Saudi Journal of Biological Sciences 27 (2020) 2651–2659

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Targeted sequencing of crucial cancer causing genes of breast cancer in Saudi patients

Mohammad Alanazi [a], Narasimha Reddy Parine [a,*], Jilani Purusottapatnam Shaik [a], Abdulrahman Al Naeem [b], Sooad Aldhaian [a]

[a] Genome Research Chair, Department of Biochemistry, College of Science King Saud, University, 11451 Riyadh, Saudi Arabia
[b] Basic Sciences Department, College of Science and Health Professions, King Saud bin, Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

ARTICLE INFO

Article history:
Received 10 November 2019
Revised 26 May 2020
Accepted 29 May 2020
Available online 6 June 2020

Keywords:
Breast cancer
Targeted sequencing
Germline mutations
BRCA1
BRCA2
TP53

ABSTRACT

Breast cancer is the most common cancer among women worldwide, causing 15% of cancer-related deaths among women. Breast cancer incidence rate is increasing in most countries. In Saudi Arabia, breast cancer constitutes nearly 22% of the newly diagnosed cancer cases in women. Breast cancer incidence in the women population of Saudi Arabia is 25.9%, with 18.2% mortality. In this study, targeted sequencing of 164 selected genes was performed on germline and somatic DNA derived from the blood and tissue samples of 50 breast cancer patients using customized panel on Ion torrent platform. This study focused on the identification of genetic variations of different cancer-causing genes, raising the hope for identification of personalized prognosis. After final filtration and validation, we found protein-truncating, non-synonymous missense, and splice site mutations in the known susceptibility genes for breast cancer. We identified a total of 14 point mutations and one deletion in *BRCA1*, *BRCA2*, and *RAD50* genes from the BRCA panel analysis of breast cancer samples. In the customized panel analysis, we identified 37 potential mutations in 25 breast cancer risk associated genes. Out of these, most mutations were observed in TP53. After filtration, we observed 7 mutations in TP53 genes (n = 7:- one stop gain (p.R81X), four non-synonymous (p.R81X, p.Y88C, p.R141H, and p.V25D), and two deletions (c.59delC and c.327delC)). Among the mutations detected in our study, TP53 (p.R81X), VHL (p.E52X), and *BRCA2* (p.K3326X) mutations, which lead to an aberrant transcript with a premature stop codon, were reported for the first time in breast cancer patients from Saudi Arabia. Our study will help in identifying the damaging mutations and predisposing genes in Saudi breast cancer patients.

1. Introduction

Breast cancer is the most common cancer among women worldwide, with 15% mortality. The incidence of this disease is higher in developing countries. According to the World Health Organization reports, more than 1.2 million women worldwide are diagnosed with breast cancer each year, resulting in almost 411,000 deaths (Parkin 2001). The incidence rate of breast cancer is still increasing in most countries. In Saudi Arabia, this disease ranks first among the malignancies occurring in women, constituting almost 22% of the newly diagnosed cancer cases (Alabdulkarim et al., 2018). Breast cancer incidence in the women population of Saudi Arabia is 25.9%, with 18.2% mortality. It is the second leading cause of cancer related deaths in the Saudi Kingdom, after colorectal cancer. The median age of onset for breast cancer patients in western countries is 65 years; however, the median age of breast cancer onset in Saudi women is 48 years. Thus, it is causing a major concern in Saudi Arabia (Alabdulkarim et al., 2018).

Breast cancer is associated with several risk factors, such as dietary, environmental, family history, genetic alterations in high penetrance genes, etc. It is believed to be a multifactorial disease, resulting from the interaction of both genetic and environmental factors (Board, 2019). Despite the increasing reports of breast cancer, the complete etiology of the disease remains unknown. Recent

* Corresponding author.
E-mail address: reddyparine@gmail.com (N.R. Parine).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2020.05.047
1319-562X/

studies have been more focused on the genetic factors linked with the risk, susceptibility, therapy, initiation, and progression of breast cancer (Fu et al., 2003). Over the past decade, significant progress has been achieved in defining few of the genetic risk factors that help in identifying women with the chance of developing breast cancer (BRCA1 and BRCA2 mutations, etc.) (Aloraifi et al., 2015) (Friedenson, 2005). Regardless of this improved knowledge, the complexity of the combination of different genetic and other external etiological factors of this disease is still unclear. A better understanding of the "genetic mechanisms," underlying the tumor initiation and development of breast cancer, would enable us to identify the prognosis and develop targeted therapies for metastatic breast cancer.

In the past few decades, significant progress has been achieved in aspects of diagnosis and treatment of breast cancer by studying the diverse nature of the tumors and their responsiveness to treatments (Velaga and Sugimoto, 2017). Signal transduction and regulatory ssystems monitor and control cell activities, including growth, function, differentiation, and apoptosis, by regulating the gene expression levels of all the genes directly or indirectly involved in these pathways (Osborne et al., 2004). Genetic variations are majorly responsible for any expression changes and hence, affect the biological diversity and behavior of human cells and tumors. Mutations in two of the most common breast cancer related genes- BRCA1 and BRCA2, are associated with a significant increase in hereditary breast cancer cases and account for approximately 25% of the families with inherited breast cancer (Devilee and Cornelisse 1994). Studies on BRCA1 and BRCA2 genes enhanced our understanding of the genetics of breast cancer, risk assessment, and hence, better management of high risk patients (Kwong et al., 2016). In addition to BRCA1 and BRCA2 mutations, mutations in the genes- PTEN, RAD50, RAD51C, STK11, TP53, ATM, and DNA repair genes such as MLH1, MSH2, MSH6, MUTYH, etc., accounting for 2 to 5% of breast cancer cases, are associated with an increase in the risk of breast cancer (Mackay et al., 2000; Rennert et al., 2012; Stephens et al., 2012; Kotsopoulos et al., 2014; Romero-Laorden and Castro 2017; Velaga and Sugimoto 2017). Pathway-based analyses with expression studies highlighted the functional effects of the gene expression changes, resulting in a better understanding of the metastatic progression and underlying somatic and germline mutations driving the expression changes (Kristensen et al., 2014; Stadler et al., 2014). Studies on genetic variation in genes of metabolic signaling pathways in different population based case control samples demonstrated significant association with colorectal cancer. Genomic approaches related to the traditional pathological approaches led to the betterment of the multidisciplinary management of breast cancer (Ogino et al., 2011). Prognosis, morphology of the disease, patient management, and therapeutics vary among individuals and populations. In the era of post-genomics, treatment strategies are being developed in western countries, and few molecules such as protein kinases are being tested in clinical trials. With the drastic increase in the diagnosis of breast cancer cases in the Kingdom per year, developing risk assessment methods and treatment strategies at the level of the individual and the population will certainly reduce the morbidity and mortality associated with breast cancer and is hence, the need of the hour. There is an immediate need to establish diagnostic, prognostic, and therapeutic importance of breast cancer by pooling all the emerging breast cancer genes of importance.

