Medical Oncology (2019) 36:71
https://doi.org/10.1007/s12032-019-1294-9

REVIEW ARTICLE

Functional coding and non-coding variants in human *BRCA1* gene and their use in genetic screening

Gulten Tuncel[1,2] · **Mahmut Çerkez Ergören**[1,2]

Received: 20 May 2019 / Accepted: 25 June 2019 / Published online: 3 July 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

BRCA1 is involved in double-strand DNA damage repair pathways, and mutations in the gene are associated with hereditary breast and ovarian cancers. With great help of the development of high-throughput DNA sequencing techniques numerous single-nucleotide polymorphisms (SNPs) and insertion deletion (Indel) mutations are detected on both coding and non-coding/regulatory regions of the *BRCA1*. Mutations may cause pathogenic or benign changes on the protein function or affect its expression. In the last decade, use of genetic screening tests to detect mutations on such genes has become greatly popular. However, it is very important to know the effect of the detected mutations, which is mostly possible by the use of predictive softwares, and also the related family history to be able to correctly analyse the screening results and to inform the patient. Therefore, use of in silico and in vitro techniques to score the pathogenicity of detected variants on genes like *BRCA1* is now of great importance. Otherwise, results obtained from screening tests and family history cannot be analysed precisely.

Keywords *BRCA1* · Variants · Genetic screening

Introduction

Genetic mutations are permanent alterations in the DNA sequence. Depending on their position, mutations can alter the function or amount of essential proteins in the body, causing various health effects. Gene mutations can be classified in a number of different ways but generally they are divided into two main groups as inherited (hereditary, germline) and somatic (acquired, sporadic) mutations. Hereditary mutations are found in the egg or sperm cells (germ cells) of the parents, which make up the zygote that forms the offspring [1]. So, the mutation is inherited to the offspring and is expected to present in all cells of the body. Somatic mutations, on the other hand, are acquired during an individual's lifetime and are present only in one or a small group of cells in the body [2]. More importantly, they are not passed to the next generation.

BRCA1 is a well-studied tumour suppressor gene encoding breast cancer type 1 susceptibility protein, which is involved in DNA damage repair pathways and is usually mutated in inherited breast and ovarian cancers [3]. The gene is located on the long arm of chromosome 17 and is composed of 24 exons spread over 80 kb [4]. To date, more than 1600 mutations are identified in the *BRCA1* gene that include deletions, insertions, missense, and nonsense mutations [4]. Generally, individuals inherit a *BRCA1* mutation from a parent and then a somatic mutation occurs on the wild-type allele leading to cancer phenotype. Studies suggest that the risk of developing breast cancer increases from 53 to 73% in women who have an oncogenic mutation in *BRCA1* and are having a familial history [5]. Around 60% of *BRCA1* mutation carriers develop breast cancer by age 70 [6].

Among other relatively large-scale mutations, single-nucleotide polymorphisms (SNPs) are single-base changes that occur normally with frequency of 0.1% throughout a person's DNA and with allele frequency of > 1% among a population [7]. Most of them do not have any effect on human health; however, some of these variations have important roles in genetic diversity, response to drugs, evolution, and susceptibility to some diseases including cancer

✉ Mahmut Çerkez Ergören
mahmutcerkez.ergoren@neu.edu.tr

Gulten Tuncel
20173423@std.neu.edu.tr

[1] Faculty of Medicine, Department of Medical Biology, Near East University, 99138 Nicosia, Cyprus

[2] Experimental Health Sciences Research Center (DESAM), Near East University, Nicosia, Cyprus

[7, 8]. So, identification and scoring of SNPs according to their pathogenicity and association with diseases are important to help develop new clinical markers as well as personalised treatment strategies.

Since the discovery of the huge variation in the human genome, it has been a challenge to identify SNPs that might be responsible for phenotype changes and diseases. Many bioinformatics prediction tools have been introduced to fulfil this aim. Database of SNP (dbSNP) in NCBI is a central and public database that contains detailed information about genetic variations obtained from the International Human Haplotype (HapMap) and 1000 Genomes Projects [9].

