

# Leucine to proline substitution by SNP at position 197 in Caspase-9 gene expression leads to neuroblastoma: a bioinformatics analysis

Arpita Kundu · Susmita Bag · Sudha Ramaiah · Anand Anbarasu

Received: 22 July 2012 / Accepted: 24 August 2012 / Published online: 18 September 2012  
© The Author(s) 2012. This article is published with open access at Springerlink.com

**Abstract** To understand the role of CASP9 (Caspase-9) gene products in relation to neuroblastoma disease, we have analyzed the single nucleotide polymorphisms (SNPs) associated with this gene. This can help us understand the genetic variations that can alter the function of the gene products. A total of 941 SNPs are investigated for CASP9 gene. To determine whether a non-synonymous SNP (nsSNP) in this gene affects its protein product, we used certain computational tools which predicted one nsSNP, rs1052574, to have deleterious phenotypic effect. This polymorphic variant results in amino acid substitution from leucine to proline at 197 position, i.e., from acyclic amino acid to a 5-membered amino acid which resides in the buried area of the protein with a high level of conservation. This amino acid substitution shows a transition from helix to coil in the mutant protein. Hence, due to the complete alteration in the structural property of the amino acid side chain, the stability of the protein is reduced which may affect the function of CASP9 protein, leading to deregulation of apoptosis and neuroblastoma development.

**Keywords** CASP9 · Leucine · Neuroblastoma · Proline · rs1052574

## Introduction

Neuroblastoma (NB) is the most common extracranial tumor of childhood, arising from neural crest cells, accounting for approximately 10 % of pediatric cancers (Gale et al. 1982; Brodeur et al. 1992; Schor 1999). NB is characterized by diverse behavior ranging from rapid malignant progression to spontaneous regression (Lastowska et al. 2001). It has been suggested that genetic susceptibility to NB is now highly probable (Shojaei-Brosseau et al. 2004). Attention has been focused on determining the specific genetic alterations in tumors affecting the prognosis and leading to targeted therapies for the individual cancer patient (Heinrichs and Look 2007). The most common form of genetic variation in the human genome are single nucleotide polymorphisms (SNPs), accounting for heritable inter-individual variability in complex phenotypes (Liaoa and Lee 2010; Suh and Vijg 2005). Several biological markers related to the outcome of NB disease have been identified (Cattelani et al. 2008). Apoptosis, the process of cell elimination play a vital role in maintaining cellular homeostasis, cell proliferation and differentiation. Disturbances in the cell death process may lead to uncontrolled cell growth and tumor formation (Zhivotovsky and Orrenius 2006). It has been proposed that aberrations in apoptosis contribute to NB progression (Abel et al. 2005). Caspase-9 (CASP9) gene, a key regulator of the apoptotic signaling system is mapped to the consensus region deleted in all NB cases with 1p deletion (Ohira et al. 2000). Thus, it is considered to be a good candidate gene for NB (Abel et al. 2002). Recent evidence suggests that CASP9, a critical member of the mitochondrial-mediated apoptotic protease cascade, is expressed to a low extent in tumors of NB patients, suggesting that dysregulation of apoptosis is likely to be instrumental in the development or progression of childhood tumor neuroblastoma (Abel

**Electronic supplementary material** The online version of this article (doi:10.1007/s13205-012-0088-y) contains supplementary material, which is available to authorized users.

A. Kundu · S. Bag · S. Ramaiah · A. Anbarasu (✉)  
Medical and Biological Computing Laboratory,  
School of Biosciences and Technology, VIT University,  
Vellore 632014, Tamil Nadu, India  
e-mail: aanand@vit.ac.in

et al. 2002, 2005). Single nucleotide changes in CASP9 gene, leading to the reduced expression of the protein, have been studied in NB tumors (Abel et al. 2002). Bioinformatics tools used to screen the potentially deleterious SNPs based on the gene of interest have been documented (Mah et al. 2011). Attention has been focused on non-synonymous SNPs (nsSNPs) for an association study of genetic diseases which can be useful to examine the potential impact of an amino acid variant on the function of the encoded protein (Johnson et al. 2005). The substitution of one amino acid for another generally results in conformational changes in the immediate vicinity of the substituted site which leads to a significant alteration in thermodynamic stability of the single mutant site from that of the corresponding native protein (Shortle and Sonddek 1995; Querol et al. 1996). A computational approach has also been employed to study the effect of the protein stability upon mutation (Guerois et al. 2002). Hydrogen bonding, being one of the major structural determinants in protein molecules, helps us understand protein structure and motions. It also contributes to the specificity of intramolecular interactions in biological systems (Kortemea et al. 2003). Such type of interactions can be affected by any amino acid variant in the protein molecule due to mutation (Wang and Moulton 2001). Thus, to analyze the intramolecular interactions upon mutations, we have carried out a computational analysis of the hydrogen bonds (H-bonds) across the modeled protein molecules.

We assume that there are no bioinformatics approaches to document the decreased expression of mutated CASP9 gene. This prompted us to carry out the analysis of phenotypic impact of nsSNPs of CASP9 gene and their effect on structural stability of the mutated protein. Our results will provide an insight to the researchers for understanding the regulatory role played by CASP9 in apoptosis and the genetic consequence of NB.

## Materials and methods

### Data source

The SNPs associated with CASP9 gene were obtained from the single nucleotide polymorphism database (dbSNP) (Wheeler et al. 2008) and the reference SNP (rs) IDs are listed in the Supplementary table. There was a total of 941 SNPs associated with CASP9.

### F-SNP identifies nsSNPs by mining Ensembl database

For selecting nsSNPs of CASP9 gene, we used functional single nucleotide polymorphism (F-SNP) database (Lee and Shatkay 2008). The F-SNP database identified nsSNPs that had deleterious effects on protein structure or function,

or impede post-translational modification (Lee and Shatkay 2008). The Ensembl database (Hubbard et al. 2009; Ramensky et al. 2002) was mined to identify nsSNPs. Ensembl provided the number of synonymous and non-synonymous SNPs related to the gene of interest. Each gene had an associated GeneSNPView page which showed SNP locations and effects related to gene and protein structures (Reumers et al. 2005).