The present study focused on the identification of genetic variations in different cancer causing genes, raising the hope for identification of personalized prognosis and therapy targets using targeted sequencing of hotspots in cancer related genes.

2. Methodology

2.1. Sample collection

The present study was approved by the IRB board of the King Fahd Medical City (KFMC), Riyadh. Following the IRB approval, 50 breast cancer tissue samples and their matching controls were obtained in RNAlater solution. Blood (5 ml) was also collected from all patients. Demographic data, age at diagnosis, tumor grade, and hormonal receptor status (ER, PR, HER) were recorded for all patients.

2.2. DNA isolation

DNA was isolated from blood samples using Qiagen blood DNA kit. The tissue samples were processed for DNA and RNA extraction with All Prep DNA/RNA mini kit from Qiagen according to the manufacturer protocol. The quality and quantification of the extracted DNA/RNA were assessed on bioanalyzer and nanodrop, respectively. Few tissue samples (5%) yielded low quantity of DNA/RNA. The low yield and low quality of DNA could be attributed to the high fat content in breast tissue. The good quality DNA isolated was used for targeted sequencing studies (n = 50). Details of breast cancer samples such as age and ER, PR, and HER status are shown in Table 1.

2.3. Targeted sequencing

Targeted sequencing of 164 selected genes was performed on germline and somatic DNA derived from the blood and tissue samples of 50 breast cancer patients using a customized panel from Thermo fisher Ion torrent plat form. Libraries were prepared using Thermofisher Ion Ampliseq kit 2.0. The mean sequencing depth of coverage was 100 × overall. Along with this customized panel, Ion AmpliSeq™ BRCA1 and BRCA2 Panel (Life Technologies, Carlsbad, CA, USA) of the respective samples was also studied. All the targeted fragments in the 164 genes were successfully sequenced in all the samples. Sanger sequencing was used to validate the mutations identified in the Targeted sequencing and BRCA panel sequencing.

2.4. Gene-gene interactions

A curated regulatory relationship was established among target genes obtained from the present study. In our analysis, we tried to establish a new association among the gene-gene interactions. We generated a curated mRNA-mRNA regulatory network using Gene mania (https://genemania.org/). The targeted genes in the network were involved in tumor suppressive or oncogenic role, and were

Table 1
Characteristics of the study subjects used in the present study.

Variable	Parameter	Cases N (%)
Total persons	–	50
Age*	≤50	21 (42%)
	>50	29 (58%)
Stage	I	13 (26%)
	II	20 (40%)
	III	11 (22%)
	IV	6 (12%)
Estrogen receptor	ER+	38 (76%)
	ER–	12 (24%)
Progesterone receptor	PR+	32 (64%)
	PR–	18 (36%)
HER Status	HER+	26 (52%)
	HER-	24 (48%)

* Median age (50 ± 7.1).

likely to play important regulatory roles in the pathogenesis of the disease.

2.5. Data analysis

Obtained data were aligned and mapped using hg19 genome and torrent aligner software. After mapping the hg19 reference sequence, high quality clean reads that matched with target regions were included, while the low quality reads were excluded. Variants were identified in all the samples using torrent variant caller software. The resulting data was saved as Variant Call Format (VCF) files. These VCF files were used for annotation to compare with dbSNP (Single Nucleotide Polymorphism Database), the 1000 Genomes Project database, COSMIC (the Catalog of Somatic Mutations in Cancer), Ensemble, and local~1000 Arab exomes databases, and identify the characteristics of the variants. Non-pathogenic (benign) mutations were excluded from further analysis, while mutations with pathogenic effects and unreported mutations were assessed for the Insilico prediction of possible damaging effects using mutation taster, polyphen-2 and SIFT.

3. Results

3.1. Sample demographics

The present study comprised of 50 breast cancer patients. The median age of the patients was 50.2 years (ranging from 24 to 72 years). Twenty nine patients (58%) were above 50 years, while the remaining 21 (42%) were below 50 years of age. Histological grade of 66% of the patients showed early-stage disease (clinical stages I and II), while the remaining 34% showed stages III and IV. Most of the tumors were hormonal receptor-positive- 76% were ER(+), 64% were PR(+), and 52% were HER2(+).The remaining were hormonal receptor-negative- 24% were ER(−), 36% were PR(−), 48% were HER2(−), and 8% were triple negative (TN).

3.2. Targeted sequencing

To identify the mutations, blood DNA sequences were compared with matching tissue DNA sequences. Resulted sequences were mapped using hg19 sequence; 93% of the obtained reads were clean, and they were exclusively matched to the target regions. These reads covered 96.30% of the targeted region with at least 98.31% of fold coverage of mean depth in each sample. The average coverage depth for target regions was 97.33%, and the maximum coverage depth was 99% in most of the samples. Therefore, adequate coverage reliably detected DNA variants within the targeted regions. dbSNP, the 1000 Genomes Project database, COSMIC, Ensemble, and local~1000 Arab exomes databases were used to identify the variants. Based on the reference genome, we observed an average of 7100 variants in the 164 genes in each sample. Among the observed variants, most of them were Single Nucleotide Polymorphisms (SNPs), matching with dbSNP, COSMIC, HGMD, and 1000 Genomes Project databases. We identified the novel variants by filtering false positive variants.

In data analysis, we observed the novel and reported variants in various genes and pathways. BRCA1 and BRCA2 mutations had the most prevalent alterations, found in 68% and 32% samples, respectively. About 82% of BRCA1 and BRCA2 mutations were matched with COSMIC, HGMD, dbSNP, and 1000 Genomes Project databases (Table 2). Further screening of the mutations was done to understand how the mutated gene was involved in various pathways. The interaction among variants, their role in pathways and networks, and their structural changes were assessed to understand their role in carcinogenesis. We considered variants with more than 5% frequency for further confirmation studies using bidirectional Sanger sequencing done by the BigDye Terminator v1.1 technology (Fig. 1).

Sequenced regions of the fifty patients contained 1055 germline variants, of which fourteen (1.33%) were insertions and deletions, and the remaining were single-nucleotide substitutions. Of the single-nucleotide substitutions, 295 (27.96%) were predicted to be synonymous, 568 (53.84%) missense, 40 (3.79%) nonsense, 13 (1.23%) stop gain, and 125 (11.85%) splice site. After final filtration and validation, we found protein-truncating and non-synonymous missense and splice site mutations in the known susceptibility genes for breast cancer. We identified a total of fourteen point mutations and one deletion related to BRCA1, BRCA2, and RAD50 genes from our BRCA panel analysis of breast cancer samples (Table 2). We also observed a few recurrent mutations affecting the BRCA1 gene (Y132C, N503H, and F439L) (Table 2). Loss of heterozygosity (LOH) was observed in BRCA1 (c.395A > G, c.1507A > C, and c.1315T > C), BRCA2 (c.1166C > A), and RAD50 (c.2651G > A).

