



OPEN

# Investigating the pathogenic SNPs in BLM helicase and their biological consequences by computational approach

Faisal A. Alzahrani<sup>1,2,15</sup>, Firoz Ahmed<sup>3,4,15</sup>✉, Monika Sharma<sup>5,15</sup>, Mohd Rehan<sup>6,7</sup>, Maryam Mahfuz<sup>8</sup>, Mohammed N. Baeshen<sup>9</sup>, Yousef Hawsawi<sup>10</sup>, Ahmed Almatrafi<sup>11</sup>, Suliman Abdallah Alsagaby<sup>12</sup>, Mohammad Azhar Kamal<sup>3,4</sup>, Mohiuddin Khan Warsi<sup>3,4</sup>, Hani Choudhry<sup>13</sup> & Mohammad Sarwar Jamal<sup>6,7,14</sup>✉

The BLM helicase protein plays a vital role in DNA replication and the maintenance of genomic integrity. Variation in the BLM helicase gene resulted in defects in the DNA repair mechanism and was reported to be associated with Bloom syndrome (BS) and cancer. Despite extensive investigation of helicase proteins in humans, no attempt has previously been made to comprehensively analyse the single nucleotide polymorphism (SNPs) of the BLM gene. In this study, a comprehensive analysis of SNPs on the BLM gene was performed to identify, characterize and validate the pathogenic SNPs using computational approaches. We obtained SNP data from the dbSNP database version 150 and mapped these data to the genomic coordinates of the “NM\_000057.3” transcript expressing BLM helicase (P54132). There were 607 SNPs mapped to missense, 29 SNPs mapped to nonsense, and 19 SNPs mapped to 3'-UTR regions. Initially, we used many consensus tools of *SIFT*, *PROVEAN*, *Condel*, and *PolyPhen-2*, which together increased the accuracy of prediction and identified 18 highly pathogenic non-synonymous SNPs (nsSNPs) out of 607 SNPs. Subsequently, these 18 high-confidence pathogenic nsSNPs were analysed for BLM protein stability, structure–function relationships and disease associations using various bioinformatics tools. These 18 mutants of the BLM protein along with the native protein were further investigated using molecular dynamics simulations to examine the structural consequences of the mutations, which might reveal their malfunction and contribution to disease. In addition, 28 SNPs were predicted as “stop gained” nonsense SNPs and one SNP was predicted as “start lost”. Two SNPs in the 3'UTR were found to abolish miRNA binding and thus may enhance the expression of BLM. Interestingly, we found that BLM mRNA overexpression is associated with different types of cancers. Further investigation showed that the dysregulation of BLM is

<sup>1</sup>Department of Biochemistry, Faculty of Science, Stem Cells Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah 21589, Saudi Arabia. <sup>2</sup>Aston Medical Research Institute, Aston Medical School, Aston University, Birmingham B4 7ET, UK. <sup>3</sup>Department of Biochemistry, College of Science, University of Jeddah, Jeddah 21589, Saudi Arabia. <sup>4</sup>University of Jeddah Centre for Scientific and Medical Research (UJ-CSMR), University of Jeddah, Jeddah 21589, Saudi Arabia. <sup>5</sup>Department of Chemical Sciences, Indian Institute of Science Education and Research (IISER), Mohali, India. <sup>6</sup>King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>7</sup>Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>8</sup>Department of Computer Science, Jamia Millia Islamia, New Delhi, Delhi, India. <sup>9</sup>Department of Biology, College of Science, University of Jeddah, Jeddah 21589, Saudi Arabia. <sup>10</sup>Department of Genetics, Research Center, King Faisal Specialist Hospital, and Research Center, MBC-03, PO Box 3354, Riyadh 11211, Saudi Arabia. <sup>11</sup>Department of Biology, Faculty of Science, University of Taibah, Medinah, Saudi Arabia. <sup>12</sup>Department of Medical Laboratories, Central Biosciences Research Laboratories, College of Science in Al Zulfi, Majmaah University, Al Majma'ah, Saudi Arabia. <sup>13</sup>Department of Biochemistry, Cancer Metabolism and Epigenetic Unit, Faculty of Science, Cancer and Mutagenesis Unit, King Fahd Center for Medical Research, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>14</sup>Integrative Biosciences Center, Wayne State University, Detroit, MI 48202, USA. <sup>15</sup>These authors contributed equally: Faisal A. Alzahrani, Firoz Ahmed, and Monika Sharma. ✉email: fahmed1@uj.edu.sa; Sarwar4u@gmail.com

associated with poor overall survival (OS) for lung and gastric cancer patients and hence led to the conclusion that BLM has the potential to be used as an important prognostic marker for the detection of lung and gastric cancer.

The BLM gene encodes an important nuclear protein, BLM helicase, which is involved in DNA replication and the maintenance of genomic integrity. BLM is a 3' to 5' DNA helicase that belongs to the evolutionarily conserved RecQ helicase family. Most mammals have five RecQ helicases (RECQL1, BLM, WRN, RECQL4, and RECQL5). Helicases are crucial for unwinding duplex DNA to produce the transient single-stranded DNA (ssDNA) intermediates necessary for replication, recombination, and repair<sup>1-3</sup>. In a complex with topoisomerase Topo IIIa and Rmi1/Rmi2, BLM helicase repairs, double-strand DNA breaks (DSBs) through a homologous recombination (HR) pathway<sup>4</sup>. Consequently, cells lacking functional BLM show an about tenfold increase in chromatid breaks, mitotic recombination, and sister chromatid crossover formation<sup>5</sup>. Bloom syndrome (BS) is a rare autosomal recessive genetic disorder caused by pathogenic variants in the BLM gene. BS belongs to OMIM entry 210900, which is characterized by genome instability that includes increased crossovers between homologous chromosomes<sup>6</sup>. The BLM gene is transcribed to a 97.93 kb-long precursor-mRNA with 21 exons, which encode a 1,417 amino acid protein. The literature shows that a large number of BS patients show insertion, deletion and missense mutations that change the amino acid sequence or nonsense mutations that introduce a premature stop codon in the BLM gene and thus inactivate the BLM helicase<sup>7-9</sup>. Symptoms of BS include low birth weight, dolichocephaly (long, narrow head), congenital short stature, growth retardation, sun-sensitive facial rash, an elevated risk of diabetes mellitus, reduced fertility and immune deficiency<sup>6,10-12</sup>. The absence of BLM protein activity causes a defect in DNA repair with a consequent increased rate of mutations and thus poses an elevated risk of cancer<sup>12-15</sup>. The average life span of BS patients is approximately 27 years, with the most common cause of death being cancer (<https://weill.cornell.edu/bsr/>).

