Genes & Diseases (2017) 4, 149–153

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: http://ees.elsevier.com/gendis/default.asp

REVIEW ARTICLE

Novel applications of next-generation sequencing in breast cancer research

Rong Ma, Jianping Gong*, Xiaowei Jiang **

Department of Hepatobiliary Surgery, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, 400010, China

Received 19 March 2017; accepted 2 July 2017
Available online 18 July 2017

KEYWORDS
Breast cancer;
The next generation sequencing technology;
The targeted sequencing;
The whole genome association studies;
The whole genome exon sequencing

Abstract With the rapid development of medicine, the studies of genes have become increasingly concerned by more people and being the contend of a great of researches. The next generation sequencing with its own advantages has been widely used in gene research nowadays. It has almost replaced the traditional sequencing methods (such as Sanger sequencing method), and played an important role in a variety of complex disease researches, including breast cancer. The next generation sequencing technology has the advantages of high speed, high throughput and high accuracy. It has been widely used in various cancers (such as prostate cancer, lung cancer, pancreatic cancer, liver cancer, etc.), especially in breast cancer. Moreover, the use of the next generation sequencing technology to make DNA sequence analysis and risk prediction has made a great contribution to the research of breast cancer. We will focus on the application of whole genome sequencing, exon sequencing and targeted gene sequencing in breast cancer gene research.

With the rapid development of medicine, the studies of diseases have no longer simply stayed at the level of individual, cell and even molecular. Since 2005, the studies of genes have become increasingly concerned by more people with the completion of human genome sequencing. A large number of researchers have been involved in the studies of diseases and genes, and have achieved good results in some complex diseases (such as cancer, autoimmune disease, diabetes, etc.) and common abnormal indexes (such as blood lipids, blood glucose and weight, etc.), discovering a large deal of gene loci associated with the pathogenesis of these diseases and bringing new ideas for the diagnosis and

* Corresponding author.
** Corresponding author.
E-mail addresses: gongjianping11@126.com (J. Gong), 674079699@qq.com (X. Jiang).
Peer review under responsibility of Chongqing Medical University.

http://dx.doi.org/10.1016/j.gendis.2017.07.003

treatment of these diseases. There is no doubt that the results of genetic studies exhilarate researchers in the field of cancer research. Nearly ten years, there have been significant genetic researches in cancers of various systems (such as prostate cancer, lung cancer, breast cancer, liver cancer, etc.).

Breast cancer and research progress of related genes

Breast cancer is one of the most common cancers in human beings, and the incidence among women is the first.[1] So far, the incidence of breast cancer has accounted for 30% of female malignant tumor.[2] Breast cancer is the most common cancer among women of the western developed countries, while it is also increasing yearly in developing countries.[3] Although the incidence of breast cancer in Chinese women is relatively low compared with the women in developed countries, it is increasing year by year with the change of lifestyle and the development of the industry. Even though the breast is less important to maintain the human life being and breast cancer in situ is not fatal, breast cancer often spread through the lymphatic system and it is often accompanied by distant metastasis, threatening life. Therefore, breast cancer brings a great harm to the body and mind of the patients (women account for 99%). The research on breast cancer is increasing both home and abroad, involving many aspects, including the cause of disease, treatment (surgery, radiotherapy, chemotherapy), gene research, and so on. The gene research on breast cancer has rapidly developed especially in recent years. It is found that mutations or deletions of many genes are associated with breast cancer, such as TP53, PTEN(Common somatic mutation), RUNX1, CCND3, PTPN22(the low frequency mutation genes), whose roles and mechanism in breast cancer are still studied.[4] Breast cancer can be divided into four primary subtypes according to its associated estrogen receptor (ER),progesterone receptor (PR), human epidermal growth factor receptor 2(Her-2). Further, the gene-expression subtypes are as following: luminal A, luminal B, Her2-enriched, and basal-like(Details in Table 1). While luminal subtypes are relate to expression of oestrogen and progesterone receptors (ER and PR) and differentiated luminal epithelial cell markers. Among them, triple negative (Basal-like) breast cancer whose estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 are all negative possessed the greatest number of mutations.[5] Therefore, triple negative breast cancer possesses a large proportion in domestic and overseas studies. Breast cancer is a disease of multifactorial inheritance, coming from the mutations of the normal cells which maintain growth. Most of the oncogenes are not inherited directly, on the contrary, they are from mutation.[6] In recent years, the researches on the genes of breast cancer have made great achievements. In particular, the next generation sequencing technology has been used to discover a large number of gene mutation sites, which enriches the research on genes of breast cancer.

