

CORRESPONDENCE

Do-it-yourself genetic testing

Steven L Salzberg* and Mihaela Pertea

Abstract

We developed a computational screen that tests an individual's genome for mutations in the BRCA genes, despite the fact that both are currently protected by patents.

As we learn more about the associations between genes and disease, a growing number of diagnostic tests have been developed to detect mutations that increase the risks of various diseases. However, anyone who wants to develop a diagnostic test or a treatment based on human genes faces a potential roadblock: gene patents. A 2005 study [1] reported that 4,382 human genes (~20% of the total number in our genome) are covered by patents or other intellectual property claims. These patents cover a wide range of methods for assaying the DNA sequence of an individual for the presence of disease-associated mutations. For example, one of the most consequential gene patents covers mutations in the *BRCA1* [2] and *BRCA2* [3] genes, which are associated with a significantly increased risk of breast and ovarian cancer [4-6]. The BRCA gene patents, which are held by Myriad Genetics, cover all known cancer-causing mutations in addition to those that might be discovered in the future. No one can develop a commercial diagnostic test or a treatment based on the BRCA gene sequences without a license from Myriad. Although a US federal court recently overturned seven of Myriad's BRCA patents, Myriad is appealing the ruling, and it holds 16 other BRCA-related patents that it claims are unaffected by the court's ruling [7].

As the cost of DNA sequencing falls, the idea of testing for mutations one gene at a time is rapidly becoming obsolete. We are also rapidly approaching the day when it will be cheaper to fully sequence a genome before testing the sequence for all known genetic mutations associated with a given disease than to conduct multiple separate tests for each gene. Currently Myriad charges more than

\$3000 for its tests on the BRCA genes, while sequencing one's entire genome now costs less than \$20,000. Furthermore, once an individual's genome has been sequenced, it becomes a resource that can be re-tested as new disease-causing mutations are discovered.

In contrast to whole-genome sequencing, standard methods for identifying mutations in *BRCA1* and *BRCA2* use PCR to amplify the genome regions containing each mutation [8]. As more mutations are discovered, these tests need to be augmented with additional PCR assays, adding to their cost. The commercial assay available from Myriad Genetics interrogates a limited number of sites by PCR and sequencing, which can miss clinically relevant mutations; for example, a recent study [9] reported that 12% of women from high-risk families with deleterious mutations in *BRCA1* or *BRCA2* had false negative results from this assay. Even if the test were perfect, a gene-centered approach will be far more expensive over time than a computational assay based on an individual's genome, because the genome only needs to be sequenced once, after which it can be used to test all 22,000+ human genes.

Regardless of how easy it might be to test for mutations, the restrictive nature of the BRCA gene patents means that anyone wishing to examine any mutation in *BRCA1* or *BRCA2* will have to obtain permission from the patent holder Myriad Genetics. This restriction applies even if testing your own genome. If you wanted to look at other genes, you would have to pay license fees for any of them that were protected by patents. In practice, although it may seem absurd, this means that before scanning your own genome sequence, you might be required by law to pay thousands of license fees to multiple patent holders.

We believe that any individual should be allowed to interrogate his or her genome for all mutations of interest, regardless of whether a private company claims to 'own' the rights to particular gene mutations. To challenge the restrictive gene patenting system, we have developed a computational assay that, as a proof-of-concept, tests for 68 known variants of the *BRCA1* and *BRCA2* genes. In other words, we empower any individual using our software (whether this is a private individual, a clinician or a clinical or basic researcher) to test for these mutations and circumvent the gene patents. Here we demonstrate the method on the publicly

*Correspondence: salzberg@umd.edu
Center for Bioinformatics and Computational Biology, University of Maryland,
College Park, MD 20742, USA

available DNA sequence from three human genomes: a Caucasian female, an African male and an Asian male [10].

We have made the software freely available (at <http://cbcb.umd.edu/software/BRCA-diagnostic>) under an open source license, allowing others to use, modify and redistribute it. The software is flexible and can easily be adapted to search for mutations in other genes. The method uses the raw sequence reads that are produced by a high-throughput sequencer; it does not require genome assembly nor any other processing of the raw data. This software provides a relatively simple, do-it-yourself home testing method for interrogating a genome for the presence of mutations in the BRCA genes. All one needs, besides the software, is the sequence data from an individual human.

