

Exome sequencing in a breast cancer family without *BRCA* mutation

Jae Myoung Noh, MD^{1*}, Jihun Kim, PhD^{2*}, Dae Yeon Cho, PhD²,
Doo Ho Choi, MD, PhD¹, Won Park, MD, PhD¹, Seung Jae Huh, MD, PhD¹

¹Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul;

²LabGenomics Clinical Research Institute, LabGenomics, Seongnam, Korea

Purpose: We performed exome sequencing in a breast cancer family without *BRCA* mutations.

Materials and Methods: A family that three sisters have a history of breast cancer was selected for analysis. There were no family members with breast cancer in the previous generation. Genetic testing for *BRCA* mutation was negative, even by the multiplex ligation-dependent probe amplification method. Two sisters with breast cancer were selected as affected members, while the mother of the sisters was a non-affected member. Whole exome sequencing was performed on the HiSeq 2000 platform with paired-end reads of 101 bp in the three members.

Results: We identified 19,436, 19,468, and 19,345 single-nucleotide polymorphisms (SNPs) in the coding regions. Among them, 8,759, 8,789, and 8,772 were non-synonymous SNPs, respectively. After filtering out 12,843 synonymous variations and 12,105 known variations with indels found in the dbSNP135 or 1000 Genomes Project database, we selected 73 variations in the samples from the affected sisters that did not occur in the sample from the unaffected mother. Using the Sorting Intolerant From Tolerant (SIFT), PolyPhen-2, and MutationTaster algorithms to predict amino acid substitutions, the *XCR1*, *DLL1*, *TH*, *ACCS*, *SPPL3*, *CCNF*, and *SRL* genes were risky among all three algorithms, while definite candidate genes could not be conclusively determined.

Conclusion: Using exome sequencing, we found 7 variants for a breast cancer family without *BRCA* mutations. Genetic evidence of disease association should be confirmed by future studies.

Keywords: Breast neoplasms, Exome, BRCA

Introduction

Although *BRCA1* and *BRCA2* mutations are well-known predispositions for breast cancer [1,2], they account for less than 20% of all familial breast cancer cases [3]. In addition to these high-penetrance breast cancer predisposition genes, moderate-penetrance breast cancer susceptibility genes,

such as *ATM*, *BRIP1*, *CHEK2*, and *PALB2*, are associated with 2- to 4-fold relative risk of breast cancer [3]. Genome-wide association studies have identified common low-penetrance breast cancer susceptibility alleles [4]. However, only 35% of the familial risk of breast cancer is explained by these high-to-low-penetrance susceptibility alleles, and a large proportion of the genetic contribution to breast cancer remains unexplained

Received 12 May 2015, Revised 8 June 2015, Accepted 9 June 2015.

Correspondence: Doo Ho Choi, MD, PhD, Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 135-710, Korea. Tel: +82-2-3410-2436, Fax: +82-2-3410-2619, E-mail: doho.choi@samsung.com

*These authors contributed equally to this work.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

www.e-roj.org

[3,5].

Exome sequencing is a technique used to sequence protein-coding regions, which constitute only 1% of the human genome, but account for 85% of known disease-related mutations. This technique can be applied to various human diseases including Mendelian diseases, and could be a promising strategy for identifying new genes associated with an increased risk of breast cancer [5-7], especially in BRCA-negative familial breast cancer.

In addition to younger age at onset of breast cancer [8], there are several distinctive features of breast cancer in Korea [9-11]. Unlike the Ashkenazi-Jewish population, highly recurrent founder mutations have not been detected, while the *BRCA2* c.7480c>T mutation has been suggested as a candidate founder mutation in Korea [9]. Some moderate-penetrance breast cancer susceptibility alleles, such as *PALB2* 1592delT and 229delT and *CHEK2* 1100delC, are not present in Korean patients [10,11]. Although the Korean Hereditary Breast Cancer (KOHBRA) study estimated the nationwide prevalence of *BRCA* mutations among a high-risk group of patients with hereditary breast cancer [12], there has been no study using exome sequencing in familial breast cancer. As a preliminary study, we performed exome sequencing in a breast cancer family without *BRCA* mutations.

without breast cancer had a history of thyroid cancer. There were no members with breast or thyroid cancer in the previous generation. The second sister underwent genetic testing for *BRCA* and *BRAF* mutations and *BRAF*^{V600E} mutation, a somatic mutation [13], was detected. However, no mutation was detected in the *BRCA1* and *BRCA2* genes, even by the multiplex ligation-dependent probe amplification method.

