**Abstract**

**Background:** Breast cancer (BC) remains one of the leading causes of death in women worldwide. The *BRCA1* deleterious mutation has a significant role in developing BC, and the risk has been estimated to be 46–87%. Many studies emphasize the need for mining *BRCA1* gene mutations that might have a role in BC pathogenesis and could affect early disease onset. This study was conducted to screen for possible pathogenic single nucleotide polymorphisms (SNPs) in *BRCA1*, targeting three regions: two in exon 11 and the third in exon 20.

**Methods:** 45 blood samples were collected from patients diagnosed with BC. DNA was extracted and selected regions were amplified by PCR using three sets of primers - two within exon 11 and one within exon 20 of *BRCA1*. Subsets of 10 samples were selected for each primer set (30 PCR products) and sequenced. Sequences were analyzed using various bioinformatics tools.

**Results:** Two missense mutations were found, Q356R (rs1799950) in one patient (27 years old) and a novel SNP, V1736D, in three premenopausal...
patients (≤45 years), which were located within exons 11 and 20, respectively. Both detected variants were heterozygous, a status found in all patients detected with such monoallelic variation. Both missense variants underwent *in silico* analysis. The well-known mutation, rs1799950, was predicted to alter the protein activity, conferred by a mutant residue (R-Arg), owing to the position with a bigger size and positive charge. The novel SNP, V1736D, was predicted to play a role in the pathogenesis of BC.

**Conclusion:** Both variants require further investigation, firstly to assess their contribution to BC and secondly to determine their potential diagnostic value when assessed in a larger population.
Corresponding authors: Mohamed Elmogtba Mouaweia Mohamed Aabdein (mogtaba8788@hotmail.com), Alsmawal Awad Mohammed Elimam (toteil1@hotmail.com)

Author roles: Aabdein MEMM: Conceptualization, Data Curation, Methodology, Resources; Elimam AAM: Conceptualization, Data Curation, Investigation, Methodology, Resources, Software, Writing – Original Draft Preparation; Altayb HN: Formal Analysis, Investigation, Software, Validation, Visualization; Eldeen MEFM: Data Curation, Formal Analysis, Methodology, Validation, Visualization; Gasemelseed MM: Investigation, Software, Writing – Original Draft Preparation; FadiAlla AA: Investigation, Software, Writing – Original Draft Preparation; Osman MM: Investigation, Software, Writing – Original Draft Preparation; Osman SA: Investigation, Software, Writing – Original Draft Preparation; Saeed HA: Investigation, Software; Ali MS: Investigation, Software, Writing – Original Draft Preparation; Siddig T: Investigation, Software, Writing – Original Draft Preparation; Osman RA: Investigation, Software; Elhadi RA: Investigation, Software; Hamid MMA: Data Curation, Formal Analysis, Methodology, Validation, Visualization, Writing – Review & Editing; Salih MA: Conceptualization, Formal Analysis, Investigation, Project Administration, Software, Supervision, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

How to cite this article: Aabdein MEMM, Elimam AAM, Altayb HN et al. BRCA1 novel mutation V1736D and in silico analysis of SNP Q356R in Sudanese patients with breast cancer [version 3; referees: 1 approved, 1 approved with reservations, 1 not approved]

Copyright: © 2017 Aabdein MEMM et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

Introduction

Breast cancer (BC) is a very serious issue worldwide, and is one of the leading causes of death in women today. In the US, it was estimated that there were approximately 232,670 new cases of BC and 40,000 BC deaths in 2014, and the number increased with 13,990 new cases and 450 deaths in 2016. In Africa, in 2012, the rate was about 94,000 women with BC, which resulted in 48,000 deaths, and studies in Africa have described a poor outcome with a late diagnosis, due to the aggressiveness of the disease and the absence of screening programs.

In Sudan, BC occurs at the highest frequency among women compared to other types of cancer. In a hospital-based statistical report in Sudan, BC was found to be the most commonly diagnosed malignant tumor and was characterized by early onset and bad prognosis. The report showed invasive ductal carcinoma to be the predominant type (82%), and 74% of patients were <50 years old with an advanced disease stage, indicating that most cases remain undiagnosed for long periods. BRCA1 (OMIM_113705) was mapped in 1994 and subsequently cloned. It is located on chromosome 17 region 2, band 1 (17q21), which is responsible for encoding 1863 amino acids. Since 1995, the BRCA1 tumor suppressor protein has been found to arrest cell proliferation, play an important role in the repairing process of DNA damage, and was suggested to have a role in cell cycle regulation through interacting directly or indirectly with other regulatory molecules. BRCA1 when it is altered becomes deficient, and such loss of function mutates the protein, which not only perturbs chromosomal integrity and genome stability, but increases the mutation rate of other genes. Therefore, it has been proposed that BRCA1 doesn’t directly initiate cancer formation, but enhances the process by making the affected cells highly susceptible to malignant transformation.