In hereditary breast cancers, identification of *BRCA1* and *BRCA2* mutations is important for better diagnosis and to better tailor the clinical management of patients in the clinic. In this review article, functional SNPs and small-scale multibase deletions or insertions (indels) within coding regions, 3′/5′ UTR sites and 3′/5′ splice sites of the human *BRCA1* gene, are examined and evaluated.

SNPs and indels in non-coding regions of the human *BRCA1* gene

Information obtained from the dbSNP shows that there are a total of 4225 indel and 18,557 SNPs within the human *BRCA1* gene. These variations include 3′/5′ splice site and 3′/5′ UTR mutations, frameshift, synonymous, missense, and nonsense mutations in the coding regions (Table 1).

There are 15 indel and 42 single-nucleotide variations (total 57) within the 3′ splice site, 26 indel, and 41 SNPs (total 67) in the 5′ splice sites. Total numbers of variations in the 3′ and 5′ splice sites are similar. These variations are important as they can potentially generate new splice sites, have an effect on exon–intron junctions, and induce cryptic sites [10]. For example, a splicing variant (c.5406 + 2T>C) in the *BRCA1* was shown to cause the loss of the canonical donor splice site and create a new splice site that causes the retention of 156 bp of intron 21 in a family suffering from hereditary breast cancer [11]. Such variations are classified as pathologic variants in the databases after confirmation of bioinformatics predictions and functional splicing assays.

Table 1 Distributions of SNP and Indel variations in the coding regions, UTR, and splice sites of the human *BRCA1* gene

Human *BRCA1*	
Indels	Total: 4225
3′ splice site	15
5′ splice site	26
3′ UTR	86
5′ UTR	51
Frameshift	1696
SNPs	Total: 18,557
3′ splice site	42
5′ splice site	41
3′ UTR	424
5′ UTR	128
Coding synonymous	697
Missense	1891
Nonsense	352

86 indel and 424 SNPs (total 510) are present in 3′ UTR site of the *BRCA1* gene. Variations in the 3′ UTR sites of genes play crucial roles in transcriptional regulation as these regions harbour binding sites for miRNAs and mRNA-binding proteins affecting the stability of the transcript. Alterations in the 3′ UTR sites in several genes have previously been associated with increased risk of tumour formation and progression [12–14]. So, it is important to analyse and experimentally test the 3′ UTR variations in the human *BRCA1* gene to detect pathogenic and likely pathogenic variants [15]. For example, a 3′ UTR variant (rs8176318 G>T) was shown to cause decreased *BRCA1* expression and result in increased breast cancer risk [16]. c.*1340_1342delTGT, on the other hand, was shown to introduce a functional miR-103 target site and reduce the stability of the *BRCA1* transcript and might be therefore pathogenic [17]. Likely, several other variants were identified to somehow reduce 3′ UTR activity of the human *BRCA1*, being potential pathogenic variants [17].

In addition to 3′ UTR mutations, some variants are also present in the mRNA 5′ UTR regions. To date, 51 indel and 128 SNP (total 179) variants are identified in the 5′ UTR of human *BRCA1* according to dbSNP data. 5′ UTR mRNA 5′ UTRs contain functional elements such as upstream open reading frames (uORFs), internal ribosome entry site (IRES), and iron-responsive elements (IRE), which are crucial in translational regulation. Variations in the 5′ UTR regions have previously been associated with diseases including hereditary cancers [8, 18]. Recently, for example, a variant that is inherited in a dominant manner in the 5′ UTR of human *BRCA1* (c.-107A>T) was shown to cause methylation-associated silencing and cause breast and ovarian cancers [19].

SNPs and indels in coding regions of the human *BRCA1* gene

The remaining mutations that will be mentioned in this section are all within the coding region of the human *BRCA1* gene. 1696 indels causing frameshift mutations, 352 nonsense, and 1891 missense SNPs have been detected in the coding region of the *BRCA1*. Other 697 SNPs were synonymous mutations, which do not alter the protein sequence.

