

Uncovering hidden genetic risk factors for breast and ovarian cancers in BRCA-negative women: a machine learning approach in the Saudi population

Nofe Alganmi^{1,2,3}, Arwa Bashanfar⁴, Reem Alotaibi⁴, Haneen Banjar^{1,2,3}, Sajjad Karim^{2,5}, Zeenat Mirza^{5,6}, Heba Abusamra², Manal Al-Attas², Shereen Turkistany⁷ and Adel Abuzenadah^{2,5}

¹ Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

² Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia

³ Centre of Artificial Intelligence in Precision Medicines, King Abdulaziz University, Jeddah, Saudi Arabia

⁴ Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

⁵ Department of Medical Lab Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

⁶ King Fahd Medical Research Center, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

⁷ Center of Innovation Personalized Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

Breast and ovarian cancers are prevalent worldwide, with genetic factors such as BRCA1 and BRCA2 mutations playing a significant role. However, not all patients carry these mutations, making it challenging to identify risk factors. Researchers have turned to whole exome sequencing (WES) as a tool to identify genetic risk factors in BRCA-negative women. WES allows the sequencing of all protein-coding regions of an individual's genome, providing a comprehensive analysis that surpasses traditional gene-by-gene sequencing methods. This technology offers efficiency, cost-effectiveness and the potential to identify new genetic variants contributing to the susceptibility to the diseases. Interpreting WES data for disease-causing variants is challenging due to its complex nature. Machine learning techniques can uncover hidden genetic-variant patterns associated with cancer susceptibility. In this study, we used the extreme gradient boosting (XGBoost) and random forest (RF) algorithms to identify BRCA-related cancer high-risk genes specifically in the Saudi population. The experimental results exposed that the RF method scored superior performance with an accuracy of 88.16% and an area under the receiver-operator characteristic curve of 0.95. Using bioinformatics analysis tools, we explored the top features of the high-accuracy machine learning model that we built to enhance our knowledge of genetic interactions and find complex genetic patterns connected to the development of BRCA-related cancers. We were able to identify the significance of HLA gene variations in these WES datasets for BRCA-related patients. We find that immune response mechanisms play a major role in the development of BRCA-related cancer. It specifically highlights genes associated with antigen processing and presentation, such as HLA-B, HLA-A and HLA-DRB1 and their possible effects on tumour progression and immune evasion. In summary, by utilizing machine learning

Submitted 6 December 2023

Accepted 26 February 2024

Published 19 April 2024

Corresponding authors

Nofe Alganmi,

nalghanimi@kau.edu.sa

Arwa Bashanfar,

abashanfar0002@stu.kau.edu.sa

Academic editor

Khalid Raza

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj-cs.1942

© Copyright

2024 Alganmi et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

approaches, we have the potential to aid in the development of precision medicine approaches for early detection and personalized treatment strategies.

Subjects Bioinformatics, Computational Biology, Algorithms and Analysis of Algorithms, Data Mining and Machine Learning

Keywords Machine learning, WES, Genetic risk factor, Bioinformatics, Breast cancer, Ovarian cancer, BRCA

INTRODUCTION

Approximately 19.3 million people were diagnosed with cancer in 2020, causing 10 million deaths, according to GLOBOCAN 2020 (*Sung et al., 2021*). Breast cancer—a BRCA-related cancer—is the most-diagnosed cancer type in the world. Breast cancer was diagnosed in 2.3 million cases, which represented 11.7%, and there were 68,496 deaths, representing 6.9% of all cancer deaths. In total, 17.8% and 3%, breast cancer and ovarian cancers, respectively, are the most common types of cancer among Saudi nationality (*National Cancer Center of the Saudi Health Council, 2024*). While genetic factors are known to play a significant role in their development, specifically mutations in the breast cancer 1 (BRCA1) and breast cancer 2 (BRCA2) genes, it is important to note that not all breast and ovarian cancer patients carry these specific mutations. This poses challenges when identifying genetic risk factors in these patients (*Kurian et al., 2011*). The BRCA1 and BRCA2 genes are well-known tumor suppressor genes that play a crucial role in maintaining the stability of the genome. Mutations in these genes significantly increase the risk of developing hereditary breast and ovarian cancers (*Petrucelli, Daly & Pal, 2022*). The prevalence of BRCA mutations in breast and ovarian cancer patients can vary among different populations. Studies have estimated that only approximately around 5% to 10% of breast and ovarian cancer patients have BRCA mutations (*Godet & Gilkes, 2017*). This indicates that a substantial number of patients who test negative for these mutations may still have an unidentified genetic risk factor contributing to their condition. There are several limitations of BRCA testing in identifying genetic risk factors in BRCA-negative patients. One limitation is the low occurrence of BRCA mutations in certain breast and ovarian cancer patients, leading to false-negative results. Moreover, BRCA testing only targets specific genes and mutations and does not analyze the entire genome for potential genetic risk factors. This means that other genetic variants that may contribute to breast and ovarian cancer susceptibility may be missed. Identifying genetic risk factors in patients who test negative for BRCA mutations requires alternative approaches. One such approach is whole exome sequencing (WES), which can identify genetic variants associated with breast and ovarian cancer susceptibility beyond BRCA1 and BRCA2 mutations. WES is an advanced DNA analysis technique that allows the sequencing of all protein-coding regions, known as exons, in one comprehensive assay. This method offers several advantages compared to other sequencing techniques such as whole-genome sequencing (WGS) and targeted gene panels (*Bartha & Györfy, 2019*). One significant advantage of WES is its efficiency and cost effectiveness (*Suwinski et al., 2019*). WES also offers the advantage of

identifying new genetic variants that could potentially contribute the susceptibility to the disease. By analyzing genes beyond those already known to cause disease, WES provides a more thorough examination of the genetic factors underlying various conditions (*Rabbani, Tekin & Mahdieh, 2014*).

Findings from studies using WES in BRCA-negative breast and ovarian cancer have identified specific genes and variants associated with the susceptibility to breast and ovarian cancer. These studies have provided valuable insights into the genetic factors underlying these cancers, potentially informing personalised treatment and risk management approaches. For instance, a recent study by *Felicio et al. (2021)* used WES analysis to identify new genes linked to the predisposition to breast and ovarian cancer. The researchers examined germline variants in cancer-related genes and conducted bioinformatic analyses to pinpoint potential genetic risk factors. Notably, they discovered a variant called c.149T > G in the FAN1 gene, present in two unrelated families, with a loss of heterozygosity observed in one family's tumor tissue. This finding suggests that FAN1 may be a promising candidate associated with hereditary breast and ovarian cancer susceptibility (*Felicio et al., 2021*). In a recent study by *Lee et al. (2022)*, the researchers used WES and case-control analyses to investigate genetic variants that might be linked to breast cancer susceptibility in individuals without BRCA mutations. Their findings uncovered new and potentially important genetic variants that could contribute to the risk of breast cancer in this particular population. Among the candidate genes identified, one of note was MUC16, which has previously been associated with increased susceptibility to ovarian cancer (*Lee et al., 2022*). Another study by *BenAyed-Guerfali et al. (2022)* focused on BRCA-negative Tunisian patients at high risk of hereditary breast/ovarian cancer. Using WES, the study aimed to identify other genes that may contribute to susceptibility. Their findings shed light on potential genetic risk factors beyond BRCA mutations, highlighting candidate genes such as RAD51C, which have been previously linked to the susceptibility to breast and ovarian cancer (*BenAyed-Guerfali et al., 2022*). Furthermore, a recent study conducted by *Grasel et al. (2020)* used WES to identify genetic variants that may contribute to an increased risk of breast cancer in individuals from families with a history of the disease who tested negative for BRCA mutations. The study revealed potential candidate genes, one of which is the ATM gene, previously linked to the susceptibility to breast cancer.

Conversely, the process of interpreting WES data and finding the disease-causing variants among thousands of variants remains a challenge due to the characteristics of these data (*Huang et al., 2022*). Thus, using machine learning (ML) with these kinds of data supports humans in dealing with these large and complex data (*Fan et al., 2022*).