Additionally, we identified 37 potential mutations in 25 breast cancer risk associated genes (other than BRCA1 and BRCA2), such as TP53 (n = 7; one stop gain, four non-synonymous, and two deletions), GPC3 (n = 1), RHBDF2 (n = 2), MLH1 (n = 1), MLH3 (n = 1), AXIN2 (n = 1), NF1 (n = 1), PTCH2 (n = 1), KIT (n = 1), RB1 (n = 1), SETBP1 (n = 1), BUB1 (n = 1), APC (n = 1), SMARCE1 (n = 1), TMEM127 (n = 1), VHL (n = 1), and WRN (n = 1) (Table 3). Out of these, thirteen were predicted to be disease causing. The mutations of TP53 (p.R81X), VHL (p.E52X), and BRCA2 (p.K3326X), which lead to an aberrant transcript with a premature stop codon, were reported for the first time in Saudi breast cancer patients.

Along with the new mutations shown above, we also observed already reported SNPs in genes (n = 82), such as FANCB, BARD1, HNF1A, PALLD, FANCA, FANCD2, MUTYH, WRN, APC, POLD1, MET, CYLD, RAD50, ATM, BRCA1, MSH6, POLE, RSPO1, BRCA2, ERCC2, MLH1, MYH8, TSC2, ALK, BUB1, MLH3, MSH2, PTCH1, RECQL4, SDHA, ATR, CBL, FLCN, KDR, MITF, PDGFRB, RHBDF2, SLX4 , AXIN2, BMPR1A, DICER1, EGFR, EXT2, FANCI, GPC3, PMS2, PTCH2, RAD51D, RUNX1, VHL, WAS, BLM, BRIP1, CDH1, CDKN1B, CHEK2, CTHRC1, EPCAM, ERCC3, FANCF, FANCG, FANCL, FANCM, FH, HNF1B, KLLN, LIG4, MEN1, MRE11A, MSR1, NBN, NF2, NTRK1, PALB2, PPM1D, PRKAR1A, PTPN11, RB1, RET, RNF168, SETBP1, SMAD4, STK11, TP53, WT1, and XPC (Supplementary Tables 1 and 2). Few of these SNPs were in the exonic region.

3.3. Analysis for structural conformation changes

The possible effect of the mutation on the protein and the stability of the mutated amino acids were analyzed using HOPE and Pop Music.

Substitution of TP53 (Val25Asp) Valine 25 by Aspartic Acid resulted in a low z score, from −6.12 to −8.01. The total energy deviation was −2.9, which might have a negative effect on the TP53 structure and function. The mutated residue was situated in the protein activity domain, where it interacted with other proteins. This mutation might affect the function of the protein (Fig. 2 and Supplementary Fig. 1).

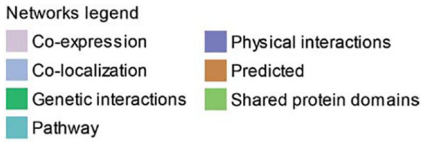
Substitution of Tyrosine 88 by Cysteine 88 resulted in the decrease in ProSA-web z score, from −6.05 to −7.92. The total energy deviation was −3.3, which might have an unfavorable effect on the TP53 structure and function. Both the wild and variant amino acids vary in size. The wild type Tyrosine is bigger than Cysteine and may not fit in the protein structure, hence, affecting the structure (Fig. 3 and Supplementary Fig. 2).

Table 2
Short listed and validated variants in BRACA1 & BRACA2.

Sample	Gene	Base Change	Amino Acid Change	Zygosity
BCB-D-1	BRCA1	c.395A > G	WT	LOH
	BRCA1	c.1507A > C	WT	LOH
	BRCA1	c.1315T > C	WT	LOH
BCB-D-13	BRCA2	c.1166C > A	WT	LOH
	BRCA2	c.7453 + 6G > A	–	Het
BCB-D-18	BRCA1	c.395A > G	p.Y132C	Het
	BRCA1	c.1507A > C	p.N503H	Het
	BRCA1	c.1315T > C	p.F439L	Het
BCB-D-19	BRCA2	c.8755-1G > C	–	Het
BCB-D-20	BRCA2	c.122C >T	p.P41L	Het
	RAD50	c.2651G > A	p.R884H	Het
BCB-D-23	BRCA2	c.122C >T	p.P41L	Homo
BCB-D-25	BRCA1	c.395A > G	p.Y132C	Homo
	BRCA1	c.1507A > C	p.N503H	Homo
	BRCA1	c.1315T > C	p.F439L	Homo
BCB-D-29	BRCA1	c.395A > G	p.Y132C	Het
	BRCA1	c.1507A > C	p.N503H	Het
	BRCA1	c.1315T > C	p.F439L	Het
BCB-D-31	–	–	–	–
BCB-D-32	BRCA1	c.395A > G	p.Y132C	Homo
	BRCA1	c.1507A > C	p.N503H	Het
	BRCA1	c.1315T > C	p.F439L	Het
BCB-D-35	BRCA2	c.5291C > G	p.S1764X	Het
	BRCA2	c.9976A > T	p.K3326X	Het
BCB-D-46	BRCA2	c.122C > T	p.P41L	Het
BCB-D-60	BRCA1	c.395A > G	p.Y132C	Het
	BRCA1	c.1507A > C	p.N503H	Het
	BRCA1	c.1315T > C	p.F439L	Het
	BRCA2	c.7534C > T	p.L2512F	Het
BCB-D-63	BRCA2	c.8382C > G	p.F2794L	Het
BCB-D-64	BRCA1	c.395A > G	p.Y132C	Het
	BRCA1	c.1507A > C	p.N503H	Homo
	BRCA1	c.1315T > C	p.F439L	Het
BCB-D-67	BRCA1	c.395A > G	WT	LOH
	BRCA1	c.1507A > C	WT	LOH
	BRCA1	c.1315T > C	WT	LOH
	BRCA2	c.6269A > G	p.H2090R	Homo
BCB-D-68	RAD50	c.2651G > A	WT	LOH
BCB-D-69	BRCA2	c.10078A > G	p.K3360E	Het

3.4. Analysis of function based on gene-gene interactions

The gene/protein-gene/protein interactions of BRCA1, BRCA2, and TP53 were analyzed using GeneMania and Cytospace tools. Most of the analyzed (n = 37) genes were co-expressed, were co-localized, interacted physically, and shared protein domains and pathways directly/indirectly with a number of proteins (Fig. 4). Top thirty seven topologically important genes were obtained from the network of 217 proteins and 216 interactions (Fig. 4). Network topology showed that the thirty seven mRNAs are clustered in a complex hub, and all of them might be regulated in a similar way, since mRNAs that have similar patterns of expression can be considered to be in the same cluster. Eighty six genes, which showed known mutations (dbSNP), were also analyzed for protein–protein interactions (Fig. 5). The network topology revealed that 33% of these genes are co-expressing and they are involved in major cell functions such as DNA recombination (27 genes), double-stranded DNA repair (19 genes), cell division (13 genes), and nuclear division (21 genes) etc.



