Journal of Medical Systems (2019) 43: 270
https://doi.org/10.1007/s10916-019-1328-z

IMAGE & SIGNAL PROCESSING

NBN Gene Analysis and it's Impact on Breast Cancer

P. Nithya [1] · A. ChandraSekar [1]

Received: 8 March 2019 / Accepted: 7 May 2019 / Published online: 5 July 2019

Abstract
Single Nucleotide Polymorphism (SNP) researches have become essential in finding out the congenital relationship of structural deviations with quantitative traits, heritable diseases and physical responsiveness to different medicines. NBN is a protein coding gene (Breast Cancer); Nibrin is used to fix and rebuild the body from damages caused because of strand breaks (both singular and double) associated with protein nibrin. NBN gene was retrieved from dbSNP/NCBI database and investigated using computational SNP analysis tools. The encrypted region in SNPs (exonal SNPs) were analyzed using software tools, SIFT, Provean, Polyphen, INPS, SNAP and Phd-SNP. The 3'ends of SNPs in un-translated region were also investigated to determine the impact of binding. The association of NBN gene polymorphism leads to several diseases was studied. Four SNPs were predicted to be highly damaged in coding regions which are responsible for the diseases such as, Aplastic Anemia, Nijmegan breakage syndrome, Microsephaly normal intelligence, immune deficiency and hereditary cancer predisposing syndrome (clivar). The present study will be helpful in finding the suitable drugs in future for various diseases especially for breast cancer.

Keywords NBN · Single nucleotide polymorphism · Double strand breaks · nsSNP · Associated diseases

Introduction

NBN (Nibrin) is a protein coding gene, it is also known as NBS1, Cell cycle regulatory Protein P95, is situated on chromosome 8 (8q21 in *Homo sapiens*. It consists of 754 amino acids which are members of MRN/NMR (NBN/Mre11/RAD50) complex also called Double strand DNA break complex regulates cellular reaction to DNA breakdown and maintenance of Chromosomal stability. Nibrin protein controls the activity of this complex by carrying the MRE11A and RAD50 proteins to the site of DNA damage through the nucleus of the cell. The MRE11A/RAD50/NBN has a more complex structure due to its interaction with large proteins formed from the ATM gene which is highly essential in identifying damaged strands of DNA and facilitating their repair [1].

Genetic deviations in Humans may happen in various nucleotide constitutions, including single nucleotide polymorphisms (SNPs) and organizational deviations such as minor additions and cancellation (indels) or large alterations in different copies of the same segment. Among these, SNPs are the wide prevalent form of human variation and it has been evaluated and estimated that one SNP exists every 290 base-pairs in the human genome [2]. The documented parameters of a protein, its organization and role can be affected by a single base replacement, insertion or deletion. SNP takes place in two regions such as Coding regions and Non coding regions. Non coding SNPS are Frame shift, Intronic SNPs, UTR. Synonymous SNPs are Single base alteration in coding regions of DNA which does not lead to amino acid changes. Non synonymous SNPs are single base changes in encryption zones of DNA which later leads to modifications in amino acid have the potential to influence of the protein structure and function.

Nijmegen breakage syndrome, which is classified as an autosomal recessive chromosomal instability syndrome,

This article is part of the Topical Collection on *Image & Signal Processing*

✉ P. Nithya
nithyaraju.r@gmail.com

A. ChandraSekar
drchandrucse@gmail.com

[1] CSE Department, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India

typically identified by mutations in the NBN gene is associated with microcephaly, retardation and stunted growth, immunodeficiency, and a predisposition to cancer. MRE11/RAD50 is an encoded protein and a strong component of double-strand break repair complex comprising of 5 proteins. The gene product is speculated to be present in DNA double-strand break repair and DNA damage-induced checkpoint activation [3]. It is a theory that DNA damage accumulates over a period of time, which can induce triggers in cells to multiply uncontrollably, giving rise to the risk of developing cancer [4]. The main objective is to investigate the change in amino acid substitution of NBN genes and their impact on various diseases by predicting and measuring the damaging effect on amino acid, based on Sequence and Structural level and analyzing changes in protein stability due to mutation in NBN genes.

