Identification of novel exonic variants responsible for hereditary breast and ovarian cancer in West Indian population

Bhargav N. Waghela
GBRC: Gujarat Biotechnology Research Centre

Ramesh J. Pandit
GBRC: Gujarat Biotechnology Research Centre

Apurvasinh Puvar
GBRC: Gujarat Biotechnology Research Centre

Franky D. Shah
GCRI: The Gujarat Cancer and Research Institute

Prabhudas S. Patel
GCRI: The Gujarat Cancer and Research Institute

Hemangini Vora
GCRI: The Gujarat Cancer and Research Institute

Harsh Sheth
FRIGE House, Ahmedabad, Gujarat

Bhoomi Tarapara
GCRI: The Gujarat Cancer and Research Institute

Shashank Pandya
GCRI: The Gujarat Cancer and Research Institute

Chaitanya G. Joshi
GBRC: Gujarat Biotechnology Research Centre

Madhvi Joshi (madhvimicrobio@gmail.com)
Gujarat Biotechnology Research Centre
https://orcid.org/0000-0003-4102-7114

Research article

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

Posted Date: October 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-952462/v1
Abstract

Background

Breast and ovarian cancers are the most common cancer types in females in India which pertain to higher mortality and morbidity due to late diagnosis and poor prognosis. Early diagnosis for better prognosis improve the patient's treatment and survival. The next-generation sequencing (NGS)-based screening has accelerated molecular diagnosis of various cancers.

Methods

We performed whole exome sequencing (WES) of 30 patients who had a first or second degree relative with breast or ovarian cancer. Further, all these patients are tested negative for BRCA1/2 or other high and moderate risk genes reported for HBOC. WES data from 30 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 by Varsome tool. The functional analysis of genes was determined by STRING analysis and disease association was determined by open target tool.

Results

We examined the variants based on the prevalence of variants among 30 patients i.e. frequency and disease association determined by the phenotype score of exomiser. From both the approaches, we found novel variants and novel gene candidates associated with HBOC conditions. The variants in HYDIN, AVIL, IWS1, PLA2G6, PRDM4, ST3GAL2, and ZNF717 were predicted highly oncogenic. Moreover, we also found 59 genes having higher phenotype score (phenotype score >0.75) and which are associated with various biological processes such as DNA integrity maintenance, transcriptional regulation, cell cycle and apoptosis.

Conclusion

The gene variants associated with HBOC condition in West Indian cohort have been revisited. Our findings provide novel as well as highly prevalent variants in the population which could be further studied in detail for their use in early diagnosis and better prognosis of HBOC patients.

Highlights:

- Hereditary Breast cancer (BC) and ovarian cancer (OC) pertain to higher mortality and needs early diagnosis.
• The whole exome sequencing (WES), one of the powerful techniques of NGS analysis has delineated highly prevalent gene variants in the west Indian population.

• The genotype-phenotype association using Exomiser also revealed some novel variants and prominent gene candidates those may involve potently in progression of BC and OC.

• The WES examination provided novel as well as highly prevalent variants in population which could be further studied.

Background:

Breast cancer (BC) and ovarian cancer (OC) are the most prominent gynecological cancers among the females worldwide as well as in India [1, 2]. In 2020, 178,361 of breast cancer and 45,701 of ovarian cancer new cases were diagnosed in India which account for 26.3% and 6.7% of the total cancer cases, respectively [3] 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 [4, 5]. However, the inherited mutations are also responsible for early onset of these cancers [6]. Hereditary breast and ovarian cancer (HBOC) syndrome is a major condition responsible for approximately 90% of inherited breast and ovarian cancer [7]. HBOC is an autosomal-dominant inherited condition associated with higher risk of early-onset of breast cancer and ovarian cancer in multiple family members [8]. The similar chances of HBOC prevalence has been reported in both females and males. In male, the HBOC attributes to 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 mainly associated with germline mutations in BRCA1 or BRCA2 genes [9]. 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 clinvar database [10]. However, the mutational profiles of BRCA 1/2 is highly variable across various populations of world and India [11].