The next generation sequencing technology

The traditional method of gene sequencing is Sanger sequencing method, including Double end termination method, Gibert and Maxam's "chemical degradation method", which was the gold standard for gene detection. However, the Sanger sequencing method has the defects of high cost, low throughput, and the deviation of the target, while the next generation sequencing technology has made up the defects of the traditional method.[7] The next generation sequencing technology is the same as high throughput sequencing and deep sequencing. Compared with the traditional Sanger sequencing method, the next generation sequencing technology has the advantages of high speed, high throughput and high accuracy.[8] For example, the next generation sequencing has a higher diagnostic and sequencing sensitivity, D'Argenio Valeria et al. confirmed that the next generation sequencing in the BRCA1/2 mutation gene detection is more sensitive than the traditional Sanger sequencing method in the latest

Table 1 The gene-expression subtypes and related genes of BC.

Subtypes of BC[a]	Definition	Common mutant genes
Luminal A	ER and/or PR+, Her2-, Ki-67weakened (<14%)	LIV-1, XBP1, GATA3, FOXA1, TFF3, ESR1, SCUBE2, αHNF3A, TREFOIL, FACTOR3, TP53 (13%)
Luminal B	Her2-, ER and/or PR+, Ki-67-enriched Her2+++, ER and/or PR+, any Ki-67	LIV-1, XBP1, GATA3, FOXA1, TFF3, ESR1, SCUBE2, TREFOIL, FACTOR3, TP53 (13%), GGH, LAPTMB4, NSEP1, SQLE, ERBB2, GRB7, αHNF3A
Her2-enriched	Her2+++, ER and PR-	ERBB2, GRB7, TRAP100, PPARBP, TP53 (40—80%)
Basal-like (Triple negative)	ER-, PR-, Her2-	KRT5, KRT17, ANNEXIN8, CX3CL1, TRIM29,c-KIT,FOXC1, P-Cadherin, LAMININ, FABP, CK5/6, CK1, BRCA1, TP53 (75%)

[a] BC: breast cancer.

article, and it will replace the traditional method in the study of this gene.[9,10] As we all known, there are hereditary susceptibility in two leading genes of breast cancer, BRCA1 and BRCA2. Lately, our scholars have used the next generation sequencing to identified a novel mutation of BRCA2: c.8946_8947delAG (p.D2983FfsX34) in a Chinese woman.[11] But in terms of the molecular diagnostic rate, exome sequencing of the next generation sequencing is higher than that of the chromosome analysis, microarray analysis, Sanger single gene sequencing and other traditional methods.[12] Moreover, the next generation sequencing technology no longer need *Escherichia coli* to being in vivo amplification, but synthesis in vitro sequencing by ligase or polymerase directly. Among them, Illumina/Solexa technology, Roche/LS454 technology, ABI/SOLiD technology and Helicos technology are the main technology.[13] The technology Illumina/Solexa is often used in GWAS (the whole genome association studies, that is to identifying sequence variation associated with diseases in the range of human whole genome—single nucleotide polymorphisms, SNPs) and exon sequencing (That is to sequence a large number of exons, and then study the areas related diseases and verify SNPs.), which has the advantages of small deviation in detecting homopolymer, simple operation and low cost.

Application of next generation sequencing in breast cancer research

In the past 10 years, a large number of researchers used the next generation sequencing technology to interpret breast cancer genome diversity, gradually revealing genomic landscape of breast cancer. The next generation sequencing in breast cancer research is mainly used in the following three aspects: genome DNA sequence analysis (including the whole genome sequencing, exon sequencing, targeting gene sequencing), RNA transcription group sequencing (including the whole transcriptome analysis, smallRNA sequencing, noncoding RNA analysis), epigenetic sequencing (including chromatin immunoprecipitation sequencing, methylation analysis sequencing). The next generation sequencing plays an important role in the study of RNA transcription group and extranuclear heredity, for example, related studies have demonstrated that specific miRNA abnormalities are associated with specific types of breast cancer, such as miRNA, which is associated with invasive papillary carcinoma.[14] Many miRNAs, such as miR-10b, miR-9, miR-31, miR-126, and miR-335 etc have been confirmed to be associated with breast cancer metastasis, while triple negative breast cancer lacks a unique miRNA expression.[15] We don't discuss the application above, the focus is on the application of whole genome sequencing, exon sequencing and targeted gene sequencing in breast cancer gene research.