4. Discussion

In the present study, we used a combination of high throughput and novel methods, such as target capture enrichment and NGS, to study all the critical genes involved in signaling pathways associated with breast cancer. This study aimed to identify genetic variants correlated with increased susceptibility to breast cancer. We used a customized panel of 164 genes that were reported to play key roles in cancer and a BRCA panel from thermos scientific. The present study comprised of fifty breast cancer patients. The median age of the patients was 50.2 years (ranging from 24 to 72 years). In contrast to the breast cancer cases reported in western countries at the median age of onset of 65 years, the breast cancer cases in Saudi women were reported at the median age of onset of 48 years, causing a major concern in Saudi Arabia. In the present study, twenty nine breast cancer patients (58%) were above 50 years, while the remaining twenty one (42%) were below 50 years. Most of the breast cancer samples (66%) showed early-stage disease (Stages I and II), while the remaining (34%) were in late stages (Stages III and IV). Most of the tumors were hormonal receptor-positive- 76% were ER(+), 64% were PR(+), and 52% were HER2(+). The remaining were hormonal receptor-negative- 24% were ER(−), 36% were PR(−), 48% were HER2(−), and 8% were TN. In the present study, we intended to identify the germline mutations in blood samples by comparing them with matching tissue samples. Resulted sequences were mapped using hg19 sequence; 93% of the obtained reads were clean, and they were exclusively matched to the target regions. These reads covered 96.30% of the targeted region with at least 98.31% of fold coverage of mean depth in each sample. The average coverage depth for target regions was 97.33%, and the maximum coverage depth was 99% in most of the samples.

The in-depth coverage of the sequencing allowed by NGS made it feasible to catalogue all levels of germline mutations attained during different stages of cancer (Beltran et al., 2013; Hagemann et al., 2013; Chen et al., 2015). Recent studies on leukemia and solid tumors concentrated only on the exome to increase the likelihood of identifying driver mutations (Garraway and Lander 2013; Huang et al., 2014). Though the mutations lead to inactivation of the signaling pathways, the efficiency remains in deconvoluting and validating the critically important mutations in heterogeneous samples and assessing them with a functional approach in association with cancer. The screening of validated SNPs, available in databases such as COSMIC, Ensemble, NCBI, etc., and those revealed from the proposed study of targeted sequencing provided information on allele frequencies and behavior of SNPs in Saudi population. This further helped us in the identification of breast cancer risk alleles. In the BRCA panel analysis, we observed recurrent mutations affecting the BRCA1 gene (Y132C, N503H, and F439L) (Table 2). LOH was observed in BRCA1 (c.395A > G, c.1507A > C, and c.1315T > C), BRCA2 (c.1166C > A), and RAD50 (c.2651G > A). Overall, BRCA1 (68%) mutations were higher than BRCA2 (32%) mutations. Our results are supporting that the 30% of the breast cancer cases are caused by BRCA1 and BRCA2 mutations and out of this 5% are deleterious (Abdulrahman and Rahman, 2012, Martínez-Ferrandis et al. 2003, Musolino et al. 2005). Around, 40% patients with BRCA1 & BRCA2 mutation carried at least one concomitant variants of DNA repair and tumor suppressor genes. A similar pattern was observed in lung cancer patients with BRCA2 mutations in Chinese population (Fang et al. 2019).

In the customized panel analysis, we identified thirty seven potential mutations in twenty five breast cancer risk associated genes (other than BRCA1 and BRCA2). Out of these, most of the mutations were observed in TP53. After filtration, we observed the mutations in TP53 (n = 7) - one stop gain (p.R81X), four non-synonymous (p.R81X, p.Y88C, p.R141H, and p.V25D), and two deletions (c.59delC and c.327delC). The other mutations were GPC3 (p.E184Q), RHBDF2 (c.1551 + 3G > C and p.A71S), MLH1 (c.208-1G > C), MLH3 (p.R224T), AXIN2 (p.L688L), NF1 (p.

