Media Farmasi: Jurnal Ilmu Farmasi (Journal of Pharmaceutical Science) e-ISSN 2503-5223
Vol. 19, No. 1, April 2024, pp. xx-xx
http://journal.uad.ac.id/index.php/Media-Farmasi/index

PROFILE MAPPING OF GENE VARIATIONS ASSOCIATED WITH BREAST CANCER AND ITS VULNERABILITY IN THE WORLD POPULATION USING A GENOMIC DATABASE

Putri Dira Nabilla[2], **Lalu Muhammad Irham**[1]

[1] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2] Undergraduate Study Program in Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
* corresponding author: putridiranabilla@gmail.com, No Wa: 085156236620

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Breast cancer, genes, genomics, SNP, gene variations.

ABSTRACT

Breast cancer is a disease that globally is the cause of cancer deaths in women. One of the risk factors for breast cancer is genetic factors. The discovery and reporting of specific gene variations of SNPs with missense mutations in breast cancer will be useful for the development and discovery of new drugs through the biomarkers obtained. So, researchers are interested in identifying the types of gene variations that are at risk for breast cancer by using a genomic database. The method used in the research is non-experimental by utilizing secondary data sources. Genomic databases accessed in this study include; Genome Wide Association Studies (GWAS) Catalog, HaploReg v4.2, Genotype Tissue Expression (GTEx) Portal, and Ensembl Genome Browser. SNP (Single Nucleotide Polymorphism) related to breast cancer was obtained through the GWAS catalog database of 2728 SNPs then filtered using the criteria of Odds Ratio > 1 and P-value < 5 x 10-8 to obtain 762 SNPs (accessed on 7/3/2024). Then the SNPs were processed using HaploReg v4.2 to obtain 10 missense SNPs (accessed on 9/3/2024). Gene expression in mammary tissue and other tissues that may have an influence is seen using the GTEx Portal (accessed on 13/03/2024) and the distribution of alleles in the world population is seen on Ensembl (accessed 18/03/2024). There are 10 missense allele locations, namely rs3184504 (SH2B3), rs3206824 (DKK3), rs16991615 (MCM8), rs6929137 (CCDC170), rs35383942 (PHLDA3), rs6964587 (AKAP9), rs1053338 (ATXN7), rs17879 961 (CHEK2), rs2363956 (ANKLE1), rs11552449 (DCLRE1B). The gene variants with the highest expression in mammary tissue are the PHLDA3 and DKK3 genes. The population and SNP estimated to be most susceptible in Africa, East Asia and South Asia is rs3184504 (Allele C) and in America and Europe it is rs3206824 (Allele C). It is hoped that these gene variations can become biomarkers to detect the risk of breast cancer and help in personalized medicine, namely an approach that adapts drug therapy to the individual genetic characteristics of breast cancer patients. Keywords: Breast cancer, genes, genomics, SNP, gene variations.

1. Introduction

Globally, breast cancer is ranked first in the cause of cancer deaths in women (Azamjah *et al*., 2019). Breast cancer is a malignant tumor that arises from cells in the breast glands, glandular ducts and supporting tissue; rarely affects the outer layer of the breast (Indonesian Ministry of Health, 2016). Based on data from GLOBOCAN (Global Burden of Cancer), the International Agency for Research on Cancer (IARC) in 2020, breast cancer is now the deadliest type of cancer, after lung cancer (Sung *et al*., 2021). In 2020 breast cancer claimed 685,000 lives and was diagnosed in 2.3 million women. Breast cancer is the most common type of cancer worldwide, with 7.8 million living women having been diagnosed in the last five years. Although breast cancer can affect any woman in the world after adolescence, its frequency tends to increase with age (World Health Organization, 2023). Based on the RISKESDAS survey, the prevalence of breast cancer ranks 7th among all types of cancer and has occurred in 61,682 cases. Breast cancer occurs in Indonesian women at a ratio of 18/100,000 and 1% in men (Ministry of Health of the Republic of Indonesia, 2018). It was found that more than 80% of breast cancer cases in Indonesia are at an advanced stage, where treatment is difficult (Indonesian Ministry of Health, 2015).