Germline mutations in BRCA1 are responsible for a large proportion of inherited predispositions for BC, and individuals that carry an inherited mutation in the BRCA1 gene have a significant risk (46–87%) for developing BC by 70 years of age. Regionally among Africa and locally in Sudan, BRCA1-associated BC has been identified among premenopausal women. This indicates the highly susceptible nature of such a mutation to enhance cancer development earlier during fertile reproductive women. African scientific literature has greatly studied the risk of BC resulting from reproductive factors (such as early menarche, late menopause, and sex hormones), but has not explored much into the genetic predisposition of the disease.

About 13108 SNPs have been reported within human BRCA1, of them 1608 were reported as missense variants and about 151 have been identified to be pathogenic, according to the SNP database at NCBI (http://www.ncbi.nlm.nih.gov/snp). Most germline mutations result from small insertions/deletions frameshift, stop codon or deletion, duplication at exon/intron sequence levels which are met by disruption of the mutated splice sites, altogether resulting in a nonfunctional truncated protein. The SNP variant rs1799950 was observed to have a negative association with BC, by favoring more frequently the control groups than BC patients. This SNP was tested among familial BRCA1 carriers for BC risk association, and a significant association was found within affected families. When the SNP was haplotype-homozygous within affected families, the risk was increased, as in the case of sporadic risk association study. The heterozygosity nature of this mutation has been noticed within different studies, and some studies found the heterozygous variant more frequently within controls, hence it was adversely associated with BC, while in another studies, the homozygous variant was found more frequently within BC patients.

In addition, heterozygosity was functionally assessed among monoallelic BRCA1 mutation carriers of rs1799950, and the results showed that such an alteration could permit variation in protein expression and activity in a haploinsufficient way, which could alter the cell’s normal behavior and result in tumor transformation by enhancing tumorigenesis. Such variation conferred by one mutated copy suggests that the wild-type copy alone is not capable of compensating the loss of the other wild allele. Turkovic et al. found that such a haplotype association was noticed more frequently within deleterious mutation carriers; however, this was observed in a small sample size.

In addition, some studies have found rs1799950 to be associated with early-onset prostate cancer.

Two genetic studies have been conducted in Sudan concerning BRCA1. One was a survey of 2370 students at a girl’s secondary school in Northern Sudan-Marawi, in which the study divided 67 students into two groups (47 students with a family history of BC and 20 with unaffected families) to analyze BRCA1 and BRCA2 mutations. In the first group, which was 2.37% of responders, the frequency of mutations was higher for BRCA1, and most mutations were within exon 11. The study continued to recommend further assessments of this region in subsequent local projects, and this formed the basis of our primer (1 and 2) selection within the present study. The other study, from central Sudan, investigated 34 early onset premenopausal women patients diagnosed with BC (<40 years) and one male patient. The study identified 60 mutations in these patients, five of which were deleterious, affecting the outcome protein.

From the same region within central Sudan, early onset BC premenopausal women have also been investigated for BRCA1 point mutations. The findings revealed the presence of one deleterious variant, 24 neutral variants and eight variants of unknown significance, within which two novel variants were discovered.
Since there have been scarce genetic studies conducted highlighting genetic characteristic and familial risk status of BC patients in Africa, our aim was to screen for the type and spectrum of germline mutations in \textit{BRCA1} by focusing on three regions within the gene, two within exon 11 and one within exon 20, using sequencing, and to further assess the detected variants using \textit{in silico} analysis tools. These regions and their selections were based on the quality of available primers (e.g. best GC content, adequate length, according to previous literature), previous local research findings revealing frequent mutations within exon 11, and the cost.

**Methods**

**Sampling**

This study was carried out in March 2015 at the Radiation and Isotope Center in Khartoum, 2–3 mls of blood were collected randomly from 45 patients diagnosed with BC who attended the center for treatment and follow-up (no other inclusion/exclusion criteria were relevant), using sterile EDTA-K3 vacutainer and kept at -20°C. All the patients were aged between 27 and 80 years old, with a mean of 45.9 years. Early onset cases were more frequent than late onset: 25 (55.6%) cases with early onset, with a mean of 36.6 years; and 20 (44.4%) cases of late onset with a mean of 57.4 years. Multiparity was high in 30/45 (66.6%). Six cases reported a family history of BC (13.3%), abortion was detected in 10 cases (22.2%). Of the presented histotypes that were available to us (16 cases unknown), ductal type tumor was more frequent and shown in 22 cases (48.8%), then lobular in five cases and mucinous in two cases. In addition, five cases showed distal cancer metastasis; lung and bone were frequent secondary sites. Right side tumor originated in 20 cases (44.4%), while the left side tumor was found in 15 cases (33.3%), and four cases presented with bilateral BC (6 cases unknown) (see Table 1).