Single nucleotide polymorphisms (SNPs) are a common genetic variation contributing greatly to phenotypic variation in the general population<sup>6</sup>. SNPs can alter the functional consequences of proteins. In the coding region of genes, SNPs may be synonymous, non-synonymous (nsSNPs) or nonsense<sup>16</sup>. Synonymous SNPs change the nucleotide base residue but do not change the amino acid residue in the protein sequence due to the degeneracy of the genetic code. The nsSNPs, also called missense variants, alter amino acid residues in protein sequences and thus change the function of proteins through altering protein activity, solubility and protein structure. Nonsense SNPs introduce premature termination in the protein sequence.

The non-coding region of the gene contains several regulatory *cis*-elements, such as miRNA binding sites, that can also affect the regulation of gene expression<sup>17-20</sup>. SNPs have emerged as genetic markers for diseases, and there are many SNP markers available in public databases. Previous reports have shown the value of defining mutations as deleterious or non-deleterious and their connection with certain diseases, thus identifying pathogenic SNPs that are functionally compromised due to structure-damaging properties<sup>21-27</sup>.

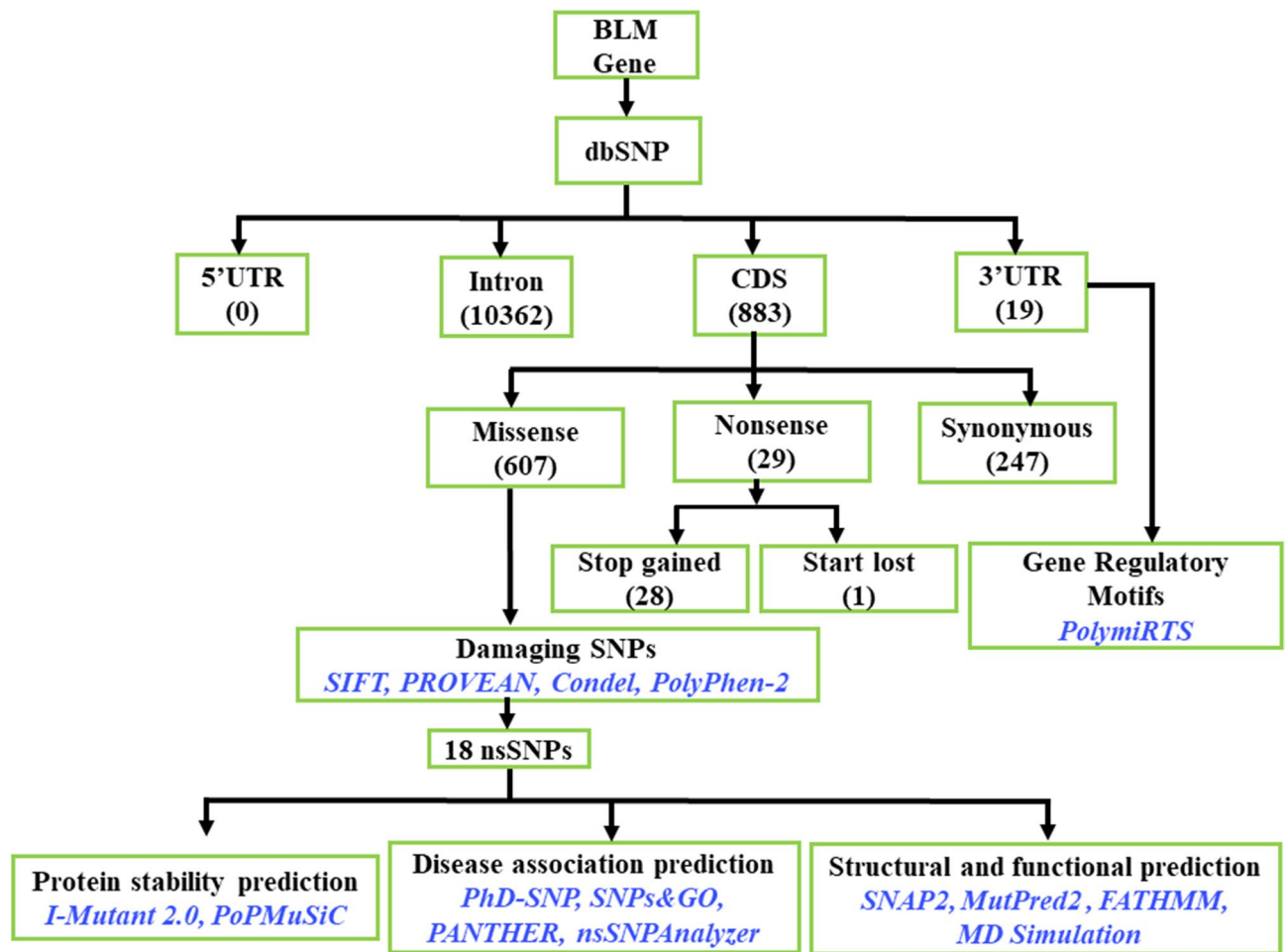
In recent decades, the computational approach has become established as an effective method that streamlines time consuming, laborious experimentation and allows researchers to shortlist the most critical or pathological SNPs. This ability leads us to focus on selectively targeted SNPs instead of scanning full genes for the identification of pathological SNPs by experimental mutational analysis. Computational studies have also evaluated and analysed genetic mutations for their pathological effects and are effective in establishing the underlying molecular mechanism<sup>26-34</sup>.

