



In silico analysis of potential structural and functional significance of human breast cancer gene BRCA2 sequence variants found in 5' untranslated region

PETAR OZRETIĆ¹
MIRELA LEVAČIĆ CVOK²
VESNA MUSANI¹
MAJA SABOL¹
DIANA CAR¹
SONJA LEVANAT¹

¹Laboratory for Hereditary Cancer,
Division of Molecular Medicine
Ruđer Bošković Institute,
HR-10000 Zagreb, Croatia

²Ruder-Medikol diagnostics Ltd,
Bijenička 54, HR-10000 Zagreb, Croatia

Correspondence:

Petar Ozretić
Laboratory for Hereditary Cancer,
Division of Molecular Medicine
Ruđer Bošković Institute,
HR-10000 Zagreb, Croatia
E-mail: pozretic@irb.hr

Key words: breast cancer, BRCA2, mRNA,
5'UTR, secondary structure

Abstract

Background and Purpose: BRCA1 and BRCA2 are major hereditary breast/ovarian cancer predisposing genes and their mutations increase the risk of developing cancer. Genetic testing of these two genes is nowadays commonly performed but almost half of found genetics alterations are declared as variants of unknown clinical significance. Interpretation of these unclassified variants is the major concern for BRCA genes. The aim of this study is to investigate potential structural and functional significance of sequence variants found in 5' untranslated region (UTR) of BRCA2 gene.

Materials and Methods: Consensus secondary structure of BRCA2 5' UTR was built based on nucleotide sequences from four different species. We collected all found human BRCA2 5' UTR variants and explored their potentials effects by folding human BRCA2 5' UTR including one of each variant, using consensus structure as a constraint. If constrained folding results in a structure that is very different from the consensus one, this may indicate that this particular sequence variant could have potential functional impact.

Results: Most of the sequence alterations are found near the 3' end of 5' UTR, what is in the vicinity of the translation initiation site. Four of them: c.-26G>A, c.-26G>C, c.-26G>T and c.-12T>C most notably disturbed consensus secondary structure by creating substructures with lower minimum free energy, thus less stable.

Conclusions: As previously deduced in the case of variant c.-26G>A, changes c.-26G>C, c.-26G>T and c.-12T>C could unstabilize the loop at the vicinity of the translation start site, which could increase the efficiency of the translation and thereby increase the expression of BRCA2. Accordingly, our study suggests this three BRCA2 5' UTR sequence variants as suitable candidates for further functional characterization and thus potentially clinically significant.

INTRODUCTION

BRCA1 (breast cancer 1, early onset) and BRCA2 (breast cancer 2, early onset) genes, located on chromosome 17q12-21 and 13q12-3, respectively, are the major hereditary breast/ovarian cancer predisposing genes and their mutations increase the risk of developing cancer (1, 2). It is assumed that germline mutations in the coding region of these two genes are responsible for familial breast/ovarian cancers while reduced expression of these genes (caused by various mecha-

nisms such as methylation of the CpG islands within the promoter region, allelic deletion of the gene locus and sequence alterations identified outside the coding region) is frequently observed in sporadic breast/ovarian tumors (3, 4).

BRCA2 gene is transcribed into an 11-kb mRNA that has 27 encoding exons and is translated into 3418 amino acids long protein that has been implicated in processes essential for all cells, including proliferation, development, DNA repair, transcription and centrosome duplication (5, 6). In addition to female and male breast cancer, mutations in the BRCA2 gene can lead to an increased risk of ovarian, prostate, pancreatic, fallopian tube cancer and melanomas (7). Mutations in the central part of the gene have been associated with a higher risk of ovarian cancer and a lower risk of prostate cancer than mutations in other parts of the gene (8).

Genetic testing for BRCA2 sequence alterations is nowadays commonly performed, more recently also in Croatia (9). There are over 1800 BRCA2 genetic variants recorded in the Breast Cancer Information Core (BIC) database (10), of which almost half are declared as variants of unknown clinical significance. Interpretation of these unclassified variants (UVs) is the major concern for BRCA genes, especially for risk assessment in genetic counseling (11). Additional difficulties for biological and clinical interpretation present genetic variants that do not change the amino acid composition of protein, like variants found within 5' and 3' untranslated regions (UTRs) of messenger RNA.

Untranslated regions are parts of the mature mRNA located before the start codon (5' UTR) and after the stop codon (3' UTR). They are transcribed with the coding region but they are not translated. Several regulatory roles have been assigned to the untranslated regions, including mRNA's localization and stability, and translational efficiency (12). These functions depend both on the sequence and structure of the UTRs (12). There are a growing number of disorders that are caused by changes in the *cis*-regulatory sequences of the UTRs (13). For human BRCA1 5' UTR it has been shown to play an important role in regulating BRCA1 translational efficiency (14) and two somatic point mutations in this region (c.-3G>C and c.-2A>T), which have been identified in a highly aggressive, sporadic breast tumors, dramatically reduced transcript translatability (15, 16). The role of BRCA2 5' UTR in post-transcriptional regulation of protein translation efficiency is not yet known. Only for the sequence variant c.-26G>A was experimentally shown that it increases the expression level of BRCA2 gene and presents risk or protection according to the genotype status in the sporadic form of breast cancer (17). But the functioning of that sequence variant at the molecular level was deduced only by computational analysis.

In the growing world of protein non-coding RNAs (ncRNAs) (18) computational methods and algorithms provide significant advancement and alleviation in research (19). In this study we tried to computationally an-

alyze potential significance of human BRCA2 5' UTR sequence variants. So far, only two studies have been conducted *in silico* to investigate functional effects of BRCA2 5' UTR sequence alterations (20, 21), but they only searched if the sequence variants change nucleotide patterns of potential regulatory regions in BRCA2 5' UTR. As it is known that the functioning of ncRNAs depends both on nucleotide sequence and their secondary structure (base pairing) (22), we tried by predicting consensus secondary structure of 5' UTR to find out if there are substructures that have been conserved by evolution, as it is far more likely that conserved structures are functionally important. Then we analyzed potential functional significance of sequence variants found within human 5' UTR of BRCA2 gene, what could be expressed by disruption of the consensus secondary structure. In essence, secondary structure is the list of base pairs, denoted by i - j for a pairing between the i^{th} and j^{th} nucleotides, where $i < j$ by convention (23). Elements of RNA secondary structure can be seen on Figure 1. RNA helices normally contain 6 base pairing combinations: The Watson-Crick pairs G-C, C-G, A-U, U-A, and the slightly weaker wobble pairs G-U and U-G.

The main aim of this study was to find out if computational screening for 5' UTR sequence variants with potential structural and functional significance can be used for narrowing a list of candidates for further (much more laborious and expensive) experimental molecular characterization.

MATERIALS AND METHODS

Sequences of BRCA2 5' UTRs from different species were collected from UTRdb (24), a curated database of 5' and 3' untranslated sequences of eukaryotic mRNAs derived from several sources of primary data.

Human BRCA2 5' UTR sequence variants were collected from three public databases: Breast Cancer Infor-

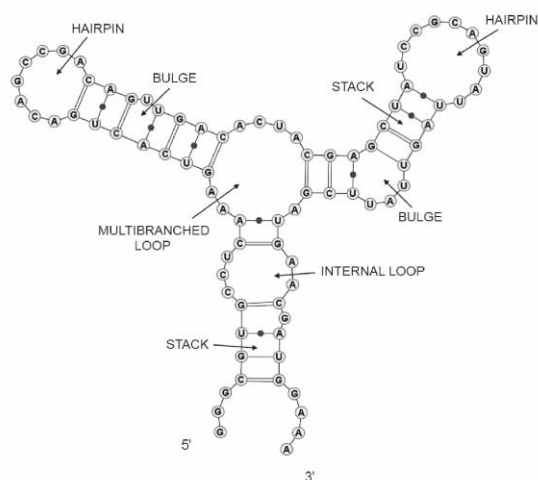


Figure 1. Elements of RNA secondary structure (base pairing). Each loop has at least one base pair. A stem consists of two or more consecutive stacks.

mation Core (BIC) (10), the biggest BRCA1 and BRCA2 sequence alterations database; dbSNP (25) and KConFab (26). In this study we included all present variants, regardless if they are found in patients or healthy controls.

