



Identification and analysis of oncogenic non-synonymous single nucleotide polymorphisms in the human *NRAS* gene: An exclusive *in silico* study



Md. Mozibullah^a, Hadieh Eslampanah Seyedi^b, Marina Khatun^a, Md Solayman^{c,*}

^a Department of Biochemistry and Molecular Biology, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh

^b Institute for Glycomics, Griffith University, Parklands Dr. Southport, QLD 4222, Australia

^c Department of Biochemistry and Molecular Biology, Primeasia University, Bangladesh

ARTICLE INFO

Keywords:

Cancer
In silico
 nsSNPs
NRAS gene
 Oncogenic
 Polymorphism

ABSTRACT

Background: N-ras protein is encoded by the *NRAS* gene and operates as GDP-GTP-controlled on/off switching. N-ras interacts with cellular signaling networks that regulate various cellular activities including cell proliferation and survival. The nonsynonymous single nucleotide polymorphism (nsSNPs)-mediated alteration can substantially disrupt the structure and activity of the corresponding protein. N-ras has been reported to be associated with numerous diseases including cancers due to the nsSNPs. A comprehensive study on the *NRAS* gene to unveil the potentially damaging and oncogenic nsSNPs is yet to be accomplished. Hence, this extensive *in silico* study is intended to identify the disease-associated, specifically oncogenic nsSNPs of the *NRAS* gene.

Results: Out of 140 missense variants, 7 nsSNPs (I55R, G60E, G60R, Y64D, L79F, D119G, and V152F) were identified to be damaging utilizing 10 computational tools that works based on different algorithms with high accuracy. Among those, G60E, G60R, and D119G variants were further filtered considering their location in the highly conserved region and later identified as oncogenic variants. Interestingly, G60E and G60R variants were revealed to be particularly associated with lung adenocarcinoma, rhabdomyosarcoma, and prostate adenocarcinoma. Therefore, D119G could be subjected to detailed investigation for identifying its association with specific cancer.

Conclusion: This *in silico* study identified the deleterious and oncogenic missense variants of the human *NRAS* gene that could be utilized for designing further experimental investigation. The outcomes of this study would be worthwhile in future research for developing personalized medicine.

1. Background

The most prevalent genetic variations that have been identified in the human genome so far are single nucleotide polymorphisms (SNPs), a polymorphism of two alternating alleles that appear in the population at a frequency of greater than one percent.⁴⁹ SNPs that alter the amino acid in the corresponding protein are called non-synonymous SNPs (nsSNPs), which are reported to be correlated with the diverse types of diseased-state.^{50,52} Disease-causing nsSNPs are accounted to cause a drastic alteration in the physicochemical features of a protein, which destabilize, weaken, and augment the flexibility of that encoded protein. In spite of that, disease-associated missense variants may

enhance the rigidity, and change the geometry, stability, and interactions with other proteins.²⁶

N-ras is a member of the Ras family (K-ras, H-ras, and N-ras) of small GTPases acting as a molecular switch during the cellular signal transduction schemes to regulate proliferation, differentiation, and survival of cells. The activated state is maintained by the association of N-ras with GTP, while the inactivated state, with GDP.^{7,24} The locus of the human *NRAS* gene is mapped to the 1p13.2 chromosomal band, which consists of seven exons encoding N-ras protein containing 189 amino acids.³⁰ The residues 1–166 constitute the catalytic domain, which is divided into two lobes namely the effector lobe (residues 1–86) and the allosteric lobe (residues 87–166).²⁴ The formation of

Abbreviations: nsSNPs, nonsynonymous Single Nucleotide Polymorphism; FATHMM, Functional Analysis Through Hidden Markov Models; PolyPhen-2, Polymorphism Phenotyping v2; SIFT, Sorting Tolerant From Intolerant; PROVEAN, Protein Variation Effect Analyzer; M-CAP, Mendelian Clinically Applicable Pathogenicity; VEST-4, Variant Effect Scoring Tool v4; PDB, Protein Data Bank; GEPIA, Gene Expression Profiling Interactive Analysis; LUAD, Lung adenocarcinoma; PAAD, Prostate adenocarcinoma; SARC, Sarcoma.

* Corresponding author at: Department of Biochemistry and Molecular Biology, Primeasia University, Star Tower, 12 Kamal Ataturk Avenue, Banani, Dhaka 1213, Bangladesh.

E-mail address: naem40thju@gmail.com (M. Solayman).

<https://doi.org/10.1016/j.jgeb.2024.100378>

Received 2 December 2023; Accepted 19 April 2024

the active site within the effector lobe for the interaction of regulators and effectors as well as hydrolysis of GTP is mediated via three specific regions like residues 10–17, 30–40, and 60–76 namely P-loop, switch I, and switch II, respectively.^{48,51}

All mammalian cells express Ras proteins, and structurally, they are closely related. Mutationally activated Ras proteins promote oncogenesis due to their proto-oncogenic nature, and they are reported to be often mutated in numerous cancers.^{14,35} To activate the Ras, guanine nucleotide exchange factors (GEFs) restructure the protein mediating the interchange of GDP to GTP. GTPase-activating proteins (GAPs)-mediated and intrinsic GTPase downregulate the Ras signaling via GTP hydrolysis. Oncogenic mutation of Ras can prolong the active state of Ras through the structural alteration that favors the GTP bound form and reduction in the Intrinsic and/or GAP-stimulated GTPase activity, leading to the constitutive activation of downstream signaling pathways e.g., RAF/MEK/ERK (Fig. 1).⁴⁴ In 1983, *NRAS* was identified in neuroblastoma as a family member of the *RAS* gene with the transforming potentiality.²⁸ Since its identification, studies have reported missense variants of N-ras are associated with different categories of diseases including autoimmune lymphoproliferative syndrome,³² colorectal cancer,²¹ and melanomas.^{33,36} A significant number of missense SNPs of *NRAS* have been enlisted in the NCBI dbSNP database. As the experimental approaches are costly and arduous to classify the deleterious missense SNPs, whilst the computational attempts are reliable, rapid, user-friendly, and inexpensive.⁶

Therefore, here we perform an extensive computational study to determine the damaging missense SNPs in *NRAS* and assess their subsequent structural and functional consequences. Moreover, we find out oncogenic missense nsSNPs in N-ras protein with their associated cancer types.

2. Method and materials

Schematical representation of the entire methodology procedure mentioned in Fig. 2.

2.1. Data retrieval

The NCBI-dbSNP database provides extensive information about the SNPs that are reported for any kind of gene. SNPs-associated information, specifically nsSNPs of *NRAS* gene including rs IDs, and allelic

variation, genomic coordinates, amino acid changes due to missense variants were retrieved from the NCBI-dbSNP database.³⁹ Basic information and protein sequence in FASTA format were obtained from the UniProtKB database.

2.2. Classifying the most damaging nsSNPs

The combination of multiple algorithmic prediction tools augments the performance, precision, and accuracy of assessing the functional and structural consequences of nsSNPs. A combination of various computational methodologies including sequence structure-based, sequence homology-based, supervised learning-based, and consensus-based prediction tools were used to discover the most damaging mutations in the *NRAS* gene.^{2,53} All *in silico* tools were accessed during the months of October to December in 2021.

2.2.1. Functional effect analysis

We employed eight tools from the aforementioned methodologies to investigate the functional consequences of nsSNPs. Two tools from each methodology such as FATHMM⁴⁰ and PolyPhen-2¹ (Sequence structure-based), SIFT⁴¹ and PROVEAN¹⁶ (Sequence homology-based), M-CAP²³ and VEST-4¹¹ (Supervised learning-based), Meta-SNP¹⁰ and PredictSNP⁵ (Consensus-based) were utilized in this comprehensive study.