BRCA testing on three human genomes

We used the Bowtie short-read alignment program [11] to screen all sequence reads against the *BRCA1* and *BRCA2* regions (located on chromosomes 17 and 13, respectively) and against a set of 68 known mutations from the Online Mendelian Inheritance in Man (OMIM) database (see Methods). The size of the datasets ranged from 2.8 to 4.1 billion reads for each genome, with most reads being 35-36 bp. The BRCA genomic regions are each about 80-90 kb; with these small target sequences Bowtie is extremely fast. Using only a single 2.4 GHz processor, Bowtie aligned reads at 127 million reads per hour, and alignment of the largest of our datasets took about 8 hours. Thus despite the enormous number of reads for each genome, screening was relatively fast.

In the Asian and African males, we found no evidence for any of the 68 deleterious mutations in *BRCA1* and *BRCA2*. The Caucasian female had no mutations at 67 of the 68 sites, but she has a heterozygous mutation at one site in *BRCA2*. At this location, 26 reads match the mutant base (C) and 24 reads match the normal base (A). This A-C mutation causes a single amino acid change, N372H, in exon 10, which in homozygous form was originally reported to carry a 30-40% increased risk of breast cancer [12,13], although a subsequent study reported no increased cancer risk [14].

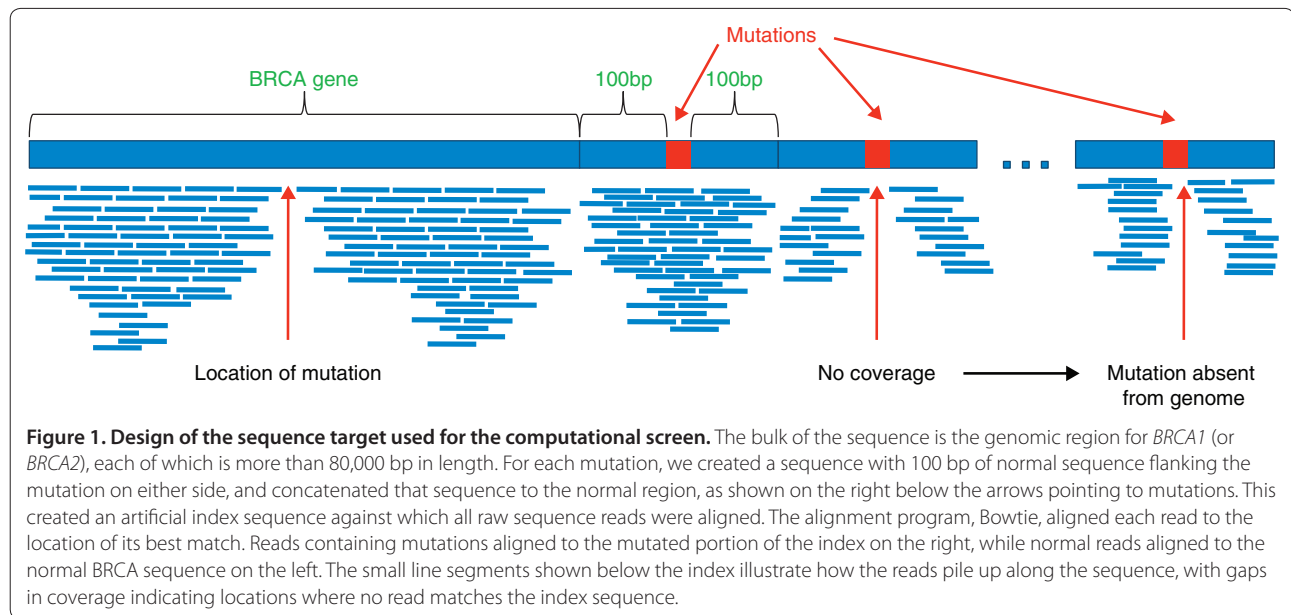
Note that the 68 mutations used in this proof-of-concept assay do not represent a comprehensive list of BRCA mutations. We used OMIM as our primary source, but other databases have much larger lists of BRCA mutations (for example the Human Gene Mutation Database [15] lists 1,215 mutations for *BRCA1* and 966 for *BRCA2*). Most of these additional mutations could easily be added to our test, simply by incorporating them in the sequence index file described below. The software can be extended to other genes by creating new index files for those genes.