We applied the linkage strategy introduced previously [7]. The second and fourth sisters with breast cancer were sequenced as affected members, while the mother of the sisters was sequenced as a non-affected member.

2. Whole exome sequencing

After informed consent was acquired, whole exome sequencing was performed through an exome sequencing service from MacroGen (South Korea), and the human exome capture by the Agilent V4+UTRs exome enrichment kit was used. Next-generation sequencing followed, using the Illumina HiSeq 2000 sequencer for captured DNA as paired end reads (100 bp). Data analysis was conducted by the MacroGen exome sequencing pipeline.

All samples were sequenced to provide mean sequence coverage of more than 150x, with more than 95% of the target bases having at least 10x coverage. Approximately 99.7% of all initial mappable reads were able to pass our thresholds for SNPs and short insertions or deletions (indels). The mean read depth of the target regions was 82.4x, 84x, and 91.4x, respectively. The throughput of the exome sequence is summarized in Table 1.

Materials and Methods

1. Patients

A Korean family that three sisters have a history of breast cancer was selected for analysis (Fig. 1). The second sister had simultaneous breast and thyroid cancer, and another sister

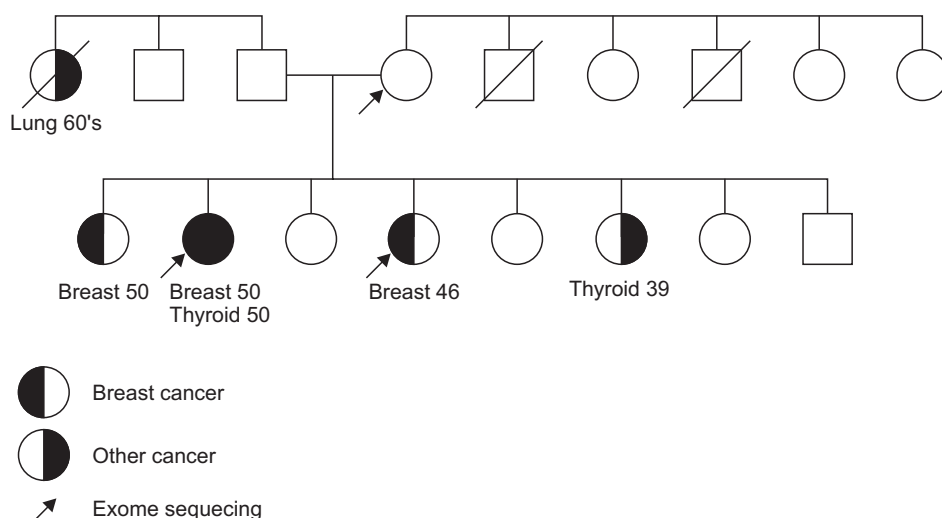


Fig. 1. The familial pedigree of the sequenced family. Three sisters had a history of breast cancer, while the second sister had simultaneous thyroid cancer. This sister received genetic testing for *BRCA* and *BRAF* mutations, and was positive for the *BRAF*^{V600E} mutation, but negative for the *BRCA* mutation. There were no family members with breast or thyroid cancer in the previous generation.