**Ethical statement**

This study was conducted under the guidelines and approval of the Research Ethics Committee of Sudan Ministry of Health – Khartoum state. All participants provided oral informed consent to participate in the study. Oral informed consent was obtained as opposed to written consent, due to the literacy levels of the patients and the time limited interaction between researchers and patients at the hospital.

**DNA extraction and PCR amplicons**

DNA was extracted using the salting-out method for 45 patients samples. In addition, proteinase K was used to enhance WBC membrane breakdown at 56°C for 1 hour. For PCR, three previously published pairs of primers were used to amplify three regions within the \textit{BRCA1} gene. All the three primers were selected for their quality performance, optimal size and GC content, after being assessed with Oligoanalyzer tool 3.1.Annealing temperatures were adjusted on several runs (Table 2). Maxime PCR PreMix Kit i-Taq 20 μl (INTRON Biotechnology, South Korea) was used for PCR - 15 μl distilled water, 3 μl sample DNA (30 ng/μl; as checked by NanoDrop 1000) and 1 μl of the final concentration of each primer (10 pmol/μl forward and reverse). PCR mixture was subjected to initial denaturation step at 96°C for 5 minutes; followed by 35 cycles of denaturation at 96°C for 30 seconds, primer annealing at

\[50 \text{ or } 55°C \text{ depending on the set used, for } 30 \text{ seconds; followed by a step of elongation at } 72°C \text{ for } 60 \text{ seconds; the final elongation was at } 72°C \text{ for } 10 \text{ minutes.} \]

After PCR amplification, the PCR products (442, 271 and 401bp) were checked by 2% gel electrophoresis at 100 V for 30–45 min (Figure 1).

The product size of the first and last primers was checked and assessed using Serial Cloner version 2.6.1 (http://serialbasics.free.fr/Serail_Cloner.html) on the known nucleotide database accession gene for \textit{BRCA1} (NG_005905) with the whole sequence size of 81189bp, both forward and reverse of each one have been found to determine regions that cover coding and non-coding sequences.

**Reference sequence NCBI nucleotide database**

The NBPI RefSeqGene NG_005905, which represents the whole \textit{BRCA1} gene, and the transcript variant 1 NM_007294 mRNA, which comprises the coding DNA reference sequence of the gene (cDNA) with 7224 bp that contains the (5'-UTR and 3'-UTR). Untranslated region sequences, and among the sequence bp, the A nucleotide of the ATG initiation codon has been identified to start at 233 and stop at 5824 to determine the coding sequence (CDS) of 5592 bp, which is commonly used in the mutation nomenclature that is supported by the Human Genome Variation Society (HGVS).

The first sequence NG_005905 was used to check that all three

<table>
<thead>
<tr>
<th>Variable</th>
<th>Frequency, n (%) (n=45)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Onset</strong></td>
<td></td>
</tr>
<tr>
<td>Early (≤45 years)</td>
<td>25 (55.6)</td>
</tr>
<tr>
<td>Late (&gt;45 years)</td>
<td>20 (44.4)</td>
</tr>
<tr>
<td><strong>Family history</strong></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>6 (13.3)</td>
</tr>
<tr>
<td>Other cancer</td>
<td>5 (11.3)</td>
</tr>
<tr>
<td>No family history of any cancer</td>
<td>34 (75.6)</td>
</tr>
<tr>
<td><strong>Parturition</strong></td>
<td></td>
</tr>
<tr>
<td>Multiparous</td>
<td>30 (66.7)</td>
</tr>
<tr>
<td>Nulliparous</td>
<td>13 (28.9)</td>
</tr>
<tr>
<td>Primiparous</td>
<td>2 (4.4)</td>
</tr>
<tr>
<td><strong>History of abortion</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>10 (22.2)</td>
</tr>
<tr>
<td>No</td>
<td>35 (77.8)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
</tr>
<tr>
<td>Currently married</td>
<td>41 (91.1)</td>
</tr>
<tr>
<td>Single</td>
<td>3 (6.7)</td>
</tr>
<tr>
<td>Previously married</td>
<td>1 (2.2)</td>
</tr>
<tr>
<td><strong>Tribe</strong></td>
<td></td>
</tr>
<tr>
<td>Ja’alya</td>
<td>5 (11.1)</td>
</tr>
<tr>
<td>Shaygeya</td>
<td>5 (11.1)</td>
</tr>
<tr>
<td>Dnagla</td>
<td>4 (8.9)</td>
</tr>
<tr>
<td>Noba</td>
<td>3 (6.7)</td>
</tr>
<tr>
<td>Rezaigat</td>
<td>3 (6.7)</td>
</tr>
<tr>
<td>Other</td>
<td>25 (55.5)</td>
</tr>
<tr>
<td><strong>Geographical region</strong></td>
<td></td>
</tr>
<tr>
<td>Central Sudan*</td>
<td>21 (46.7)</td>
</tr>
<tr>
<td>Western Sudan</td>
<td>15 (33.3)</td>
</tr>
<tr>
<td>Northern Sudan</td>
<td>6 (13.3)</td>
</tr>
<tr>
<td>Eastern Sudan</td>
<td>3 (6.7)</td>
</tr>
<tr>
<td><strong>Tumor site</strong></td>
<td></td>
</tr>
<tr>
<td>Unilateral</td>
<td>35 (77.8)</td>
</tr>
<tr>
<td>Bilateral</td>
<td>4 (8.9)</td>
</tr>
<tr>
<td>Unknown</td>
<td>6 (13.3)</td>
</tr>
</tbody>
</table>

