In silico analysis of deleterious single nucleotide polymorphisms in human BUB1 mitotic checkpoint serine/threonine kinase B gene

Fatemeh Akhoundi a,b, Nikpour Parvaneh c, Emadi-Baygi Modjtaba a,b,⁎

a Department of Genetics, Faculty of Basic Sciences, Shahrekord University, Shahrekord, Iran
b Research Institute of Biotechnology, Shahrekord University, Shahrekord, Iran
c Department of Genetics and Molecular Biology, Faculty of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran

Abstract

One of the major challenges in the analysis of human genetic variation is to distinguish mutations that are functionally neutral from those that contribute to disease. BubR1 is a key protein mediating spindle-checkpoint activation that plays a role in the inhibition of the anaphase-promoting complex/cyclosome (APC/C), delaying the onset of anaphase and ensuring proper chromosome segregation. Owing to the importance of BUB1B gene in mitotic checkpoint a functional analysis using different in silico approaches was undertaken to explore the possible associations between genetic mutations and phenotypic variation. In this work we found that 3 nsSNPs I82N, P334L and R814H have a functional effect on protein function and stability. A literature search revealed that R814H was already implicated in human diseases. Additionally, 2 SNPs in the 5′UTR region was predicted to exhibit a pattern change in the internal ribosome entry site (IRES), and eight MicroRNA binding sites were found to be highly affected due to 3′UTR SNPs. These in silico predictions will provide useful information in selecting the target SNPs that are likely to have functional impact on the BUB1B gene.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords:
Single nucleotide polymorphism (SNP) Pathogenic variants Spindle assembly checkpoint BUB1B gene

1. Introduction

The spindle assembly checkpoint (SAC) is a cell-cycle surveillance mechanism that prevents premature anaphase entry until all chromosomes have completely aligned at the metaphase plate. SAC is composed of the checkpoint proteins BubR1, Bub3, and Mad2, associated with the APC/C coactivator Cdc20. The checkpoint system acts to inhibit the activity of the large multi-protein E3 ubiquitin ligase known as the anaphase promoting complex/cyclosome (APC/C), by binding to the co-activating subunit Cdc20. BubR1 is a key protein mediating spindle-checkpoint activation That directly binds to Cdc20 and inhibits APC/C activity (Kaisari et al., 2016). The corresponding BUB1B gene is located on chromosome 15q15 and is composed of 23 exons that encodes 1050 amino acids (Davenport et al., 1999; Hanks et al., 2012). BUB1B mutated in several cancers including colorectal, lung, breast, hematopoietic malignancies and in a rare human hereditary condition called premature chromatid separation syndrome (mosaic variegated aneuploidy) (Kapanidou et al., 2015; Kops et al., 2005; Hanks et al., 2004, 2006; Matsuura et al., 2006; Suijkerbuijk et al., 2010; Ohshima et al., 2000).

Single nucleotide polymorphisms (SNPs) found in any position throughout the genome in exons, introns, intergenic regions, promoters and enhancers (Drazen et al., 1999). Many SNPs are phenotypically neutral. However, others could predispose human to disease or influence their response to a drug. Nonsynonymous SNPs (nsSNPs) that lead to an amino acid substitution in the corresponding protein product are of particular interest as they are responsible for nearly half of the known gene lesions responsible for human inherited disease (Krawczak et al., 2000).

Computational analyses of BUB1B gene for harmful nsSNPs have not been carried out until now; therefore, we applied different publicly available computational tools according to Fig. 1. The value and novelty of this study is to prioritize SNPs with functional significance from an enormous number of neutral non-risk alleles of BUB1B and provides new insights for further genetic association studies.

2. Material and method

2.1. Dataset

The NCBI database of SNPs (Sherry et al., 2001), dbSNP available at http://www.ncbi.nlm.nih.gov/SNP and SWISSProt databases (Bairoch and Apweiler, 1996) were used to obtain the SNP information [SNP ID, amino acid position, mRNA accession number NM_001211, and Protein accession number NP_001202.4] of the human BUB1B gene for our computational analyses.

http://dx.doi.org/10.1016/j.mgene.2016.05.002
2214-5400/© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
2.2. Predicting functional context of missense mutation

The functional context of nsSNPs was predicted using SIFT (Sorting Intolerant from Tolerant), PolyPhen 2.0, I-Mutant 3.0 and PROVEAN (Protein Variation Effect Analyzer) (Table 1).