The whole genome association studies (GWAS) and application of risk prediction in breast cancer

The whole genome association studies (GWAS) is an important method in the study of human genes, nearly 2000 SNPs or gene loci were found and verified in a variety of common diseases in the first five years since 2005. In addition to the advantages of the next generation sequencing as noted above, the whole genome association study also has the advantages of less sample requirements and low frequency mutation etc. Epidemiologic studies of breast cancer have identified some risk factors of environmental and lifestyle (for example, age of menarche and menopause age, the age of first birth, body mass index (BMI) and exogenous hormone use), and more and more evidence has displayed that in addition to the known risk factors above, breast cancer is related to the existence of the risk gene closely.[16] Since 2007, a large scale of GWAS has reported 5 risk loci are associated with breast cancer, there have already more than 90 risk loci being confirmed to be associated with the occurrence of breast cancer, and it is predicted that at least more than 1000 risk loci associated with breast cancer genes are not known yet.[17] Moreover, in addition to the two studies in Africa and Asia, the rest findings of studies are from women in European ancestry. Fang Chen et al.[18] confirmed that two gene loci (rs4322600 at 14q31; rs10510333 at 3p26) were closely related to the occurrence of breast cancer in African women in the whole genome association studies of female breast cancer in African ancestry. 5 years later, Cai Qiuyin et al. identified that three gene loci were related to breast cancer of Asian women in the whole genome association studies of breast cancer in East Asia.[19] Although the single SNP identified by the whole genome association studies (GWAS) has little effect in the mechanism of breast cancer occurrence, it should not be overlooked that the role of multiple SNP superposition effect is played in breast cancer risk prediction. As early as the year 2007, J. Hunter David et al. had verified that there was high correlation between a group of 4 SNP of intron 2 of FGFR2 (tyrosine kinase receptor) and the occurrence of breast cancer.[20] The latest research shows that most of the detected genetic risk loci are associated with ER positive breast cancer, while only eight of the genes are associated with ER negative breast cancer, which are 1q32.1 (*MDM4*, $P = 2.1 \times 10^{-12}$ and *LGR6*, $P = 1.4 \times 10^{-8}$), 2p24.1 ($P = 4.6 \times 10^{-8}$), 16q12.2 (*FTO*, $P = 4.0 \times 10^{-8}$), 5p15.33 (*TERT*, $P = 1.0 \times 10^{-1}$), 20q11.22 (*RALY*, $P = 1.1 \times 10^{-8}$), 16q12.2 (*FTO*, $P = 4.0 \times 10^{-8}$ and 6q25.1 (*TAB2*).[21] And, BRCA1-associated breast cancers are usually highly proliferative and TP53-mutated, and usually lack expression of ESR1 and ERBB2(Her2 is also known as ERBB2). Moreover, multiple novel mutations were identified by NGS in patients of familial Breast/Ovarian Cancer. Other than the gene *BRCA1/2*, *PALB2* was the second most frequently mutated gene according to the research above.[22] Although many genes have been confirmed to be associated with breast cancer, the incidence of breast cancer is the result of many factors. However, the proportion of the risk genes in the carcinogenic factors is still limited, and the mechanism needs to be further explored.

Application of whole genome exon sequencing in breast cancer research

Although the region of encoding proteins by exon only accounts for 1% in the human genome, 85% of the mutations meaningful to human locate here.[23] Therefore, the whole

genome exon sequencing has great significance for the detection of gene and the risk prediction of malignant tumor. In recent years, it has been used in breast cancer research and promoted the research of breast cancer gene. Previous studies have demonstrated that women with BRCA1 or BRCA2 mutations carry a risk of breast cancer which is 50%—80%.[24] The latest study using exon sequencing found that the rare male breast cancer risk was closely related to BRCA2 gene deletion.[25] Although the patients of breast cancer are mainly distributed in the population, 15% of them are in the form of family aggregation. Thompson et al. have demonstrated in the study that the rare mutations of DNA repair gene FANCC and BLM are the potential susceptibility alleles of breast cancer, and this outcome is from the method whole genome exon sequencing.[26] Johanna I. Kiiski et al.[27] have also discovered the gene FANCM is a susceptible gene locus of triple negative breast cancer by using the exon sequencing in a family research of breast cancer in Finland. This year, Noh et al.[28] also found that 7 mutations (XCR1,DLL1, TH, ACCS, SPPL3, CCNF and SRL) may be associated with the potential breast cancer by the same method exon sequencing in a study of familial breast cancer with negative BRCA. Recently, Sarah L Maguire et al.[29] found that the mutation of SF3B1 was proved to be a new target for anticancer therapy by the method which was the combination of the exon sequencing and the whole genome sequencing. The discovery of these mutations not only enriched the research of breast cancer genes, but also brought new ideas for the diagnosis and treatment of breast cancer. In addition, exon sequencing can also be used widely for the researches of many aspects in patients with breast cancer, such as the differentiation of stem cells, genetic susceptibility, and so on.

