Breast Cancer Genomics in the Deep Sequencing Era

Nir Pillar MD and Noam Shomron PhD

Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

KEY WORDS: breast cancer, deep sequencing, germline mutations, *BRCA*, genomics

IMAJ 2014; 16: 783–784

Breast cancer is the most common malignancy among females. Approximately 12% of women will be diagnosed with breast cancer at some point during their lifetime [1]. Currently, there are about 3 million women in the United States living with breast cancer (roughly 70,000 in Israel); 40,000 of them die each year due to disease complications (roughly 900 in Israel) [1,2].

Most breast cancer patients are considered sporadic, as familial susceptibility to breast cancer accounts for < 25% of all cases. About 20% of the inherited cases are caused by mutations in the *BRCA1* and *BRCA2* genes [3]. *BRCA1* and *BRCA2* are tumor suppressor genes involved in repair of double-strand DNA breaks, control of cell cycle checkpoint responses, and genome integrity. The *BRCA1* and *BRCA2* genes are located on chromosomes 17q21 and 13q12.3 and are translated into 1863 and 3418 amino acid long proteins, respectively. Over 2000 different mutations have been reported in *BRCA1/2* genes including deletions and insertions, in addition to many single nucleotide polymorphisms (SNPs). Mutations occur throughout the entire coding region, while most of them lead to a truncated protein when *BRCA1* and *BRCA2* genes are translated.

Despite the large number of different mutations on *BRCA1/2*, there are a few groups in which a small subset of these mutations occurs at a higher rate than expected. The recurrence of these mutations is related to a founder effect caused by interbreeding due to geographic or religious isolation. An example of such a group is the Ashkenazi Jewish population in which the rare mutations – 5382insC and 185delAG in *BRCA1* and 6174delT in *BRCA2* – were detected at a higher rate than expected [4].

In the past 20 years, additional genes and mutations conferring increased susceptibility to breast cancer were discovered and *BRCA1* and *BRCA2*-derived predisposition seems to be only a part of the general picture. Genes carrying highly penetrant germline mutations that increase breast cancer risk include the following mutations: *TP53* (Li-Fraumeni syndrome), *STK11* (Peutz-Jeghers syndrome) and *PTEN* (Cowden syndrome) [5].

Additional, more frequent, less penetrant mutations have been identified in familial breast cancer cases, in genes such as *CHEK2*, *ATM*, *PALB2*, and *BRIP1* [5]. Large case-control studies defined additional mutations associated with a minor increase in breast cancer relative risk. These included methods that scan large amounts of polymorphisms or copy number variations (CNVs). Molecular studies of breast cancer focus on high throughput platforms, including mRNA expression profiling (microarrays) or DNA comparative genomic hybridization, and more recently ultra-high throughput sequencing (also known as Deep Sequencing or Next Generation Sequencing, NGS).

In this issue of *IMAJ*, Cornejo-Moreno et al. [6] highlight the fact that despite enormous investment in research, funding and clinical collaboration, additional high penetrance breast cancer driving mutations and genes (so called "*BRCA3*") are yet to be discovered.

Deep sequencing enables the sequencing of millions of bases in a single run, by processing multiple reactions in a massively parallel fashion. The advent of deep sequencing enabled researchers to explore biological systems at an unprecedented level. Since the adoption of deep sequencing to common use (around 2008) the technology's data output has increased at a tremendous pace, more than doubling itself each year. Over 250 novel genes related to rare Mendelian inherited diseases were discovered using deep sequencing, and it is predicted that most of the remaining disease-causing genes will be identified by the year 2020 [7]. Researchers can now sequence several human genomes in a single run, producing data in less than a week, for a fraction of the price tag per genome (by comparison, the first human genome required roughly 10 years to sequence).

Deep sequencing technology has greatly expedited cancer genetics research by facilitating identification of novel mutated cancer-related genes [8], uncovering treatment resistance mechanisms [9], genomics-based classification of tumors [10], and developing cancer prediction scores for high risk patients [11]. The rapid advances in terms of higher throughput and lower cost, coupled with the development of multiple genomic sequence enrichment methods, have contributed significantly to both research and clinical applications of cancer genome sequencing. The concept of an integrative approach for a range of "omics" data (genomics/epigenomics/proteomics, etc.) is not new, but in recent years it has resurfaced and become vividly active through technological advances that have made possible the 'profiling of everything' (i.e., the characterization of informational macromolecules at the genomic, transcriptomic and epigenomic levels) using deep sequencing technologies [12].

Cancer genomic analysis by deep sequencing is carried out by detecting varia-

tions within the tumor (somatic mutations) and heritable ones in the affected individual (germline mutations). In contrast to the sequencing of cancer genomes through tumor samples to identify somatic mutations, studies describing the sequencing of non-tumor tissue to identify germline causal mutations for familial cancers are relatively few in number [13]. This is partly due to thousands to millions of cases that are necessary to encompass minor changes in populations and to reach statistical significance in classifying rare cancer variants [14].

Large-scale exploratory studies that aim to generate a comprehensive list of all major cancer-driving mutations were based on genome-wide association studies (GWAS), and deep sequencing. GWAS focused on SNPs and traits like major diseases using a very large number of samples (in the 10,000s) [15]. On the other hand, deep sequencing first focused on individual samples and then a small cohort of tens of them [16]. Large consortiums and databases embarked on using deep sequencing on mammoth sample cohorts, such as The Cancer Genome Atlas (TCGA) [17] and International Cancer Genome Consortium (ICGC) [18].