SIFT predict whether an amino acid substitution in a protein would be tolerated or damaging. The amino acid substitution is predicted damaging when the score is below or equal to 0.05, and tolerated if the score is greater than 0.05 (Ng and Henikoff, 2003). PolyPhen input is the amino acid sequence of protein or SNP identifier with the nsSNP. The output levels of probably damaging and possibly damaging were classified as functionally significant (≥0.5) and the benign level being classified as tolerated (≥0.5) (Ramensky et al., 2002). I-Mutant 3.0 performed analyses based on the protein sequence combined with mutational position and correlated new residue and the output result of the predicted free energy change (DDG) classifies the prediction into one of three classes: largely unstable (DDG < −0.5 kcal/mol), largely stable (DDG > −0.5 kcal/mol), or neutral (−0.5 ≤ DDG ≥ 0.5 kcal/mol) (Capriotti et al., 2008). PROVEAN is able to provide predictions for any type of protein sequence variations including amino acid substitutions, and in-frame insertions and deletions (Choi et al., 2012). The PROVEAN predict a protein variant to be neutral if the score is above the threshold. The cutoff score −2.5 indicates a deleterious substitution (Manickam et al., 2014).

Furthermore, we used SNP&GO, PHD-SNP Predictor of human deleterious single nucleotide polymorphisms and PANTHER (Protein Analysis Through Evolutionary Relationships) tools to filter the disease-associated nsSNPs (Table 1).

SNP&GO predict SNPs are or are not disease-associated with including the protein FASTA sequence and Gene Ontology terms. The probability score higher than 0.5 indicates the disease related effect of mutation on the parent protein function (Calabrese et al., 2009). PhD-SNP predicts whether the given amino acid substitution leads to disease associated or neutral along with the reliability index score (Capriotti et al., 2006). PANTHER comprehensive software system predicts the likelihood of a particular nsSNP to cause a functional impact on a protein. The cutoff subPSEC score −3 indicates a deleterious substitution (Thomas et al., 2003).

2.3. Biophysical validation of nsSNPs

NetSurfP predicts the surface and, solvent accessibility of amino acids, using the amino acid FASTA sequence. The solvent accessibility has been predicted in two classes as either buried or exposed, based on the accessibility of the amino acid residues to the solvent, respectively. The reliability of this prediction method is in the form of Z-score. The Z-score highlights the surface prediction reliability, but is not associated with the secondary structure (Petersen et al., 2009).

Finding 3D structure of proteins is helpful in predicting the impact of SNPs on the structural level and in showing the degrees of alteration. I-TASSER generates a full length model of proteins by excising continuous fragments from threading alignments and then reassembling them using replica-exchanged Monte Carlo simulations. Low temperature replicas (decays) generated during the simulations are clustered by SPICKER and the top five cluster centroids are selected for generating full atomic models. The quality of prediction models was reflected in the form of c-scores (−5 to 2) (Roy et al., 2010, 2012). The native structure was mutated with the most deleterious amino acid substitution predicted in this study, using Swiss PDB viewer and Chimera (Kaplan and Littlejohn, 2001; Pettersen et al., 2004).

In addition, we used HOPE Project that provides the 3D structural visualization of mutated proteins, and gives the results by using UniProtKB and predictions from DAS-servers. FASTA sequence of whole protein and selection of mutant variants is considered to be an input option, the output is based on the structural variation between the mutant and the wild-type residues (Venselaar et al., 2010).

2.4. Predictions of protein-protein interactions

STRING (Search Tool for the Retrieval of Interacting proteins) is a database and web resource dedicated to protein–protein interactions, including direct (physical) and indirect (functional) interactions (Jensen et al., 2009); the database contains information from: genomic context, experimental repositories, co-expression and public text collections (Szklarczyk et al., 2011).

2.5. Functional SNPs in UTR found by the UTRscan

The UTRscan program allows one to search the user-submitted sequences for any of the patterns collected in the UTR site (Grillo Fig. 1. Schematic representation of computational tools for in silico analysis of BUB1B gene.

Fig. 1. Schematic representation of computational tools for in silico analysis of BUB1B gene.
et al., 2010). If different sequences for each UTR SNP are found to have different functional patterns, that the particular UTR SNP is predicted to have functional significance.

2.6. PolymiRTS database (version 3.0) for polymorphism in microRNA target site

PolymiRTS database was designed specifically for the analysis of non-coding SNPs namely 3′ UTR. The polymorphic microRNA target sites are classified into four classes according to Table 1 (Bhattacharya et al., 2013). PolymiRTS of ‘D’ may cause loss of normal repression; PolymiRTS of class ‘C’ may cause abnormal gene repression control. Therefore, these two classes of PolymiRTS are most likely to have functional impacts.

2.7. 1000 Genomes Project

The 1000 Genomes Project (Consortium, 2010) is sequencing the entire genome of approximately 2500 individuals from different worldwide populations. The aim of the 1000 Genomes Project is to determine most of the genetic variation that occurs at a population frequency greater than 1%.

3. Results

3.1. SNP dataset from dbSNP

The BUB1B gene investigated in this work was retrieved from dbSNP database (Table 4). It contained a total of 827 SNPs: 90 were non-synonymous SNPs (nsSNPs), 57 were in non-coding regions, which comprises of 21 SNPs in 5′ UTR region and 36 SNPs in 3′ UTR region.