2.2.2. Protein stability analysis

The nsSNPs-mediated alteration in the corresponding protein causes variation in the structure. The difference in the free energy of a mutated protein compared to a wild-type is a fundamental parameter for assessing the protein stability.^{27,54} We utilized I-Mutant3.0¹² using physiological parameters (temperature 37 C and pH 7.4) and MU-pro¹⁵ for the prediction of structural impacts in terms of protein stability. Both tools rely on the algorithm of Support Vector Machines for assessing the stability of a protein due to the missense variants, where we used the FASTA sequence of N-ras protein as an input.

2.3. Conservation analysis

ConSurf server has long been a widely accepted tool for analyzing the evolutionary conservation scenario of each residue of macromolecules like protein, DNA, and RNA.^{4,3} ConSurf can reveal the vital

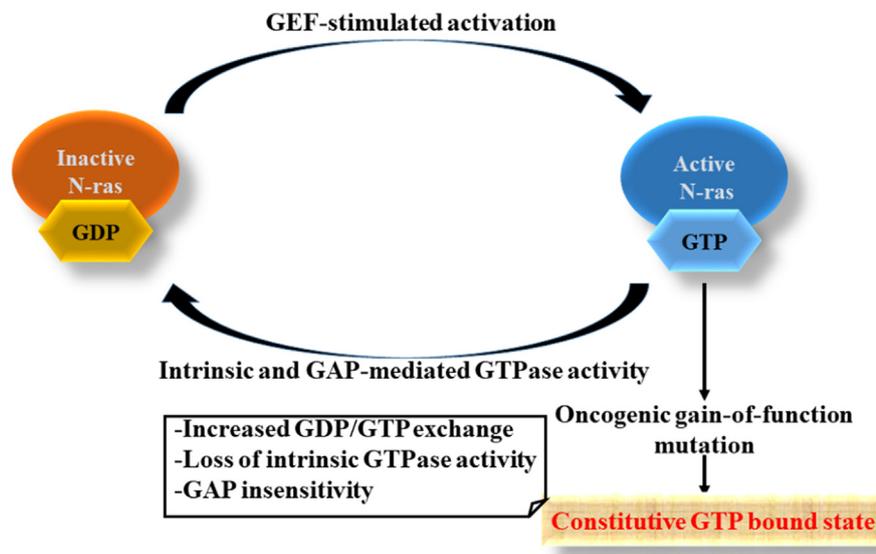


Fig. 1. The active and inactive state of N-ras with the mechanism of oncogenic mutation leads to constitutive active GTP bound state. Oncogenic mutation-mediated structural alteration can result in the augmented GDP/GTP exchange or loss of intrinsic GTPase activity. Oncogenic mutation can also alter the GEF and GAP regulatory mechanisms.⁴⁴

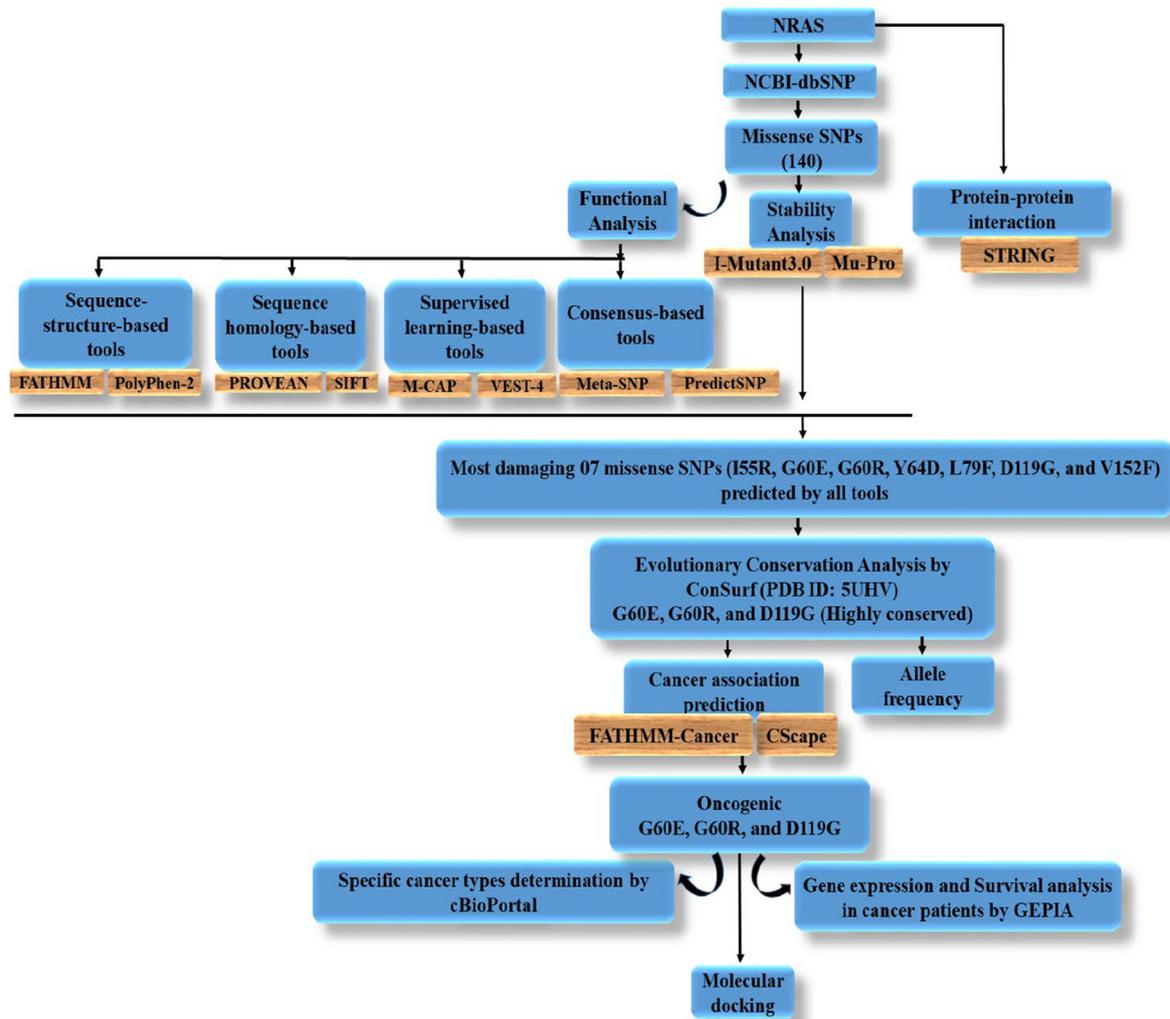


Fig. 2. The diagram presents the overall methodology. Missense SNPs retrieved from the dbSNP database were subjected to functional and stability analysis to classify the most damaging SNPs. ConSurf server was employed to determine the conservation nature of human N-ras protein which identified the highly conserved residues. Further downstream analysis including their potentiality to be oncogenic, gene expression, and molecular docking was accomplished.

region of the macromolecules relating to the structure and/or function. ConSurf employs the Bayesian computation technique to assign an evolutionary conservation score based on the multiple sequence alignment and a phylogenetic tree. The conservation scores of 1 to 4, 5 and 6, and 7 to 9 are considered as variable, intermediate, and conserved, respectively. This web-server was deployed to assess the evolutionary conservation profile of N-ras protein using PDB ID: 5UHV as input structure. 5UHV was selected due to no mutation, no interim missing amino acids, higher resolution (1.67 Å), and 166 amino acids solved coverage compared to other available X-ray crystallography structures.²⁴

2.4. Allele frequency (AF) analysis

AF data was retrieved by utilizing the genome aggregation database (gnomAD v2.1.1) for the mutants residing in the highly conserved region of human N-ras protein. The gnomAD database is developed from a plethora of sequencing projects to harmonize and aggregate data of both genome and exome sequencing for the availability to the scientific community. The gnomAD v2.1.1 is based on the population genetic and disease-associated investigation of unrelated individuals spanning whole-genome (15,708) and exome sequencing (125,748).²⁵

2.5. Protein-protein interaction analysis

The functions of protein largely rely on the interactions with other proteins. Hence, the STRING database was utilized to unveil the functional interaction of N-ras protein with other proteins. The STRING collects the interaction data from both the experimental and predicted information.⁴²⁻⁴³

2.6. Prediction of cancer-associated SNPs

CScape utilizes a statistical approach to predict the missense mutation as a cancer driver with 91 % accuracy.³⁷ FATHMM-cancer is a species-independent approach, which has different classifiers to predict the disease-causing, cancer driver, and disease-specific mutations.⁴⁰ Here, we selected the cancer option for the prediction of cancer-causing variants. The cBioPortal database¹³ was utilized to identify the specific types of cancer-related to our oncogenic missense SNPs.