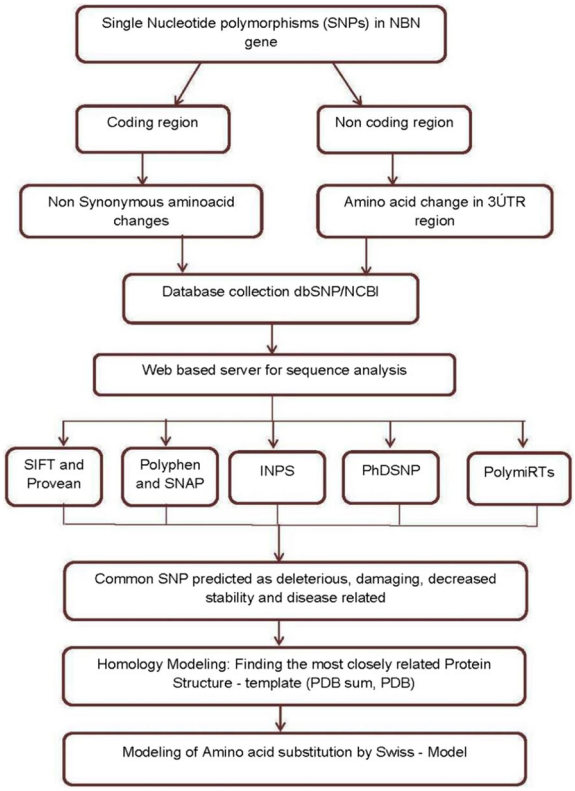
Methods and materials

Tools and databases used

The online tools and databases were used for data mining process, to find various validation score, structure prediction and validating the structure. The databases Genecards, dbSNP & UNIPROT for data mining, the online tool for finding the deleterious SNP's is SIFT, PROVEAN for obtaining various scores INPS & SNAP the online tools for forecasting the protein structure is done in Swiss Model the validation of the structures is done with PDB Sum. Pymol offline tool was used to induce mutagenesis. Function, interaction and Network of NBN gene was done by GENEMANIA [5]. Retrieval of gene Id and uniprot Id was done by GENECARDS: (Gene cards Version 3: The Human Gene Integrator [5].UNIPROTKB: (Universal Protein Knowledgebase Resource) (The Uniprot Consortium 2015).

Retrieval of synonymous and non-synonymous SNP DBSNP dbSNP is hosted by NCBI [6]. The rs ID for NBN-HUMAN genes are obtained from dbSNP. Predictions of Damaging Amino acid Substitution by SIFT (Sorting Intolerant from Tolerant) [7] Tolerated if the score is found to be higher than 0.05: Deleterious if the score is less than 0.05. PROVEAN (Protein Variation Effect Analyser) [8]. Estimate the functional modification in deleterious *nsSNP* by Polyphen-v2 (polymorphism phenotypingv2) [9]. SNAP (Screening of non-acceptable polymorphism) [10, 11]. Vijayakumar et.al proposed various data classification methods for data analysis [12, 13]. Prediction of change in stability due to mutation by INPS (Impact of nonsynonymous mutations on protein stability) (http://inpsmd.biocomp.unibo.it) [14]. Association of nsSNP to disease by PhDSNP (Prediction of Deleterious Single Nucleotide Polymorphism) [15].

WORK FLOW



Result and discussion

GENEMANIA

GENEMANIA is a networked server providing the functionality and interactivity of the NBN gene predominantly on data sets of genomics and proteomics with a massive data center and very high precision rates (Fig. 1).

Above Table 1 represents the functions of NBN gene such as DNA damage, Double-stranded break repair, chromosomal integrity maintenance, cell cycle checkpoint control and meiosis.

Above Table 2 represents gene and its descriptions are mentioned.

Database collection from DbSNP

SNP analysis

GENECARDS It is available as an integrated database for searchable comprehensive and convenient information access on all annotated and forecast human genes from this database. Entrez gene ID: 4683 is obtained for NBN gene.

