Gene 852 (2023) 147070

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Identification of novel exonic variants contributing to hereditary breast and ovarian cancer in west Indian population

Bhargav N. Waghela [a], Ramesh J. Pandit [a], Apurvasinh Puvar [a], Franky D. Shah [b], Prabhudas S. Patel [b], Hemangini Vora [b], Harsh Sheth [c], Bhoomi Tarapara [b], Shashank Pandya [b], Chaitanya G. Joshi [a], Madhvi N. Joshi [a,*]

[a] Gujarat Biotechnology Research Centre, Department of Science and Technology, Government of Gujarat, Gandhinagar, Gujarat 382011, India
[b] Gujarat Cancer Research Institute, Civil Hospital, Ahmedabad, Gujarat 380016, India
[c] Frige House, Jodhpur Gam Rd, Satellite, Ahmedabad, Gujarat 380015, India

A R T I C L E I N F O

Edited by John Randolph Hawse

Keywords:
Autosomal dominant diseases
HBOC
Next-generation sequencing
Non-BRCA genes
Whole-exome sequencing

A B S T R A C T

Breast and ovarian cancers are the most common cancer types in females worldwide and in India. Patients with these cancers require an early diagnosis which is essential for better prognosis, treatment and improved patient survival. Recently, the utilization of next-generation sequencing (NGS)-based screening has accelerated molecular diagnosis of various cancers. In the present study, we performed whole-exome sequencing (WES) of 30 patients who had a first or second-degree relative with breast or ovarian cancer and are tested negative for *BRCA1/2* or other high and moderate-risk genes reported for HBOC. WES data from patients were analyzed and variants were called using bcftools. Functional annotation of variants and variant prioritization was performed by Exomiser. The clinical significance of variants was determined as per ACMG classification using Varsome tool. The functional analysis of genes was determined by STRING analysis and disease association was determined by open target tool. We found novel variants and gene candidates having significant association with HBOC conditions. The genes identified by exomiser (phenotype score > 0.75) are associated with various biological processes such as DNA integrity maintenance, transcription regulation, cell cycle regulation, and apoptosis. Our findings provide novel and prevalent gene variants associated with the HBOC condition in the West Indian population which could be further studied for early diagnosis and better prognosis of HBOC.

1. Introduction

Breast cancer (BC) and ovarian cancer (OC) are the most prominent gynecological cancers among females worldwide as well as in India (Edlich et al., 2005; Bray et al., 2018). In 2020, 178,361 breast cancer and 45,701 ovarian cancer new cases were diagnosed in India which account for 26.3% and 6.7% of the total cancer cases, respectively (Sung et al., 2021) and the number of death due to breast and ovarian cancer is 90,408 and 32,077, respectively. The incidences of sporadic breast cancer or ovarian cancer are frequent in adult females which increase with the age (Anders et al., 2009; Bakkach et al., 2017). However, inherited mutations are also responsible for the early onset of these cancers (Szabo and King, 1995). Hereditary breast and ovarian cancer (HBOC) syndrome is a major condition responsible for approximately 90% of inherited breast and ovarian cancer (Grabenstetter et al., 2020). HBOC is an autosomal-dominant inherited condition associated with a higher risk of early-onset of breast cancer and ovarian cancer in multiple family members (Lynch et al., 2013). Similar chances of HBOC prevalence have been reported in both females and males. In males, the HBOC attributes to a higher risk of male breast cancer, melanoma, prostate cancer, and pancreatic cancer. HBOC accounts for 5–10% of cancer patients having breast and/or ovarian cancer and is mainly associated with germline mutations in *BRCA1* or *BRCA2* genes (Foulkes, 2008). The pathogenic variants of *BRCA 1/2* are major responsible for the HBOC condition. More than 3000 variants of *BRCA 1/2* are reported in the clinvar database (Jarhelle et al., 2019). However, the mutational profiles of *BRCA 1/2* are highly variable across various populations of the world and India (Sharma-Oates et al., 2018).

Abbreviations: HBOC, Hereditary Breast and Ovarian Cancer; WES, Whole-Exome Sequencing; NGS, Next-Generation Sequencing; OC, Ovarian Cancer; BC, Breast Cancer; HPO, Human Phenotype Ontology; NCCN, National Comprehensive Cancer Network.

* Corresponding author.
E-mail address: madhvimicrobio@gmail.com (M.N. Joshi).

https://doi.org/10.1016/j.gene.2022.147070
Received 25 May 2022; Received in revised form 21 October 2022; Accepted 18 November 2022
Available online 24 November 2022
0378-1119/© 2022 Published by Elsevier B.V.

The mutational landscape of multiple genes apart from *BRCA1* and *BRCA2* are involved in the predisposition of HBOC (Jarhelle et al., 2019). The National Comprehensive Cancer Network (NCCN) has provided guidelines for the clinical management of HBOC. To date, NCCN has compiled 21 genes in which, the pathogenic mutations propagate to HBOC condition (Daly et al., 2020). According to the NCCN guidelines, individuals having a strong family history of breast and/or ovarian cancers should undergo the genetic assessment test for early diagnosis and better clinical management (Daly et al., 2020).

Several studies have reported that the deleterious mutations in genes other than *BRCA1* or *BRCA2* also promotes HBOC conditions (Yoshida, 2020). In our previous study, we have designed and validated the next generation sequencing (NGS) based multigene panel covering the mutations in 14 genes (high to moderate risk genes associated with HBOC) which including *BRCA1*, *BRCA2* and several non-BRCA genes for the diagnosis of HBOC in the west Indian population (Kadri et al., 2020). We also identified a set of pathogenic variants, VUS, and novel variants which readily associated with HBOC risk. Further, we suggested that the mutational profiles of Indian HBOC patients are different from other populations suggesting clinical guidelines and gene-disease relations reported globally may partly support clinical management of HBOC in the Indian population (Kadri et al., 2020). Interestingly, we noticed that several individuals were negative for the multi-gene panel but suffering from breast and/or ovarian cancer. This observation indicates that there may be several uncharacterized genes which progresses the familial breast and/or ovarian cancer. Therefore, we were keen to elucidate the novel gene variants involve in HBOC. To test the hypothesis, we selected 30 patients whom were negative for the multi-gene panel for the study.

Whole-exome sequencing (WES) is an efficient technique for clinical diagnosis and is used to identify the mutational profiling of genetic diseases (Berberich et al., 2018). The WES promisingly utilized for the screening of mutational profiles of genes involved in BC and/or OC (Felicio et al., 2021). In the present study, we performed WES of 30 HBOC patients having a strong family history of BC and/or OC (i.e., the patients having at least one first- or second-degree relative with breast cancer diagnosed before age 70) but are negative for our previously designed population-specific multi-gene panel. We annotated and prioritized the gene variants by the exomiser tool (Smedley et al., 2015) and further analyzed them based on their disease association and prevalence.