### Functional effects of nsSNPs are assessed using F-SNP database

We used F-SNP database which provided a comprehensive collection of functional information about SNPs with respect to four biomolecular functional categories: protein coding, splicing regulation, transcriptional regulation and post-translation effects. To obtain the functional effects of SNPs, it made use of large variety of publicly available tools and resources (Lee and Shatkay 2008).

### Mining different tools by F-SNP to predict the functional effects of nsSNPs in protein coding region

The tools that were utilized for identifying the coding nsSNPs include Sorting Intolerant from Tolerant (SIFT) (Ng and Henikoff 2003; Shen et al. 2006), Polymorphism Phenotyping (PolyPhen) (Johnson et al. 2005; Zhu et al. 2008), SNPeffect (Reumers et al. 2005, 2006), large-scale annotation of coding nsSNPs (LS-SNP) (Karchin et al. 2005; Ryan et al. 2009) and SNPs3D (Yue et al. 2006).

### Identifying changes in the splicing regulation system by F-SNP

F-SNP database mined certain computational prediction tools that were developed for locating splicing elements and identifying the exon/intron structures of genes (Cartegni et al. 2003). Exonic splicing enhancer (ESEfinder) (Cartegni et al. 2003), relative enhancer and silencer classification by unanimous enrichment (RESCUE-ESE) (Fairbrother et al. 2004), exonic splicing regulator (ESR-Search) (Fairbrother et al. 2002) and putative exonic splicing enhancer (PESE) (Zhang et al. 2005) were used for locating SNPs in exonic-splice regions.

### F-SNP mines different tools to examine transcriptional regulation and post-translational modification sites

Golden Path was mined to identify SNPs in transcriptional regulatory regions (Kuhn et al. 2007). OGPET (Gerken et al. 2004) was used to examine post-translational

modification sites. OGPET predicted the O-glycosylation sites (Lanver et al. 2010).

### Predicting the phenotypic effect of nsSNPs

We employed the nsSNPAnalyzer (Bao et al. 2005) to predict nsSNP's phenotypic effect (disease associated vs. neutral). nsSNPAnalyzer annotated the structural environment of each SNP site using the ENVIRONMENT program (Bowie et al. 1991). The program combined three parameters: (i) area of the side chain which was classified as buried, partially buried and exposed according to its solvent accessibility. The buried class was subdivided into three classes (B1, B2 and B3) in order of increasing environmental polarity. The partially buried class was subdivided into P1 and P2 in order of increasing polarity. The exposed side chain was labeled as E; (ii) fraction polar score, giving a measure of environmental polarity related to hydrogen bond formation; and (iii) secondary structure (helix, sheet or coil). The server also integrated evolutionary information and the normalized probability of the substitution was calculated using SIFT program (Ng and Henikoff 2003). nsSNPAnalyzer then used a machine learning method, Random Forests (Bao et al. 2005), to predict the phenotypic class of nsSNPs (Bao and Cui 2005).

### Identifying highly conserved positions in protein sequence

The ConSurf server was used for calculating the evolutionary conservation of amino acid positions in proteins using an empirical Bayesian inference (Ashkenazy et al. 2010). It automated the algorithmic tools to identify the functionally important regions in query proteins by surface mapping of the level of conservation of the amino acid sites among their close sequence homologs (Glaser et al. 2003). The conservation scores were divided into a discrete scale of nine grades for visualization, from the most variable positions (grade 1), through intermediately conserved positions (grade 5), to the most conserved positions (grade 9) (Ashkenazy et al. 2010).

### Comparative modeling of wild and mutant CASP9 proteins

Modeller 9.10, comparative modeling software was executed to build protein models from the templates obtained from sequence similarity with the target protein sequence (Fisher and Sali 2003). The prediction process consisted of target-template alignment, model building and model evaluation (Eswar et al. 2007).

### Analyzing the structural effect of protein upon mutation

SNPeffect server (Reumers et al. 2005) was employed to analyze the structural effect of the nsSNP. The server utilized the force-field FoldX (Guerois et al. 2002) to evaluate the protein stability upon mutation. FoldX program estimated the mutational free energy change on the stability of a protein (Schymkowitz et al. 2005).

### Hydrogen-bonding interactions

To understand intramolecular interactions upon mutations, we have carried out an analysis of H-bonds across the wild and mutant protein molecules, respectively. We employed the Swiss-PDB viewer program to visualize protein complexes and compute H-bonding (Guex 1996; Guex and Peitsch 1996).

## Results

### Identification of nsSNPs and their selection using F-SNP database

The F-SNP database resource (Lee and Shatkay 2008) lists 11 nsSNPs out of the 941 SNPs for CASP9 gene. They are rs1052576, rs1052574, rs2308939, rs2308949, rs2308950, rs1820204, rs2020897, rs2308938, rs2308941, rs9282624 and rs4646008. The corresponding allele change of these nsSNPs is also obtained from F-SNP. SIFT and PolyPhen tools are employed to obtain the corresponding changes in the amino acid residues of these nsSNPs. F-SNP calculates a specific functional significance (FS) score for each of these nsSNPs which signifies their damaging effects (Lee and Shatkay 2008, 2009). FS scores computed by F-SNP database are found to be 0.774, 0.789, 0.977, 0.136, 0.318, 0.888, 0.916, 0.533, 0.749, 0.919 and 0.774 for the 11 nsSNPs rs1052576, rs1052574, rs2308939, rs2308949, rs2308950, rs1820204, rs2020897, rs2308938, rs2308941, rs9282624 and rs4646008, respectively, as depicted in Table 1.