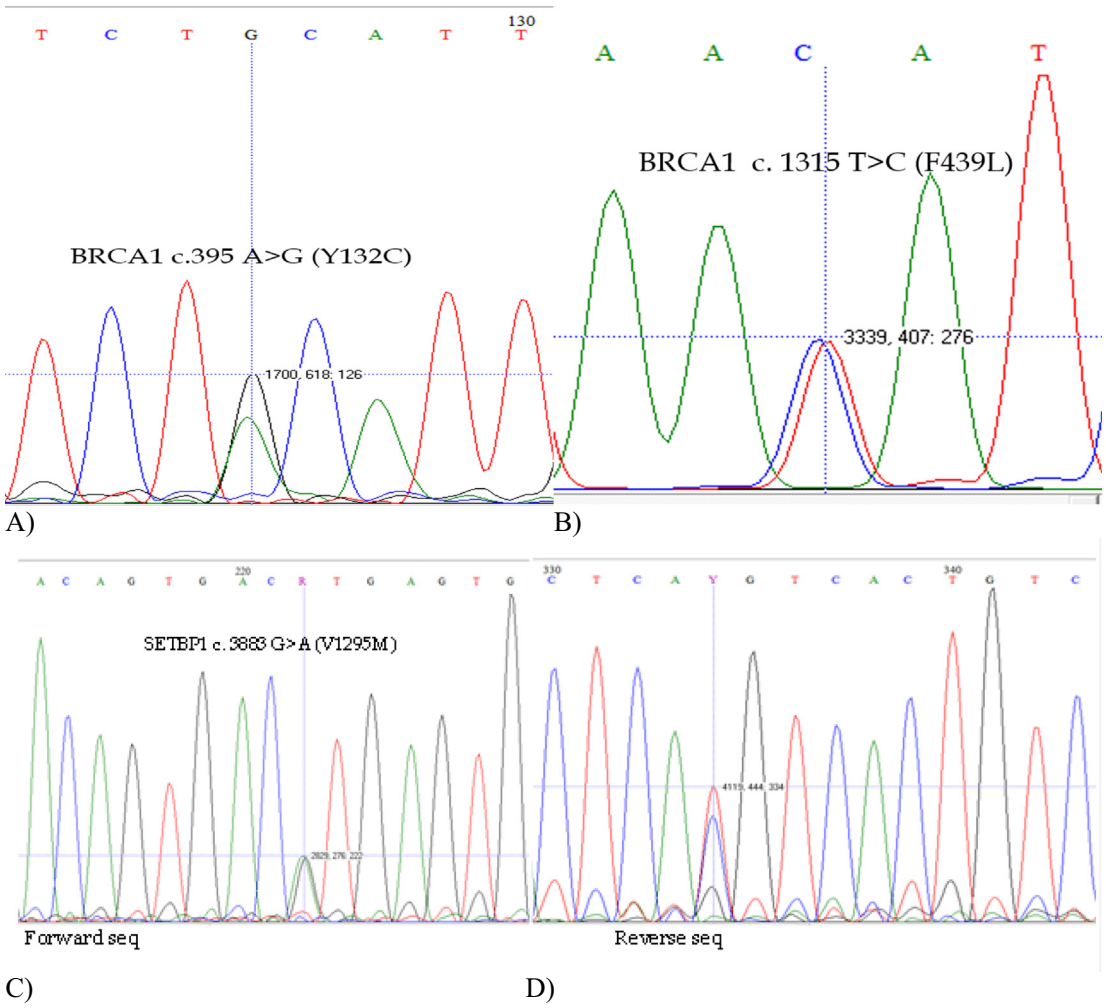


Fig. 1. Validation and confirmation of mutations observed in breast cancer patients. The position of the variants and the amino acid change is in–dicated in exons of BRCA1 and SETBP1 genes, A). BRCA1 A > G (Y132C), B) BRCA1 T > C (F439L), C) SETBP1 G > A (V1295M) Forward sequence, and D) SETBP1 T > C (V1295M) Reverse sequence respectively.

E2210Q), PTCH2 (c.1083 + 3C > T), KIT (p.G803S), RB1 (c.1498 + 1 G > A), SETBP1 (p.V1295M), BUB1 (c.1699-2A > G), APC (p. R445T), SMARCE1 (p.K271fs), TMEM127 (p.R94Q), VHL (p.E52X), and WRN (p.V537fs) (Table 3). Out of these, thirteen were predicted to be disease causing. The mutations of TP53 (p.R81X), VHL (p.E52X), and BRCA2 (p.K3326X), which lead to an aberrant transcript with a premature stop codon, were reported for the first time in Saudi breast cancer patients. We observed seven splicing mutations, out of which BUB1 (c.1699-2A > G) mutation was splicing and damaging. Exonic mutations in NF1 (c.6628G > C (p. E2210Q)) and TP53 (c.263A > G (p.Y88C)) were observed to be splicing and damaging. The most significant mutations were located in the exonic regions, which made these findings potentially strong. The functional significance of identifying mutations in splicing regions could be mediated through mapping of these mutations to transcription sites (Barash et al., 2010). Interestingly, some significant variants in our data, e.g. mutations within BRCA1 and TP53, may affect the structure of their genes during the mechanism of a stop codon, essential splice, or amino acid change. It would therefore be interesting to investigate the role of these SNPs in carcinogenesis.

Due to less number of samples, we also focused on common variants, defined as minor variant frequency of at least 0.05. We identified 15 genes with potential functional consequences mapped to a number genes involved in cell motility and repair eg: GPC3, RHBDF2, MLH1, MLH3, AXIN2, NF1, PTCH2, KIT, RB1, SETBP1, APC, SMARCE1, TMEM127, VHL, and WRN. Structural changes in most of these genes are reported to inactivate their function. Mainly tumor suppressor genes such as APC, RB1, NF1, TP53, TMEM127, VHL, PTCH2 (Herschkowitz et al. 2008, Pharoah et al. 1999, Li et al. 2017) and DNA repair genes MLH1. MLH3, WRN (Fang et al. 2019) are associated with poor prognosis of breast cancer patients. In the present study we observed the NF1 (E2210Q) mutation in one of the younger aged patient (<50 Year). This is supporting previous studies which reported that NF1 mutations might increase breast cancer incidence in younger aged patients (Suarez-Kelly et al. 2019). SMARCE1 gene which showed an exonic splicing mutation in our study has been reported to involve in regulation of metastasis in breast cancer through HIF1A/PTK2 pathway (Sethuraman et al. 2016). The other genes are not well known to cause mutations in breast cancer and they may represent specific variant driver genes, out of these GPC3 gene has been reported to play a critical role in epithelial-mesenchymal transition mechanism which is vital in tumor metastasis in hepatocellular carcinoma (Guo et al., 2020).

Due to limited number of samples in the compared groups, we also focused on common variants, which were defined as minor variants with frequency of at least 0.05. The variant call algorithm was originally set to increase the sensitivity of detection of the SNPs, so that potentially relevant variants were not missed. Top ranked SNPs with potential functional consequences were mapped to a number genes involved in recombination and DNA repair, e.g.

Table 3
Shortlisted and validated variants from targeted sequencing.

Sample	Germline mutations						
	Gene	Base Change	Amino Acid Change	Zygosity	Mutationtaster prediction/score	Polyphen score	Region/ damage
BCB-D-1	AXIN2	c.2062C > T	p.L688L (POLY)	Homo	–	–	–
BCB-D-2	TP53	c.263A > G	p.Y88C	Het	Disease causing (0.99)	1	Exonic; splicing; Damaging
BCB-D-4	KIT	c.2407G > A	p.G803S	Het	Disease causing (0.99)	1	Exonic damaging
BCB-D-8	TP53	c.327delC	p.S109fs	Het	Disease causing (0.99)	1	Exonic damaging
BCB-D-20	SETBP1	c.3883G > A	p.V1295M	Het	Polymorphism (0.99)	–	
BCB-D-25	MLH3	c.671G > C	p.R224T	Het	Disease causing (0.99)	0.962	Exonic Damaging
	APC	c.1334G > C	p.R445T	Het	Disease causing (0.99)	1	Exonic damaging
	TP53	c.74 T > A	p.V25D	Het	Disease causing (0.99)	1	Damaging
BCB-D-32	RHBDF2	c.1551 + 3G > C	–	Het			splicing
	TMEM127	c.281G > A	p.R94Q	Het		0.681	Exonic Probably damaging
	FANCM	c.808C > T	p.R270C	Het	Disease causing (0.99)	0.996	Exonic Damaging
BCB-D-43	GPC3	c.550G > C	p.E184Q	Het	Disease causing (0.99)	0.991	Damaging
BCB-D-45	GPC3	c.550G > C	p.E184Q	Het	Disease causing (0.99)	0.991	Damaging
BCB-D-48	TP53	c.241C > T	p.R81X	Het			Stop Gain Mutation
BCB-D-54	SMARCE1	c.813_828del	p.K271fs	Homo			Exonic;Splicing
BCB-D-59	TP53	c.241C > T	p.R81X	Het			Stopgain
	VHL	c.154G > T	p.E52X	Homo			Stopgain
	MLH1	c.208-1G > C	–	Het			Splicing
BCB-D-64	TP53	c.59delC	p.P20fs	Het			Deletion
	BUB1	c.1699-2A > G	–	Homo	Disease causing (0.99)	1	Splicing Damaging
	RB1	c.1498 + 1G > A	–	Het			Splicing
BCB-D-67	TP53	c.422G > A	p.R141H	Homo	Disease causing (0.99)	1	Damaging
	NF1	c.6628G > C	p.E2210Q	Het	Disease causing (0.99)	0.989	Exonic splicing Damaging
BCB-D-68	PTCH2	c.1083 + 3C > T	–	Het			Splicing
BCB-D-69	ATM	c.7330G > A	p.E2444K	Het	Disease causing (0.99)	0.996	Exonic D
BCB-D-71	WRN	c.1612dupA	p.V537fs	Het			Frameshift
	RHBDF2	c.211G > T	p.A71S	Het		0	ns SNV