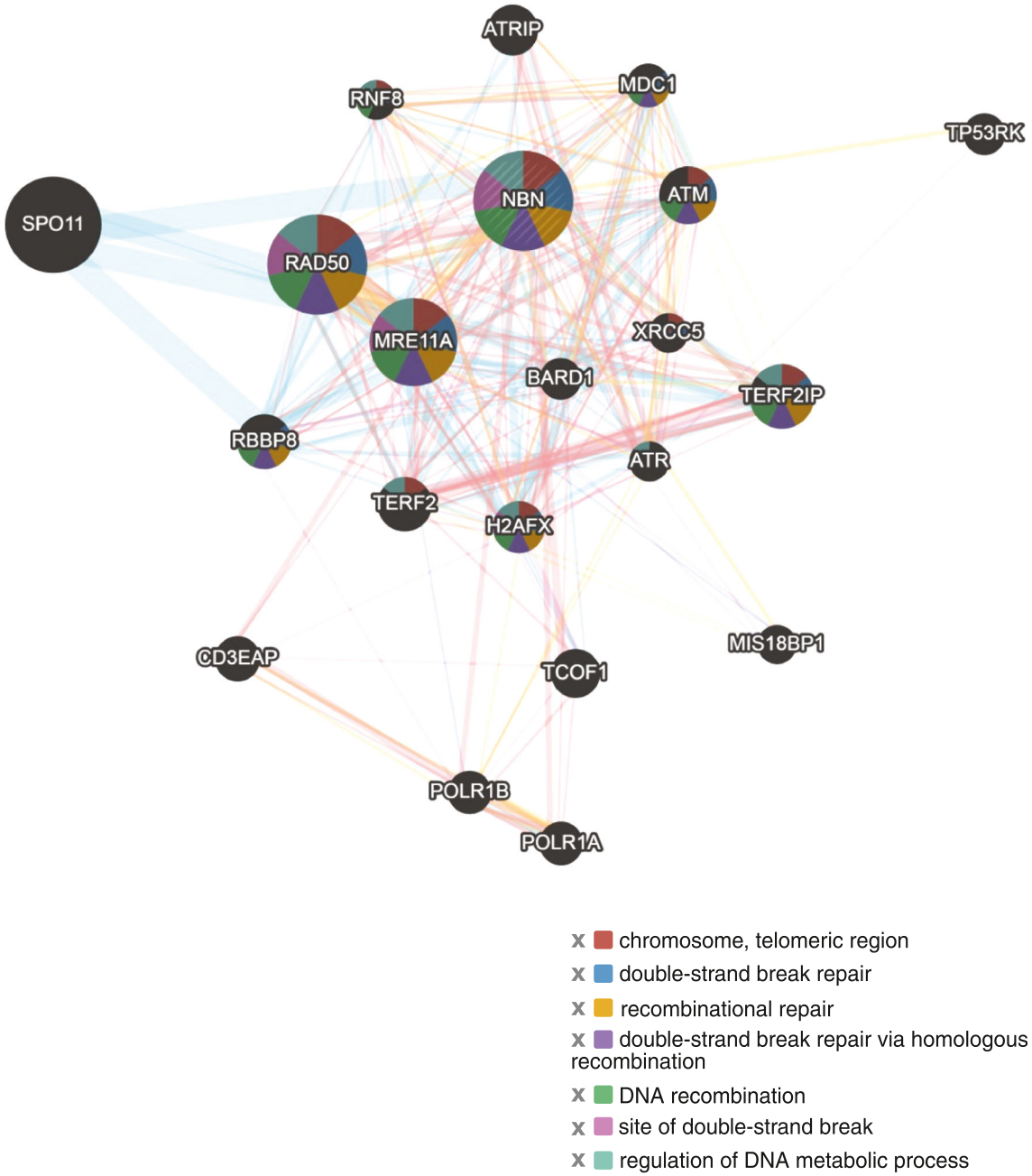


Fig. 1 Represents the function, interaction and network of NBN genes

UNIPROT It is a database with a large selection of freely accessible protein sequence and functional data with items being obtained from ordered genome databases. The Uniprot id for NBN-HUMAN gene is 060934. The fasta sequence for NBN gene was establish from UniprotKB database. It comprises of 754 amino acids.

>NP_002476.2 nibrin isoform 1 [*Homo sapiens*]

MWKLLPAAGPAGGEPYRLLTGVEYVVGRKNCAIL
IENDQSISRNHAVLTANFSVTNLSQTDEIPVLTLKD
NSKYGTFVNEEKMQNGFSRTLKSGDGITFGVFGS
KFRIEYEPLVACSSCLDVSGKTALNQAILQLGGFTV
NNWTEECTHLVMVSVKVTIKTICALICGRPIVKPEY
FTEFLKAVESKKQPPQIESFYPPLDEPSIGSKNVDL
SGRQERKQIFKGKTFIFLNAKQHKKLSSAVVFGGG
EARLITEENEEEHNFFLAPGTCVVDTGITNSQTLI
PDCQKKWIQSIMDMLQRQGLRPIPEAEIGLAVIFMT
TKNYCDPQGHPSTGLKTTTPGPSLSQGVSVDEKL
MPSAPVNTTTYVADTESEQADTWDLSERPKEIKV
SKMEQKFRMLSQDAPTVKESCKTSSNNNSMVSN
TLAKMRIPNYQLSPTKLPSINKSKDRASQQQQTNS
IRNYFQPSTKKRERDEENQEMSSCKSARIETSCSL
LEQTQPATPSLWKNKEQHLSENEPVDTNSDNNL