2.7. Expression pattern and overall survival analysis

Gene Expression Profiling Interactive Analysis (GEPIA) works based on the Cancer Genome Atlas (TCGA) and the Genotype-Tissue

Expression (GTE_x) to provide interactive functions of gene expression, analysis of patient's survival, correlation, and profiling plotting.⁴⁶ The expression profile of *NRAS* gene as well as overall survival pattern were studied with specific type of cancers such as Lung adenocarcinoma (LUAD), Sarcoma (SARC), and Prostate adenocarcinoma (PAAD) due to their association with our studied oncogenic mutants.

2.8. Molecular docking study

Analysis of wild-type and mutant (D119G) proteins-ligand (GDP and GTP) binding was executed by molecular docking. The 3D structure of human N-ras protein was retrieved from the Protein Data Bank (PDB)⁹ in PDB format (5UHV). Initially, the 3D structure was prepared by Discovery Studio 2020 to remove the attached ligand, ions, and

water. The Swiss-PDB Viewer (SPDV)⁹ was used to eliminate and add bad contacts and missing atoms, respectively. The wild-type protein was then subjected to energy minimization with the integrated GRO-MOS96 force field. The SVDB was also employed to generate the mutant structure by replacing the wild-type residue with the mutant residue and energy minimization.

The 3D structure of ligand guanosine diphosphate (GDP) and guanosine triphosphate (GTP) was downloaded in a mole file from the ChemSpider database and converted to sdf format by Discovery Studio 2020. The ligand was then energy minimized using mmff94 force field and steepest descent option.

Binding affinity (kJ/mol) of best poses between protein and ligand (GDP and GTP) was generated by the Autodock Vina tool,⁴⁷ a widely used and accepted utilized tool for analysis of docking. The protein

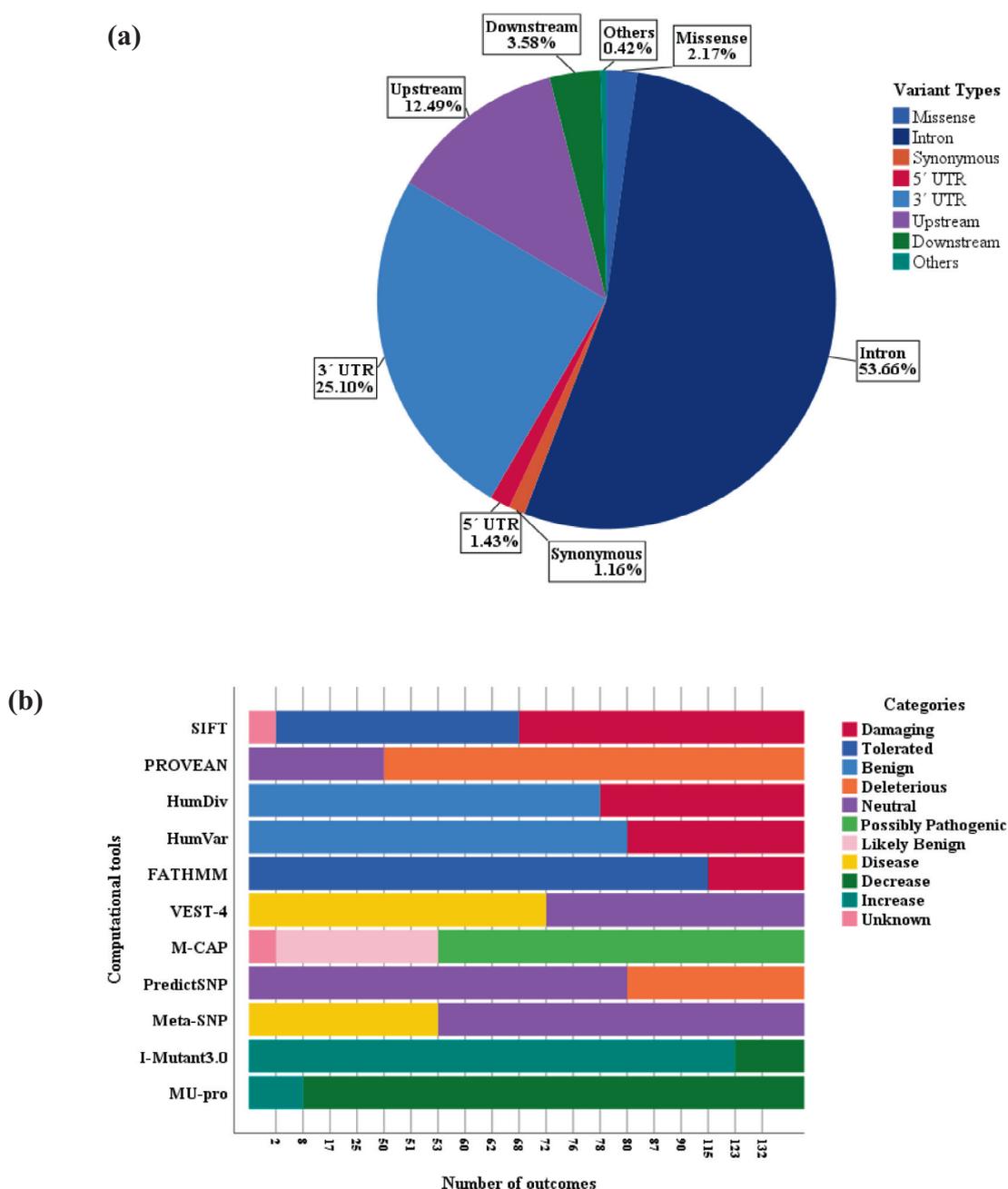


Fig. 3. SNP related data of human *NRAS* gene. (a) The pie chart displays variant types with their percentage frequency distribution deposited in the NCBI-dbSNP database. (b) The stacked bar chart describes the comprehensive outcome of utilized algorithmic tools for studied missense variants.

Table 1The most damaging missense variants predicted in the human *NRAS* gene.

No	SNP IDs	AA Change	Sequence homology-based tools		Sequence-structure-based tools				Supervised learning-based tools		Consensus-based tools		Stability analysis tools	
			SIFT	PROVEAN	PolyPhen-2		FATHMM	M-CAP	VEST-4 (p-value*)	Meta-SNP	PredictSNP	I-Mutant3.0	MU-pro	
					HumDiv	HumVar								
1	rs1246727247	I55R	D	D	PD	PD	D	PP	D	D	D	D	D	D
2	rs267606920	G60E	D	D	PD	PD	D	PP	D	D	D	D	D	D
3	rs1557982817	G60R	D	D	PD	PD	D	PP	D	D	D	D	D	D
4	rs752508313	Y64D	D	D	PD	PD	D	PP	D	D	D	D	D	D
5	rs1659097188	L79F	D	D	PD	PD	D	PP	D	D	D	D	D	D
6	rs754428086	D119G	D	D	PD	PD	D	PP	D	D	D	D	D	D
7	rs757968407	V152F	D	D	PD	PD	D	PP	D	D	D	D	D	D

D: Deleterious/Damaging/Disease/Decrease; PD: Probably damaging; PP: Possibly pathogenic.

* p-value range: 0.00162 – 0.02287.

region that will be used for docking is determined by the size of the grid box. During docking, no location outside of the box will be investigated.¹⁸ In our study, molecular docking was done by keeping the grid box size at X: 42.2867 Å, Y: 36.0680 Å, Z: 41.4736 Å to enclose the specified area of N-ras allowing GDP and GTP binding with the best possible conformation.