With recent advances in high-throughput sequencing technology, hundreds of new SNPs have been mapped to human BLM genes. However, not all SNPs are functionally important. Despite extensive studies of helicase proteins in humans and the effect of their polymorphisms in cancer<sup>11</sup>, no attempt was previously made to comprehensively and systematically analyse and establish the functional consequences of SNPs of the BLM gene. The aim of this study was to identify the high-confidence pathogenic SNPs of the BLM gene and to determine their structural, molecular and functional consequences using computational approaches. This work may be useful in the development of precision medicine-based treatment for diseases caused by these genomic variations. In the future, this report can be used for biomarker discovery by establishing the importance of SNPs for the diagnosis of Bloom syndrome, as well as cancer, and will further aid in targeted therapies.

## Materials and methods

Application of computational methodological approaches for acquiring biological insight is well established<sup>35-41</sup>. Earlier studies support the notion that the application of various powerful tools and algorithms leads to increased prediction accuracy<sup>42-45</sup>. To ensure that the results are of the highest accuracy, we utilized several computational algorithms that can be used for the prediction of nsSNPs of BLM helicases because of their disease-related properties. For this purpose, we used tools such as SIFT<sup>15</sup>, PROVEAN<sup>46</sup>, Polyphen-2.0<sup>47</sup>, SNAP2-2<sup>48</sup>, Muscle<sup>49</sup>, Weblogo<sup>50,51</sup>, SNP&GO<sup>52</sup>, nsSNPAnalyzer<sup>53</sup>, and Mutpred-2<sup>54</sup>. This approach enabled greater accuracy for the prediction of most disease-associated mutations in BLM genes and their structural consequences.

**SNP dataset.** The SNPs of the BLM helicases were retrieved from the dbSNP database build 150 and mapped on genome assembly GRCh38 using Variation Viewer<sup>12</sup>. We used “BLM” as our search term and filtered for SNPs (<https://www.ncbi.nlm.nih.gov/variation/view/?q=BLM>). Furthermore, we mapped these SNPs to the genomic coordinates of the “NM\_000057.3” transcript expressing the BLM helicase (P54132) for computational analysis of the effect of missense variants and nonsense SNPs as well as SNPs in the UTR region. The protein sequences of the genes for BLM (P54132) were retrieved from the UniProt database (<https://www.UniProt.org>).



**Figure 1.** Flow chart of the in silico analysis of pathogenic SNPs in the BLM gene and their biological consequences.

We employed various sets of computational tools that together increased the accuracy and reliability of the identification of pathogenic SNPs and their effects on the structural and functional consequences of BLM (Fig. 1).

**Tools used to predict the SNP effects.** *Predicting deleterious and damaging nsSNPs.* SIFT. An algorithm that predicts tolerant and intolerant coding base substitutions based upon amino acid properties and sequence homology<sup>15</sup>. The tool considered the vital positions in the protein sequence that have been conserved throughout evolution, and therefore, substitutions at conserved alignment positions are expected to be less tolerated and to affect protein function more than those at diverse positions. We used SIFT version 2.0 (<https://sift.jcvi.org/>), which predicted an amino acid substitution score from zero to one. SIFT predicted substituted amino acids to be damaging at a default threshold score  $< 0.05$ , while a score  $\geq 0.05$  was predicted to be tolerated.

PROVEAN. An online tool (<https://provean.jcvi.org/>) that uses an alignment-based scoring method for predicting the functional consequences of single and multiple amino acid substitutions and in-frame deletions and insertions<sup>46</sup>. The tool has a default threshold score, i.e., -2.5. If a protein variant is below the threshold, it is predicted as deleterious; above that threshold, a protein variant is considered neutral.

Condel. A tool (<https://bbglab.irbbarcelona.org/fannsd/>) that predicts the consequences of non-synonymous SNPs as neutral or deleterious<sup>55</sup>. It uses a consensus deleteriousness (Condel) score calculated by integrating the normalized score of five predictive tools: SIF<sup>15</sup>, PolyPhen-2<sup>47</sup>, Logre<sup>56</sup>, MAPP<sup>57</sup>, and MutationAssessor<sup>58</sup>. The Condel score could vary between 0 to 1, where a higher score indicates SNPs as deleterious.

PolyPhen-2. A tool that predicts the structural and functional consequences of a particular amino acid substitution in a human protein<sup>47</sup>. The prediction of the PolyPhen-2 server (<https://genetics.bwh.harvard.edu/pph2/>) is based on a number of features, including structural and sequence comparison information. Its score varies between 0.0 (benign) and 10.0 (damaging). The PolyPhen-2 prediction output categorizes SNPs into three

basic categories: benign (score < 0.2), possibly damaging (score between 0.2 and 0.96), and probably damaging (score > 0.96).

**BLM helicase sequence retrieval and SNP position sequence logo.** BLM helicase sequences were retrieved from the UniProt protein sequence database using a search option with the key words 'BLM helicase'. The sequences containing the key words 'Bloom' and/or 'Blm' with the names were considered, and then the fragments and irrelevant sequences were removed. Finally, 153 sequences were obtained and considered for multiple sequence alignment. The multiple sequence alignment was performed by Muscle v.3.8.31<sup>49</sup>, and all the respective sites for human BLM helicase SNPs were extracted from the alignment using an in-house Perl script. The sequence logo for the SNP positions was prepared using Weblogo v.2.8.2<sup>50,51</sup>. A sequence logo is a graphical representation of a multiple sequence alignment, whereby a stack of amino acid symbols corresponds to a column position in the alignment. Within a stack, the varying heights of amino acid symbols show the relative frequency of each amino acid at that position.