Table 1 Represents the funtions of NBN gene and their interaction with other gene

Function	Role of NBN with other genes
DNA damage	NBN renders modulation of the DNA damage-signal sensing by employing PI3/PI4-Kinase family memebers (ATM, ATR) and at times DNA-Pkcs to the DNA damage sites and activating their functions.
Double-strand break repair(DSBs)	NBN employs MRE11 and RAD50 to the proximity of DSBs by an interaction with the H2AFX.NBN with RBBP8 links DNA double strand break sensing to resection.
Maintenance of chromosomal integrity	NBN also performs telomere length maintenance by establishing the overhang serving as a primer of telomerase dependent telomere elongation.
Cell cycle checkpoint control and meiosis	NBN is a huge partaker in the control of intra-s-phase checkpoint and NBN is involved in G1 and g2 checkpoint.

FTDTDLKSIVKNSASKSHAAEKLRSNKKREMDDV AIEDEVLEQLFKDTKPELEIDVKVQKQEEDVNVRK RPRMDIETNDTFSDEAVPESSKISQENEIGKKREL KEDSLWSAKEISNNDKLQDDSEMLPKKLLLTEFRS LVIKNSTSRNPSGINDDYGQLKNFKKFKKVTYPGA GKLPHIIGGSDLIAHHARKNTELEEWLRQEMEVQN QHAKEESLADDLFRYNPYLKRRR.

DBSNP The Single Nucleotide Polymorphism Database (dbSNP) is a genetic data center hosting information on genetic variation with the same and different species available to the public free of cost. It was developed and put up by the National Center for Biotechnology Information (NCBI) in collaboration with the National Human Genome Research Institute (NHGRI) (Fig. 2).

Using the dbSNP database, the SNPs of the NBN gene was analysed. The existence of human NBN gene contained a total number of 14,469 SNPs; from that 4067 were found in *Homo sapiens*, which 465 where missense, 209 were coding synonymous, 21 were nonsense, 2835 were in the intronic,149 in the 3'UTR, 369 were in 5'UTR, 60 were in frameshift,14 in 3′ spice site and 16 in the 5'spice site.

Prediction of damaging effects of NSSNP

SIFT analysis

The forecasting is carried out by observing the extent of safekeeping of each amino acid leftover of the specific query sequence, and in order to do so, SIFT establishes a concise compilation of similar protein sequences by traversing the protein databases TrEMBL and UniProt with the aid of the PSI-BLAST algorithm, the query sequence and builds an arrangement of the found sequences.

After the completion of the first step, a regularized probability for each substitution at the corresponding specific place in the arrangement is evaluated and recorded in the scaled probability matrix. This score is also called a SIFT score and substitutions are **Tolerated if the score is greater than 0.05**; and considered **Deleterious if the score is less than 0.05.**

The SIFT approach banks on the assumption that a largely conserved point is found to be intolerant to most substitutions, while a poorly conserved position is found to tolerate most substitutions.

PROVEAN analysis

PROVEAN is comprised of two major steps; the first one involves the assimilation of both alike and sequences of distant relations from the NCBI NBN protein database with the help of BLASTP. To eliminate unwanted repetitions, the composed sequences are clustered, based on the sequence identity.

The second step consists of computing a delta score for each specific sequence in the supporting sequence set, using the BLOSUM62 substitution matrix. For every individual cluster, an average delta score is evaluated and the averaged delta scores are again averaged for all clusters. This analysis results in an unbiased averaged delta score that is the final PROVEAN score.