The mutational landscape of multiple genes apart from BRCA1 and BRCA2 are involved in predisposition of HBOC [10]. The National Comprehensive Cancer Network (NCCN) has provided guidelines for clinical management of HBOC. To date, NCCN has reported 21 genes in which, prevalence of pathogenic mutation leading to HBOC condition [12]. According to the NCCN guidelines, individuals having strong family history of breast and/or ovarian cancers should undergo the genetic assessment test for early diagnosis and better clinical management [12].

Several studies have reported that the deleterious mutations in non-BRCA genes leading to HBOC conditions [13]. In our previous study, we have designed the next generation sequencing (NGS) based multigene panel to detect the mutations in 18 14 high to moderate risk genes including BRCA and non-BRCA genes for west Indian population [14]. In that, we 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 population suggesting clinical guidelines and gene-disease relations reported globally may partly support clinical management of HBOC in Indian population.
Interestingly, we noticed few patients which were negative for the multi-gene panel but having breast and/or ovarian cancer and selected 30 patients for the study.

Whole-exome sequencing (WES) is an efficient technique for clinical diagnosis and used to identify the mutational profiling of genetic diseases [15]. The WES promisingly utilized for the screening of mutational profile of genes involved in BC and/or OC [16]. In present study, we performed WES of 30 HBOC patients having 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 variants by the score of exomiser tool and further analyzed based on frequency in patients and disease-associated genetic variants by exomiser score.

## Methods:

### 2.1 Patient selection:

The patients were selected based on their genetic counselling 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 relative were selected for study. the patients were selected on the basis of 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 the medical records. All patients signed informed consent from approved by the institutional review board at GCRI. Further, 30 HBOC patients whom were negative to 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 using QIAamp DNA Blood Mini kit (QIAGEN, Germany). The DNA concentration was determined by Qubit Fluorometer 4.0® (Thermo Fisher Scientific, USA) and purity of DNA was determined by QIAxpert (QIAGEN, Germany).

### 2.3 Whole exome library preparation and sequencing

In this study, we have selected 30 samples for amplicon based exome sequencing. For library preparation, genomic DNA (approx. 100ng) of each subject was amplified using Ampliseq RDY panel kit (Thermo Fisher Scientific, USA). Ampliseq exome kit includes 2,93,903 primer pairs that cover 97% of CCDS with 5bp 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 100pmol 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 200bp 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 FASTQC toolkit (v.0.11.5) (Andrews and others, 2010). Raw sequences were further trimmed and filtered using PRINSEQ-lite v.0.20.4 (Schmieder and Edwards, 2011) in which 5bp from left end and 10bp from right end were trimmed, sequence length lower than 50bp and quality mean less than 20 were removed. clean reads were mapped on 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 Figure 1).

2.4.2 Variant annotation:

The functional annotation of variants and variant prioritization was performed by Exomiser (Version 12.1.0 available at https://github.com/exomiser/Exomiser). The VCF files obtained were used as an input in YML file in exomiser. The HPOIDs for breast and ovarian cancer were entered and inheritance mode was autosomal dominant (0.1) in exomiser and was run with default settings. We have included the Ashkenazi Jewish population in frequency sources and entered the pathogenicity prediction tools such as 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:

The variants were analysed in varsome and classified according to the American College of Medical Genetics and Genomics (ACMG) recommendations. The variants were classified into five categories such as pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign and benign.

Results:

Patient selection:

In present study, we have selected 30 unrelated patients with breast cancer and/or ovarian cancer and were found negative to panel genes designed previously for west Indian population. Distribution of the patients on the basis of disease and age of onset has been represented in supplementary figure 2. The whole exome analysis of 30 patients was performed and the data were analysed 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 total of 4,56,741 unique variants of 18,594 genes among the 30 patients with the average 40,000 variants per patient. We examined the variants
based on two major criteria, (i) prevalence of variants among the patients and (ii) the disease association using phenotype score of exomiser (Figure 1).