Like the mutations in the regulatory regions, coding sequence polymorphisms in *BRCA1* also have pathogenic effects leading to increased susceptibility to hereditary breast and ovarian cancers. When we analysed the missense SNPs (total 1891) in human *BRCA1* using the SIFT (sorting intolerant from tolerant) [20], which is a bioinformatics server that predicts the damaging effect of small mutations on protein function, we found out that 3555 of total 7987 possible variants are predicted to be "deleterious". This means that the mutation results in the generation of a protein product that is different from the wild type or protein translation is interrupted due to the formation of a premature stop codon. Other 4432 variants are either tolerated or no information is yet available about their effect on protein function.

Nonsense SNPs (total 352) in human *BRCA1* were analysed for their potential effects in protein level using the SIFT software. 122 out of total 2310 possible variants are predicted to be deleterious and 155 as tolerated. No information was found to be available about the other variants.

A total of 1696 insertion/deletion (indel) mutations resulting in frameshifts within the coding region of *BRCA1* are listed in the dbSNP database. Among these, 1515 are classified as pathogenic and an additional 12 as likely pathogenic. Therefore, not surprisingly, a total of 1527 out of 1696 indels within the *BRCA1* gene cause severe changes in the protein structure resulting in possible deleterious effects. For example, the rs80357697 variant, which was identified in four unrelated women suffering from hereditary breast cancer from Kurdish Jewish descent, is a result of a deletion (c.224_227delAAAG, p.Glu75fs) causing a frameshift in *BRCA1* [21].

In this context, it is important to note that defining the role and scoring the pathogenicity of a variant is an intense bioinformatics and laboratory work, which demand intensive labour and financial sources. Considering the huge amount of data being obtained from the next-generation sequencing (NGS) technology, it is not surprising that an important percentage of detected variants are still of unknown significance (VUS) [15].

Discussion

Detecting variations among the human genome has become an easy procedure after the introduction of next-generation DNA sequencing technology and bioinformatics tools over the last decade. However, detection is only the first step and the main challenge has become to score the detected variants to reveal their significance.

First genetic tests for detecting known mutations in monogenic disorders such as sickle cell anaemia and Tay–Sachs disorder was introduced as routine tests in different regions starting from 1970s [22, 23]. As the genotype–phenotype associations of such diseases are well defined, genetic tests are diagnostic and give certain results about the condition of the tested individual. Also, in countries including Cyprus, where Thalassemia is a common monogenic disorder, premarital screening for the disease was made compulsory in 1980s and preventive strategies including prenatal genetic testing in 1990s resulted in a sharp decrease in the number of affected birth rates [24]. Then, the rapid development in genetic analysis technologies and data obtained from projects like 1000 Genomes project and HapMap project resulted in the emergence of various genetic screening kits that are marketed in dietary clinics, some pharmacies, and even in online platforms. Such screens are designed to detect any variation in tested genes and provide a risk score for the individual to develop an associated disease. Tests designed for *BRCA1* and *BRCA2* are being used widely to detect the presence of any mutations that may result in increased risk for hereditary breast and ovarian cancers [25]. In complex diseases, like cancers, the genotype–phenotype correlation is not as clear as it is in monogenic disorders. Analysis and interpretation of genetic screen/test results must be carried out professionally, taking the patient's medical and family history into account. As noted above, an important percentage of *BRCA1* variants are still VUS. Additionally, these tests screen only for known mutations on the *BRCA* genes. Patients may carry a previously unidentified single variation, which may cause a false-negative result.

Conclusion

Having a variant in the *BRCA* genes, if inherited from a previously affected parent increases the risk score of the tested individual, as well as detecting a variant with a known pathogenic significance reported in the databases. During variant annotation, as bioinformatics tools work on the basis of predictions, it is important to use either more than two prediction tools or design functional assays to confirm the pathogenic effect of the variant in silico or in vitro. Such as, applications such as 'BRCA exchange' will be significant to improve our understanding of the molecular genetic basis of breast cancer and other diseases by collecting data on the *BRCA* genes variants [26]. Uncovering the effects of SNPs in the human genome is of high importance, as detection of pathogenic variants will help clinicians in diagnosis, preventive treatment strategies, and to enable personalised medicine for patients.

Compliance with ethical standards

Conflict of interest All authors certify that they have no affiliations with or involvement in any organisation or entity with any financial

or non-financial interest in the subject matter or materials discussed in this manuscript.