Table 2 Represents the gene and their description

Gene	Description
NBN	Nibrin
MRE11A	MRE11 meiotic recombination 11 homolog A
RAD50	RAD50 homolog
H2AFX	H2A histone family.member of X
ATM	Ataxia telangiectasia mutated
RBBP8	Retinoblastoma binding protein 8
ATR	Ataxia telangiectasia mutated and Rad3 related
TERF2	Telomeric repeat binding factor 2

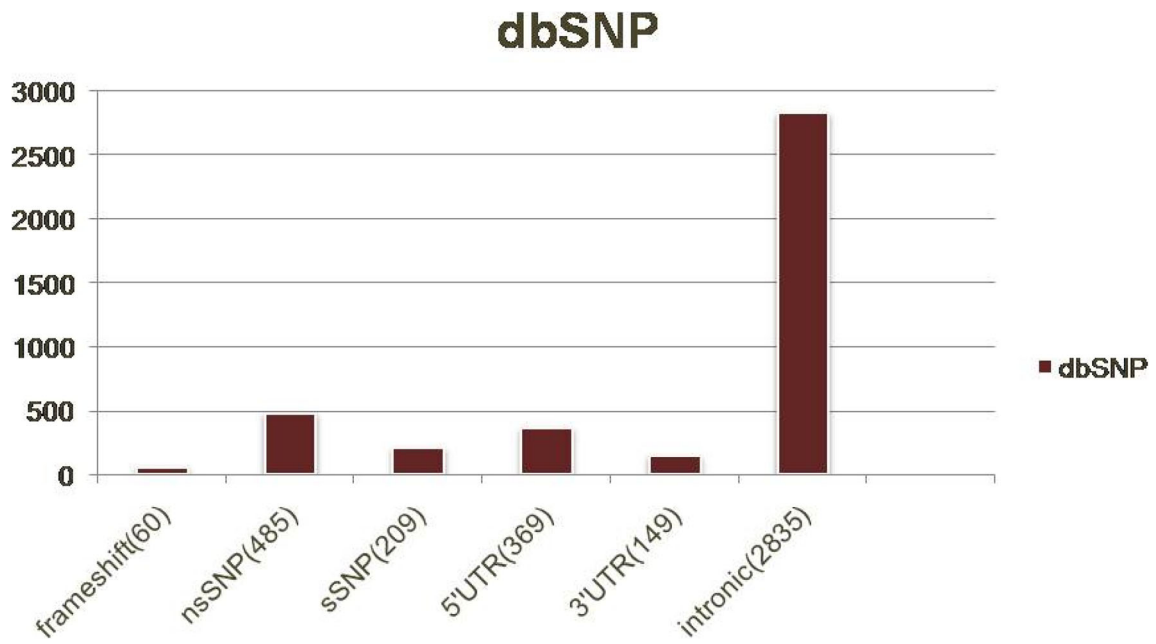


Fig. 2 Distribution of SNPs NBN gene based on dbSNP databases

Table 3 Represents SIFT and PROVEAN result

rs ID	Amino acid change	Nucleotide change	SIFT		PROVEAN	
			Prediction	Score	Prediction	Score
rs769420	**P266L**	**A/G**	**Deleterious**	**0.003**	**Deleterious**	**−5.692**
rs1805794	E185Q	C/G	Tolerated	1	Neutral	0.525
rs3026268	T497A	C/T	Tolerated	0.337	Neutral	−0.934
rs34120922	K408E	A/G	Tolerated	0.58	Neutral	−0.130
rs34767364	R215W	C/G/T	Tolerated	0.138	Deleterious	−4.100
rs61754967	T253I	A/G	Tolerated	0.21	Neutral	−2.433
rs61754966	I171V	C/G/T	Deleterious	0	Neutral	−0.795
rs61754796	V210F	G/T	Deleterious	0.02	Neutral	−2.406
rs61753720	**D95N**	**C/T**	**Deleterious**	**0.04**	**Deleterious**	**−4.3**
rs61753718	R215Q	C/T	Tolerated	0.481	Neutral	−1.3
rs61612852	A265V	A/G	Tolerated	0.099	Neutral	−2
rs769414	N142S	C/T	Tolerated	1	Neutral	−1.022
rs6413508	P672L	C/T	Tolerated	0.687	Neutral	−0.604
rs12721593	**S93 L**	**C/T**	**Deleterious**	**0.025**	**Deleterious**	**−3.083**
rs28538230	I439M	C/T	Tolerated	0.16	Neutral	−0.079
rs72550742	E564K	A/G	Deleterious	0.037	Neutral	−1.439
rs72563785	N716D	A/G	Tolerated	0.1	Neutral	−1.716
rs78870221	I35M	A/C/T	Deleterious	0.01	Neutral	−1.749
rs104895031	D527Y	G/T	Deleterious	0.037	Neutral	−2.334
rs104895032	L421S	C/T	Tolerated	0.096	Neutral	−0.789
rs104895033	P401R	A/C/G/T	Tolerated	0.119	Neutral	0.296
rs115321485	E628K	C/T	Tolerated	1	Neutral	−3.089
rs769416	Q216K	G/T	Tolerated	1	Neutral	−0.713
rs13312858	**K105 N**	**A/T**	**Deleterious**	**0.22**	**Deleterious**	**−3.1**
rs111239312	D519Y	A/C	Deleterious	0.016	Neutral	−2.384
rs111244949	L133S	A/G	Tolerated	0.382	Neutral	−1.497
rs112524180	F603 V	A/C	Tolerated	0.493	Neutral	−0.323