2. Materials and methods

2.1. Patient selection

The patients were selected based on their genetic counseling of breast and ovarian cancer at Gujarat Cancer and Research Institute (GCRI), Ahmedabad, Gujarat, India. Patients having disease earlier or undergoing treatment with familial breast or ovarian cancer in first/second-degree relatives were selected for the study. The patients were selected based on ICMR guidelines (https://www.icmr.nic.in/sites/default/files/guidelines/ICMR_Ethical_Guidelines_2017). Clinical and pathologic details of all the patients were retrieved from their medical records. All patients signed an informed consent form approved by the institutional review board at GCRI. Further, 30 HBOC patients who were negative for BRCA mutations were selected for the study. Out of 30 patients, 23 patients were diagnosed with breast cancer, 6 patients with ovarian cancer, 1 patient with breast and ovarian cancer. The age of 20 patients was below 50 years with a mean age of 47 years at the time of diagnosis. All 30 cases were unrelated individuals from singular various families.

2.2. Sample preparation

Genomic DNA was isolated from blood samples of selected patients using QIAamp DNA Blood Mini kit (QIAGEN, Germany) as per the manufacturer's instructions. The DNA concentration was determined by Qubit Fluorometer 4.0® (Thermo Fisher Scientific, USA) and the purity of DNA was determined by QIAxpert (QIAGEN, Germany).

2.3. Whole-exome library preparation and sequencing

In this study, we performed amplicon-based exome sequencing. For library preparation, genomic DNA (approx. 100 ng) of each subject was amplified using Ampliseq RDY panel kit as per manufacturer's instructions (Thermo Fisher Scientific, USA). Ampliseq exome kit includes 2,93,903 primer pairs that cover 97% of CCDS with a 5 bp padding region around exons. Further, the libraries were prepared with Ion AmpliSeq™ Library Kit plus (Thermo Fisher Scientific, USA). The library profile was checked using DNA high sensitivity assay kit on Bio-analyser 2100 (Agilent Technologies, USA) and library quantification was further done with Ion Library TaqMan™ Quantitation Kit on qPCR (Thermo Fisher Scientific, USA). Thereafter, each library was diluted to 100 pmol and all the libraries were pooled in equimolar concentration and sequencing was carried out on the Ion Proton and Ion S5 platform with Ion PI and 540 chip respectively with 200 bp chemistry.

2.4. Data analysis

2.4.1. Raw data quality assessment, genome assembly, and variant calling

Raw sequence data in FASTQ format was assessed using the FASTQC toolkit (v.0.11.5)(Andrews, 2010). Raw sequences were further trimmed and filtered using PRINSEQ-lite v.0.20.4 (Schmieder and Edwards, 2011) in which 5 bp from left end and 10 bp from the right end were trimmed, the sequence length lower than 50 bp and quality mean values less than 20 were removed. clean reads were mapped on the hg19 reference genome with MEM algorithms of BWA software. Aligned BAM files were further proceeding for variant calling with mpileup and call algorithms of bcftools (Supplementary Fig. 1).

2.4.2. Variant annotation & prioritization

The functional annotation of variants and variant prioritization was performed by Exomiser (Version 12.1.0 available at https://github.com/exomiser/Exomiser). Exomiser annotates, filters and prioritize the disease causing variants based on the HPOIDs (Human Phenotype Ontology identifiers/terms). The VCF files obtained from the analysis pipeline (Supplementary Fig. 1) were used as an input in the (.yml) file in exomiser. The HPOIDs for breast and ovarian cancer were entered and inheritance mode was autosomal dominant (0.1) in exomiser. The variants prioritization was performed based on defined criteria. We have defined as follows: (a) variant frequency data using THOUSAND GENOMES, TOPMED, UK10K, ESP and GNOMAD, (b) pathogenicity source such as POLYPHEN, MUTATION_TASTER, SIFT and CADD The scores of pathogenicity prediction tools were Polyphen ($>0.956|>0.446$), Mutation_Taster (>0.94), SIFT (<0.06), and CADD (>0.483). We also included several phenotype similarity algorithms such as human phenotypes in hiPhivePrioritiser.

2.5. Variant Classification and functional analysis:

The variants were analyzed in varsome suite (https://varsome.com/) (Kopanos et al., 2019) and Clinvar database (https://www.ncbi.nlm.nih.gov/clinvar/) (Landrum et al., 2014) and classified according to the American College of Medical Genetics and Genomics (ACMG) recommendations (Richards et al., 2015). The variants were classified into five categories such as the pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign. The oncogenic potential of variants was predicted using CScape tool (http://CScape.biocompute.org.uk/) (Rogers et al., 2017). Further, the analysis of functional association of genes were performed in STRING (https://string-db.org/). Moreover, functional annotation was performed in the Database for Annotation, Visualization and Integrated Discovery

(DAVID) v6.8 (https://david.ncifcrf.gov/tools.jsp) (Huang et al., 2007).

3. Results

3.1. Patient selection

In the present study, we have selected 30 unrelated patients with breast cancer and/or ovarian cancer and were negative to panel genes designed previously for the West Indian population. The distribution of the patients based on disease and age of onset has been represented in supplementary Fig. 2. The whole-exome analysis of 30 patients was performed and the data were analyzed using various *in silico* tools. Further, the variants were annotated for functional consequences by exomiser based on the Human Phenotype Ontology (HPO) for BC and OC. Initially, we found a total of 4,56,741 unique variants (more than 1.2 million variant entries) of 18,594 genes among the 30 patients with an average of 40,000 variants per patient. We examined the variants based on two major criteria, (i) the disease association using phenotype score of exomiser and (ii) prevalence of disease associated variants among the patients (Fig. 1).

3.2. Variants analysis using Exomiser

We analyzed the variants in Exomiser and selected the top 50 variants (Based on Phenotype score) from each patient sample. By examining those variants; we identified a total of 687 variants of 81 genes from 30 patients. From this, we identified 223 variants of 22 genes which are associated with high to moderate risk of HBOC (reported in NCCN guidelines and/or included in previously designed customized gene panel for HBOC). The remaining 464 variants of 59 genes have not been reported for their association with HBOC (not included in the NCCN guideline for HBOC).