### Functional prediction of nsSNPs using F-SNP database

F-SNP database predicts the deleterious effect of SNPs with respect to protein coding, splicing regulation, transcriptional regulation and post-translational effects (Lee and Shatkay 2008). To obtain the damaging functional effect of nsSNPs, F-SNP integrates multiple tools that are based on different algorithms and computes the FS score for each of them. The deleterious SNP has FS score value between 0.5 and 1 (Lee and Shatkay 2009). Nine nsSNPs (rs1052576, rs1052574, rs2308939, rs1820204, rs2020897,

**Table 1** Lists of nsSNPs identified by F-SNP database with their corresponding alleles and amino acid changes

SNP ID	F-SNP			Amino acid change with position	
	SNP type	Allele change	FS score	SIFT	PolyPhen
rs1052576	Non-synonymous	T <sup>a</sup> /G <sup>b</sup>	0.774	Q <sup>c</sup> (221)R <sup>d</sup>	Q <sup>c</sup> (221)R <sup>d</sup>
rs1052574	Non-synonymous	T <sup>a</sup> /C <sup>e</sup>	0.789	L <sup>f</sup> (197)P <sup>g</sup>	L <sup>f</sup> (197)P <sup>g</sup>
rs2308939	Non-synonymous	C <sup>e</sup> /A <sup>h</sup>	0.977	R <sup>d</sup> (192)S <sup>i</sup>	R <sup>d</sup> (192)S <sup>i</sup>
rs2308949	Non-synonymous	G <sup>b</sup> /A <sup>h</sup>	0.136	G <sup>j</sup> (176)R <sup>d</sup>	G <sup>j</sup> (176)R <sup>d</sup>
rs2308950	Non-synonymous	G <sup>b</sup> /A <sup>h</sup>	0.318	R <sup>d</sup> (173)H <sup>k</sup>	R <sup>d</sup> (173)H <sup>k</sup>
rs1820204	Non-synonymous	T <sup>a</sup> /A <sup>h</sup>	0.888	F <sup>l</sup> (136)L <sup>f</sup>	F <sup>l</sup> (136)L <sup>f</sup>
rs2020897	Non-synonymous	G <sup>b</sup> /C <sup>e</sup>	0.916	E <sup>m</sup> (114)D <sup>n</sup>	E <sup>m</sup> (114)D <sup>n</sup>
rs2308938	Non-synonymous	C <sup>e</sup> /T <sup>a</sup>	0.533	L <sup>f</sup> (106)F <sup>l</sup>	L <sup>f</sup> (106)F <sup>l</sup>
rs2308941	Non-synonymous	C <sup>e</sup> /T <sup>a</sup>	0.749	T <sup>o</sup> (102)I <sup>p</sup>	T <sup>o</sup> (102)I <sup>p</sup>
rs9282624	Non-synonymous	C <sup>e</sup> /G <sup>b</sup>	0.919	I <sup>p</sup> (185)M <sup>q</sup>	I <sup>p</sup> (185)M <sup>q</sup>
rs4646008	Non-synonymous	C <sup>e</sup> /T <sup>a</sup>	0.774	S <sup>i</sup> (99)L <sup>f</sup>	S <sup>i</sup> (99)L <sup>f</sup>

*F<sub>s</sub>* Functional significance

<sup>a</sup> Thymine

<sup>b</sup> Guanine

<sup>c</sup> Glutamine

<sup>d</sup> Arginine

<sup>e</sup> Cytosine

<sup>f</sup> Leucine

<sup>g</sup> Proline

<sup>h</sup> Adenine

<sup>i</sup> Serine

<sup>j</sup> Glycine

<sup>k</sup> Histidine

<sup>l</sup> Phenylalanine

<sup>m</sup> Glutamate

<sup>n</sup> Aspartate

<sup>o</sup> Threonine

<sup>p</sup> Isoleucine

<sup>q</sup> Methionine

rs2308938, rs2308941, rs9282624 and rs4646008) are found to have significant FS scores in the range of 0.5–1 as depicted in Table 1.

The nsSNP, rs1052576, with FS score of 0.774 is found to have a deleterious effect on the protein coding region as predicted by SNP-effect tool. Splicing regulation system is found to be altered by ESEfinder. Changes in transcriptional regulatory region and in post-translational modification site are examined by Golden Path and OGPET, respectively. The nsSNP, rs1052574, with FS score of 0.789 is predicted to be damaging by SIFT, PolyPhen, SNPeffect, LS-SNP and SNPs3D tools, which in turn are queried by F-SNP database to predict the functional impact on protein coding region. ESEfinder and ESRSearch predict changes in the splicing regulation system. Golden Path examines a change in transcriptional regulatory region. For nsSNP, rs2308939, with FS score of 0.977, protein coding

region is predicted to be deleterious by PolyPhen, SNPeffect and SNPs3D tools. Splicing regulation system, transcriptional regulatory region and post-translational modification site for rs2308939 are examined to have changes by their respective tools (Table 2).

For nsSNP, rs1820204 with FS score 0.888, PolyPhen predicts a deleterious effect in protein coding region. Splicing regulation system, transcriptional regulatory region and post-translational modification site for rs1820204 are examined to have changes by their respective tools. The nsSNP, rs2020897 with FS score 0.916, protein coding region is predicted to be deleterious by SIFT and SNPs3D. Changes in transcriptional regulatory region and in post-translational modification site are examined by Golden Path and OGPET, respectively. SNPs3D predicts a deleterious effect on protein coding region for rs2308938 with FS score of 0.533. A change in exonic-splice region is predicted by