3.2. Prediction of functional mutations

Of the 90 nsSNPs used in our analysis, 18 nsSNPs were identified to be deleterious with SIFT and the results were listed in Table 2.

A total of 46 nsSNPs was predicted to be damaging and the remaining 44 nsSNPs were categorized as benign with Polyphen 2.0 and the results were listed in Table 2. Out of 90 nsSNPs, 27 nsSNP were predicted to be neutral mutation (< 0.5 ≤ DDG ≤ 0.5 kcal/mol), 60 nsSNP were predicted to be “large decrease” (≤ 0.5 kcal/mol) and 3 nsSNP were predicted to be “large increase” (> 0.5 kcal/mol). I-Mutant 3.0 predicted 63 of SNPs to affect the stability of protein structure (Table 2). All the nsSNPs submitted to SIFT and PolyPhen 2.0 and I-Mutant 3.0 were submitted as input to PROVEAN. Out of 90 nsSNPs, 17 nsSNP were predicted to be deleterious and 73 were found to be neutral (Table 2).

The accuracy of the in silico techniques for prioritizing deleterious SNPs can be increased by combining different computational methods. Out of 90 nsSNPs, SIFT, PolyPhen, I-Mutant 3.0 and PROVEAN predicted 8 nsSNPs as deleterious (Fig. 2).

SNP&GO, PhD-SNP and PANTHER were performed to validate the results obtained from four tools. Out of 8 nsSNPs that predicted to be deleterious with SIFT, Polyphen, I-Mutant and PROVEAN; SNP&GO predicted 3 nsSNP, PhD-SNP predicted 4 nsSNP and PANTHER predicted 5 nsSNP to be associated with disease (Table 3).

Finally out of 90 nsSNP, we found 3 nsSNPs namely rs55355571 (I82N), rs141953425 (P334L) and rs28989182 (R814H) that are common in all (SIFT, Polyphen, I-Mutant, PROVEAN, PHD-SNP, SNP&GO, PANTHER) prediction.

| Table 1 | In silico approaches as online tools. |
|----------|--------------------------------------|
| Server   | Feature                              | URL                                   | Reference           |
| SIFT     | Focuses more on the sequence preservation over the evolutionary time in predicting the effect of residue substitutions on function. | http://sift.bii.a-star.edu.sg/index.html | Kumar et al. (2009), Magesh and Doss (2014) |
| PolyPhen 2.0 | Sequence and structure based method that predicts the possible impact of an amino acid substitution on the structure and function of a protein. | http://genetics.bwh.harvard.edu/php2 | Adzhubei et al. (2010) |
| 1-Mutant 3.0 | Support vector machine (SVM) based predictors of protein stability changes upon single amino acid substitution. | http://gpcr2.biocomp.unibo.it/cgi/predictors/1-Mutant3.0/1-Mutant3.0.cgi | Capriotti et al. (2008) |
| PROVEAN  | Sequence based predictor that estimates whether a protein sequence variation affects protein function. | http://provean.jcvi.org | Choi et al. (2012) |
| SNP&GO   | Support vector machine (SVM) based web server that combine protein structural/functional parameters and sequence analysis derived information. | http://snps.biofold.org/snps-and-go/snps-and-go.html | Magesh and Doss (2014) |
| PhD-SNP  | SVM based on evolutionary information. | http://gpcr.biocomp.unibo.it/cgi/predictors/PhD-SNP/PhD-SNP.cgi | Magesh and Doss (2014) |
| PANTHER  | Protein family and subfamily database that predicts the frequency of occurrence of amino acid at a particular position in evolutionary related protein sequences. | http://pantherdb.org/tools/csnpScoreForm.jsp | Mi et al. (2005) |
| UTRscan  | Analyzing functional impacts of UTR SNPs. | http://www.ba.ibb.cnrs.fr/Big/UTRscan | Pesole et al. (1999) |
| NetSurfP | Analysis of SNP effects on surface and solvent accessibility of protein. | http://www.cbs.dtu.dk/services/NetSurfP | Petersen et al. (2009) |
| I-TASSER | Protein structure prediction server. | http://zhanglab.ccmb.med.umich.edu/I-TASSER | Zhang (2008) |
| HOPE Project | An automatic mutant analysis server for studying the structural features of native protein and the variant models. | http://www.cmbi.ru.nl/hope/home | Venselaar et al. (2010) |
| STRING  | Database of known and predicted protein-protein interactions. | http://string-db.org | Von Mering et al. (2005) |
| PolymiRTS | The polymorphic microRNA target sites are classified into four classes: ‘D’ (the derived allele disrupts a conserved microRNA site), ‘N’ (the derived allele disrupts a nonconserved microRNA site), ‘C’ (the derived allele creates a new microRNA site) and ‘O’ (other cases when the ancestral allele cannot be determined unambiguously). | http://comphio.uthsc.edu/miRSNP/ | Bhattacharya et al. (2013) |
| 1000 Genomes Project | A resource about human genetic variation that will be used in many studies of particular phenotypes, such as complex diseases or drug response. | http://www.1000genomes.org | Via et al. (2010) |
## Table 2
List of nsSNP analysis by SIFT, PolyPhen-2, I-Mutant 3.0 PROVEAN respectively.