Many studies profit from ML algorithms using WES for different diseases. For example, in *Trakadis et al. (2019)*, they applied extreme gradient boosting (XGBoost), random forest (RF), L1.Logistic and a support vector machine (SVM) on WES data to identify the high-risk genes for the schizophrenia (SCZ) disease. XGBoost yielded ideal results with an accuracy equal to 85.7%, a specificity equal to 86.6%, a sensitivity equal to 84.9% and an area under the receiver operator characteristic (AUC) equal to 0.95. Moreover, they analyzed the top 50 genes of the algorithms that are associated with SCZ using

bioinformatic resources. In addition, in *Hooshmand (2020)*, they applied a naive Bayesian algorithm on 1,091 ovarian cancer samples from The Cancer Genome Atlas (TCGA) research network and 179 healthy people from the Genotype-Tissue Expression (GTEx) project to distinguish between cancerous and noncancerous cells. They yielded amazing results with 100% accuracy, 100% specificity, 100% sensitivity and an AUC equal to one. Additionally, in *Wadapurkar et al. (2023)*, they applied five supervised machine learning algorithms-SVM, decision tree (DT), RF, Naive Bayes and XGBoost - to identify variants that are associated with ovarian cancer. XGBoost yielded 94.64% in accuracy and 0.97 as the AUC. RF and XGBoost are efficient machine learning methods because they work well for problems with high dimensions and they generally allow nonlinear relationships between the features (ref).

In conclusion, it is crucial to continue exploring genetic risk factors for breast and ovarian cancers in individuals who do not have BRCA1 and BRCA2 mutations. While these mutations are significant risk factors, a considerable number of patients lack them, underscoring the need to identify other robust biomarkers. Machine learning techniques for analyzing WES data have emerged as a powerful tool for uncovering hidden genetic variant patterns associated with the susceptibility to cancer beyond the established BRCA genes.

MATERIALS AND METHODS

The model, which contains three stages-data pre-processing, machine learning and pathway analysis-is shown in *Fig. 1*. The complete source code and data files are available on Zenodo at <https://zenodo.org/records/10720553>, <https://doi.org/10.5281/zenodo.10926612>.

Exome data set

We utilized WES data obtained from the Center of Excellence in Genomic Medicine Research (CEGMR) at King Abdulaziz University. The dataset includes 76 women who were diagnosed with BRCA-related cancer but tested negative for known risk variants (BRCA genes). Ethical approval for this study was obtained under IRB No. 32-CEGMR-Bioeth-2021, and written informed consent was obtained from all participants. Among the participants, 21 individuals had a positive BRCA genetic status, while 55 individuals had a negative BRCA genetic status. The data files were provided in the VCF file, containing a total of 249,430 germline short variants (single nucleotide polymorphisms (SNP) and indel) calls. *Figure 2* illustrates the screenshot of the VCF file.

VCF file pre-processing

Raw whole-exome sequencing data were aligned with the human reference genome GRCh38 and then called using GATK's (v.4.1.2) HaplotypeCaller. After that, BCFtools V.1.13.1 was used to merge all VCF files into one cohort VCF. In the merged VCF file, all SNPs are biallelic SNPs that have been filtered to remove calls with a depth of less than 20. The data were annotated by ANNOVAR (*Wang, 2010*) using the reference genome hg38/GRCh38.

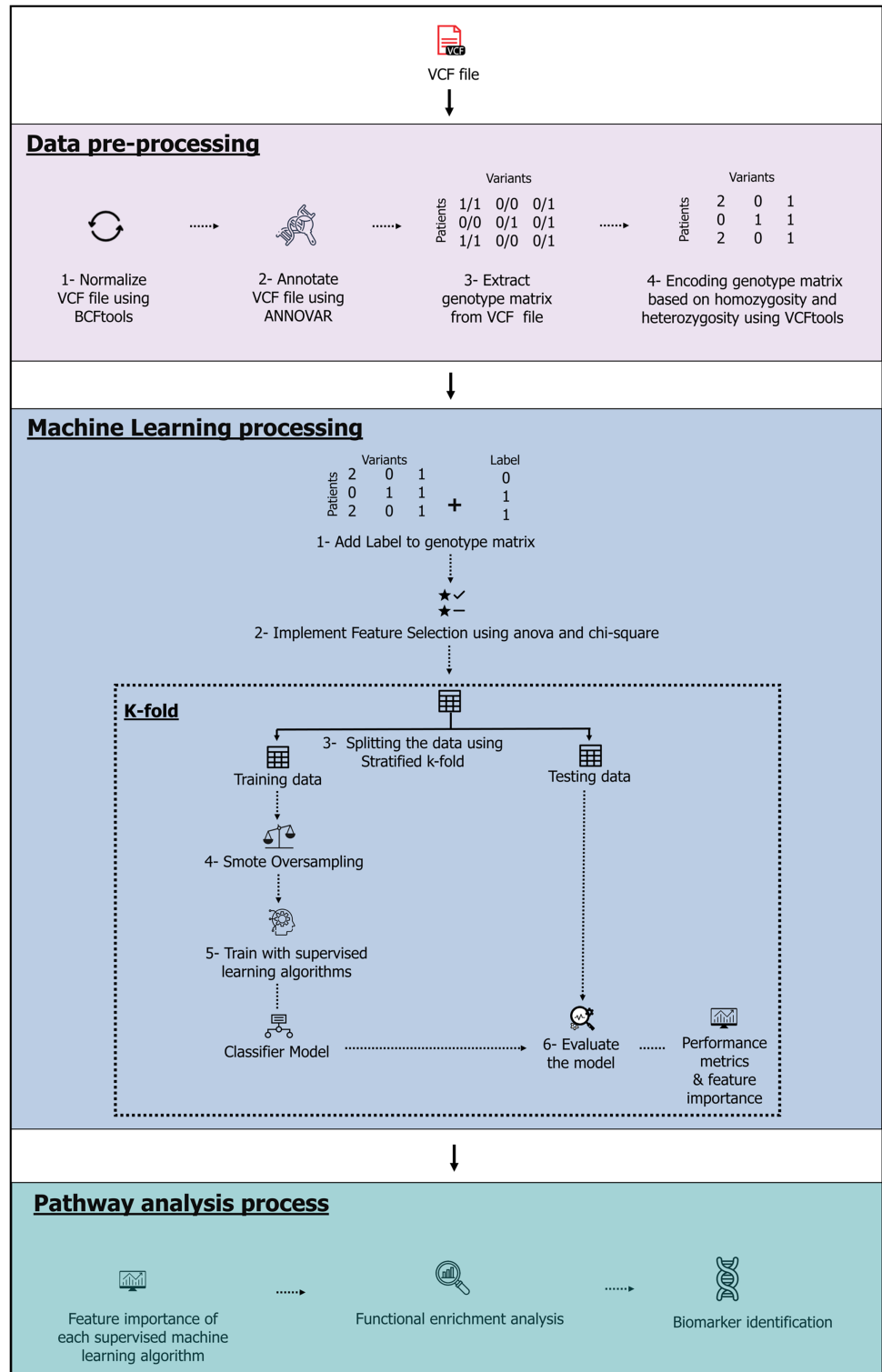


Figure 1 The machine learning pipeline. Image source credits: Feature selection icons created by herikus-Flaticon. Data table icon, Designed by Freepik. Machine learning icon, Designed by Freepik. Magnifying glass icons created by Freepik-Flaticon. Classifier model icon, Designed by Freepik. Benchmark icons created by srip-Flaticon. Genes icons created by Smashicons-Flaticon.

Full-size  DOI: 10.7717/peerj-cs.1942/fig-1

	CHROM	POS	ID	REF	ALT_1	ALT_2	ALT_3	QUAL
0	chr1	13380	rs571093408	C	G	NaN	NaN	2064.10009
1	chr1	13417	rs777038595	C	CGAGA	NaN	NaN	17749.19921
2	chr1	13418	rs75175547	G	A	NaN	NaN	43.90000
3	chr1	13445	rs558318514	C	G	NaN	NaN	513.20001
4	chr1	13550	rs554008981	G	A	NaN	NaN	139.10000
...
52817	chrM	15812	rs200336777	G	A	NaN	NaN	1284.09997
52818	chrM	15884	rs527236195	G	C	NaN	NaN	306.29998
52819	chr22_K1270879v1_alt	271020	rs2266637	C	T	NaN	NaN	210.00000
52820	chr22_K1270879v1_alt	277188	rs8140585	C	T	NaN	NaN	142.50000
52821	chr22_K1270879v1_alt	278394	rs56106137	G	A	NaN	NaN	307.60000

249430 rows × 9 columns

VCF file

Figure 2 The snapshot of the dataset (VCF file).