**Table 2** Functional effect of nsSNPs by F-SNP database

Functional category	Prediction tools	Prediction results of SNPs				
		rs1052576	rs1052574	rs2308939	rs1820204	rs2020897
		FS score: 0.774	FS score: 0.789	FS score: 0.977	FS score: 0.888	FS score: 0.916
Protein coding	PolyPhen	Benign	Damaging	Damaging	Damaging	Benign
	SIFT	Tolerated	Damaging	Tolerated	Tolerated	Damaging
	SNP-effect	Deleterious	Deleterious	Deleterious	No entry	Benign
	LS-SNP	Benign	Deleterious	Benign	Benign	Benign
	SNPs3D	Benign	Deleterious	Deleterious	Benign	Deleterious
Splicing	ESEfinder	Changed	Changed	Changed	Changed	Changed
Regulation	ESRSearch	Not changed	Changed	Changed	Changed	Changed
	PESE	Not changed	Not changed	Changed	Changed	Changed
	RESCUE-ESE	Not changed	Not changed	Changed	Changed	Changed
Transcriptional regulation	Golden Path	Exist	Exist	Exist	Exist	Exist
Post-translation	OGPET	Exist	Not exist	Exist	Exist	Not exist
		rs2308938	rs2308941	rs9282624	rs4646008	
		FS score: 0.533	FS score: 0.749	FS score: 0.919	FS score: 0.774	
Protein coding	PolyPhen	Benign	Damaging	Damaging	Damaging	
	SIFT	Tolerated	Tolerated	Tolerated	Tolerated	
	SNP-effect	Benign	Deleterious	Benign	Deleterious	
	LS-SNP	Benign	Benign	Benign	Benign	
	SNPs3D	Deleterious	Deleterious	Benign	No entry	
Splicing regulation	ESEfinder	Not changed	Changed	Changed	Not changed	
	ESRSearch	Changed	Changed	Changed	Changed	
	PESE	Changed	Not changed	Changed	Not changed	
	RESCUE-ESE	Changed	Not changed	Changed	Changed	
Transcriptional regulation	Golden Path	Exist	Exist	Exist	Exist	
Post-translation	OGPET	Not exist	Exist	Not exist	Exist	

FS Functional significance

ESRSearch, PESE and RESCUE-ESE. Golden Path examines a change in the transcriptional regulatory region (Table 2).

The nsSNP, rs2308941, with FS score 0.749 is predicted to be deleterious by PolyPhen, SNPeffect and SNPs3D tools. A change exists in exonic-splice region as predicted by ESEfinder and ESRSearch. Changes in transcriptional regulatory region and in post-translational modification site are examined by Golden Path and OGPET, respectively. For nsSNP, rs9282624 with FS score 0.919, PolyPhen predicts a deleterious effect on protein coding region. Changes in splicing regulation system and transcriptional regulatory region are examined by their respective tools. For nsSNP, rs4646008 with FS score 0.774, protein coding region is predicted to be deleterious by PolyPhen and SNP-effect. ESRSearch and RESCUE-ESE predict changes in splicing regulation system. Changes in transcriptional regulatory region and in post-translational modification site are examined by Golden Path and OGPET, respectively (Table 2).

Phenotypic effect of nsSNPs predicted by nsSNPAnalyzer

The nine nsSNPs (rs1052576, rs1052574, rs2308939, rs1820204, rs2020897, rs2308938, rs2308941, rs9282624 and rs4646008) having significant FS scores are annotated by nsSNPAnalyzer to evaluate their phenotypic class (disease associated vs. neutral) (Bao et al. 2005). The ENVIRONMENT program (Bowie et al. 1991) evaluates the structural environment of the SNP site based on three structural parameters (area buried, fraction polar and secondary structure). Among these nine nsSNPs, rs1052574 is annotated to be in the buried area (B1) with a score of 0.637 and fraction polar score of 0.188. A deleterious SIFT score of 0.00 is also predicted for rs1052574 and hence, this nsSNP is classified to be disease associated by nsSNPAnalyzer. The nsSNP, rs2308939 is classified to be neutral which is in the partially buried category of P2 with area buried score of 0.351, fraction polar score of 0.771 and SIFT score of 0.58. Another nsSNP, rs9282624 is classified



to be neutral which is in the partially buried category of P2 with area buried score of 0.401, fraction polar score of 0.760 and SIFT score of 0.07. The other six nsSNPs (rs1052576, rs1820204, rs2020897, rs2308938, rs2308941 and rs4646008) are not evaluated by the ENVIRONMENT program. Hence, rs1052574 is predicted to be functionally important. The phenotypic classes of nsSNPs are depicted in Table 3.

Functionally important regions of protein identified by the ConSurf server

To estimate the degree of conservation of the amino acid residues among their close sequence homologs, we have accessed the ConSurf server. Multiple sequence alignment is obtained by CLUSTALW for the query protein sequences with their respective homologous sequences, and a position-specific conservation level score is assigned for each amino acid residue in the alignment using the empirical Bayesian algorithms (Ashkenazy et al. 2010). The amino acid residue leucine at 197 position (L197) in the native protein is found to be conserved with low negative normalized score of  $-1.048$  and a high conservation score of '9'. The mutated amino acid residue proline at 197 position (P197) for rs1052574 is also found to be conserved with low negative normalized score of  $-0.920$  and a high conservation score of '8'.

Wild and mutant protein structures obtained from Modeller 9.10

The 3D protein structures of both wild and mutant proteins are obtained by comparative homology modeling (Eswar et al. 2007). The program assigns the target sequence and the database of sequences of known structure from PDB as the input to obtain target-template alignments. A better measure of the significance of the alignment is given by the lower expected value (E-value) of the alignment which helps to choose the suitable template candidates (Eswar et al. 2007). The templates (1JXQA, 1NW9A and 2AR9A) having E-value of 0.0 with better sequence coverage with the query protein are chosen for significant sequence alignments. A dendrogram is obtained, which calculates a clustering tree from the input matrix of pairwise distances, to understand the differences among the template candidates (Eswar et al. 2007). The most appropriate template,

1JXQA, with 99 % sequence identity is selected for target-template alignment and construction of the final model. We have shown the native and mutant protein structures of SNP rs1052574 in Fig. 1, where leucine (Leu) residue at 197 position in the native protein is replaced by proline (Pro) residue in the mutant protein, showing helix to coil transition.