3.3. Analysis of disease-associated variants identified from exomiser for their Pathogenicity

To analyze the pathogenic effects of respective variation in the exome of HBOC patients, we analyzed the 464 variants of 59 genes by their respective SIFT score (<0.06) and Polyphen scores (>0.956|> 0.446) and found 12 variants of 9 genes. The genes include *COL14A1, FAN1, GNAS, OPCML, PHB, PIK3CA, POLE, PPM1D, RAD54L, RNF43, TERT,* and *TWIST1*. We evaluated these gene variants in varsome and classified them based on the ACMG guidelines. Out of the 12 variants, 11 variants, *COL14A1*:c.529G > T, *GNAS*:c.478A > G, *PHB*:c.505 T > C, *PIK3CA*:c.31 T > G, *PIK3CA*:c.32G > T, *POLE*:c.6302C > A, *POLE*: c.6344A > G, *PPM1D*:c.1579G > A, *RAD54L*:c.345A > C, *RAD54L*: c.579C > G, and *RNF43*:c.379C > T, were found VUS and 1 variant *FAN1*:c.1589 T > C was likely benign. Further, we analyzed the oncogenic properties of variants and found variants of *COL14A1, OPCML, PHB, PIK3CA, POLE, PPM1D, RAD54L, RNF43, TERT,* and *TWIST1* were oncogenic, *FAN1* and *GNAS* were benign. We also confirmed our results by visualizing the chromosomal position of each variation in IGV

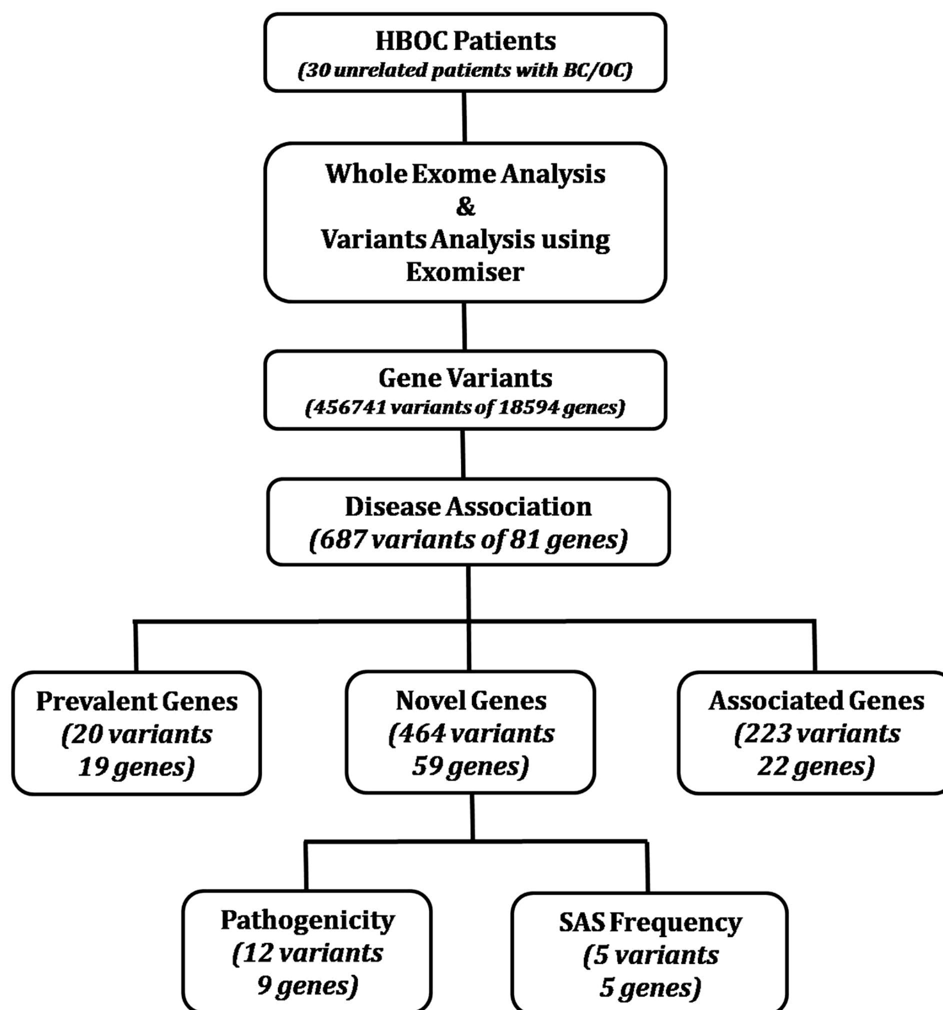

Fig. 1. Various approaches utilized to analyze the gene variants in the west Indian patients. We utilized majorly two approaches to analyze gene variants, (i) Based on prevalence among the patients, and (ii) Based on the exomiser output (top 50 entries of variants were considered for analysis).

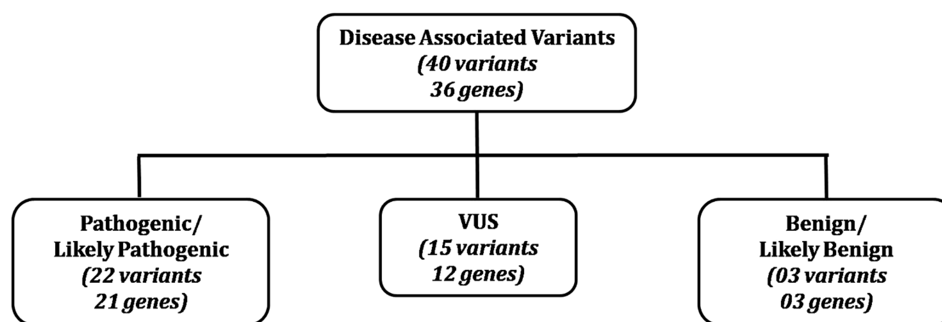

Fig. 2. Graphical presentation of the disease associated variants based on their clinical significance.

(Table 1). Further, the protein–protein interaction and functional association analysis using STRING suggested that out of 59 genes, 54 genes were found to be associated with high and moderate risk genes reported for HBOC except *COL14A1, AAGAB, OPCML, SEC23B,* and *DMPK*. (Supplementary Figure 3). The analysis of disease association with BC and OC revealed that the majority of them are strongly associated with BC and OC. Moreover, the functional annotation analysis suggested that the majority of genes are also involved in the biological processes associated with DNA integrity maintenance, transcriptional regulation, cell cycle, and apoptosis.