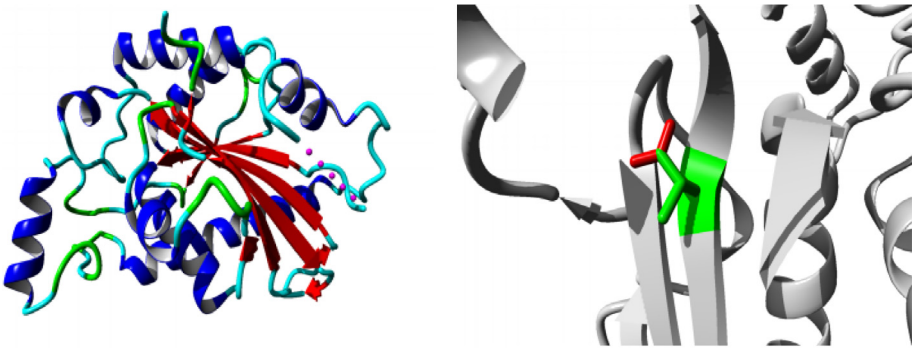


Fig. 2. The TP53 Val25Asp protein region structures shown with both the wild type and the mutant residues.

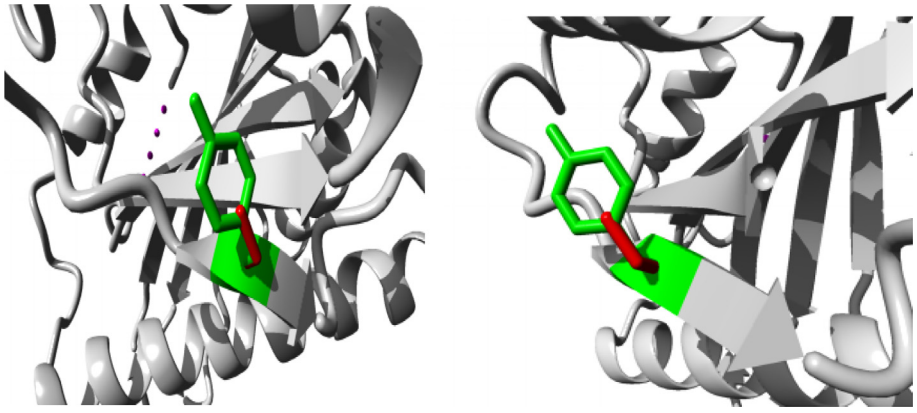


Fig. 3. The TP53 Tyr88Cys protein region structures shown with both the wild type and the mutant residues.

TP53, ATM, XPC, and XRCC3. Our study identified several SNPs that were associated with increased risk of breast cancer in the population of Saudi Arabia; yet the generalization of these results to other populations remains to be established. It also enhanced our knowledge of risk assessment, early detection, therapy, and prevention of breast cancer in the Kingdom. The stringent study of the well-

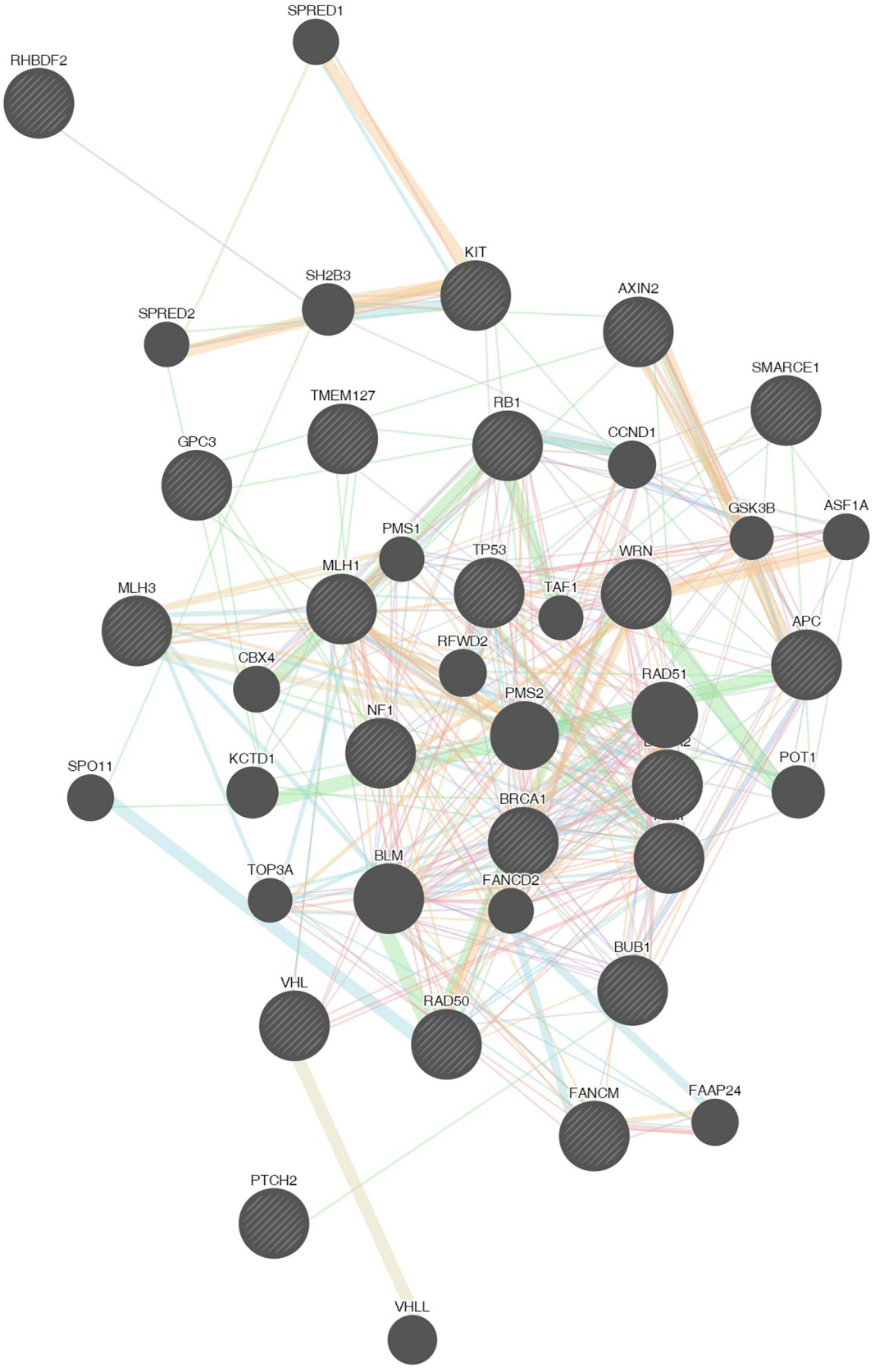


Fig. 4. Protein-Protein interaction network of BRCA1, BRCA2, and TP53 genes.