Full-size  DOI: 10.7717/peerj-cs.1942/fig-2

The last step in this phase is encoding the genotype matrix to be ready to use in the machine learning phase. The encoding step was done using VCFtools (*Danecek et al., 2011*) Version 0.1.16-3 through the '012 matrix' command (*Figueiredo de Sá et al., 2019; Magi et al., 2015; Carvalho et al., 2020*). VCFtools provides simple tools for working with genetic variation data stored in VCF files. The encoding accomplished was as follows:

- 0/0 denotes the reference homozygote that is coded as 0.
- 0/1 and 1/0 denote the heterozygote that is coded as 1.
- 1/1 denotes the alternative homozygote that is coded as 2.
- All missing values are coded as -1.

Figure 3 illustrates the final data used in the machine learning process.

Machine learning process

We used random forest (RF) and XGBoost supervised machine learning algorithms with the labeled genotype matrix for the 76 individual datasets to identify features or genes that might correlate with BRCA-related risk factors. The following subsections discuss the feature selection process and machine learning in detail.

Feature selection process

As a preliminary step, we employed ANOVA and chi-square feature selection techniques in order to prevent overfitting. Using the top 5,000 and 7,000 variants of each technique, we reduced the number of variants in the genotype matrix. Thus, after applying feature selection techniques, we obtained the top 500 and 700 variants for each feature selection method (ANOVA and chi-square).

	Variants							Lable / Target	
	V1	V2	V3	V4	V5	...	V249430	Lable	
P1	0	1	0	2	0	1	0	
P2	2	-1	0	2	-1	0	1	
P3	-1	2	0	1	2	2	1	
P4	2	0	1	2	1	1	0	
P5	0	1	1	-1	1	1	1	
....	
P76	0	2	1	-1	0	2	0	

Dataset

Figure 3 The final data set that is used in the machine learning process.

Full-size  DOI: 10.7717/peerj-cs.1942/fig-3

Machine learning process

For splitting data into training and testing datasets, we applied stratified k-fold cross-validation with the k equal to four (Bukhari et al., 2022; Bukhari, Webber & Mehbodniya, 2022). For the training data, the Synthetic Minority Oversampling Technique (SMOTE) was used to avoid problems of imbalanced data using the imbalanced-learn package Version 0.10.1. To evaluate the classifier model, the testing datasets were used. Table 1 shows an illustration of the parameters of the supervised machine learning algorithms which that chosen through GridSearchCV.

Molecular pathway and network analysis

Molecular pathway analysis

The machine learning models were used to identify the top 100 significant variants through feature importance. Gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were conducted on the significant top variants and their corresponding genes using DAVID and R to explore the potential mechanisms and risk factors associated with BRCA-related cancer, including both BRCA-positive and BRCA-negative patients. The statistically significant enriched terms were considered for an adjusted *P*-value less than 0.01.

RESULTS AND DISCUSSION

Machine learning process

In order to identify the most important variants and genes related to ovarian cancer and breast cancer in the Saudi population, a prediction model was trained using machine learning techniques. Specifically, we employed the XGBoost and RF algorithms. The performance of these models was evaluated using various metrics, including accuracy, precision, recall, F1 score and AUC.

Table 2 presents the results of the feature selection techniques used to improve the performance of the machine learning algorithms. The feature selection technique ANOVA, using the top 5,000 variants, achieved an accuracy of 78.95% for XGBoost and 88.16% for RF. On the other hand, chi-squared feature selection demonstrated the best

Table 1 List of the machine learning algorithms' parameters with optimized values.

Machine learning algorithm	Parameter name	Parameter value
XGBoost	Loss function	log_oss
	Number of boosting stages	100
	Function to measure the quality of a split	friedman_mse
	Minimum number of samples required to split	2
	Minimum number of samples required to be at a leaf node	1
RF	Criterion of trees	gini
	Number of trees in the forest	100
	Minimum number of samples required to split	2
	Minimum number of samples required to be at a leaf node	1
	Number of features for the best split	sqrt

Table 2 Performance metrics among supervised machine learning algorithms with ANOVA and chi-squared feature selection (5,000 features).

Model\Metric		Accuracy	Precision	Recall	F1 score	AUC
ANOVA	XGBoost	78.95%	72.50%	46.67%	52.63%	0.79
	RF	88.16%	100%	55.83%	68.56%	1.0
Chi-squared	XGBoost	82.89%	81.67%	52.5%	64.64%	0.81
	RF	88.16%	95%	60.83%	71.30%	0.95

performance in this study, particularly in combination with RF. Figure 4 shows the ROC curve for the best-performing algorithm. The accuracy of XGBoost was 82.89%, while RF achieved an accuracy of 88.16%. Additionally, the F1 score improved from 52.63% to 64.64% with XGBoost and from 68.56% to 71.30% with RF.

Gene ontology analysis

The analysis revealed an enrichment in the biological processes regarding the immune response, such as antigen processing and presentation, T cell-mediated cytotoxicity and immune cell differentiation (Fig. 5). These findings highlighted the crucial role of immune mechanisms in the evolution of BRCA-related cancer. Additionally, genes associated with MHC class I and II molecules, including HLA-B, HLA-A and HLA-DRB1, were identified, emphasizing the significance of antigen presentation and immune recognition in this cancer (Liu et al., 2021). Disruptions in these processes may facilitate immune evasion and contribute to tumour advancement. Moreover, the enrichment of the genes involved in protein hetero-tetramerization and cell adhesion suggested the potential involvement of cellular interactions and communication in BRCA-related cancers (Godet & Gilkes, 2017). Alterations of these aspects may disrupt cell-cell adhesion, promoting cancer cell metastasis and invasiveness. Unexpectedly, the analysis also revealed the enrichment of genes involved in tissue-specific processes such as muscle organ development, heart development and spermatogenesis. This suggested potential correlations between these