3.4. Analysis of disease-associated variants identified from exomiser for the South Asian prevalence

We analyzed the 464 variants of 59 genes for their frequency in the South Asian population. We found 5 variants of 5 genes encompassing South Asian population frequency. The genes include *KRAS, MRE11, PPM1D, RAD54L,* and *RNF43*. Out of these 5 variants, *KRAS*:c.547A > G was benign, *MRE11*:c.1441del and *RAD54L*:c.2209C > T were pathogenic, and *PPM1D*:c.1579G > A and *RNF43*:c.379C > T were VUS as per ACMG guidelines in varsome. Further, we analyzed the oncogenic properties of variants and found that variants of *KRAS, PPM1D, RAD54L,* and *RNF43* are oncogenic (Table 2).

3.5. Analysis of prevalent disease-associated variants identified from exomiser among patients

We analyzed the 687 variants of 81 genes for their prevalence in west Indian population and scrutinized the variants based on their prevalence (≥25%) patients. We found 20 variants of 19 genes having higher prevalence (variants identified>8 patients out of 30 patients). Out of these 20 variants, 16 variants, *CTNNB1*:c.718del, *WNT10A*:c.307del, *SMAD4*:c.130_131insA, *PALLD*:c.88del, *PRLR*:c.1251del, *HMMR*:c.470del, *MITF*:c.598_599del, *GNAS*:c.106del, *POLD1*:c.262del, *KEAP1*:c.1652del, *ESR1*:c.677dup, *TERF2IP*:c.1116del, *RAD51*:c.60dup, *POT1*:c.1789dup, *SEC23B*:c.82_83del, , , *FGFR2*:c.2096del, were pathogenic, 3 variants, *AKT1*:c.722del, , *POLD1*:c.66_67insG and *TERT*:c.3327del were likely pathogenic and 01 variant, *CDKN2B*:c.173del, was VUS in varsome (as per ACMG Guidelines). (Table 3). These genes were further analyzed by STRING and found that the 29 genes showed prominent association with high to moderate HBOC genes (Supplementary Figure 4). Therefore, the identified gene variants may possess the potential for diagnosis purpose for the early detection of HBOC. Moreover, the functional annotation analysis suggested that the majority of genes are also involved in the biological processes associated with DNA integrity maintenance, transcriptional regulation, and cell cycle (Supplementary Figure 5).

3.6. Analysis of variants identified from exomiser based on high/ moderate-risk genes of HBOC

We analyzed 687 variants of 81 genes for their frequency. We scrutinized the variants based on their frequency in patients. For the present study, we scrutinized genes and their variants having a high to moderate risk of HBOC and found 223 variants of 22 genes. The genes include *BRCA1, BRCA2, CDH1, BRIP1, NBN, PALB2, TP53, ATM, STK11, BARD1, CHEK2, RAD51C, PTEN, PMS2, MSH2, MSH6, CDKN2A, MLH1, EPCAM, RAD51D, RAD50* and *CASP8*. Based on the pathogenicity score, we identified the 3 variants of 3 genes includes *TP53, , STK11,* and *CASP8*. Of them, *TP53*:c.787A > G was predicted benign, , *STK11*:c.191A > G was likely pathogenic, and *CASP8*:c.811 T > C was VUS. The *TP53*:c.787A > G has been reported in the South Asian population (Table 4).

4. Discussion

Breast and ovarian cancer are the leading cause of cancer-related death in females (Bray et al., 2018). Conventional therapies such as chemotherapy and radiation therapy are utilized to treat the patients however, it concerned with severe side effects, low rate of patient survival, and poor health quality (Moo et al., 2018). These problems instigate to identify alternative approaches such as early diagnosis for efficient prognosis and improved patient survival. Recently, identification of the mutational landscape of cancer patients using next-generation sequencing (NGS) has been well practiced in the field of cancer genomics and also essential for the early diagnosis of breast and ovarian cancer (Castera et al., 2014). Previously, to identify the effects of potent variants in the patients, we have designed a customized gene panel comprising 14 genes. We validated 144 patient samples and identified prominent gene variants having clinical significance. Interestingly, few patients were negative for panel genes and diagnosed with HBOC (Kadri et al., 2020). This observation led us to investigate novel mutation/s in the whole exonic region which may have a role in these cancers. The WES analysis is utilized for the screening of mutational profiles of genes involved in BC and/or OC (Felicio et al., 2021). In the present study, we have screened 30 patients having breast and/or ovarian cancer for amplicon-based exome sequencing in which previously no mutation was identified with our customized gene panel of 14 gene genes (*BRCA1, BRCA2, TP53, PTEN, CDH1, STK11, BARD1, ATM, BRIP1, CHEK2, ERBB2, NBN, PALB2, RAD51C*). We identified novel gene variants involved in HBOC progression in the West Indian patient cohort based on their disease association and their prevalence. We found various gene variants that may be closely related to *BRCA1* and *BRCA2* in breast and/or ovarian cancer. We found novel variants having a higher disease association and prevalence in the west India patient cohort which requires further characterization. Moreover, our analysis based on the functional consequences identified the gene candidates involved in various cancer signal transduction pathways which may progress breast and/or ovarian cancer.

In the present study, we screened 4,56,741 variants of 18,594 genes

Table 1
Variants identified from Exomiser and their pathogenicity among West Indian patients.