In contrast to cervical cancer, the course of the disease and the etiology of breast cancer are not yet clearly known. However, many researchers have successfully conducted research and shown related risk factors that increase the likelihood of breast cancer. Hormonal disorders (characterized by excess estrogen) and hereditary or genetic factors are the main risk factors for breast cancer (Indonesian Ministry of Health, 2016). A study that was conducted in the UK regarding genetic testing for breast cancer susceptibility stated that protein truncating variants in 5 genes (*ATM*, *BRCA1*, *BRCA2*, *CHEK2*, and *PALB2*) were associated with overall breast cancer risk with a P value of less than 0.0001. Protein truncating variants in 4 other genes (*BARD1*, *RAD51C*, *RAD51D*, and *TP53*) were associated with overall breast cancer risk with P values less than 0.05. For gene variants in *ATM* and *CHEK2*, OR values were higher in estrogen receptor (ER)–positive disease than in ER-negative disease; for gene variants in *BARD1*, *BRCA1*, *BRCA2*, *PALB2*, *RAD51C*, and *RAD51D*, ORs were higher for ER-negative disease than for ER-positive disease. Missense variants in ATM, *CHEK2*, and *TP53* were associated with overall breast cancer risk with P values less than 0.001. For *BRCA1*, *BRCA2*, and *TP53*, missense variations that meet pathogenic classification criteria are associated with increased risk of breast cancer (Dorling et al., 2021)

Identifying the genes that cause a disease, genomic databases can help the development and discovery of new drugs to maximize therapeutic achievements. Thus, the appropriate dose and degree of therapeutic success are obtained in medical therapy. Genetic identification aims to identify inherited genetic risk factors for disease (Irham et al., 2022). Single Nucleotide Polymorphisms (SNPs) are one of the common genetic variations. Therefore, the researcher's aim in conducting this research is to find and report types of specific gene variations/SNPs with missense mutations. This is due to the fact that missense variations can cause changes in the amino acid sequence, which can have an impact on pathophysiological aspects (Ramayanam et al., 2022). The gene variations obtained not only explain disease susceptibility, but can be used to assist in the discovery and development of new drugs through biomarkers obtained in breast cancer (Lifia *et al*., 2023).

2. Materials and Methods

In the research, the type of method used is non-experimental by utilizing secondary data sources in the form of data obtained from Genome-wide Association Studies (GWAS) https://www.ebi.ac.uk/gwas/ which was designed by the National Human Genome Research Institute (NHGRI). Researchers carry out integration of genomic databases. Data integration is the process of combining or uniting two or more data from various different database sources. In this case, the databases used are the Genome-wide Association Studies (GWAS) catalog, HaploReg v4.2, Genotype Tissue Expression (GTEx Portal), and Enssembl Genome Browser to identify gene variations that play a role in the pathogenesis of breast cancer.

2.1. Sample

The sample used was secondary data in the form of breast cancer SNPs in the GWAS catalog database which was accessed in March 2024 and met the inclusion criteria. The following are the inclusion and exclusion criteria in this study:

1. Inclusion criteria:
   a) SNP in the GWAS catalog with the keyword "Breast Cancer"
   b) SNP with p value 5 x 10-8 and SNP with Odd Ratio > 1
   c) missense SNP
2. Exclusion criteria
   a) SNP with incomplete rs number

Namely SNPs that are written as protein positions such as chr10:123340431 or written as gene codes.

b) SNP with the same rs number

2.2. Tools and Materials

The tools used in this research are an HP laptop with the Windows 11 operating system, genomic databases such as the Genome-wide Association Studies (GWAS) catalog, Haploreg v4.2, Genotype Tissue Expression (GTEx Portal), and Emsembl Genome Browser. The material used in this research is Breast Cancer SNP data obtained from databases such as the Genome-wide Association Studies (GWAS) Catalog.

2.3. Data Analizes

- SNPs associated with Breast Cancer.
  SNP data obtained from GWAS with keywords related to breast cancer were collected using Microsoft Excel, then the data was analyzed using inclusion criteria with P-Value 5 x10-8 and odds ratio > 1, then the data was processed using HaploReg v4.2. HaploReg v4.2 is a database used to review genome annotations such as candidate SNP regulatory loci associated with certain diseases. Missense is a variant of one of the functional annotation methods of HaploReg v4.2. The criteria for SNP annotation are SNPs that have missense mutations.

- Frequency of breast cancer alleles in various populations
  Allele frequencies in various populations were obtained from Ensembl Genome Browser data. Ensembl is a resource for obtaining genomic information and annotation with a number of visualization tools, one of which is the distribution of allele frequencies in the world population.