| rsID          | Amino acid change | SIFT Score | PolyPhen-2 Score | I-Mutant 3.0 Score | PROVEAN Score | Score |
|---------------|-------------------|------------|------------------|--------------------|---------------|-------|
| rs202114756   | S384G             | Tolerated  | 0.31             | Possibly damaging  | 0.86          | Large decrease |
| rs202136016   | E21K              | Tolerated  | 0.09             | Probably damaging  | 1.00          | Large decrease |
| rs143346774   | H850R             | Tolerated  | 0.61             | Probably damaging  | 0.99          | Neutral       |
| rs28989181    | L844F             | Tolerated  | 0.98             | Large decrease     | 0.82          | Large decrease |
| rs28989182    | R814H             | Damaging   | 1.00             | Possibly damaging  | 1.43          | Large decrease |
| rs28989187    | S691L             | Tolerated  | 0.08             | Large increase     | 0.78          | Large decrease |
| rs289891235   | D576E             | Tolerated  | 0.37             | Large decrease     | 0.09          | Neutral       |
| rs57153880    | G316D             | Tolerated  | 0.60             | Large decrease     | 0.02          | Neutral       |
| rs57759191    | R727C             | Tolerated  | 1.00             | Large decrease     | 0.46          | Neutral       |
| rs1017842     | E390D             | Tolerated  | 0.01             | Large decrease     | 0.09          | Neutral       |
| rs1801528     | V618A             | Tolerated  | 1.00             | Large decrease     | 0.16          | Neutral       |
| rs200788206   | Q350K             | Tolerated  | 0.79             | Large decrease     | 0.72          | Neutral       |
| rs200060772   | S691L             | Tolerated  | 0.08             | Large increase     | 0.08          | Neutral       |
| rs14705245    | A764V             | Tolerated  | 0.38             | Large decrease     | 0.09          | Neutral       |
| rs14346774    | H850R             | Tolerated  | 0.61             | Large decrease     | 0.98          | Neutral       |
| rs143559902   | D675E             | Tolerated  | 0.01             | Large decrease     | 0.09          | Neutral       |
| rs145020343   | C825F             | Tolerated  | 0.35             | Large decrease     | 0.96          | Neutral       |
| rs145020854   | E184Q             | Damaging   | 0.01             | Large decrease     | 0.36          | Neutral       |
| rs14536174    | A335G             | Tolerated  | 0.84             | Large decrease     | 0.95          | Neutral       |
| rs14557829    | I567V             | Tolerated  | 0.3            | Large decrease     | 0.52          | Neutral       |
| rs14636789    | L258F             | Tolerated  | 0.74             | Large decrease     | 0.18          | Neutral       |
| rs14679555    | T493I             | Tolerated  | 0.64             | Large decrease     | 0.07          | Neutral       |
| rs146821149   | R886S             | Damaging   | 0.3            | Large decrease     | 0.02          | Neutral       |
| rs14725072    | G376V             | Tolerated  | 0.17             | Large decrease     | 0.18          | Neutral       |
| rs14749987    | V4M               | Tolerated  | 0.23             | Large decrease     | 0.07          | Neutral       |
| rs14783258    | S83G              | Damaging   | 0.01             | Large decrease     | 0.06          | Neutral       |
| rs148159407   | N26D              | Tolerated  | 0.52             | Large decrease     | 0.93         | Neutral       |
| rs148348158   | T648H             | Tolerated  | 0.19             | Large decrease     | 0.00          | Neutral       |
| rs149628229   | D579G             | Tolerated  | 0.11             | Large decrease     | 0.32         | Neutral       |
| rs149955447   | E813A             | Tolerated  | 0.07             | Large decrease     | 0.97          | Neutral       |
| rs15070673    | S797A             | Tolerated  | 0.13             | Large decrease     | 0.45          | Neutral       |
| rs150983783   | R241Q             | Tolerated  | 0.15             | Large decrease     | 0.18          | Neutral       |
| rs181352808   | H836Q             | Tolerated  | 0.29             | Large decrease     | 0.95          | Neutral       |
| rs184484975   | M626V             | Tolerated  | 0.61             | Large decrease     | 0.06          | Neutral       |
| rs190900940   | Y134F             | Tolerated  | 0.14             | Large decrease     | 0.92          | Neutral       |
| rs199509124   | P222L             | Damaging   | 0.01             | Large decrease     | 0.77          | Large increase |
| rs199746555   | V274A             | Damaging   | 0.01             | Large decrease     | 0.86          | Large decrease |
| rs200060772   | S691L             | Tolerated  | 0.08             | Large decrease     | 0.99          | Neutral       |
| rs20078206    | Q350K             | Tolerated  | 0.79             | Large decrease     | 0.98          | Neutral       |
| rs20097833    | K542R             | Tolerated  | 0.35             | Large decrease     | 0.05          | Neutral       |
| rs201251790   | R421W             | Tolerated  | 0.09             | Large decrease     | 0.02          | Neutral       |
| rs20130106    | E21K              | Tolerated  | 0.09             | Large decrease     | 1.00          | Neutral       |
| rs20214756    | S384G             | Tolerated  | 0.31             | Large decrease     | 0.80          | Neutral       |
| rs202182335   | A735G             | Damaging   | 0.06             | Large decrease     | 0.72          | Neutral       |
| rs367543489   | L258F             | Tolerated  | 0.11             | Large decrease     | 0.99          | Neutral       |
| rs368032159   | K488N             | Tolerated  | 0.15             | Large decrease     | 0.04          | Neutral       |
| rs368079817   | Q42R              | Tolerated  | 0.25             | Large decrease     | 0.99          | Negative      |

(continued on next page)
3.3. In silico biophysical validation of nsSNPs

Based on the in silico analyses performed, 3 nsSNPs was selected for further analyses. The location and the type of a mutated residue affect the stability of the protein due to mutation decreases. NetSurfP Z-score allows the identification of the most reliable/unreliable predictions for both buried and exposed amino acids. A huge drift in the Z-score was not observed for 3 nsSNPs as given in Table 5. For any of 3 nsSNPs, the class assignment does not change.

The I-TASSER tool created the 5 full-length models for Bub1 protein (with C-scores: −2.83, −2.95 and −3.00) by excising top 10 structures with C-scores after targeting the PDB library hits (Table 5). Top 10 proteins in the PDB which are structurally closest to the predicted models. Among the 5 predicted models, model 1 (Fig. 3) carried the high-quality confidence in the form of C-score (−0.24), TM-score (0.68 ± 0.12), and the RMSD (9.6 ± 4.6 Å) (Table 6). We did not perform any molecular dynamics structure optimization; therefore, our 3D homology model is a preliminary model implicating the disruptive role of the SNPs.

| rsID     | Amino acid change | SIFT | PolyPhen-2 | I-Mutant 3.0 | PROVEAN | Score |
|----------|-------------------|------|------------|--------------|---------|-------|
| rs358996088 | F781L             | 0.16 | Benign     | 0.075        | Large decrease | −1.32 |
| rs370388424 | P640L             | 0.29 | Benign     | 0.020        | Large decrease | −0.60 |
| rs370655726 | C356S             | 0.74 | Probably damaging | 1.000 | Large decrease | −0.69 |
| rs371124423 | C700R             | 0.38 | Benign     | 0.068        | Large decrease | −0.26 |
| rs371305662 | T291K             | 0.84 | Benign     | 0.100        | Large decrease | −0.46 |
| rs372003254 | D846E             | 0.1  | Possibly damaging | 0.626 | Neutral | 0.24 |
| rs372569297 | I755T             | 0.75 | Benign     | 0.000        | Large decrease | −1.59 |
| rs372566676 | K454R             | 0.21 | Probably damaging | 1.000 | Neutral | 0.12 |
| rs377785523 | T658I             | 0.16 | Benign     | 0.000        | Large decrease | 0.25 |
| rs378302562 | A108T             | 0.49 | Probably damaging | 0.982 | Large decrease | −0.71 |
| rs374682772 | V333I             | 0.35 | Benign     | 0.002        | Large decrease | −0.84 |
| rs375105548 | I854V             | 0.44 | Benign     | 0.007        | Large decrease | −0.65 |
| rs375388175 | I703T             | 0.76 | Probably damaging | 0.985 | Large decrease | −1.94 |
| rs375798678 | Q181R             | 1    | Benign     | 0.000        | Neutral | 0.06 |
| rs375885859 | C51R              | 0.11 | Benign     | 0.000        | Neutral | 0.07 |
| rs376072341 | P632L             | 0.1  | Benign     | 0.001        | Large decrease | −0.40 |

Project Hope revealed the 3D structure of the proteins with its new residue. Furthermore, it described the reaction and physicochemical properties of these candidates. Here we present the results upon each candidate and discuss the conformational variations and interactions with the neighboring amino acids:

A/G Mutation (rs28989182) caused substitution of the amino acid from Arginine into a Histidine at position 814 (R814H). For this variant the mutated residue is smaller (Fig. 1); this might lead to loss of interactions. The wild-type residue was positively charged, the mutated residue is neutral. Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein. Additionally, the structural analysis of H814 showed some clashes for Phe822 which may contribute to the extra energy in the protein structure, and hence the decrease in stability (Fig. 4).