studied signaling pathway genes associated with cancer, and their screening, by following recent advance technologies like next-generation DNA sequencing (NGS) and customized arrays, and validation in the population of Saudi, will be extremely helpful in initiating the development of better prognostic tools and identification of novel therapeutic targets.

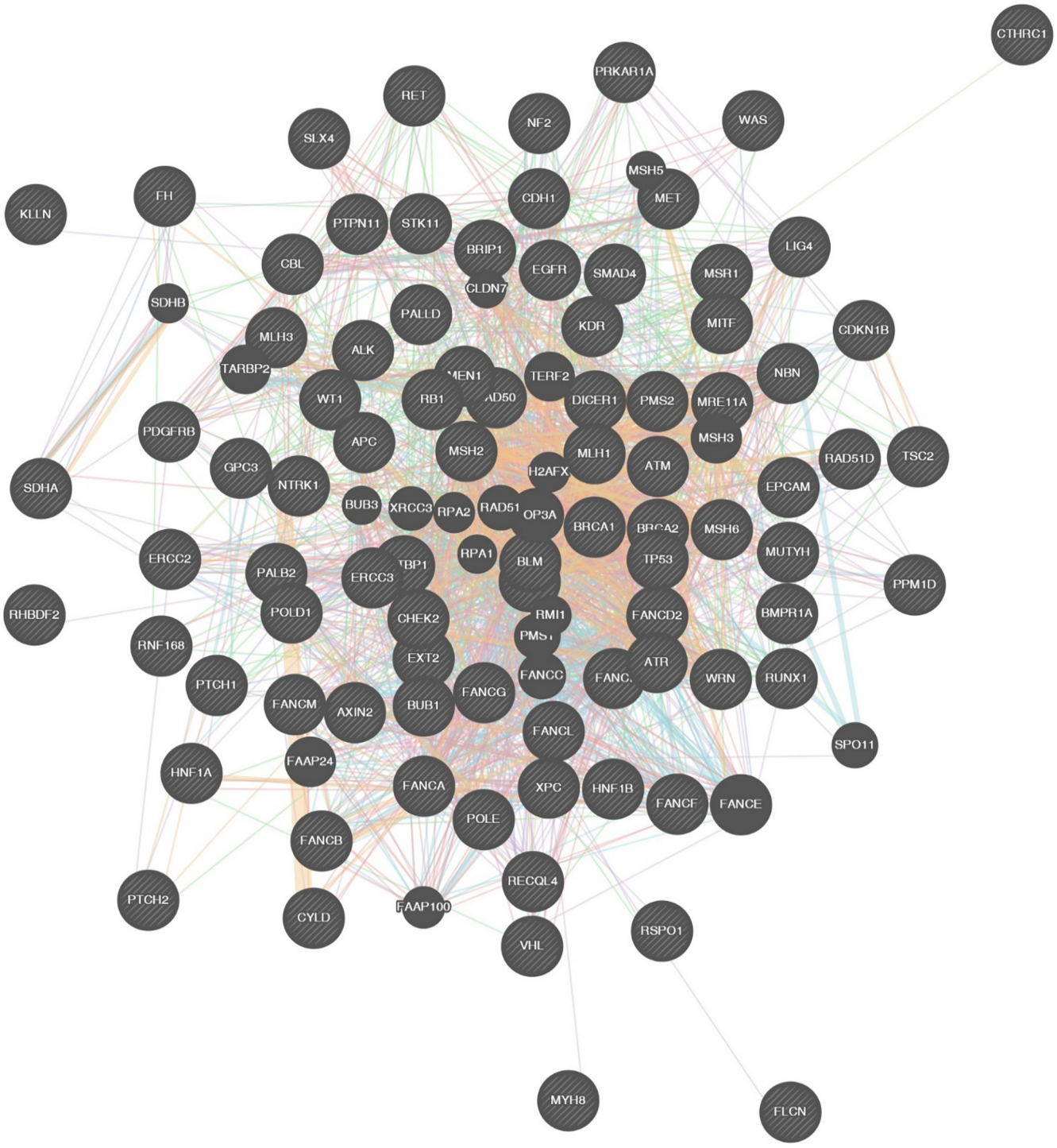


Fig. 5. Extended GeneMania network from the present project- FANCB, BARD1, HNF1A, PALLD, FANCA, FANCD2, MUTYH, WRN, APC, POLD1, MET, CYLD, RAD50, ATM, BRCA1, MSH6, POLE, RSPO1, BRCA2, ERCC2, MLH1, MYH8, TSC2, ALK, BUB1, MLH3, MSH2, PTCH1, RECQL4, SDHA, ATR, CBL, FLCN, KDR, MITF, PDGFRB, RHBDF2, SLX4, AXIN2, BMPR1A, DICER1, EGFR, EXT2, FANCI, GPC3, PMS2, PTCH2, RAD51D, RUNX1, VHL, WAS, BLM, BRIP1, CDH1, CDKN1B, CHEK2, CTHRC1, EPCAM, ERCC3, FANCF, FANCG, FANCL, FANCM, FH, HNF1B, KLLN, LIG4, MEN1, MRE11A, MSR1, NBN, NF2, NTRK1, PALB2, PPM1D, PRKAR1A, PTPN11, RB1, RET, RNF168, SETBP1, SMAD4, STK11, TP53, WT1, and XPC were used as input for the GeneMania app in Cytoscape.

5. Conclusion

Targeted genome capture, enrichment, and sequencing with high coverage using NGS during screening would have a potential impact in personalized genome sequencers. Targeted sequencing of cancer causing genes using NGS approach with sufficient coverage to detect the vast majority of variations would be a very powerful tool for screening and identification of variants causing cancer. This would facilitate analysis of large patient cohorts within short periods of time and at lower cost, compared with whole genome and whole exome sequencing. SNPs, studied in various GWAS studies, mostly occur in Caucasians, and hence, they are not well known in the Middle East. The present study demonstrated that the targeted genomic profiling of breast cancer could reveal new molecular markers. Validation of mutations reported in the present study in other Arab populations and further utilizing

them in the evaluation of the risk-predicting model is an important aspect. This will assist clinicians in assessing the risk, prognosis, and treatment of breast cancer.