3. Results

3.1. Extraction of SNPs data

As reported by the NCBI-dbSNP database in December 2021, a total of 5252 SNPs existed related to the *NRAS* gene. Among them, intron variants were the most abundant (53.66 %), which was followed by 3' UTR variants (25.10 %), upstream variants (12.49 %), downstream variants (3.58 %), missense variants (2.17 %), 5' UTR variants (1.43 %), synonymous variants (1.16 %), and other variants (0.42 %) such as frameshift, initiator codon, splice donor, and splice acceptor variants. (Fig. 3a). Out of all SNPs, missense nsSNPs (due to their association with diverse types of diseases) were considered to analyze their consequences on human N-ras protein.

3.2. Identification of the most damaging nsSNPs

Missense variants retrieved from the NCBI-dbSNP database were investigated to classify the most deleterious variants using 10 computational algorithmic tools (Supplementary Table S1). We filtered the most damaging variants from 140 missense variants based on the classification of all employed tools as damaging/deleterious (Fig. 3b).

3.2.1. Functional analysis

Eight predicting algorithms were employed to investigate the functional consequences of 140 missense variants in the N-ras protein. Out of eight tools, PROVEAN classified the utmost number of deleterious missense nsSNPs (90 in number), while FATHMM predicted the fewest number of missense nsSNPs as damaging (25 in number). The detailed outcomes of predicting algorithms for functional analysis of missense variants are depicted in Fig. 3b.

3.2.2. Stability analysis

According to I-Mutant3.0 and MU-pro, 108 and 132 missense variants were found to have decreasing effects on the stability of N-ras protein, respectively as illustrated in Fig. 3b.

3.2.3. Selected most damaging variants

Seven variants as rs1246727247 (I55R), rs267606920 (G60E), rs1557982817 (G60R), rs752508313 (Y64D), rs1659097188 (L79F), rs754428086 (D119G), and rs757968407 (V152F) were classified as

deleterious/ damaging by all employed computational prediction tools (Table 1 and Supplementary Table S2)). Hence, these were determined as the most damaging missense variants and considered for further investigation.

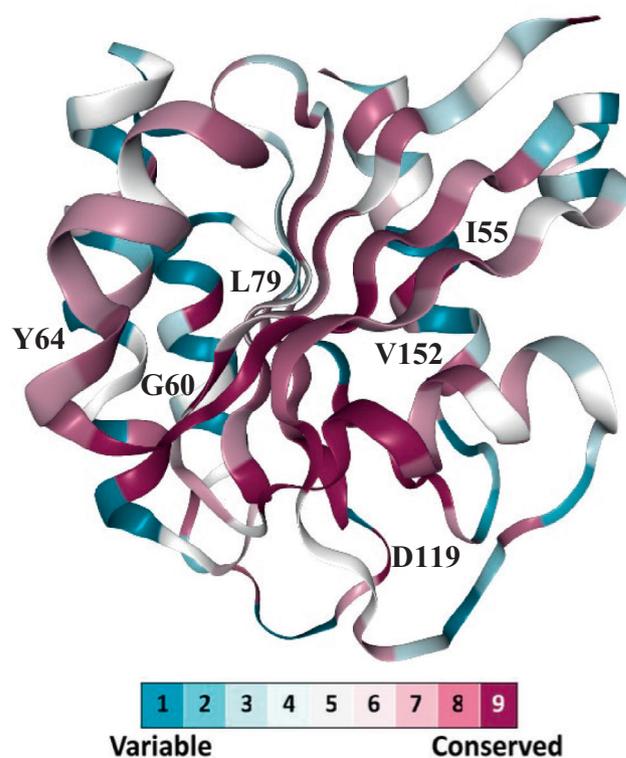


Fig. 4. Evolutionary conservation profile of human N-ras protein. Residue color is based on the conservation scale.

Table 2

Evolutionary conservation profile of native residues at the position of high-risk missense variants.

AA position	Native residue	Mutant residue	Conservation Score
55	I	R	8
60	G	E	9 (Highly conserved)
60	G	R	9 (Highly conserved)
64	Y	D	8
79	L	F	5
119	D	G	9 (Highly conserved)
152	V	F	8

Table 3
Allele frequency data according to gnomAD v2.1.1 database.

SNPs ID	Amino acid change	Protein consequence	Allele count	Allele number	Allele frequency
rs267606920	G60E	p.Gly60Glu	1	31,392	3.19e-5
rs1557982817	G60R	-	-	-	-
rs754428086	D119G	p.Asp119Gly	1	251,488	3.98e-6

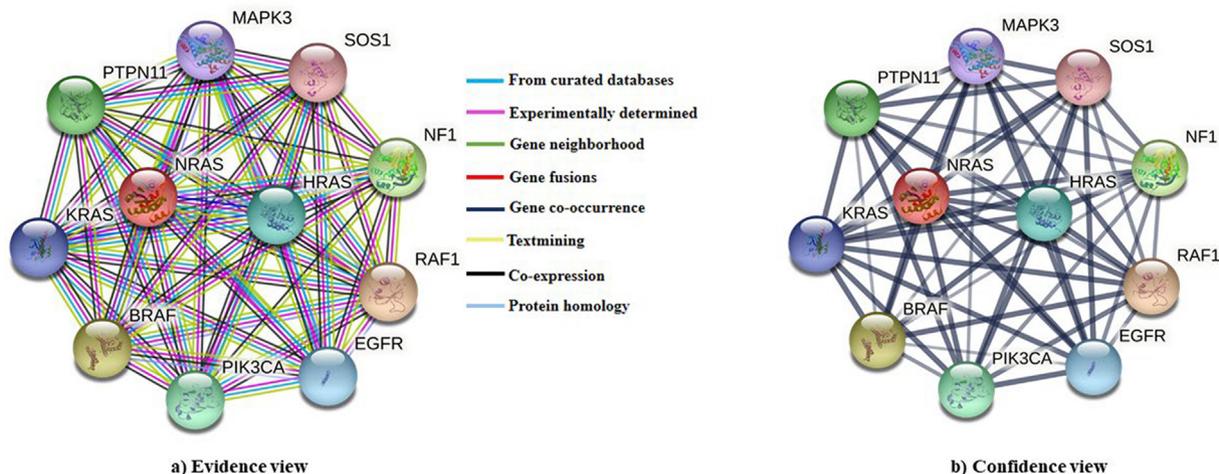


Fig. 5. The interactions of N-ras protein with other proteins determined by STRING database. (a) Colors of line indicate the type of evidence interaction. (b) Thickness of line indicates the strength of data support.

Table 4
Prediction of cancer-associated missense SNPs.

No.	SNPs ID	Amino acid change	CScape		FATHMM-Cancer	
			Prediction	Coding score	Prediction	Score
1	rs267606920	G60E	Oncogenic (high confidence)	0.93197	Cancer	-4.39
2	rs1557982817	G60R	Oncogenic	0.884552	Cancer	-4.39
3	rs754428086	D119G	Oncogenic (high confidence)	0.920309	Cancer	-4.55

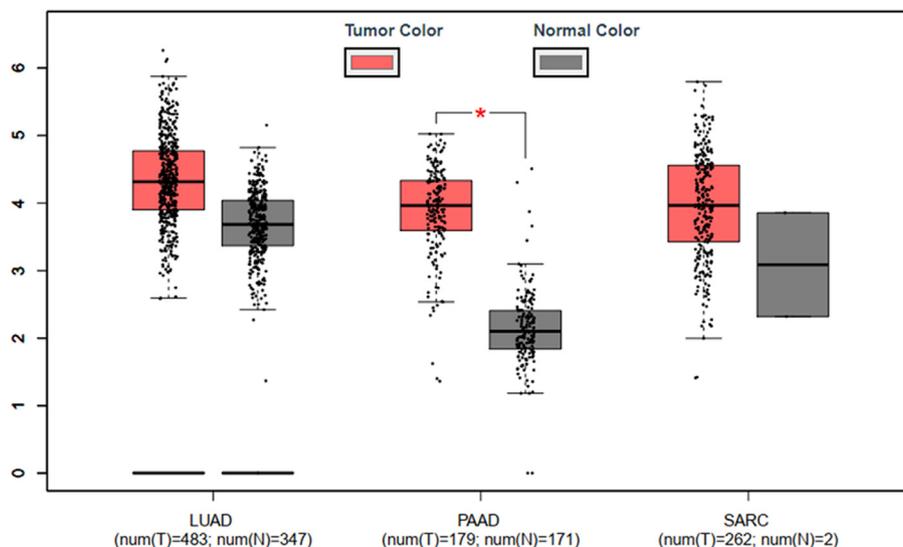


Fig. 6. Expression profile of NRAS gene in Lung adenocarcinoma (LUAD), Prostate adenocarcinoma (PAAD), and Sarcoma (SARC) ($P < 0.05$). The signature score is calculated by mean value of \log_2 (TPM + 1) of each gene in NRAS-like signature gene set. * $P < 0.01$.