**Predicting disease-associated nsSNPs.** SNPs&GO. A webserver that predicts whether an amino acid substitution is associated with a disease or not (<https://snps.biofold.org/snps-and-go>)<sup>52</sup>. It is a support vector machine (SVM)-based tool that considers protein sequence features, evolutionary information, and functional annotation according to Gene Ontology terms. We input the Swiss-Prot Code of BLM helicase (P54132) and provided the list of amino acid mutations. The results predicted whether helicase polymorphisms would be disease associated or not by three methods: (a) SNPs&GO, (b) PhD-SNP, and (c) PANTHER. A probability score > 0.5 is predicted as a disease-associated variation.

**nsSNPAnalyzer.** A random forest classifier developed using curated SNP datasets from SwisProt to predict the phenotypic effects of nsSNPs (<https://snpanalyser.uthsc.edu/>)<sup>53</sup>. It can predict (a) the structural environment of SNPs; (b) the normalized probability of substitution in the multiple sequence alignment; and (c) the similarity and dissimilarity between the variant and the original amino acid.

**Predicting the molecular and phenotypic impact of nsSNPs.** SNAP2. A tool used to predict the functional consequences of non-synonymous SNPs using a neural network (<https://roslab.org/services/snap/>)<sup>48</sup>. SNAP2 incorporates various features, including evolutionary similarity from multiple sequence alignments and secondary structure and solvent accessibility, to predict whether a substitution is likely to alter the protein effect. It predicts a score from -100, considered strongly neutral, to + 100, considered as a strong effect. A threshold score > 0 is considered an effect. We input the protein sequences of BLM helicases and obtained the score of SNPs.

**MutPred2.** A neural network-based method to predict the molecular and phenotypic impact of amino acid variants as pathogenic or benign in humans (<https://mutpred.mutdb.org/>)<sup>54</sup>. It is programmed on 53,180 pathogenic and 206,946 putatively benign amino acid substitutions from HGMD, Swiss-Prot, dbSNP, and orthologous alignments. It also incorporated the impact of amino acid substitutions on over 50 different local structural and functional protein properties and thus helped to infer the molecular mechanisms of pathogenicity. The outcome of MutPred2 includes the following: (A) a general pathogenicity score (g), which is the likelihood that a substituted amino acid is pathogenic; (B) predicted molecular mechanism; (C) property score (pr) of molecular mechanism and its P-value (P); and (D) affected PROSITE and ELM Motifs. General scores vary from 0.0 (benign) to 1.0 (pathogenic), where with  $g \geq 0.5$  would suggest pathogenicity; however,  $g \geq 0.68$  yields a false positive rate (fpr) of 10%, whereas  $g \geq 0.80$  yields an fpr of 5%. The higher the property score (Pr), the more likely that the molecular mechanism of the disease involves the alteration of the property.

**Structural characterization of predicted nsSNPs.** *Modelling of mutant BLM helicase proteins.* The coordinates of the native human BLM helicase were retrieved from the PDB database with PDB id '4O3M'<sup>59,60</sup>, referred to in the article as the wild type (WT). The WT structure consists of the BLM protein (640–1,290) bound to 1 ADP molecule, 1 Ca<sup>2+</sup> ion, 1 Zn<sup>2+</sup> ion, and a 3'-overhang DNA duplex. A few residues were missing in the crystal structure: 799–807, 1,011–1,013, 1,069–1,071, 1,093–1,104, 1,195–1,206, and 1,292–1,298. We modelled these residues using Modeller9v15<sup>61</sup>. The modelled structure was minimized and refined using simulations. Corresponding point mutations were introduced in the obtained structure using Modeller9v15<sup>61</sup> to generate mutant structures. For the simulations, the protein and nucleic acids were represented by Amber forcefields: ff14SB<sup>62</sup> and DNA (OL15)<sup>63</sup>, respectively. For bound ADP and two ions, GAFF charges provided with the AMBER forcefields were used. The modelled systems were solvated in a cubic box with TIP3P potential at a 10 Å marginal radius and neutralized with Na<sup>+</sup> or Cl<sup>-</sup> ions. Long-range interactions were modelled using particle mesh Ewald (PME) with a tolerance of 1e-05 and a grid spacing of 1.2 Å. A non-bonded cut-off of 12 Å was applied to the Lennard–Jones potential, and the cut-off for the direct-space part of the Coulomb forces was a switching function starting at a distance of 10 Å and reaching zero at 12 Å. A 2 fs time step was used for all simulations while constraining bonds involving hydrogens with the Lincs algorithm. Periodic boundary conditions were employed to eliminate surface effects. To reach the desired temperature and pressure, minimized systems were simulated for 5 ns while coupled to a *v-rescale* thermostat and Parinello–Rahman barostat implemented in Gromacs2016.5<sup>64</sup>. The equilibrated structures of the WT and 18 mutants were simulated further for 20 ns each.

**Trajectory and structural analyses.** Structural analyses were performed using built-in programs of Gromacs2016.5 and VMD-1.9.3<sup>65</sup>. Root mean square deviations (rmsd) and radius of gyration values for all



backbone atoms were calculated with respect to the crystal structure. Root mean square fluctuations (rmsf) of Ca atoms in all trajectories were also calculated. The changes in the secondary structure of proteins during the course of simulation were analysed using the DSSP program as implemented in the Bio3D package written in R<sup>66–68</sup>. Domain-wise rmsd values were calculated for each domain after aligning the rest of the protein.

**Assessment of the effect of BLM genes on survival by Kaplan–Meier plots.** Kaplan–Meier plots were analysed using online KM plotter software (<https://kmplot.com/analysis/>)<sup>69</sup>. The tool analyses the effect of 54,675 genes on the survival outcome of patients using 10,293 cancer samples from the Affymetrix microarray data in the Gene Expression Omnibus (GEO: <https://www.ncbi.nlm.nih.gov/geo/>), the European Genome-phenome Archive (EGA: <https://ega.crg.eu/>) and The Cancer Genome Atlas (TCGA: <https://cancergenome.nih.gov/>) databases. We analysed the potential effect of BLM gene expression on overall survival in a large number of cancer patients, including patients with gastric (1,065), ovarian (1816), lung (2,437), and breast (5,143) cancer. The hazard ratio (HR) with 95% confidence intervals and log rank *P*-value (below 0.05 were considered significant) were calculated. Biased arrays were excluded for quality control.