**Variants selection based on the prevalence of variants among West Indian patients:**

To analyse the variants based on their prevalence in west Indian patients, we analysed all the entries obtained from Exomiser. Firstly, we obtained total more than 1.2 million variants from 30 patients. Further, we examined the variants on the basis of their prevalence in ≥90% of patients (27/30 or more patients). We found the 2,481 variants of 2,150 genes were present in ≥90% patients. Therefore, we indicated these 2,481 variants as highly prevalent variants among the West Indian population for HBOC.

**Analysis of prevalent variants and their pathogenicity:**

We analyzed the highly prevalent variants on the basis of their mutational pathogenicity predicted in SIFT tool and found 17 variants of 8 genes. The genes are HYDIN, AVIL, IWS1, PLA2G6, PRDM4, ST3GAL2, TROAP, and ZNF717. The genes HYDIN having 5 variants, IWS1 having 2 variants and ZNF717 having 5 variants. Rest of the genes have one variant each. We analyzed the variants and found 8 variants, HYDIN:c.12478G>C, HYDIN:c.11224G>A, HYDIN:c.2119A>G, TROAP:c.1676G>A, ZNF717:c.1994A>G, ZNF717:c.1980A>T, ZNF717:c.1980A>T and ZNF717:c.1021T>G and ZNF717:c.526T>C were likely benign and 9 variants, HYDIN:c.6892C>G, HYDIN:c.6259C>T, AVIL:c.1364C>T, IWS1:c.713A>G, IWS1:c.712G>A, PLA2G6:c.2283C>A, PRDM4:c.82C>G, ST3GAL2:c.611G>C and ZNF717:c.2083C>T were VUS. None of them were reported in clinvar database. The variants of HYDIN, AVIL, IWS1, PLA2G6, PRDM4, ST3GAL2, and ZNF717 were predicted highly oncogenic and the variant from TROAP and 2 variants of ZNF717 were predicted benign. None of the gene has shown the significant association with cancer, breast cancer or ovarian cancer (Table 1).

**Analysis of prevalent variants having South Asian frequency:**

We looked into the allele frequency of these variants in the South Asian population and found 9 variants of 9 genes. The genes are TCF20, SOST, MALT1, LRIT2, MAN2C1, SLC4A3, ZFR2, ZNF717 and FAM104B. We analysed the variants in varsome tool and found that the 2 variants, SOST: c.122del and MALT1: c.2406del were pathogenic, 1 variant, SLC4A3: c.470del was likely pathogenic, 3 variants, TCF20: c.5853C>T, MAN2C1:c.2246+5A>G and ZNF717:c.959T>C were likely benign and 3 variants, LRIT2:c.726del, ZFR2:c.943del and FAM104B:c.331C>T were VUS. None of the variants are reported to be associated with BC and OC in clinvar database. The MAN2C1:c.2246+5A>G was predicted highly oncogenic by cScape tool. None of the gene is significantly associated with cancer, breast cancer or ovarian cancer (Table 2).

**Analysis of prevalent variants for high to moderate HBOC Genes:**
We analyzed the highly prevalent variants whether that particular gene or the variants have been previously associated with high or moderate risk of HBOC condition. First, we gone through the set of genes reported in NCCN guidelines and thereafter the set of genes which were included in our previously customized gene panel. From NCCN guideline, we included 22 genes which are associated with high or moderate risk of HBOC condition and from the gene panel we included 18 genes for the present analysis. We found 3 variants of 3 genes from NCCN genes and 3 variants of 3 genes from customized panel genes set. Further, we found that the \textit{BRCA1}:c.3214del and \textit{BRIP1}:c.356del were common in both sets, whereas \textit{NF1}:c.3093_3094del was from NCCN genes set and \textit{ERBB2}:c.2694del was from customized panel genes set. The \textit{BRCA1}:c.3214del was found pathogenic in varsome and shows significant association with HBOC in clinvar. The \textit{BRIP1}:c.356del variant was found likely pathogenic but not reported in clinvar. The \textit{NF1}:c.3093_3094del was found pathogenic but not reported in clinvar. The \textit{ERBB2}:c.2694del variant was found likely pathogenic but was not reported in clinvar. All the four genes, \textit{BRCA1}, \textit{BRIP1}, \textit{NF1} and \textit{ERBB2} have shown strong functional relation with breast and ovarian cancer (Table 3).