3. Results and Discussion

A. Identification of Breast Cancer Gene Variations

This research uses non-experimental methods by utilizing secondary data sources. The secondary data used is a genomic database, including GWAS Catalog, HaploReg v4.2, GTEx Portal, and Ensembl Genome Browser. Based on the results of research that has been carried out, 2728 SNPs related to breast cancer were obtained through the GWAS Catalog database with the keyword "Breast Cancer" https://www.ebi.ac.uk/gwas/ (accessed on 7/3/2024). After sorting the SNPs based on P-value < 5 x 10-8 and Odds Ratio > 1, 762 SNPs were obtained. Then the SNP was developed using HaploReg v4.2 to see the risk variant of the missense allele https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php (accessed on (9/3/2024). Variants with missense mutations can result in Changes in the amino acid sequence that can have an effect on the pathophysiology of the disease. In this study, what was tested was tissue expression that affects breast cancer with missense variants. Based on the research results, 10 missense allele locations were obtained which had a risk of association with breast cancer, namely rs3184504, rs3206824, rs6929137, rs35383942, rs6964587, rs1053338, rs16991615, rs17879961, rs2363956, rs1155 2449. The gene code for each variant is SH2B3, DKK3, MCM8, CCDC170, PHLDA3, AKAP9, ATXN7, CHEK2, ANKLE1, DCLRE1B.

Table 1. Variant risk alleles coding for 11 genes

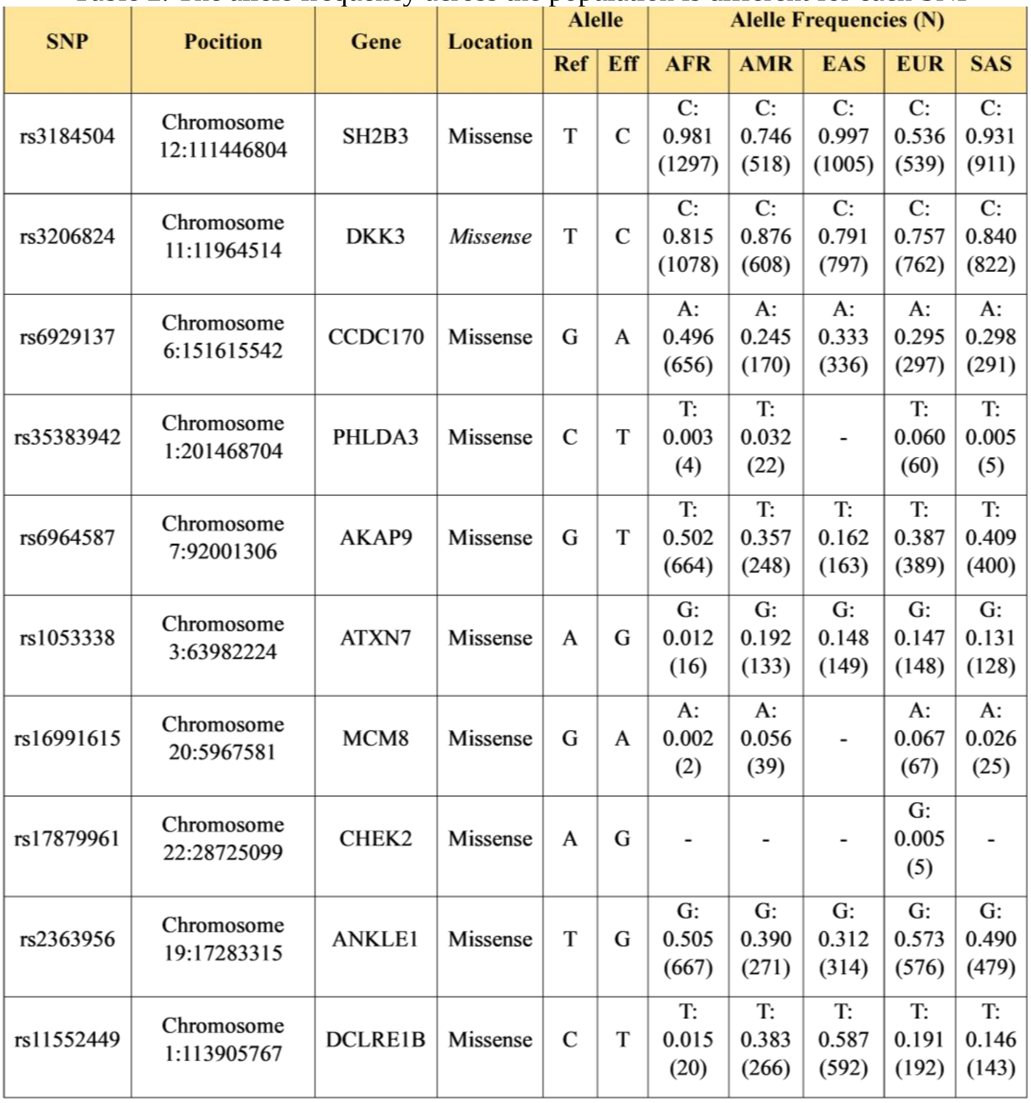
Variant Risk Allele	*P-Value*	OR	Genkode	Lokasi Allele
rs3184504	2 x 10^{-39}	1,12	*SH2B3*	*Missense*
rs3206824	2 x 10^{-20}	1,06	*DKK3*	*Missense*
rs6929137	2 x 10^{-13}	1,12	*CCDC170*	*Missense*
rs35383942	4 x 10^{-13}	1,12	*PHLDA3*	*Missense*
rs6964587	3 x 10^{-12}	1,04	*AKAP9*	*Missense*
rs1053338	5 x 10^{-11}	1,05	*ATXN7*	*Missense*
rs16991615	2 x 10^{-09}	1,10	*MCM8*	*Missense*
rs17879961	1 x 10^{-08}	1,26	*CHEK2*	*Missense*
rs2363956	2 x 10^{-08}	1,22	*ANKLE1*	*Missense*