Thus, the overall effect of alterations on a protein's role can be estimated accurately as the difference in this arrangement score, the delta score.

Less delta scores imply that the changes have more of a destructive effect on the protein's function and the default threshold of delta score < = −2.5 will have a detrimental effect. Whereas increased delta scores are given the interpretation as variations with neutral effect (default threshold of delta score > = − 2.5) is neutral.

Based on the SIFT score it is deleterious if the score is lesser than 0.05; tolerated if the score is more than 0.05. Out of 465nsSNP obtained from dbSNP, predicted SIFT results are represented in this Table 3 among that 17 nsSNPs were damaged i.e. deleterious and 9 nsSNP were tolerated.

Based on the PROVEAN score if the threshold score is less than or equal to −2.5 is deleterious; if the score is more than −2.5 is said to be neutral. Based on the threshold of delta score, among the 26 nsSNP, 4 were predicted as deleterious and 22 were neutral effect.

Based on the SIFT and PROVEAN score gained from table it is analyzed that 4nsSNP were said to be extremely damaged (rs769420), (rs61753720), (rs12721593) and (rs13312858).

Prediction of functional effect and stability

Polyphen analysis

The program performs a search for three-dimensional protein models, carries out several alignments of similar sequences and examinations in numerous protein structure databases from the data of amino acid contact.

Thus, position-specific independent count (PSIC) scores will be considered for each of the two amino acid residues after which the PSIC score transformation will be evaluated between them.

PSIC score is said to be damaging if the score is 1.0 (Probably damaging).

Table 4 Prediction of functional effect by Polyphen

rs ID	Amino acid change	Nucleotide change	POLYPHEN-V2	
			Prediction	Score
rs769420	**P266L**	**A/G**	**Probably damaging**	**1.000**
rs1805794	E185Q	C/G	Benign	0
rs3026268	T497A	C/T	Benign	0.007
rs34120922	K408E	A/G	Possibly damaging	0.722
rs34767364	**R215W**	**C/G/T**	**Probably damaging**	**1.000**
rs61754967	T253I	A/G	Possibly damaging	0.948
rs61754966	**I171V**	**C/G/T**	**Probably damaging**	**1.000**
rs61754796	V210F	G/T	Benign	0.215
rs61753720	**D95N**	**C/T**	**Probably damaging**	**1.000**
rs61753718	**R215Q**	**C/T**	**Probably damaging**	**0.999**
rs61612852	**A265V**	**A/G**	**Probably damaging**	**0.969**
rs769414	N142S	C/T	Benign	0.129
rs6413508	P672L	C/T	Benign	0
rs12721593	S93 L	C/T	Possibly damaging	0.744
rs28538230	I439M	C/T	Benign	0.015
rs72550742	E564K	A/G	Possibly damaging	0.657
rs72563785	N716D	A/G	Benign	0.142
rs78870221	**I35M**	**A/C/T**	**Probably damaging**	**0.961**
rs104895031	D527Y	G/T	Possibly damaging	0.788
rs104895032	L421S	C/T	Benign	0.381
rs104895033	P401R	A/C/G/T	Benign	0.030
rs115321485	E628K	C/T	Benign	0
rs769416	Q216K	G/T	Benign	0.003
rs13312858	**K105N**	**A/T**	**Probably damaging**	**1.000**
rs111239312	**D519Y**	**A/C**	**Probably damaging**	**0.989**
rs111244949	**L133S**	**A/G**	**Probably damaging**	**0.975**
rs112524180	F603 V	A/C	Benign	0.064