Predicting protein stability changes upon mutation

The empirical protein design force-field FoldX (Guerois et al. 2002) is used by SNPeffect (Reumers et al. 2005) to calculate the difference in free energy of the mutation:  $\Delta\Delta G$  (ddG). The homology model build for the query protein sequence is evaluated by the FoldX program and found that the mutation from leucine to proline at position 197 results in a ddG of 7.74 kcal/mol.

Intramolecular H-bonding interactions

The significant nsSNP, rs1052574 of CASP9 producing the amino acid variant L197P shows a helix to coil transition. The effect of intramolecular interactions upon mutation is analyzed by computing H-bonding in the wild and mutant protein molecules. In the native CASP9 protein, leucine197 (L197) interacts with arginine (R193), showing a strong H-bonding with a distance of 2.17 Å between N of L197 and O of R193, as depicted in Fig. 2. Amino acid substitution from L197 to P197 shows a weak H-bonding in the mutant protein with a distance of 3.16 Å between N of P197 and O of R193, as represented in Fig. 2.

## Discussion

We retrieved a total of 941 SNPs associated with CASP9 gene and identified 11 SNPs to be non-synonymous (rs1052576, rs1052574, rs2308939, rs2308949, rs2308950, rs1820204, rs2020897, rs2308938, rs2308941, rs9282624 and rs4646008). Among these, nine nsSNPs (rs1052576, rs1052574, rs2308939, rs1820204, rs2020897, rs2308938, rs2308941, rs9282624 and rs4646008) are found to have significant FS scores in the range of 0.5–1. They are then annotated to identify the disease-associated nsSNPs (Bao et al. 2005). Among them, rs1052574 is annotated to be in the buried area (B1) with a deleterious SIFT score of 0.00

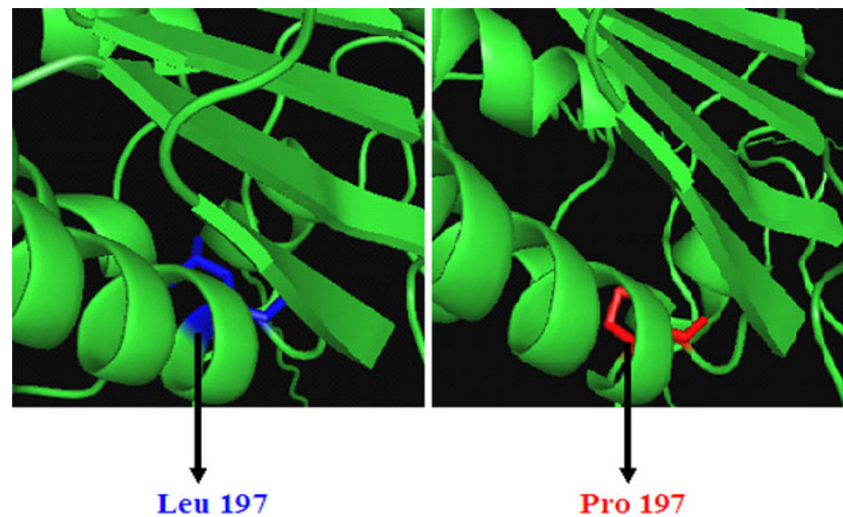
**Table 3** Phenotypic classes of nsSNPs obtained from nsSNPAnalyzer

SNP ID	Area buried	Fraction polar	Environment	Secondary structure	SIFT score	Phenotype
rs1052574	0.637	0.188	B1 <sup>a</sup>	Helix	0.00	Disease
rs2308939	0.351	0.771	P2 <sup>b</sup>	Helix	0.58	Neutral
rs9282624	0.401	0.760	P2 <sup>b</sup>	Helix	0.07	Neutral

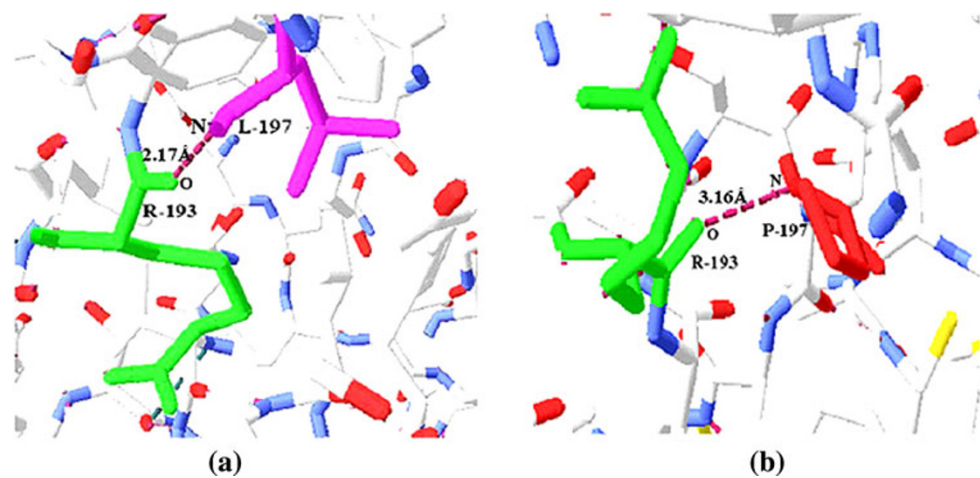
<sup>a</sup> Buried

<sup>b</sup> Partially buried

**Fig. 1** The native and mutant protein structures of SNP rs1052574 where leucine (*Leu*) residue at 197 position in the native protein is replaced by proline (*Pro*) residue in the mutant protein, showing helix to coil transition