rs11552449	2 x 10^{-08}	1,07	*DCLRE1B*	*Missense*

B. Allele Frequencies of Gene Variation in the World Population

Identification of allele frequencies in diverse populations. Variant allele frequencies were evaluated in diverse populations from Europe, America, East Asia, South Asia, and Africa. Allele frequencies were extracted from Ensembl Genome Browser https://www.ensembl.org/Homo_sapiens/Info/Index (accessed 03/18/2024).

Table 2. The allele frequency across the population is different for each SNP



SNP	Pocition	Gene	Location	Alelle		Alelle Frequencies (N)				
				Ref	Eff	AFR	AMR	EAS	EUR	SAS
rs3184504	Chromosome 12:111446804	SH2B3	Missense	T	C	C: 0.981 (1297)	C: 0.746 (518)	C: 0.997 (1005)	C: 0.536 (539)	C: 0.931 (911)
rs3206824	Chromosome 11:11964514	DKK3	*Missense*	T	C	C: 0.815 (1078)	C: 0.876 (608)	C: 0.791 (797)	C: 0.757 (762)	C: 0.840 (822)
rs6929137	Chromosome 6:151615542	CCDC170	Missense	G	A	A: 0.496 (656)	A: 0.245 (170)	A: 0.333 (336)	A: 0.295 (297)	A: 0.298 (291)
rs35383942	Chromosome 1:201468704	PHLDA3	Missense	C	T	T: 0.003 (4)	T: 0.032 (22)	-	T: 0.060 (60)	T: 0.005 (5)
rs6964587	Chromosome 7:92001306	AKAP9	Missense	G	T	T: 0.502 (664)	T: 0.357 (248)	T: 0.162 (163)	T: 0.387 (389)	T: 0.409 (400)
rs1053338	Chromosome 3:63982224	ATXN7	Missense	A	G	G: 0.012 (16)	G: 0.192 (133)	G: 0.148 (149)	G: 0.147 (148)	G: 0.131 (128)
rs16991615	Chromosome 20:5967581	MCM8	Missense	G	A	A: 0.002 (2)	A: 0.056 (39)	-	A: 0.067 (67)	A: 0.026 (25)
rs17879961	Chromosome 22:28725099	CHEK2	Missense	A	G	-	-	-	G: 0.005 (5)	-
rs2363956	Chromosome 19:17283315	ANKLE1	Missense	T	G	G: 0.505 (667)	G: 0.390 (271)	G: 0.312 (314)	G: 0.573 (576)	G: 0.490 (479)
rs11552449	Chromosome 1:113905767	DCLRE1B	Missense	C	T	T: 0.015 (20)	T: 0.383 (266)	T: 0.587 (592)	T: 0.191 (192)	T: 0.146 (143)

Description: AFR, Africa; AMR, America; EAS, East Asia; EUR, Europe; SAS, South Asia; N, total number of samples; Ref, reference; Eff, Effects of AFR, AMR, EAS, EUR, SAS alleles extracted from Ensembl.org (https://www.ensembl.org/Homo_sapiens/Info/Index?redirect=no). *Allelic effects are associated with higher expression.