Table 1. Patient demographics and characteristics. *Comprising both Khartoum 16 cases and AlGezirah 5 cases.
Table 2. *BRCA1* sets of primers used for PCR. **F**: forward; **R**: reverse; **bp**: base pair; **CDS**: coding sequence; **V1**: variant one.

| PRIMERS 5’→3’ | EXON | ANNEALING Temp. (°C) | PRIMER SIZE (bp) | PRODUCT SIZE (bp) | PRODUCT REGION |
|--------------|------|----------------------|------------------|-------------------|----------------|----------------|
| Primer 1     | F: 5’CCA AGG TGT ATG AAG TAT GT’3 | 11 | 50 | 20 | 442 (NG_005905: 81189bp) | Only 367bp CDS (V1 NM_007294) |
|              | R: 5’GAT CAG CAT TCA GAT CTA CC’3 | 11 | 50 | 20 | 442 (NG_005905: 81189bp) | Only 367bp CDS (V1 NM_007294) |
| Primer 2     | F: 5’CTC ACT AAA GAC AGA ATG’3 | 11 | 50 | 18 | 271 | 271bp CDS (V1 NM_007294) |
|              | R: 5’CTT TCT GAA TGC TGC TAT’3 | 11 | 50 | 18 | 271 | 271bp CDS (V1 NM_007294) |
| Primer 3     | F: 5’ATA TGA CTT CCG TGC TCC AC’3 | 20 | 55 | 20 | 401 (NG_005905: 81189bp) | Only 86bp CDS (V1 NM_007294) |
|              | R: 5’GGG AAT CCA AAT TAC ACA GC’3 | 20 | 55 | 20 | 401 (NG_005905: 81189bp) | Only 86bp CDS (V1 NM_007294) |

**Figure 1.** PCR amplification of *BRCA1* gene. Left-side, PCR amplicons of primer 2, size 271 bp; right-side, PCR amplicons of primer 3, size 401 bp; L, ladder of 100bp each.

primers amplicons within the *BRCA1* gene region, while the second sequence NM_007294 was used for assessing all the three primers amplicons within the *BRCA1* coding sequence. The gene sequence of *BRCA1* has marked all the primers set amplicons to be within the gene region sequence. The transcript variant 1 (NM_007294) mRNA has marked (only 367 and 86bp) nucleotide sequences within primers sets 1 and 3 amplicons, respectively, to be within coding sequences, and the whole set of primer 2 amplicon (271bp) was within the coding sequence of the *BRCA1* gene (Table 2).

**Sequencing of *BRCA1* gene**

The PCR products of the 10 best bands yielded from the patient samples for each primer set, a total of 30 resulted amplicons, were sent for Sanger dideoxy sequencing. Partial standard sequencing for the three regions within the gene, including both forward and reverse nucleotide sequencing, was performed by Macrogen Company (Seoul, South Korea), using the same pairs of primers.

**Computational analysis**

**Sequence analysis.** The sequence results for the 30 sequence chromatogram files were viewed by FinchTV program version 1.4.0\(^6\), which was used to check both nucleotide sequences of the patients forward and reverse sequences to be free of errors. Any errors were excluded during processing. The Basic Local Alignment Search Tool (BLAST; https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to assess nucleotide and protein sequence similarities\(^6\). In ExPASy translate tool - SIB Bioinformatics Resource Portal, the gene sequences were translated into amino acid sequences\(^5\). For primers 2 and 3, the *BRCA1* nucleotide sequences from the patients, with their translated proteins, underwent multiple sequence alignment using BioEdit software version 7.0.9.0\(^6\). Multiple sequence alignment included the reference sequence with the highest similarity, as obtained by BLAST (RefSeq transcript mRNA - NM_007294 transcript variant 1), two additional nucleotide sequences (NM_007297,3; transcript variant 3 mRNA and JN686490.1; Figure 2A), and the gene sequence NG_005905, which is mainly the sequence between positions 68120-68810 (Figure 2B).