References

1. Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. Trends Genet. 2013;29(10):575–84.
2. Griffiths AJ, et al. Somatic versus germinal mutation. An introduction to genetic analysis. Stuttgart: W. H. Freeman; 2000.
3. Wu J, Lu LY, Yu X. The role of BRCA1 in DNA damage response. Protein Cell. 2010;1(2):117–23.
4. Petrucelli N, Daly MB, Feldman GL. Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. Genet Med. 2010;12(5):245–59.
5. Lippi G, Mattiuzzi C, Montagnana M. BRCA population screening for predicting breast cancer: for or against? Ann Transl Med. 2017;5:13.
6. Godet I, Gilkes DM. BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. Integr Cancer Sci Ther. 2017. https://doi.org/10.15761/ICST.1000228.
7. Hassan MM, et al. Bioinformatics approach for prediction of functional coding/noncoding simple polymorphisms (SNPs/indels) in human BRAF gene. Adv Bioinf. 2016. https://doi.org/10.1155/2016/2632917.
8. Deng N, et al. Single nucleotide polymorphisms and cancer susceptibility. Oncotarget. 2017;8(66):110635–49.
9. Smigielski EM, et al. dbSNP: a database of single nucleotide polymorphisms. Nucleic Acids Res. 2000;28(1):352–5.
10. Thery JC, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. Eur J Hum Genet. 2011;19(10):1052–8.
11. Esposito MV, et al. A novel pathogenic BRCA1 splicing variant produces partial intron retention in the mature messenger RNA. Int J Mol Sci. 2016;17(12):2145.
12. López de Silanes I, Paz Quesada M, Esteller M. Aberrant regulation of messenger RNA 3 and M. Esteller, r Intron Retention. Cell Oncol. 2007;29(1):1–17.
13. Singh P, et al. Global changes in processing of mRNA 3′ untranslated regions characterize clinically distinct cancer subtypes. Cancer Res. 2009;69(24):9422–30.
14. Mao G, Pan X, Gu L. Evidence that a mutation in the MLH1 3′-untranslated region confers a mutator phenotype and mismatch repair deficiency in patients with relapsed leukemia. J Biol Chem. 2008;283(6):3211–6.
15. Karbassi I, et al. A standardized DNA variant scoring system for pathogenicity assessments in mendelian disorders. Hum Mutat. 2016;37(1):127–34.
16. Dorairaj JJ, et al. A germline mutation in the BRCA1 3′ UTR predicts stage IV breast cancer. BMC Cancer. 2014;14:421.
17. Brewster BL, et al. Identification of fifteen novel germline variants in the BRCA1 3′ UTR reveals a variant in a breast cancer case that introduces a functional miR-103 target site. Hum Mutat. 2012;33(12):1665–75.
18. Gochhait S, et al. Implication of BRCA2-26G>A 5′ untranslated region polymorphism in susceptibility to sporadic breast cancer and its modulation by p53 codon 72 Arg > Pro polymorphism. Breast Cancer Res. 2007;9(5):R71.
19. Evans DGR, et al. A dominantly inherited 5′ UTR variant causing methylation-associated silencing of BRCA1 as a cause of breast and ovarian cancer. Am J Hum Genet. 2018;103(2):213–20.
20. Sim NL, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:W452–7.
21. Zick A, et al. A BRCA1 frame shift mutation in women of Kurdish Jewish descent. Open Med J. 2015;2:31–6.
22. Benson JM, Therrell BL Jr. History and current status of newborn screening for hemoglobinopathies. Semin Perinatol. 2010;34(2):134–44.
23. Kaplan F. Tay–Sachs disease carrier screening: a model for prevention of genetic disease. Genet Test. 1998;2(4):271–92.
24. Bozkurt G. Results from the north cyprus thalassemia prevention program. Hemoglobin. 2007;31(2):257–64.
25. Lynch J, Venne V, Berse B. Genetic tests to identify risk for breast cancer. Semin Oncol Nurs. 2015;31(2):100–7.
26. Cline MS, et al. BRCA challenge: BRCA exchange as a global resource for variants in BRCA1 and BRCA2. PLoS Genet. 2018;14(12):e1007752.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.