If free software can be used to diagnose human genetic mutations, then individuals will be able to run their own tests in the privacy of their own homes. Fundamentally, this seems no different from measuring one's temperature or blood pressure, but because of gene patents, the act of reading one's own genome may require the permission of a private company. It is hard to envision how the patent holders can enforce their claims in this scenario. Our contention is that these patents never should have been awarded, and that no private entity should have rights to the naturally occurring gene sequences in every human individual.

Computational methods

A list of mutations in *BRCA1* and *BRCA2* were compiled from the OMIM database of human genetic diseases [16], identifiers 113705 and 600185. We created indexes for the Bowtie program [11] using the *BRCA1* and *BRCA2* genomic regions including introns that span 81,155 bp and 84,193 bp, respectively. A Bowtie index is a specialized, compressed representation of a genome sequence that enables very fast alignment. At the end of each region, we concatenated DNA sequences corresponding to each of the 35 (*BRCA1*) and 33 (*BRCA2*) mutations listed in OMIM (Figure 1). These extra sequences included 100 bp on either side of the mutant site. The mutations include insertions, deletions and base pair changes.

All three genomes were sequenced using the Illumina platform. The Asian genome (3,334,275,294 reads) was the first sequence of an Asian individual to be published [10]. The African (4,055,510,372 reads) and Caucasian (2,807,568,082 reads) genome data were generated for the 1000 Genomes Project; the African male is a member of the Yoruba population in Ibadan, Nigeria (individual NA18507) and the Caucasian female is from a set of Utah residents (CEPH) with European ancestry (individual NA12892). The Asian, African and Caucasian genomes were sequenced to 40x, 50x and 35x coverage, respectively, which means that for each genomic position, an average of 40, 50 and 35 sequence reads covered that position. The DNA samples from the 1000 Genomes Project are anonymous and have no associated medical or phenotype data, and all sample collection followed ethical guidelines developed for that project, which permits the use of these data to study genetic diseases [17]. We then aligned all reads for each genome to both *BRCA1* and *BRCA2* using Bowtie version 0.12.3 [11] with default parameters, which reported only the best match for each read, allowing up to two mismatches. Because the indexes included both normal and mutant versions for each known sequence variant, the best match for a read aligned to the normal version unless that read derived from a mutant locus. Additional mutations can



be added simply by concatenating them to the target sequence and rebuilding the Bowtie index.

We created new programs to process all matching reads and report which if any reads matched each of the 68 mutations in the diagnostic screen. For each mutation, the program reports whether the individual has the mutation, and whether the individual is homozygous or heterozygous for that mutation. In creating this software, we are not violating the *BRCA* patents directly but any user would be, because even a noncommercial use (such as examining one's own genome) is considered to be patent infringement [18].

Preparing for the genomic age

Finally, we recognize that there may be some controversy about giving ordinary individuals the ability to test their own DNA, without also providing expert genetic counseling. As pointed out in a recent *New England Journal of Medicine* article: "health care providers are increasingly bypassed ... as patients embrace direct-to-consumer (DTC) genetic tests and turn to social networks for help in interpreting their results. In the future, a primary role of health care professionals may be to interpret patients' DTC genetic test results and advise them about appropriate follow-up" [19]. The same article points out that "most primary care providers struggle to interpret single-gene tests (e.g., for *BRCA1* and *BRCA2*) and are unprepared for the genomic age." Nonetheless, the door to this new technology is already open and it cannot be closed. Rather than trying to keep patients in the dark, we need to embrace the technology and work harder to educate both physicians and patients about the power and the limitations of genetic tests.