**SNP information.** SNP information [SNP ID, MIM: 113705, RefSeq Gene accession No.: NG_005905 on chromosome 17, mRNA accession NO.: NM_007294 transcript variant 1 with 7224 bp and Protein accession numbers : (NP_009225) protein isoform 1 with 1863 a.a. and P38398 UniProt entries] concerning the human *BRCA1* gene, which was used in our computational analysis, was retrieved from the NCBI database of SNPs: dbSNP (https://www.ncbi.nlm.nih.gov/snp).

**SNP prediction.** SNPs were analyzed using five prediction online tools: SIFT (http://sift.bii.a-star.edu.sg/)\(^7\), Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/)\(^8\), I-MutantDDG-Seq Suite (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) and PhD-SNP (http://snpdb.biofold.org/phd-snp/phd-snp.html)\(^9\). The tertiary model of protein and mutation analysis was done online using Hope software (http://www.cmbi.ru.nl/hope/input)\(^10\). MutationTaster2 (http://www.mutationtaster.org) was used to assess the protein features of the resulting variants, with comprehensive (input/output) criteria, which predicts potential disease-causing mutations\(^4\).

**Results**

**Patient sequences**

Two missense variants were detected within the study, one patient with Q356R and three patients with V1736D. Both variants were heterozygous (Figure 2) and were detected within premenopausal patients, with a mean age of 37 years. Three patients were multiparous; the one case of (Q356R) and two cases of (V1736D) were multiparous (mean parity, 2.8). There was no family history of BC in patients with the two variants (Table 3).
Bioinformatics assessment

The patients’ sequences, and the additional nucleotide sequences with the gene sequence, were aligned against the reference standard sequence in the RefSeqGene and nucleotide databases (accession, NM_007294), which is BRCA1 gene transcript variant 1, and was introduced to replace the previously existing sequence (gi: 6325287) in May 2009. Two substitutions-bearing monoallelic alterations were found, the first one at position 1299 (A/G) located in exon 11 (Figure 2), and the second one at position 5439 (T/A) in exon 20 (Figure 2). In conformity with the HGVS nomenclature, the sequence variations 1299 (A/G) and 5439 (T/A) met by (n.1299A>G, c.1067A>G) and (n.5439T>A, c.5207T>A) respectively. The (n) describes the noncoding which highlights the position within the complete transcript that possesses the UTRs where the numbering includes the 5’-UTR. The (c) describes mainly the alteration position within the derived coding sequence that does not possess the UTRs where the numbering starts from the initiation codon ATG. Primer 1 has been excluded from the study because of the errors that have been noticed within all patient sequence data chromatogram results.

After translation to amino acid sequences, the samples were aligned against BRCA1 protein isoform 1 (accession, NP_009225). Q356R was found to meet its corresponding nucleotide change and position n.1299A>G in which Glutamine (Gln) replaced by Arginine (Arg) in patient 23, and V1736D was found to meet its corresponding nucleotide change and position n.5439T>A, in which Valine (Val) was replaced by Aspartic acid (Asp) in patients 2, 22 and 26 (Figure 2). These variants (Q356R, V1736D) were then predicted with SIFT, Polyphen-2, I-Mutant-3, PhD-SNP and Mutation-Taster2 software to obtain their pathological effects, and results are provided in Table 4. Amino acid properties for the wild Val and the mutant Asp residues and the 3D structure of the variant V1736D were obtained using Project Hope software (Figure 3A).

Table 3. Characteristics and clinical data of the patients with the BRCA1 variants.