Table 1. The throughput of the exome sequence in this study

| Data | Sister #2 | Sister #4 | Mother | Average |
|--|----------------|----------------|----------------|----------------|
| Total reads | 110,987,862 | 114,589,014 | 111,186,892 | 112,254,589 |
| Total yield (bp) | 11,209,774,062 | 11,573,490,414 | 11,229,876,092 | 11,337,713,523 |
| Average throughput depth of target regions (X) | 157.3 | 162.4 | 157.6 | 159.1 |
| Initial mappable reads (mapped to human genome) | 110,741,070 | 114,218,994 | 110,931,930 | 111,963,998 |
| Non-redundant unique reads (uniquely mapped to human genome) | 79,141,236 | 82,826,069 | 88,347,648 | 83,438,318 |
| On-target reads (=reads mapped to target regions) | 65,569,825 | 66,867,301 | 72,737,605 | 68,391,577 |
| % On-target reads (out of non-redundant unique reads) | 82.9 | 80.7 | 82.3 | 82.0 |
| % Coverage of target regions (more than 10X) | 96.4 | 96.4 | 96.7 | 96.5 |
| Number of on-target genotypes (more than 10X) | 68,690,302 | 68,687,989 | 68,904,167 | 68,760,819 |
| Mean depth of target regions (X) | 82.4 | 84 | 91.4 | 85.9 |
| Number of SNPs | 84,942 | 84,868 | 84,876 | 84,895 |
| Number of coding SNPs | 19,436 | 19,468 | 19,345 | 19,416 |
| Number of synonymous SNPs | 10,181 | 10,194 | 10,087 | 10,154 |
| Number of non-synonymous SNPs | 8,759 | 8,789 | 8,772 | 8,773 |
| Number of indels | 9,678 | 9,649 | 9,715 | 9,681 |
| Number of coding indels | 429 | 451 | 426 | 435 |

SNP, single-nucleotide polymorphism; indel, insertional-deletional variations.

3. Variant calling and analysis

Sequence variants including single-nucleotide variations (SNVs) and small insertional-deletional variations (indels) were called using SAMtools (ver. 0.1.18). The SNVs and indels were annotated by the ANNOVAR program (ver. November 2011) [14], where the files were converted to their input format, variations within exonic or splicing regions were selected, and synonymous SNVs were filtered out. Then, an in-house program was applied to filter out the indels and additional SNVs that have been reported in common variant databases including dbSNP135 or 1000 Genomes SNP call release (20101109, 628 individuals) [15]. Lastly, all selected variants were annotated using dbNSFP [16], which is a database developed for functional prediction and annotation of all potential non-synonymous single-nucleotide variants (nsSNVs) in the human genome.

We used Sorting Intolerant From Tolerant (SIFT), PolyPhen-2, and MutationTaster to assess the impact of mutations on protein function. SIFT examines the degree of conservation for amino acid residues across species, and PolyPhen-2 finds change in protein structure and function. MutationTaster checks both, and additionally looks at effects on splicing or mRNA expression.

Results

We identified 19,436, 19,468, and 19,345 SNPs in the coding

regions, respectively (8,759, 8,789, and 8,772 non-synonymous SNPs). After identification of variants, we merged variations from the three samples, and then extracted 124,440 total variations. We focused only on the relevant variation for each phenotype, including exonic variants, which are variants that overlap a coding exon, and splicing variants, which are variants within 2 bp of a splicing junction. A total of 25,435 variants remained after selection process. We then filtered out an additional 12,843 synonymous variations, and 12,105 known variations, which are represented in the dbSNP135, 1000 Genomes Project with indels. Finally, we selected 73 variations that did not exist in the sample from the unaffected mother, but were in the samples from the affected sisters by assuming a dominant genetic model, and we identified 64 genes that contained these variations.

The missense prediction programs, SIFT, PolyPhen-2, and MutationTaster were used and the different scores from these tools were derived for the final 73 candidate variants. SIFT predicts whether an amino acid substitution in a protein will have a phenotypic effect, and the score less than 0.05 are predicted to be deleterious, those greater than or equal to 0.05 are predicted to be tolerated. PolyPhen-2 predicts the functional significance of an allele replacement, and the score more than 0.85 are interpreted as probably damaging and scores 0.15–0.85 as possibly damaging. MutationTaster predicts the disease potential of an alteration, and the predicted value close to 1 indicates a high 'security' of the prediction.

