

In-silico analysis of BCL2 gene using multiple bioinformatics tools to identify the most lethal mutations that are crucial for its structural and functional integrity

Usman Ghani

Center for Applied Molecular Biology and Forensic University of Punjab Lahore

Yasir Ali

Abdul Wali Khan University Mardan

Karim Gul

Abdul Wali Khan University Mardan

Aamir sohail

Department of Biotechnology Abdul Wali Khan University Mardan

Rahmat Ullah

Center for Applied Molecular Biology and Forensic University of Punjab Lahore

Tayyaba Anjum

Center for applied molecular and forensic Sciences University of Punjab Lahore

Dr. Fazal jalil (✉ fazaljalil@awkum.edu.pk)

Department of Biotechnology Abdul Wali Khan University Mardan

Research Article

Keywords: BCL2, antiapoptotic activity, BCL2 lethal variants, BCL2 and cancer, structural and functional variants of BCL2

Posted Date: August 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-310735/v4>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

BCL2 was the first ever known gene for anti-apoptotic activity, that encodes for essential proteins of the external mitochondrial membrane. Regarding tumorigenesis, deregulated BCL2 expression and related proteins have been recognized as characteristic of several human cancers and there is concrete evidence that the deregulated expression of BCL2 like proteins plays a vital role in tumor development, persistence and therapeutic resistance. Therefore, it is important to identify the polymorphisms of BCL2 that are both structurally and functionally important for research to find their possible malfunctions and therapeutics. For this reason, in our research, we have used a variety of bioinformatics tools to recognize the most destructive nsSNPs that may be important for the structure and function of BCL2. *In silico* tools, PROVEAN, SIFT, SNP&GO, PhD SNP, and PolyPhen2 included a variety of other tools such as I Mutant, MutPred, and ConSurf, to study their conservation profiles to validate their stability, structural, and functional impacts. Post-transcriptional alteration sites were also predicted followed by application of 3-D mapping with I-TASSER and Phyre2 tools. Furthermore, the gene interactions were mapped via STRING and GeneMANIA. We also found that nsSNPs Q118R (rs759928495), G193R (rs1197820694), R129C (rs777784952), and L181V (rs752310933) are the most destructive nsSNPs in BCL2 genes that can have a vital part in BCL2 protein defects and possibly cause different cancers. Gene-gene interactions showed relation of BCL2 with other genes depicting its importance in several pathways and co-expressions. This research is the first of its kind and offers future prospects for the development of dedicated medicines as well. In the animal models, the effects of BCL2 can also be tested in diseases. Such should be the study of BCL2 proteins from cancer patients. The effects of BCL2 can also be tested on animal models.

Introduction

There are different categories of mutations that may occur in the DNA sequence of human. Among them the single base pair changes termed as SNPs (Single Nucleotide Polymorphisms) constitutes ~90% of it. It is estimated that around 500,000 SNPs occur in the coding region of human genomes (Collins et al., 1998). The nsSNPs (nonsynonymous Snp) are some of the SNPs that results in the functional and structural diversity of human proteins (Lander 1996). Such nsSNPs have a very important effect in changing the amino acids that cause favorable or deleterious effects on protein structure and/or function (Capriotti and Altman, 2011). These alterations include shift in gene regulation, instability of protein structure (Boroso et. al., 1999), hydrophobicity, impacting structure and protein charge (Petukh et. a., 2015), translation, relative stability, protein interrelationships and dynamics (Inter/Intra) (Chasman and Adams, 2001, Kucukkal et. al., 2015) are other harmful effects for structural and functional cell integrity (Thomas et. al., 1999).

Single Nucleotide Polymorphisms in *BCL2* gene have been also found that encodes an integral outer mitochondrial membrane protein that controls apoptosis, upregulate in nearly 50 percent of all human cancers and fulfill its role as an apoptotic control system. They are the principal regulators of apoptosis mitochondrial pathways. This mechanism is important for the normal development of embryos and cancer prevention. The mitochondrial external membrane, which releases cytochrome c and further apoptotic factors into the cytosol, is regulated by these proteins. The BCL2 family are generally classified into two groups, with up to four BCL2 homology domains (BH1, BH2, BH3 & BH4), anti-apoptotic and pro-apoptotic proteins. Their intracellular function and sequence homology are based on anti-apoptotic and pro apoptotic proteins. BCL2 plays a major role as an apoptotic regulation agent for nearly one half of all human cancers (Cory et al, 2003, Yip and Reed, 2008, Reed 2008). In small cell lymphomas, namely chronic lymphocytic leukemia (CLL), peripheral lymphoma and cell lymphoma mantle, BCL2 are over-expressed, although less than 5 percent are BCL2-related (Tomite, N., 2011) Elevated BCL2 expression is often reported in nearly all patients suffering

from acute lymphocytis and acute myeloid leukemia (Yip and Reed in 2008). Though, BCL2 protein has inappropriate expression in most adult follicular lymphoma FL events, the pediatric kind of follicular lymphoma (FL) is negative to expression of BCL2. About 30 percent of the patients who are suffering from large-diffuse B cell lymphoma (DLBCL), are regarded as relatively exposed to BCL (Schuetz et al., 2012). BCL2 is also suspected of being an inappropriate expression in non-hematic tumours, and in solid tumors such as prostate, breast, small cells and non-small cells lung cancers. Among small cells lung cancer, strong BCL2 expression was identified for almost for >90 percent of patients (Verdoodt et al., 2013, Anagnostou et al., 2010 Oakes et al., 2012). The BCL2 inappropriate expression has been widely recognized for ovarian, neuroblastoma, lung, colorectal and certain cancers of the head and neck (Beierle et al., 2002 Yasmeen et al., 2011 Cho, H.J et al., 2006 Carter, R.J., 2019 Koehler, B.C., 2013)

We have also proposed 3D model of wild type BCL2 protein and its possible deleterious ns SNPs protein models as well. For BCL2, this is first study of its type which covers *in silico* analysis for protein and can be helpful in future for treatment of BCL2 associated diseases caused by nsSNPs.

Methodology

The work was carried out in many steps for which a schematic flow is given in figure 1.

Recruiting nsSNPs

SNPs in BCL2 gene were recruited from NCBI dbSNP ([https:// www.ncbi.nlm.nih.gov/projects/SNP/](https://www.ncbi.nlm.nih.gov/projects/SNP/)) along with related information such as location, global minor allele frequencies (MAF), residual transition etc. (Bhagwat, 2010) For investigation all 159 nsSNPs were processed/filtered.

Deleterious nsSNPs Identification

To classify the impact of nsSNPs that were chosen from dbSNP, four bioinformatics tools were used. These tools were SIFT (Sorting Intolerant From Tolerant) (PC and Henikoff, 2006 and Kumar et. al., 2009), PROVEAN (Protein Variation Effect Analyzer) (Choi et. al., 2012), PhD-SNP (Human Deleterious SNP Predictor) (Capriotti et. al., 2006) and SNPs&GO (Capriotti et. al., 2013). Those SNPs were selected which are predicted by all four tools to be deleterious or intolerant, and were further screened through polyPen2 (Polymorphism Phenotyping 2) (Adzhubei et. al., 2010).

Predicting Functional and Structural Effects

Mutpred 1.2 was used to model the functional and structural impact of nsSNPs (residual amino acid change) (Li et. al., 2009). It is a web-based platform that tracks amino acid substitution and also forecasts molecular cause of a disease. This tool works on evaluating numerous structural and functional properties, such as increase in helical inclination or lack of site phosphorylation. The protein sequence of BCL2 in FASTA format was sent and the deleterious nsSNPs (change of amino acids) were chosen. $P < 0.05$ was deemed optimistic, and $p < 0.01$ was considered quite positive.

Protein stability Analysis

I-Mutant 2.0 was used to test the stability of our target protein (Capriotti et. al., 2005). It is a web server that is based on vector machine technology. It predicts improvement in mutated protein stability and offers RI (Reliability Index) predictions that vary from 0 to 10 where 0 and 10 represent the lowest and highest reliability respectively. In order to predict the impact of deleterious nsSNPs on BCL2 protein, its protein sequence has been submitted with conditions set at 25 μ C and 7.0 pH.

Prediction of Evolutionary Conservation of Protein

The determination of evolutionary conservation of all amino acids in a protein chain has been accomplished by ConSurf (Berezin et al., 2004). ConSurf analysis is based on phylogenetic connections between homologous sequences (Ashkenazy et. al., 2010; Celniker et. al., 2013; and Ashkenazy et. al., 2016).

Predicting 3D Protein Structure

Phyre2 is a 3D simulation tool for prediction of 3D protein structures (Kelly et. al., 2015). 3D models were created for wild type BCL2 and its 12 mutants associated with most deleterious nsSNPs. TM-align was used for contrast of wild type BCL2 with chosen mutants. It predicts TM ranking, RMSD (Root Mean Square Deviation), and structural superposition. TM scores are given in the range from 0 to 1, where 1 indicates greater structural similarity. Higher the RMSD values, the difference between mutant and wild form systems will be greater (Carugo and Pongor, 2010 and Zhang and Skolnick, 2005). For further analysis of protein 3D structure differences (Zhang, 2008, Roy et. al., 2010 and Yang et. al., 2015), three mutants with higher RMSD values were sent to the I-TASSER along with the wild type of BCL2. Chimera 1.11 has been used to analyze molecular dynamics and digital analysis of the corresponding protein structure (Petterson et. al., 2004).

Post-Transcriptional Modification (PTM) Sites Prediction

A thorough study on PTM in protein will help in the prediction of protein activity. GPS-MSP 3.0 had expected methylation sites in BCL2 protein (Deng et. al., 2017). Predicting phosphorylation sites using GPS 3.0 and NetPhos 3.1 was performed at serine, tyrosine, and threonine residual locations in the BCL2 protein chain. GPS 3.0 expected a more specific outcome with a higher phosphorylation ability than NetPhos 3.1 (Xue et. al., 2008). NetPhos 3.1 has used neural network ensembles and set a threshold of 0.5 (Blom et. al., 1999). Residues with a score higher than the phosphorylated threshold are expected. BDM-PUB and UbPred were used for the prediction of ubiquitination sites in BCL2 protein. For UbPred, a healthy cut-off was chosen (Li et. al., 2009) and UbPred projected lysine residues as ubiquitinated getting a value equivalent to or above 0.62 threshold (Radivojac et al., 2010).

Gene-Gene Interaction of BCL2

GeneMANIA and STRING researched the relationship of the BCL2 gene to identify its relation with other genes and to forecast the impact of BCL2's nsSNPs on other associated genes (Farley et. al., 2010, Gasteiger et. al., 2003). GeneMANIA predicts gene-gene interaction dependent on co-expression, mechanisms, co-localization, similarities to the protein domain and association between genes and proteins. STRING projections were limited to the top 10 best interactive genes for which the criteria were; gene fusions, co-occurrence, co-expression, experimental, and biochemical data. The combined score for each target gene interacting was between 0 and 1, where 0 was the lowest interaction and 1 was the highest interaction.

BCL2 submitted as in put gene for generating gene-gene interaction network.

Results And Discussion

Recruited nsSNPs

A total of 44931 SNPs were recruited from dbSNP i.e. the largest SNP pool comprising of 159 non-synonymous SNPs, 398 found in 5'UTR, 1842 existed in 3'UTR and other forms of SNPs (Figure 2). We only investigated nsSNPs for further review.

Identification of deleterious nsSNPs

All of the 159 nsSNPs recruited from dbSNP were exposed to four separate bioinformatics tools that were used to predict the impact of these nsSNPs on BCL2 protein function and structure. It composed of PROVEAN, PhD SNP, SIFT, and SNP & GO in silico devices. In PROVEAN, the threshold value was set at -2.5 and the version having the final score below this threshold was deleterious. According to PROVEAN findings 35 nsSNPs is shown to have deleterious effects. In SIFT, a baseline value of 0.05 was assumed to be the TI (Tolerance Index), while the tests below this value were deemed to be efficient or intolerant. SIFT had expected the hostility of 35 nsSNPs. 4 nsSNPs were expected to be diseased by PhD SNP. SNPs and GO concluded in 98 nsSNPs categorized as sick. Result shown below in Figure 3. We picked 4 nsSNPs that were shown to be deleterious in all four tools and 0.00 TI in SIFT estimation as well. These 4 chosen nsSNPs have been sent to PolyPhen2 which predicts benign, probably damaging and possibly damaging effects. Next to positive, and possibly harmful, the most optimistic forecasts are likely to be detrimental. It also offers 0 to 1. count rate for further study, 3 nsSNPs with score 1 were picked, as they are deemed the most destructive.

Table 1: Identification of Deleterious nsSNPs using 5 Insilco tools

rs ID	Amino acid Change	PROVEAN	Polyphen2 (HumDiv)	SNPs&GO		SIFT	PhD-SNP
		Score	Score	RI	Probability	Score	Score
rs759928495	Q11R	-3.349	1.000	6	0.813	0.06	2
rs777784952	R129C	-3.199	1.000	6	0.809	0	1
rs752310933	L181V	-2.706	1.000	4	0.714	0	7
rs1197820694	G193R	-7.039	1.000	10	0.992	0	4

MutPred Prediction for Structural and Functional Modification

The selected 4 nsSNPs from PolyPhen2 were subjected to server MutPred. It provided results that are shown in table 2 with the likelihood scores. The projections have shown that certain nsSNPs can induce alteration of the protein and can influence its role or structure.

Table 2: Probability values of deleterious nsSNPs identified in BCL2

ID	Substitution	MutPred2 score	Remarks	Affected PROSITE and ELM Motifs
rs759928495	Q118R	0.383	-	-
rs777784952	R129C	0.487	-	-
rs752310933	L181V	0.713	-	ELME000052, ELME000084, ELME000182
		Molecular mechanisms with P-values <= 0.05	Probability	P-value
rs759928495	G193R	0.893	-	PS00008, PS01258
		Molecular mechanisms with P-values <= 0.05	Probability	P-value
Loss of Strand		0.26	0.05	Loss of Strand
Altered Transmembrane protein		0.13	0.02	Altered Transmembrane protein

BCL2 Stability Prediction

I-Mutant was used to determine the stability of BCL2 proteins for the chosen nsSNPs and their replacement of amino acids. All selected nsSNPs were submitted separately and their outcome of decreasing / increasing stability was obtained with RI varying from 0 to 10 presented in table 3. In 3 selected nsSNPs, no improvement in stability and all showed decrease in stability is shown. This result predicted that by reducing its stability, these 4 nsSNP may cause greater harm to BCL2 protein.

Table 3: I-Mutant Result for the selected nsSNPs

SNP id	Amino acid Change	Stability	RI
rs759928495	Q118R	Decrease	1
rs777784952	R129C	Decrease	9
rs752310933	L181V	Decrease	6
rs1197820694	G193R	Decrease	9

Note: The pH was set at 7.0 and Temperature at 25°C.

Evolutionary Conservation of BCL2 Protein

ConSurf provided us evolutionary profile for all the amino acids of BCL2 protein. According to ConSurf predictions, R129C Q118R and G193R were highly conserved, exposed and functionally active residue. L181V was highly conserved and buried as structural residue. Conservation scores for each of the selected nsSNPs are given in table 4.

These results show that nsSNPs, which are located at highly conserved regions are the most damaging to the BCL2 protein function and structure.

Table 4: Conservation profiles of the selected Most Damaging nsSNPs:

SNP ID	Amino Acid Chang	Conservation Score	Prediction
rs759928495	Q118R	9	Highly Conserved and Exposed(f)
rs777784952	R129C	8	Highly Conserved and Exposed
rs752310933	L181V	9	highly conserved and buried(S)
rs1197820694	G193R	9	Highly Conserved and Exposed(f)

3D-Modelling of BCL2 Protein

I-mutant estimated that the stability of 4 nsSNP protein in BCL2 protein will decrease and have been chosen for final protein modeling. Protein sequences apply to I-TASSER for generation of BCL2 protein structure with single wild-type amino acid and mutants. It is the most accurate and sophisticated method to determine the structure of proteins. Wild type and mutant protein sequences were sent to I-Tasser for the development of BCL2 protein structure and for the creation of five models for each BCL2 protein and mutant. I-TASSER used 2co9 (83 per cent identity) and 2nbiA (85 per cent thread alignment coverage) models. I-TASSER used ten SPARKS-X, Rally, HHSEARCH-2, FFAS-3D, HHSEARCH 1, HHSEARCH, Neff-PPAS, pGenTHREADER, and wdPPAS threading programs. The result of I-TASSER was submitted to TM-align for measurement of TM-score and RMSD value for each of the mutant models specified in table 5. Chimera 1.11 was used to show the structure of proteins and to analyze the molecular characterization shown in Figure 2.

Table 5: TM-score and RMSD values of 3 selected nsSNPs

SNP ID	Residual Change	TM Score	RMSD Values
rs759928495	Q118R	0.95526	1.41
rs777784952	R129C	0.95707	1.39
rs752310933	L181V	0.96152	1.28
rs1197820694	G193R	0.93910	1.67

0.0 < TM-score < 0.30, random structural similarity 0 ± 0.3 and $0.5 < \text{TM-score} < 1.00$, in about the same fold 0.5 ± 1

Predicted PTMs (Post Transcriptional Modifications)

Protein structures and function are regulated by PTMs that have been shown to be involved in cell signaling and protein-protein interactions such as key events in biological systems (Dai and Gu, 2010; Shiloh and Ziv, 2013). In our analysis we examined whether the chosen nsSNPs displays some alteration of BCL2 protein throughout PTMs. Hence different bioinformatics tools were used to model the PTM locations in this protein.

Methylation: Methylation is an essential PTM as it influences DNA binding and alters gene expresses in certain proteins, lysine residues when methylated. For this reason,

GPS-MSP 3.0 was used for this purposes, BCL2 no sites would be methylated.

Phosphorylation:

GPS 3.0 and NetPhos 3.1 predicted BCL2 protein phosphorylation sites. GPS 3.0 estimates that 34 residues (Serine 43 %, Thr:34% and Tyr:23%) will have the target for phosphorylation. On the other side, Netphose 3.1 had estimated 25 residues (Ser:14, Thr:09, Tyr:02) that would be able to get phosphorylated. The findings of GPS 3.0 and NetPhos 3.1 are shown in the table to compare and common residues are given in table in fig 6....

Table 6: Prediction of phosphorylation sites in BCL2 protein using GPS3.0 and NetPhos3.1.

Position	Peptide	GPS 3.0			NetPhos3.1	
		Kinase	Score	Cutoff	kinase	Score
S70	RDPVARTSPLQTPAA	AGC/PKA/PKACA	3.24	3.21	Unsp	0.868
T7	*MAHAGRTGYDNREI	AGC/DMPK/GEK/DMPK	7.392	3.509	Unsp	0.550

Ubiquitination:

For ubiquitination prediction BDM-PUB and UbPred were used. BDM-PUB predicted lysine residues to get ubiquitinated, while UbPred predicted four of the lysine residues to get ubiquitinated

Table 7: Prediction of Ubiquitination sites in BCL2 protein using BDM-PUB and UbPred

Peptide	Position	BDM-PUB Score (Threshold 0.3)	UbPred Score (Threshold 0.62)
DFSWLSLKTLLSLAL	218	0.97	No
LGAYLGHK*****	239	4.89	No

GeneMANIA and STRING were used to model activity between the BCL2 and other genes. Analysis of GeneMANIA revealed that BCL2 physical associations are with FKBP8 BBC3 BECN1 TMBIM6 BAK1 BAX BAD VDAC1 BID BCL2L11 BNIP3 BIK BAG1 PMAIP1 NR4A1 and BNIP3. The co-expressions of BCL2 are BNIP3 PMAIP1 and BECN1. Bcl2 have no co-localization with any other gene. It has a connection to BAG2 PAMIP1 BCL2L11 AIFM1 BBC3 FKBP8 NR4A1 BID BAD BIK BAX BAK1 BOK in pathways. Genetic relationships are with ITLN1 and with BIK BID BCL2L11 BAX BAK1 and BOK common protein domain. Chain projections showed combined score for each gene. It is clear from these projections that TP53 is the most responsive gene with BCL2 accompanied by geneMANIA and STRING expected gene interactions as in Fig 5 and 6 respectively.

Discussion:

The protein encoded by *BCL2* gene play a very important role in controlling cell death mechanisms such as apoptosis, necrosis and autophagy. The gene product act as apoptotic regulator by controlling mitochondrial apoptosis, normal embryonic cells growth and protection from cancer. *BCL2* is one of the major factors responsible for 50% of all human cancers such lymphocytic leukemia, acute myeloid leukemia, follicular lymphoma (FL), prostate cancer, breast cancer, lung cancer, neuroblastoma, colorectal cancer etc. The aim of this study is to identify the potential SNPs in *BCL2* gene

area which maybe have effects on regulatory regions of the gene and its protein product. From the former analysis we predict that the analyze SNPs have significant role in above mention diseases.

The analysis of total SNPs in a selected region reveal that 159 were nsSNPs, while 398 and 1842 were contained in 5' UTR and 3' UTR regions respectively. The remaining SNPs were of certain forms (nonsense, uncategorized, intronic etc.) and were not included in the analysis. This data suggests that a very small number of SNPs are exposed to BCL2 protein to directly affect their protein. Observing all nsSNPs in a given region these four, Q118R (rs759928495), G193R (rs1197820694), R129C (rs777784952) and L181V (rs752310933) were shown to be the most damaging. PROVEAN expected G193R with the highest score of -7.039 while L181V had projected the lowest score of -2.706 among these four nsSNPs. PolyPhen2 predicted nsSNPs on a scale of 0-1 which included Q118R, R129C, L181V and G193R with exact 1 ranking. PhD-SNP and SNP&GO forecast, G193R as the most detrimental with the highest probability score of 10 and 0.992.

We have cross-checked these nsSNPs on several additional tools such as CADD, REVEL, Mutation Assessor and MetaR in Ensemble genome explorer 96 (Accessed: 25th June 2019). CADD, REVEL and MetaR expected that all three mutations (out of 4 nsSNPs) would be deleterious and harmful. Mutation Assessor estimated that the selected four mutations would be detrimental. Out of these CADD ratings varied from 29 to 34. G193R, Q118R, L181V and R129C is 33, 26, 29 and 34 respectively in which G193R and R129C having score 33 and 34 are consider to be expected deleterious. (CADD score of 20 indicates it is among 1 percent of the most damaging SNPs in the human genome and CADD value 30 means it is among the most damaging of 0.1 percent and so on).

MutPred1.2 system predicts impact dependent on many characteristics such as methylation loss, acetylation increase, altered organized interface, altered disordered interface, and intrinsic condition benefit. Mutation Q118R has the minimum P-value of and 0.487 and R129C, L181V with P-values of 0.383 and 0.713 respectively. Mutations G193R, have the highest P-values of 0.893. These findings suggest that these nsSNPs can have major effects on the structure and function of the BCL2 protein.

I-Mutant provided us results predicting the influence of nsSNPs with index of reliability (RI) values for protein stability prediction. The RI values are from 0-10 where 0 indicates zero reliability while 10 implies full reliability. Three SNPs lead to three G193R, Q118R, L181V and R129C amino acid substitutions expected to decrease protein but our result was not projected to improve SNPs protein stability. The G193R, Q118R, L181V and R129C reliability indexes were strong (see Table 3). But we are keeping our outcomes so predictable that we have tested these mutations in the CUPSAT registry. (<http://cupsat.tu-bs.de/>) further. CUPSAT server predictions is entirely in line with I-Mutant predictions.

ConSurf projected BCL2 protein conservation profile providing prediction to maintain, embedded, functionally or structurally significant growing amino acids. Based on the role of amino acids in protein, those that were highly conserved were expected to be either functionally or structurally significant (Berezin et. al., 2004). Highly conserved and exposed amino acids are expected to have many important functions such as associations in protein activities. Our data were based on our 3 selected nsSNPs G193R, Q118R and R129C which are highly conserved, exposed and functionally essential. The L181V, which is highly conserved, buried and structurally significant. This further reinforces certain nsSNPs' deleterious effect on BCL2 protein.

The protein structures were modelled using I-TASSER. Only FASTA sequence of protein was used as a data. The method itself does prototype collection and protein simulation. It used 2nbiA and 2co9 models with a range of 85 percent and 83 percent, respectively. Using Ramachandran Plot Analysis (RAMPAGE server) (Lovell et al., 2002) we measured R AMPAGE values for the modelled protein structures (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). Wild-type BCL2 protein had a RAMPAGE value of 92.8%

preferred and allowed for residues and 7.2% for outlier residues. For the Q118R mutant structure, the preferred and allowed residues were 92.8% and 7.2% of the outer residues. The preferred and allowed residues of R129C, L181V, and G193R were 91.5%, 95% and outlier values are 92.4%, 8.4%, 5.1% and 7.6%, respectively. The configurations of proteins will be regarded as stronger structures if their RAMPAGE values surpass 80 percent (Morris et al, 1992).

These four replacement amino acids have Root mean square deviation values which indicate that these nsSNPs are the most damaging SNPs in BCL2 protein. nsSNP Q118R (rs759928495) has an RMSD (1.41Å) of 0.95526 TM score. R129C (rs777784952) has RMSD Value (1.39 Å) with 0.95707 TM Score, L181V (rs752310933) RMSD Value (1.28Å with 0.96152 TM Score and G193R (rs1197820694) RMSD Value (1.67Å) 0.93910 TM Score such nsSNPs will cause considerable harm to the BCL2 protein.

Post Transcriptional Modifications (PTMs) are key factors that guide proteins to conduct other important functions such as PPIs (Protein-Protein interactions) and Cell signaling (Dai and Gu 2010; Shiloh and Ziv 2013). We also searched for PTM locations in BCL2 protein whether or not it had possible PTMs at such places in nsSNPs. Interestingly, none of the phosphorylation and ubiquitination sites have been shown to be at the most destructive location of nsSNPs.

We looked for other nsSNPs that identified as GPS3.0 34 and NetPhos3.1 at 25 positions expected to be phosphorylated (Table 4). At two possible phosphorylation sites predicted by GPS3.0 34 and NetPhos3.1 we find one nsSNPs at amino acid positions 7. Only one nsSNPs was also found at amino acid location at the two possible ubiquitination sites expected by BDM-PUB and UbPred at amino acid position 17.

Our study was conducted in depth, and all findings were cross-checked in order to avoid confusion. Every research has some drawbacks and so does ours. While our analysis was in depth, it is focused on web servers and computational software powered by mathematical and statistical algorithms that need experimental examination for validation.

Conclusion

Our study found bcl2 gene SNPs that could be vital in nearly 50 percent of human cancers. This study shows that different nsSNPs can disturb the structure and/or function of bcl2 protein. Four major mutations were found in the native protein bcl2 gene: glutamine! Arginine (rs759928495), glycine at position 118! Position 193 Arginine (rs1197820694), Arginine! Position 129 Cysteine (rs777784952), Leucine! The most damaging of all predicted were valine L181V (rs752310933). These SNPs can have deleterious effects on bcl2 protein and can play a key role in diseases. These four nsSNPs can therefore be strongly regarded as key candidates. It will contribute towards effective drug discovery and the development of accurate medicines when causing diseases related to bcl2 malfunctions. Careful investigations and experiments in the wet laboratory are necessary for exploring the structure and function of the protein effects of these polymorphisms. Different animal models that consist of these important mutations in a bcl2 protein can also be extremely helpful in exploring in this disease. Our study can help in personalized medicines that are designed for various types of cancer patients and can also help in future cancer prediction and can help in cancer diagnosis with family history of cancer. Our study was carried out in detail, however it is in silicon study so that wet laboratories are required to study cell cultures and animal models containing these SNPs.

Declarations

Conflict of Interest

The authors have declared no conflict of interest.

References

- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7: 248±249.
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016; 8: W344±350.
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010; 38: W529±533.
- Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, et al. ConSeq: The Identification of Functionally and Structurally Important Residues in Protein Sequences. *Bioinformatics*. 2004; 20: 1322±1324.
- Bhagwat M. Searching NCBI's dbSNP database. *Curr Protoc Bioinformatics*, Chapter 1: Unit 1. 2010;19.
- Blom N, Gammeltoft S, Brunak S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*. 1999; 294: 1351±1362.
- Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006; 22: 2729±2734.
- Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*. 2013; 3: S6.
- Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*. 2005; 33: W306±310.
- Celniker G, Nimrod G, Ashkenazy H, Glaser F, Martz E, Mayrose I, et al. ConSurf: Using Evolutionary Data to Raise Testable Hypotheses about Protein Function. *Israel Journal Of Chemistry*. 2013; 53: 199±206.
- Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One*. 2012; 7: e46688.
- Ciesla J, Fraczyk T, Rode W. Phosphorylation of basic amino acid residues in proteins: important but easily missed". *Acta Biochim Pol*. 2011; 58: 137±147.
- Dai C, Gu W. p53 post-translational modification: deregulated in tumorigenesis. *Trends Mol Med*. 2010; 16: 528±536.
- Deutscher J, Saier MH. Ser/Thr/Tyr protein phosphorylation in bacteria for long time neglected, now well established. *J Mol Microbiol Biotechnol*. 2005; 9: 125±131.
- Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. *ExpPASy: the proteomics server for in-depth protein knowledge and analysis* *Nucleic Acids Res*. 2003; 31:3784-3788.
- Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015; 10: 845±858.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4: 1073±1081.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. 2009; 25: 2744±2750.

Lovell S.C, Davis I.W, Arendall III W.B, de Bakker P.I.W, J.M. Word, M.G. Prisant, J.S.Richardson and D.C. Richardson. 2002. Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*. **50**: 437-450.

Morris A L, MacArthur M W, Hutchinson E G and Thornton J M. 1992. Stereochemical quality of protein structure coordinates. *Proteins* 12:345-364.

Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function. *Annu RevGenomics Hum Genet*. 2006; 7: 61±80.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF ChimeraDa visualization system for exploratory research and analysis. *J Comput Chem*. 2004; 13:1605±1612.

Puttick J, Baker EN, Delbaere LT. Histidine phosphorylation in biological systems. *Biochem Biophys Acta*. 2008; 1784: 100±105.

Radivojac P, Vacic V, Haynes C, Cocklin RR, Mohan A, Heyen JW, et al. Identification, Analysis and Prediction of Protein Ubiquitination Sites. *Proteins*. 2010; 78: 365±380.

Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: Server and survey. *Nucleic Acids Res*. 2002; 30: 3894±3900.

Sawicka A, Seiser C. Sensing core histone phosphorylationDa matter of perfect timing. *Biochim Biophys Acta*. 2014; 1839: 711±718.

Shiloh Y, Ziv Y. The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol*. 2013; 14: 197±210.

[Wankun Deng Yongbo Wang Lili Ma Ying Zhang Shahid Ullah Yu Xue](#). Computational prediction of methylation types of covalently modified lysine and arginine residues in proteins. *Briefings in Bioinformatics*, 2017; 18(4); 647–658,

Warde-Farley D, Donaldson SL, Comes O, et al.: The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*. 2010; 38(Web Server issue): W214–20.

Xue Y, Ren J, Gao X, Jin C, Wen L, Yao X. GPS 2.0, a tool to predict kinase-specific phosphorylation sites in hierarchy. *Mol Cell Proteomics*. 2008; 7: 1598±1608.

Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*. 2015; 12: 7±8.

Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM- score. *Nucleic Acids Res*. 2005; 33: 2302±2309.

Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008; 9:40.

Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res*. 1998; 8: 1229±1231.

Lander ES. The new genomics: global views of biology. Science. 1996; 274: 536±539.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics. 2009; 25: 2744±2750.

Capriotti E, Altman RB. Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics. 2011; 12: S3.

Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, et al. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature. 1999; 402: 880±883.

Chasman D, Adams RM. Predicting the functional consequences of nonsynonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation. J Mol Biol. 2001; 307: 683±706.

Kucukkal TG, Petukh M, Li L, Alexov E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. Curr Opin Struct Biol. 2015; 32: 18±24.

Thomas R, McConnell R, Whittacker J, Kirkpatrick P, Bradley J, Sandford R. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. Am J Hum Genet. 1999; 65: 39±49.

Figures

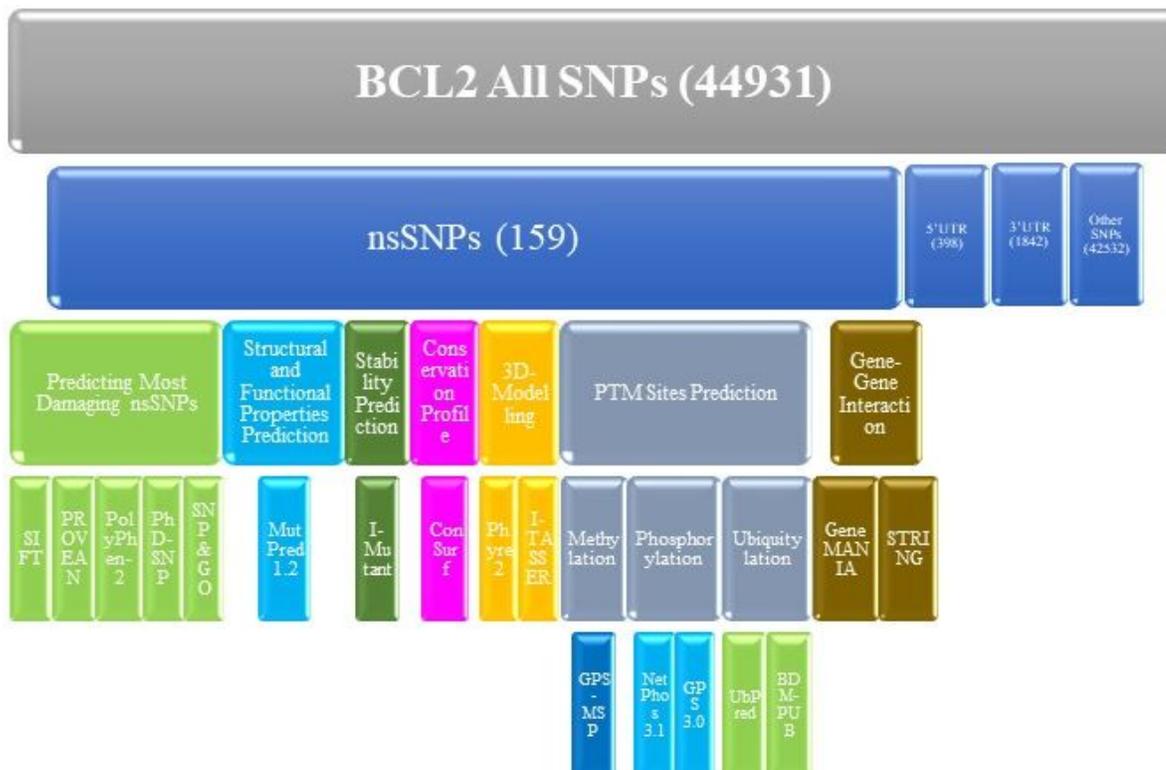


Figure 1

A schematic flow of the Methodology.

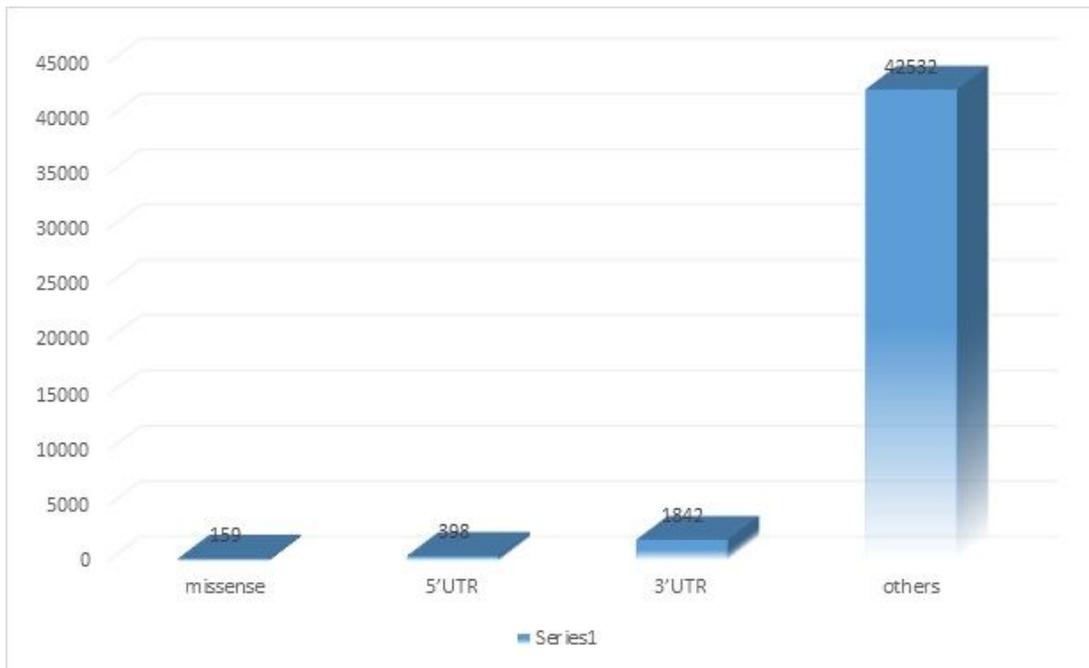


Figure 2

All SNPs in BCL2 gene

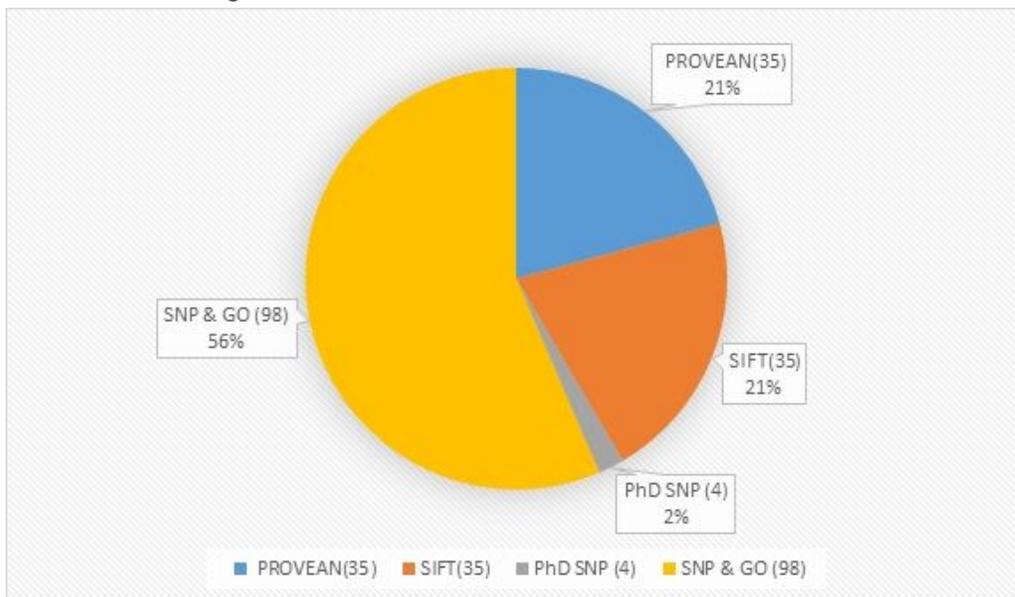


Figure 3

Pie chart showing percentage of nsSNPs predicted by PROVEAN, SIFT, PhD SNP and SNP&GO.

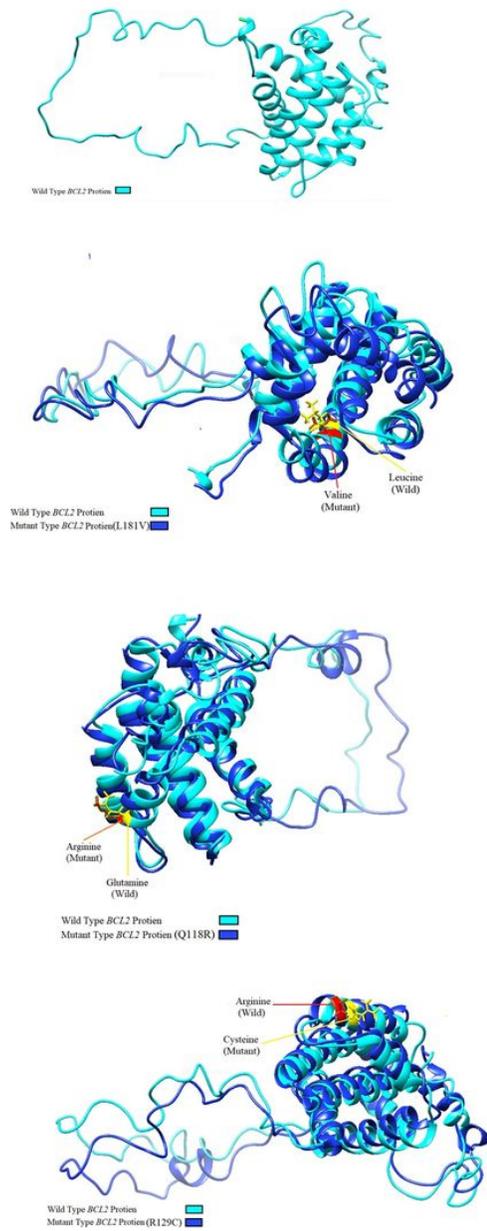


Figure 4

A). Wild type BCL2 Protein Structure. B) Superimposed Structure of BCL2 and its G193R mutant. C) Superimposed structure of BCL2 and its Q118R mutant. D) Superimposed structure of BCL2 and its R129C mutant. E) Superimposed structure of BCL2 and its L181V mutant.

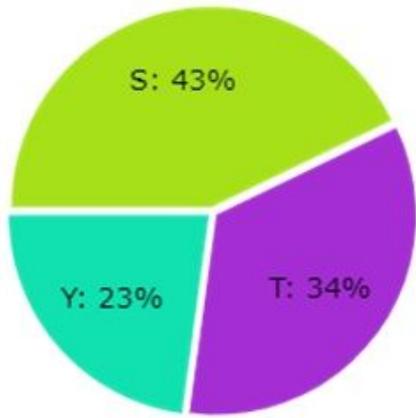


Figure 5

Graphical presentation of phosphorylation site (Serine 43 %, Thr:34% and Tyr:23%) by GPS 3.0

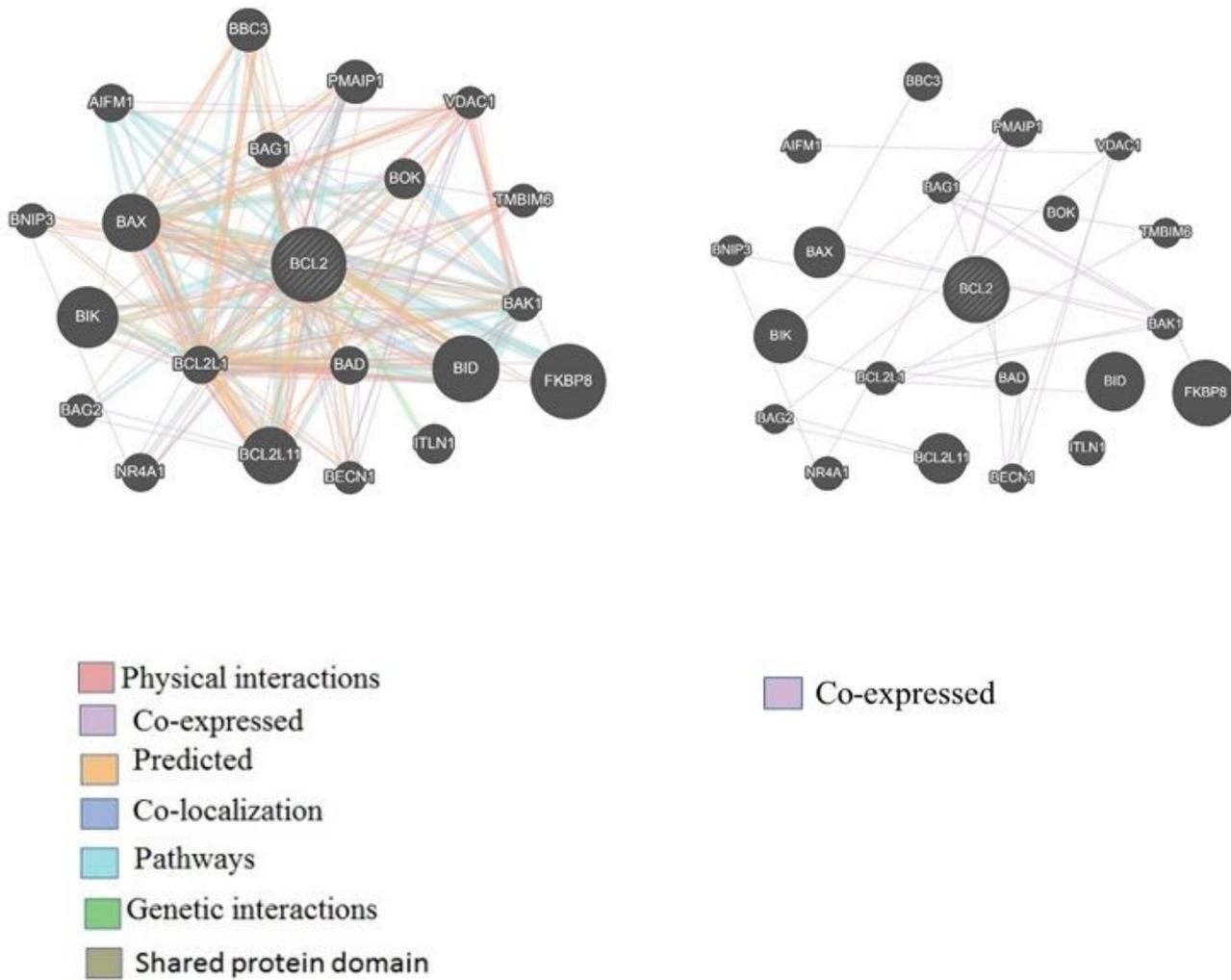


Figure 6

Gene-gene Interaction of BCL2 with other genes proposed by GeneMANIA

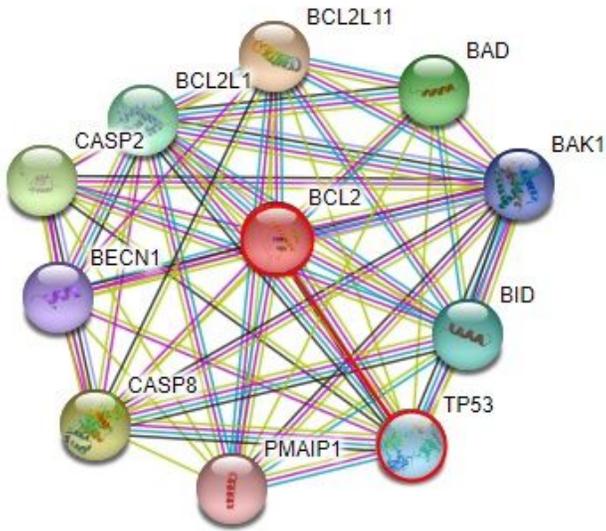


Figure 7

Gene-gene Interaction of BCL2 with other genes proposed by STRING

Supplementary Files

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

- [GraphicalPresentation.png](#)