<table>
<thead>
<tr>
<th>Patient no.</th>
<th>Age (years)</th>
<th>Parturition</th>
<th>Tribe/Residency</th>
<th>Mutation</th>
<th>Grading/Staging</th>
<th>Breast cancer type</th>
<th>Family history</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-23</td>
<td>27</td>
<td>Multiparous – 3</td>
<td>Ja’alya/Central-AlGezeerah</td>
<td>Q356R</td>
<td>T,N,M</td>
<td>Unilateral-Right</td>
<td>None</td>
</tr>
<tr>
<td>3-2</td>
<td>45</td>
<td>Multiparous – 4</td>
<td>Noba/Western-Kadoggly</td>
<td>V1736D</td>
<td>T,N,M</td>
<td>Bilateral</td>
<td>None</td>
</tr>
<tr>
<td>3-22</td>
<td>40</td>
<td>Multiparous – 4</td>
<td>Kawahla/Central-JabalAwlya’a</td>
<td>V1736D</td>
<td>G III</td>
<td>Unilateral-Right</td>
<td>None</td>
</tr>
<tr>
<td>3-26</td>
<td>37</td>
<td>Nulliparous–Single</td>
<td>Mahas/Central-Khartoum</td>
<td>V1736D</td>
<td>T,N,M</td>
<td>Unilateral-Right</td>
<td>None</td>
</tr>
</tbody>
</table>
Table 4. SNP prediction obtained by various sequencing softwares. SNP: single nucleotide polymorphism; RI: reliability; DDG: ΔΔG; SVM: support vector; SVM2 value: DDG < 0: decrease stability, DDG >0 increase stability machine; DDG value: DG (New Protein)-D (Wild Type) in Kcal/mol.

<table>
<thead>
<tr>
<th>SNP</th>
<th>SIFT Score</th>
<th>Polyphen-2 Prediction</th>
<th>I-Mutant SVM2 Prediction</th>
<th>I-Mutant RI</th>
<th>SVM2 DDG value Kcal/mol</th>
<th>PHD-SNP Prediction</th>
<th>PHD-SNP RI</th>
<th>MutationTaster2 Score</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q356R</td>
<td>0.01</td>
<td>Affect protein function</td>
<td>0.998</td>
<td>Decrease protein stability</td>
<td>1</td>
<td>-0.08</td>
<td>Disease-related polymorphism</td>
<td>4</td>
<td>43</td>
</tr>
<tr>
<td>V1736D</td>
<td>0.00</td>
<td>Affect protein function</td>
<td>0.984</td>
<td>Decrease protein stability</td>
<td>8</td>
<td>-1.82</td>
<td>Disease-related polymorphism</td>
<td>8</td>
<td>152</td>
</tr>
</tbody>
</table>

Figure 3. 3D structure of mutant BRCA1 protein due to nonsense mutation at position 1736 and the structural unit configurations of both wild and mutant amino acid residues of the Q356R variant. (A) This illustration shows the BRCA1 protein in grey colour on the (left) with the magenta coloured side chain of the mutated residue, and on the (right) the higher magnification of the area of mutation displays the side chains of both amino acid residues more clearly, the green highlights the wild type Valine, and the red highlights the mutant type Aspartic acid with their schematic structures, as both contain the same backbone (red) but they differ in their side chains (black). (B) The illustration shows the schematic structural units of the neutral wild type Glutamine (left) and the positive mutant Arginine (right). Both amino acids contain the same backbone (Red) but they differ in their side chains (Black).

Discussion
In this study, we found two mutations-bearing monoallelic features. One missense mutation rs1799950 in exon 11 BRCA1 gene in patient 23: position number 1299 (A/G) as (CAG A) to (CGG A), using transcript variant 1 (NM_007294), n.1299A>G, which led to change in the coding amino acid (Glutamine-Gln-Q) to (Arginine-Arg-R) at position 356 (Figure 2A).

The other was a novel SNP, situated at exon 20, with higher frequency in patients 2, 22 and 26: position 5439 (T/A) as (GTC A) to (GAC A), n.5439A>G, which led to change in the coding amino acid (Valine-Val-V) to (Aspartic acid-Asp-D) at position 1736 (Figure 2B). Both SNPs (Q356R) and (V1736D) were found to affect the resulting translated protein. According to Project Hope software, starting with the novel (V1736D) variant, we found that the mutant residue Aspartic acid (D) of the outcome
**BRCA1** mutated protein is located near a highly conserved position. The mutant residue is bigger and less hydrophobic than the wild-type residue. The wild-type residue was neutral, the mutant residue is negatively charged. The mutated residue is located in a domain that is import for binding of other molecules, as it is in contact with residues in another domain. It is possible that the mutation can disturb these contacts. It may also disturb the interaction between these two domains, which would affect the function of the protein; therefore, it might disturb signal transfer from binding domain to the active domain. This novel mutation was detected in patients from different tribes from Sudan, who were all ≤45 years old.

In the case of the Q356R mutation, the mutant residue of Arg is positively charged compared to the neutral wild type Gln residue. This could lead to the repulsion of ligands or other similar charged residues. The size of the end variant is bigger than the wild-type residue and this might lead to bumps, as reported by Project Hope software (Figure 3B). In addition, the mutated residue is located on a domain responsible for protein activity; hence the activity could be altered by its physical variation conferred by its new charge and size. The Arg 356 variant does, however, generate a run of three positively charged residues (Lys Arg Lys), and as a result it could alter the properties of the protein, which is composed of 16% negatively charged residues overall⁷. In the current study, we found the same result when it was described as a disease-related mutation of altered protein stability (Table 4).