Results from the three different tools were compared, and we identified that 7 variants were reported risky among all three algorithms: *XCR1*, *DLL1*, *TH*, *ACCS*, *SPPL3*, *CCNF*, and *SRL* (Fig. 2, Table 2). The variants were confirmed by Sanger sequencing.

Discussion and Conclusion

In the current study, we found 7 variants that could affect protein function through exome sequencing followed by subsequent filtering and selection by SIFT, Polyphen-2, and MutationTaster. We searched literatures regarding potential relationship between the variants and breast cancer. Among the 7 variants, 3 variants were related to breast cancer. The *XCR1*, a chemokine receptor belonging to the G protein-coupled receptor superfamily, has been known to be involved in cytotoxic immune response [17,18]. With regard to breast cancer, Gantsev et al. [19] demonstrated the increase in

expression of the genes *CCL16*, *XCR1*, *CYFIP2*, and *TNFSF14* in newly formed lymph nodes in breast cancer. The *DLL1* gene encodes for delta-like protein 1, which acts as a ligand for Notch receptors that engage in oncogenic conversion of human breast epithelial cells [20]. Furthermore, inhibition of Notch signaling is suggested to be beneficial for breast cancer [21]. The *CCNF*, a member of the cyclin family, is known to be involved in mitosis and genome integrity during the G2 phase of the cell cycle in association with CP110 [22]. Roy et al. [23] demonstrated that etodolac, a member of the cyclooxygenase inhibitor, altered 6 cell cycle regulatory protein genes of mammary epithelial cells, including *CCNF*.

Several genes are related to breast cancer as described above, and could be potential candidates for breast cancer predisposition. However, definite candidate genes could not be conclusively determined because only one family with two affected members was examined. However, this preliminary

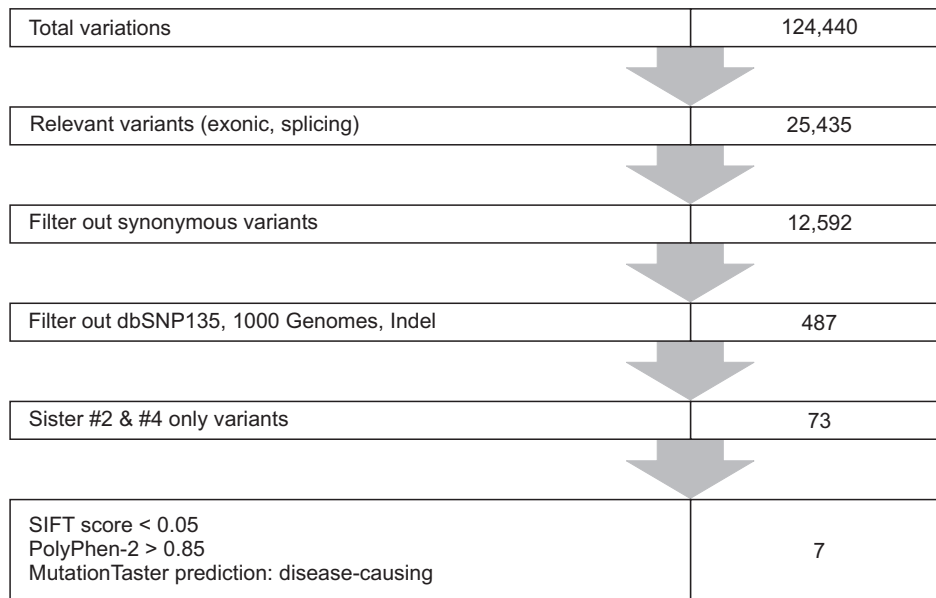


Fig. 2. Number of variants filtered against several criteria determining causative alleles. SIFT, Sorting Intolerant From Tolerant.