Table 5 Prediction of functional effect by SNAP and prediction of change in protein stability by INPS

rs ID	Position	Wild type amino acid	Variant type amino acid	SNAPPrediction effect	Score	INPS(DDG in kcal/mol)
rs769420	P266L	P	L	Effect	85	−0.945724
rs34767364	R215W	R	W	Effect	72	−0.6867
rs61754967	T253I	T	I	Neutral	−3	−0.20953
rs61754966	I171V	I	V	Effect	80	−0.82285
rs61754796	V210F	V	F	Effect	78	−0.170534
rs61753720	D95N	D	N	Effect	92	−0.461093
rs61753718	R215Q	R	Q	Effect	66	−0.989829
rs61612852	A265V	A	V	Neutral	−51	−0.707974
rs12721593	S93 L	S	L	Effect	4	−0.351289
rs72550742	E564K	E	K	Effect	17	−0.548225
rs72563785	N716D	N	D	Effect	6	−0.474765
rs78870221	I35M	I	M	Neutral	−12	−1.42311
rs104895031	D527Y	D	Y	Effect	24	−0.160905
rs104895032	L421S	L	S	Neutral	−2	−1.38732
rs13312858	K105 N	K	N	Effect	72	−0.818373
rs111239312	D519YL	D	Y	Neutral	−1	−0.297731
rs111244949	133S	L	S	Neutral	−32	−1.38732

PSIC score is said to be benign if the score is 0 to 0.05 (benign).

PSIC score is said to be possibly harmful if the score is 0.5 to 0.9 (possibly damaging).

These scores are mainly based on Sensitivity and Specificity.

From the above Table 4, the given 26nsSNP which are obtained from 12nsSNP, SIFT were considered to be 10nsSNP, Benign were considered to be Probably Damaging and 2nsSNP were considered to be possibly damaging.

SNAP analysis

The network computes a total value for every individual substitution and the values will be converted into binary predictions of neutral or non-neutral effect.

The reliability score (Reliability Index: RI) is calculated from the absolute prediction score. This measure is meant to simplify assessing prediction strength and to immediately convey the reliability of a prediction. The predicted values are shown in the table.

INPS analysis

INPS gives predictions on the thermionic free energy change brought about by the amino acid residue change found in protein sequences by employing the aid of two types of features that deal with the mutation type (molecular weight, mutability, hydrophobicity) and the hereditary details.

The thermodynamic free energy DDG is measured in kcal/mol has a constant value −0.581. The predicted values are shown in table.

From the above Table 5, the SNAP analysis is made among 17 deleterious nsSNP predicted from SIFT, 6 nsSNP were identified as a neutral effect (i.e.there is no damage based on

Table 6 Association of nsSNP to disease by PhdSNP

Position	Wild type protein	New amino acid type after mutation	Effect –PhdSNP	RI
266	**P**	**L**	**Disease**	2
215	**R**	**W**	**Disease**	5
253	**T**	**I**	**Disease**	0
171	I	V	Neutral	5
210	V	F	Neutral	1
95	**D**	**N**	**Disease**	3
215	R	Q	Neutral	0
265	A	V	Neutral	5
93	**S**	**L**	**Disease**	5
564	E	K	Neutral	0
716	N	D	Neutral	3
35	I	M	Neutral	4
527	**D**	**Y**	**Disease**	1
421	**L**	**S**	**Disease**	4
105	**K**	**N**	**Disease**	6
519	**D**	**Y**	**Disease**	1
133	**L**	**S**	**Disease**	3

functional effect) and 11 nsSNP were identified as a non-neutral effect (i.e there is a damage).

SNAP score is said to be neutral, if the value falls 0 or below 0.

The score is said to be non neutral, if the value is based on reliability index.

In INPS analysis protein stability is analysed for 17 deleterious nsSNP predicted from SIFT, (if the DDG score < 0 kcal/mol) increase in energy. so all deleterious nsSNP has a change in protein stability. The thermodynamic DDG score is −0.581.

Association on NSSNP to disease

PhdSNP analysis

The output comprises of a table listing the total count of the modified position in the protein sequence, the new residue, the wild-type residue and if the related mutation is predicted as neutral polymorphism (Neutral) or as disease-related (Disease).

The RI value (Reliability Index) is evaluated from the output of the provision vector machine O as in the Eq. (1):

$$RI = 20*\text{abs}\,(0-0.5) \quad (1)$$

The score ranges from 0 (very low reliability) to 9 (very high reliability).