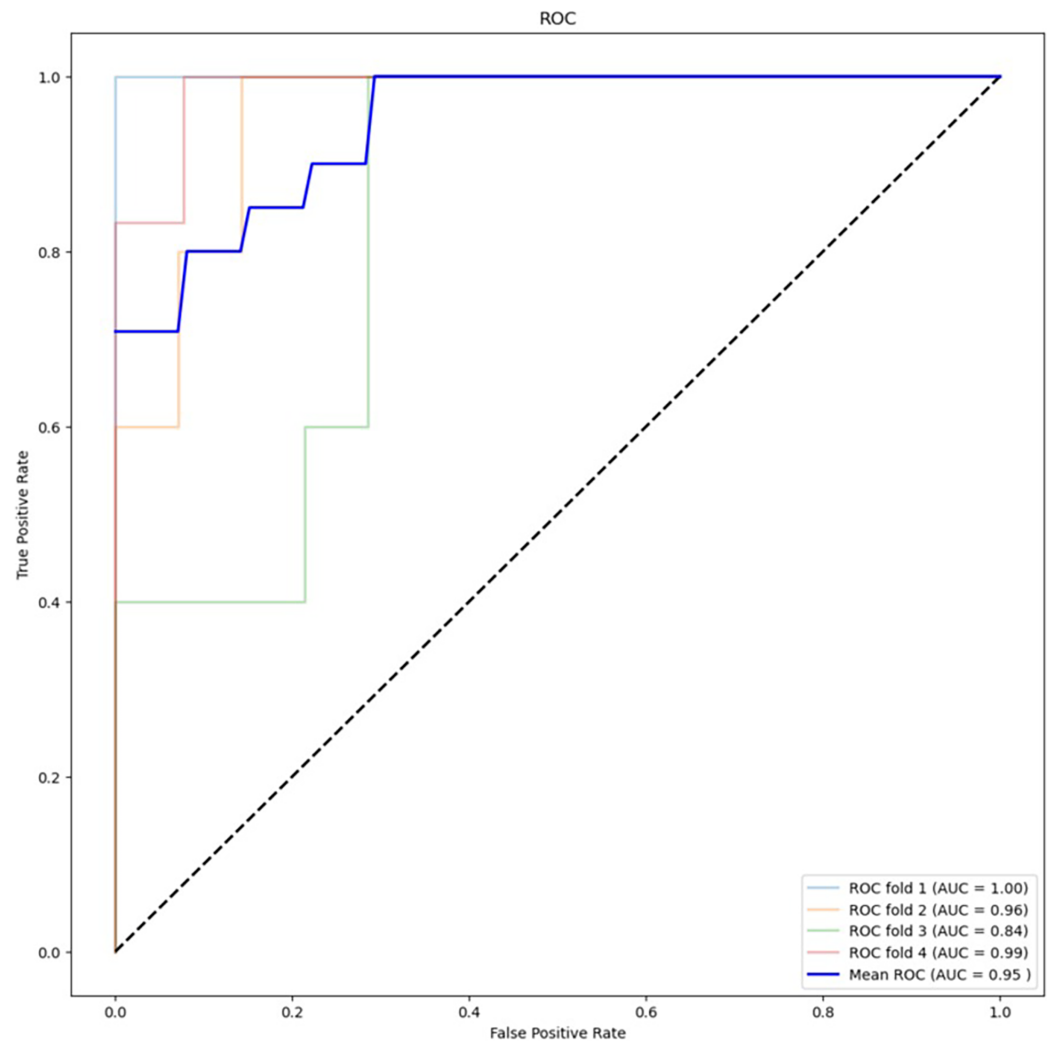


Figure 4 ROC curve for best performing algorithm: RF with chi-square (5,000 variants).

Full-size DOI: 10.7717/peerj-cs.1942/fig-4

processes and BRCA-related cancers, indicating the possible role of tissue-specific dysfunctions in cancer susceptibility or phenotype expression (Zhang & Li, 2018). Table 3 shows the key findings (genes) and their pivotal functions.

KEGG pathway analysis

The KEGG pathway analysis using DAVID (Sherman et al., 2022; Huang, Sherman & Lempicki, 2009) identified several pathways of potential relevance to BRCA-related cancers. The enrichment of pathways associated with the immune response, immunological diseases, natural killer cell-mediated cytotoxicity and allograft rejection further emphasises the importance of immune cell interactions in these cancers. Additionally, pathways related to hormone signaling, including the estrogen-signalling pathway, were enriched, suggesting the influence of estrogen signalling even in BRCA-negative patients (Yin et al., 2020).

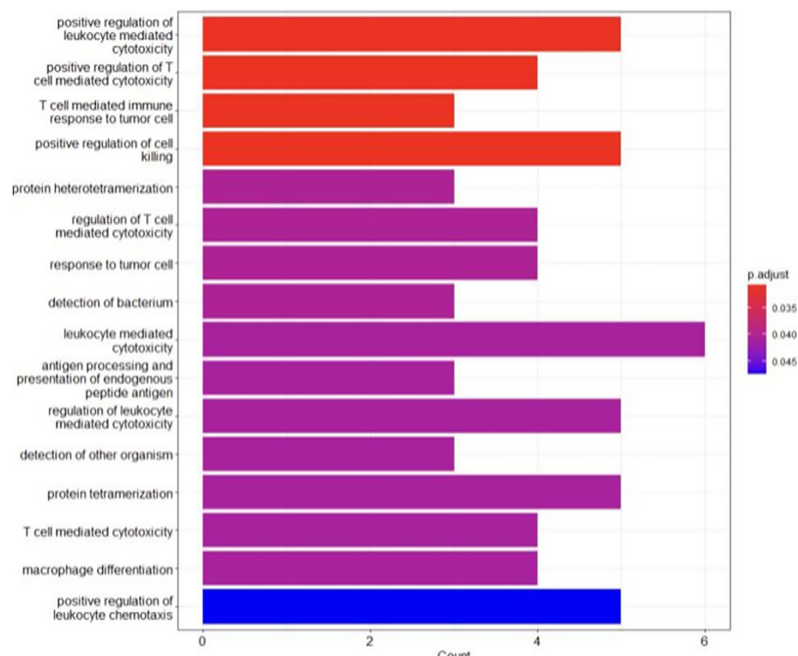


Figure 5 The GO analysis conducted in R reveals significant biological processes linked to the genes of interest.

Full-size DOI: 10.7717/peerj-cs.1942/fig-5

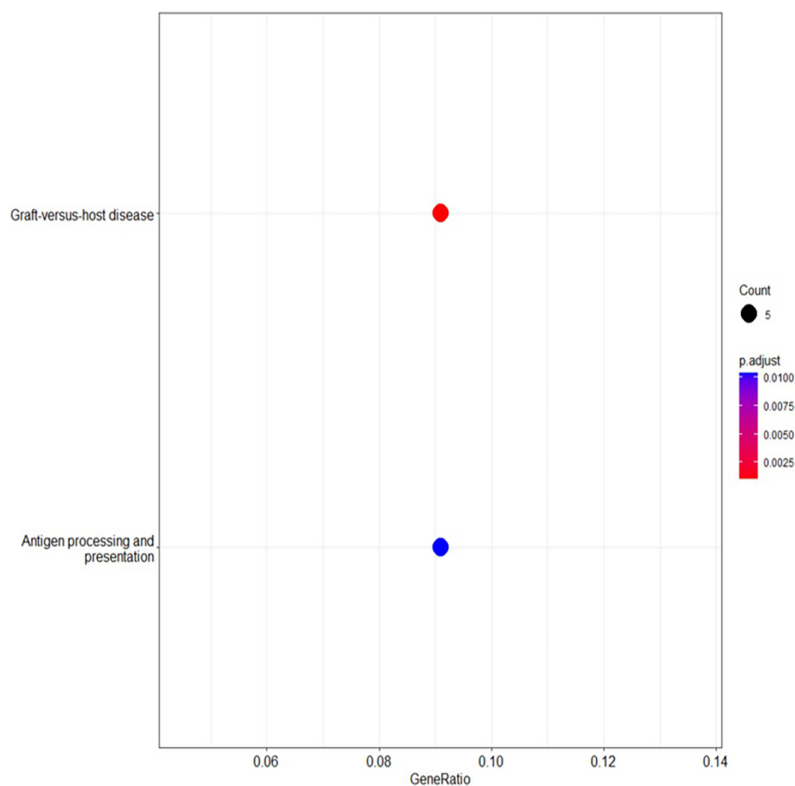
Table 3 Table encapsulating the regarding genes and their pivotal functions.

Gene	Process involved in	Description/Pivotal function
HLA-A	Antigen processing and presentation	Crucial for immune response, presents peptide antigens to T cells
HLA-B	Graft-versus-host disease, allograft rejection	Plays a key role in immune system recognition, impacting graft compatibility
HLA-DRB1	Antigen processing and presentation	Involved in presenting antigens to immune cells, crucial to the immune response
KIR2DL1	Natural killer cell-mediated cytotoxicity	Regulates natural killer cell activity, important for immune surveillance
KIR2DL3	Natural killer cell-mediated cytotoxicity	Influences the activity of natural killer cells, key in immune defense mechanisms
KIR2DS1	Natural killer cell-mediated cytotoxicity	Affects natural killer cell function, important in immune system's response
KIR2DS5	Natural killer cell-mediated cytotoxicity	Modulates the activity of NK cells, playing a role in immune response
GHR	Estrogen signalling pathway	While primarily known for growth hormone receptor, may interact with estrogen signaling
SHC1	Estrogen signalling pathway	Involved in signal transduction, potentially linking to estrogen pathways
SIRT1	Estrogen signalling pathway	Known for roles in aging and metabolism, may have links to estrogen signaling

These findings highlighted the complex interplay of factors, including immune-related processes, hormone signalling and natural killer cell-mediated cytotoxicity, as potential risk factors for BRCA-related cancers, extending beyond BRCA mutations' cell targeting. For example, genes involved in antigen processing and presentation, TAP binding and T cell receptor binding were implicated, indicating their potential roles in immune recognition and cancer (Reeves & James, 2017). The enrichment of the graft-vs.-host disease pathway suggests that immune dysregulation may contribute to BRCA-related

Table 4 The KEGG pathways associated with the analysed genes, highlighting their involvement in specific biological pathways and networks.