3.3. Conservation analysis of selected variants

Highly conserved residues of a protein throughout its evolution are mainly crucial for the maintenance of structure, function, and interaction with the effectors and regulators.¹⁷ The conservation profile of N-ras protein (Fig. 4) showed that the native positions of the selected mostly damaging variants (I55, G60, Y64, L79, D119, and V152) were highly conserved, and the conservation scores were varying from 5 to 9 (Table 2). More specifically, 2 native positions (G60 and D119) creating three mostly deleterious mutations (G60E, G60R, and D119G) were located in the highest conserved area (conservation score of 9)

and D119G was found to be a substrate binding residue. Therefore, these three mutants were subjected to further downstream analysis.

3.4. Allele frequency data

Allele frequency (AF) related to the most damaging nsSNPs residing in the highly conserved wild-type residue was collected from the gnomAD database. Table 3 provided the data of allele frequency with protein consequences, allele number, and allele count. The allele frequency for the G60E variant (3.19×10^{-5}) was higher than D119G (3.98×10^{-6}).

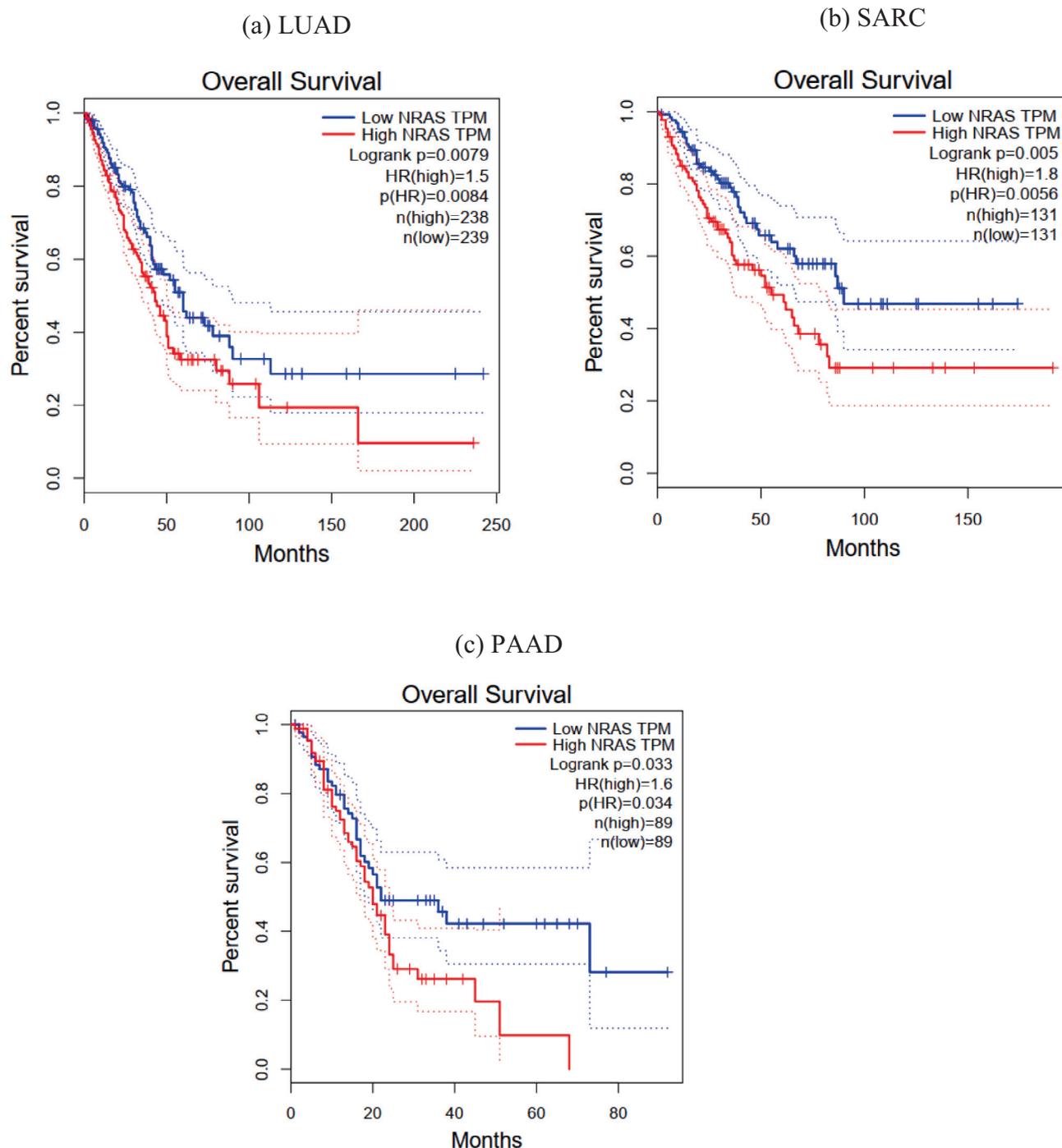


Fig. 7. Overall survival analysis of Lung adenocarcinoma (LUAD), Sarcoma (SARC), and Prostate adenocarcinoma (PAAD) patients. TPM and HR stand for transcripts per million and hazard ratio, respectively.

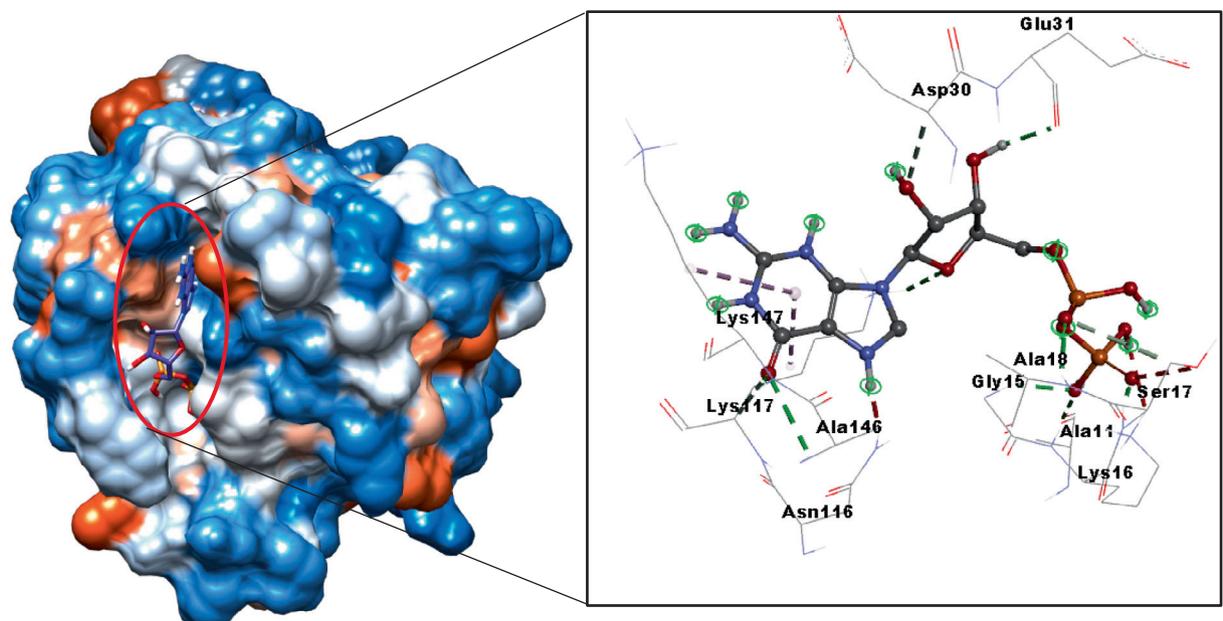
3.5. Investigation of protein–protein interactions