Based on the research results, it was found that on the continents of Africa, East Asia and South Asia the highest allele frequency was the C allele at rs3184504, which can be seen that the SNP with this rs number encodes the SH2B3 gene. In the Americas and Europe, the highest allele frequency is the C allele at rs3206824 which encodes the DKK3 gene. Knowing the distribution of gene expression on different continents can be very useful in the development of drugs that are marketed globally. It is hoped that by knowing the genes that are common in a population, the development of more specific therapies for the treatment and prevention of breast cancer can be more organized. The genetic variations that have been identified in this study have the potential to increase breast cancer susceptibility which was carried out using a genomic database, so there are still many limitations in data collection. Therefore, experimental research and clinical studies are needed to confirm these findings. This research provides insight for future findings and can be used as a potential early marker or biomarker regarding breast cancer susceptibility.

4. Conclusion

The conclusions of this research are:

A. There are 10 gene variations that are thought to influence the pathogenesis of breast cancer, namely rs3184504 (SH2B3), rs3206824 (DKK3), rs16991615 (MCM8), rs6929137 (CCDC170), rs35383942 (PHLDA3), rs6964587 (AKAP9), rs1053338 (ATXN7), rs1 7879961 (CHEK2) , rs2363956 (ANKLE1), rs11552449 (DCLRE1B). The PHLDA3 and DKK3 genes have the highest expression in mammary tissue so they are thought to have the strongest influence on the pathogenesis of breast cancer.

B. The gene variations identified in this study show different SNP distributions in various populations. The populations along with the SNP estimated to be most susceptible are Africa, East Asia and South Asia with the C allele at rs3184504. The American and European populations and their SNPs are estimated to be most susceptible to the C allele with rs3206824.

5. Suggegstions

A. Research related to drugs for breast cancer that targets genes that influence breast cancer can be further developed.

B. Data on the distribution of gene variations that cause breast cancer can be studied further in the world population. The use of genomic databases for research using similar databases but with different diseases is also expected to be useful for the discovery and development of treatments in the future.

References

Azamjah, N., Soltan-Zadeh, Y., & Zayeri, F. (2019). Global trend of breast cancer mortality rate: A 25-year study. Asian Pacific Journal of Cancer Prevention, 20(7), 2015–2020. https://doi.org/10.31557/APJCP.2019.20.7.2015

Dorling, L., Carvalho, S., Allen, J., González-Neira, A., Luccarini, C., Wahlström, C., Pooley, K. A., Parsons, M. T., Fortuno, C., Wang, Q., Bolla, M. K., Dennis, J., Keeman, R., Alonso, M. R., Álvarez, N., Herraez, B., Fernandez, V., Núñez-Torres, R., Osorio, A., … Easton, D. F. (2021). Breast Cancer Risk Genes — Association Analysis in More than 113,000 Women. New England Journal of Medicine, 384(5), 428–439. https://doi.org/10.1056/nejmoa1913948

Kemenkes RI. (2018). Hasil Riset Kesehatan Dasar Tahun 2018. Kementrian Kesehatan RI, 53(9), 1689–1699.

Kementerian Kesehatan RI. (2015). Panduan Nasional Penanganan Kanker Payudara. http://kanker.kemkes.go.id/guidelines/PNPKPayudara.pdf

Kementrian Kesehatan RI. (2016). Pedoman Teknis Pengendalian Kanker Payudara dan Kanker Leher Rahim. Igarss, 1, 1–5. http://www.p2ptm.kemkes.go.id/dokumen-ptm/pedoman-teknis-pengendalian-kanker-payudara-kanker-leher-rahim

Lifia, A., Kartikasari, N., Devi, A., Wibowo, K., & Budiawan, H. (2023). Identifikasi Variasi Gen dan Ekspresi Gen Yang Berhubungan Dengan Anemia Aplastik Menggunakan Pendekatan Genomik Dan Bioinformatika. 4(2), 300–306.

Puspitaningrum, A. N., Perwitasari, D. A., Adikusuma, W., Djalilah, G. N., Dania, H., Maliza, R., Faridah, I. N., Sarasmita, M. A., Rezadhini, M., Cheung, R., & Irham, L. M. (2022). Integration of genomic databases and bioinformatic approach to identify genomic variants for sjogren's syndrome on multiple continents. Media Farmasi: Jurnal Ilmu Farmasi, 19(2), 71. https://doi.org/10.12928/mf.v19i2.23706

Ramayanam, N. R., Manickam, R., Mahalingam, V. T., Goh, K. W., Ardianto, C., Ganesan, P., Ming, L. C., & Ganesan, R. M. (2022). Functional and Structural Impact of Deleterious Missense Single Nucleotide Polymorphisms in the NR3C1, CYP3A5, and TNF-α Genes: An In Silico Analysis. Biomolecules, 12(9). https://doi.org/10.3390/biom12091307

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660

World Health Organization. (2023). Breast Cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer