Molecular Neurobiology
https://doi.org/10.1007/s12035-023-03460-2

Genomic Variation Affecting MPV and PLT Count in Association with Development of Ischemic Stroke and Its Subtypes

Abhilash Ludhiadch[1] · Sulena[2] · Sandeep Singh[3] · Sudip Chakraborty[4] · Dixit Sharma[5] · Mahesh Kulharia[6] · Paramdeep Singh[7] · Anjana Munshi[1]

Received: 1 December 2022 / Accepted: 22 June 2023

Abstract

Platelets play a significant role in the pathophysiology of ischemic stroke since they are involved in the formation of intravascular thrombus after erosion or rupture of the atherosclerotic plaques. Platelet (PLT) count and mean platelet volume (MPV) are the two significant parameters that affect the functions of platelets. In the current study, MPV and PLT count was evaluated using flow cytometry and a cell counter. SonoClot analysis was carried out to evaluate activated clot timing (ACT), clot rate (CR), and platelet function (PF). Genotyping was carried out using GSA and Sanger sequencing, and expression analysis was performed using RT-PCR. *In silico* analysis was carried out using the GROMACS tool and UNAFold. The interaction of significant proteins with other proteins was predicted using the STRING database. Ninety-six genes were analyzed, and a significant association of *THPO* (rs6141) and *ARHGEF3* (rs1354034) was observed with the disease and its subtypes. Altered genotypes were associated significantly with increased MPV, decreased PLT count, and CR. Expression analysis revealed a higher expression in patients bearing the variant genotypes of both genes. *In silico* analysis revealed that mutation in *the THPO* gene leads to the reduced compactness of protein structure. mRNA encoded by mutated *ARHGEF3* gene increases the half-life of mRNA. The two significant proteins interact with many other proteins, especially the ones involved in platelet activation, aggregation, erythropoiesis, megakaryocyte maturation, and cytoskeleton rearrangements, suggesting that they could be important players in the determination of MPV values. In conclusion, the current study demonstrated the role of higher MPV affected by genetic variation in the development of IS and its subtypes. The results of the current study also indicate that higher MPV can be used as a biomarker for the disease and altered genotypes, and higher MPV can be targeted for better therapeutic outcomes.

Keywords Platelets · MPV · Ischemic stroke · Genotypes

✉ Anjana Munshi
anjana.munshi@cup.edu.in

[1] Complex Disease Genomics and Precision Medicine Laboratory, Department of Human Genetics and Molecular Medicine, Central University of Punjab, Ghudda, Bathinda, Punjab 151401, India

[2] Department of Neurology, Guru Gobind Singh Medical College and Hospital, Sadiq Road, Faridkot, Punjab 151203, India

[3] MD Anderson Cancer Center, Houston, TX, USA

[4] Department of Computational Sciences, School of Basic and Applied Sciences, Central University of Punjab, Ghudda, Bathinda, Punjab 151401, India

[5] Department of Animal Sciences, School of Life Sciences, Central University of Himachal Pradesh, Kangra, Himachal Pradesh 176206, India

[6] Centre for Computational Biology and Bioinformatics, School of Life Sciences, Central University of Himachal Pradesh, Kangra, Himachal Pradesh 176206, India

[7] Department of Radiodiagnosis, All India Institute of Medical Sciences, Bathinda, Punjab 151001, India

Published online: 15 July 2023

Introduction

Stroke or 'brain attack' is one of the major causes of mortality and morbidity worldwide after myocardial infarction [1]. Stroke is mainly divided into two types: ischemic and hemorrhagic stroke [2]. Eighty-seven percent of the cases are of ischemic stroke (IS) type, 10% intracerebral hemorrhage (ICH), and 3% of subarachnoid hemorrhage (SAH) [3]. IS occurs by an obstruction in the blood supply to the brain, which is usually caused by the formation of a thrombus leading to cell death in the brain [4]. Various modifiable and non-modifiable risk factors are associated with the development of stroke. The modifiable risk factors, including hypertension, obesity, hyperglycemia, hyperlipidemia, atherosclerosis, thrombosis, renal dysfunction, and Sickle Cell Disease, contribute to an 87% risk of stroke. The non-modifiable risk factors include family history, age, and ethnicity. It has also been established that stroke has a strong genetic component [5–7]. Genes involved in various pathways, including homocysteine metabolism, coagulation and hemostasis, rennin angiotensin aldosterone system, cAMP degradation pathway, inflammation, extracellular matrix, and lipid metabolism, are known to be associated with stroke susceptibility [7–9].

Platelet traits such as mean platelet volume (MPV) and platelet count (PLT) and pathways involved in the recruitment of platelets have also been implicated in the disease pathophysiology [10–12]. Platelets are known to play an important role in the pathophysiology of IS by virtue of their capability in the formation of intravascular thrombus after the erosion or rupture of atherosclerotic plaques [13]. We have already established the association of increased MPV with a degree of disability and rate of clot formation in IS patients [14]. PLT count and MPV are markers of platelet function and activation and are positively associated with platelet reactivity and aggregation [15–17]. An increase in MPV occurs when platelets become activated and swollen spheres instead of quiescent discs. Large platelets are more adhesive and likely to aggregate more than smaller ones [18]. These traits and other platelet functions have been reported to be highly influenced by genetic variation. Various genome-wide association studies (GWAS) involving different populations have demonstrated that genomic alterations are associated with PLT count and MPV [19–21]. Variation involved in the genes taking part in significant processes such as megakaryopoiesis, megakaryocyte/platelet adhesion, platelet formation, and cell cycle regulation has been reported to influence platelet physiology [10, 22, 23]. PLT count and MPV altered by genetic profile have not been evaluated in association with the development of IS and its subtypes. Therefore, the current study has been carried out with an aim to explore the alterations in genes affecting MPV and PLT count and their functional implications associated with IS and its subtypes.

Material and Methods

Study Population

Two hundred IS patients were recruited from Guru Gobind Singh Medical College and Hospital, Faridkot, Punjab, India. The study was approved by the ethical committee of the Central University of Punjab as well as the study hospital. Patients confirmed to have suffered an IS as diagnosed by a CT scan or MRI were included in this study. Patients having HS or TIA were excluded from this study. Patients with major secondary problems like renal, hepatic skeletal, and other neurological disorders were also excluded from the study. As a control group, 200 age and sex-matched healthy individuals without a history of any other medical condition, especially cardiovascular and neurological diseases, were also employed in the study. The present study involved patients between the age group of 27–94 years with a mean age of 61.8±13.3. The controls were between 24 and 72 years, with a mean age of 60.2±11.6. The male: female ratio of patients and controls was 156:44. Written informed consent was obtained from all the recruited subjects. Stroke subtypes were stratified as per TOAST classification [24].

Blood Sample Collection

A total of 5 ml of blood was collected in EDTA and sodium citrate-containing vacutainers with the written informed consent of the participants.

Measurements of Platelet Count and Mean Platelet Volume

MPV and PLT counts were evaluated using an automated cell counter (ABX Micros 60 Hematology System). The values were confirmed using BD Accuri C6 flow cytometer in 50% of the patients, as reported previously [14].

Sonoclot Signature Analysis

The assessment of clot timing in IS patients was carried out using Sonoclot Coagulation and Platelet Function Analyzer (Sienco Inc.: Model no. SCP1) as described in our previous study [14].

DNA Isolation

DNA isolation was carried out using the organic method (phenol-chloroform method) (Russell and Sambrook 2001).

Evaluation of Genomic Alterations Affecting MPV and PLT

The Screening for genetic variations affecting MPV and PLT count was carried out in 17% of IS patients to figure out the gene variants occurring at a higher frequency. This screening was carried out using Global Screening Array (GSA) v3.0 microchip (Illumina Inc.). After analyzing the GSA results, variants of three genes *THPO (*rs6141), *WDR66* (rs7961894), and *ARHGEF3* (rs1354034) were filtered out and all the samples; i.e., 200 patients and 200 controls were screened for these variants using Sanger Sequencing.

Expression Analysis

RNA was isolated from the platelets using the Trizol method. cDNA synthesis was carried out using a cDNA synthesis kit (iScript™ cDNA Synthesis Kit Bio-Rad) as per the manufacturer's instructions, with an equal amount of RNA from each sample. 18s rRNA was used as a housekeeping gene.

Statistical Analysis

All significant variants were tested for Hardy-Weinberg equilibrium. The association of genotypes and alleles with IS (univariate analysis) was estimated by the odds ratio with a 95% confidence interval (CI) and χ^2 analysis using OpenEpi software (version 2.3.1; Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA). The association of gene variants with the disease was confirmed by Multiple Logistic Regression analysis (MLR). The independent variables were decided as the following dummy variables: MPV, 0 for standard and 1 for elevated; hypertension, 0 for normotension and 1 for hypertension; diabetes, 0 for normal and 1 for diabetic; tobacco use, 0 for no tobacco use and 1 for tobacco use; alcohol consumption, 0 for non-alcoholics and 1 for alcohol consumers; family history, 0 for no family history of IS and 1 for family history of IS. The dummy variables for gene variants were 0 for normal homozygous and 1 for heterozygous and mutant homozygote. All the statistical analysis was carried out using SPSS (Version: 28.0.1.1 (15)). Statistical significance was defined as $p<0.05$.

Molecular Modelling and Simulation of ARHGEF3 (Q9NR81) and THPO (P40225)

The amino acid sequence of the proteins was retrieved from UniProt [25]. The AlphaFold Protein Structure Database was used to predict the protein structures [26]. GROMACS 2018.3 modules were used to simulate the system of proteins as well as for the analysis to identify the potential role of mutation on the structural integrity of the two proteins with accession numbers Q9NR81 (Rho guanine nucleotide exchange factor 3, ARHGEF3) and P40225 (Thrombopoietin, THPO). In order to identify any structural deviation and/or stability, we have calculated the RMSD (root mean squared deviation) for the wild and mutated proteins along with the Radius of Gyration (Rg) that again tell us about any changes in the folded structure of the protein or conformational jump during the course of the simulation. As far as the rs1354034 variant of the ARHGEF3 gene is concerned, it is an intronic variant; it was difficult to simulate the mutated ARHGEF3 protein as it does not show any change in the amino acid sequence. Therefore, we used UNAFold to predict the changes caused by this SNP at the mRNA level [27].

Protein-Protein Interaction Networks

The STRING web-based server was used to generate a protein-protein interaction network of THPO and ARHGEF3 proteins at the highest confidence score of 0.700 [28].

Results

In our recent study, we already established the association of various risk factors with IS and its subtypes [14]. A total of 106 variants affecting MPV, PLT count, and platelet reactivity were screened using GSA in 17% of IS patients (Fig. 1a). Three genes *ARHGEF3* (rs1354034, T>C), *THPO* (rs6141, C>T), and *WDR66* (rs7961894, C>T) showed variation based on the results of GSA. These were further validated by subjecting all the samples (200 patients and 200 controls) to Sanger sequencing. No significant association of WDR66 (rs7961894) was found with the disease. In the case of *THPO*, rs6141 (C>T) polymorphism, a significant difference was observed in genotypic distribution between IS patients and controls [for TT vs CC, $X^2=37.09$; $p<0.001$, OR=6.98 (95% CI; 3.60–13.22)] (Fig. 1b and Tables 1 and 2). TT and CT genotypes showed a significant association with the disease (for TT vs CC+CT, $X^2=6.323$; $p<0.005$, OR=1.89 (95% CI; 1.17–3.07)). However, we did not find any significant difference in the distribution of T and C alleles between IS patients and controls. Even after

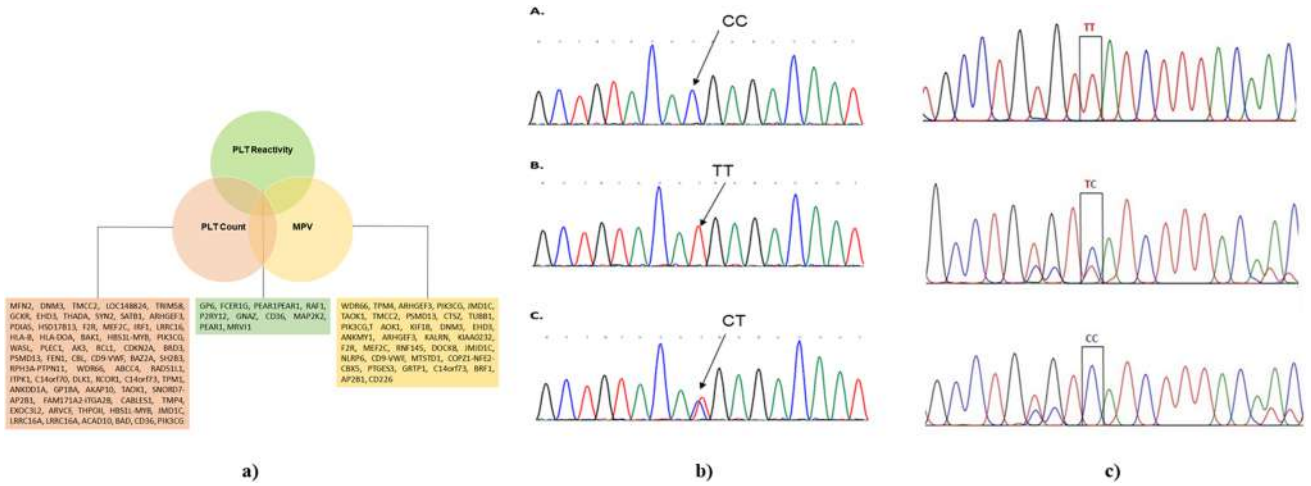

Fig. 1 **a** Genes involved in MPV, PLT Count, and Platelet Reactivity analysed by GSA. **b** Chromatogram showing CC TT and CT genotypes of THPO gene (rs6141). **c** Chromatogram showing TT, TC and CC genotypes of ARHGEF3 gene (rs1354034)

Table 1 Distribution of THPO (rs6141), genotypes and allelic frequencies in ischemic stroke patients

Gene	Genotype	Patients	Controls	Allele	Patients	Controls
THPO (rs6141)	CC	21 (10.50)	89 (44.50)	C	165 (0.41)	255 (0.63)
	CT	123 (61.50)	77 (38.50)			
	TT	56 (28)	34 (17)	T	235 (0.58)	145 (0.36)

Table 2 Chi-square (X^2), crude odds ratio, and p-value for THPO (rs6141) gene variant

Gene	Patients vs controls	Chi-square (X^2)	Odds ratio (95% CI)	p value
THPO (rs6141)	TT vs CC	37.09	6.98 (3.6-13.22)	<0.001
	TT vs CC+CT	6.323	1.89 (1.17-3.07)	<0.005
	T vs C	0.064	1.049 (0.78-1.39)	<0.3

controlling all the confounding risk factors using MLR, a significant association of TT genotype with IS was found ((p = 0.008; adjusted odds ratio −3.834; 95% CI; 1.431–10.277)) (Table 5).

Evaluating the association of the ARHGEF3 (rs1354034, T>C) gene with the disease, a significant difference was observed in the genotypic distribution of CC and TT genotype between IS patients and controls (for CC vs TT, X^2=81.02; p<0.001, OR=20.7 (95% CI; 10.1–42.39)) (Fig. 1c). A significant difference between CC and TT+TC genotypes between IS patients and controls (for CC vs TT+TC, X^2=47.2; p<0.001, OR=5.688 (95% CI; 3.392–9.39)) was also observed (Tables 3 and 4). The C allele was also associated significantly with the disease (C vs T, X^2=84.54; p<0.001, OR=3.9 (95% CI; 2.908–5.23)).

After controlling all the confounding risk factors using MLR significant association of CC genotype with IS was observed ((p = 0.003; adjusted odds ratio- 4.419; 95% CI; 1.657–11.785)) (Table 5).

Evaluating the association of these two variants with IS subtypes, a significant association of TT genotypes of *the THPO* gene, whereas for *the ARHGEF3* gene, a significant association was observed with the C allele as well as CC genotypes (p<0.001) (Tables 6 and 7).

The association of variants of *THPO* (rs6141) and *ARHGEF3* (rs1354034) genes with MPV and PLT count was also evaluated. The altered genotypes of *THPO* (rs6141) and *ARHGEF3* (rs1354034) gene showed a significant association with increased MPV, whereas, for PLT count, an association of the variant genotypes with decreased PLT count was observed, although it did not reach statistical significance (Table 8 and Figs. 2a and b). This is also in consensus with our observation, where an inverse correlation was observed between these two variables [14].

The clotting parameters, including ACT, CR, and PF, were also compared among various genotypes of *THPO* (rs6141) and *ARHGEF3* (rs1354034) genes. A significant

Table 3 Distribution of ARHGEF3 (rs1344034), genotypes, and allelic frequencies in ischemic stroke patients and controls

Gene	Genotype	Patients	Controls	Allele	Patients	Controls
ARHGEF3 (rs1354034)	TT	15 (7.50)	84 (42)	T	130 (0.325)	261 (0.652)
	TC	100 (50)	93 (46.5)			
	CC	85 (42.5)	23 (11.5)	C	270 (0.675)	139 (0.347)

Table 4 Chi-square (X^2), crude odds ratio, and p value for ARHGEF3 (rs1344034) gene variant

Gene	Patients vs controls	Chi-square (X^2)	Odds ratio (95% CI)	p value
ARHGEF3 (rs1354034)	CC vs TT	81.02	20.7 (10.1-42.39)	<0.001
	CC vs TT+TC	47.2	5.688 (3.392- 9.539)	<0.001
	C vs T	84.54	3.9 (2.908- 5.23)	<0.001

Table 5 Independent association of genotypes with IS

Gene	Risk factor (genotype/s)	Adjusted odds ratio	p value
THPO (rs6141)	CT+TT genotype (Heterozygous+ mutant homozygotes)	3.834 (1.431–10.277	p<0.008
ARHGEF3 (rs1354034)	TC+CC (heterozygous + mutant homozygotes)	4.419 (1.657– 11.785)	p<0.003

Table 6 Chi-square (X^2), odds ratio, and p-value for THPO (rs6141) gene variant in IS subtypes

Subtype	THPO (rs6141)	Chi-square (X2)	Odds ratio (95% CI)	p value
Large Artery atherosclerosis	TT vs CC	21.59	5.399 (2.639–11.05)	0.001
	T vs C	25.46	2.274 (1.657–3.121)	0.001
Small artery occlusion	TT vs CC	10.89	8.725 (2.264–33.63)	0.001
	T vs C	13.12	2.528 (1.541–4.147)	0.001
Cardioembolism	TT vs CC	13	13.09 (2.727–62.83)	0.001
	T vs C	1.03	0.2512 (0.03061–2.062)	0.156

Table 7 Chi-square (X^2), odds ratio and p-value for ARHGEF3 (rs1344034) gene variant in IS subtypes

Subtype	ARHGEF3 (rs1354034)	Chi-square (X^2)	Odds Ratio (95% CI)	p value
Large Artery Atherosclerosis	CC vs TT	51.09	15.27 (6.839, 34.1)	<0.001
	C vs T	50.75	3.219 (2.33, 4.447)	<0.001
	CC vs TT	32.87	19.17 (5.984, 61.43)	<0.001
Small artery occlusion	C vs T	35.57	4.78 (2.801, 8.156)	<0.001
	CC vs TT	-	-	-
Cardioembolism	C vs T	37.5	8.919 (4.049, 19.65)	<0.001
Acute stroke of other determined etiology	CC vs TT	-	-	-
	C vs T	7.34	13.14 (1.601, 107.9)	<0.003

Table 8 Association of mean MPV and PLT count with THPO (rs6141) and ARHGEF3 (rs1354034) genotypes

	Genotype	Mean MPV	p value	Mean PLT count	p value
THPO (rs6141)	CC	10.41±1.789	Reference	321.3±126.931	<0.3351
	CT	11.00±1.720	<0.001	313.55±90.025	<1.0000
	TT	11.76±1.794	<0.001	310.66±90.531	Reference
ARHGEF3 (rs1354034)	TT	10.04±1.422	Reference	298.8±87.682	Reference
	TC	11.13±1.871	P < 0.001	315.88±97.854	<0.0667
	CC	11.19±1.767	P < 0.001	313.44±91.988	<0.1041

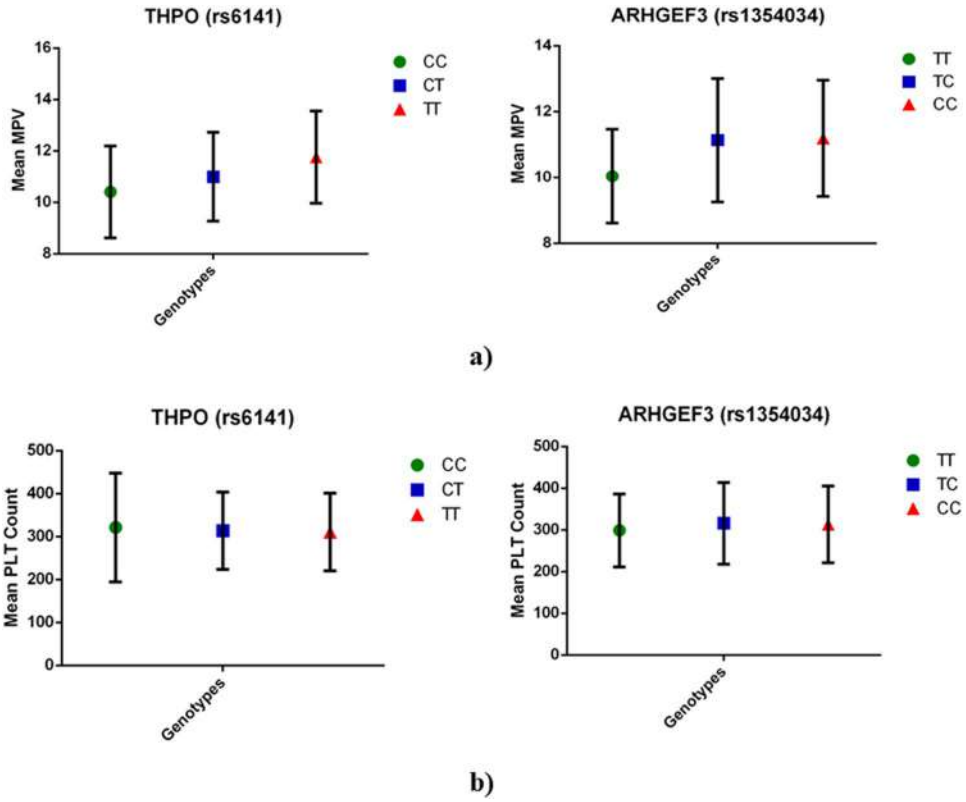

Fig. 2 **a** Association of Mean MPV with different genotypes of THPO (rs6141) and ARHGEF3 (rs1354034) genes. **b** Association of Mean PLT count, and different genotypes of THPO (rs6141) and ARHGEF3 (rs1354034) genes

increase in CR was observed in the patients bearing the altered TT genotype of *THPO* and CC genotype of *ARHGEF3* gene in comparison with normal genotypes of both these genes, CC and TT, respectively. However, we did not find significant differences in ACT and PF values among the variant genotypes of both genes (Tables 9 and 10 and Figs. 3a and b).

Expression analysis carried out by qPCR of *THPO* (rs6141) and *ARHGEF3* (rs1354034) genes revealed higher expression of both genes in patients bearing altered genotypes. The altered genotypes showed significantly higher expression in comparison with heterozygous and normal genotypes ($p<0.05$). Furthermore, the heterozygous genotypes showed a significantly higher expression as compared to the normal genotypes ($p<0.05$). Similarly, the expression of the CC genotype of *the ARHGEF3* (rs1354034) gene was significantly higher in comparison with the heterozygous and normal genotypes ($p<0.05$). After comparing the heterozygous genotype with the normal genotype, it showed significantly higher expression ($p<0.05$) (Fig. 4). The results were normalized against 18s rRNA, which was used as a housekeeping gene to evaluate the expression of both genes.

Table 9 Association of different genotypes of THPO (rs6141) with ACT, CR, and PF

THPO (rs6141)	ACT	*p* value	CR	*p* value	PF	*p* value
CC	104.5±22.86	-	28.25±13.32	-	3.47±1.132	-
CT	140±90.24	0.003	34.7±16.28	0.01	2.36±1.36	<0.0001
TT	105.62±34.48	0.834	59±25.90	<0.001	3.3±0.96	<0.3768

Table 10 Association of different genotypes of ARHGEF3 (rs1354034) with ACT, CR and PF

ARHGEF3 (rs1354034)	ACT	*p* value	CR	*p* value	PF	*p* value
TT	93.5±5.5	-	25.5±1.5	-	2.25±0.25	-
TC	133.58±83.98	0.0003	38.41±19.71	<0.001	2.55±1.257	0.07
CC	121±64.57	0.0013	42.73±26.31	<0.001	2.94±1.40	0.0003

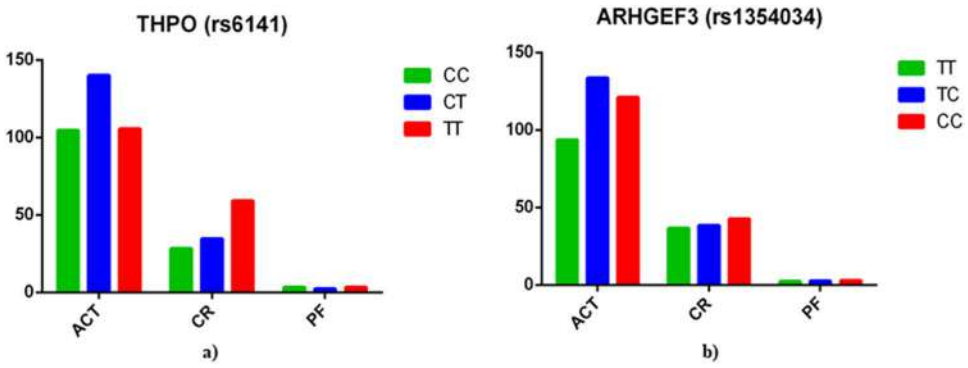

Fig. 3 **a** ACT, CR and PF values in patients bearing CC, CT and TT genotypes of THPO gene. **b** ACT, CR, and PF values in patients bearing TT TC and CC genotypes of ARHGEF3

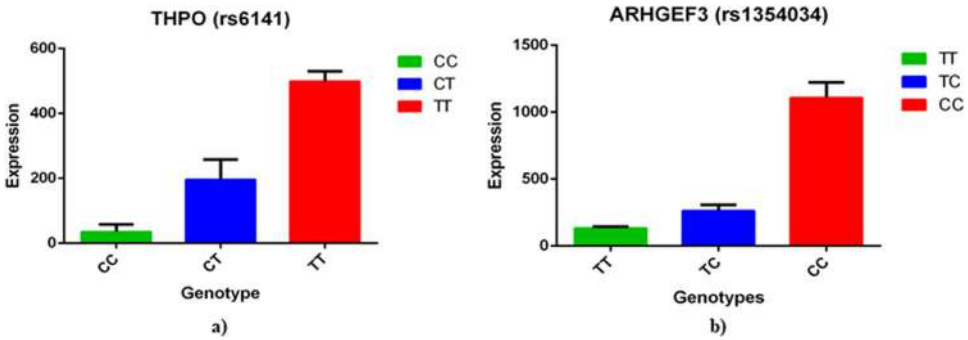

Fig. 4 Expression analysis of different genotypes of **a** THPO (rs6141) and **b** ARHGEF3 (rs1354034) genes

The amino acid sequence of the proteins was retrieved from UniProt [1]. The AlphaFold Protein Structure Database was used to predict the protein structures of both *THPO* and *ARHGEF3* genes. RMSD for both the Proteins (wild and mutated) were calculated using the GROMACS module gmx rms module with respect to a crystal structure as a reference. The stability of a protein relative to the reference structure can be determined by measuring the deviation produced during the simulation. The smaller the deviations, the more stable the simulated structure. RMSD values for all atoms of the mentioned three proteins were calculated for 100ns simulation. It can be observed for P40225 (Thrombopoietin), the wild-type structure was stable around 2.0 nm (3.5 ns) with respect to its crystal structure, but the mutated (R38C) structure showed a sudden jump to a higher rmsd value (2.5 nm) at 4.0 ns and maintained a constant elevation of RMSD value during the course of its simulation (Fig. 5). A similar indication was observed in the case of the Radius of gyration calculation that suggests the P40225 wild-type protein structure is stable during the course of the simulation, whereas in the case of mutated structure, moderate fluctuations were observed. However, the mutated protein shows reduced compactness in the protein structure with respect to the wild variant till the end of the simulation (Figs. 6 and 7).

As far as the rs1354034 variant of the ARHGEF3 gene is concerned, it is an intronic variant; it was difficult to simulate

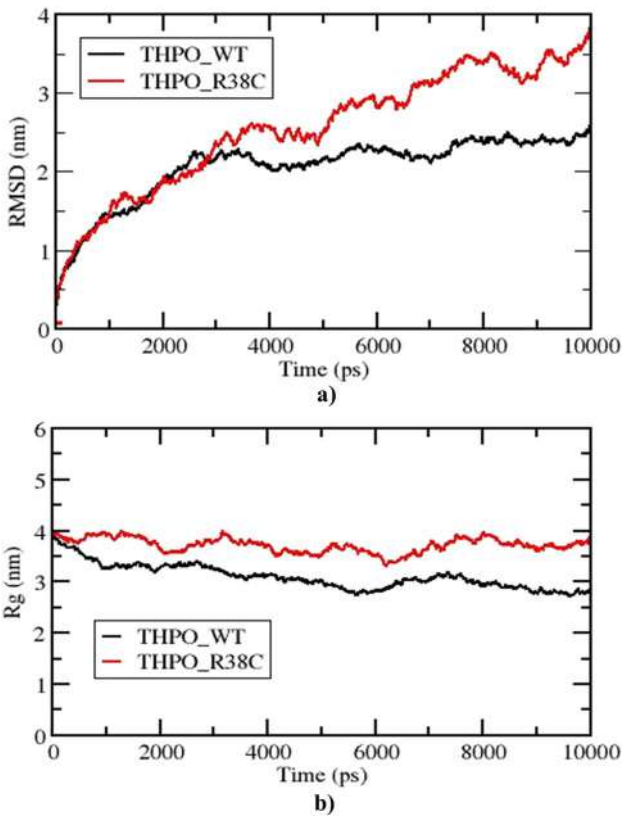

Fig. 5 **a** RMSD values **b** Rg evolutions of THPO (rs6141) (Accession number: P40225) with simulation time in wild (black) and mutated (red) proteins

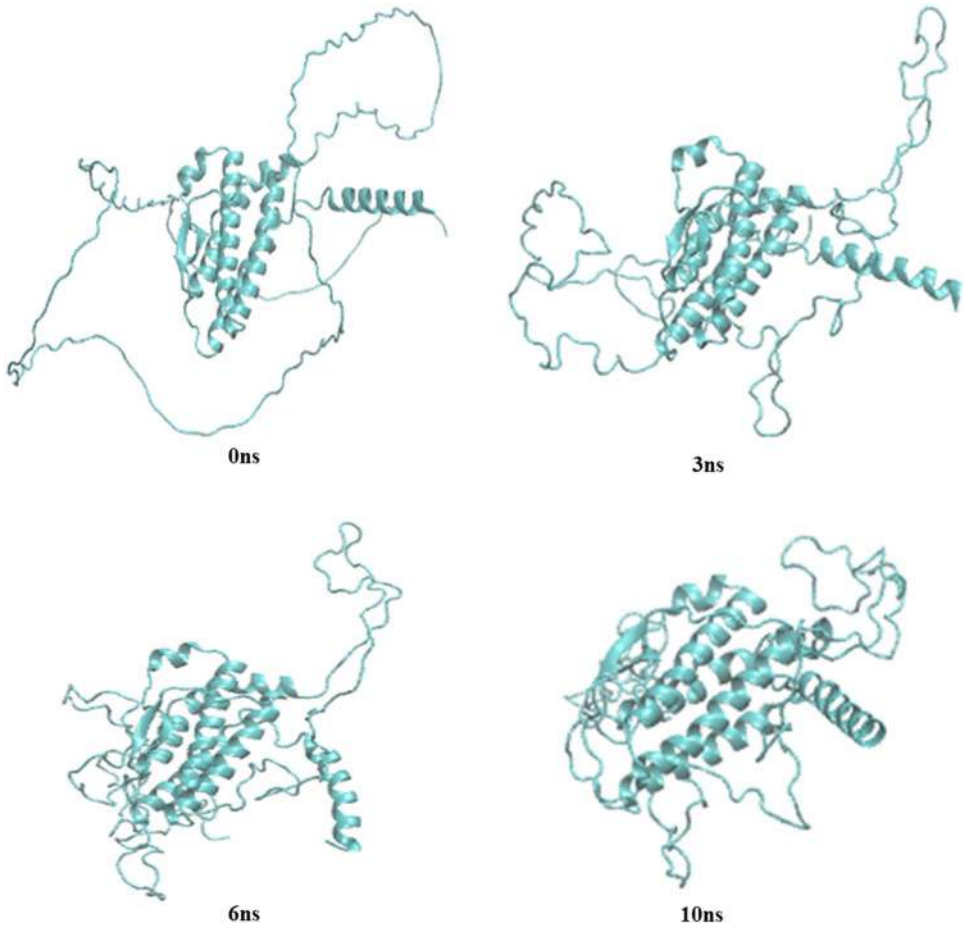

Fig. 6 Ribbon model of wild type protein of Thrombopoietin (Accession number: P40225) during the course of a 100 ns simulation

the mutated ARHGEF3 protein as it does not show any change in the amino acid sequence. Therefore, we used UNAFold to predict the changes caused by this SNP at the mRNA level. UNAFold software package is used to create the simulations of folding, hybridization, and melting pathways for one or two single-stranded RNA or DNA molecules. It combines free energy minimization, partition function calculations, and stochastic sampling to predict the folding of single-stranded RNA or DNA. Further, for melting simulations, the package computes entire melting profiles, not just the melting temperatures [27].

The results showed that the mutant (rs1354034) variant RNA exhibits qualitatively greater stability wrt the free energy associated with the secondary structure. This increased stability manifests in the reduced decay rate of the mutant RNA and, as a consequence, it increased processing into mRNA and its translation into the protein. The examination of the secondary structure of the mutant variant clearly indicates a more organized structure (Fig. 8) with 54 helices as compared to the wild-type sequence (50 helices). The average stem-loop size is also less for the mutant variant vis-à-vis wild type. This reduces the probability of its decay by RNases.

The protein-protein interaction network of THPO showed that it interacts closely with ten proteins including the signal transducer and activator of transcription 5A (STAT5A), signal transducer and activator of transcription 3 (STAT3), SHC-transforming protein 1 (SHC1), tyrosine-protein kinase (JAK2), erythropoietin (EPO), signal transducer and activator of transcription 5B (STAT5B), granulocyte colony-stimulating factor (CSF3), kit ligand (KITLG), thrombopoietin receptor (MPL), and interleukin-3 (IL3) (Fig. 9a). These proteins are mainly involved in cell growth, development, differentiation, and mediating cellular responses to cytokines and other growth factors. Furthermore, the relationship was noticed with the increased level of THPO protein which increases both platelet size and platelet count (Balcik *et al.,* 2013; Kapur *et al.,* 2020).

The Rho guanine nucleotide exchange factor 3 (ARHGEF3) acts as guanine nucleotide exchange factor for RhoA and RhoB GTPases and it showed interaction with five proteins based on STRING analysis including Ring finger protein 145 (RNF145), Rho-related GTP-binding protein RhoB (RHOB), Rsa homology gene family (RHOA), Rho-related GTP-binding protein RhoC (RHOC) and Catenin beta-1 (CTNNB1) (Fig. 9b).

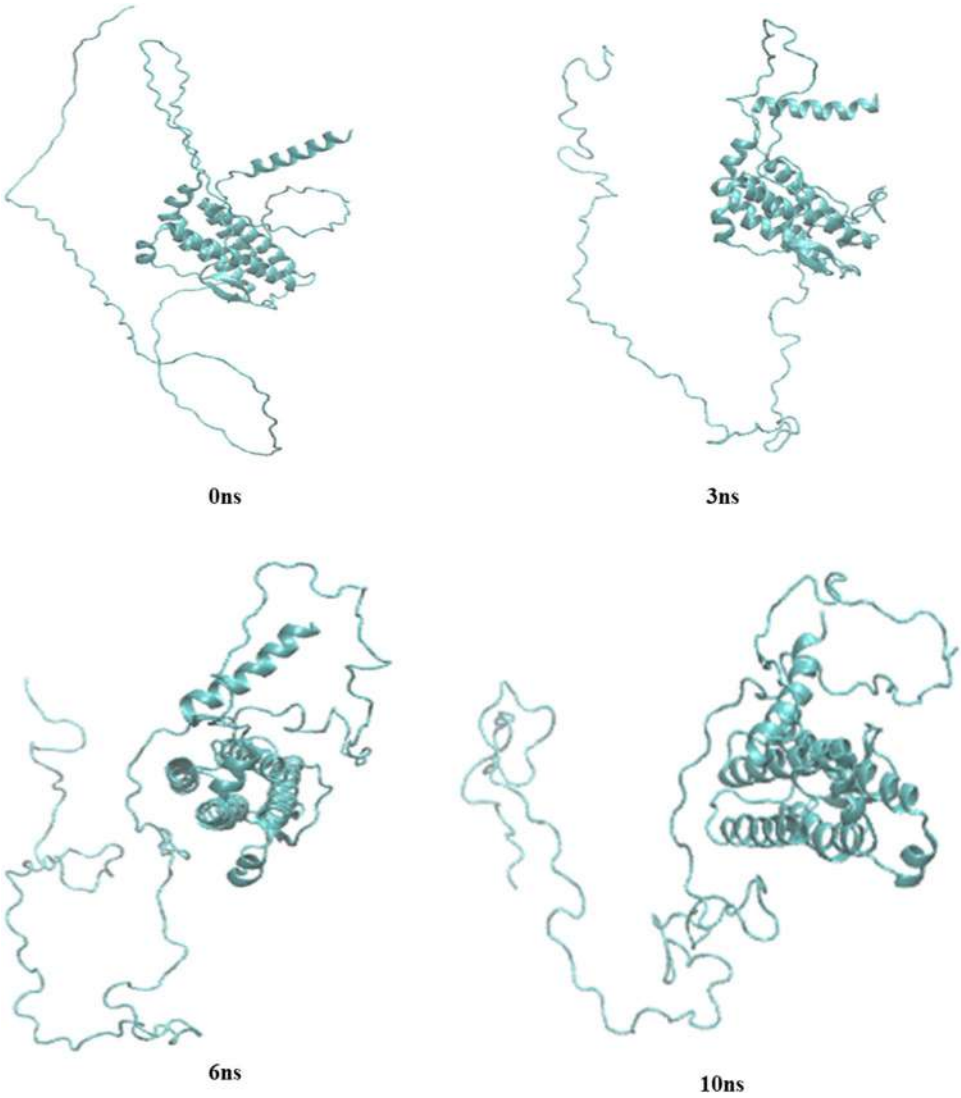

Fig. 7 Ribbon model of mutant protein of Thrombopoietin (Accession number: P40225) during the course of a 100 ns simulation

Discussion

IS affects a large number of individuals. In spite of the availability of various treatment modalities, the disease causes significant morbidity and mortality among patients. Therefore, there is a need to explore new biomarkers that can be used as diagnostic and prognostic markers and also targeted for better therapeutic approaches. Platelets are significant players in the formation of thrombus. Although the patients are prescribed various anti-platelet agents like aspirin and clopidogrel to prevent a stroke in individuals with a strong family history and also to prevent a recurrent stroke [13]. Platelet parameters like MPV and PLT count are considered significant determinants of platelet function [29]. There is evidence that increased platelet size and count reflect increased platelet activity and are useful predictive and prognostic biomarkers for cerebrovascular events. There is no study available on MPV and PLT count affected by a specific genetic alteration in association with the development of IS and its subtypes. Therefore, the current study is a novel approach to evaluate these parameters as potential diagnostic and prognostic markers in IS. In a recent study published by our lab, an elevated MPV was found to be significantly associated with an increased risk of IS, and also higher clot rate, and a higher degree of disability based on mRS. [14]. These results are in accordance with previous studies where an increase in MPV has been associated significantly with an increased risk of IS [11, 29–31]. In the current study, we screened the genetic variants involved in PLT count, MPV, and platelet reactivity in IS patients. Based on the previous reports, a total of 106 variants in 96 genes involved in PLT count, MPV and Platelet reactivity were initially screened using GSA in 17% of patients. Out of these, 62 variants have been reported to affect PLT count; 33 were found to affect MPV and 11 variants reported to affect platelet reactivity (Tables 11, 12, and 13). Most of these genes were found to be

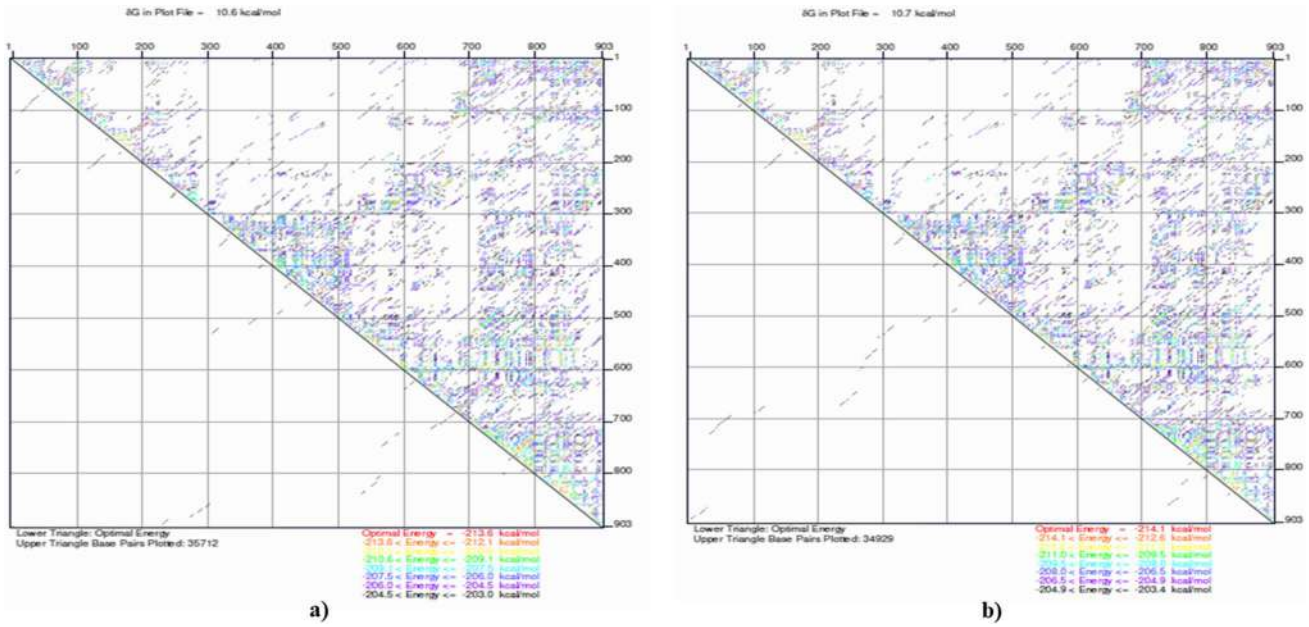

Fig. 8 Simulations of folding, hybridization, and melting pathways of **a** rs1354034 wild type and **b** rs1354034 (mutant) at 37^{O} C

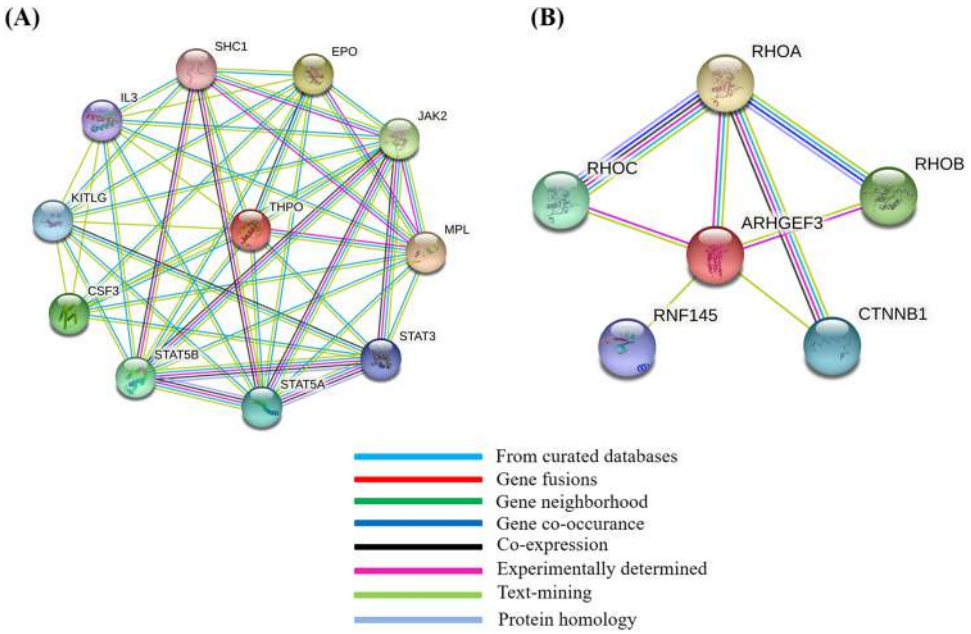

Fig. 9 Protein-protein interaction network analysis of THPO and ARHGEF3 proteins. **A** Represents interaction pattern of THPO. **B** Represents interaction pattern of ARHGEF3. The network of THPO and ARHGEF3 proteins was generated with high confidence level (0.700) using STRING. The selected proteins form large clusters with other functional proteins. The node of the network represent proteins and functional association among proteins was represented by edge. The association of edge was shown by different colored lines i.e., red line-gene fusions, sky blue line-curated database, purple line-experimentally determined, green line-gene neighborhood, black line-co-expression, dark blue line-gene co-occurrence, light blue line-protein homology and light green line-text-mining

either normal homozygous or showed very minor frequency of heterozygosity except for variants of two genes *ARHGEF3* (rs1354034, T>C) and *THPO* (rs6141, C>T). Therefore, these were evaluated further by Sanger Sequencing after amplifying the specific regions of these genes bearing the variation in all the subjects.

Table 11 SNPs associated with PLT count

SNP	Gene	SNP	Gene
rs2336384	MFN2	rs373121156	CDKN2A
rs10914144	DNM3	rs117899880	BRD3
rs1668871	TMCC2	rs505404	PSMD13
rs7550918	LOC148824	rs4246215	FEN1
rs3811444	TRIM58	rs4938642	CBL
rs12603268	GCKR	rs7342306	CD9-VWF
rs625132	EHD3	rs941207	BAZ2A
rs17030845	THADA	rs3184504	SH2B3
rs76160061	SYN2	rs17824620	RPH3A-PTPN11
rs7641175	SATB1	rs7961894	WDR66
rs1354034	ARHGEF3	rs4148441	ABCC4
rs3792366	PDIA5	rs8022206	RAD51L1
rs7694379	HSD17B13	rs8006385	ITPK1
rs17568628	F2R	rs7149242	C14orf70-DLK1
rs700585	MEF2C	rs11628318	RCOR1
rs2070729	IRF1	rs2297067	C14orf73
rs441460	LRRC16	rs3809566	TPM1
rs3819299	HLA-B	rs1719271	ANKDD1A
rs399604	HLA-DOA	rs6065	GP1BA
rs210134	BAK1	rs397969	AKAP10
rs9399137	HBS1L-MYB	rs55997232	TAOK1
rs342275	PIK3CG	rs10512472	SNORD7-AP2B1
rs4731120	WASL	rs708382	FAM171A2-ITGA2B
rs6995402	PLEC1	rs11082304	CABLES1
rs409801	AK3	rs8109288	TMP4
rs13300663	RCL1	rs17356664	EXOC3L2
rs1034566	ARVCF	rs12526480	LRRC16A
rs6141	THPOII	rs6490294	ACAD10
rs9494145	HBS1L-MYB	rs477895	BAD
rs7896518	JMD1C	rs13236689	CD36
rs151361	LRRC16A	rs342293	PIK3CG

Table 12 SNPs associated with MPV

SNP	Gene	SNP	Gene
rs7961894	WDR66	rs10512627	KALRN
rs8109288	TPM4	rs117341321	KIAA0232
rs1354034	ARHGEF3	rs2227831	F2R
rs342293	PIK3CG	rs4521516	MEF2C
rs7075195	JMD1C	rs10076782	RNF145
rs8076739	TAOK1	rs10813766	DOCK8
rs117213068	TMCC2	rs7075195	JMJD1C
rs17655730	PSMD13	rs17655730	NLRP6
rs4812048	CTSZ-TUBB1	rs1558324	CD9-VWF
rs342296	PIK3CG	rs2015599	MTSTD1
rs11653144	TAOK1	rs10876550	COPZ1-NFE2-CBX5
rs17396340	KIF1B	rs2950390	PTGES3
rs10914144	DNM3	rs7317038	GRTP1
rs649729	EHD3	rs944002	C14orf73
rs4305276	ANKMY1	rs3000073	BRF1
rs1354034	ARHGEF3	rs16971217	AP2B1
rs12969657	CD226		

Thrombopoietin (also known as THPO, TPO) is a major cytokine that plays a crucial role in platelet production. This humoral substance controls MK proliferation and differentiation to maintain normal thrombopoiesis [32]. The SNP rs6141 (C>T) situated at the 3'-UTR region of the THPO gene is reported to be involved in the post-transcriptional control of the gene expression, mainly affecting mRNA splicing [33]. Studies have also shown that microdeletions involving this SNP in the THPO gene cause mild congenital thrombocytopenia [34, 35]. Furthermore, two gain-of-function mutations in THPO gene G>C transversion and a G>T transversion have been reported to produce mRNAs with shortened 5′–untranslated regions (UTR) that are more efficiently translated in comparison with transcripts produced by wild-type THPO. These transcripts with a gain of function mutation result in elevated PLT count, which might lead to thrombosis and bleeding [36, 37]. THPO variants with bi-allelic loss-of-function cause multilineage bone marrow failure and severely reduced platelet counts [38–40]. Another study identified a one-base deletion in the 5′-untranslated region of the *THPO* gene. In vitro experiments showed that this mutation increased TPO production and suggested that this region of the *THPO* gene may play a crucial role in regulating *THPO* expression [41]. Based on the results of different GWAS studies, it has been established that rs6141 of the *THPO* gene is a significant determinant of MPV and PLT count [33, 42, 43]. As far as the role of *THPO* in IS is concerned, it has been reported that elevated levels of TPO are associated with increased MPV and PLT counts in these patients [44–46].

In the current study evaluating the association of rs6141 (THPO) with IS, we found a significant association of TT genotype with the disease, which was confirmed by MLR analysis showing an independent association of TT genotype

Table 13 SNPs associated with PLT reactivity

SNP	Gene
rs1613662	GP6
rs3557	FCER1G
rs3737224	PEAR1
rs11264579	PEAR1
rs3729931	RAF1
rs147212241	P2RY12
rs3788337	GNAZ
rs10496541	CD36
rs35091628	MAP2K2
rs12566888	PEAR1
rs7940646	MRVI1

with the disease ($p<0.05$). However, we did not find a significant difference in the distribution of T and C alleles between patients and controls. As far as the association of this variant with IS subtypes is concerned, the T allele showed a significant association with LAA, small artery occlusion, and cardioembolism. We also evaluated the association of variant genotype with MPV, Clot rate, and PLT count. The TT and CT genotypes showed a significant association with elevated MPV, higher clot rate, and reduced PLT count in comparison with the CC genotype-bearing patients. This association was also confirmed by MLR controlling all other confounding factors.

Since the focus of the current study was on platelet parameters, therefore, the expression analysis of *THPO* (rs6141) was carried out using mRNA isolated from platelets of IS patients. Patients bearing TT genotype showed the highest expression, followed by CT and CC genotypes. Aged platelets induce the production of TPO in the hepatocytes. TPO increases the number of circulating platelets once released into the bloodstream. Most of the studies have linked higher TPO levels with increased platelet activity [46]. A study carried out by Balcik *et al*. (2013) reported that patients with IS have higher TPO and MPV levels and concluded that increased TPO levels elevate both PLT count and MPV resulting in a higher thrombotic capacity of platelets [44]. Another study carried out in acute myocardial infarction (AMI) patients and unstable angina pectoris also reported that increased TPO and MPV levels are positively associated with each other in AMI patients [47]. Yang *et al*. (2008) evaluated the role of Severe Acute Respiratory Syndrome (SARS) in affecting the normal functions of hematopoietic stem cells and megakaryocytic cells. They found that increased TPO levels in the plasma of these patients lead to thrombocytosis and hyperactive platelets [48]. Recently a study demonstrated that platelets in COVID-19 patients aggregate faster and showed increased spreading of both fibrinogen and collagen during clot retraction. It was also found that TPO levels were elevated in the serum of SARS-CoV-2 patients [49].

Previous studies have mostly explored the association of *THPO* with PLT count. However, its role in MPV has not been studied much. Since MPV and PLT count is inversely correlated, therefore, it is obvious that studies showing its association with increased PLT count might not have observed its impact on MPV [50].

The proposed mechanism by which TT genotype (rs6141) might lead to the higher expression of *the THPO* gene in bone marrow producing hyperactive platelets, has been depicted in Fig. 10a. TPO binds to the megakaryocytes or platelets and controls their production through a feedback mechanism. During normal hemostasis, TPO concentrations remain normal. Platelets experience mechanical stress that shortens their life span under the conditions like atherosclerosis which indirectly activates platelet biogenesis [44, 51]. *In vivo*, the injection of recombinant adenoviral vectors, or transgenes, resulted in variable thrombocytosis [52–55]. In addition, studies have also demonstrated that patients bearing gain of function mutations in the THPO gene enhance TPO mRNA translation which elevates its expression inducing lineage-selective effects in patients affected with thrombocytosis and polyclonal hematopoiesis. TPO levels were also observed to be higher in the serum [37, 41, 56, 57]. It has been reported that around 1000 to 3000 platelets are produced from a single MK [58]. The mechanism leading

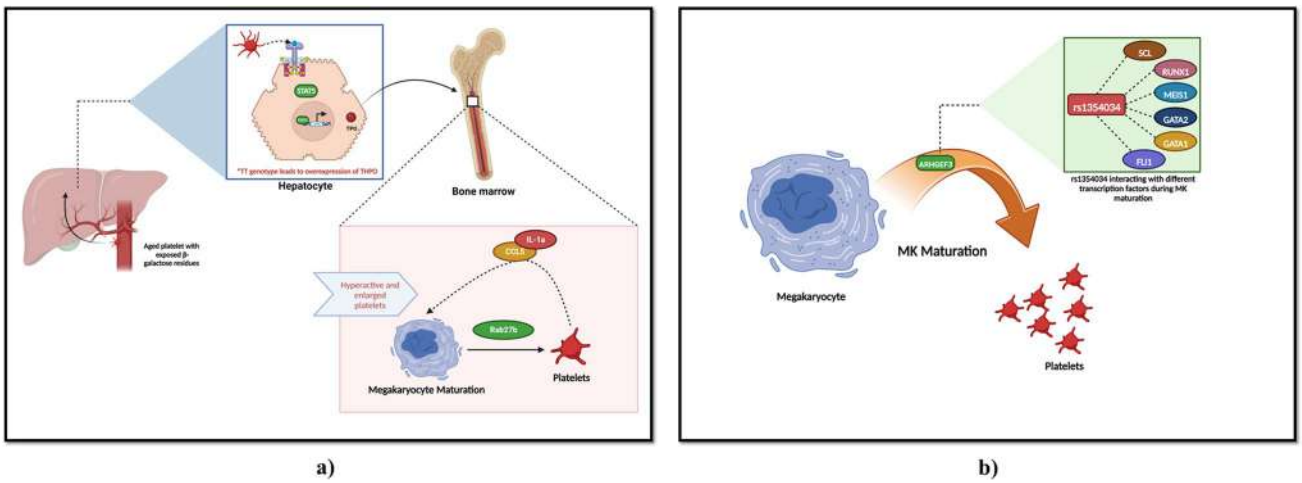

Fig. 10 **a** Desialylated platelets from circulation activates the JAK-STAT cascade in the hepatocytes leading to the activation og THPO gene, which further induces the MKs to produce platelets through c-mpl receptor in the bone marrow. **b** ARHGEF3 upregulated during MK maturation, genomic region of ARHGEF3 where rs1354034 located may influence the binding of certain transcription factors promoting MK maturation to form platelets

to the production of platelets through MKs involves a lot of reorganization of cytoskeletal components like actin and tubulin. THPO is the major watchdog of this process [59, 60]. Although the process of thrombopoiesis is understood well, there are still many unanswered questions as to how some transcription factors like GATA and FOG1 affect the size of the platelets [61].

The impact of rs6141 of *the THPO* gene on protein structure was evaluated by protein dynamics studies using GROMACS. It showed that the wild-type structure of THPO was stable around 2.0 nm (3.5 ns) with respect to its crystal structure. On the other hand, mutated (R38C) structure showed a sudden jump to a higher RMSD value (2.5 nm) at 4.0 ns and maintained a constant higher RMSD value during the course of its simulation. The mutant protein encoded by *THPO* (rs6141) gene also showed reduced compactness in structure in comparison with the wild-type protein. This suggested that a minor deviation between the wild and mutated RMSD values affects the original protein structure.

ARHGEF3, also known as XPLN, is an exchange factor found in platelets, leukemics, and neuronal tissues [62]. It was first identified as RhoGEF (Rho guanine nucleotide exchange factor) for Rho GTPases through an expressed sequence tag database search using the diffused B-cell lymphoma (Dbl) homology (DH) domain query in the BLASTN system [62]. Skeletal muscles and the brain have the highest levels of ARHGEF3 protein expression, followed by the heart, kidneys, platelets, and macrophages [63]. It plays a non-canonical role by inhibiting mTORC2 kinase activity through Akt signaling [63, 64]. It is also involved in various primary cellular functions, including cell adhesion, motility, polarity, growth, cell differentiation, and cytoskeleton rearrangements [63, 65].

