clinical management recommendations. Problems arise when interpretable genetic variants with well-established clinical significance are mixed up with VUSs. When a well-established deleterious *BRCA1* mutation—a cause of hereditary breast and ovarian cancer syndrome—is put in the same results category as a *BRCA1* VUS that one laboratory calls out as a polymorphism and another a "likely deleterious" variant or VUS, patients and doctors get confused. Often the VUSs are mistakenly called deleterious mutations by doctors (Domchek et al. 2013). These hazy results remain in the domain of research, not patient care. And note that these VUS reports are coming from CLIA-approved laboratories. Incentives to share high-quality genotype–phenotype data in publications (e.g., *CSH Molecular Case Studies*) and in public databases (e.g., ClinVar, dbGaP, and NIH's Genetic Testing Registry) may help governments reach their regulatory goals.

Understandably, some commercial laboratories are concerned about what the additional oversight and sharing could mean. They assert that more regulation will squash innovation (Evans and Watson 2015). They claim that the current landscape in which there is little regulation of test interpretation is one of "vibrant competition" leading to the development of new and improved tests. They are wrong. In this situation where genetic test results are uninterpretable without large numbers of patients, competition is caustic. To quote the current Secretary of Health and Mental Hygiene, Joshua Sharfstein (Sharfstein 2015), "Innovation is not just novelty; it is novelty that works."

Keeping genetic data too close to the chest leads to misdiagnoses. What if the testing laboratories for our patient with the *CHEK2* variant had compared her *CHEK2* sequence with a common database of genotypes and phenotypes? Both laboratories may have seen that our patient's genome had an alternate, likely nonpathogenic sequence of *CHEK2* found in patients who do not show an increased incidence of cancer (assuming the second laboratory was correct when they claimed they had more data). The first laboratory would not have made the error of calling it harmful, and she would not have been at risk for taking unnecessary steps to manage a risk she does not have.

We need a system that discourages clinical laboratories from considering their data proprietary. When it comes to genomes and patient care, laboratories simply cannot compete to

see who can “win” by having the most authoritative interpretation of sequence data first. Beyond competition, there are also policy, legal, and information technology issues that impede sharing and must also be worked through (Collins and Hamburg 2013). Nonetheless, data sharing among laboratories is critical and mechanisms have to be found to facilitate it. The data should be shared so that all clinicians and researchers can draw on that information to make the most informed interpretations for their patients.

Across the world, countries like the United Kingdom, China, and Australia are setting up their own banks of data. Public databases in the United States are also building up steam and include the NIH’s ClinVar, the Genetic Testing Registry, and dbGaP. Recently, our genetics program has worked to put our linked clinical and whole-genome sequence “research” data into dbGaP. This database is available for authorized scientists to chew on the raw sequence data, to reinterpret the data, to compare to their data, and to come up with their own hypotheses. Putting linked clinical and sequence data into databases is not trivial but it can be done. Uploading terabytes of sequence data and entering each of the subject’s personal and family histories into the database requires resources. If all clinical laboratories share their data with a few public databases, we will know more quickly about the clinical implications of the sequences and we will be able to use that information to help patients in the future.

Imagine if the FDA had required Myriad Genetics to share the million-plus *BRCA1/2* test results that have been paid for by patients’ insurance companies. This would make it possible for all research and clinical laboratories as well as health-care providers to have the most comprehensive understanding of the clinical significance of each genetic variant. We would all better understand how the myriad alternative spellings of *BRCA1* and *BRCA2* affect risk, and patients would be able to make more informed choices about managing their risks.

There is hope that the FDA will build a data sharing plan into their guidelines as they develop over the years. If the FDA approval requires reporting of test results and clinical outcomes, not only will “uncertain” test results become interpretable more quickly, but our overall understanding of human genetics will increase. This requirement, together with established or future data sharing initiatives such as the Global Alliance’s BRCA Challenge and the importance of journals like *CSH Molecular Case Studies*, is the only way for us to make progress.

If we do not quickly fill up these databases, we will be missing out on what could be the health-care highlight of the 21st century. In FDA-approved repositories of gene sequences and patient data from clinical tests, there could be libraries full of genetic codes, helping scientists, doctors, and patients improve their ability to improve the care of current and future patients.

Let us hope that the FDA’s stamp of approval of laboratory-developed genetic tests will also require and facilitate sharing of test results linked to clinical data that are the basis for test interpretation (Fig. 2). This will help laboratories standardize their test interpretations as well as ensure interpretations are based on the most current data. Data sharing, although not free of logistical hurdles, will lead to more enlightened patient care and innovation.

## ADDITIONAL INFORMATION

---

### Competing Interest Statement

The author has declared no competing interest.

### Acknowledgments

I thank the Cancer Genetics team and patients at UT Southwestern Medical Center, members of my laboratory, and Sean Morrison for several critical discussions and comments that led to the initiation and development of this essay.

## REFERENCES

---

- Bell CJ, Dinwiddie DL, Miller NA, Hateley SL, Ganusova EE, Mudge J, Langley RJ, Zhang L, Lee CC, Schilkey FD, et al. 2011. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* **3**: 65ra64.
- Collins FS, Hamburg MA. 2013. First FDA authorization for next-generation sequencer. *N Engl J Med* **369**: 2369–2371.
- Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME. 2013. Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? *J Clin Oncol* **31**: 1267–1270.
- Evans JP, Watson MS. 2015. Genetic testing and FDA regulation: overregulation threatens the emergence of genomic medicine. *JAMA* **313**: 669–670.
- Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* **23**: 5866–5878.
- Foley SB, Rios J, Mgbemena VE, Robinson L, Hampel HL, Toland AE, Duham L, Ross TS. 2015. Use of whole genome sequencing for diagnosis and discovery in the cancer genetics clinic. *EbioMedicine* **2**: 74–81.
- MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**: 469–476.
- Sharfstein J. 2015. FDA regulation of laboratory-developed diagnostic tests: protect the public, advance the science. *JAMA* **313**: 667–668.
- Stenson PD, Mort M, Ball EV, Shaw K, Phillips A, Cooper DN. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* **133**: 1–9.
- Yorczyk A, Robinson LS, Ross TS. 2014. Use of panel tests in place of single gene tests in the cancer genetics clinic. *Clin Genet*. doi: 10.1111/cge.12488.