We used UTRscan server (27) to search human BRCA2 5' UTR sequence for any of the patterns collected in UTRSite (24), a collection of experimentally validated functional sequence patterns located in eukaryotic 5' or 3' mRNA UTRs, that are crucial for many aspects of gene regulation and expression. If different BRCA2 5' UTR sequence of each UTR sequence variants is found to have different functional pattern(s), this UTR sequence variant could have functional significance.

Structural analysis of BRCA2 5' UTR and its sequence variants was performed using different tools from the Vienna RNA websuite (28). There are two main computational approaches for predicting RNA secondary structure: the first is a thermodynamic method, which assumes that a given sequence will fold into the structure with the minimum free energy (MFE) (29) and the second approach compares multiple orthologous sequences to identify patterns of co-evolution between sites that could be indicative of compensatory mutations (30) to maintain complementary base pairing within stacked base pairs (31, 32). Since the accuracy of thermodynamic secondary structure predictions for individual sequences is rather limited, computing the consensus structure common to several related RNA sequences can drastically improve the prediction (33). Consensus secondary struc-

ture of BRCA2 5' UTR was predicted using RNAalifold web server (34). RNAalifold works like single sequence folding algorithms with the main difference being the energy model is augmented by covariance information. Compensatory mutations (e.g., a U·A pair mutates to a C·G pair) and consistent mutations (e.g., A·U mutates to G·U) give a »bonus« energy, whereas inconsistent mutations (e.g., C·G mutates to C·A) yield a penalty. This results in a consensus MFE structure common to most of the sequences in an alignment (34). Multiple sequence alignment, prerequisite for consensus structure prediction, was constructed using ClustalX2 version 2.0.12 software (35). To explore potentials effects of human BRCA2 5' UTR sequence variants on consensus 5' UTR secondary structure, folding of human BRCA2 5' UTR including one of each variant was predicted with RNAfold (28) using consensus structure as a constraint. If constrained folding results in a structure that is very different from the consensus one, this may indicate that this particular sequence variant could have potential impact on normal BRCA2 5' UTR.

RESULTS

We retrieved nine different BRCA2 5' UTR sequences from eight species (mouse has two different BRCA2 5' UTR transcripts) and according to the sequence identity and length, four BRCA2 5' UTR sequences were included in further analysis: human (*Homo sapiens*, 227 nucleotides, accession nr. BR431031), chimpanzee (*Pan troglodytes*, 212 nt, BR307535), rhesus monkey (*Macaca*