GWAS studies have identified newer roles of the *ARHGEF3* gene in modulating bone mineral density (BMD), platelet differentiation, and Hirschsprung disease [22, 66]. Another GWAS carried out to evaluate the association of significant variants with platelet traits reported the association of rs1354034 with MPV in association with other genes, including *WDR66*, TAOK1, and Phospatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit gamma [22]. A meta-analysis including the results of various GWAS studies on 66 867 European individuals also demonstrated that rs1354034, located at 3p14.3, is associated significantly with PLT count and MPV [50]. Zou *et al*. (2017) found that this SNP is present in the regulatory region (non-coding region upstream of the transcription start site) of *the ARHGEF3* gene and proposed that it may influence the binding of certain transcription factors like RUNX1, MEIS1, and GATA2, GATA1, and FLI1 during MK maturation (Fig10b). However, it is not clear if this SNP is directly involved in influencing the binding sites of these transcription factors [66]. The C allele of rs1354034 has been associated with lower ARHGEF3 mRNA expression, higher PLT count and lower MPV in humans [66]. In a very recent report, researchers tried to investigate the genetic overlap of platelet parameters with an endophenotype of Parkinson's disease. They found that various genes, including *ARHGEF3*, are associated with MPV as well as the age of onset and Parkinson's disease susceptibility [67].

In the current study evaluating the association of rs1354034 (ARHGEF3) with IS and its subtypes, we found a significant association of CC genotype and C allele with the disease, which was confirmed by MLR analysis showing an independent association of CC genotype and C allele with the disease. As far as the association of this variant with IS subtypes is concerned, the CC genotype and C allele were found to be associated significantly with LAA, cardioembolism, small artery occlusion, and stroke of undetermined etiology.

We also evaluated the association of variant genotypes of *the ARHGEF3* gene with MPV, Clot rate, and PLT count. The variants CC and TC genotypes showed a significant association with elevated MPV, higher clot rate, and reduced PLT count in comparison with the TT genotype. This association was also confirmed by MLR. A GWAS reported that the same SNP (rs1354034) is associated *in trans* with the expression of vWF, which is an important factor in the blood coagulation pathway in humans [68].

Expression analysis revealed that patients bearing the CC genotype of *the ARHGEF3* gene showed the highest expression, followed by the TC genotype in comparison with the TT genotype in platelets.

Since this variant is an intronic variant, its impact on ARHGEF3 mRNA was evaluated using UNAFold. This analysis revealed that the mRNA encoded by the CC genotype of the *ARHGEF3* gene (rs1354034) leads to qualitatively greater stability with respect to the free energy associated with the secondary structure as compared to the mRNA encoded by the normal genotype.

This SNP present upstream to the *ARHGEF3* gene has been associated significantly with higher expression of *ARHGEF3* during MK maturation both in murine and humans [66, 69, 70]. Based on previous studies, it has been reported that *ARHGEF3* is involved in platelet shape change and function. It has also been demonstrated that ARHGEF3 might be a missing link between ADP-mediated platelet shape change and activation via P2Y1 and P2Y2 receptors [66]. The proposed mechanism by which CC genotype (rs1354034) might lead to the higher expression of *the ARHGEF3* gene activating the MK maturation in bone marrow producing enlarged platelets, has been depicted in Fig. 11. Studies on mice lacking P2Y1 receptors do not show shape change of platelets in response to ADP, suggesting that the ADP signaling is associated with shape change mechanisms in these blood cells [71]. This observation was further aided by another study that reported platelet shape change occurs through Rho signaling and actin reorganization [70, 72].

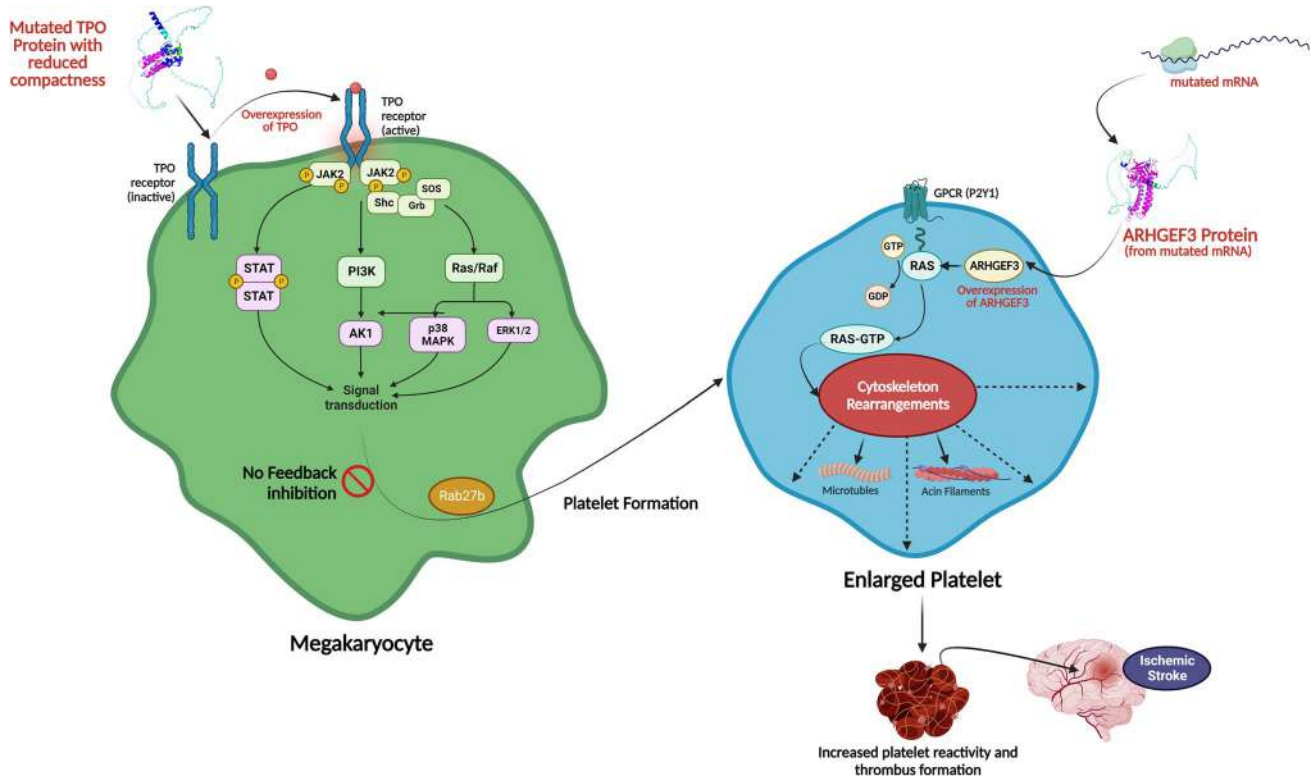

Fig. 11 Mutated TPHO leads to an over expressed TPO cytokine, which activates the JAK-STAT pathway through its receptor (c-MPL) in Megakaryocytes. This pathway leads to the formation of platelets and also controls this formation as a feedback loop. Overexpression of THPO hampers this feedback loop and thereby results in the continuous formation of enlarged platelets. Overexpression of ARHGEF3 activates RAS-GTP through Rho GEF after the activation of platelet receptor (P2Y1). RAS pathway in turn leads to the changes in cytoskeleton of platelets and thereby controls the platelet shape

As mentioned previously, THPO is a glycoprotein produced primarily in the liver that stimulates the formation of megakaryocytes and platelets. THPO protein was found to interact with 10 other proteins mainly involved in platelet activation (c-MPL, IL3), platelet aggregation (IL3), erythropoiesis (EPO, STAT5B), megakaryocyte development (c-MPL, JAK2, STAT3), cytoskeleton organization (CSF3), cell survival and proliferation (KITLG, STAT5A) [73–75]. For the production of platelets, MKs undergo a series of remodeling events that result in thousands of platelets being released from a single cell [76]. All the proteins found to interact with THPO protein are known to regulate platelet formation and functions [77–79]. The interaction of THPO protein with other proteins suggested that there might be potential alternate mechanisms that could affect platelet production, morphology, and function. The THPO, along with other interacting proteins, might be explored as a significant biomarker affecting platelet parameters and functions and thereby a potential therapeutic target.

ARHGEF3 activates two members of the Rho family GTPases, RHOA and RHOB, which are involved in osteoblast maintenance [70]. Various other cellular processes, including cytoskeleton reorganization, are activated and inactivated by Rho-like GTPases, as discussed previously [80, 81]. By catalyzing the release of bound GDP, guanine nucleotide exchange factors (GEFs) accelerate Rho GTPase activity. ARHGEF3 inhibits mTORC2 kinase activity, primarily for Akt, by binding the mTORC2 complex. ARHGEF3 protein has been found to interact with five other proteins involved in cytoskeleton organization (RHOA, RHOC), cell adhesion, migration (RHOA, RHOC, AHOB, and CTNNB1), apoptosis (RHOB), and cholesterol homeostasis (RNF145)**.** Many physiological and pathological functions of platelets are mediated by Rho GTPase proteins [80]. Actin cytoskeleton regulation is one of the main functions of Rho GTPases, although they also participate in several other biochemical pathways [82]. When platelets interact with vWF and collagen via the cell-surface receptors GpIb-IX-V and GPVI, respectively, a dramatic change in shape occurs due to the reorganization of the actin cytoskeleton. When platelet morphology is altered, more surface area is available for interactions with the ECM and other cells [83, 84]. Initial shape changes include discoid loss, sphering, and filopodia extension. The interaction of ARHGEF3 and THPO with other proteins significantly involved in various platelet parameters and functions suggests that the

genotype-phenotype correlation should not be based on one protein but rather than a complete network should be analysed to explore their role as biomarkers or therapeutic targets. The current study is highly relevant as it sheds light on the underlying genetic mechanisms affecting platelet parameters contributing to the development of IS and its subtypes. This study is the first of its kind, focusing on these parameters. Overall the study is a step forward in the field of stroke, providing important insights into platelet parameters, especially MPV, altered by variation in significant genes and thereby highlighting the role of platelet markers in the development of IS. However, the limitation of the study is that the entire genomic landscape has not been explored. NGS approaches should be used to find out the association of variants, including novel ones in other genes employing a larger cohort and replicating the study in populations of different ethnicities. In addition, a more robust functional study needs to be conducted in order to interpret the potential impact of the significant genetic alterations observed in this study and understand the mechanisms involved. Knock-in and knock-out animal models can be used to determine the functional implication of significant variants observed in the current study as well as the ones that will be delineated by future omics studies.

Conclusion

We observed an association of *ARHGEF3* (rs1354034) and *THPO* (rs6141) genes with higher MPV, higher rate of clot formation, and risk of developing IS. Further, we also observed that MPV and PLT count showed an inverse relationship with mutant alleles of both genes. Expression analysis of both *THPO* and *ARHGEF*3 genes revealed a higher expression of variant genotypes in the platelets. In silico analysis carried out for *THPO* (rs6141) gene showed that the mutated protein has reduced compactness in the protein structure in comparison with the wild type, which might be resulting in the higher expression of the *THPO* gene in the platelets. We used UNAFold to predict the changes caused by the variant *ARHGEF3* (rs1354034) at the mRNA level because it is difficult to simulate the mutated *ARHGEF3* (rs1354034) since it is an intronic variant. It showed that the mutant (rs1354034) variant RNA exhibits qualitatively greater stability with respect to the free energy associated with the secondary structure as compared to normal *ARHGEF3* (rs1354034), which might lead to a higher expression in the platelets. Based on the STRING analysis, it was observed that these two significant proteins interact with other proteins which are involved in various pathways such as platelet activation, aggregation, erythropoiesis, megakaryocyte development, cytoskeleton organization, and cell adhesion. Cell migration, vascular development, apoptosis, cell proliferation, and cholesterol homeostasis. The current study is a step forward to establish MPV as a diagnostic or prognostic marker for IS. There is a need to develop specific treatment strategies that can particularly reduce MPV. Furthermore, establishing the specific genotype-phenotype correlation of markers affecting MPV in a particular population might help in devising better or specific treatment strategies.

Acknowledgements Financial assistance from the Council for Scientific and Industrial Research (CSIR), India, and DST-FIST is highly acknowledged.

Author Contribution Anjana Munshi (AM) and Abhilash Ludhiadch (AL) conceived and planned the experiments. Sandeep Singh (SS) helped in mRNA expression analysis. AL carried out the experiments. Sudip Chakraborty (SC), and Mahesh Kulharia (MK) planned and carried out the protein and mRNA simulations. Sulena (S) and Paramdeep Singh (PS) helped in the sample collection and identification of patients. Dixit Sharma (DS) helped in protein-protein interaction studies. AM, AL, and SS contributed to the interpretation of the results. AL and AM took the lead in writing the manuscript. Overall the manuscript was shaped with invaluable and critical feedback from all the authors.

Funding Financial assistance from DST-FIST (SR/FST/LS-I/2017/49) is acknowledged with thanks. Financial support to Mr. Abhilash Ludhiadch (Award No-09/ 1051(0029)/2 019-EMR-1) from the Council for Scientific and Industrial Research (CSIR) India is highly acknowledged.

Data Availability The data generated or analysed during this study has been included in this article. The data relating to MPV, PLT count, and demographic profile of the study participants have already been published.

Declarations

Ethics Approval The study was approved by the Institutional ethics committee of the University (CUPB/CC/RO/18/2316) as well as the study hospital (GGS/IEC/56).

Consent to Participate Consent from each participant was taken prior to the sample collection.

Consent for Publication Not applicable.

Conflict of Interest The authors declare no competing interests.