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.

Acknowledgments

This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award no. 12-MED2622-02.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2020.05.047.

References

Abdulrahman, Ganiy Opeyemi, Rahman, Ganiyu Adebisi, 2012. Epidemiology of Breast Cancer in Europe and Africa. J. Cancer Epidemiol. 2012, 1–5. https://doi.org/10.1155/2012/915610.

Alabdulkarim, B., Hassanain, M., Bokhari, A., AlSaif, A., Alkarji, H., 2018. Age distribution and outcomes in patients undergoing breast cancer resection in Saudi Arabia: A single-institute study. Saudi Med. J. 39 (5), 464.

Aloraifi, F., Boland, M.R., et al., 2015. Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. Surg. Oncol. 24, 100–109.

Barash, Y., Calarco, J.A., et al., 2010. Deciphering the splicing code. Nature 465, 53.

Beltran, H., Yelensky, R., et al., 2013. Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. Eur. Urol. 63, 920–926.

Board, P.C.G.E., 2019. Cancer Genetics Risk Assessment and Counseling (PDQ®). PDQ Cancer Information Summaries [Internet], National. Cancer Institute (US).

Chen, K., Meric-Bernstam, F., et al., 2015. Clinical actionability enhanced through deep targeted sequencing of solid tumors. Clin. Chem. 61, 544–553.

Devilee, P., Cornelisse, C.J., 1994. Somatic genetic changes in human breast cancer. BBA Rev. Cancer. 1198, 113–130.

Fang, W., Cai, X., Zhou, H., Wang, Y., Zhang, Y., Hong, S., Shao, Y., Zhang, L. 2019. BRCA1/2 germline mutations and response to PARP inhibitor treatment in lung cancer. J. Clin. Oncol. 37 (15), e13007.

Friedenson, B., 2005. BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. Med. Gen. Med. 7, 60.

Fu, Y.-P., Yu, J.-C., et al., 2003. Breast cancer risk associated with genotypic polymorphism of the nonhomologous end-joining genes: a multigenic study on cancer susceptibility. Cancer Res. 63, 2440–2446.

Garraway, L.A., Lander, E.S., 2013. Lessons from the cancer genome. Cell 153, 17–37.

Guo, M., Zhang, H., Zheng, J., Liu, Y., 2020. Glypican-3: A New Target for Diagnosis and Treatment of Hepatocellular Carcinoma. J. Cancer 11 (8), 2008.

Hagemann, I.S., Cottrell, C.E., et al., 2013. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. Cancer Gen. 206, 420–431.

Herschkowitz, J.I., He, X., Fan, C., Perou, C.M., 2008. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. Breast Cancer Res. 10 (5), R75.

Huang, Y., Zheng, J., et al., 2014. Discovery of somatic mutations in the progression of chronic myeloid leukemia by whole-exome sequencing. Genet. Mol. Res. 13, 945–953.

Kotsopoulos, J., Metcalfe, K., et al., 2014. Prospective study of high-risk, BRCA1/2-mutation negative women: the 'negative study'. BMC cancer. 14, 221.

Kristensen, V.N., Lingjærde, O.C., et al., 2014. Principles and methods of integrative genomic analyses in cancer. Nat. Rev. Cancer. 14, 299–313.

Kwong, A., Shin, V.Y., et al., 2016. Comprehensive spectrum of BRCA1 and BRCA2 deleterious mutations in breast cancer in Asian countries. J. Med. Gen. 53 (1), 15–23.

Li, S., Shen, Y., Wang, M., Yang, J., Lv, M., Li, P., Chen, Z., Yang, J., 2017. Loss of PTEN expression in breast cancer: association with clinicopathological characteristics and prognosis. Oncotarget 8 (19), 32043.

Mackay, H., Cameron, D., et al., 2000. Reduced MLH1 expression in breast tumors after primary chemotherapy predicts disease-free survival. J. Clin. Oncol. 18, 87.

Martínez-Ferrandis, J.I., Vega, A., Chirivella, I., Marín-García, P., Insa, A., Lluch, A., Carracedo, A., Chaves, F.J., García-Conde, J., Cervantes, A., Armengod, M.E., 2003. Mutational analysis of BRCA1 and BRCA2 in Mediterranean Spanish women with early-onset breast cancer: Identification of three novel pathogenic mutations. Hum. Mutat. 22 (5), 417–418.

Musolino, A., Naldi, N., Michiara, M., Bella, M.A., Zanelli, P., Bortesi, B., Capelletti, M., Savi, M., Neri, T.M., Ardizzoni, A., 2005. A breast cancer patient from Italy with germline mutations in both the BRCA1 and BRCA2 genes. Breast Cancer Res. Treat. 91 (2), 203–205.

Ogino, S., Chan, A.T., et al., 2011. Molecular pathological epidemiology of colorectal neoplasia: an emerging transdisciplinary and interdisciplinary field. Gut 60, 397–411.

Osborne, C., Wilson, P., et al., 2004. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. Oncologist. 9, 361–377.

Parkin, D.M., 2001. Global cancer statistics in the year 2000. Lancet Oncol. 2, 533–543.

Pharoah, P.D.P., Day, N.E., Caldas, C., 1999. Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. Br. J. Cancer 80 (12), 1968–1973.

Rennert, G., Lejbkowicz, F., et al., 2012. MutYH mutation carriers have increased breast cancer risk. Cancer 118, 1989–1993.

Romero-Laorden, N., Castro, E., 2017. Inherited mutations in DNA repair genes and cancer risk. Curr. Prob. Cancer. 41, 251–264.

Sethuraman, A., Brown, M., Seagroves, T.N., Wu, Z.H., Pfeffer, L.M., Fan, M., 2016. SMARCE1 regulates metastatic potential of breast cancer cells through the HIF1A/PTK2 pathway. Breast Cancer Res. 18 (1), 81.

Stadler, Z.K., Schrader, K.A., et al., 2014. Cancer genomics and inherited risk. J. Clin. Oncol. 32 (7), 687.

Stephens, P.J., Tarpey, P.S., et al., 2012. The landscape of cancer genes and mutational processes in breast cancer. Nature 486 (7403), 400.

Suarez-Kelly, L.P., Yu, L., Kline, D., Schneider, E.B., Agnese, D.M., Carson, W.E., 2019. Increased breast cancer risk in women with neurofibromatosis type 1: a meta-analysis and systematic review of the literature. Hereditary Cancer Clin. Pract. 17 (1), 12.

Velaga, R., Sugimoto, M., 2017. Future Paradigm of Breast Cancer Resistance and Treatment. Resistance to Targeted Therapies in Breast Cancer. Springer, pp. 155–178.