Published: 7 October 2010

References

1. Jensen K, Murray F: Intellectual property. Enhanced: intellectual property landscape of the human genome. *Science* 2005, **310**:239-240.
2. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, et al: A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 1994, **266**:66-71.
3. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, Nguyen K, Seal S, Tran T, Averill D, Fields P, Marshall G, Narod S, Lenoir GM, Lynch H, Feunteun J, Devilee P, Cornelisse CJ, Menko FH, Daly PA, Ormiston W, McManus R, Pye C, Lewis CM, Cannon-Albright LA, Peto J, Ponder BAJ, Skolnick MH, Easton DF, Goldgar DE, Stratton MR: Localization of a breast cancer susceptibility gene, *BRCA2*, to chromosome 13q12-13. *Science* 1994, **265**:2088-2090.
4. The Breast Cancer Linkage Consortium: Cancer risks in *BRCA2* mutation carriers. *J Natl Cancer Inst* 1999, **91**:1310-1316.
5. Thompson D, Easton DF: Cancer incidence in *BRCA1* mutation carriers. *J Natl Cancer Inst* 2002, **94**:1358-1365.
6. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, Bishop DT, Weber B, Lenoir G, Chang-Claude J, Sobol H, Teare MD, Struwing J, Arason A, Scherneck S, Peto J, Rebbeck TR, Tonin P, Neuhausen S, Barkardottir R, Eyfjord J, Lynch H, Ponder BA, Gayther SA, Zelada-Hedman M, et al: Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am J Hum Genet* 1998, **62**:676-689.
7. Wadman M: Breast cancer gene patents judged invalid. *Nature* 2010, doi:10.1038/news.2010.160.
8. Frank TS, Manley SA, Olopade OI, Cummings S, Garber JE, Bernhardt B, Antman K, Russo D, Wood ME, Mullineau L, Isaacs C, Peshkin B, Buys S, Venne V, Rowley PT, Loader S, Offit K, Robson M, Hampel H, Brenner D, Winer EP, Clark S, Weber B, Strong LC, Thomas A, et al: Sequence analysis of *BRCA1* and *BRCA2*: correlation of mutations with family history and ovarian cancer risk. *J Clin Oncol* 1998, **16**:2417-2425.
9. Walsh T, Casadei S, Coats KH, Swisher E, Stray SM, Higgins J, Roach KC, Mandell J, Lee MK, Ciernikova S, Foretova L, Soucek P, King MC: Spectrum of mutations in *BRCA1*, *BRCA2*, *CHEK2*, and *TP53* in families at high risk of breast cancer. *JAMA* 2006, **295**:1379-1388.
10. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, et al: The diploid genome sequence of an Asian individual. *Nature* 2008, **456**:60-65.

11. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol* 2009, **10**:R25.
12. Healey CS, Dunning AM, Teare MD, Chase D, Parker L, Burn J, Chang-Claude J, Mannermaa A, Kataja V, Huntsman DG, Pharoah PD, Luben RN, Easton DF, Ponder BA: **A common variant in BRCA2 is associated with both breast cancer risk and prenatal viability.** *Nat Genet* 2000, **26**:362-364.
13. Spurdle AB, Hopper JL, Chen X, Dite GS, Cui J, McCredie MR, Giles GG, Ellis-Steinborner S, Venter DJ, Newman B, Southey MC, Chenevix-Trench G: **The BRCA2 372 HH genotype is associated with risk of breast cancer in Australian women under age 60 years.** *Cancer Epidemiol Biomarkers Prev* 2002, **11**:413-416.
14. Cox DG, Hankinson SE, Hunter DJ: **No association between BRCA2 N372H and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2005, **14**:1353-1354.
15. Stenson PD, Ball E, Howells K, Phillips A, Mort M, Cooper DN: **Human Gene Mutation Database: towards a comprehensive central mutation database.** *J Med Genet* 2008, **45**:124-126.
16. Amberger J, Bocchini CA, Scott AF, Hamosh A: **McKusick's Online Mendelian Inheritance in Man (OMIM).** *Nucleic Acids Res* 2009, **37**:D793-D796.
17. **1000 Genomes** [<http://www.1000genomes.org/>]
18. **BRCA: Genes and Patents** [<http://www.aclu.org/free-speech/brca-genes-and-patents>]
19. Evans JP, Dale DC, Fomous C: **Preparing for a consumer-driven genomic age.** *N Engl J Med*, **363**:1099-1103.

doi:10.1186/gb-2010-11-10-404

Cite this article as: Salzberg SL, Pertea M: **Do-it-yourself genetic testing.** *Genome Biology* 2010, **11**:404.