**Fig. 2** Intramolecular interactions in the native (a) and mutant (*L197P*) (b) of CASP9 protein models. **a** Hydrogen bond with a distance of 2.17 Å between oxygen (*O*) of arginine (*R193*) and nitrogen (*N*) of leucine (*L197*) in the native CASP9 protein. **b** Hydrogen bond with a distance of 3.16 Å between oxygen (*O*) of arginine (*R193*) and nitrogen (*N*) of proline (*P197*) in the mutant CASP9 protein



and is classified to be disease associated. If the area of the side chain ( $A$ )  $>114 \text{ \AA}^2$ , the residue is placed in the environment class B1 when fraction polar ( $f$ )  $<0.45$ . If  $0.45 \leq f < 0.58$ , then the environment class is B2 and the environment class is B3 when  $f \geq 0.58$ . If  $40 \text{ \AA}^2 < A \leq 114 \text{ \AA}^2$ , the residue is placed in environment category P1 when  $f < 0.67$  and environment class P2 when  $f \geq 0.67$ . A residue is placed in the exposed environment category E if less than  $40 \text{ \AA}^2$  of the side chain is buried (Bowie et al. 1991). SIFT program (Ng and Henikoff 2003) uses an empirical threshold: substitutions with normalized probabilities  $<0.05$  are predicted as deleterious while others are predicted as tolerated. Buried area reflects the solvent accessibility constraint and it is known that disease-associated nsSNPs tend to occur at buried sites (Sunyaev et al. 2000). The SIFT score (Ng and Henikoff 2003) measures the tolerance for a substitution in a multiple sequence alignment and hence incorporates evolutionary information. Thus, rs1052574, residing in the buried site (B1), is predicted to be functionally important. The core residues

play a key role in protein folding and stability, and core mutations are considered more deleterious than surface mutations (Cordes and Sauer 1999). The other eight nsSNPs (rs2308939, rs9282624, rs1052576, rs1820204, rs2020897, rs2308938, rs2308941 and rs4646008) are not classified to be disease associated and are not considered in our study.

The significant change in protein coding region is found for rs1052574 which replaces leucine197 in CASP9 protein with proline due to the nucleotide change at 835th position where CTG is replaced with CCG, i.e., **T** (thymine) to **C** (cytosine). This indicates a transformation of acyclic amino acid to a 5-membered amino acid residue. (The base represented in bold caption is the nsSNP.) Putative ESEs are also predicted for rs1052574 by the change in splicing regulation region. ESEs are short oligonucleotide sequences other than splice sites that enhance splicing from an exonic location. ESEs are recognized by proteins of the SR (serine–arginine) family, which recruit components of the core splicing machinery to nearby splice sites (Fairbrother

et al. 2004). Numerous disease associated polymorphisms exert their effects by disrupting the activity of ESEs (Smith et al. 2006). The functionally important regions in these proteins are identified by quantifying the conservation status of amino acid residues among their close sequence homologs (Ashkenazy et al. 2010; Glaser et al. 2003). Both the wild and mutated amino acid residues at position 197 are found to have high conservation score and low (negative) normalized score. Low (negative) normalized scores indicate the conserved positions, while the high scores indicate the variable ones (Goldenberg et al. 2009). If an amino acid in a particular position of a particular protein is conserved, it indicates that this amino acid may be located in an important or functional region of the protein and that its mutation may cause a significant change of the protein's structure and function (Huang et al. 2010). Comparative study of protein sequences among species infers preliminary information about the protein function, and this has been reported in many genetic diseases. Evolutionarily conserved amino acid residues may serve as a hallmark to identify the functionally critical amino acid of cancer-related genes that may be mutated in tumors (Greenblatt et al. 2003).

Structural information is needed to fully understand the effects and consequences of mutations (Khan and Vihinen 2007). The 3D protein structures of both wild and mutants of CASP9 are obtained by executing comparative homology modeling (Fisher and Sali 2003; Eswar et al. 2007). The L197P substitution for SNP, rs1052574, shows a transition from helix to coil in the modeled protein as shown in Fig. 1. Proline is a very unusual amino acid, in that the side chain cyclizes back on to the backbone amide position. Proline is detrimental to the  $\alpha$ -helical conformation for several reasons. First, the amide proton is replaced by a  $-\text{CH}_2$  group, so it is unable to participate in helix stabilization through intramolecular H-bonding. Second, the bulkiness of its pyrrolidine ring places steric constraints on the conformation of the preceding residue in the helix (Williamson 1994; Li et al. 1996). Recent evidence suggests that L166P variant in parkinson protein 7 confers reduced protein stability which may be the underlying cause of early onset Parkinson's disease (Moore et al. 2003). It is reported that the L166P mutation is located in the  $\alpha$ -helix 7 and is predicted to lead to the unfolding of the C-terminal portion of parkinson protein 7 due to the potent helix breaking properties of the substituted proline (Moore et al. 2003). To analyze the structural effect of the nsSNP, rs1052574 on CASP9 protein, we evaluated the energetic impact of the protein upon mutation from leucine to proline at position 197 (Guerois et al. 2002). The mutational free energy change (ddG) is evaluated by FoldX and is found to be 7.74 kcal/mol. The FoldX error margin is around 0.5 kcal/mol. If the mutation destabilizes the structure, ddG is increased, whereas stabilizing mutations decrease the ddG (Schymkowitz et al.

2005). Our result implies that the mutation severely reduces the protein stability. Correct folding and stability are essential for protein function (Gidalevitz et al. 2009). It is also interesting to note that this residue (L197) located in  $\alpha$  helix (Fig. 1) resides in the buried region of the native protein (Table 3) and thus, mutation occurring in this residue is critical for protein stability.

The proline substitution may be detrimental to the  $\alpha$ -helical conformation by disrupting intramolecular H-bonding (Williamson 1994; Li et al. 1996). Therefore, intramolecular interactions upon mutation are analyzed by computing H-bonds in the native and mutant protein molecules. We have found strong H-bonding in the native protein with a distance of 2.17 Å between L197 and R193, whereas the substitution of L197 to P197 leads to weak H-bonding in the mutant protein with a distance of 3.16 Å between P197 and R193 (Fig. 2). Thus, the substitution of an acyclic amino acid (L197) to a 5-membered amino acid residue (P197), occurring in the buried region of CASP9 protein with weak H-bond interactions, can be critical for the stability in protein monomers. A considerable change in the structure of the protein due to the nsSNP rs1052574 is also observed which shows helix to coil transition in Fig. 1. Hence, due to the change in the structure of the protein and reduction in its stability, the function of the protein may be altered, leading to decreased expression. Protein instability caused due to mutation represents degradation of mutant proteins which can be a prevalent disease-causing mechanism. Degradation of mutant proteins has now been strongly implicated in the etiology of a number of genetic disorders (Waters 2001).