The application of targeted sequencing in breast cancer research

Targeted sequencing, which targets at a few genes or genomic regions to carry out targeted sequencing analysis. In the study of breast cancer, it primarily do some further targeted researches of sequencing on gene loci which has already been found. For example, MED 12 are the gene known to be related to breast cancer, Then Chieko Mishima et al. make a comparison by targeted sequencing between the mutation volume of mammary gland fibroma and phyllodes tumors that the mutation volume of phyllodes tumors is higher than mammary gland fibroma apparently.[30] In addition, a latest study conducted among women in southern India shows that targeted sequencing can effectively improve the detection rate of deleterious mutation genes of breast cancer and/or ovarian cancer.[31] Meanwhile, targeted sequencing also improves the accuracy of breast cancer targeted therapy. In short, its application is an important supplement to breast cancer research following the whole genome sequencing and exon sequencing.

Outlook

The technology of the next generation sequencing has played an important role in detecting risk gene loci of breast cancer. Compared to gene chip, they do not rely on the information of previous sequencing used for microarray probes, and the expression level is not segmented, but based on the entire text to provide a higher confidence in the quantitative analysis of gene expression on the level of the whole genome.[32] The discovery of the meaningful mutation sites in the pathogenesis of breast cancer can be used for the further study on the levels of molecule, cell, protein and animal. The discovered molecular markers (such as Her-2, ER, D1, PR, cyclin, p53, Bcl-2 and so on) are closely related to the prognosis and treatment of breast ductal carcinoma in situ. It can be combined with the follow-up data of patients with clinical prognosis research of breast cancer, whose prospects are not to be underestimated. However, it is difficult to find out the new mutation sites by sequencing. Even though we set aside the above difficulties temporarily and are successful to discover the mutation site, one of the most prominent features of malignant tumor is malignant proliferation and extremely easy to mutate. Therefore, the mutation site found by sequencing is extremely likely to mutate before that is successful in clinical verification, leading to the work in vain. The whole genome association studies (GWAS) is widely applied in the research of gene locus in breast cancer, but it possesses its own disadvantages that it is prone to false negative or false positive, not sensitive to find rare variants, and SNPs found by GWAS often located in intergene or introns. Therefore, in future studies, the combination of gene exon sequencing and the whole genome association studies is a complementation of advantages and disadvantages. In short, the next generation sequencing technology is one of the main methods of gene research, and it has replaced the traditional sequencing methods. Although there are still many shortcomings, with the progress of science and technology, we will gradually overcome its shortcomings to apply genetic research to research, diagnosis and treatment of diseases.

Conflicts of interest

There is no conflict of interest.

Acknowledgments

The junior author would like to thank her mentor for allowing her participation in this review.

References

1. Lindström S, Ablorh A, Chapman B, et al. Deep targeted sequencing of 12 breast cancer susceptibility regions in 4611 women across four different ethnicities. *Breast Cancer Res*. 2016;18(1):109.
2. Wang Z, Liao H, Deng Z, et al. MiRNA-205 affects infiltration and metastasis of breast cancer. *Biochem Biophys Res Commun*. 2013;441(1):139—143.
3. Youlden DR, Cramb SM, Dunn NA, et al. The descriptive epidemiology of Female breast cancer: an international comparison of screening, incidence, survival and mortality. *Cancer Epidemiol*. 2012;36(3):237—248.