*The GEPIA.* A web-based tool for analysing the differential expression of mRNA in tumour and normal cells (<https://gepia.cancer-pku.cn/index.html>)<sup>70</sup>. The tool uses expression data from 8,587 normal and 9,736 tumour samples obtained from the TCGA and the Genotype-Tissue Expression (GTEx) projects. The mRNA expression on box plot was calculated using “Expression DIY” module using parameters: Gene “BLM”; |Log2FC| cut-off “1.5”; *p*-value cut-off “0.01”; and match TCGA normal and GTEx data.

*PolymiRTS 3.* A database of variants in miRNA and miRNA target sites (<https://compbio.uthsc.edu/miRSNP/>)<sup>71</sup>. It was used to examine the impact of SNPs on miRNA-target (BLM) binding and on BLM gene expression.

## Results and discussion

The application of computational biology in genome research is well established<sup>18,72</sup> and is frequently used to filter for the most potential deleterious nsSNPs in target genes to understand the aetiology of various diseases<sup>28,73–75</sup>. Computational analysis can provide molecular insight into the changes in protein structure and function due to point mutations in the target genes. In the current study, we report nsSNPs in BLM genes, which are most likely associated with pathogenic conditions involving BLM and other BS-associated diseases, such as cancers. Our findings suggest that the application of multiple powerful algorithms to SNP datasets identifies the most deleterious SNPs, which might be associated with diseases. The accuracy of nsSNP prediction was also supported by various reports discussed below.

For this study, we retrieved 11,983 rsIDs of SNPs mapped in the human BLM gene from dbSNP (Table S1). However, these rsIDs fall in different molecular consequence classes. For instance, some rsIDs are associated with multiple SNPs and therefore belong to different classes. We mapped these SNPs to the genomic coordinates of the “NM\_000057.3” transcript expressing BLM helicase (P54132). There were 883 SNPs (825 #rsID) mapped to the CDS region, 19 SNPs (17 #rsID) to the 3’UTR, and 10,362 SNPs (9,956 #rsID) mapped to introns. We did not find any SNPs in the 5’UTR of the NM\_000057.3 transcript. In the CDS region, 608 SNPs (570 #rsID) were missense mutations, 28 SNPs (28 #rsID) were nonsense, and 247 SNPs (233 #rsID) belonged to the synonymous group (Table S2). For the study of the functional consequences of nsSNPs in BLM genes, we selected the “missense” and “nonsense” variant map to reference transcript “NM\_000057.3”. In addition, we analysed the effects of SNPs in the 3’UTR of the “NM\_000057.3” transcript on miRNA binding and poly (A) signals. The selection gave us a final dataset of 636 nsSNPs with 597 rsIDs and 19 SNPs (17 #rsID) mapped to the 3’UTR, which were then used for further analysis.

**Predicting deleterious and damaging nsSNPs.** To predict the damaging nsSNPs, we employed multiple consensus tools. Initially, we used the online tool VEP (<https://www.ensembl.org/Tools/VEP>). VEP has advantages, such as using the latest human genome assembly, GRCh38.p10, and predicting thousands of SNPs from multiple tools, including *SIFT*, *PROVEAN*, *Condel*, and *PolyPhen-2*, to retrieve accurate results. We uploaded the 597 nsSNP accession numbers to the VEP tool, and the prediction results were used for further analysis.

We found 28 nsSNPs predicted as “stop gained” and one nsSNP as a “start lost” mutation (Table S3). The remaining nsSNPs were filtered on default scores of consensus tools based on sequence and structure homology methods: (a) *SIFT* (score < 0.5) and (b) *PROVEAN* (score < -2.5) and *Condel* (score > 0.522), which showed 136 nsSNPs as damaging (Table S4). To obtain very high-confidence nsSNPs that impact the structure and function of the BLM helicase, we considered highly stringent scores across different consensus tools at parameters of *SIFT* (score = 0), *PROVEAN* (score < -8.0) and *Condel* (score > 0.9), we obtained 18 nsSNPs (Table 1). These 18 nsSNPs were further analysed by *PolyPhen-2*, which gave a score greater than 0.96 for all 18 nsSNPs and therefore placed them in the predicted category of probably damaging. Additional analysis of these 18 nsSNPs with *SNAP2* showed that these amino acid substitutions have a damaging effect on protein structure with high accuracy (Detail information: Table S5).

The sequence logo for nsSNP positions in BLM helicase sequences show that these positions are highly conserved among various species (Fig. 2). The reported 18 nsSNP variants were looked into the sequence logo, and three of them, W803R, H805L, and G891V, were found to naturally occur in other sequences; however, their frequency was very low. The two positions G952 and G978 were found to be 100% conserved, as shown in the sequence logo. The complete sequence alignment of BLM helicases is also provided in Supplementary\_file1.