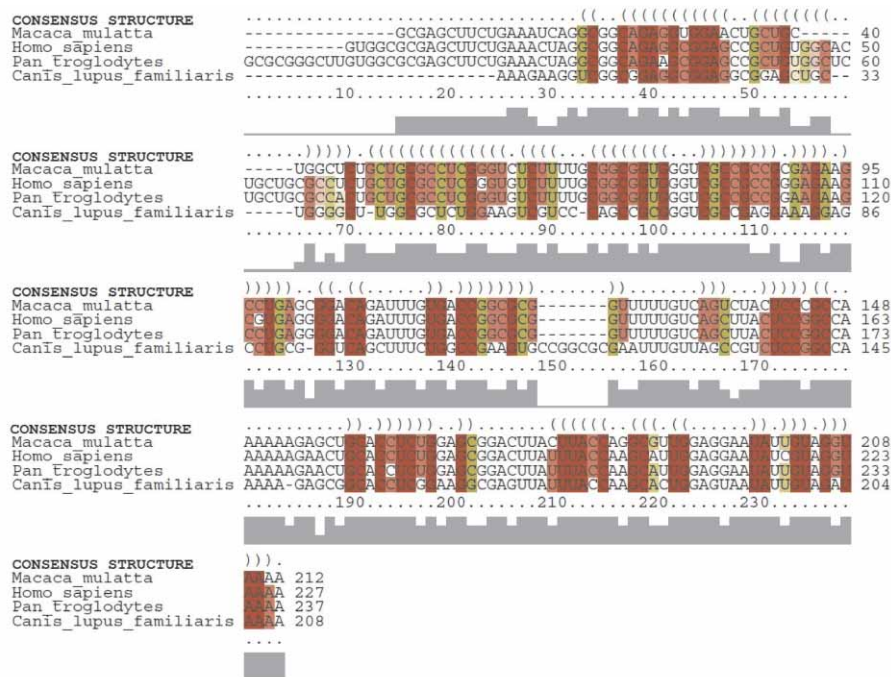


Figure 2. Structure annotated alignment. Consensus secondary structure of BRCA2 5' UTR based on alignment of sequences from four different species is presented in the first line in dot-bracket notation. Gray bars under the multiple sequence alignment present level of base conservation within that position in four 5' UTR sequences. The colors of the shadings indicate the number of different types of letter combinations that form a base pair. Red means that there is 1 and ochre that there are 2 different base-pair combinations. If a base pair cannot be formed in one or more sequences, the colors are shown faded in different levels.

mulatta, 212 nt, BR134108) and dog (*Canis lupus familiaris*, 208 nt, BR118937).

We found thirteen different human BRCA2 5' UTR variants: c.-175C>T, c.-123G>T, c.-52A>G, c.-34T>C, c.-26G>A, c.-26G>C, c.-26G>T, c.-15A>C, c.-14T>C, c.-12T>C, c.-12T>G, c.-11C>T and c.-9T>C.

According to UTRscan, human wild-type BRCA2 5' UTR sequence does not contain any known functional 5' UTR sequence pattern, nor any of sequence variants create a new one.

Based on the BRCA2 5' UTR sequences from four different species, RNAalifold software built consensus secondary structure (Figures 2, 3) with a minimum free energy of -98.63 kcal/mol (-50.63 kcal/mol from MFE averaged over all sequences in the alignment plus -48.00 kcal/mol from covariance contributions). Figure 3 shows conventional secondary structure drawing with conservation annotation.

Folding of human wild type BRCA2 5' UTR using consensus structure as a constraint gave a secondary structure with MFE of -73.60 kcal/mol (Figure 4a) and that structure greatly resembles consensus one (Figure 3). Folding of BRCA2 5' UTR with one of each sequence variant with the same constraint gave secondary structures that ranged in MFE from -76.20 to -72.10 kcal/mol

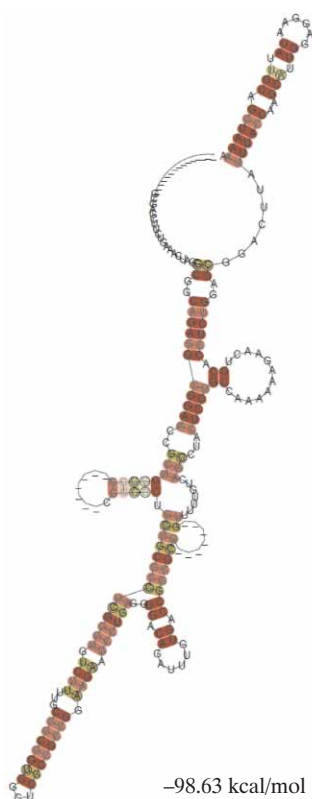


Figure 3. Classical graphical representation of BRCA2 5' UTR consensus secondary structure with minimum free energy (MFE). Variable positions are circled (one circle, consistent mutation; two circles, compensatory mutation). The coloring scheme is the same as for Figure 2.