Category	Term	RT	Count	%	P-value	Benjamini
KEGG_PATHWAY	Antigen processing and presentation	RT	8	5.7	6.9E-7	1.1E-4
KEGG_PATHWAY	Graft-vs-host disease	RT	6	4.3	7.5E-6	5.7E-4
KEGG_PATHWAY	Natural killer cell-mediated cytotoxicity	RT	6	4.3	1.5E-3	7.4E-2
KEGG_PATHWAY	Estrogen signalling pathway	RT	5	3.5	1.3E-2	4.9E-1
KEGG_PATHWAY	Allograft rejection	RT	3	2.1	2.6E-2	7.9E-1
KEGG_PATHWAY	Type I diabetes mellitus	RT	3	2.1	3.3E-2	8.3E-1
KEGG_PATHWAY	Autoimmune thyroid disease	RT	3	2.1	4.8E-2	1.0E0
KEGG_PATHWAY	Human T-cell leukemia virus 1 infection	RT	5	3.5	5.9E-2	1.0E0
KEGG_PATHWAY	Viral myocarditis	RT	3	2.1	6.0E-2	1.0E0

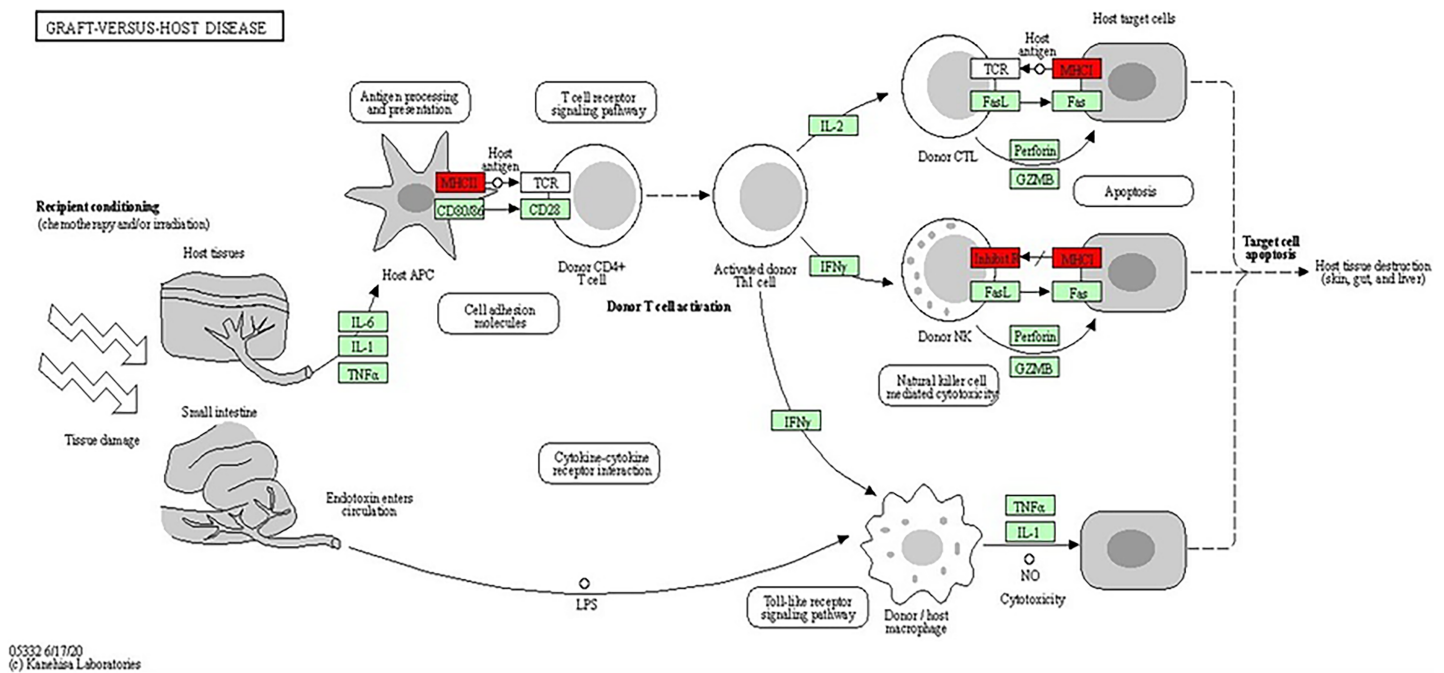
**Figure 6** The enriched KEGG pathways analysed using R linked to the selected genes, providing insights into their functional relationships and potential roles in biological processes.

Full-size DOI: 10.7717/peerj-cs.1942/fig-6

cancer susceptibility (Table 4). Additionally, metalloendopeptidase activity and calcium channel activity were identified as potential mechanisms, pointing to the involvement of extracellular matrix remodeling, cell migration and calcium signalling in cancer progression (Chen *et al.*, 2013).

The analysis of the KEGG pathway using the KEGG Mapper, as shown in Figs. 6 and 7 (Kanehisa & Sato, 2020; Kanehisa, Sato & Kawashima, 2022; Kanehisa & Goto, 2000;

GRAFT-VERSUS-HOST DISEASE



05332 6/17/20
 © Kanehisa Laboratories

Figure 7 KEGG pathway diagram for graft-vs.-host disease depicting the complex interplay of cellular and molecular interactions involved in the pathogenesis of graft-vs.-host disease (GVHD). Key components include donor T cells, antigen-presenting cells (APCs) and various cytokines. The detailed interactions between these components highlighted the roles of immune response and inflammation in GVHD progression (Image source credit: Kanehisa Laboratories). [Full-size !\[\]\(ab8f7a9d25e63edc6ae9f62ddaa1d31c_img.jpg\) DOI: 10.7717/peerj-cs.1942/fig-7](https://doi.org/10.7717/peerj-cs.1942/fig-7)

Kanehisa et al., 2023) revealed only two significant pathways with implications for BRCA-related cancers: the graft-vs.-host disease pathway and the antigen processing and presentation pathway. The ‘graft-vs.-host disease pathway’ pathway was identified as an enriched pathway in the analysis. This pathway is typically associated with complications that can occur after haematopoietic stem cell transplantation (HSCT), where immune cells from the donor (graft) attack the recipient’s (host) tissues (*Zeiser & Blazar, 2017*). The involvement of this pathway in BRCA-related cancers suggests a potential role for immune dysregulation in contributing the susceptibility to cancer. The presence of genes associated with this pathway pointed to the importance of immune-related factors beyond BRCA mutations.