S.N	#rsID	VSc	Amino acid	SIFT	PROVEAN	Condel
1	rs761589072	C>T	Pro690Leu	0	- 9.32	0.935
2	rs770625327	C>T	Pro702Leu	0	- 9.24	0.945
3	rs148394770	T>C	Trp803Arg	0	- 13.05	0.945
4	rs761938011	G>T	Trp803Leu	0	- 11.92	0.935
5	rs766292814	A>T	His805Leu	0	- 10.26	0.935
6	rs145029382	A>G	Tyr811Cys	0	- 8.26	0.945
7	rs749632465	C>T	Pro825Leu	0	- 8.37	0.902
8	rs763471784	G>T	Gly891Val	0	- 8.8	0.906
9	rs367543034	G>T	Gly952Val	0	- 8.34	0.935
10	rs150475674	G>T	Gly972Val	0	- 8.15	0.911
11	rs1051102270	A>G	Tyr974Cys	0	- 8.35	0.945
12	rs750210123	G>T	Gly978Val	0	- 8.55	0.945
13	rs137853153	G>T	Cys1036Phe	0	- 10.13	0.945
14	rs747571272	A>G	Tyr1044Cys	0	- 8.29	0.902
15	rs746218707	T>C	Cys1055Arg	0	- 11.05	0.935
16	rs367543029	G>A	Cys1055Tyr	0	- 10.13	0.945
17	rs367543032	A>T	Asp1064Val	0	- 8.29	0.945
18	rs367543025	G>A	Cys1066Tyr	0	- 8.63	0.935

**Table 1.** List of deleterious missense SNPs in the BLM gene using consensus bioinformatics tools. These SNPs are selected based upon the parameters *SIFT* (score = 0), *PROVEAN* (score < -8.0) and *Condel* (score > 0.9).

Afterwards, these 18 high-confidence nsSNPs were further assessed for functional and structural consequences using different bioinformatics tools.

**Predicting the effect of 18 deleterious nsSNPs.** *Effect of nsSNPs on protein stability.* Studies have shown that most disease-associated missense mutations change the stability of proteins<sup>76,77</sup>. Therefore, we analysed these nsSNPs in terms of amino acid substitutions and their effect on the stability of mutant BLM proteins by IMutant2 (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>) and PoPMuSiC v3.1 (<https://soft.dezymy.com>) tools<sup>78,79</sup>.

Both IMutant2 and PoPMuSiC predict the effect of mutations on proteins based on protein 3D structures. Because the whole protein 3D structure of the BLM gene is not yet available, partial protein structures, such as PDB id 4O3M with a length of 640–1,290 amino acids, i.e., 613 amino acids long, were used for analysis. We submitted the 18 amino acid substitutions to IMutant2 and PoPMuSiC, which predicted the stability of the BLM protein variants. We found that three mutants (P690L, P702L, and P825L) and two mutants (P702L and G891V) were predicted to stabilize the mutant proteins by IMutant2 and PoPMuSiC, respectively (Table 2). However, twelve mutants (Y811C, G891V, G952V, G972V, Y974C, G978V, C1036F, Y1044C, C1055R, C1055Y, D1064V, and C1066Y) were predicted to destabilize the proteins by IMutant2; while thirteen mutants (P690L, Y811C, P825L, G952V, G972V, Y974C, G978V, C1036F, Y1044C, C1055R, C1055Y, D1064V, and C1066Y) were predicted to destabilize the proteins by PoPMuSiC. We found 12 mutants out of 15 predicted results showed consensus effect on the protein stability between IMutant2 and PoPMuSiC; while three mutants (W803R, W803L, H805L) were not predicted by both tools because the amino acid residues from 799 to 807 are missing in the BLM helicase crystal structure (PDB id: 4O3M) (Table 2).

*Identifying disease-associated nsSNPs.* Eighteen selected amino acid substitutions in the BLM protein were used to analyse disease association. BLM protein ID “P54132” and its amino acid mutations were submitted to the “SNPs&GO” tool (<https://snps.biofold.org/snps-and-go/snps-and-go.html>). The predicted disease associations from three different tools were analysed. The output of (a) SNPs&GO and (b) PhD-SNP predicted that all the tested SNPs were associated with diseases, while (c) PANTHER predicted 16 SNPs as disease associated and two as neutral (Table S6). In addition, these nsSNPs were also analysed with the nsSNPAnalyzer tool to predict disease association. This tool also provides supplementary information about nsSNPs, such as secondary structure, structural environment, area buried and fraction polar. We uploaded the BLM protein sequence, list of 18 amino acid substitutions, and PDB structure (4O3M) to the nsSNPAnalyzer. The output of nsSNPAnalyzer reported 17 nsSNPs as disease associated and one nsSNP reported as neutral (Table 3).

Eighteen amino acid substitutions were further analysed with the Fathmm server (<https://fathmm.biocompute.org.uk/>) at the default threshold for cancer-promoting mutations and disease-causing mutations<sup>80</sup>. The output results of these tools showed that two (G952V and G978V) were associated with cancer (Table 4), while three (G952V, P702L and G978V) were also associated with cancer (Table 5).