Gene	Gene Name	Chromosome	Position	Ref	Alt	Functional Class	HGVS	ACMG Classification	Prevalence	Polyphen score	SIFT Score	Phenotype Score	cScape	Major Allele Frequency	Minor Allele Frequency
COL14A1	Collagen type XIV alpha 1 Chain	8	121,209,122	G	T	Missense variant	ENST00000247781.3: c.529G > T:p. (Val177Phe) (NM_021110.4)	Uncertain Significance	1	0.999	0.999	0.747134	Oncogenic (high conf.)	34 (45%)	41 (55%)
FAN1	FANCD2 and FANCI associated nuclease 1	15	31,203,030	T	C	Missense variant	ENST00000561594.1: c.1589 T > C:p. (Leu530Pro) (NM_001146096.1)	Likely Benign	1	0.016	1	0.720564	Benign	70 (51%)	66 (49%)
GNAS	GNAS complex locus	20	57,415,639	A	G	Missense variant	ENST00000313949.7: c.478A > G:p. (Thr160Ala) (NM_016592.5)	Uncertain Significance	1	0.015	1	0.755637	Benign	15 (71%)	6 (29%)
PHB	Prohibitin	17	47,486,409	A	G	Missense variant	ENST00000300408.3: c.505 T > C:p. (Ser169Pro) (NM_002634.4)	Uncertain Significance	1	0.988	0.999	0.842652	Oncogenic (high conf.)	36 (55%)	30 (45%)
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha	3	178,916,644	T	G	Missense variant	ENST00000263967.3: c.31 T > G:p.(Trp11Gly) (NM_006218.4)	Uncertain Significance	1	0.994	1	0.968161	Oncogenic (high conf.)	91 (57%)	69 (43%)
		3	178,916,645	G	T	missense_variant	ENST00000263967.3: c.32G > T:p.(Trp11Leu) (NM_006218.4)	Uncertain Significance	1	0.994	0.997	0.968161	Oncogenic (high conf.)	33 (73%)	12 (27%)
POLE	DNA polymerase epsilon, catalytic subunit	12	133,208,929	G	T	missense_variant	ENST00000320574.5: c.6302C > A:p. (Ala2101Asp) (NM_006231.4)	Uncertain Significance	1	0.898	0.993	0.755901	Oncogenic	33 (43%)	44 (57%)
		12	133,202,890	T	C	missense_variant	ENST00000320574.5: c.6344A > G:p. (Asp2115Gly) (NM_006231.4)	Uncertain Significance	1	0.973	0.989	0.755901	Oncogenic (high conf.)	48 (38%)	77 (62%)
PPM1D	protein phosphatase, Mg2+/Mn2 + dependent 1D	17	58,740,674	G	A	missense_variant	ENST00000305921.3: c.1579G > A:p. (Glu527Lys) (NM_003620.4)	Uncertain Significance	1	0.985	1	0.842652	Oncogenic	88 (50%)	87 (50%)
RAD54L	DNA Repair And Recombination Protein RAD54-Like	1	46,725,709	A	C	missense_variant	ENST00000371975.4: c.345A > C:p. (Lys115Asn) (NM_001142548.1)	Uncertain Significance	1	0.002	0.963	0.842652	Oncogenic	71 (82%)	16 (18%)
		1	46,726,500	C	G	missense_variant	ENST00000371975.4: c.579C > G:p. (Cys193Trp) (NM_001142548.1)	Uncertain Significance	1	0.999	1	0.842652	Oncogenic	49 (51%)	47 (49%)
RNF43	Ring finger protein 43	17	56,440,958	G	A	missense_variant	ENST00000407977.2: c.379C > T:p. (Arg127Trp) (NM_017763.6)	Uncertain Significance	1	0.987	1	0.788733	oncogenic (high conf.)	67 (51%)	64 (49%)

Ref.: Reference base present in genome sequence, Alt: Altered base present in sequence, HGVS: Human Genome Variation Society nomenclature, ACMG Classification: Classification of variants as per ACMG Guidelines, Prevalence: Prevalence of variant among number of patients, Polyphen score: (>0.956|>0.446), SIFT Score (<0.06), Phenotype Score: Score generated by Exomiser using HPO-annotated disease and other annotation present in humans, cScape: a tool used to predict the variant (single-point mutations) for their oncogenic potential.
* Major allele frequency and minor allele frequency were indicated based on the visualization in IGV.

Table 2
Variants identified from Exomiser present in south Asian population.

Gene	Gene Name	Functional Class	HGVS	ACMG Classification	Prevalence	Polyphen Score	SIFT Score	DBSNP ID	ExAC SAS Frequency	Gene Pheno Score	cScape Analysis	Major Allele Frequency	Minor Allele Frequency
KRAS	KRAS proto-oncogene, GTPase	Missense variant	KRAS: ENST00000256078.4: c.547A > G:p.(Ile183Val)	Benign	1	0.009	0.442	rs529925358	0.030285	0.842652	Oncogenic	65 (57%)	50 (43%)
MRE11	MRE11 homolog, double-strand break repair nuclease	Frameshift truncation	MRE11: ENST00000323929.3: c.1441del:p.(Thr481Hisfs*43)	Pathogenic	1	.	.	rs137852762	0.006058	0.844771	–	16 (62%)	10 (38%)
PPM1D	Protein phosphatase, Mg2+/Mn2 + dependent 1D	Missense variant	PPM1D: ENST00000305921.3: c.1579G > A:p.(Glu527Lys)	Uncertain Significance	1	0.985	1	rs564827577	0.012126	0.842652	Oncogenic	88 (50%)	87 (50%)
RAD54L	DNA Repair And Recombination Protein RAD54-Like	Stop gained	RAD54L: ENST00000371975.4: c.2209C > T:p.(Gln737*)	Pathogenic	1	.	.	rs758653425	0.006691	0.842652	Oncogenic (high conf.)	40 (53%)	36 (47%)
RNF43	Ring finger protein 43	Missense variant	RNF43: ENST00000407977.2: c.379C > T:p.(Arg127Trp)	Uncertain Significance	1	0.987	1	rs369636118	0.006816	0.788733	Oncogenic (high conf.)	67 (51%)	64 (49%)

HGVS: Human Genome Variation Society nomenclature, ACMG Classification: Classification of variants as per ACMG Guidelines, Prevalence: Prevalence of variant among number of patients, Polyphen score: (>0.956|>0.446), SIFT Score (<0.06), DBSNP ID: Unique identifier assigned to the SNP in database, ExAC_SAS_Frequency: South Asian in Exome Aggregation Consortium (Available in gnomAD), Gene Pheno Score: The score indicated in exomiser output. cScape: a tool used to predict the variant (single-point mutations) for their oncogenic potential.
* Major allele frequency and minor allele frequency were indicated based on the visualization in IGV.

to identify the novel variants having higher disease association and prevalence, we analyzed the mutations based on the Phenotype score of exomiser (top 50 entries) and identified 687 variants of 81 genes which were further screened based on their pathogenicity. We found 15 variants of 12 genes, of them 14 were identified as VUS and had oncogenic potential. *COL14A1* have been reported to associated with breast tumors and proliferation and migration of breast cancer cells (Guo et al., 2014). *FAN1* functions as nuclease and involved in DNA inter-strand cross-link repair. Mutation in *FAN1* impairs DNA Repair and causes hereditary colorectal cancer (Seguí et al., 2015). *GNAS*:c.478A > G was also found VUS however predicted benign (Table 1 & Supp. Table 1). *GNAS* encodes the α-subunit of the stimulatory G protein which is associated with the actions of various hormones and endogenous molecules. The mutation in the *GNAS* is associated with several pathological consequences (Turan and Bastepe, 2013). *PHB* regulates cell proliferation, resistance and metastasis signalling in various malignancies (Rajalingam et al., 2005). *PIK3CA* Mutations contributes to the metastatic breast cancer and the mutational landscape has been utilized for luminal breast cancer, HER2-Negative and Metastatic Breast Cancer (Fusco et al., 2021). The germline pathogenic variants in the *POLE* has been reported to involve in familial cancers (Mur et al., 2020). *PPM1D* encodes a serine threonine phosphatase which modulates tumour suppressor pathways has been reported to be amplified in approximately 8% of breast cancers (Lambros et al., 2010). The mutational patterns in *RAD54L* has been included for the study of breast cancer (https://www.mycancergenome.org/). The *RNF43*, an E3 ligase that ubiquitinates and degrades Wnt receptors and inhibits Wnt signalling. It is frequently mutated in various malignancies such as colon, stomach and endometrial cancers.