A/T mutation (rs553555716) resulted in a change of the Isoleucine to Asparagine at position 82 (I82N). The wild type residue is smaller and more hydrophobic than the mutated residue (Fig. 6). The residue

### Table 3

List of nsSNP predicted as disease associated by PHD-SNP, SNP&GO and PANTHER server.

| rsID       | Amino acid change | PHD-SNP | SNP&GO | PANTHER subPSEC score |
|------------|-------------------|---------|--------|-----------------------|
| rs357987333 | I82N              | Disease | Disease | Deleterious | −5.89613 |
| rs533967444 | I727N             | Disease | Disease | Deleterious | −2.73516 |
| rs199743655 | V274A             | Neutral | Neutral | Tolerated | −2.50206 |
| rs199509124 | P222L             | Neutral | Neutral | Deleterious | −3.07979 |
| rs141953425 | P334L             | Disease | Disease | Deleterious | −3.73233 |
| rs56158360  | R244H             | Neutral | Neutral | Deleterious | −4.34981 |
| rs17851677  | P378S             | Neutral | Neutral | −3.97828 |
| rs2589182   | R814H             | Disease | Disease | Deleterious | −7.86508 |

### Table 4

Surface accessibility of wild-type and mutant variants in BUB1B.

| Amino acid | Class assignment | Position | Relative surface accessibility | Absolute surface accessibility | Z-fit score for RSA prediction |
|------------|------------------|----------|--------------------------------|-------------------------------|-------------------------------|
| 1          | Buried           | 82       | 0.02                           | 3.829                         | 1.418                         |
| N          | Buried           | 0.02    | 3.177                         | 1.838                         |
| P          | Exposed          | 0.546   | 77.520                        | 1.915                         |
| L          | Exposed          | 0.249   | 63.975                        | 0.441                         |
| R          | Exposed          | 0.436   | 99.798                        | 0.340                         |
| H          | Exposed          | 0.423   | 76.998                        | 0.280                         |

### Table 6

Top 10 templates used by I-TASSER to create the high quality models for human BUB1B secondary structure.

| Rank | PDB hit | Iden1 | Iden2 | Cov. | Norm. Z-score |
|------|---------|-------|-------|------|---------------|
| 1    | 3e7eA   | 0.22  | 0.08  | 0.31 | 1.71          |
| 2    | 1w1A    | 0.08  | 0.20  | 0.87 | 2.36          |
| 3    | 3e7eA   | 0.08  | 0.31  | 0.92 | 1.44          |
| 4    | 4spB    | 0.09  | 0.19  | 0.92 | 1.44          |
| 5    | 3e7eA   | 0.23  | 0.08  | 0.30 | 2.68          |
| 6    | 3cm9S   | 0.07  | 0.11  | 0.55 | 1.19          |
| 7    | 3e7eA   | 0.22  | 0.08  | 0.31 | 1.51          |
| 8    | 4spB    | 0.08  | 0.19  | 0.89 | 2.06          |
| 9    | 3e7eA   | 0.22  | 0.08  | 0.31 | 1.51          |
| 10   | 4k77A   | 0.10  | 0.19  | 0.85 | 2.31          |
is buried in the core of a domain, annotated in UniProt as: “BUB1 N terminal”. The mutation will cause loss of hydrophobic interactions in the core of the protein. This residue is part of an interprotein domain named “Mitotic checkpoint serine/threonine protein kinase Bub1/Mitotic spindle checkpoint component Mad3” (IPR015661).

C/T (rs141953425) leads to conversion of Proline into a Leucine at position 334 (P334L). The wild-type and mutant amino acid differs in sizes; the mutated residue is bigger and this might lead to displace of the mutant residue. Prolines are known to have a very rigid structure, sometimes forcing the backbone in a specific conformation. The mutation can disturb this special conformation. This residue is part of an interprotein domain named “Mitotic checkpoint serine/threonine protein kinase Bub1/Mitotic spindle checkpoint component Mad3”.

Fig. 3. 3D structure of BubR1 predicted with I-TASSER.

Table 6
I-TASSER results carrying C-score, TM-score and RMSD regarding selected secondary structure (native protein model 1).

| Model   | C-score | Exp. TM-score | Exp. RMSD | No. of decoys | Cluster density |
|---------|---------|---------------|-----------|---------------|----------------|
| Model 1 | −0.24   | 0.68 ± 0.12   | 9.6 ± 4.6 Å | 312           | 0.3039         |
| Model 2 | −1.01   |               |           | 258           | 0.1404         |
| Model 3 | −2.83   |               |           | 60            | 0.0228         |
| Model 4 | −2.95   |               |           | 57            | 0.0202         |
| Model 5 | −3.00   |               |           | 57            | 0.0191         |

Fig. 4. Deep view of superimposed structure of wild and mutant residue at 814 position. The main protein core is shown in gray color while the wild type and mutated residues are shown in green and red colors respectively. SNP ID: rs28989182, protein position 814 changed from Arginine to Histidine.

Fig. 5. SNP ID: rs28989182, protein position 814 changed from Arginine (green) to Histidine (red).