Missense mutation may cause the alteration of the protein structure, consequently, functional interaction with other proteins may also be changed. Hence, the STRING server was utilized to unveil the interaction of N-Ras with other proteins (Fig. 5) employing minimum required interaction scores of 0.400 (medium confidence). The STRING server revealed that N-ras functionally interacts with other proteins with PPI enrichment *p*-value and clustering coefficient as 3.43×10^{-9} and 1, respectively. According to the STRING server, N-ras interacted protein partners were GTPase H-ras, GTPase K-ras, B-raf (proto-oncogene serine/threonine-protein kinase), Phosphatidylinositol 4, 5-bisphosphate 3-kinase catalytic subunit alpha isoform

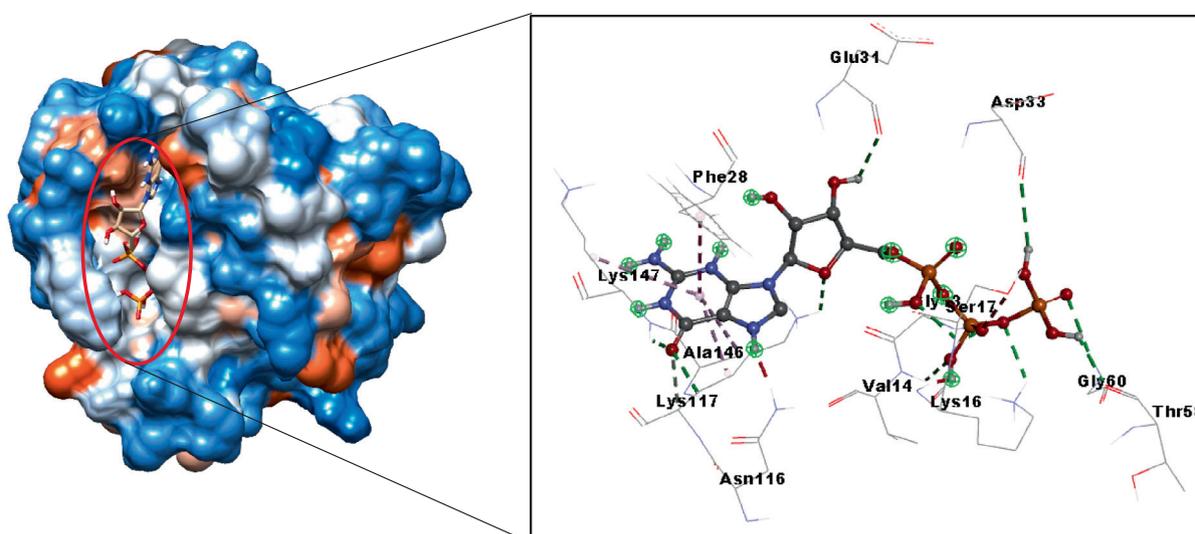
(PIK3CA), Epidermal growth factor receptor (EGFR), RAF (proto-oncogene serine/threonine-protein kinase), Neurofibromin 1 (NF1), Son of sevenless homolog 1 (SOS1), and Mitogen-activated protein kinase 3 (MAPK3). Therefore, deleterious variants in N-ras protein could hamper the functional interaction with these proteins leading to alteration of the corresponding cellular activities.

3.6. Prediction of cancer-associated nsSNPs

Variants residing in the highly conserved residues were studied to determine their cancer-promoting role due to the oncogenic nature of N-ras protein by two online tools (FATHMM-Cancer and CScape). A variant, for which CScape coding score is close to one, is predicted



(a) GDP bound D119G variant



(b) GTP bound D119G

Fig. 8. Schematic representation of protein–ligand docked complexes. (a) D119G variant-GDP docked complex (left) and the interacted amino acid residues of D119G mutant (right). (b) D119G variant-GTP docked complex (left) and the interacted amino acid residues of D119G mutant (right). The green interacting color indicates the hydrogen bond, and the rest of the colors indicate different hydrophobic interactions (ink color: pi-pi T shaped, yellow color: pi-alkyl interaction). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as a cancer driver with the highest confidence and accuracy. CScape predicted all query variants (G60E, G60R, and D119G) as oncogenic, while G60E and D119G variants predicted with the highest confidence and accuracy (CScape coding scores: 0.93197 and 0.920309, respectively). In the case of FATHMM-Cancer, a score less than -0.75 indicates the mutant is significantly related to cancer. The prediction revealed that G60E, G60R, and D119G mutants are potentially associated with cancer having the score of -4.39 , -4.39 , and -4.55 , respectively (Table 4).

3.7. Determination of cancer types and expression profile

Predicted oncogenic nsSNPs were subjected to assess the association with specific types of cancers. cBioPortal database revealed the association of G60E with Lung Adenocarcinoma and Rhabdomyosarcoma, whilst G60R with Prostate Adenocarcinoma. To elucidate the N-ras expression pattern in these cancers, the box plot analysis was performed which demonstrated that the *NRAS* gene is overexpressed in Lung adenocarcinoma, Sarcoma, and Prostate adenocarcinoma (Fig. 6). Therefore, overexpression of these mutated *NRAS* variant could result these types of cancers.

3.8. Overall survival analysis

The GEPIA database was employed to analyze the relationship between *NRAS* gene expression and survival of patients with LUAD,

Table 5
Binding affinity of protein ligand complexes in Molecular docking.

Docked complexes	Binding affinity (kJ/mol)
Wildtype-GDP	-9.9
Wildtype-GTP	-10
D119G-GDP	-9.6
D119G-GTP	-10.5

Table 6
Non-bond interaction analysis of D119G-GDP and D119G-GTP docked complexes.

Docked complex	Binding affinity (kJ/mol)	Interacting AA	Distance (Å)	Bond category	Bond type		
D119G-GDP	-9.6	GLY15	2.10606		Conventional Hydrogen Bond		
		SER17	1.94699				
		ALA18	2.05301				
		LYS117	2.86878				
		ALA146	2.70073				
		LYS147	2.15136				
		GLU31	2.08537				
		ALA11	2.89436				
		SER17	3.41768				
		ASP30	3.78621			Hydrogen Bond	
		LYS117	3.77153				
		LYS117	4.02388			Hydrophobic	Pi-Alkyl
		LYS147	4.74728				
		D119G-GTP	-10.5	GLY13		2.17502	
VAL14	2.80057						
LYS16	2.497						
SER17	2.39915						
SER17	1.97616						
GLY60	2.64362						
LYS117	2.89321						
ALA146	2.68436						
LYS147	2.20948						
GLU31	2.72902						
ASP33	2.95475						
THR58	2.21455						
LYS117	3.62153						
PHE28	4.57016				Hydrophobic	Carbon Hydrogen Bond Pi-Pi T-shaped Pi-Alkyl	
LYS117	4.00211						
ALA146	5.45919						
LYS147	4.6959						

SARC, and PAAD. The subjects were categorized into two groups as such high and low expression levels based on the median expression of the *NRAS* gene. The patients with LUAD, PAAD, and SARC exhibited poor overall survival with high expression levels of *NRAS* gene compared to the patients with lower expression of *NRAS* gene (Fig. 7). All results were statistically significant. More specifically, PAAD patients showed substantially lower survival time with higher expression of the *NRAS* gene (Fig. 7c).