Variants analysis using Exomiser score:

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

Analysis of disease associated variants identified from Exomiser for their Pathogenicity:

To analyse the pathogenic effects of respective variation in exome of HBOC patients, we analysed the 464 variants of 59 genes by their respective SIFT and Polyphen scores and found 15 variants of 12 genes. The genes include \textit{COL14A1}, \textit{FAN1}, \textit{GNAS}, \textit{OPCML}, \textit{PHB}, \textit{PIK3CA}, \textit{POLE}, \textit{PPM1D}, \textit{RAD54L}, \textit{RNF43}, \textit{TERT} and \textit{TWIST1}. We found these gene variants in varsome and classified the variants based on the ACMG guidelines. Out of the 15 variants, 14 variants, \textit{COL14A1}:c.529G>T, \textit{GNAS}:c.478A>G, \textit{OPCML}:c.458T>C, \textit{PHB}:c.505T>C, \textit{PIK3CA}:c.31T>G, \textit{PIK3CA}:c.32G>T, \textit{POLE}:c.6302C>A, \textit{POLE}:c.6344A>G, \textit{PPM1D}:c.1579G>A, \textit{RAD54L}:c.345A>C, \textit{RAD54L}:c.579C>G, \textit{RNF43}:c.379C>T, \textit{TERT}:c.2711T>A and \textit{TWIST1}:c.332T>C were found VUS and 1 variant \textit{FAN1}:c.1589T>C was likely benign. Further, we analysed the oncogenic properties of variants and found variants of \textit{COL14A1}, \textit{OPCML}, \textit{PHB}, \textit{PIK3CA}, \textit{POLE}, \textit{PPM1D}, \textit{RAD54L}, \textit{RNF43}, \textit{TERT} and \textit{TWIST1} were oncogenic, \textit{FAN1} and \textit{GNAS} were benign. We also confirmed our results by visualizing the chromosomal position of each variation in IGV (Table 4). Further, the protein-protein interaction analysis using STRING suggested that out of 59 genes, 53 showed functional association. These 59 genes were also analysed for their association with high and moderate risk genes reported for HBOC, and found that 54 genes except \textit{COL14A1}, \textit{AAGAB}, \textit{OPCML}, \textit{SEC23B} and \textit{DMPK} genes, were functionally associated (Supplementary Figure 3). The analysis of disease association with BC and OC revealed that majority of them are strongly associated with BC and OC.
Moreover, the functional annotation analysis suggested that the majority of genes also involves in the biological processes associated with DNA integrity maintenance, transcriptional regulation, cell cycle and apoptosis.

Analysis of disease associated variants identified from Exomiser for the South Asian prevalence:

We analysed the 464 variants of 59 genes for their frequency in 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 in varsome. Further, we analysed the oncogenic properties of variants and found that variants of KRAS, PPM1D, RAD54L and RNF43 are oncogenic (Table 5).

Analysis of prevalent disease associated variants identified from Exomiser among patients:

We analyzed the 687 variants of 81 genes for their frequency. We scrutinized the variants on the basis of their frequency in patients. For the present study, we scrutinized variants with ≥25%. We found 33 variants of 30 genes having higher frequency among patients. Out of these 33 variants, 25 variants, CTNNB1:c.718del, WNT10A:c.307del, NBN:c.2027del, 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, STK11:c.326dup, ESR1:c.677dup, TERF2IP:c.1116del, RAD51:c.60dup, PALB2:c.3199del, POT1:c.1789dup, SEC23B:c.82_83del, ATM:c.1102del, CDKN2A:c.238del, MLH1:c.718del, RAD50:c.2863del, FGFR2:c.2096del, CDH1:c.1955del, PALB2:c.3351del were pathogenic, 5 variants, ACD:c.1217del, AKT1:c.722del, BRCA2:c.1597dup, POLD1:c.66_67insG and TERT:c.3327del were likely pathogenic and 03 variants, CDKN2B:c.173del, CDKN2B:c.310del and IDH1:c.994del were VUS in varsome. These 3 genes having 2 variants each are CDKN2B- (2) VUS, PALB2- (2) Pathogenic, POLD1- (1) Pathogenic, (1) Likely Pathogenic (Table 6). 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 also involves in the biological processes associated with DNA integrity maintenance, transcriptional regulation and cell cycle (Supplementary Figure 5).