Fig. 6. Superimposed structure of (wild type) Ile and (mutant) Asn residues at 82 position.

Fig. 7. SNP ID: rs553552716, protein position 82 changed from Isoleucine (left image) to Aspargine (right image).
Chimera (Figs. 5, 7, 8) and Swiss PDB viewer were used to visualize the structural features of amino acids in native and mutant protein chains. During structural visualization for all 3 mutations, only mutant residue (Histidine) at 814 position showed a network of clashes with Phe822 (Fig. 9).

3.4. Protein-protein interactions analysis

The interaction analysis revealed that BUB1B is related to Cell Division Cycle 20 homolog (CDC20), Budding Uninhibited By Benzimidazoles 3 (BUB3), Cancer Susceptibility Candidate 5 (CASC5), MAD2 Mitotic Arrest Deficient-Like 1 (MAD2L1), Cell Division Cycle 27 homolog (CDC27), Centromere Protein E (CENPE), BUB1 Mitotic Checkpoint Serine/Threonine Kinase (BUB1), ZW10 Interacting Kinetochore Protein (ZWINT), Anaphase Promoting Complex Subunit 2 (ANAPC2), Cell Division Cycle 16 homolog (CDC16) (Fig. 10). Furthermore, our literature search
demonstrated that BubR1 interacts with Bub3, Cdc20, and Mad2 (Kapanidou et al., 2015).

### 3.5. Functional SNPs in UTR found by UTRscan server

All of the 57 UTR SNPs was analyzed using UTRscan. After comparing the functional elements for each UTR SNP, we predicted that two SNPs, namely rs375434078 and rs538302864 in 5’ UTR are related to the functional pattern change of IRES (Table 7). Internal ribosome entry site (IRES) is bound by an internal mRNA ribosome that is an alternative mechanism of translation initiation compared to the common 5’-cap dependent ribosome scanning mechanism (Pickering and Willis, 2005).

#### 3.6. Functional SNPs in 3’ untranslated regions (UTR) predicted by PolymiRTS database 3.0

Among 36 SNPs in 3’ UTR region of BUB1B gene, 3 functional SNPs was predicted that among them, only one SNP disrupts 8 miRNAs conserved site (ancestral allele with support ≥2), while all of them create 8 new miRNA site. The results are listed in Table 8.

### 4. Discussion

The identification of SNPs responsible for specific phenotypes with molecular approaches seems to be expensive and time-consuming (Chen and Sullivan, 2003), hence computational approaches can help in narrowing down the number of missense mutations to be screened in genetic association studies and for a better understanding of the functional and structural aspects of the parent protein.

Previous studies on polymorphisms screening using in silico analysis helped in predicting the functional nsSNPs associated with genes such as G6PD (Rajith, 2011), ATM (Doss and Rajith, 2012), PTEN (Doss and Rajith, 2013), BRAF (Hussain et al., 2012). Our results also revealed that implementations of different algorithms often serve as powerful tools for prioritizing candidate functional nsSNPs. Recent work by Thusberg and Vihinen (2009) compared different in silico tools, out of which SIFT and PolyPhen were reported to have better performance in identifying deleterious nsSNPs. The accuracy of SIFT and PolyPhen 2.0 was further validated by Hicks et al. (2011), which makes these tools more applicable for the prediction. I-Mutant 3.0 was used which evaluate the stability change upon single amino acid mutation that ranked as one of the most reliable predictor based on the work performed by Khan and Vihinen (2010).

Based on these in silico studies, we select SIFT, PolyPhen, I-Mutant, PROVEAN, SNP&GO, PHD-SNP and PANTHER for the screening of functional mutation in BUB1B gene. By comparing the scores of all 7 methods, 3 nsSNPs with positions I82N, P334L and R814H were found to be highly significant.

The 5’ and 3’ UTR SNPs was analyzed using UTRscan. Due to the importance of the translational regulation of miRNAs, we further studied whether the 3’ UTR SNPs changes the profile of microRNA binding to the BUB1B gene using PolymiRTS. Two SNPs in the 5’ UTR was predicted to influence the translation pattern of the BUB1B gene through UTRscan analysis, and three 3’ UTR SNPs may affect microRNA binding sites, as determined through PolymiRTS. Protein-protein interaction analysis showed the interaction of BUB1B with ten different genes. Therefore, any changes in the protein function would have an impact on many pathways involved in disease.