The antigen processing and presentation pathway, a fundamental process for immune recognition and elimination of pathogens or abnormal cells, including cancer cells, was also enriched. This pathway (shown in [Fig. 8](#)) involves the presentation of antigens to immune cells, initiating an immune response. Antigens are processed and presented on the cell’s surface through two major pathways: the major histocompatibility complex class I (MHC I) pathway and the major histocompatibility complex class II (MHC II) pathway (*Wieczorek et al., 2017*). In the MHC I pathway, intracellular proteins, such as viral or tumour antigens, are broken down into small peptide fragments within the cytosol. These peptides are then transported into the endoplasmic reticulum (ER), where they bind to MHC I molecules. The MHC I-peptide complexes are subsequently presented on the cell’s

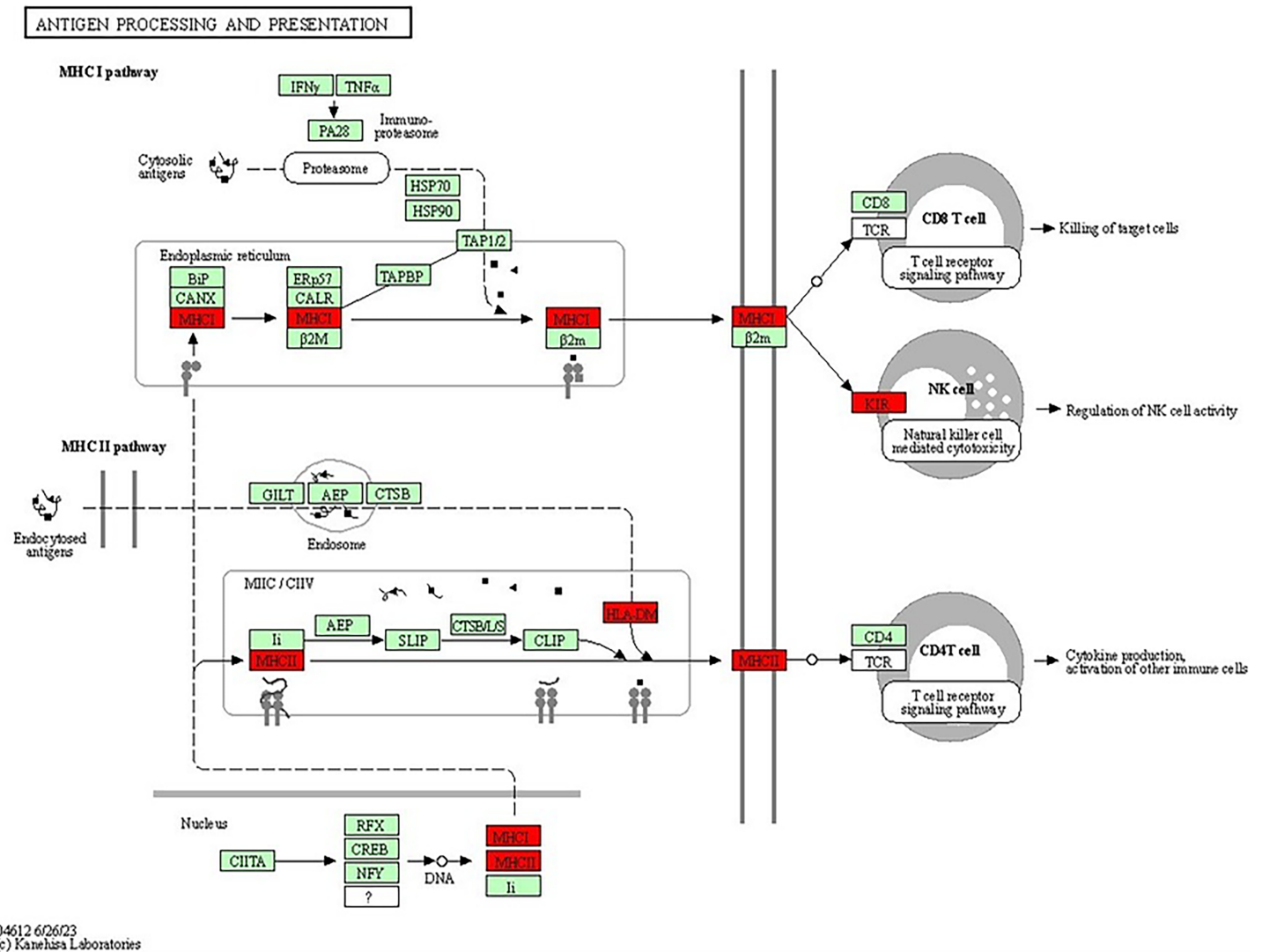


Figure 8 KEGG pathway diagram for antigen processing and presentation illustrating the central processes involved in antigen processing and presentation, which are essential for immune system function. The diagram focuses on the breakdown of pathogens into peptide fragments, their subsequent association with MHC molecules and their presentation to T cells, thereby triggering an immune response (Image source credit: Kanehisa Laboratories). [Full-size DOI: 10.7717/peerj-cs.1942/fig-8](https://doi.org/10.7717/peerj-cs.1942/fig-8)

surface, allowing cytotoxic T cells to recognise and eliminate infected or malignant cells. The presence of genes associated with MHC I-mediated antigen presentation, including HLA-B, HLA-A and HLA-DRB1, suggested their potential roles in BRCA-related cancers. Dysregulation or abnormalities in this pathway may impact the immune system's ability to recognise and combat cancer cells, independent of BRCA mutations. The MHC II pathway involves the presentation of antigens derived from extracellular sources, such as bacteria or proteins released from dying cells (Abualrous, Sticht & Freund, 2021). Antigen-presenting cells (APCs), including dendritic cells, macrophages and B cells, take up the extracellular antigens. Within the APCs, the antigens are processed into peptide fragments and loaded

The search for cancer risk-associated genes was accomplished by focusing on variants located especially in known cancer-related pathways that result in non-synonymous mutations, potentially affecting protein function, and are more prevalent in cancer patients compared to BRCA-negative patients. The role of non-BRCA genes in cancer-related pathways is increasingly being recognized (*Evans et al., 2021*). These genes, which include PALB2, ATM and CHEK2 among others, are known to interact with BRCA1 and BRCA2 in DNA repair pathways and mutations in these genes have been associated with a moderate increase in the risk of breast and ovarian cancer (*Couch et al., 2017; Cybulski et al., 2015; Easton et al., 2015*). Non-synonymous mutations, which result in changes to the protein sequence, have a higher likelihood of affecting protein functioning compared to synonymous mutations (*Miosge et al., 2015*). These mutations can alter protein activity, stability or interaction with other proteins, leading to the dysregulation of cellular pathways. In the context of BRCA-related cancers in patients who are BRCA negative, non-synonymous mutations may represent additional genetic risk factors.

By observing non-synonymous mutations in genes known to be involved in DNA repair or cell cycle control, we identified potential disruptions in normal cellular functions that could contribute to an increased risk of cancer. For example, a non-synonymous variant in the RNF223 gene, which changes the proline at position 231 to arginine (P231R), may impact the protein's function. BRCA-related cancer patients exhibit a diverse range of genetic variants, emphasising the polygenic nature of cancer susceptibility (*Vihinen, 2022*). These variants occur in multiple genes, some of which are already known to be associated with cancer, while others may represent novel candidate susceptibility genes. For instance, the dataset revealed a nonsynonymous variant in the RNF223 gene, which encodes an E3 ubiquitin-protein ligase (*Feng et al., 2021*). Disruptions in such genes can interfere with the protein degradation process, potentially leading to the accumulation of harmful proteins that promote cancer. Additionally, the GABRD gene encodes a subunit of the gamma-aminobutyric acid (GABA) receptor (*Wu et al., 2023*). Perturbations in GABA signalling have been implicated in various cancers, suggesting that alterations in this pathway could contribute to the development of cancer.

Additionally, several variants were found in non-coding regions or led to synonymous changes (no change in the encoded amino acid). While these variants do not alter protein sequences, they can still influence gene regulation and expression, thereby potentially contributing to cancer susceptibility (*Sharma et al., 2019*). These regions, once considered 'junk DNA', are now recognized as crucial regulatory regions capable of influencing gene expression. Variants in these regions can impact the binding of transcription factors or other regulatory molecules, leading to changes in the expression levels of associated genes. Such regulatory alterations can profoundly affect cellular processes and potentially contribute to the development or progression of cancer. They can influence mRNA stability, splicing and translation, thereby impacting protein levels within the cell. For instance, the analysis identified a synonymous variant in the FAM131C gene. Although this variant does not change the protein sequence it could still influence the expression or function of FAM131C, which may be relevant to cancer (*Gotea et al., 2015*). Taking this into consideration, further research is required to validate these potential risk factors,

investigate their interactions with each other and determine their interplay with non-genetic factors. Such endeavors will contribute to a more comprehensive understanding of the cancer risk in BRCA-negative patients.