Further, by analyzing the South Asian population frequency, we found 5 variants of 5 genes, of them *MRE11*:c.1441del and *RAD54L*:c.2209C > T were pathogenic. The variant *RAD54L*:c.2209C > T was predicted as an oncogenic. Further, we analyzed the prevalent variants among patients having frequency $\geq$ 25% (8/30) with higher phenotype scores and found 20 variants of 20 genes. Of them, 16 variants were pathogenic, 3 variants were likely pathogenic and 1 variant was VUS in varsome. The functional analysis revealed that the 29 genes have a prominent association with high to moderate HBOC genes and are also involved in the biological processes associated with DNA integrity maintenance, transcriptional regulation, and cell cycle (Kobayashi et al., 2013; Cury et al., 2020). Moreover, we also found the 223 variants of 22 genes involved in a high to moderate risk of HBOC.

In conclusion, the whole exome sequencing analysis of HBOC patients identified the number of disease associated gene variants which are novel variants among the west Indian population. The prevalence of variants with the phenotypic association has shown prominent gene candidates and variants that may involve potently in the progression of many cancer including breast cancer and ovarian cancer. The deep analysis resulted in novel variants and novel gene which needs to be warranted and functional study is required to fully characterize their role in breast cancer and/or ovarian cancer.

Disclosure statements

Ethics approval and consent to participate

Informed consent was obtained from all patients.

Consent for publication

Written informed consent for publication was obtained.

Funding

Gujarat Biotechnology Research Centre (GBRC) and Gujarat Cancer Research Institute (GCRI) were funded by Gujarat State Biotechnology Mission (GSBTM), Department of Science and Technology, Government of Gujarat.

Submission declaration

The work described has not been published previously.

Data Availability Statement:

The datasets of the current study are available from the corresponding author on reasonable request.

Table 3
Prevalent disease-associated variants identified from Exomiser among patients.

Gene	Gene Name	Chromosome	Position	Ref	Alt	Functional Class	HGVS	Prevalence	ACMG Classification	Pathogenicity Score	DepthRange	Reference Allele Range	Altered Allele Range
CTNNB1	Catenin beta 1	3	41,267,044	GC	G	frameshift_variant	CTNNB1: ENST00000349496.5: c.718del:p. (Leu240Trpfs*2)	24	Pathogenic	No	20–173	17–101	2–68
WNT10A	Wnt family member 10A	2	219,747,074	AC	A	frameshift_truncation	WNT10A: ENST00000258411.3: c.307del:p. (Gln103Serfs*52)	21	Pathogenic	No	22–128	11–96	7–36
SMAD4	SMAD family member 4	18	48,573,546	G	GA	frameshift elongation	SMAD4: ENST00000342988.3: c.130_131insA:p. (Val44Aspfs*10)	19	Pathogenic	No	9–197	5–139	3–58
PALLD	Palladin, cytoskeletal associated protein	4	169,432,741	GC	G	frameshift_truncation	PALLD: ENST00000261509.6: c.88del:p. (Leu30Phefs*10)	17	Pathogenic	No	21–115	17–89	4–35
PRLR	Prolactin Receptor	5	35,065,808	AG	A	frameshift_truncation	PRLR: ENST00000382002.5: c.1251del:p. (Pro417Profs*35)	16	Pathogenic	No	15–188	11–121	4–89
HMMR	Hyaluronan mediated motility receptor	5	162,898,197	GA	G	frameshift_truncation	HMMR: ENST00000358715.3: c.470del:p. (Asn157Metfs*7)	15	Pathogenic	No	4–215	3–145	1–70
MITF	Melanocyte inducing transcription factor	3	70,001,015	GAA	G	frameshift_truncation	MITF: ENST00000314557.6: c.598_599del:p. (Lys200Glyfs*7)	14	Pathogenic	No			
CDKN2B	Cyclin dependent kinase inhibitor 2B	9	22,006,229	GC	G	frameshift_truncation	CDKN2B: ENST00000276925.6: c.173del:p. (Ser58Thrfs*107)	14	VUS	No	3–64	2–49	1–15
GNAS	GNAS complex locus	20	57,428,422	AG	A	frameshift_truncation	GNAS: ENST00000371100.4: c.106del:p. (Ala38Profs*652)	13	Pathogenic	No	8–164	2–58	6–108
POLD1	DNA polymerase delta 1, catalytic subunit	19	50,902,685	TG	T	frameshift_truncation	POLD1: ENST00000595904.1: c.262del:p. (Asp88Thrfs*81)	13	Pathogenic	No	12–149	8–87	4–71
KEAP1	Kelch like ECH associated protein 1	19	10,599,923	CT	C	frameshift_variant	KEAP1: ENST00000171111.5: c.1652del:p. (Lys551Serfs*44)	12	Pathogenic	No	41–266	29–157	9–190
ESR1	Estrogen receptor 1	6	152,332,887	A	AC	frameshift elongation	ESR1: ENST00000427531.2: c.677dup:p. (Gly227Argfs*8)	11	Pathogenic	No	5–45	3–36	2–27
TERF2IP	TERF2 interacting protein	16	75,690,420	CA	C	frameshift_variant	TERF2IP: ENST00000300086.4: c.1116del:p. (Lys372Lysfs*19)	11	Pathogenic	No	35–261	24–165	8–95

(*continued on next page*)

Table 3 (*continued*)