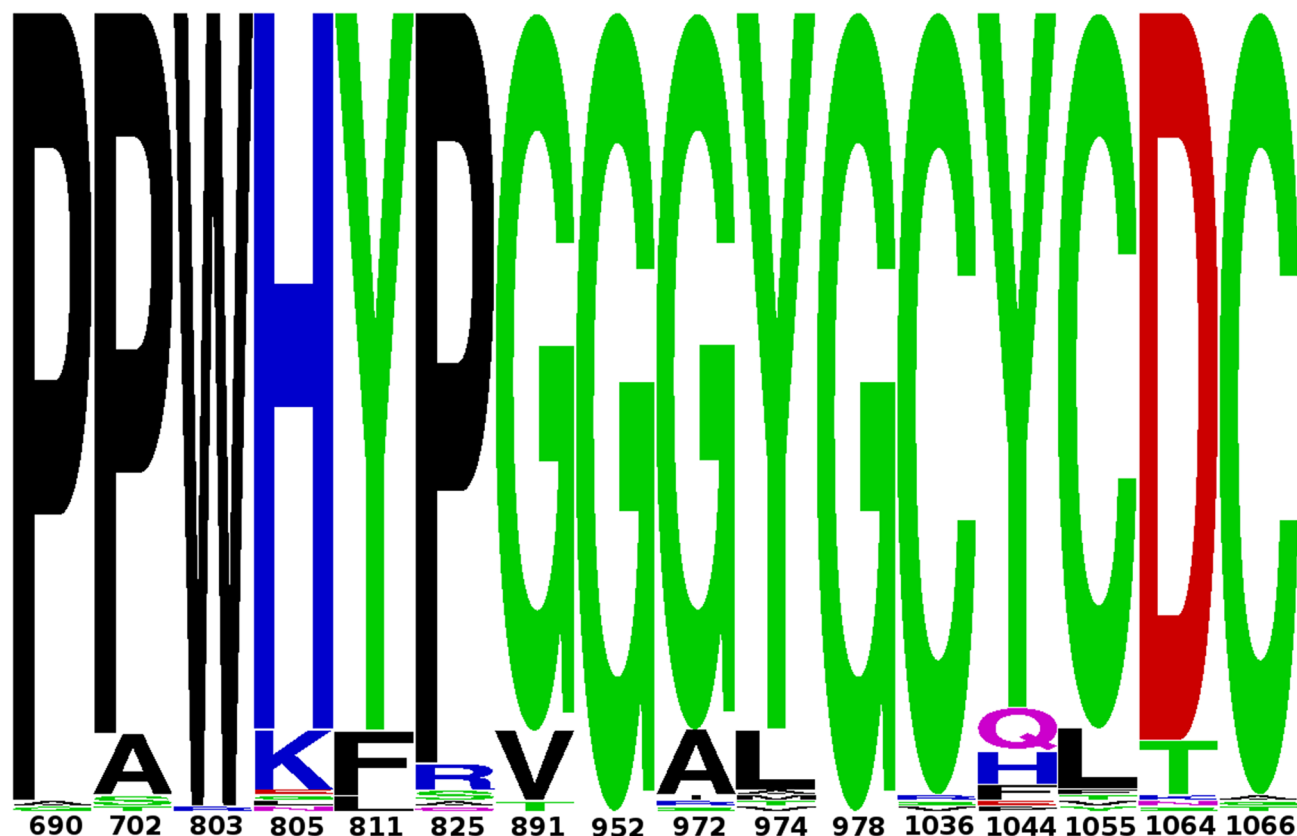
*Predicting protein structural and functional consequences.* MutPred2 was used to infer the structural, molecular and phenotypic impacts of amino acid variants. The results for MutPred2 are shown in Table S7. There are 18 nsSNPs reported with high confidence to affect the structure and function of BLM proteins.

*Substitution effect on helicase structure of BLM via molecular dynamics simulations.* Bloom syndrome proteins comprise three domains: ATPase domain (642–1,068), RecQ family-specific C-terminal (RQC) domain (1,074–1,194), and Helicase and RNase D C-terminal (HRDC) domain (1,208–1,290). The C-terminus of the ATPase domain, known as the Zn subdomain, consists of Zn-binding residues (994–1,068). One ADP molecule is observed to be bound to the inter-subdomain cleft between two ATPase subdomains, 1A (642–857) and 2A (858–1,068). In the BLM WT crystal structure, the duplex region of DNA is bound to the RQC domain surface via an unconventional winged-helix domain. This interaction differs from that known in conventional winged-helix domains, where the minor and major grooves of dsDNA are recognized by one recognition helix and  $\beta$  wing<sup>81</sup>. In the BLM RQC domain, dsDNA does not form direct interactions with the recognition helix. Instead, a loop connecting two helices serves as the prominent DNA-interacting entity of the RQC domain. The terminal region of the dsDNA duplex interacts with the  $\beta$  wing of the RQC domain, which seems to act as a scalpel for splitting the dsDNA duplex. The third domain of BLM, the HRDC domain, folds as a helical bundle of five  $\alpha$ -helices and one  $3_{10}$  helix connected by a short loop. However, much remains unknown regarding the functional aspects of the HRDC domain and how it occurs in two of the five human RecQ-family helicases (BLM and WRN), as well as in RecQ helicases in bacteria and yeast. BLM mutants lacking the HRDC domain possess core helicase and ATPase activities similar to those of the wild-type protein<sup>82–84</sup> but are defective in both strand annealing<sup>74</sup> and double Holliday junction dissolution<sup>83</sup>.

We evaluated the 18 predicted deleterious amino acid substitutions for possible damage to the 3D structure of the BLM helicase using molecular dynamics simulations. When these SNP positions were mapped onto the structure of the BLM helicase (Fig. 3), we found that the SNP positions were distributed in the helicases and coil structures. Three SNP positions, C1036, C1055, and C1066 (interestingly all cysteines), were in the Zn subdomain, directly interacting with the Zn atom. The SNP position, D1064, was also found close to the Zn atom. We then compared the deviations by rmsd (Fig. 4) and rmsf values (Fig. 5) observed during simulations within mutant proteins with those observed for native (or WT) proteins and classified the mutants into two categories: (a) mutants with considerably higher rmsd values, suggestive of instability induced in the protein; and (b) mutants with similar rmsd values, and thus, possessing significantly more or equally stable conformations compared to those of the WT. Mutants G891V, G952V, G972V, G978V, C1036F, and C1055R destabilize the BLM structure and belong to category (a), Mutants P690L, P702L, W803R, W803L, H805L, Y811C, P825L, Y974C, Y1044C, C1055Y, D1064V, and C1066Y belong to category (b). Interestingly, the mutation of C1055 seems to be residue specific. If C1055 is mutated to arginine, it destabilizes the complete BLM structure, whereas mutation to tyrosine is tolerated. The mutant C1055Y shows similar deviations from the WT. Below, we will describe the effects of each mutant in detail. For visual inspection, we compared the final structure obtained after 20 ns of sampling of each mutant with the final structure obtained for the WT or native state (Figs. S1, S2). We calculated rmsd values for each domain for these mutants (Figs. 6, S3). We further analysed the variation in secondary structure for mutants and WT with respect to time (Fig. S4).