According to MutationTaster2, the novel variant (V1736D) was predicted to be ‘disease-causing’, and the software assessed and calculated the probability of such variation on the resulting protein, and showed that the new features would be disease-causing and displayed it with a higher score. By contrast, the software predicted that in the case of Q356R, the resulting Arg was a polymorphism of little harm, and reported it with a lower score compared to the protein feature of the other variant, and reports from the other softwares used for assessment (Table 4). Q356R has not yet been classified in terms of clinical significance in NCBI dbSNP.

Two novel variants were identified to be deleterious within the **BRCA1** gene among premenopausal women patients in two local studies⁸⁹. One variant was a truncated stop codon and the other was predicted computationally: c.4986+6T>C, located in intron-exon boundary was found in the youngest patient of 25 years in one of the studies⁸⁹. These findings, compared to our present study finding (V1736D), showed that early onset BC is associated with a deleterious nature of the identified variants. In addition, our study included pre and postmenopausal patients, all detected variants were mainly confined to premenopausal cases. Therefore, genetic studies are highly recommended to highlight the genetically susceptible nature of patients diagnosed with early onset disease, who may harbor the deleterious variants that could have a significant role in developing BC. Two patients within the present study were screened previously in a local study targeting pathological SNPs within **BRCA2** gene selected regions, which identified a stop codon (L1053X)⁹⁰. One of these patients was the youngest patient identified with Q356R, and the other presented with the bilateral disease and identified with V1736D. Both patients were reported to have stop codons at position L1053X with nucleotide and protein sequences identifiers of KT901810 and ALQ44030, and KT901807 and ALQ44027, respectively⁹⁰.

The SNP found in the present study, n.1299A>G, has been highlighted within matched reports as c.1067A>G⁹¹ and also with similar altered nucleotide position 1299⁹², and the same altered outcome protein position, Q356R⁹⁴,⁹⁵. The same protein position (Q356R)⁹⁴ with the same altered nucleotide (A/G) has also been described previously, but with a different nucleotide position 1186 (A/G) which is being more consistent with the nucleotide sequence transcript accession U14680.1⁹⁶,⁹⁷. Some studies found that this SNP is associated with patients under 40 years old⁹⁶,⁹⁷, which agreed with our study - the youngest premenopausal patient (27 years old) was detected with this variant. Although this variant’s effect remains uncertain, a previous finding to our study has described the same mutation, Q356R in Moroccan BC patients, which was the first study that described this mutation in a North African population⁹⁷. They found that most previous studies that described this mutation were within western European populations⁹⁸ and there were no studies found in a North African population to confirm this, thus more research is needed to investigate this. Our study is closer to the Moroccan study regarding the early onset characteristic this variant has, revealed by the similarity in the background history of the patient with a Q356R polymorphism in our study and those detected in the Moroccan study, who also did not have a family history of BC⁹⁷. This variant has been reported to be independently minor or leads to a very slightly increased BC risk, but a risk that is cumulatively significant⁹⁷. In another study, this mutation was found in patients with a family history of ovarian cancer, suggesting that this variant may increase ovarian cancer risk.⁹⁶,⁹⁷ Both detected variants in the present study were identified to have a pathological effect, with the exception of the results from MutationTaster2 in the case of Q356R. Therefore, theoretically, the Q356R and the novel SNP variants have a prediction to pathologically affect protein function which is in need for further verifications in order to confirm their roles in enhancing the tumorigenesis of the disease practically.

In order to only have clear DNA sequence results for comparison with NCBI references, primer 1 results were excluded due to sequencing errors (see Supplementary File 1). In addition, three patient sequences of primer 2 set and one patient of primer 3 set have been excluded for the same sequence errors.

The limitation of this study was the small sample size and the functional assessment facilities available to assess the protein of monoallelic alteration for their pathological contribution to the disease. Financial constraints also limited the study. Therefore, we recommend further studies in a larger number of Sudanese patients to further explore these findings.

**Conclusions**

In the present study, Sudanese BC patients were investigated for **BRCA1** mutations. Two different mutations were found in young patients of ≤45 years old, with no family history of BC. To conclude, the study has highlighted a need for further research of
these mutations amongst a larger population (including patients and controls), so as to investigate the variants’ distribution through the population and their potential diagnostic value. This will aid the understanding of a variant’s frequency and clinical significance. In addition, both variants identified, require in vitro functional and protein level assessment.