(Figures 4b to 4n). Compared to wild type sequence, differences in MFE ranged from $+1.50$ to -2.60 kcal/mol. Most of the 5' UTR sequence alterations (10 of 13) are found near the 3' end of 5' UTR, what is in the vicinity of the translation initiation site. All three nucleotide changes at position c.-26 disrupted conserved substructure with three small loops and created one big loop with smaller MFE. The same situation was with sequence variant c.-12T>C (Figure 4l) but not c.-12T>G (Figure 4k). These four sequence variants most notably changed consensus secondary structure. Variant c.-14T>C completely resembles wild type structure but has a lower (more stable) MFE (Figure 4j). Sequence variant c.-123G>T is the only alteration positioned more upstream from the translation start site that changed consensus secondary structure (Figure 4c).

DISCUSSION

Among nine different BRCA2 5' UTR sequences, only four had enough sequence identity and similarity for building multiple sequence alignment, what is prerequisite for predicting consensus secondary structure. It is interesting that mouse and rat, although not so evolutionary distant from human, have completely different sequences of BRCA2 5' UTR. That could also imply differences in regulation of BRCA2 expression. The same situation is with mouse and human BRCA1 5' UTR (36).

The amount of sequence variants found in BRCA2 5' UTR is much higher than in BRCA1 5' UTR (13 compared to 4). Since genetic testing of BRCA1 and BRCA2 5' and 3' UTRs is not so commonly performed as for their coding region, the reason for that discrepancy is not clear. This could potentially imply two things: first, that the role of BRCA2 5' UTR in post-transcriptional regulation of protein translation efficiency is not so important and greater sequence variability is allowed; or second, that more breast cancers could be connected with mutations in BRCA2 5' UTR. Unfortunately, it is not known for all BRCA2 5' UTR variants how they co-segregate with breast/ovarian cancer and what is their penetrance.

Most of the 5' UTR sequence alterations (10 of 13) are found near the 3' end of 5' UTR, what is in the vicinity of the translation initiation site. Our study confirmed finding from Gochhait *et al.* (17) that substitution of G with A at position -26 from the first AUG codon created substructures with lower minimum free energy, thus less stable. They also showed that this c.-26G>A variant is functional in a way so that the A allele increased the reporter gene expression by twice that of the wild type H allele (17). As three another sequence variants c.-26G>C, c.-26G>T and c.-12T>C in the same manner disturbed consensus secondary structure near the translation initiation site, they could also increase the efficiency of the translation and thereby increase the expression of BRCA2. This is in agreement with the work of Vega Laso and colleagues (37) and Kozak (38), who argued that the extent of a hairpin negative effect on eukaryotic mRNA translation *in vivo* depends upon its stability and localization within the molecule. Also, according to their findings, for