Common variants (frequency > 5%) in approximately 70 loci have been identified over the past several years as breast cancer risk factors via GWAS [19]. However, these common variants together explained only a small portion of the heritability for breast cancer. It has been recognized that the missing heritability for breast cancer and other complex diseases may be partially explained by low frequency variants (minor allele frequency < 1%) [20]. A large number of low frequency variants exists in the human genome, and these rare coding variants are under-represented in current databases that are based on small population samples.

In conclusion, deep sequencing technology created a revolution in cancer research, leading to the discovery of hundreds of novel gene mutations, CNVs and SNPs, all involved in malignancy formation. Despite these studies, no single, high prevalence driver-mutation such as *BRCA1/2* was annotated, and as the data accumulate at a tremendous pace with thousands of breast tumors being sequenced the chances of detecting "*BRCA3*" seem low. It is unlikely that most rare missense variants will be classifiable in the near future, and accurate relative risk estimates may never be available for very rare variants. Currently 70% of familial breast cancer cases cannot be related to known genomic mutations. Enlarged patient cohorts, improved data-mining algorithms, and equitable translation of genomics into cancer care should help dissipate some of the fog behind hereditary breast cancer.

Acknowledgments

We thank Ofer Isakov and Avital Gilam for their helpful suggestions.

Funding: The Shomron laboratory is supported by the Israel Cancer Research Fund (ICRF), Research Career Development Award (RCDA); Wolfson Family Charitable Fund; Earlier.org—Friends for an Earlier Breast Cancer Test; Claire and Amedee Maratier Institute for the Study of Blindness and Visual Disorders; I-CORE Program of the Planning and Budgeting Committee, The Israel Science Foundation (grant number 41/11); Saban Family Foundation-Melanoma Research Alliance.

Correspondence

Dr. N. Shomron

Phone: (972-3) 640-6594

Fax: (972-3) 640-7432

email: nshomron@post.tau.ac.il

References

1. Edwards BK, Noone AM, Mariotto AB, et al. Annual Report to the Nation on the status of cancer, 1975-2010, featuring prevalence of comorbidity and impact on survival among persons with lung, colorectal, breast, or prostate cancer. *Cancer* 2014; 120 (9): 1290-314.
2. National Cancer Registry. Annual Reports: Breast cancer incidence among Jews: trend 1980–2010. *Israel National Cancer Registration*, 2011.
3. Foulkes WD. Inherited susceptibility to common cancers. *N Engl J Med* 2008; 359 (20): 2143-53.
4. Karami F, Mehdipour P. A comprehensive focus on global spectrum of BRCA1 and BRCA2 mutations in breast cancer. *Biomed Res Int* 2013; 2013: 928562.
5. Apostolou P, Fostira F. Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int* 2013; 2013: 747318.
6. Cornejo-Moreno B, Uribe-Escamilla D, Salamanca-Gómez F. Breast cancer genes: looking for BRACA's lost brother. *IMAJ* 2014; 16: 787-92.
7. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet* 2013; 14 (10): 681-91.
8. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014; 505 (7484): 495-501.
9. Ria R, Catacchio I, Berardi S, et al. HIF-1alpha of bone marrow endothelial cells implies relapse and drug resistance in patients with multiple myeloma and may act as a therapeutic target. *Clin Cancer Res* 2014; 20 (4): 847-58.
10. Clinical Lung Cancer Genome Project (CLCGP), Network Genomic Medicine (NGM). A genomics-based classification of human lung tumors. *Sci Transl Med* 2013; 5 (209): 209ra153.
11. Laitman Y, Simeonov M, Keinan-Boker L, Liphshitz I, Friedman E. Breast cancer risk prediction accuracy in Jewish Israeli high-risk women using the BOADICEA and IBIS risk models. *Genet Res* (Camb) 2013; 95 (6): 174-7.
12. McShane LM, Cavenagh MM, Lively TG, et al. Criteria for the use of omics-based predictors in clinical trials. *Nature* 2013; 502 (7471): 317-20.
13. Ku CS, Cooper DN, Wu M, et al. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. *Mod Pathol* 2012; 25 (8): 1055-68.
14. Shirts BH, Jacobson A, Jarvik GP, Browning BL. Large numbers of individuals are required to classify and define risk for rare variants in known cancer risk genes. *Genet Med* 2014; 16 (7): 529-34
15. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009; 106 (23): 9362-7.
16. Balko JM, Giltnane JM, Wang K, et al. Molecular profiling of the residual disease of triple-negative breast cancers after neoadjuvant chemotherapy identifies actionable therapeutic targets. *Cancer Discov* 2014; 4 (2): 232-45.
17. Li Q, Seo JH, Stranger B, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell* 2013; 152 (3): 633-41.
18. Ramsay AJ, Martínez-Trillos A, Jares P, Rodríguez D, Kwarciak A, Quesada V. Next-generation sequencing reveals the secrets of the chronic lymphocytic leukemia genome. *Clin Transl Oncol* 2013; 15 (1): 3-8.
19. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* 2013; 45 (4): 353-61, 361e1-2.
20. Dowty JG, Lee E, McKean-Cowdin R, et al. Estimating risks for variants of unknown significance according to their predicted pathogenicity classes with application to BRCA1. *Breast Cancer Res Treat* 2014; 144 (1): 171-7.

"You cannot shake hands with a clenched fist"

Indira Gandhi (1917-1984), third Prime Minister of India and a central figure of the Indian National Congress