References

1. Feigin VL et al (2022) World Stroke Organization (WSO): global stroke fact sheet 2022. Int J Stroke 17(1):18–29
2. Amarenco P et al (2009) Classification of stroke subtypes. Cerebrovasc Dis 27(5):493–501
3. Virani SS et al (2021) Heart disease and stroke statistics—2021 update: a report from the American Heart Association. Circulation 143(8):e254–e743
4. Orellana-Urzúa S et al (2020) Pathophysiology of ischemic stroke: role of oxidative stress. Curr Pharm Des 26(34):4246–4260
5. Munshi A et al (2009) Phosphodiesterase 4D (PDE4D) gene variants and the risk of ischemic stroke in a South Indian population. J Neurol Sci 285(1-2):142-145

6. Munshi A et al (2012) Association of LPL gene variant and LDL, HDL, VLDL cholesterol and triglyceride levels with ischemic stroke and its subtypes. J Neurol Sci 318(1-2):51–54
7. Traylor M et al (2021) Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. The Lancet Neurology 20(5):351–361
8. Boehme AK, Esenwa C, Elkind MS (2017) Stroke risk factors, genetics, and prevention. Circ Res 120(3):472–495
9. Ludhiadch A, Vasudeva K, Munshi A (2020) Establishing molecular signatures of stroke focusing on omic approaches: a narrative review. Int J Neurosci 130(12):1250–1266
10. Vasudeva K, Munshi A (2019) Genetics of platelet traits in ischaemic stroke: focus on mean platelet volume and platelet count. Int J Neurosci 129(5):511–522
11. Greisenegger S et al (2004) Is elevated mean platelet volume associated with a worse outcome in patients with acute ischemic cerebrovascular events? Stroke 35(7):1688–1691
12. Miller MM, Henninger N, Słowik A (2020) Mean platelet volume and its genetic variants relate to stroke severity and 1-year mortality. Neurology 95(9):e1153–e1162
13. Del Zoppo GJ (1998) The role of platelets in ischemic stroke. Neurology 51(3 Suppl 3):S9–S14
14. Ludhiadch A, Yadav P, Singh SK, Sulena, Munshi A (2022) Evaluation of mean platelet volume and platelet count in ischemic stroke and its subtypes: focus on degree of disability and thrombus formation. Int J Neurosci 30:1–8
15. Mayda-Domaç F, Mısırlı H, Yılmaz M (2010) Prognostic role of mean platelet volume and platelet count in ischemic and hemorrhagic stroke. J Stroke Cerebrovasc Dis 19(1):66–72
16. Zarmehri B et al (2020) Association of platelet count and mean platelet volume (MPV) index with types of stroke. Caspian J Intern Med 11(4):398
17. Sotero FD et al (2021) Mean platelet volume is a prognostic marker in acute ischemic stroke patients treated with intravenous thrombolysis. J Stroke Cerebrovasc Dis 30(6):105718
18. Cho SY et al (2013) Mean platelet volume/platelet count ratio in hepatocellular carcinoma. Platelets 24(5):375–377
19. Meisinger C et al (2009) A genome-wide association study identifies three loci associated with mean platelet volume. Am J Hum Genet 84(1):66–71
20. Eicher JD et al (2016) Platelet-related variants identified by exome chip meta-analysis in 157,293 individuals. Am J Hum Genet 99(1):40–55
21. Schick UM et al (2016) Genome-wide association study of platelet count identifies ancestry-specific loci in Hispanic/Latino Americans. Am J Hum Genet 98(2):229–242
22. Eicher JD, Lettre G, Johnson AD (2018) The genetics of platelet count and volume in humans. Platelets 29(2):125–130
23. Kunicki TJ, Williams SA, Nugent DJ (2012) Genetic variants that affect platelet function. Curr Opin Hematol 19(5):371–379
24. Adams HP Jr et al (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 24(1):35–41
25. UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49(D1):D480–D489
26. Varadi M et al (2022) AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res 50(D1):D439–D444
27. Nicholas R, Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. Bioinformatics 453:3–31
28. Szklarczyk D et al (2021) The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res 49(D1):D605–D612
29. Du J et al (2016) Association of mean platelet volume and platelet count with the development and prognosis of ischemic and hemorrhagic stroke. Int J Lab Hematol 38(3):233–239
30. Sadeghi F et al (2020) Platelet count and mean volume in acute stroke: a systematic review and meta-analysis. Platelets 31(6):731–739
31. Bath P et al (2004) Association of mean platelet volume with risk of stroke among 3134 individuals with history of cerebrovascular disease. Stroke 35(3):622–626
32. Hitchcock IS, Kaushansky K (2014) Thrombopoietin from beginning to end. Br J Haematol 165(2):259–268
33. Garner C et al (2006) Two candidate genes for low platelet count identified in an Asian Indian kindred by genome-wide linkage analysis: glycoprotein IX and thrombopoietin. Eur J Hum Genet 14(1):101–108
34. Dasouki M et al (2014) Confirmation and further delineation of the 3q26. 33–3q27. 2 microdeletion syndrome. Eur J Med Genet 57(2-3):76–80
35. Mandrile G et al (2013) 3q26. 33–3q27. 2 microdeletion: a new microdeletion syndrome? Eur J Med Genet 56(4):216–221
36. Ghilardi N et al (1999) Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene. Br J Haematol 107(2):310–316
37. Wiestner A et al (1998) An activating splice donor mutation in the thrombopoietin gene causes hereditary thrombocythaemia. Nat Genet 18(1):49–52
38. Dasouki MJ et al (2013) Exome sequencing reveals a thrombopoietin ligand mutation in a Micronesian family with autosomal recessive aplastic anaemia. Blood, J Am Soc Hematol 122(20):3440–3449
39. Pecci A et al (2018) Thrombopoietin mutation in congenital amegakaryocytic thrombocytopenia treatable with romiplostim. EMBO Mol Med 10(1):63–75
40. Seo A et al (2017) Bone marrow failure unresponsive to bone marrow transplant is caused by mutations in thrombopoietin. Blood, J Am Soc Hematol 130(7):875–880
41. Kondo T et al (1998) Familial essential thrombocythemia associated with one-base deletion in the 5′-untranslated region of the thrombopoietin gene. Blood, J Am Soc Hematol 92(4):1091–1096
42. Shameer K et al (2014) A genome-and phenome-wide association study to identify genetic variants influencing platelet count and volume and their pleiotropic effects. Hum Genet 133(1):95–109
43. Kamatani Y et al (2010) Genome-wide association study of hematological and biochemical traits in a Japanese population. Nat Genet 42(3):210–215
44. Balcik ÖS et al (2013) Thrombopoietin and mean platelet volume in patients with ischemic stroke. Clin Appl Thromb Hemost 19(1):92–95
45. Varol E (2013) Increased thrombopoietin and mean platelet volume in patients with ischemic stroke. SAGE Publications Sage CA, Los Angeles, CA, pp. 342–343
46. Sedky HAA et al (2015) Value of thrombopoietin level and platelet size in patients with ischemic stroke. The Egyptian Journal of Haematology 40(1):24
47. Şenaran H et al (2001) Thrombopoietin and mean platelet volume in coronary artery disease. Clin Cardiol 24(5):405–408
48. Yang M et al (2008) Thrombopoietin levels increased in patients with severe acute respiratory syndrome. Thromb Res 122(4):473–477
49. Manne BK et al (2020) Platelet gene expression and function in patients with COVID-19. Blood 136(11):1317–1329
50. Gieger C et al (2011) New gene functions in megakaryopoiesis and platelet formation. Nature 480(7376):201–208
51. Sinzinger H, Virgolini I, Fitscha P (1990) Platelet kinetics in patients with atherosclerosis. Thromb Res 57(4):507–516

52. Kaushansky K et al (1994) Promotion of megakaryocyte progenitor expansion and differentiation by the c-Mpl ligand thrombopoietin. Nature 369(6481):568–571
53. Lok S et al (1994) Cloning and expression of murine thrombopoietin cDNA and stimulation of platelet production in vivo. Nature 369(6481):565–568
54. Yan X-Q et al (1995) Chronic exposure to retroviral vector encoded MGDF (mpl-ligand) induces lineage-specific growth and differentiation of megakaryocytes in mice. Blood 86(11):4025–4033
55. Zhou W et al (1997) Transgenic mice overexpressing human c-mpl ligand exhibit chronic thrombocytosis and display enhanced recovery from 5-fluorouracil or antiplatelet serum treatment. Blood, J Am Soc Hematol 89(5):1551–1559
56. Jorgensen M et al (1998) Familial thrombocytosis is associated with the overproduction of thrombopoietin due to a novel splice donor site mutation. in Blood. WB SAUNDERS, Co Independence Square West Curtis Center, Ste 300, Philadelphia
57. Cazzola M, Skoda RC (2000) Translational pathophysiology: a novel molecular mechanism of human disease. Blood, J Am Soc Hematol 95(11):3280–3288
58. Stenberg P, Levin J (1989) Mechanisms of platelet production. Blood Cells 15(1):23–47
59. Italiano JE Jr et al (1999) Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. J Cell Biol 147(6):1299–1312
60. Levin J (2019) The evolution of mammalian platelets. In Platelets. Academic Press, pp. 1–23
61. Geddis AE, Kaushansky K (2004) Inherited thrombocytopenias: toward a molecular understanding of disorders of platelet production. Curr Opin Pediatr 16(1):15–22
62. Thiesen S et al (2000) Isolation of two novel human RhoGEFs, ARHGEF3 and ARHGEF4, in 3p13-21 and 2q22. Biochem Biophys Res Commun 273(1):364–369
63. Arthur WT et al (2002) XPLN, a guanine nucleotide exchange factor for RhoA and RhoB, but not RhoC. J Biol Chem 277(45):42964–42972
64. Rossman KL, Der CJ, Sondek J (2005) GEF means go: turning on RHO GTPases with guanine nucleotide-exchange factors. Nat Rev Mol Cell Biol 6(2):167–180
65. Snyder JT et al (2002) Structural basis for the selective activation of Rho GTPases by Dbl exchange factors. Nat Struct Biol 9(6):468–475
66. Zou S et al (2017) SNP in human ARHGEF3 promoter is associated with DNase hypersensitivity, transcript level and platelet function, and Arhgef3 KO mice have increased mean platelet volume. PloS One 12(5):e0178095
67. Tirozzi A et al (2021) Genomic overlap between platelet parameters variability and age at onset of Parkinson disease. Appl Sci 11(15):6927
68. Zhang X et al (2014) Genetic associations with expression for genes implicated in GWAS studies for atherosclerotic cardiovascular disease and blood phenotypes. Hum Mol Genet 23(3):782–795
69. Simon LM et al (2014) Human platelet microRNA-mRNA networks associated with age and gender revealed by integrated platelet omics. Blood, J Am Soc Hematol 123(16):e37–e45
70. Khaliq SA, Umair Z, Yoon M-S (2021) Role of ARHGEF3 as a GEF and mTORC2 regulator. Front Cell Dev Biol 9:806258–806258
71. Kim YB, Jin J, Dangelma C (1999) The P2Y1 receptor is essential for ADP-induced shape change and aggregation in mouse platelets. Platelets 10(6):399–406
72. Klages B et al (1999) Activation of G12/G13 results in shape change and Rho/Rho-kinase–mediated myosin light chain phosphorylation in mouse platelets. J Cell Biol 144(4):745–754
73. Drachman JG, Griffin JD, Kaushansky K (1995) The c-Mpl ligand (Thrombopoietin) stimulates tyrosine phosphorylation of Jak2, Shc, and c-Mpl (∗). J Biol Chem 270(10):4979–4982
74. Hoffmann O et al (2011) Thrombopoietin contributes to neuronal damage in experimental bacterial meningitis. Infect Immun 79(2):928–936
75. Baker JE et al (2008) Human thrombopoietin reduces myocardial infarct size, apoptosis, and stunning following ischaemia/reperfusion in rats. Cardiovasc Res 77(1):44–53
76. Saintillan D (2020) Physical mechanisms of platelet formation. Proc Natl Acad Sci 117(36):21841–21843
77. Mbiandjeu S, Balduini A, Malara A (2021) Megakaryocyte cytoskeletal proteins in platelet biogenesis and diseases. Thromb Haemost 122(05):666–678
78. Thon JN, Italiano JE (2010, July) Platelet formation. In Seminars in hematology (Vol. 47, No. 3, pp. 220–226). WB Saunders
79. Machlus KR, Italiano JE Jr (2013) The incredible journey: From megakaryocyte development to platelet formation. J Cell Biol 201(6):785–796
80. Goggs R et al (2015) Platelet Rho GTPases–a focus on novel players, roles and relationships. Biochem J 466(3):431–442
81. Ulu A, Frost JA (2016) Regulation of RhoA activation and cytoskeletal organization by acetylation. Small GTPases 7(2):76–81
82. Hotta K et al (1996) Interaction of the Rho family small G proteins with kinectin, an anchoring protein of kinesin motor. Biochem Biophys Res Commun 225(1):69–74
83. Hensler M et al (1992) Platelet morphologic changes and fibrinogen receptor localization. Initial responses in ADP-activated human platelets. Am J Pathol 141(3):707
84. Maxwell MJ et al (2006) Shear induces a unique series of morphological changes in translocating platelets: effects of morphology on translocation dynamics. Arterioscler Thromb Vasc Biol 26(3):663–669

Medical Oncology (2019) 36:70
https://doi.org/10.1007/s12032-019-1290-0

ORIGINAL PAPER

A study of mechanistic mapping of novel SNPs to male breast cancer

Raman Preet Kaur[1] · Vicky Kumar[2] · Gowhar Shafi[3] · Rajesh Vashistha[4] · Mahesh Kulharia[2] · Anjana Munshi[1]

Received: 11 April 2019 / Accepted: 1 June 2019 / Published online: 15 June 2019

Abstract

Alterations in *BRCA2*, *PALB2*, *CHEK2*, and *p53* genes have been identified for their association with male breast cancer in various studies. The incidence of male breast cancer in India is consistent with its global rate. The present study was carried out with an aim to evaluate the genetic alterations in male breast cancer patients from Malwa region of Punjab, India. Four male breast cancer patients belonging to different families were recruited from Guru Gobind Singh Medical College and Hospital, Faridkot, India. A total of 51 genes reported with implications in the pathogenesis of breast cancer were screened using next generation sequencing. Germline variations were found in *BRCA1*, *BRCA2*, *PMS2*, *p53*, and *PALB2* genes, previously reported to be associated with MBC as well as FBC. In addition to these, 13 novel missense alterations were detected in eight genes including *STK11*, *FZR1*, *PALB2*, *BRCA2*, *NF2*, *BAP1*, *BARD1*, and *CHEK2*. Impact of these missense alterations on structure and function of protein was also analyzed through molecular dynamics simulation. Structural analysis of these single nucleotide polymorphisms (SNPs) revealed significant impact on the encoded protein functioning.

Keywords SNPs · Male breast cancer · Molecular dynamics simulation · Germline variations · Novel missense alterations

Introduction

Male breast cancer (MBC) is a rare/uncommon malignancy representing 1% of all types of cancer and it is increasing at a pace of 1.1% every year [1–3]. The incidence rate of the MBC in Indian population is consistent with the global rate, i.e., one case per 100,000 men/year [4, 5]. The pathogenesis of MBC has been attributed to the genetic, hormonal, and environmental risk factors as in the case of female breast cancer (FBC). Among these, the family history of cancer is considered as the main predisposing factor because the risk of breast cancer increases twofold in case of family history [6]. As far as epidemiological prospective is concerned, MBC resembles post-menopausal FBC. However, it is different from FBC in clinical and pathological characteristics. In addition, the prognosis of MBC is poor in comparison with FBC as the symptoms appear in advanced stage in case of the former [5]. Around 10% of MBC cases are hereditary in nature and are due to germline mutations in tumor suppressor genes including *BRCA2, p53, ATM*, *CHEK2,* and *PALB2*. The mutation 1100delC of *CHEK2* has been linked with tenfold increased risk for MBC, whereas *PALB2* is reported to increase MBC risk by eightfold [7–11]. Genetic profiling of MBC patients can aid in the identification of genetic markers associated with MBC and could potentially aid in the identification of new therapeutic strategies [12].