From the above Table 6, we predicted that SNP in the rs76940(P266L), rs34767364(R215W), rs61754967(T253I), rs12721593(S93L), rs104895031(D527Y), rs104895032(L421S), rs13312858(K105N), rs111239312(D519Y), rs111244949(L133S) were predicted to be related to infections and disorders. The score series from 0 (very low reliability) to 9 (very high reliability).

Conclusion

The NBN SNPs were analyzed using bioinformatics SNP prediction tools to check the effect of its mutation on the protein function and structure. Four SNPs were predicted to be highly damaged in coding regions such as rs769420 where Phenylalanine is mutated at 266 position into leucine which leads to Aplastic Anemia and Nijmegan breakage syndrome (obtained from gene test records) rs61753020 (Aspartic acid is mutated into Asparigine) and rs12721593(S93L) leads to Microsephaly normal intelligence and immune deficiency and hereditary cancer predisposing syndrome (clivar). Forrs13312858 (K105 N) still the disease association with this region is not confirmed.

SNPs evaluation of the NBN gene point has brought to light the presence of multiple polymorphic sites, some of which cause a change in the functionality of the receptor and were found to be associated to phenotypictraits and higher chances of infection. However, given the large count of SNPs in this gene, association studies should be conducted on genetic mutants that possess a major significance. The relation between obtained results and received data from association studies suggests that an alternate approach to select functional SNPs may be developed with the application of computational tools in inter-related studies. Since SNP prioritization and association studies are two non-redundant source of knowledge, the reflection is that a good correlation between them can benefit the use of computational tools for the selection of SNPs to be examined by association studies. This research aids the evaluation of the genetic mutants corresponding to NBN gene and diseases caused due to mutation. Thus, we conclude that the mutational occurrence in NBN gene leads to various types of diseases and surely it will help to find the suitable drugs in future for various diseases.

Compliance with Ethical Standards

Conflict of interest The Authors declares that there is no conflict of interest.

Research Involving Human Participants and/or Animals Article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent No humans are involved.

References

1. Nichols, K. E., Malkin, D., Garber, J. E., Fraumeni, J. F., and Li, F. P., Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. *Cancer Epidemiol. Biomark. Prev.* 10(2):83–87, 2001.
2. Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., and de Bakker, P. I., SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24:2938–2939, 2010.
3. Chompret, A., Brugières, L., Ronsin, M. et al., P53 germline mutations in childhood cancers and cancer risk for carrier individuals. *Br. J. Cancer* 82(12):1932–1937, 2000.
4. Waters, C. M., Hoover, A. C., McClain, L. C., Moore, T. T., Rogers, C. T., and Thornton, K., Current guidelines and best practice evidence for intensified/enhanced breast cancer screening in women with BRCA mutations. *J. Nurse Pract.* 5(6):447–453, 2009.
5. Safran, M., Dalah, I., Alexander, J. et al., GeneCards version 3: The human gene integrator. *Database (Oxford)* 2010:baq020, 2010.
6. Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielskand, E. M., and Sirotkin, K., dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 29:308–311, 2001.
7. Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, S. C., SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40:W452–W457, 2012.
8. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., and Chan, A. P., Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7(10):e46688, 2012.

9. Adzhubei, I. A., Schmidt, S., Peshkin, L. et al., A method and server for predicting damaging missense mutations. *Nat. Methods* 7(4): 248–249, 2010.
10. Marchini, J., Howie, B., Myers, S., McVean, G., and Donnelly, P., Anew multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* 39(7):906–913, 2007.
11. Zeggini, E., Scott, L. J., Saxena, R. et al., Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* 40(5):638–645, 2008.
12. Vijayakumar, K., and Arun, C., Continuous security assessment of cloud based applications using distributed hashing algorithm in SDLC. *Clust. Comput.*, 2017. https://doi.org/10.1007/s10586-017-1176-x, Springer.
13. Vijayakumar, K., and Arun, C., Automated risk identification using NLP in cloud based development environments. *J. Ambient. Intell. Humaniz. Comput.*, 2017. https://doi.org/10.1007/s12652-017-0503-7, Springer.
14. Savojardo, C., Fariselli, P., Martelli, P. L., and Casadio, R., INPS-MD: A web server to predict stability of protein variants from sequence and structure. *Bioinformatics* 32(16):2542–2544, 2016.
15. Capriotti, E., Fariselli, P., and Casadio, R., I-Mutant2.0: Predicting stability changes upon mutation from the protein sequence or structure. *J. Nucleic Acid Res.* 33:306–310, 2013.