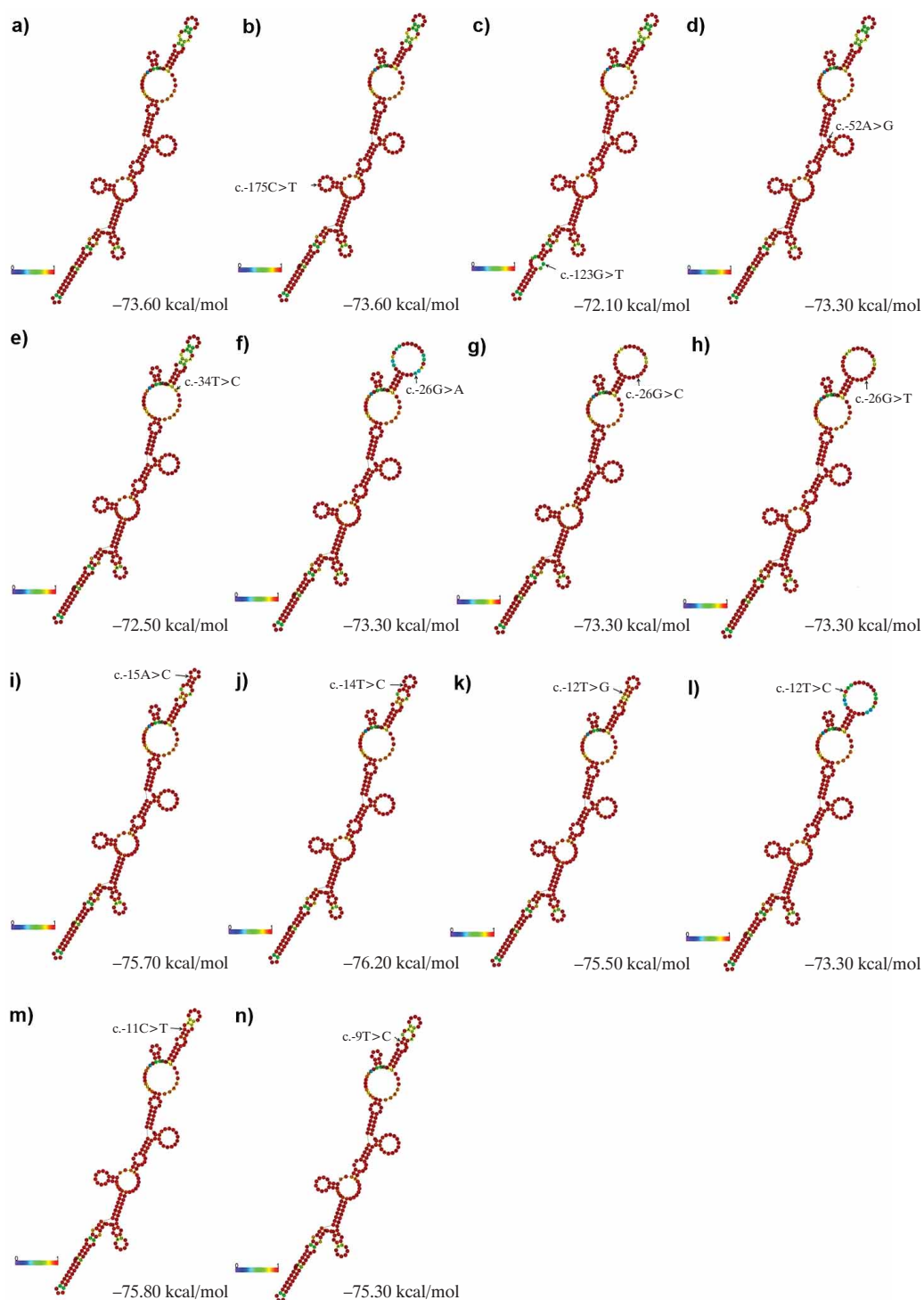


Figure 4. Potential effect of each BRCA2 5' UTR sequence variant on consensus secondary structure. For wild type and each sequence variant folding into the structure with minimum free energy (MFE) was predicted using consensus secondary structure (Figure 2 and 3) as a constraint. Structure colors encode base-pair probabilities.

sequence variant c.-123G>T is less likely to be functional because it is positioned closer to the 5' end of 5' UTR. All these findings suggested that these three BRCA2 5' UTR sequence variants, c.-26G>C, c.-26G>T and c.-12T>C, as the c.-26G>A variant, are suitable candi-

dates for further functional characterization and that they could be potentially clinically significant.

In conclusion, as there is no more experimental validation of possible effects of human BRCA2 5' UTR sequence variants, besides that for variant c.-26G>A, *in*

silico analysis made the first step in that elucidation. Computational screening for 5' UTR sequence variants with eventual structural and functional significance is a valuable tool for narrowing a list of candidates for further (much more laborious and expensive) experimental molecular characterization.