CONCLUSION

In this study, we used WES samples from 76 women with BRCA-related cancer who tested negative for BRCA gene mutations. We present a machine-learning approach to detect the high-risk genes in the Saudi population. The RF algorithm achieves 88% accuracy and an AUC of 0.95 on the testing dataset. The learned model has been analysed to extract the features as variants to be studied. The analysis of the data and the results indicate that changes in HLA genes affecting the immune system are crucial to the development of cancers related to BRCA. Thus, understanding the interplay between immune dysregulation and BRCA-related cancers could provide valuable insights for the development of immunotherapeutic strategies and improved treatment outcomes. Further research is required to validate these potential risk factors, investigate their interactions with each other and determine their interplay with non-genetic factors. Such endeavours will contribute to a more comprehensive understanding of cancer risk in BRCA-negative patients.

ACKNOWLEDGEMENTS

The authors extend their appreciation to Dr. Robert Hoehndorf, Dr. Ashraf Dallol and Dr. Heba Alkatabi for their interesting discussions on this work.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia through the project number IFPNC-008-141-2020 and King Abdulaziz University, DSR, Jeddah, Saudi Arabia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia: IFPNC-008-141-2020.

King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

Competing Interests

Sajjad Karim is an Academic Editor for PeerJ.

Author Contributions

- Nofe Alganmi conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

- Arwa Bashanfar conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Reem Alotaibi conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Haneen Banjar conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Sajjad Karim analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Zeenat Mirza performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Heba Abusamra analyzed the data, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Manal Al-Attas analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Shereen Turkistany analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Adel Abuzenadah conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

The Center of Excellence in Genomics Medical Research bioethical committee approval to carry out the study within its facilities. (License #ot KACST: HA-02-J-003).

Data Availability

The following information was supplied regarding data availability:

The data and source code are available at Zenodo: Bashanfar, A. (2024). Uncovering Hidden Genetic Risk Factors for Breast and Ovarian Cancers in BRCA-Negative Women: A Machine Learning Approach in the Saudi Population. Zenodo. DOI [10.5281/zenodo.10926611](https://doi.org/10.5281/zenodo.10926611).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.1942#supplemental-information>.

REFERENCES

- Abualrous ET, Sticht J, Freund C. 2021. Major histocompatibility complex (mhc) class i and class ii proteins: impact of polymorphism on antigen presentation. *Current Opinion in Immunology* 70:95–104 DOI [10.1016/j.coi.2021.04.009](https://doi.org/10.1016/j.coi.2021.04.009).
- Bartha Á, Györfly B. 2019. Comprehensive outline of whole exome sequencing data analysis tools available in clinical oncology. *Cancers* 11(11):1725 DOI [10.3390/cancers11111725](https://doi.org/10.3390/cancers11111725).

- BenAyed-Guerfali D, Kifagi C, BenKridis-Rejeb W, Ammous-Boukhris N, Ayedi W, Khanfir A, Daoud J, Mokdad-Gargouri R. 2022. The identification by exome sequencing of candidate genes in BRCA-negative Tunisian patients at a high risk of hereditary breast/ovarian cancer. *Genes* 13(8):1296 DOI 10.3390/genes13081296.
- Bukhari SNH, Jain A, Haq E, Mehbodniya A, Webber J. 2022. Machine learning techniques for the prediction of B-cell and T-cell epitopes as potential vaccine targets with a specific focus on sars-cov-2 pathogen: a review. *Pathogens* 11(2):146 DOI 10.3390/pathogens11020146.
- Bukhari SNH, Webber J, Mehbodniya A. 2022. Decision tree based ensemble machine learning model for the prediction of Zika virus T-cell epitopes as potential vaccine candidates. *Scientific Reports* 12(1):7810 DOI 10.1038/s41598-022-11731-6.
- Carvalho KSS, da Silva Júnior WJ, da Silveira Regueira Neto M, Silva VC, de Sá Leitão Paiva Júnior S, Balbino VQ, Costa DL, Costa CHN. 2020. Application of next generation sequencing (ngs) for descriptive analysis of 30 genomes of *leishmania infantum* isolates in middle-north brazil. *Scientific Reports* 10(1):12321 DOI 10.1038/s41598-020-68953-9.
- Chen Y-F, Chen Y-T, Chiu W-T, Shen M-R. 2013. Remodeling of calcium signaling in tumor progression. *Journal of Biomedical Science* 20(1):1–10 DOI 10.1186/1423-0127-20-23.
- Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J, Hallberg E, Moore R, Thomas A, Lilyquist J, Feng B, McFarland R, Pesaran T, Huether R, LaDuca H, Chao EC, Goldgar DE, Dolinsky JS. 2017. Associations between cancer predisposition testing panel genes and breast cancer. *JAMA Oncology* 3(9):1190–1196 DOI 10.1001/jamaoncol.2017.0424.
- Cybulski C, Kluźniak W, Huzarski T, Wokołorczyk D, Kashyap A, Jakubowska A, Szwiec M, Byrski T, Dębniak T, Górski B, Sopik V, Akbari MR, Sun P, Gronwald J, Narod SA, Lubiński J. 2015. Clinical outcomes in women with breast cancer and a palb2 mutation: a prospective cohort analysis. *The Lancet Oncology* 16(6):638–644 DOI 10.1016/S1470-2045(15)70142-7.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and vcf tools. *Bioinformatics* 27(15):2156–2158 DOI 10.1093/bioinformatics/btr330.
- Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A, Couch FJ, Southey M, Goldgar DE, Evans DGR, Chenevix-Trench G, Rahman N, Robson M, Domchek SM, Foulkes WD. 2015. Gene-panel sequencing and the prediction of breast-cancer risk. *New England Journal of Medicine* 372(23):2243–2257 DOI 10.1056/NEJMSr1501341.
- Evans DG, van Veen EM, Woodward ER, Harkness EF, Ellingford JM, Bowers NL, Wallace AJ, Howell SJ, Howell A, Lalloo F, Newman WG, Smit MJ. 2021. Gene panel testing for breast cancer reveals differential effect of prior brca1/2 probability. *Cancers* 13(16):4154 DOI 10.3390/cancers13164154.
- Fan B, Wang G, Han H, Li B, Wang L, Qu C, Zhang X, Liu H, Sun Z, Zhou Y, Zhang Y, Wu W. 2022. Machine learning model identifies genomic variation in noise-induced hearing loss via whole-exome sequencing. DOI 10.21203/rs.3.rs-1829497/v2.
- Felicio PS, Grasel RS, Campacci N, de Paula AE, Galvão HCR, Torrezan GT, Sabato CS, Fernandes GC, Souza CP, Michelli RD, Andrade CE, Barros BDDF, Matsushita MM, Revil T, Ragoussis J, Couch FJ, Hart SN, Reis RM, Melendez ME, Tonin PN, Carraro DM, Palmero EI. 2021. Whole-exome sequencing of non-brca1/brca2 mutation carrier cases at high-risk for hereditary breast/ovarian cancer. *Human Mutation* 42(3):290–299 DOI 10.1002/humu.24158.
- Feng L, Wang J, Zhang J, Diao J, He L, Fu C, Liao H, Xu X, Gao Y, Zhou C. 2021. Comprehensive analysis of e3 ubiquitin ligases reveals ring finger protein 223 as a novel

oncogene activated by klf4 in pancreatic cancer. *Frontiers in Cell and Developmental Biology* 9:738709 DOI 10.3389/fcell.2021.738709.