Apoptosis is responsible for the maintenance of homeostasis in tissues as well as in embryonic development. CASP9, belonging to a family of cysteine proteases, is a key regulator of the apoptotic signaling system (Abel et al. 2002). It is an interesting potential tumor suppressor gene in NB, in part due to its localization to human chromosome 1 band p36.1 (Soengas et al. 1999). Thus, the nsSNP, rs1052574, with amino acid variant (L197P) showing a deleterious phenotypic effect and reduced protein stability can decrease the function of the protein which may lead to reduced expression of CASP9. Thereby, it can alter the apoptotic signaling system which can be detrimental and result in the development of NB. Dysregulation of apoptosis is likely to be instrumental in the development of childhood tumor neuroblastoma, and it has been studied that polymorphism in CASP9 gene leads to reduced expression of the protein in NB tumors (Abel et al. 2005; Abel et al. 2002). Mutations in genes encoding proteases, as well as inhibitors and regulators of proteolytic pathways, have been shown to cause numerous human diseases, both inherited and acquired (oncogenic) (Kato 1999; Vu and Sakamoto 2000).



Hence, the results of the nsSNP, rs1052574, for CASP9 gene which is predicted to have a deleterious phenotypic effect and reduced protein stability may be considered to be significant, reducing the function of the apoptotic protease cascade and thereby, leading to the development of neuroblastoma. Our results obtained from these *in silico* studies, may provide an insight into the genetic mechanism of the disease.

**Acknowledgments** Dr. Anand Anbarasu gratefully acknowledges the Indian council of Medical Research (ICMR), Government of India Agency for the research grant-IRIS ID: 2011-03260. We would like to thank the management of VIT for providing us the necessary funds and infrastructure for conducting this project.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Abel F, Sjöberg R-M, Ejeska K et al (2002) Analyses of apoptotic regulators CASP9 and DFFA at 1P36.2, reveal rare allele variants in human neuroblastoma tumours. *Br J Cancer* 86:596–604
- Abel F, Sjöberg R-M, Nilsson S et al (2005) Imbalance of the mitochondrial pro- and anti-apoptotic mediators in neuroblastoma tumours with unfavourable biology. *Eur J Cancer* 41:635–646
- Ashkenazy H, Erez E, Martz E et al (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 38:529–533
- Bao L, Cui Y (2005) Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. *Bioinformatics* 21:2185–2190
- Bao L, Zhou M, Cui Y (2005) nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res* 33:480–482
- Bowie JU, Luthy R, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164–170
- Brodeur GM, Azar C, Brother M et al (1992) Neuroblastoma. *Cancer* 70:1685–1694
- Cartegni L, Wang J, Zhu Z et al (2003) ESEfinder: a web resource to identify exonic splicing enhancers. *Nucleic Acids Res* 31:3568–3571
- Cattalani S, Defferrari R, Marsilio S et al (2008) Impact of a single nucleotide polymorphism in the MDM2 gene on neuroblastoma development and aggressiveness: results of a pilot study on 239 patients. *Clin Cancer Res* 14:3248–3253
- Cordes MH, Sauer RT (1999) Tolerance of a protein to multiple polar-to-hydrophobic surface substitutions. *Protein Sci* 8:318–325
- Eswar N, Webb B, Renom MAM et al (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 2(9):1–31
- Fairbrother WG, Yeo RF, Sharp PA et al (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297:1007–1013
- Fairbrother WG, Yeo GW, Yeh R et al (2004) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. *Nucleic Acids Res* 32:187–190
- Fisher A, Sali A (2003) Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol* 374:461–491
- Gale GB, D'Angio GJ, Uri A et al (1982) Cancer in neonates: the experience at the Children's Hospital of Philadelphia. *Pediatrics* 70:409–413
- Gerken T, Tep C, Rarick J (2004) The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 50-diphosphate-alpha-nacetylgalactosamine: polypeptide *N*-acetylgalactosaminyl transferases T1 and T2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. *Biochemistry* 43:9888–9900
- Gidalevitz T, Krupinski T, Garcia S et al (2009) Destabilizing protein polymorphisms in the genetic background direct phenotypic expression of mutant SOD1 toxicity. *PLoS Genet* 5:399–498
- Glaser F, Pupko T, Paz I et al (2003) ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 19:163–164
- Goldenberg O, Erez E, Nimrod G, Ben-Tal N (2009) The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures. *Nucleic Acids Res* 37:323–327
- Greenblatt MS, Beaudet JG, Gump JR et al (2003) Detailed computational study of p53 and p16: using evolutionary sequence analysis and disease-associated mutations to predict the functional consequences of allelic variants. *Oncogene* 22:1150–1163
- Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320:369–387
- Guex N (1996) Swiss-PdbViewer: a new fast and easy to use PDB viewer for the Macintosh. *Experientia* 52:A26
- Guex N, Peitsch MC (1996) Swiss-PdbViewer: a fast and easy-to-use PDB viewer for Macintosh and PC. *Protein Data Bank Q Newsletter* 77:7
- Heinrichs S, Look AT (2007) Identification of structural aberrations in cancer by SNP array analysis. *Genome Biol* 8:219.1–219.5
- Huang T, Wang P, Ye ZQ et al (2010) Prediction of deleterious non-synonymous SNPs based on protein interaction network and hybrid properties. *PLoS ONE* 5:11900–11907
- Hubbard TJP, Aken BL, Ayling S et al (2009) Ensembl 2009. *Nucleic Acids Res* 37:690–697
- Johnson MM, Houck J, Chen C (2005) Screening for deleterious non-synonymous single-nucleotide polymorphisms in genes involved in steroid hormone metabolism and response. *Cancer Epidemiol Biomark Prev* 4:1326–1329
- Karchin R, Diekhans M, Kelly L et al (2005) LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources. *Bioinformatics* 21:2814–2820
- Kato GJ (1999) Human genetic diseases of proteolysis. *Hum Mutat* 13:87–98
- Khan S, Vihinen M (2007) Spectrum of disease-causing mutations in protein secondary structures. *BMC Struct Biol* 7:56–74
- Kortemmea T, Morozova AV, Baker D (2003) An orientation dependent hydrogen bonding potential improves prediction of specificity and structure for proteins and protein–protein complexes. *J Mol Biol* 326:1239–1259
- Kuhn R, Karolchik D, Zweig AS et al (2007) The UCSC genome browser database: update 2007. *Nucleic Acids Res* 35:668–673
- Lanver D, Mendoza-Mendoza A, Brachmann A et al (2010) Sho1 and Msb2-related proteins regulate appressorium development in the smut fungus *Ustilago maydis*. *Plant Cell* 22:2085–2101