4. Banerji S, Cibulskis K, Rangel-escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*. 2012;486(7403):405—409.
5. Wang Y, Waters J, Leung ML, et al. Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*. 2014;512(7513):155—160.
6. Rich TA, Woodson AH, Litton J, et al. Hereditary breast cancer syndromes and genetic testing. *J Surg Oncol*. 2015;111(1):66—80.
7. Moorcraft SY, Gonzalez D, Walker BA. Understanding next generation sequencing in oncology: a guide for oncologists. *Crit Rev Oncol Hematol*. 2015;96(3):463—474.
8. Fan H, Ives AR, Surget-groba Y, et al. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. *BMC Genomics*. 2015;16(1):522.
9. D'argenio V, Frisso G, Precone V, et al. DNA sequence capture and next-generation sequencing for the molecular diagnosis of genetic cardiomyopathies. *J Mol Diagn*. 2014;16:32—44.
10. D'arenio V, Esposito MV, Telese A, et al. The molecular analysis of BRCA1 and BRCA2: next-generation sequencing supersedes conventional approaches. *Clin Chim Acta*. 2015; 446:221—225.
11. Ma J, Yang J, Jian W, et al. A novel loss-of-function heterozygous BRCA2 c.8946_8947delAG mutation found in a Chinese woman with family history of breast cancer. *J Cancer Res Clin Oncol*. 2017;143(4):631—637.
12. Ahmadloo S, Nakaoka H, Hayano T, et al. Rapid and cost-effective high-throughput sequencing for identification of germline mutations of BRCA1 and BRCA2. *J Hum Genet*. 2017; 62(5):561—567.
13. Geskin A, Legowskie E, Chakka A, et al. Needs assessment for research use of high-throughput sequencing at a large academic medical center. *PLoS One*. 2015;10(6), e0131166.
14. Li S, Yang C, Zhai L, et al. Deep sequencing reveals small RNA characterization of invasive micropapillary carcinomas of the breast. *Breast Cancer Res Treat*. 2012;136(1):77—87.
15. Farazi TA, Horlings HM, Ten Hoeve JJ, et al. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. *Cancer Res*. 2011;71(13):4443—4453.
16. Lin Y, Fu F, Chen M, et al. Associations of two common genetic variants with breast cancer risk in a Chinese population: a stratified interaction analysis. *PLoS One*. 2014;9(12), e115707.
17. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447:1087—1093.
18. Chen F, Chen GK, Stram DO, et al. A genome-wide association study of breast cancer in women of African ancestry. *Hum Genet*. 2013;132(1):39—48.
19. Cai Q, Zhang B, Sung H, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. *Nat Genet*. 2014;46(8):886—890.
20. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007; 39(7):870—874.
21. Fachal L, Dunnig AM. From candidate gene studies to GWAS and post-GWAS analyses in breast cancer. *Curr Opin Genet Dev*. 2015;30C:32—41.
22. Kraus C, Hoyer J, Vasileiou G, et al. Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int J Cancer*. 2017;140(1):95—102.
23. Kaur H, Mao S, Shah S, et al. Next-Generation Sequencing: a powerful tool for the discovery of molecular markers in breast ductal carcinoma in situ. *Expert Rev Mol Diagn*. 2013;13(2): 151—165.
24. Pan H, He Z, Ling L, et al. Reproductive factors and breast cancer risk among BRCA1 or BRCA2 mutation carriers: results from ten studies. *Cancer Epidemiol*. 2014;38(1):1—8.
25. Timoteo AR, Albuquerque BM, Moura PC, et al. Identification of a new BRCA2 large genomic deletion associated with high risk male breast cancer. *Hered Cancer Clin Pract*. 2015;13(1):2.
26. Thompson ER, Doyle MA, Ryland GL, et al. Exome sequencing identifies rare deleterious mutations in DNA repair genes FANCC and BLM as potential breast cancer susceptibility alleles. *PLoS Genet*. 2012;8(9), e1002894.
27. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. *Proc Natl Acad Sci USA*. 2014;111(42):15172—15177.
28. Noh JM, Kim J, Cho DY, et al. Exome sequencing in a breast cancer family without BRCA mutation. *Radiat Oncol J*. 2015; 33(2):149—154.
29. Maguire SL, Leonidou A, Wai P, et al. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J Pathol*. 2015;235(4):571—580.
30. Mishima C, Kagara N, Tanei T, et al. Mutational analysis of MED12 in fibroadenomas and phyllodes tumors of the breast by means of targeted next-generation sequencing. *Breast Cancer Res Treat*. 2015;152(2):305—312.
31. Rajkumar T, Meenakumari B, Mani S, et al. Targeted resequencing of 30 genes improves the detection of deleterious mutations in south indian women with breast and/or ovarian cancers. *Asian Pac J Cancer Prev*. 2015;16(13):5211—5217.
32. Kim K, Seong MW, Chung WH, et al. Effect of next-generation exome sequencing depth for discovery of diagnostic variants. *Genomics Inf*. 2015;13(2):31—39.