The present study was taken up with an aim evaluate the incidence, clinical, and genetic profile of MBC patients from the Malwa region of Punjab, India, where cancer is widely feared on account of its steeply rising graph. Genetic alterations reported from this study were also examined using molecular dynamics-based approach to determine their impact on structure and function of protein.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12032-019-1290-0) contains supplementary material, which is available to authorized users.

✉ Mahesh Kulharia
kulharia@gmail.com

✉ Anjana Munshi
anjanadurani@yahoo.co.in

[1] Department of Human Genetics and Molecular Medicine, Central University of Punjab, Bathinda 151001, India

[2] Department of Computational Sciences, Central University of Punjab, Bathinda 151001, India

[3] Department of Clinical Genomics and Bioinformatics, Positive Bioscience, Mumbai, India

[4] Max Super-specialty Hospital, Bathinda 151001, India

Materials and methods

Four MBC patients evaluated by an oncologist at Guru Gobind Singh Medical College and Hospital (GGSMCH), Faridkot, India were included in the study. The proposed work was carried out with the approval from the Ethics committees of GGSMCH and Central University of Punjab. Information on demographic profile and breast cancer-associated risk factors was collected by using a structured questionnaire. 5 ml of venous blood was collected from patients in EDTA-coated vacutainers with written informed consent of the patients. DNA was extracted from whole blood by using phenol–chloroform method. After the qualitative and quantitative analysis of DNA, the samples were subjected to next generation sequencing (NGS).

NGS assay

Library preparation for NGS was accomplished using the HaloPlex PCR target enrichment system (Agilent Technologies Inc.) [13, 14]. Using Sure Design studio (Agilent Technologies Inc.), probes were generated to cover the exons of all 51 genes frequently implicated in the pathogenesis of breast cancer. Using this design, libraries from four samples were generated. Design was made for Illumina 150-bp paired-end sequencing.

Library preparation and sequencing

Amplicon libraries were prepared from genomic DNA of all the patients using the HaloPlex PCR target enrichment system according to the manufacturer's recommendations. In brief, 50 ng of DNA was used for restriction reactions and hybridization was performed for 3 h at 54 °C. All the DNA samples were individually indexed. Libraries were quantified using the Qubit fluorometer and DNA concentration was calculated using the formula [1 ng/mL = 3 nmol/L/(library average size in bp/500)], where average fragment length was obtained from bioanalyzer data. Amplicon libraries were diluted to 2 nmol/L and indexed samples were pooled at a final concentration of 6 pmol/L. Sequencing was performed using Next Seq reagent kit 300 cycles on the NextSeq instrument (Illumina Inc.).

Bioinformatics analysis

The raw data quality was checked using NextSeq Sequencing Analysis Viewer software (Illumina Inc.). The sequences obtained were quality controlled using FastQC software (Babraham Bioinformatics, Cambridge, UK). SAMtools flagstat was used to obtain metrics of the analyzed reads. Raw FASTQ files were first evaluated using quality control checks from FastQC. NGSQC toolkit was then employed for read trimming and filtering. After removing low-quality bases, adapters, and other technical sequences, each library was aligned to the human reference genome (GRCh37) using BWA-mem generating sorted BAM files with SAMtools. Reads from the same libraries were then merged. The regions with < 30 reads of base quality ≥ 30 and mapping quality ≥ 20 were considered as target-sequence gaps. For the identification of germline point mutations, filtering parameters were total read depth of ≥ 6, mutated allele count of ≥ 3, variant frequency of ≥ 0.1, base quality of ≥ 30, mapping quality of ≥ 20. SAMtools was used for conversion of .sam files to .bam files and for generation of .mpileup files. From the .mpileup file, coverage was calculated using a combination of awk and custom scripts in R. Genome Analysis Toolkit (Broad Institute, Cambridge, MA) was used for raw variant calling (UnifiedGenotyper), realignment (RealignerTargetCreator and IndelRealigner), recalibration (BaseRecalibrator and PrintReads), and variant calling (Haplotype caller, Unified Genotyper, Samtools, SNVer and Mutect). Variants were looked for in exons and 20 bp flanking each exon because we wanted to identify variants in splice sites also. Even if flanking regions had not been targeted, this was possible because HaloPlex amplification, as a rule, includes a substantially larger region than the targeted exons, resulting in satisfactory coverage also in the flanking regions. Visual inspection of the alignment was also performed as a final step for all reported variants using Integrative Genomics Viewer. The process of variant calling includes an extra evaluation layer for variants present in provided data sources (COSMIC for somatic and HGMD for germline variants), regardless of the filtering thresholds (including the 0.025 minimum variant frequency otherwise required). Variants called according to these criteria are further evaluated by the filtering thresholds, but instead of being discarded when quality measures are not met, the variant is registered in a different output, facilitating further inspection. The objective of this added flexibility is to identify relevant mutations that would otherwise be missed when cutting thresholds are not met. Variants were annotated using several databases containing functional (Ensembl, CCDS, RefSeq), populational (dbSNP, 1000 Genomes, ESP, ExAC), and disease-related (COSMIC, ICGC, HGMD professional, Clinvar) information, as well as 11 scores from algorithms for prediction of the impact caused by nonsynonymous variants on the structure and function of the protein (SIFT, PolyPhen2, PROVEAN, Mutation Assessor, Mutation Taster, LRT, MetaLR, MetaSVM, FATHMM, and FATHMM-MKL), and to evaluate evolutionary conservation of a particular variant, two tools were used, the Genomic Evolutionary Rate Profiling (GERP) and the PhastCons. The GERP score of > 2.0 and the PhastCons score of > 0.3 indicate a good level of

conservation of the variants and these scores were considered in the screening of variant.

Structural analysis of the novel mutations

Generation of protein models

In order to determine the effect of mutations at the molecular level, SNPs were considered for their structural influence. Since no crystal structure was available for encoded form of genes (i.e., proteins) under study, software tool YASARA (version: 17.11.22, License No.: 476925318) which creates homology models was used [15]. The wild-type protein sequence (i.e., unmutated form) was modeled by YASARA to get the protein model. For mutated proteins, the amino acid corresponding to SNP was swapped with its mutated version in the structural model. Subsequently, the obtained protein models (both unmutated and mutated) were refined using molecular dynamics simulation.

Quality check of the protein models

In order to ascertain the quality of protein models, various checks and validations were performed such as Q mean score [16]. The orientation of residues in protein models was examined using RAMPAGE [17]. All the protein models were reported to be well modeled and did not violate Ramachandran plot.

Energy minimization and molecular dynamics simulation of protein models

GROMACS was used to derive the energy-minimized structure of protein models and their molecular dynamics simulation [18]. As all the proteins under study were water soluble, proteins were simulated in the presence of water. The charge of system was neutralized by adding sufficient number of sodium and chlorine ions. Steepest-descent algorithm, with 50,000 steps and energy step-size of 0.01 kJ/mol/nm, was used for energy minimization. The energy-minimized system was equilibrated with respect to pressure and volume using leap-frog integrator algorithm for 100 ps with a step-size of 2 fs. At last, the equilibrated system was considered for molecular dynamics (MD) simulation using the leap-frog integrator algorithm for 2 ns with a step-size of 2 fs.

Cluster analysis of simulated protein models

The cluster analysis of MD-simulated protein model generated a number of conformational states. To select one conformational state as representative of the protein model, the conformational state with least amount of RMSD with respect to other conformational states was selected and analyzed further.

Neighborhood and structural analysis of wild-type and mutated protein models

The neighborhood (within radius of 5 Å) and structural analysis of wild type and their mutated version of protein models were examined in order to see the changes induced by the mutation at the atomic level.

Results

The age of four patients recruited in the present study was 58, 63, 67, and 72 years at the time of disease diagnosis. The incidence of MBC in this region was observed to be 1.3%, since out of 300 breast cancer patients, only four were found to be MBC cases, the rest were FBC cases. All the patients were positive for ER/PR and one patient was HER2 positive as well. Family history of breast cancer was observed in one patient (Fig. 1b) and rest of the cases were sporadic (Fig. 1a, c, d). Germline missense alterations were detected in *BRCA1* (rs1799966, rs1799967 & rs16942), *BRCA2* (rs45469092), *PMS2* (rs18005321 & rs18005323), *p53* (rs1042522), and *PALB2* (rs152451 & rs45551636) genes previously reported to be associated with MBC and FBC [8, 9]. In addition, the genetic profiling of these four patients has unraveled 13 novel missense alterations in eight genes including *STK11, FZR1, PALB2, BRCA2, NF2, BAP1, BARD1*, and *CHEK2*. These novel alterations were detected in all the four patients.

Two novel mutations Val63Met and His168Asp were detected in STK11 protein. The former was observed in ATP binding domain of the protein, whereas the latter was located in the MO25-binding interface. Structural analysis of H168D revealed changes in interaction network of the amino acids. In the wild-type STK11 protein, Lys175 was observed maintaining hydrogen bonding with Gln214, Thr212, and

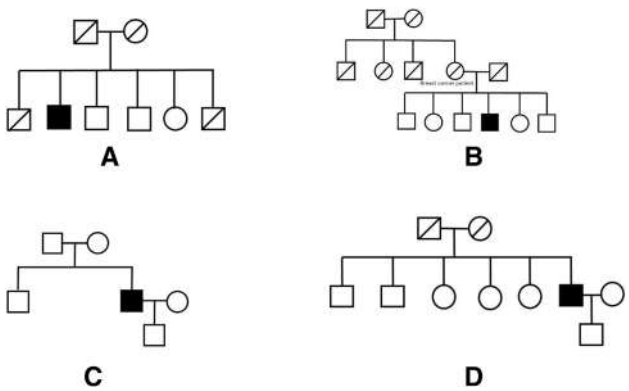

Fig. 1 Pedigree MBC patients

Glu199, whereas the mutation His168Asp in STK11 led to the loss of these interactions. On the other hand, Lys175 was not observed to interact with any residues in the STK11 protein-bearing Val63Met. The hydrogen bonding observed in Cys73 and Thr71 in wild-type STK11 also disappeared in both His168Asp- and Val63Met-mutated STK11 (Supplementary sheet: S1, S2 & S3). These mutations resulted in the local structural configuration modification in the vicinity, thereby probably reducing/altering the ATP and MO25-binding ability.

A novel mutation Ser163Ile was detected in the FHA domain of CHEK2 protein. The structural analysis of Ser163Ile showed the change in interactions among various amino acids through disruption in hydrogen bonding and hydrophobic interaction. Arg191 was seen interacting with Asn155 in the wild type but an increase in hydrogen bonding in mutated form allowed Arg191 to interact with an additional residue, i.e., Asp154. Additionally, in mutated CHEK2, Thr164 starts interacting with other residues through hydrogen bonding which could not be observed in wild type. Hydrophobic interactions of Cys164 were also observed in mutated CHEK2. The amino acids displaying hydrophobic interaction were different in both mutated and wild-type CHEK2. The mutation had also changed the orientation of the oxygen atoms of Agr160, Asp161, and Ile163 (Supplementary sheet: S4 & S5), leading to variations in the interacting pattern of the protein.

The mutation Thr2968Asn in the tumor suppressor gene *BRCA2* was detected in its DNA binding domain. The neighborhood analysis revealed that Thr2968Asn posed a significant disturbance in the interacting networks of the residues Trp2830 and Arg2991. Thr2968 did not show any hydrogen bonding with any other residues in the wild type (Supplementary sheet: S6 & S7). However, mutated BRCA2 interacts with Gln2829 and Arg2991 through hydrogen bonding. Strong hydrophobic interactions of Trp2990, absent in wild type, were also observed in the mutated BRCA2.

Two novel mutations Gly158Val and Gly158Arg were observed in BAP1 protein, 10 amino acids prior to the catalytic site. The structural analysis of the Gly158Arg-mutated BAP1 protein revealed an increase in the hydrogen bonding between Glu7 and Arg163. On the other hand, Gly158Val mutation resulted in the loss of original hydrogen bond between Glu7 and Arg163. Additionally, the hydrophobic interactions increased in both of these mutated forms of BAP1 in comparison with the wild type (Supplementary sheet: S8, S9 & S10).

Two mutations Leu299Ile and Leu299Arg in *NF2* gene were detected on its actin binding site. Significant changes were observed in the interaction pattern of amino acids in wild-type and mutated *NF2* gene as the residues Thr71, Glu7, Asp237 in the wild-type NF2 were seen interacting with other residues through hydrogen bonding but the same interactions were missing in NF2 with L299I and L299R mutations. Trp74 interacted with Leu297 via hydrogen bonding in the wild-type NF2 protein. However, Leu299Ile alteration resulted in the disappearance of this interaction, whereas Leu299Arg mutation changed the interacting partner of Trp74 from Leu297 to Glu298. In addition, the amino acids exhibiting the hydrophobic interactions were different in normal and mutated NF2 protein (Supplementary sheet: S11, S12 & S13).

The WD40 domains of *PALB2* and *FZR1* genes were observed to be mutated on account of Lys899Asn and Lys899Ile, and Leu284Met and Leu284Arg, changes, respectively. Alteration Lys670Asn in *BARD1* gene was present on the C-terminal domain (Table 1).

Table 1 List of domain containing novel alterations

Gene	HGVSp	Domain
PALB2	p.Lys899Asn	Partner and localizer of BRCA2 WD40 domain
	p.Lys899Ile	Partner and localizer of BRCA2 WD40 domain
STK11	p.Val63Met	ATP binding; catalytic domain
	p.His168Asp	MO25 interface on conserved domain; catalytic domain
FZR1	p.Leu284Met	WD40 domain
	p.Leu284Arg	WD40 domain
NF2	p.Leu299Ile	FERM domain; actin binding site on conserved domain
	p.Leu299Arg	FERM domain; actin binding site on conserved domain
BAP1	p.Gly158Val	Cysteine peptidase C12 containing ubiquitin carboxyl-terminal hydrolase
	p.Gly158Arg	Cysteine peptidase C12 containing ubiquitin carboxyl-terminal hydrolase
BARD1	p.Lys670Asn	C-terminal domain
CHEK2	p.Ser163Ile	Fork-head associated domain
BRCA2	p.Thr2968Asn	DNA binding domain

Discussion

In the present study, demographic and molecular analyses of four cases of MBC from the Malwa region of Punjab (India) were carried out. To the best of our knowledge, the present study is the first one to analyze the genetic profiling of MBC cases from North India. Further, the study is the first of its kind from Malwa region of Punjab, where breast cancer is the second most common cancer.

MBC incidence in this region was observed to be 1.3%, which was in accordance with the previous reports [5]. A previous study carried out by Chikaraddi et al. reported the frequency of MBC to be 0.4% in Indian populations, whereas another study carried out in the North Indian population reported the MBC incidence to be 0.5% [19]. On the contrary, the incidence of MBC was reported to be 4.1% in the Kashmiri population which is quite high in comparison with the reports from other regions of India as well as other countries [20, 21].