Gene	Gene Name	Chromosome	Position	Ref	Alt	Functional Class	HGVS	Prevalence	ACMG Classification	Pathogenicity Score	DepthRange	Reference Allele Range	Altered Allele Range
RAD51	RAD51 paralog B	15	40,991,013	C	CT	frameshift_variant	RAD51: ENST00000267868.3: c.60dup:p. (Gly21Trpfs*44)	11	Pathogenic	No	15–202	10–152	4–49
AKT1	AKT serine/ threonine kinase 1	14	105,239,636	AC	A	frameshift_truncation	AKT1: ENST00000544168.1: c.722del:p. (Gly241Valfs*4)	10	Likely Pathogenic	No	33–89	26–70	7–20
POT1	Protection of telomeres 1	7	124,465,308	A	AT	frameshift_variant	POT1: ENST00000357628.3: c.1789dup:p. (Ile597Asnfs*2)	10	Pathogenic	No	9–81	6–60	3–21
SEC23B	SEC23 homolog B, COPII coat complex component	20	18,491,560	CCG	C	frameshift_truncation	SEC23B: ENST00000262544.2: c.82_83del:p. (Arg28Alafs*29)	10	Pathogenic	No	22–66	17–53	5–17
POLD1	DNA polymerase delta 1, catalytic subunit	19	50,902,174	C	CG	frameshift_variant	POLD1: ENST00000440232.2: c.66_67insG:p. (Trp23Valfs*3)	9	Likely Pathogenic	No	23–164	17–126	6–38
TERT	Telomerase reverse transcriptase	5	1,253,914	TC	T	frameshift_truncation	TERT: ENST00000310581.5: c.3327del:p. (Gly1109Glyfs*4)	9	Likely Pathogenic	No			
FGFR2	Fibroblast growth factor receptor 2	10	123,245,001	GC	G	frameshift_truncation	FGFR2: ENST00000346997.2: c.2096del:p. (Gly699Alafs*16)	8	Pathogenic	No	21–216	15–153	6–75

Ref.: Reference base present in genome sequence, Alt: Altered base present in sequence, HGVS: Human Genome Variation Society nomenclature, Prevalence: Prevalence of variant among number of patients, ACMG Classification: Classification of variants as per ACMG Guidelines, Pathogenicity Score: The score indicated in Varsome. Depth Range: it indicates the range of depth of sequences among patients, Reference allele range and Altered allele range: It indicated based on the visualization in IGV using respective (.bed) file.

Table 4
High to moderate risk of HBOC associated gene variants identified from Exomiser.

Gene	Gene Name	Chromosome	Position	Ref	Alt	Functional Class	HGVS	ACMG Classification	Polyphen Score	Mutationtaster Score	SIFT (<0.06)	DBSNP ID	EXAC SAS FREQ	Major Allele Frequency	Minor Allele Frequency
TP53	Tumor protein p53	17	7,577,151	T	C	Missense variant	TP53:ENST00000269305.4: c.787A > G:p.(Asn263Asp)	Benign	0.055	.	0.99	rs72661119	0.071974	11 (37%)	19 (63%)
STK11	Serine/ threonine kinase 11	19	1,207,103	A	G	Missense variant	STK11:ENST00000326873.7: c.191A > G:p.(Lys64Arg)	Likely Pathogenic	0.681	1	0.93	.	.	41 (72%)	16 (28%)
CASP8	Caspase 8	2	202,149,799	T	C	Missense variant	CASP8:ENST00000264274.9: c.811 T > C:p.(Phe271Leu)	VUS	0.998	1	1	.	.	91 (82%)	20 (18%)

Ref.: Reference base present in genome sequence, Alt: Altered base present in sequence, HGVS: Human Genome Variation Society nomenclature, ACMG Classification: Classification of variants as per ACMG Guidelines, Polyphen score: (>0.956|>0.446), Mutationtaster Score (>0.94), SIFT Score (<0.06), DBSNP ID: Unique identifier assigned to the SNP in database, ExAC_SAS_Frequency: South Asian in Exome Aggregation Consortium (Available in gnomAD).
* Major allele frequency and minor allele frequency (Variant Allele Frequency) were indicated based on the visualization in IGV.

CRediT authorship contribution statement

Bhargav N. Waghela: Data curation, Formal analysis, Writing – original draft preparation. **Ramesh J. Pandit:** Data curation, Writing – review & editing. **Apurvasinh Puvar:** Formal analysis, Writing – review & editing. **Franky D. Shah:** Resources, Writing – review & editing. **Prabhudas S. Patel:** Resources, Writing – review & editing. **Hemangini Vora:** Resources, Writing – review & editing. **Harsh Sheth:** Formal analysis, Writing – review & editing. **Bhoomi Tarapara:** Resources, Writing – review & editing. **Shashank Pandya:** Resources, Writing – review & editing. **Chaitanya G. Joshi:** Conceptualization, Methodology, Supervision. **Madhvi N. Joshi:** Conceptualization, Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We are thankful to the patients and their families for registering themselves and providing consent for participation in this study. We also thank Gujarat State Biotechnology Mission (GSBTM), Department of Science and Technology, Government of Gujarat, Gandhinagar for their financial assistance. We would also like to thank the Gujarat Cancer and Research Institute for providing ethical committee approval for conducting research. We would like to thank Ms. Komal Patel for her constant support in various laboratory work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.147070.

References

Anders, C.K., Johnson, R., Litton, J., Phillips, M., Bleyer, A., 2009. Breast cancer before age 40 years. Semin. Oncol. 36, 237–249.
Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data. Version 0.11. 2. Website: http://www. bioin forma tics. babra ham. ac. uk/proje cts/fastqc.
Bakkach, J., Mansouri, M., Derkaoui, T., Loudiyi, A., Fihri, M., Hassani, S., Barakat, A., Ghailani Nourouti, N., Bennani Mechita, M., 2017. Clinicopathologic and prognostic features of breast cancer in young women: a series from North of Morocco. BMC. Womens. Health 17, 106.
Berberich, A.J., Ho, R. and Hegele, R.A., 2018. Whole genome sequencing in the clinic: empowerment or too much information? CMAJ 190, E124-E125.
Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer. J. Clin 68, 394–424.
Castera, L., Krieger, S., Rousselin, A., Legros, A., Baumann, J.J., Bruet, O., Brault, B., Fouillet, R., Goardon, N., Letac, O., Baert-Desurmont, S., Tinat, J., Bera, O., Dugast, C., Berthet, P., Polycarpe, F., Layet, V., Hardouin, A., Frebourg, T., Vaur, D., 2014. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. Eur. J. Hum. Genet 22, 1305–1313.
Cury, N.M., Brotto, D.B., de Araujo, L.F., Rosa, R.C.A., Texeira, L.A., Plaça, J.R., Marques, A.A., Peronni, K.C., de Cássia Ruy, P., Molfetta, G.A., 2020. Germline variants in DNA repair genes associated with hereditary breast and ovarian cancer syndrome: Analysis of a 21 gene panel in the Brazilian population. BMC Med. Geno. 13, 1–24.
Daly, M.B., Pilarski, R., Yurgelun, M.B., Berry, M.P., Buys, S.S., Dickson, P., Domchek, S. M., Elkhanany, A., Friedman, S., Garber, J.E., Goggins, M., Hutton, M.L., Khan, S., Klein, C., Kohlmann, W., Kurian, A.W., Laronga, C., Litton, J.K., Mak, J.S., Menendez, C.S., Merajver, S.D., Norquist, B.S., Offit, K., Pal, T., Pederson, H.J., Reiser, G., Shannon, K.M., Visvanathan, K., Weitzel, J.N., Wick, M.J., Wisinski, K.B., Dwyer, M.A., Darlow, S.D., 2020. NCCN Guidelines Insights: Genetic/Familial High-