Data availability
The BRCA1 sequence data of the novel variant (V1736D) from this study has been submitted to NCBI GenBank under the accession numbers and protein identifiers found in Table S1.

Dataset 1: BRCA1 sequence results in a zipped file. These sequencing results as received from Macrogen Company (Seoul, South Korea) comprise all the breast cancer patients in this study using the three sets of primers (1, 2 and 3). Each patient has the sequencing data in different file formats (a sequencing data file that needs to be viewed by a sequencing viewer software, i.e FinchTV; a PDF; and a FASTA format text document). doi, 10.5256/f1000research.11395.d172445

Dataset 2: Patient demographics (non-identifying) according to primer. Patient demographics, clinical data, and histological parameters with highlighted detected missense variants (primers 2 and 3) of each 10 selected subset patients. doi, 10.5256/f1000research.11395.d172446

Supplementary material

Click here to access the data.

Table S1: NCBI GenBank accession numbers for the novel BRCA1 sequence data from this study.

Click here to access the data.

References


Open Peer Review

Current Referee Status:  

Khalid Dafaallah Awadelkarim  
Department of Molecular Biology, National Cancer Institute (NCIUG), University of Gezira, Wad Madani, Sudan

This manuscript is emphasized in a good way. However, it will benefit from some more revisions.

Major points:
1. I recommend the usage of standard HGVS nomenclature for documenting the detected variants. This will resolve the issue of different nucleotide position for the Q356R variant. Moreover, the novel V1736D need to be re-checked for its novelty after applying the standard HGVS nomenclature.

2. The Q356R is widely regarded as polymorphism (Borg et al. (2010)\(^1\)), or at least of very minor risk. What the additional evidence provided by this paper giving it’s the evident lack of control population and functional study. I suggest designating this variant as polymorphism or at least as unclassified variant based on the bioinformatics tools. However, Hadjisavvas et al. (2002)\(^2\) suggested that S1512I in combination with Q356R may be disease-associated, since these two missense mutations are simultaneously present only in this family, and in none of the other samples from the breast cancer or control groups.

3. The V1736D variant need to be investigated more, at least in control populations. I seems to me it is very hard to say that a novel variant that is detected with a frequency of 3/10 (30%) is probably a pathogenetic mutation based only on bioinformatics studies. Thus verification on control populations will provide more information about this variant.

Minor points:
1. In many part of the manuscript there is many comments (The mutated residue is located in a domain), which domain and what is the predicted change on function?

2. The introductory part about SNPs page 3, third paragraph, second sentence “These” I think refer to BRCA1 mutations, whereas the paragraph about SNPs. This is mainly because frameshift mutations are not part of SNPs. Please resolve.

3. The PCR figure appears to have a smear. If no good image is available, this figure could be removed as DNA sequence chromatogram were provided.
References

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Referee Report 09 October 2017
doi:10.5256/f1000research.13766.r26244

Mahmoud Balbaa
Department of Biochemistry, Faculty of Science, Alexandria University, Alexandria, Egypt

Figure 1 is still in a bad quality and should be changed to show clear band with a bp scale. The band of housekeeping gene should be shown in the figure.

**Competing Interests:** No competing interests were disclosed.
Referee Expertise: Enzymes and cell signalling, protein expression, gene expression

I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.

---

Version 1

Referee Report 14 September 2017

doi:10.5256/f1000research.12300.r25416

Mahmoud Balbaa
Department of Biochemistry, Faculty of Science, Alexandria University, Alexandria, Egypt

BRCA1 was highly investigated around the whole world. The authors just applied the research of BRCA1 on Sudanese patients. Generally, the authors have obtained good results, but Figure 1 that shows the PCR amplification of the BRCA1 gene has a bad quality and should be changed to show clear band with a bp scale. The band of housekeeping gene e.g beta actin should be shown in the figure.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Referee Expertise: Enzymes and cell signalling, Protein expression, gene expression.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

---

Referee Report 04 September 2017
Habiba Chaabouni-Bouhamed 1,2
1 Laboratory of Human Genetics, Faculté de Médecine, University Tunis El Manar, Tunis, Tunisia
2 Centre Urbain Nord, Tunis, Tunisia

The authors report the molecular analysis of 2 exons (11 & 20) of BRCA1 in a Sudanese population. The paper is well written.

It gives a truncated information about BRCA1 mutations in a population but the work is correctly done; it has to be completed by analysing the whole gene; using HRM method could help them to reduce cost analysis.

One remark about reference 1& 2, I didn't find information about African statistics which are available in references 4 & 5.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Not applicable

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Genetics, molecular genetics, cytogenetics, genetic counseling

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com