REFERENCES

- MIKI Y, SWENSEN J, SHATTUCK-EIDENS D, FUTREAL P A, HARSHMAN K, TAVTIGIAN S, LIU Q, COCHRAN C, BENNETT L M, DING W, BELL R, ROSENTHAL J, HUSSEY C, TRAN T, MCCLURE M, FRYE C, HATTIER T, PHELPS R, HAUGEN-STRANO A, SKOLNICK M A 1994 Strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266: 66–71
- WOOSTER R, BIGNELL G, LANCASTER J, SWIFT S, SEAL S, MANGION J, COLLINS N, GREGORY S, GUMBS C, MICKLEM G 1995 Identification of the breast cancer susceptibility gene BRCA2. *Nature* 378: 789–792
- THOMPSON M E, JENSEN R A, OBERMILLER P S, PAGE D L, HOLT J T 1995 Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. *Nat Genet* 9(4): 444–50
- CHAN K Y, OZÇELİK H, CHEUNG A N, NGAN H Y, KHOO U S 2002 Epigenetic factors controlling the BRCA1 and BRCA2 genes in sporadic ovarian cancer. *Cancer Res* 62(14): 4151–6
- POWELL S N, KACHNIC L A 2003 Roles of BRCA1 and BRCA2 in homologous recombination, DNA replication fidelity and the cellular response to ionizing radiation. *Oncogene* 22(37): 5784–91
- YOSHIDA K, MIKI Y 2004 Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci* 95(11): 866–71
- The Breast Cancer Linkage Consortium 1999 Cancer Risks in BRCA2 Mutation Carriers. *J Natl Cancer Inst* 91(15): 1310–6
- LUBINSKI J, PHELAN C M, GHADIRIAN P, LYNCH H T, GARBER J, WEBER B, TUNG N, HORSMAN D, ISAACS C, MONTEIRO A N, SUN P, NAROD S A 2004 Cancer variation associated with the position of the mutation in the BRCA2 gene. *Fam Cancer* 3(1): 1–10
- CVOK M L, CRETNİK M, MUSANI V, OZRETIĆ P, LEVANAT S 2008 New sequence variants in BRCA1 and BRCA2 genes detected by high-resolution melting analysis in an elderly healthy female population in Croatia. *Clin Chem Lab Med* 46(10): 1376–83
- SZABO C, MASIELLO A, RYAN J F, BRODY L C 2000 The breast cancer information core: database design, structure, and scope. *Hum Mutat* 16(2): 123–31 WEB: <http://research.nhgri.nih.gov/bic/>
- WU K, HINSON S R, OHASHI A, FARRUGIA D, WENDT P, TAVTIGIAN S V, DEFFENBAUGH A, GOLDFAR D, COUCH F J 2005 Functional evaluation and cancer risk assessment of BRCA2 unclassified variants. *Cancer Res* 65(2): 417–26
- PESOLE G, GRILLO G, LARIZZA A, LIUNI S 2000 The untranslated regions of eukaryotic mRNAs: structure, function, evolution and bioinformatic tools for their analysis. *Brief Bioinform* 1(3): 236–49
- CHATTERJEE S, PAL J K 2009 Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. *Biol Cell* 101(5): 251–62
- SOBCZAK K, KRZYZOSIAK W J 2002 Structural determinants of BRCA1 translational regulation. *J Biol Chem* 277: 17349–58
- SIGNORI E, BAGNI C, PAPA S, PRIMERANO B, RINALDI M, AMALDI F, FAZIO V M 2001 A somatic mutation in the 5' UTR of BRCA1 gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene* 20: 4596–600
- WANG J, LU C, MIN D, WANG Z, MA X 2007 A mutation in the 5' untranslated region of the BRCA1 gene in sporadic breast cancer causes downregulation of translation efficiency. *J Int Med Res* 35(4): 564–73
- GOCHHAIT S, BUKHARI S I, BAIRWAN, VADHERA S, DARVISHI K, RAISH M, GUPTA P, HUSAIN S A, BAMEZAI R N 2007 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* 9(5): R71
- FRITH M C, PHEASANT M, MATTICK J S 2005 The amazing complexity of the human transcriptome. *Eur J Hum Genet* 13(8): 894–7
- MACHADO-LIMA A, DEL PORTILLO H A, DURHAM A M 2008 Computational methods in noncoding RNA research. *J Math Biol* 56(1–2): 15–49