- Figueiredo de Sá BSL, Rezende AM, A. de-Melo-Neto OP, Brito MEFd, Brandao Filho SP. 2019.** Identification of divergent *leishmania* (*Viannia*) *braziliensis* ecotypes derived from a geographically restricted area through whole genome analysis. *PLOS Neglected Tropical Diseases* 13(6):e0007382 DOI 10.1371/journal.pntd.0007382.
- Godet I, Gilkes DM. 2017.** Brca1 and brca2 mutations and treatment strategies for breast cancer. *Integrative Cancer Science and Therapeutics* 4(1):1–7 DOI 10.15761/ICST.1000228.
- Gotea V, Gartner JJ, Qutob N, Elnitski L, Samuels Y. 2015.** The functional relevance of somatic synonymous mutations in melanoma and other cancers. *Pigment Cell & Melanoma Research* 28(6):673–684 DOI 10.1111/pcmr.12413.
- Gracia-Aznarez FJ, Fernandez V, Pita G, Peterlongo P, Dominguez O, de la Hoya M, Duran M, Osorio A, Moreno L, Gonzalez-Neira A, Rosa-Rosa JM, Sinilnikova O, Mazoyer S, Hopper J, Lazaro C, Southey M, Odefrey F, Manoukian S, Catucci I, Caldes T, Lynch HT, Hilbers FSM, van Asperen CJ, Vasen HFA, Goldgar D, Radice P, Devilee P, Benitez J. 2013.** Whole exome sequencing suggests much of non-brca1/brca2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PLOS ONE* 8(2):e55681 DOI 10.1371/journal.pone.0055681.
- Grasel RS, Felicio PS, de Paula AE, Campacci N, de Oliveira Garcia FA, de Andrade ES, Evangelista AF, Fernandes GC, da Silva Sabato C, Marchi PD, de Pádua Souza C, de Paula CAA, Torrezan GT, de Campos Reis Galvão H, Carraro DM, Palmero EI. 2020.** Using cosegregation and loss of heterozygosity analysis to define the pathogenicity of unclassified variants in hereditary breast cancer patients. *Frontiers in Oncology* 10:571330 DOI 10.3389/fonc.2020.571330.
- Hooshmand A. 2020.** Naive bayesian machine learning to diagnose breast cancer. DOI 10.21203/rs.3.rs-60997/v1.
- Huang Y-S, Hsu C, Chune Y-C, Liao I-C, Wang H, Lin Y-L, Hwu W-L, Lee N-C, Lai F. 2022.** Diagnosis of a single-nucleotide variant in whole-exome sequencing data for patients with inherited diseases: Machine learning study using artificial intelligence variant prioritization. *JMIR Bioinformatics and Biotechnology* 3(1):e37701 DOI 10.2196/37701.
- Huang DW, Sherman BT, Lempicki RA. 2009.** Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 4(1):44–57 DOI 10.1038/nprot.2008.211.
- Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023.** KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* 51(D1):D587–D592 DOI 10.1093/nar/gkac963.
- Kanehisa M, Goto S. 2000.** KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28(1):27–30 DOI 10.1093/nar/28.1.27.
- Kanehisa M, Sato Y. 2020.** KEGG mapper for inferring cellular functions from protein sequences. *Protein Science* 29(1):28–35 DOI 10.1002/pro.3711.
- Kanehisa M, Sato Y, Kawashima M. 2022.** KEGG mapping tools for uncovering hidden features in biological data. *Protein Science* 31(1):47–53 DOI 10.1002/pro.4172.
- Kurian AW, Gong GD, John EM, Johnston DA, Felberg A, West DW, Miron A, Andrulis IL, Hopper JL, Knight JA, Ozcelik H, Dite GS, Apicella C, Southey MC, Whittemore AS. 2011.** Breast cancer risk for noncarriers of family-specific brca1 and brca2 mutations: findings from the breast cancer family registry. *Journal of Clinical Oncology* 29(34):4505–4509 DOI 10.1200/JCO.2010.34.4440.

- Lee NY, Hum M, Amali AA, Lim WK, Wong M, Myint MK, Tay RJ, Ong P, Samol J, Lim CW, Ang P, Tan M, Lee S, Lee ASG. 2022. Whole-exome sequencing of brca-negative breast cancer patients and case-control analyses identify variants associated with breast cancer susceptibility. *Human Genomics* 16(1):61 DOI 10.1186/s40246-022-00435-7.
- Liu L, Sun X, Yuan C, Liu H. 2021. Relationship between hla-drb1 gene polymorphism and breast cancer: a protocol for systematic review and meta-analysis. *Medicine* 100(12):e25078 DOI 10.1097/MD.00000000000025078.
- Magi A, D'Aurizio R, Palombo F, Cifola I, Tattini L, Semeraro R, Pippucci T, Giusti B, Romeo G, Abbate R, Gensini GF. 2015. Characterization and identification of hidden rare variants in the human genome. *BMC Genomics* 16:1–16 DOI 10.1186/s12864-015-1481-9.
- Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A, Balakishnan B, Liang R, Zhang Y, Lyon S, Beutler B, Whittle B, Bertram EM, Enders A, Goodnow CC, Andrews TD. 2015. Comparison of predicted and actual consequences of missense mutations. *Proceedings of the National Academy of Sciences of the United States of America* 112(37):E5189–E5198 DOI 10.1073/pnas.1511585112.
- National Cancer Center of the Saudi Health Council. 2024. Cancer incidence report 2020. Available at <https://shc.gov.sa/Arabic/NCC/Activities/AnnualReports/Cancer%20Incidence%20Report%202020.pdf>.
- Petrucelli N, Daly MB, Pal T. 2022. Brca1-and brca2-associated hereditary breast and ovarian cancer. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Amemiya A, eds. *GeneReviews*[®] [Internet]. Seattle (WA): University of Washington, Seattle, 1993–2024.
- Rabbani B, Tekin M, Mahdieh N. 2014. The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics* 59(1):5–15 DOI 10.1038/jhg.2013.114.
- Reeves E, James E. 2017. Antigen processing and immune regulation in the response to tumours. *Immunology* 150(1):16–24 DOI 10.1111/imm.12675.
- Sharma Y, Miladi M, Dukare S, Boulay K, Caudron-Herger M, Groß M, Backofen R, Diederichs S. 2019. A pan-cancer analysis of synonymous mutations. *Nature communications* 10(1):2569 DOI 10.1038/s41467-019-10489-2.
- Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. 2022. David: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Research* 50(W1):W216–W221 DOI 10.1093/nar/gkac194.
- Sung H, Ferlay J, Siegel RL, 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 DOI 10.3322/caac.21660.
- Suwinski P, Ong C, Ling MH, Poh YM, Khan AM, Ong HS. 2019. Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Frontiers in Genetics* 10:49 DOI 10.3389/fgene.2019.00049.
- Trakadis YJ, Sardaar S, Chen A, Fulginiti V, Krishnan A. 2019. Machine learning in schizophrenia genomics, a case-control study using 5,090 exomes. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 180(2):103–112 DOI 10.1002/ajmg.b.32638.
- Vihinen M. 2022. When a synonymous variant is nonsynonymous. *Genes* 13(8):1485 DOI 10.3390/genes13081485.
- Wadapurkar R, Bapat S, Mahajan R, Vyas R. 2023. Machine learning approaches for prediction of ovarian cancer driver genes from mutational and network analysis. *Data Technologies and Applications* 58(1):62–80 DOI 10.1108/DTA-03-2022-0096.

- Wang K. 2010.** Annovar documentation. Available at <https://annovar.openbioinformatics.org/en/latest/>.
- Wieczorek M, Abualrous ET, Sticht J, Álvaro-Benito M, Stolzenberg S, Noé F, Freund C. 2017.** Major histocompatibility complex (mhc) class i and mhc class ii proteins: conformational plasticity in antigen presentation. *Frontiers in Immunology* **8**:292 DOI [10.3389/fimmu.2017.00292](https://doi.org/10.3389/fimmu.2017.00292).
- Wu L, Zhang Y, Zheng C, Zhao F, Lin Y. 2023.** Gamma-aminobutyric acid type a receptor subunit delta (gabrd) inhibits breast cancer progression by regulating the cell cycle. *Iranian Journal of Public Health* **52(3)**:542 DOI [10.18502/ijph.v52i3.12137](https://doi.org/10.18502/ijph.v52i3.12137).
- Yin L, Duan J-J, Bian X-W, Yu S-c. 2020.** Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research* **22(1)**:1–13 DOI [10.1186/s13058-020-01296-5](https://doi.org/10.1186/s13058-020-01296-5).
- Zeiser R, Blazar BR. 2017.** Acute graft-versus-host disease—biologic process, prevention, and therapy. *New England Journal of Medicine* **377(22)**:2167–2179 DOI [10.1056/NEJMra1609337](https://doi.org/10.1056/NEJMra1609337).
- Zhang X, Li R. 2018.** Brca1-dependent transcriptional regulation: implication in tissue-specific tumor suppression. *Cancers* **10(12)**:513 DOI [10.3390/cancers10120513](https://doi.org/10.3390/cancers10120513).