Table 2. Candidate variants shared by affected individuals and selected by SIFT, PolyPhen-2, and MutationTaster algorithms

| Chr:Position ^{a)} | Gene | Reference sequence | Sequence variation | Protein variation | SIFT score | PolyPhen-2 score | MutationTaster prediction |
|----------------------------|--------------|--------------------|--------------------|-------------------|------------|------------------|---------------------------|
| chr03:46063060 | <i>XCR1</i> | NM_005283 | c.380G>A | p.R127H | 0 | 1 | Disease-causing |
| chr06:170594187 | <i>DLL1</i> | NM_005618 | c.1069G>A | p.G357S | 0.03 | 0.998 | Disease-causing |
| chr11:2188215 | <i>TH</i> | NM_000360 | c.743G>C | p.C248S | 0 | 1 | Disease-causing |
| chr11:44102802 | <i>ACCS</i> | NM_032592 | c.1043C>T | p.S348F | 0.03 | 0.916 | Disease-causing |
| chr12:121206820 | <i>SPPL3</i> | NM_139015 | c.545G>C | p.R182P | 0.02 | 1 | Disease-causing |
| chr16:2503415 | <i>CCNF</i> | NM_001761 | c.1592T>G | p.L531R | 0.01 | 0.988 | Disease-causing |
| chr16:4253212 | <i>SRL</i> | NM_001098814 | c.214G>A | p.E72K | 0 | 0.998 | Disease-causing |

SIFT, Sorting Intolerant From Tolerant.

^{a)}Genomic position from HG build 19.

study could help explain non-*BRCA* familial breast cancer. Lynch et al. [24] suggested that genetic predisposition for familial breast cancer could be family-specific, which would be difficult to detect using a population-based approach. Therefore, exome sequencing data from a single family could be valuable for determining genetic predisposition.

Future studies should seek to identify candidate families who could benefit from exome sequencing. According to the KOHBRA study, the prevalence of *BRCA* mutations among familial breast cancer probands was 21.7% [12]. The remaining 78.3% of the familial breast cancer probands who were negative for *BRCA* mutations could be potential candidates for exome sequencing. In addition to a family history of breast cancer, Han et al. [12] suggested that age at diagnosis (<50 years) should also be considered when selecting patients for genetic testing. In this regard, genetic anticipation, a phenomenon of early onset of disease in subsequent generations, should also be considered, because the generational difference in age at diagnosis could be attributed to genetic predisposition, with or without *BRCA* mutation [25]. Exome sequencing could explain the actions of moderate to low penetrance susceptibility alleles of non-*BRCA* breast cancer families, which are considered high risk for breast cancer and include several affected women across generations [6].

Several limitations should be considered in regard to the current study. First, only one affected breast cancer patient, the second sister, underwent genetic testing for *BRCA* mutations. Second, distant affected young relative and nearby unaffected old relative should be added to the analysis to validate susceptible variants. But the selection of the family was not satisfactory for such analysis. Third, we chose the unaffected mother rather than the breast cancer-free sister. Since the sister relations are likely to be genetically identical, the unaffected sister should be considered to be chosen.

In summary, we found 7 variants by exome sequencing for breast cancer family without *BRCA* mutations. The *XCR1*, *DLL1*, *TH*, *ACCS*, *SPPL3*, *CCNF*, and *SRL* genes could be potential candidates for breast cancer predisposition. Genetic evidence of disease association should be confirmed by validation through additional non-*BRCA* breast cancer families and comparison with general population.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

The present research has been supported by the Korea Breast Cancer Foundation (12-02) and supported by a grant of the Korea Health Technology R&D Project through KHIDI, funded by the Ministry of Health & Welfare, Republic of Korea (No. HI13C2096).