Analysis of variants identified from Exomiser pertaining to high/moderate-risk genes of HBOC:

We analysed 687 variants of 81 genes for their frequency. We scrutinized the variants on the basis of their frequency in patients. For the present study, we scrutinized genes and their variants having 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,
**Discussion:**

Breast and ovarian cancer are the leading cause of cancer related death in females [2]. Conventional therapies are used to treat the patients however, it majorly concerned with side effects, low rate of patient survival, and poor health quality [17]. These problems instigate to identify alternative approaches such as early diagnosis for efficient prognosis that may result in improved patient survival. Identification of mutational landscape of cancer patients using the next generation sequencing (NGS) has accelerated the field of cancer genomics and essential for early diagnosis of Breast and ovarian cancer [18].

The mutations in *BRCA1* and *BRCA2* are associated with the HBOC patients. According to NCCN guidelines, apart from *BRCA1* and *BRCA2*, mutations in approximately 20 other genes are also reported in pathologies associated with HBOC in various population of the world [12]. Several studies have been conducted to identify prevalent gene variants playing deleterious role in HBOC patients [10]. To identify the mutations of those gene variants in West Indian population, we have designed a customised gene-panel comprising 18 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 [14]. This observation led us to investigate novel mutation/s in the whole exonic region which may have role in these cancers. Recently, WES is utilized for the screening of mutational profile of genes involved in BC and/or OC [16]. 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 customised gene panel of 18 genes (*BRCA1, BRCA2, TP53, PTEN, CDH1, STK11, BARD1, ATM, AR, TGFB1, BRIP1, CASP8, CHEK2, ERBB2, NBN, PALB2, RAD50, RAD51C*). We identified novel gene variants involved in HBOC progression in West Indian patient cohort based on their prevalence and phenotype association score. We found various gene variants those may closely related to *BRCA1* and *BRCA2* in breast and/or ovarian cancer. We found novel variants having higher prevalence in the west India patient cohort which requires further characterization. Moreover, our analysis based on the exomiser phenotype score and associated functional consequences identified the gene candidates involve 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 for their higher prevalence i.e. ≥90% (27/30 patients) and found 2,481 variants of 2,150 genes. These were further screened for their pathogenicity by their SIFT and Polyphen Score. We found 9 VUS variants, *HYDIN:c.6892C>G, HYDIN:c.6259C>T, AVIL:c.1364C>T, IWS1:c.713A>G, IWS1:c.712G>A, PLA2G6:c.2283C>A, PRDM4:c.82C>G, ST3GAL2:c.611G>C and ZNF717:c.2083C>T* and having oncogenic potential. However, these variants are not reported in Clinvar, GnomAD and LOVD indicating that the variants are novel. The previous study using whole exome sequencing suggested that the somatic mutation in *HYDIN* has been associated with
breast cancer in Chinese patients' cohort [19]. The HYDIN has been found mutated with profound cancer genes such as TP53, KRAS, PTEN in large intestine cancer [20]. The overexpression of AVIL has been reported to be involved in cell proliferation and migration and drives tumorigenesis in glioblastoma [21, 22]. The IWS1 has been overexpressed in lung, skin and breast cancers [23, 24]. The PLA2G6 dysfunction was reported to involve in hepatocellular carcinoma [25]. The PRDM4 has been reported to be associated with progression and recurrence of gastric cancer and potent prognostic biomarker for the prediction of gastric cancer [26]. The mRNA level of PRDM4 has been significantly increased metastatic prostate cancer [27]. The higher expression of ST3GAL2 (the rate-limiting enzyme of SSEA4 synthesis) has been associated with poor clinical outcome in breast and/or ovarian cancer patients treated with chemotherapy [28]. Cumulatively, this suggests that, the novel mutations which are identified in the present study may have important role in HBOC.