3.9. Molecular docking study

N-ras with GTP bound state is active (GDP bound state is inactive) and mediates downstream effector functions. Constitutively, prolonged active N-ras leads to various cancers. Therefore, binding strength and conformations of GDP and GTP with mutant (D119G) N-ras protein was determined to reveal the mechanism on how D119G variants might be oncogenic by a molecular docking study using Autodock Vina (v1.2.0) (Fig. 8). The results of the docking study showed that binding affinity of D119G-GTP (10.5 kJ/mol) docked complex were higher than D119G-GDP (9.6 kJ/mol) docked complex and even higher than the wild-type N-ras-GTP (Table 5). D119G-GTP docked complex manifested increased hydrogen bond and hydrophobic interaction compared to D119G-GDP, indicating the rational for higher binding affinity. Overall, D119G-GTP formed 13 hydrogen bonds and 4 hydrophobic interactions, while D119G-GDP showed 11 hydrogen bonds and 2 hydrophobic interactions (Table 6).

4. Discussion

The abundance of nsSNPs has overflowed the relevant databases due to the advancement of sequencing techniques. Moreover, increasing trend of inundated nsSNPs turns the situation in an impossible state to characterize all nsSNPs employing experimental approaches. Conversely, computational approaches are a great way of differentiating the most damaging and disease-associated nsSNPs from a large num-

ber of nsSNPs.²² The presence of numerous deleterious nsSNPs in oncogenes promoting cancer has increased the attention to a serious concern.³⁸ The *NRAS* gene encodes a small GTPase that acts as a molecular regulator of various cellular activities (including cell proliferation and survival) and has been reported to be associated with various types of cancer.^{7, 8, 24, 36}

The computational approach to classify the disease-associated nsSNPs of human *NRAS* gene has yet to be performed. To predict the most deleterious nsSNPs more accurately and reliably, a combination of algorithmic prediction tools with different methodologies should be employed. In this study, 10 algorithmic tools (FATHMM, PolyPhen-2, SIFT, PROVEAN, M-CAP, VEST-4, Meta-SNP, PredictSNP, I-Mutant3.0, and MU-pro) were exploited to analyze the functional consequences of nsSNPs in human *NRAS* gene. Generally, the native function of a protein relies on the stability of the corresponding protein.³⁸ To assess the impact of the nsSNPs in N-Ras protein stability, I-Mutant3.0, and MU-pro tools were used. Rigorous assessment of nsSNPs in the *NRAS* gene by all utilized tools has made identification of 7 nsSNPs as the most deleterious.

The information related to evolutionary conservation is crucial for the identification of disease-associated nsSNPs.²⁹ The missense variants at highly conserved residues have more tendency to be deleterious than variants in less conserved regions. As stated by the ConSurf web server, wild-type residues of three missense variants out of seven studied variants are located in the highly conserved region (G60E, G60R, and D119G). Furthermore, G60E showed higher allele frequency than D119G indicating that it (G60E) is very often noticed in people. Protein-protein interaction is crucial to execute the cellular processes. The STRING database revealed that N-ras interacts with different cellular proteins, most of which are involved in the signaling process of cell proliferation, differentiation, and survival. STRING also showed the association of N-ras with NF1, which is a regulator of Ras GTPase activity. Our identified deleterious missense variants of N-ras could alter the interaction with these proteins to promote the development of diseases e.g., cancers.

Genomic instability due to mutations may cause numerous cancer types. Various studies have demonstrated the association of Ras variants with different cancers.³¹ To predict cancer-causing mutants, FATHMM-Cancer and CScape tools were used and predicted that all three (G60E, G60R, and D119G) were oncogenic/cancer drivers. Furthermore, the cBioPortal database showed that more than 75 types of cancers are associated with the missense variants of the *NRAS* gene. The G60E variant was revealed to be associated with Lung Adenocarcinoma, Rhabdomyosarcoma and G60R with Prostate Adenocarcinoma. The D119G variant was not found to be associated with any cancer types. This D119G mutant could be a subject of detailed study to find out the association with specific cancer types. Approximately 25% of cancers are due to gain-of-function mutation of *RAS* genes.²⁰ The GEFs stimulated activation of Ras is maintained with GTP bound state, which preferentially binds with its downstream signaling molecules, and GAP stimulated inactivation is mediated by the GTPase activity to form GDP bound inactive Ras. The mechanisms of the gain-of-function variant of Ras in cancers are missense variant mediated conformational change of protein that favors GTP binding and loss of GTPase activity that also favors GTP bound activated state.^{20,31} To unveil the possible mechanism of D119G (residue in the ligand binding site) oncogenic mutant of *NRAS* gene, molecular docking of D119G variant with GDP and GTP using AutoDock Vina was accomplished. The higher binding affinity as well as greater hydrogen bonds and hydrophobic interactions of D119G with GTP in comparison to GDP indicated that D119G might alter the N-ras protein structure in a way that favored the GTP bound state. Therefore, cancer-driven property of D119G variant might be explained by the constitutive GTP bound activated conformation.

Mutation can alter the expression of a gene such as mutated *RUNX1* was demonstrated to be strongly expressed.^{19,34} Another study also

revealed that missense mutated *TP53* and *ZNF750* genes in patients of esophageal squamous cell carcinoma (ESCC) were highly expressed.⁴⁵ Box plot analysis of expression of *NRAS* gene in LUAD, SARC, and PAAD cancers by GEPIA web server unveiled that these cancers were caused by the higher expression of *NRAS*. Survival analysis also showed that upregulated expression of *NRAS* gene in all LUAD, SARC, and PAAD patients showed poorer survival time than the patients with downregulated *NRAS* expression.

Although this comprehensive study predicts that G60E, G60R, and D119G are potential oncogenic mutants, robust *in vitro* and *in vivo* experiments are required to validate the present outcomes in the future.

5. Conclusion

Based on *in silico* analysis, the current study reported 3 nsSNPs like rs267606920 (G60E), rs1557982817 (G60R), and rs754428086 (D119G) as the potentially damaging considering their functional and structural consequences in N-ras protein and their presence in the highly conserved region. Furthermore, three variants were predicted to be oncogenic and two (G60E and G60R) of them were found to be associated with Lung adenocarcinoma, Rhabdomyosarcoma, and Prostate adenocarcinoma. Therefore, it appears that other mutant D119G might be involved in distinct types of cancers. The outcomes of the present study will potentially create a guideline to filter cancer driver nsSNPs. However, extensive studies on population and wet-lab investigation are crucial to characterize and decipher the consequences of the three oncogenic mutants in N-ras protein and development of an effective personalized medicine.

CRedit authorship contribution statement

Md. Mozibullah: Conceptualization, Data curation, Formal analysis, Writing – original draft. **Hadih Eslampanah Seyedi:** Data curation, Writing – review & editing. **Marina Khatun:** Validation, Writing – review & editing. **Md Solayman:** Conceptualization, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jgeb.2024.100378>.

References

- Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–249. <https://doi.org/10.1038/nmeth0410-248>.
- Alam MS, Saleh MA, Mozibullah M, Riham AT, Solayman M, Gan SH. Computational algorithmic and molecular dynamics study of functional and structural impacts of non-synonymous single nucleotide polymorphisms in human DHFR gene. *Comput Biol Chem*. 2021;107587.
- Ashkenazy H, Abadi S, Martz E, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res*. 2016;44(W1):W344–W350. <https://doi.org/10.1093/nar/gkw408>.
- 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(SUPPL. 2):529–533. <https://doi.org/10.1093/nar/gkq399>.
- Bendl J, Stourac J, Salanda O, et al. PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. *PLoS Comput Biol*. 2014;10(1):1–11. <https://doi.org/10.1371/journal.pcbi.1003440>.
- Bhatnager R, Dang AS. Comprehensive *in-silico* prediction of damage associated SNPs in Human Prolidase gene. *Sci Rep*. 2018;8(1):1–14.

7. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature*. 1990;348(6297):125–132.
8. Brown R, Marshall CJ, Pennie SG, Hall A. Mechanism of activation of an N-ras gene in the human fibrosarcoma cell line HT1080. *EMBO J*. 1984;3(6):1321–1326.
9. Burley SK, Bhikadiya C, Bi C, et al. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res*. 2021;49(D1):D437–D451.
10. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics*. 2013;14 Suppl 3(Suppl:3). <https://doi.org/10.1186/1471-2164-14-s3-s2>.
11. Carter H, Douville C, Stenson PD, Cooper DN, Karchin R. Identifying Mendelian disease genes with the variant effect scoring tool. *BMC Genomics*. 2013;14(3):1–16.
12. Casadio R, Calabrese R, Capriotti E, Fariselli P, Martelli PL. Protein folding, misfolding and diseases: the I-mutant suite. *HIBIT*. 2008;2008:67.
13. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–404. AACR.
14. Chai C-Y, Maran S, Thew H-Y, et al. Predicting deleterious non-synonymous single nucleotide polymorphisms (nsSNPs) of HRAS gene and in silico evaluation of their structural and functional consequences towards diagnosis and prognosis of cancer. *Biology*. 2022;11(11):1604.
15. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins Struct Funct Bioinf*. 2006;62(4):1125–1132.
16. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745–2747.
17. Dash R, Mitra S, Munni YA, et al. Computational insights into the deleterious impacts of missense variants on N-acetyl-d-glucosamine kinase structure and function. *Int J Mol Sci*. 2021;22(15):8048.
18. Lohning EA, Levonis MS, Williams-Noonan B, Schweiker SS. A practical guide to molecular docking and homology modelling for medicinal chemists. *Curr. Topics Med. Chem.*. 2017;17(18):2023–2040.
19. Gerstung M, Pellagatti A, Malcovati L, et al. Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun*. 2015;6:5901. <https://doi.org/10.1038/ncomms6901>.
20. Hobbs GA, Der CJ, Rossman KL. RAS isoforms and mutations in cancer at a glance. *J Cell Sci*. 2016;129(7):1287–1292.
21. Irahara N, Baba Y, Noshio K, et al. NRAS mutations are rare in colorectal cancer. *Diagn Mol Pathol*. 2010;19(3):157–163. <https://doi.org/10.1097/PDM.0b013e3181c93fd1>.
22. Islam R, Rahaman M, Hoque H, et al. Computational and structural based approach to identify malignant nonsynonymous single nucleotide polymorphisms associated with CDK4 gene. *PLoS One*. 2021;16(11):e0259691.
23. Jagadeesh KA, Wenger AM, Berger MJ, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. 2016;48(12):1581–1586. <https://doi.org/10.1038/ng.3703>.
24. Johnson CW, Reid D, Parker JA, et al. The small GTPases K-Ras, N-Ras, and H-Ras have distinct biochemical properties determined by allosteric effects. *J Biol Chem*. 2017;292(31):12981–12993.
25. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434–443. <https://doi.org/10.1038/s41586-020-2308-7>.
26. 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. <https://doi.org/10.1016/j.sbi.2015.01.003>.
27. Kulshreshtha S, Chaudhary V, Goswami GK, Mathur N. Computational approaches for predicting mutant protein stability. *J Comput Aided Mol Des*. 2016;30(5):401–412.
28. Malumbres M, Barbacid M. RAS oncogenes: the first 30 years. *Nat Rev Cancer*. 2003;3(6):459–465.
29. Miller MP, Kumar S. Understanding human disease mutations through the use of interspecific genetic variation. *Hum Mol Genet*. 2001;10(21):2319–2328.
30. Mitchell ELD, Jones D, White GRM, Varley JM, Koref MFS. Determination of the gene order of the three loci CD2, NGFB, and NRAS at human chromosome band 1p13 and refinement of their localisation at the subband level by fluorescence in situ hybridisation. *Cytogenet Genome Res*. 1995;70(3–4):183–185.
31. Muñoz-Maldonado C, Zimmer Y, Medová M. A Comparative Analysis of Individual RAS Mutations in Cancer Biology. *Front Oncol*. 2019;9. <https://doi.org/10.3389/fonc.2019.01088>.
32. Oliveira JB, Bidère N, Niemela JE, et al. NRAS mutation causes a human autoimmune lymphoproliferative syndrome. *Proc Natl Acad Sci*. 2007;104(21):8953–8958.
33. Pang JM, Chien P-C, Kao M-C, et al. MicroRNA-708 emerges as a potential candidate to target undruggable NRAS. *PLoS One*. 2023;18(4):e0284744.
34. Pimanda JE, Donaldson IJ, de Bruijn MFTR, et al. The SCL transcriptional network and BMP signaling pathway interact to regulate RUNX1 activity. *PNAS*. 2007;104(3):840–845. <https://doi.org/10.1073/pnas.0607196104>.
35. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. *Cancer Res*. 2012;72(10):2457–2467.
36. Raybaud F, Noguchi T, Marics I, et al. Detection of a low frequency of activated ras genes in human melanomas using a tumorigenicity assay. *Cancer Res*. 1988;48(4):950–953.
37. Rogers MF, Shihab HA, Gaunt TR, Campbell C. CScape: a tool for predicting oncogenic single-point mutations in the cancer genome. *Sci Rep*. 2017;7(1):11597. <https://doi.org/10.1038/s41598-017-11746-4>.
38. Rozario LT, Sharker T, Nila TA. In silico analysis of deleterious SNPs of human MTUS1 gene and their impacts on subsequent protein structure and function. *PLoS One*. 2021;16(6):e0252932.
39. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308–311. <https://doi.org/10.1093/nar/29.1.308>.
40. Shihab HA, Gough J, Cooper DN, et al. Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models. *Hum Mutat*. 2013;34(1):57–65. <https://doi.org/10.1002/humu.22225>.
41. Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(W1):W452–W457.
42. Szklarczyk D, Franceschini A, Kuhn M, et al. The STRING database in 2011: Functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011;39(SUPPL. 1):561–568. <https://doi.org/10.1093/nar/gkq973>.
43. Szklarczyk D, Morris JH, Cook H, et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res*. 2017;45(D1):D362–D368. <https://doi.org/10.1093/nar/gkw937>.
44. Takács T, Kudlik G, Kurilla A, Szeder B, Buday L, Vas V. The effects of mutant Ras proteins on the cell signalome. *Cancer Metastasis Rev*. 2020;39(4):1051–1065.
45. Takahashi M, Hosomichi K, Nakaoka H, et al. Biased expression of mutant alleles in cancer-related genes in esophageal squamous cell carcinoma. *Esophagus*. 2022;19(2):294–302. <https://doi.org/10.1007/s10388-021-00900-7>.
46. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98–W102.
47. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31(2):455–461.
48. Volmar AY, Gutierrez H, Zhou H, et al. Mechanisms of isoform-specific residue influence on GTP-bound HRas, KRas, and NRas. *Biophys J*. 2022;121(19):3616–3629.
49. Wang DG, Fan J-B, Siao C-J, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*. 1998;280(5366):1077–1082.
50. Wang Z, Moul J. SNPs, protein structure, and disease. *Hum Mutat*. 2001;17(4):263–270.
51. Wittinghofer A, Vetter IR. Structure-function relationships of the G domain, a canonical switch motif. *Annu Rev Biochem*. 2011;80:943–971.
52. Wu J, Jiang R. Prediction of deleterious nonsynonymous single-nucleotide polymorphism for human diseases. *Scientific World Journal*. 2013;2013. <https://doi.org/10.1155/2013/675851>.
53. Yazar M, Özbek P. In silico tools and approaches for the prediction of functional and structural effects of single-nucleotide polymorphisms on proteins: an expert review. *OMICS: A J Integr Biol*. 2021;25(1):23–37.
54. Yue P, Li Z, Moul J. Loss of protein structure stability as a major causative factor in monogenic disease. *J Mol Biol*. 2005;353(2):459–473.