All the four cases of MBC reported in this study were found to be affected with infiltrating ductal carcinoma. According to the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) database results, more than 90% of the MBC have been reported to be ER positive receptor subtype [22]. All the four MBC patients from the proposed study were ER/PR positive.

Genetic profiling of the patients was carried out by massively parallel sequencing (MPS) of the coding regions of 51 genes which have been reported to be most frequently altered in breast cancer including tumor suppressor, proto-oncogenes, and DNA repair genes. These 51 genes were screened for SNPs and mono/biallelic alterations in *BRCA1, BRCA2, PALB2, PMS2,* and *p53* genes were observed. In addition, 13 novel missense alterations were also detected in 8 genes including *STK11, FZR1, PALB2, BRCA2*, *NF2, BAP1, BARD1,* and *CHEK2*. For their in-depth analysis, these novel alterations were considered for their molecular dynamics simulation-based structural analysis.

According to the previous reports, *BRCA1, BRCA2, ATM, PMS2, p53*, and *PALB2* gene alterations are reported to be associated with FBC as well as MBC. The association of *ATM* gene with MBC has been reported by three previous studies in Greek, Caucasians, and Ashkenazi populations [8, 9]. The alterations found in *ATM* gene are also in consensus with previous studies. A common mutation 1100delC in *CHEK2* gene has been reported to be associated with FBC as well as MBC [9]. However, in the current study, none of the MBC patients was observed to carry this mutation or other mutations reported in the *CHEK2* gene.

CHEK2 gene is responsible for regulation of cell cycle and phosphorylation of genes in response to DNA damage. CHEK2 inhibits the phosphatase activity of CDC25A, CDC25B, and CDC25C. Inhibition of phosphatase activity of CDCs increases the concentration of tyrosine phosphorylated form of CDK-cyclin complexes and thus halts cell cycle progression. It also stimulates transcription of various genes associated with DNA repair including BRCA2 by phosphorylating and activating transcription factor FoxM1 [23]. The CHEK2 protein binds to phosphothreonine, phosphoserine, and sometimes phosphotyrosine with the help of residues Arg160, Leu172, Ser183, Lys184, Gly208, Asn209, and Gly210 of its polypeptide binding site, to carry out various functions. A novel alteration Ser163Ile was observed in the FHA domain of *CHEK2* gene. This domain is responsible for the establishment and maintenance of cell cycle checkpoints [24]. The neighborhood analysis of Ser163Ile revealed the change in interacting network of one of the binding site residues Lys184 which in turn perturbed the interacting network of Arg160, another binding residue of polypeptide binding site of the FHA domain. Therefore, it can be concluded that Ser160Ile has led to disruption in CHEK2 functioning by disturbing the binding residues of FHA domain and interacting pattern among residues which may lead to the loss of function of the other genes or encoded proteins associated with FHA domain.

NF2 gene regulates the hippo signaling pathway, a pathway that plays a pivotal role in tumor suppression by inhibiting cell proliferation and enhances apoptosis. It suppresses cell proliferation and tumorigenesis by inhibiting the ubiquitin–protein ligase complex, CUL4A-RBX1-DDB1-VprBP/DCAF1 E3 [25]. In addition, it also inhibits the activity of PI3kinase as it binds with AGAP2, a factor that stimulates activity of kinase [26]. Two novel alterations Leu299Arg and Leu299Ile observed in MBC patients involved in the current study were located in the actin binding site of FERM domain of *NF2* gene. In addition to actin binding site, FERM domain of *NF2* gene also hosts another three important binding sites: phosphoinositide binding site, peptide binding site, and a homodimer interface. The mutation Leu299Arg was found to reduce the size of interacting network of binding residues of peptide and actin binding sites, and increased the size of interacting network of the binding residues of homodimeric interface. The other mutation Leu299Ile showed reduction in interacting network of binding residues of peptide and actin binding sites of FERM domain. The variability seen in the interaction network of amino acids, because of these mutations, may result in change in structural complexity and thereby affect the functional properties of NF2 protein.

BRCA2, a tumor suppressor gene, is reported to be frequently altered in familial and sporadic breast cancer. Alterations in the conserved domain might affect the DNA repair function of the protein [27]. The novel mutation Thr2968Asn lies on the most conserved region second

OB fold (OB2) of the 800-amino acid C-terminal ssDNA binding domain (DBD) of the BRCA2. BRCA2 binds to the ssDNA with the help of Ile2828, Gln2829, Trp2830, Met2831, Val2966, Thr2967, Thr2968, Val2969, Trp2970, Ile2989, Trp2990, Arg2991, and Pro2992 residues. Hence, this alteration might affect local orientation of the neighboring residues and thereby disrupt or disturb the functions of BRCA2 encoded protein.

Two missense novel alterations Val63Met and His168Asp were detected in *STK11*, a tumor suppressor gene which activates both AMP-activated protein kinase (AMPK) family proteins and non-AMPK family proteins by phosphorylation [28]. These mutations were found to be located in ATP binding site, MO25 interface, STRAD interface, and active site of the STK11. The residue Val63Met is a binding residue on both ATP binding site and active site of the STK11. It disturbed the interaction network of other binding residues, i.e., Gly58, Lys78, Met129, and Leu183 of ATP binding site and active site domain. The surface analysis of STK11 with respect to Val63Met alteration showed that methionine created an occlusion on the binding site leading to change in geometric complementarity of the site. The other missense alteration His168Asp is located on the MO25 interface of STK11 and results in alteration of interacting network with other binding residues like Glu165, Pro203, Phe234, and Arg301 of MO25 interface. Both the variations changed the interaction network of amino acids resulting in structural and functional variabilities that might abolish the tumor suppressor function of the gene.

BAP1, a BRCA1-associated protein, acts as deubiquitinating enzyme and regulates cellular proliferation, chromatin dynamics, and response to DNA damage [29]. Two missense mutations, Gly158Arg and Gly158Val, were found to be located on ssDNA binding domain of this protein which has been reported to be a cancer-causing mutation site [29]. Gly158Arg and Gly158Val were observed to be exactly ten amino acids upstream to the catalytic domain. Both of these variations altered the neighboring amino acids of Gly158 by 4-6 Å which resulted in decrease in volume of the active site and increased the compactness of protein. So, these changes in structural and interacting network of the protein are likely to cause the loss of tumor suppressor function of BAP1 gene.

PALB2, a partner and localizer of BRCA2 gene, is a tumor suppressor that plays a crucial role in homologous recombination. Heterozygous carriers of germline mutations in *BRCA2*, *PALB2*, and *RAD51C* genes have an increased lifetime risk of developing breast, ovarian, and other cancers [30]. PALB2 interacts with BRCA2 and RAD51D via WD40 domain and recruits the complex at the site of DNA damage. In addition, PALB2 also plays a crucial role in stability and localization of BRCA2. The novel mutations Lys899Ile and Lys899Asn are located on WD40 domain of PALB2 protein. Some malfunctioning may be associated with these alterations. However, based on structural and neighborhood analyses of these mutations, we could not conclude it precisely. Similar difficulties were faced in analyzing two novel variations Leu284Met and Leu284Arg in FZR1 protein and Lys670Asn in BARD1 protein.

In conclusion, we observed the incidence of MBC to be 1.3% in Malwa region of Punjab. Although alterations in various genes including *BRCA1*, *BRCA2*, *PALB2*, *PMS2*, and *p53* genes were observed in MBC patients employed in the current study, we found 13 novel alterations in *STK11*, *FZR1*, *PALB2*, *BRCA2*, *NF2*, *BAP1*, *BARD1*, and *CHEK2* genes. Surprisingly, these alterations were observed in all the patients from this region in MBC cases, which is a rare disease in comparison with FBC. These alterations if confirmed in more MBC patients from this region might guide us in terms of management of the disease in this region. No doubt MBC, like other multifactorial rare diseases, does suffer from the absence of comprehensive studies, thereby restricting the translation of these research finding into a personalized management of the disease. However, studies investigating the appropriate screening and risk management tools for MBC patients might lead to appropriate treatment strategies for this rare disease in future.

Acknowledgments The authors are thankful to Central University of Punjab, Bathinda, India for providing academic, administrative, and financial support to carry out this work. University Grants Commission (UGC), New Delhi, India, is also highly acknowledged for providing financial assistance to VK in the form of UGC-NET JRF award.

Compliance with ethical standards

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

Ethical approval The blood sample collection of breast cancer patients for the present study was approved by Institutional Ethics Committee (IES) of CUPB.

References

1. Staruch RM, Rouhani MJ, Ellabban M. The surgical management of male breast cancer: time for an easy access national reporting database? Ann Med Surg. 2016;1(9):41–9.
2. Maguire SL, Tomczyk K, Perrakis E, Saunders E, Leongamornlert D, Kote-Jarai Z, Eeles R, Garcia-Closas M, Pharoah P, Easton D, Lord CJ. Rare variants in DNA damage repair genes are associated with male breast cancer predisposition. Cancer Res. 2018. https://doi.org/10.1158/1538-7445.AM2018-1225.
3. Yoney A, Kucuk A, Unsal M. Male breast cancer: a retrospective analysis. Cancer/Radiothérapie. 2009;13(2):103–7.
4. Hodgson NC, Button JH, Franceschi D, Moffat FL, Livingstone AS. Male breast cancer: is the incidence increasing? Ann Surg Oncol. 2004;11(8):751–5.
5. Sundriyal D, Kotwal S, Dawar R, Parthasarathy KM. Male breast cancer in India: series from a cancer research centre. Indian J Surg Oncol. 2015;6(4):384–6.

6. Ojara EA. Carcinoma of the male breast in Mulago Hospital, Kampala. East Afr Med J. 1978;55(10):489–91.
7. Rizzolo P, Silvestri V, Tommasi S, Pinto R, Danza K, Falchetti M, Gulino M, Frati P, Ottini L. Male breast cancer: genetics, epigenetics, and ethical aspects. Ann Oncol. 2013;24(suppl_8):viii75–82.
8. Pritzlaff M, Summerour P, McFarland R, Li S, Reineke P, Dolinsky JS, Goldgar DE, Shimelis H, Couch FJ, Chao EC, LaDuca H. Male breast cancer in a multi-gene panel testing cohort: insights and unexpected results. Breast Cancer Res Treat. 2017;161(3):575–86.
9. Fostira F, Saloustros E, Apostolou P, Vagena A, Kalfakakou D, Mauri D, Tryfonopoulos D, Georgoulias V, Yannoukakos D, Fountzilas G, Konstantopoulou I. Germline deleterious mutations in genes other than BRCA2 are infrequent in male breast cancer. Breast Cancer Res Treat. 2018;169(1):105–13.
10. Huang H, Shen H, Wang Y, Wang X, Chen Q, Yan H, Liu Z, Shi X. LOH12CR10 is a novel tumor suppressor inhibiting tumor growth through deregulation of G1/S checkpoint in human colorectal carcinoma. Curr Mol Med. 2018;18(1):25–35.
11. Tang X, Chen Z, Deng M, Wang L, Nie Q, Xiang JW, Xiao Y, Yang L, Liu Y, Li DW. The sumoylation modulated tumor suppressor p53 regulates cell cycle checking genes to mediate lens differentiation. Curr Mol Med. 2018;18(8):556–65.
12. Bell R, Barraclough R, Vasieva O. Gene expression meta-analysis of potential metastatic breast cancer markers. Curr Mol Med. 2017;17(3):200–10.
13. Cabanillas R, Diñeiro M, Castillo D, Pruneda PC, Penas C, Cifuentes GA, de Vicente A, Durán NS, Álvarez R, Ordóñez GR, Cadiñanos J. A novel molecular diagnostics platform for somatic and germline precision oncology. Mol Genet Genom Med. 2017;5(4):336–59.
14. Gréen A, Gréen H, Rehnberg M, Svensson A, Gunnarsson C, Jonasson J. Assessment of HaloPlex amplification for sequence capture and massively parallel sequencing of arrhythmogenic right ventricular cardiomyopathy-associated genes. J Mol Diagn. 2015;17(1):31–42.
15. Krieger E, Vriend G. New ways to boost molecular dynamics simulations. J Comput Chem. 2015;36(13):996–1007.
16. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics. 2010;27(3):343–50.
17. Lovell SC, Davis IW, Arendall WB III, De Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins Struct Funct Bioinform. 2003;50(3):437–50.
18. Berendsen HJ, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel molecular dynamics implementation. Comput Phys Commun. 1995;91(1–3):43–56.
19. Chikaraddi SB, Krishnappa R, Deshmane V. Male breast cancer in Indian patients: is it the same? Indian J Cancer. 2012;49(3):272.
20. Rai B, Ghoshal S, Sharma SC. Breast cancer in males: a PGIMER experience. J Cancer Res Ther. 2005;1(1):31.
21. Shah P, Robbani I, Shah O. Clinicopathological study of male breast carcinoma: 24 years of experience. Ann Saudi Med. 2009;29(4):288.
22. Anderson WF, Jatoi I, Tse J, Rosenberg PS. Male breast cancer: a population-based comparison with female breast cancer. J Clin Oncol. 2010;28(2):232.
23. Tan Y, Raychaudhuri P, Costa RH. Chk2 mediates stabilization of the FoxM1 transcription factor to stimulate expression of DNA repair genes. Mol Cell Biol. 2007;27(3):1007–16.
24. Apostolou P, Fostira F. Hereditary breast cancer: the era of new susceptibility genes. Biomed Res Int. 2013. https://doi.org/10.1155/2013/747318.
25. Grigoraş A, Rîşcanu L, Amălinei C. Meningiomas–insights into genetics and correlations with histological features. Arch Clin Cases. 2018. https://doi.org/10.22551/2018.18.0501.10122.
26. Kim Y, Jho EH. Regulation of the Hippo signaling pathway by ubiquitin modification. BMB Rep. 2018;51(3):143.
27. Hart SN, Hoskin T, Shimelis H, Moore RM, Feng B, Thomas A, Lindor NM, Polley EC, Goldgar DE, Iversen E, Monteiro AN. Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. Genet Med. 2018;8:1.
28. Herzig S, Shaw RJ. AMPK: guardian of metabolism and mitochondrial homeostasis. Nat Rev Mol Cell Biol. 2018;19(2):121.
29. Campagne A, Zielinski D, Michaud A, Le Corre S, Dingli F, Chen H, Vassilev I, Lee MK, Servant N, Loew D, Pasmant E. The BAP1 deubiquitinase complex is a general transcriptional co-activator. bioRxiv. 2018. https://doi.org/10.1101/244152.
30. Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H, Meetei AR, Andreassen PR. Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. Oncogene. 2014;33(40):4803.