- Lastowska M, Cullinane C, Variend S et al (2001) Comprehensive genetic and histopathologic study reveals three types of neuroblastoma tumors. *J Clin Oncol* 19:3080–3090
- Lee PH, Shatkay H (2008) F-SNP: computationally predicted functional SNPs for disease association studies. *Nucleic Acids Res* 36:820–824
- Lee PH, Shatkay H (2009) An integrative scoring system for ranking SNPs by their potential deleterious effects. *Bioinformatics* 25:1048–1055
- Li S-C, Goto NK, Williams KA et al (1996)  $\alpha$ -Helical, but not  $\beta$ -sheet, propensity of proline is determined by peptide environment. *Biochemistry* 93:6676–6681
- Liaoa PY, Lee KH (2010) From SNPs to functional polymorphism: the insight into biotechnology applications. *Biochem Eng J* 49:149–158
- Mah JTL, Low ESH, Lee E (2011) In silico SNP analysis and bioinformatics tools: a review of the state of the art to aid drug discovery. *Drug Discov Today* 16:800–809
- Moore DJ, Zhang L, Dawson TM et al (2003) A missense mutation (L166P) in DJ-1, linked to familial Parkinson's disease, confers reduced protein stability and impairs homo oligomerization. *J Neurochem* 87:1558–1567
- Ng P, Henikoff S (2003) SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812–3814
- Ohira M, Kageyama H, Mihara M et al (2000) Identification and characterization of a 500-kb homozygously deleted region at 1p36.2-p36.3 in a neuroblastoma cell line. *Oncogene* 19:4302–4307
- Querol E, Perez-Pons JA, Mozo-Vularias A (1996) Analysis of protein conformational characteristics related to thermostability. *Protein Eng* 9:265–271
- Ramensky V, Bork P, Sunyaev S (2002) Human nonsynonymous SNPs: server and survey. *Nucleic Acids Res* 30:3894–3900
- Reumers J, Schymkowitz J, Ferkinghoff-Borg J et al (2005) SNPeff: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. *Nucleic Acids Res* 33:527–532
- Reumers J, Maurer-Stroh S, Schymkowitz J et al (2006) SNPeff v2.0: a new step in investigating the molecular phenotypic effects of human non-synonymous SNPs. *Bioinformatics* 22:2183–2185
- Ryan M, Diekhans M, Lien S et al (2009) LS-SNP/PDB: annotated non-synonymous SNPs mapped to Protein Data Bank structures. *Bioinformatics* 25:1431–1432
- Schor NF (1999) Neuroblastoma as a neurobiological disease. *J Neurooncol* 41:159–166
- Schymkowitz J, Borg J, Stricher F et al (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33:382–388
- Shen J, Deininger PL, Zhao H (2006) Applications of computational algorithm tools to identify functional SNPs in cytokine genes. *Cytokine* 35:62–66
- Shojaei-Brosseau T, Chompret A, Abel A et al (2004) Genetic epidemiology of neuroblastoma: a study of 426 cases at the Institute Gustave-Roussy in France. *Pediatr Blood Cancer* 42:99–105
- Shortle S, Sondak J (1995) The emerging role of insertions and deletions in protein engineering. *Curr Opin Biotechnol* 6:387–393
- Smith PJ, Zhang C, Wang J et al (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. *Hum Mol Genet* 15:2490–2508
- Soengas MS, Alarcon RM, Yoshida H et al (1999) Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science* 284:156–159
- Suh Y, Vijg J (2005) SNP discovery in associating genetic variation with human disease phenotypes. *Mutat Res* 573:41–53
- Sunyaev S, Ramensky V, Bork P (2000) Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 16:198–200
- Vu PK, Sakamoto KM (2000) Ubiquitin-mediated protein degradation in genetic diseases proteolysis and human disease. *Mol Genet Metab* 71:261–266
- Wang Z, Moul J (2001) SNPs, protein structure, and disease. *Hum Mutat* 17:263–270
- Waters PJ (2001) Degradation of mutant proteins, underlying “loss of function” phenotypes, plays a major role in genetic disease. *Curr Issues Mol Biol* 3:57–65
- Wheeler DL, Barrett T, Benson DA et al (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 36:13–21
- Williamson MP (1994) The structure and function of proline-rich regions in proteins. *Biochem J* 297:249–260
- Yue P, Melamud E, Moul J (2006) SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* 7:166–181
- Zhang XHF, Kangsamaksin T, Chao MSP et al (2005) Exon inclusion is dependent on predictable exonic splicing enhancers. *Mol Cell Biol* 25:7323–7332
- Zhivotovsky B, Orrenius S (2006) Carcinogenesis and apoptosis: paradigms and paradoxes. *Carcinogenesis* 27:1939–1945
- Zhu Y, Hoffman A, Wu X et al (2008) Correlating observed odds ratios from lung cancer case-control studies to SNP functional scores predicted by bioinformatic tools. *Mutat Res* 639:80–88