- RAJASEKARAN R, DOSS G P, SUDANDIRADOSS C, RAMANATHAN K, RITURAJ P, SETHUMADHAVAN R 2008 Computational and structural investigation of deleterious functional SNPs in breast cancer BRCA2 gene. *Sheng Wu Gong Cheng Xue Bao* 24(5): 851–6.
- TOMMASI S, PILATO B, PINTO R, MONACO A, BRUNO M, CAMPANA M, DIGENNARO M, SCHITTULLI F, LACALAMITA R, PARADISO A 2008 Molecular and in silico analysis of BRCA1 and BRCA2 variants. *Mutat Res* 644(1–2): 64–70
- PESOLE G, MIGNONE F, GISSI C, GRILLO G, LICCIULLI F, LIUNI S 2001 Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 276(1–2): 73–81
- ZUKER M 2000 Calculating nucleic acid secondary structure. *Curr Opin Struct Biol* 10(3): 303–10
- GRILLO G, TURI A, LICCIULLI F, MIGNONE F, LIUNI S, BANFI S, GENNARINO V A, HORNER D S, PAVESI G, PICCARDI E, PESOLE G 2010 UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res* 38(suppl 1): D75–D80 WEB: <http://utrdb.ba.itb.cnr.it/>
- SHERRY S T, WARD M H, KHOLODOV M, BAKER J, PHAN L, SMIGIELSKI E M, SIROTKIN K 2001 dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research* 29(1): 308–311
- OSBORNE R, HOPPER J, KIRK J, CHENEVIX-TRENCH G, THORNE H, SAMBROOK J and the KATHLEEN CUNNINGHAM Consortium for Research into Familial Breast Cancer 2000 KConFab: a resource for studies into the genetics, epidemiology, biology, pathology and psychosocial aspects of breast cancer families. I. Identification and characterisation of eligible families. *Medical Journal of Australia* 172: 463–464 WEB: <http://www.kconfab.org/Progress/Mutations.aspx>
- PESOLE G, LIUNI S 1999 Internet resources for the functional analysis of 5' and 3' untranslated regions of eukaryotic mRNAs. *Trends Genet* 15(9): 378 WEB: <http://itbtools.ba.itb.cnr.it/utrscan>
- GRUBER A R, LORENZ R, BERNHART S H, NEUBÖCK R, HOFACKER I L 2008 The Vienna RNA website. *Nucl Acids Res* 36(suppl 2): W70–W74 WEB: <http://rna.tbi.univie.ac.at>
- ZUKER M, STIEGLER P 1981 Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9(1): 133–48
- KIMURA M 1985 The role of compensatory neutral mutations in molecular evolution. *J Genet* 64: 7–19
- HIGGS P G 1998 Compensatory neutral mutations and the evolution of RNA. *Genetica* 102–103(1–6): 91–101
- CHEN Y, CARLINI D B, BAINES J F, PARSCH J, BRAVERMAN J M, TANDA S, STEPHAN W 1999 RNA secondary structure and compensatory evolution. *Genes Genet Sys* 74: 271–286
- HOFACKER I L, FEKETE M, STADLER P F 2002 Secondary structure prediction for aligned RNA sequences. *J Mol Biol* 319(5): 1059–66
- BERNHART S H, HOFACKER I L, WILL S, GRUBER A R, STADLER P F 2008 RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 9: 474
- LARKIN M A, BLACKSHIELDS G, BROWN N P, CHENNA R, McGETTIGAN P A, MCWILLIAM H, VALENTIN F, WALLACE I M, WILMA, LOPEZ R, THOMPSON J D, GIBSON T J, HIGGINS D G 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948
- CHAMBERS J A, SOLOMON E 1996 Isolation of the murine Nbr1 gene adjacent to the murine Brca1 gene. *Genomics* 38: 305–313
- VEGA LASO M R, ZHU D, SAGLIOCCO F, BROWN A J, TUITE M F, MCCARTHY J E 1993 Inhibition of translational initiation in the yeast *Saccharomyces cerevisiae* as a function of the stability and position of hairpin structures in the mRNA leader. *J Biol Chem* 268(9): 6453–62
- KOZAK M 1994 Determinants of translational fidelity and efficiency in vertebrate mRNAs. *Biochimie* 76: 815–821