References

1. Wooster R, Bignell G, Lancaster J, et al. Identification of the breast cancer susceptibility gene *BRCA2*. *Nature* 1995;378:789-92.
2. Miki Y, Swensen J, Shattuck-Eidens D, et al. A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* 1994;266:66-71.
3. Turnbull C, Rahman N. Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet* 2008;9:321-45.
4. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007;447:1087-93.
5. Snape K, Ruark E, Tarpey P, et al. Predisposition gene identification in common cancers by exome sequencing: insights from familial breast cancer. *Breast Cancer Res Treat* 2012;134:429-33.
6. Gracia-Aznarez FJ, Fernandez V, Pita G, et al. Whole exome sequencing suggests much of non-*BRCA1/BRCA2* familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLoS One* 2013;8:e55681.
7. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 2012;20:490-7.
8. Ahn SH, Yoo KY. Chronological changes of clinical characteristics in 31,115 new breast cancer patients among Koreans during 1996-2004. *Breast Cancer Res Treat* 2006;99:209-14.
9. Kim H, Cho DY, Choi DH, et al. Characteristics and spectrum of *BRCA1* and *BRCA2* mutations in 3,922 Korean patients with breast and ovarian cancer. *Breast Cancer Res Treat* 2012;134:1315-26.
10. Kim JH, Choi DH, Cho DY, Ahn SH, Son BH, Haffty BG. *PALB2* mutations 1592delT and 229delT are not present in Korean breast cancer patients negative for *BRCA1* and *BRCA2* mutations. *Breast Cancer Res Treat* 2010;122:303-6.
11. Choi DH, Cho DY, Lee MH, et al. The *CHEK2* 1100delC mutation is not present in Korean patients with breast cancer cases tested for *BRCA1* and *BRCA2* mutation. *Breast Cancer Res Treat* 2008;112:569-73.
12. Han SA, Kim SW, Kang E, et al. The prevalence of *BRCA* mutations among familial breast cancer patients in Korea: results of the Korean Hereditary Breast Cancer study. *Fam*

- Cancer 2013;12:75-81.
13. Jeong D, Jeong Y, Park JH, et al. BRAF (V600E) mutation analysis in papillary thyroid carcinomas by peptide nucleic acid clamp real-time PCR. *Ann Surg Oncol* 2013;20:759-66.
 14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164.
 15. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491:56-65.
 16. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat* 2011;32:894-9.
 17. Lei Y, Takahama Y. XCL1 and XCR1 in the immune system. *Microbes Infect* 2012;14:262-7.
 18. Yoshida T, Imai T, Kakizaki M, Nishimura M, Takagi S, Yoshie O. Identification of single C motif-1/lymphotactin receptor XCR1. *J Biol Chem* 1998;273:16551-4.
 19. Gantsev SK, Umezawa K, Islamgulov DV, et al. The role of inflammatory chemokines in lymphoid neorganogenesis in breast cancer. *Biomed Pharmacother* 2013;67:363-6.
 20. Ayyanan A, Civenni G, Ciaroni L, et al. Increased Wnt signaling triggers oncogenic conversion of human breast epithelial cells by a Notch-dependent mechanism. *Proc Natl Acad Sci U S A* 2006;103:3799-804.
 21. Han J, Hendzel MJ, Allalunis-Turner J. Notch signaling as a therapeutic target for breast cancer treatment? *Breast Cancer Res* 2011;13:210.
 22. D'Angiolella V, Donato V, Vijayakumar S, et al. SCF (Cyclin F) controls centrosome homeostasis and mitotic fidelity through CP110 degradation. *Nature* 2010;466:138-42.
 23. Roy D, Arason GA, Chowdhury B, Mitra A, Calaf GM. Profiling of cell cycle genes of breast cells exposed to etodolac. *Oncol Rep* 2010;23:1383-91.
 24. Lynch H, Wen H, Kim YC, et al. Can unknown predisposition in familial breast cancer be family-specific? *Breast J* 2013;19:520-8.
 25. Noh JM, Choi DH, Baek H, et al. Genetic anticipation of familial breast cancer with or without BRCA mutation in the Korean population. *Cancer Genet* 2014;207:160-3.