GENETIC TESTING AND MOLECULAR BIOMARKERS
Volume 18, Number 1, 2014

Pp. 1–2
DOI: 10.1089/gtmb.2014.1551

**PERSPECTIVE**

# Free the Data

Katherine Lambertson and Sharon F. Terry

It is probable that there are both genetic and environmental risk factors for breast and/or ovarian cancer. Adequate evidence suggests that mutations in the *BRCA1* and *BRCA2* genes are associated with an increased risk of breast and ovarian cancer, and perhaps other cancers as well (U.S. Preventive Services Task Force, 2013). Although estimates vary widely, in the general population about 5%–10% of all breast cancer and 10%–15% of ovarian cancer cases can be attributed to variation in these genes. This rate is higher in some populations and families (Fackenthal and Olopade, 2007; Ferla *et al.*, 2007).

Millions of individuals worldwide have a mutation in the *BRCA1* or *BRCA2* gene. Mutations in one of these genes can significantly increase an individual's chances of developing hereditary breast or ovarian cancer. A positive result asks for difficult decisions from the individual. The individual needs as much validated information as possible to make the best choices from options that are difficult to begin with. Clinicians need to know which mutations have been identified and whether they are putative. Researchers need information about the genes and associated clinical data to help refine the predictive value of the genetic test, potential allele-specific therapies, and further understanding of disease progression. None of this is possible with the mutations alone, but knowing what mutations have been discovered is an important part of discovery and clinical services.

Some genes have substantial variation represented in open databases. Some of these databases have received attention from the Human Variome Project, which seeks to ''collect and curate all human variation affecting human health (Ring *et al.,* 2006).'' There are groups working on tools to curate the data, such as the Leiden Open Variation Database (LOVD) (Fokkema *et al.* 2011), whose goal is to provide an open tool to display DNA variation. LOVD at NCBI (National Center for Biotechnology Information) offers the LOVD system installed on an NCBI server at the National Library of Medicine of the National Institutes of Health. PXE International, a small disease-specific advocacy organization for pseudoxanthoma elasticum (PXE) founded and managed by this author, has created a comprehensive database of mutations using LOVD at NCBI (PXE International, 2013, available at http://pxe.org/mutation-database). The database contains 288 *ABCC6* mutations from nearly 500 people representing 20% of the known affected population and almost all of the tested population. With permission from individuals who have been tested, PXE International collects mutations from GeneDx, the laboratory with the largest *ABCC6* volume of testing. Individuals from all over the world send in their test results to be placed in the database. In a bygone era, these mutations would be published, but now, unless there are other important considerations, they go largely unpublished. There have been some appeals for ''microattribution'' to encourage crowdsourcing variation for the human genome (Giardine *et al.*, 2013; Patrinos *et al.*, 2012; Giardine *et al.*, 2011; Axton, 2011), building on a plea Axton made as early as 2008 (Axton, 2008, Nature Genetics Blog, available at http://blogs.nature.com/freeassociation/2008/03/microattribution_for_community_I.html).

In recent years, NCBI has taken on the effort of cataloging variation. ClinVar provides an open-access archive of reports of relationships among clinically important variants and phenotypes. The database is integrated with other NCBI assets: dbSNP and dbVar, which maintain information about the location of variation in the genome (Landrum *et al.*, 2013). Other efforts include the International Collaboration for Clinical Genomics (ICCG) and an effort newly funded by the National Institutes of Health, called the ClinGen Project, in which a very large consortium of these laboratories will share variants and clinical data.

In June 2013, the *BRCA1/2* genes received a great deal of attention. In the lawsuit *Association for Molecular Pathology et al. v. Myriad Genetics, Inc., et al,* the U.S. Supreme Court determined that isolated genomic DNA is not patent-eligible. This opened the door for many other laboratories to test for *BRCA1/2* (Myriad is suing). However, the Supreme Court ruling did nothing to open access to the proprietary *BRCA1/2* mutation database, which is considered Myriad's ''trade secret'' and legal according to common business practices. Myriad is not the only testing laboratory that doesn't put all of the variation it discovers into open-access databases.

For variation in all genes, this is unacceptable. It is time to Free the Data!

Free the Data is a grassroots movement collecting mutations and depositing them in ClinVar and other open-access databases. It was originally launched as a component of the ICCG by Robert Nussbaum of the University of California, San Francisco, and is supported by the Laboratory for Molecular Medicine at the Harvard Partners Healthcare Center for Personalized Genetic Medicine. The project invites clinicians to submit *BRCA1/2* reports so that the variation can be catalogued in ClinVar. Free the Data expanded this year to invite individuals to donate their mutations and associated

Genetic Alliance, Washington, District of Columbia.

clinical data. The steering committee includes many of the brave pioneers from the ICCG and others who have worked on open access for a long time. A list of the steering committee and members of the consortium can be found on the Free the Data Web site (available at www.free-the-data.org).

Free the Data starts with the *BRCA1/2* genes, arguably the most sequenced genes in the human genome. The campaign will expand to encompass all genes and all mutations, powering open-access research and better health for all. In the information age, data are not scarce and cannot be treated as a commodity. It is time to Free the Data!

## References

Axton M (2011) Crowdsourcing human mutations. Nat Genet. 43: 279.

Fackenthal JD, Olopade OI (2007) Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat Rev Cancer. 7: 937–948.

Ferla R, Calo V, Cascio S, *et al.* (2007) Founder mutations in BRCA1 and BRCA2 genes. Ann Oncol. 18(Suppl 6):vi93–98.

Fokkema IF, Taschner PE, Schaafsma GC, Celli J, Laros JF, den Dunnen JT (2011) LOVD v.2.0: the next generation in gene variant databases. Hum Mutat. 32: 557–563.

Giardine B, Borg J, Higgs DR, *et al.* (2011) Systematic documentation and analysis of human genetic variation in hemoglobinopathies using the microattribution approach. Nat Genet. 43: 295–301.

Giardine B, Borg J, Viennas E, *et al.* (2013) Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic Acids Res. Published online October 16, 2013.

Landrum MJ, Lee JM, Riley GR, *et al.* (2013) ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. Published online November 14, 2013.

Patrinos GP, Cooper DN, van Mulligen E, *et al.* (2012) Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. Hum Mutat. 33: 1503–1512.

Ring HZ, Kwok PY, Cotton RG (2006) Human Variome Project: an international collaboration to catalogue human genetic variation. Pharmacogenomics 7: 969–972.

U.S. Preventive Services Task Force (2013). Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer: U.S. Preventive Services Task Force Draft Recommendation Statement. Available at: http://www.uspreventiveservicestaskforce.org/uspstf12/brcatest/draftrecbrcatest.htm

Address correspondence to:
*Sharon F. Terry, MA*
*President & CEO*
*Genetic Alliance*
*4301 Connecticut Avenue, NW*
*Suite 404*
*Washington, DC 20008*

*E-mail:* sterry@geneticalliance.org