Risk Assessment: Breast, Ovarian, and Pancreatic, Version 1.2020. J. Natl. Compr. Canc. Netw 18, 380–391.
Edlich, R.F., Winters, K.L., Lin, K.Y., 2005. Breast cancer and ovarian cancer genetics. J. Long. Term. Eff. Med. Implants 15, 533–545.
Felicio, P.S., Grasel, R.S., Campacci, N., de Paula, A.E., Galvão, H.C., Torrezan, G.T., Sabato, C.S., Fernandes, G.C., Souza, C.P., Michelli, R.D., 2021. Whole-exome sequencing of non-BRCA1/BRCA2 mutation carrier cases at high-risk for hereditary breast/ovarian cancer. Human. mutation 42, 290–299.
Foulkes, W.D., 2008. Inherited susceptibility to common cancers. N. Engl. J. Med 359, 2143–2153.
Fusco, N., Malapelle, U., Fassan, M., Marchiò, C., Buglioni, S., Zupo, S., Criscitiello, C., Vigneri, P., Dei Tos, A.P., Maiorano, E., 2021. PIK3CA mutations as a molecular target for hormone receptor-positive, HER2-negative metastatic breast cancer. Front. Oncol. 11, 562.
Grabenstetter, A., Lazaro, C., Turashvili, G., 2020. Editorial: Hereditary Breast and Ovarian Cancer: Current Concepts of Prevention and Treatment. Front. Oncol 10, 618369.
Guo, D.-Q., Zhang, H., Tan, S.-J. and Gu, Y.-C., 2014. Nifedipine promotes the proliferation and migration of breast cancer cells. PloS one 9, e113649.
Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W. and Lane, H.C., 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic acids research 35, W169-W175.
Jarhelle, E., Stensland, H.M.F.R., Hansen, G.Å.M., Skarsfjord, S., Jonsrud, C., Ingebrigtsen, M., Strømsvik, N., Van Ghelue, M., 2019. Identifying sequence variants contributing to hereditary breast and ovarian cancer in BRCA1 and BRCA2 negative breast and ovarian cancer patients. Sci. Rep. 9, 1–12.
Kadri, M.S.N., Patel, K.M., Bhargava, P.A., Shah, F.D., Badgujar, N.V., Tarapara, B.V., Patel, P.S., Shaikh, M.I., Shah, K., Patel, A., Pandya, S., Vora, H., Joshi, C.G., Joshi, M.N., 2020. Mutational Landscape for Indian Hereditary Breast and Ovarian Cancer Cohort Suggests Need for Identifying Population Specific Genes and Biomarkers for Screening. Front. Oncol 10, 568786.
Kobayashi, H., Ohno, S., Sasaki, Y., Matsuura, M., 2013. Hereditary breast and ovarian cancer susceptibility genes. Oncology. reports 30, 1019–1029.
Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C.E., Aguilera, M.A., Meyer, R., Massouras, A., 2019. VarSome: the human genomic variant search engine. Bioinformatics 35, 1978.
Lambros, M.B., Natrajan, R., Geyer, F.C., Lopez-Garcia, M.A., Dedes, K.J., Savage, K., Lacroix-Triki, M., Jones, R.L., Lord, C.J., Linardopoulos, S., 2010. PPM1D gene amplification and overexpression in breast cancer: a qRT-PCR and chromogenic in situ hybridization study. Modern. Pathology 23, 1334–1345.
Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M. and Maglott, D.R., 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research 42, D980-D985.
Lynch, H.T., Snyder, C. and Casey, M.J., 2013. Hereditary ovarian and breast cancer: what have we learned? Ann Oncol 24 Suppl 8, viii83-viii95.
Moo, T.A., Sanford, R., Dang, C., Morrow, M., 2018. Overview of Breast Cancer Therapy. PET. Clin 13, 339–354.
Mur, P., García-Mulero, S., Del Valle, J., Magraner-Pardo, L., Vidal, A., Pineda, M., Cinnirella, G., Martín-Ramos, E., Pons, T., López-Doriga, A., 2020. Role of POLE and POLD1 in familial cancer. Genet. Med. 22, 2089–2100.
Rajalingam, K., Wunder, C., Brinkmann, V., Churin, Y., Hekman, M., Sievers, C., Rapp, U. R., Rudel, T., 2005. Prohibitin is required for Ras-induced Raf–MEK–ERK activation and epithelial cell migration. Nat. Cell. Biol. 7, 837–843.
Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–423.
Rogers, M.F., Shihab, H.A., Gaunt, T.R., Campbell, C., 2017. CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. Sci. Rep. 7, 1–10.
Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.
Seguí, N., Mina, L.B., Lázaro, C., Sanz-Pamplona, R., Pons, T., Navarro, M., Bellido, F., López-Doriga, A., Valdés-Mas, R., Pineda, M., 2015. Germline mutations in FAN1 cause hereditary colorectal cancer by impairing DNA repair. Gastroenterology 149, 563–566.
Sharma-Oates, A., Shaaban, A.M., Tomlinson, I., Wynne, L., Cazier, J.-B., Sundar, S., 2018. Heterogeneity of germline variants in high risk breast and ovarian cancer susceptibility genes in India. Precision. Clinical. Medicine 1, 75–87.
Smedley, D., Jacobsen, J.O., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O.J., Washington, N.L., 2015. Next-generation diagnostics and disease-gene discovery with the Exomiser. Nat. Protocols 10, 2004–2015.
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:. Cancer J. Clin. 71, 209–249.
Szabo, C.I., King, M.C., 1995. Inherited breast and ovarian cancer. Hum Mol Genet 4 Spec No, 1811-7.
Turan, S., Bastepe, M., 2013. The GNAS complex locus and human diseases associated with loss-of-function mutations or epimutations within this imprinted gene. Hormone Res. Paediatr. 80, 229–241.
Yoshida, R., 2020. Hereditary breast and ovarian cancer (HBOC): Review of its molecular characteristics, screening, treatment, and prognosis. Breast Cancer 1–14.