Next, we found 9 variants of 9 genes having allele frequency in South Asian population include 2 pathogenic variants, SOST: c.122del and MALT1: c.2406del, 1 likely pathogenic variant, SLC4A3: c.470del, and 3 VUS variants, LRIT2:c.726del, ZFR2:c.943del and FAM104B:c.331C>T which is not reported to be in Clinvar database. The SOST acts as a Wnt antagonist and potent inhibitor of prostate cancer invasion [29]. Moreover, SOST inhibits osteoblast differentiation during breast cancer [30]. MALT1 is a component of CBM signalosome (triad of CARMA3, Bcl10, and MALT1) which activates NF-κB signalling and promotes aggressive behaviour of breast cancer [31]. Apart from novel variants, we also found 4 variants of 4 genes having high to moderate risk for HBOC such as BRCA1:c.3214del, BRIP1:c.356del, NF1: c.3093_3094del and ERBB2: c.2694del. The BRCA1:c.3214del is well reported for HBOC in clinvar. The BRIP1:c.356del and ERBB2: c.2694del variant were found likely pathogenic but not reported in clinvar. Previously, we included ERBB2 gene in our customized gene panel [32]. Moreover, Neurofibromatosis 1 causing NF1 gene has been associated with breast cancer progression [33] and included in NCCN genes however, the variant NF1: c.3093_3094del has not been reported in clinvar. Further, to identify the novel variants having higher disease potential despite lower prevalence, we analysed the mutations based on 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 having oncogenic potential. GNAS:c.478A>G was also found VUS however predicted benign. By analysing 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 analysed the prevalent variants among patients having frequency ≥25% (8/30) with higher phenotype score and found 33 variants of 30 genes. Of them, 25 variants were pathogenic, 5 variants were likely pathogenic and 03 variants were VUS in varsome. The functional analysis revealed that the 29 genes have prominent association with high to moderate HBOC genes and also involved in the biological processes associated with DNA integrity maintenance, transcriptional regulation and cell cycle [34, 35]. Moreover, we also found the 223 variants of 22 genes having high to moderate risk of HBOC.

Conclusion:
In conclusion, our study depicts that the exome analysis of HBOC patients resulted in large number of gene variants which are prevalent in the West Indian population. Further, the phenotypic association also revealed some novel variants among the population. The prevalence of variants with phenotypic association has shown prominent gene candidates and variants those may involve potently in progression of many cancer including BC and OC. 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 BC and/or OC.

**List Of Abbreviations:**


**Declarations:**

**Ethics approval and consent to participate**

Informed consent was obtained from all patients.

**Consent for publication**

Written informed consent for publication was obtained.

**Conflict of Interest:**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**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.

**Author Contributions:**

CGJ and MNJ: Conceptualization, Methodology, Supervision; BW: Data curation, analysis and interpretation of data, Writing- Original draft preparation; RP: Data curation, interpretation of data and
Data Availability Statement:

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

Acknowledgments:

We are thankful to the patients and their families for registering themselves and providing consent for participation in this study. We also thank to Gujarat State Biotechnology Mission (GSBTM), Department of Science and Technology, Government of Gujarat, Gandhinagar for the 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.

References:


**Tables:**

Due to technical limitations, Tables 1,2,3,4,5,6 and 7 are only available as a download in the Supplemental Files section.

**Figures**
Figure 1

Approaches used to analyze the gene variants.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1.docx
- Table2.docx
- Table3.docx
- Table4.docx
- Table5.docx
- Table6.docx
- Table7.docx
- Supp.Figure1.tif
- Supp.Figure2.tif
- Supp.Figure3.tif
- Supp.Figure4.tif
- Supp.Figure5.tif