In conclusion, we surveyed and compared available databases such as NCBI, dbSNP, 1000 Genome project along with in silico prediction programs to assess the effects of deleterious functional variants on the protein functions. Analyzing deleterious nsSNPs by both sequence and structure level has the added advantage of being able to assess the reliability of the generated prediction results by cross-referencing the results from both approaches. One striking observation was the identification of rs28989182 (R814H), that associated with Mosaic Variegated Aneuploidy Syndrome (Bairoch and Apweiler, 1996), lie within a serine/threonine kinase domain of BubR1 protein. Only this residue type was found at this position. Mutation of a 100% conserved residue is usually damaging for the protein. Both I82N and P334L mutations occurred in the N terminal region of BubR1; Therefore, these mutations may compromise its binding to Bub1, Mad2 and cdc20 resulting plausible failure of the corresponding checkpoint. In addition rs149437374 and rs143807849 in 3’ UTR that disrupts a conserved of 8 miRNAs site are genotyped by

### Table 7

List of mRNA UTR SNPs that were predicted to be of functional significance by UTRscan server.

| SNP ID       | Nucleotide change | UTR position | Functional element change |
|--------------|------------------|--------------|---------------------------|
| rs375434078  | C/T              | 5’ UTR       | IRES → no pattern         |
| rs538302864  | A/G              | 5’ UTR       | IRES → no pattern         |

### Table 8

Prediction result of PolymiRTS database.

| dbSNP ID | miR ID       | Conservation | miRSite | Function class |
|----------|--------------|--------------|---------|----------------|
| rs149437374 | hsa-miR-130a-3p | 2            | ATGCACTAccatt | D              |
| rs143807849 | hsa-miR-539-5p | 4            | CaATTTCTccct | C              |
| rs1047193   | hsa-miR-4450  | 4            | CcatttCTCTCA  | C              |
|            | hsa-miR-3666  | 5            | CcatttCTCTCA  | C              |
|            | hsa-miR-130b-3p | 2          | ATGCACTAccatt | D              |
|            | hsa-miR-301a-3p | 2          | ATGCACTAccatt | D              |
|            | hsa-miR-301b   | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-301b   | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-3955   | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-454-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-4671-3p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-323a-5p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-876-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-454-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-4671-3p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-323a-5p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-876-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-454-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-4671-3p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-323a-5p | 2         | ATGCACTAccatt | D              |
|            | hsa-miR-876-3p | 2           | ATGCACTAccatt | D              |
|            | hsa-miR-454-3p | 2           | ATGCACTAccatt | D              |
Adzhubei, I.A., et al., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7 (4), 248–249.

Bairoch, A., Apweiler, R., 1996. The SWISS-PROT protein sequence data bank and its new supplement TREMBL. Nucleic Acids Res. 24 (1), 21–25.

Bhattacharya, A., Ziebarth, J.D., Cui, Y., 2013. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. Nucleic Acids Res. gkt1028.

Calabrese, R., et al., 2009. Functional annotations improve the predictive score of human disease-related mutations in proteins. Hum. Mutat. 30 (8), 1237–1244.

Capiroti, E., Calabrese, R., Casadio, R., 2006. Predicting the insurmount of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics 22 (22), 2729–2734.

Capiroti, E., et al., 2008. A three-state prediction of single point mutations on protein stability changes. BMC Bioinf. 9 (Suppl. 2), S6.

Chen, X., Sullivan, P., 2003. Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. Pharmacogenomics J. 3 (2), 77–96.

Choi, Y., et al., 2012. Predicting the functional effect of amino acid substitutions and indels. PLoS One 7 (10), e46688.

Consortium, G.P., 2010. A map of human genome variation from population-scale sequencing. Nature 467 (7319), 1061–1068.

Davenport, J.W., et al., 1999. The mouse mitotic checkpoint gene bub1b, a novel bub1 regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res. 27 (12), 4900.

Drazen, J.M., et al., 1999. Pharmacogenetic association between ALOX5 promoter genotype and the response to anti-asthma treatment. Nat. Genet. 22 (2), 168–170.

Grillo, G., et al., 2010. UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. Nucleic Acids Res. 38 (Suppl. 1), D75–D80.

Hanks, S., Snape, K., Rahman, N., 2012. BUB1B (Budding Uninhibited by Benzimidazoles 1 Homolog Beta (Yeast)).

Hanks, S., et al., 2004. Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. Nat. Genet. 36 (11), 1159–1161.

Hanks, S., et al., 2006. Comparative genomic hybridization and BUB1B mutation analyses in childhood cancers associated with mosaic variegated aneuploidy syndrome. Cancer Lett. 239 (2), 234–238.

Hickey, S., et al., 2011. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. Hum. Mutat. 32 (6), 661–668.

Hussain, M.R.M., et al., 2012. In silico analysis of single nucleotide polymorphisms (SNPs) in human Braf gene. Gene 508 (2), 188–196.

Ljung, J.I., et al., 2009. STRING 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37 (Suppl. 1), D412–D416.

Kaisari, S., et al., 2016. Intermediates in the assembly of mitotic checkpoint complexes and their role in the regulation of the anaphase-promoting complex. Proc. Natl. Acad. Sci. 201524551.

Kapanidou, M., Lee, S., Bolanos-Garcia, V.M., 2015. BubR1 kinase: protection against aneuploidy and premature aging. Trends Mol. Med.