**Mutations in ATPase domain 1A.** Mutations P690L and P702L are present in ATPase domain 1A, near the ADP binding pocket. Our computational results suggest that their properties are similar to those of the WT. However, Mirzaei et al., reported six mutant residues (P690L, R717T, W803R, Y811C, F857L, G972V) that cause loss of function for the BLM protein<sup>85</sup>. Modelling studies of the BLM residues confirm that P690 along with R717, W803, and Y811 are located in the first lobes of the helicase domain. Among the point mutations of BLM, P690L does not affect the extremely conserved GK(T/S) residues of motif I (Walker A). Another report from Shashtri et al. showed, using a HU-based hypersensitivity-based study, that the P690L mutant of the BLM protein was unable to rescue the DNA damage response in the BS cell line compared to the WT<sup>5</sup>. The reason behind the loss of function could be due to leucine (hydrophobic) replacing proline (helix breaker), which may culminate in reduced flexibility and hydrophobicity, causing weakened dsDNA binding and altering the location of the dsDNA within the motif, which leads to failure to bind ATP and/or Mg<sup>2+</sup><sup>5,85</sup>.

Studies from Mirzaei et al. showed that the W803 mutation leads to impaired BLM function associated with Bloom syndrome. This mutation site is located in the highly conserved region called the aromatic-rich (AR) loop and is crucial for ATP binding, hydrolysis, DNA binding, ssDNA binding and unwinding<sup>86–89</sup>. Because it is situated as a key residue in a loop region, it plays a crucial role in various functions. Another set of mutations, W803L, W803R, H805L, Y811C, P825L in ATPase domain 1A show similar dynamics as WT. W803, a residue that is conserved in all five human RecQ homologs (BLM, WRN, RecQL1, RecQL4, and RecQL5), suggests that this residue plays a key role in helicase activity. Hence, the W803R mutation might hamper helicase activity, which is one of the major causes of BS development. Though our molecular dynamics simulation-based study suggests similar dynamics to those of the WT, other studies suggest that these residues are crucial for the above-mentioned activities. Hence, we also speculate that this mutation may lead to a deterioration in BLM activity, which may lead to the development of BS. The downstream Y811C and P825L mutations showed no structural differences as WT. Though there is no evidence for P825 or Y811, it is identified as a key residue for the functioning of BLM proteins<sup>85,91,93,96–98</sup>. Another report<sup>5</sup> suggested based on HU hypersensitivity that, similar to P602L, the Y811C mutant of the BLM protein was unable to rescue the DNA damage response in the BS cell line compared to the WT.



**Figure 2.** Sequence logo for 16 chosen SNP positions in BLM helicase sequence alignment. The height of the amino acid symbols shows the relative frequency of amino acids in the alignment. The amino acids are coloured according to their chemical properties. The polar amino acids (G, S, T, Y, C, Q, N) are green, basic (K, R, H) are blue, acidic (D, E) are red and hydrophobic (A, V, L, I, P, W, F, M) amino acids are black.

**Mutations in ATPase domain 1B.** Mutations in ATPase domain 1B, destabilizing the BLM structure, are observed for glycines (G891V, G952V, G972V, G978V). The structural influence of these mutants can be observed by larger variations in domain-wise rmsd plots (Fig. 6). A report from Rong et al.<sup>90</sup>, also endorsed our results, which suggests that the G891 mutation along with Q672R, I841T and C901Y inactivates the helicase domain<sup>7,8</sup>. There are also point mutations, such as K803A, Q680P and I849T, in mouse BLM<sup>7,8,90,91</sup>. All these point mutations hamper ATPase and DNA unwinding activity. G972V mutation, in the N terminal region of BLM, may hamper the replicative role of BLM as reported by Selak et al.<sup>92</sup>, a major issue in Bloom Syndrome<sup>93–97</sup>. 972 is proximal to the nucleic acid overhang. Mutating G972 to valine pushes the nearby K968, which is directly interacting with the nucleic acid overhang, and this further affected the downstream RCQ domain and to some extent to the Zn subdomain, as observed from the domain-wise rmsd values. To depict that the nsSNP leads to full inactivation of the BLM protein, Mirzaei et al.<sup>85</sup>, exploited a yeast model system by designing a chimera BLM gene consisting of the yeast and human *BLM* genes. The introduction of nsSNP into the chimera BLM leads to impaired BLM function, which was examined and confirmed by a hypersensitivity assay of cells to hydroxyurea (HU), a DNA-damaging agent<sup>85</sup>. The hypersensitivity assay showed that six-point mutations (P690L, R717T, W803R, Y811C, F857L, G972V) cause total loss-of-function of BLM, and three cause partial loss-of-function (R791C, P868L, G1120R). The location of G972 is in close proximity to the arginine finger; therefore, this mutation might impair coordination among the ATP binding of lobe 1 with lobe 2. G978V mutation resulted in a change in the backbone orientation of the ATPase domain 1B, induced structural deviations in this interface area, as well as in the Zn subdomain, and influenced the downstream HRDC domain (Fig. 6). Since there is no report available for this mutation, in future studies, it may be informative to sequence this pathogenic mutation site, especially in BS patients and in other related diseases, and/or study it using mutational approaches using various models. We speculate that this mutation may also be a hypomorphic mutation in which, instead of causing BS, it may be connected with other BS-associated diseases, such as cancer or type 2 diabetes<sup>5</sup>.

Another set of mutations, Y974C and Y1044C in ATPase 1B domain, showed no significant destabilizing structural effects during the sub-nano seconds simulations. Y974 is present at the interface, pointing towards the cleft between ATPase domains 1A and 1B and forming stacking interactions with F1045 of the nearby helix. In the native structure, Y1044 forms hydrogen-bonding interactions with E971 of the nearby helix, and this interaction is maintained throughout the simulations of the WT. Mutating tyrosine to cysteine disrupts these hydrogen-bonding interactions but results in another set of stable interactions with S1025 of a nearby helix, thus maintaining the structural dynamics of the BLM. Although our computational results suggest that Y1044C is similar to the WT, other studies suggest that Y1044C is present along with S897C in Japanese male patients



and white female patients with metachronous colon cancer in the C-terminal of BLM in CRC. This issue can be further explored in the future using BS cells and model organisms. Highly conserved BLM variants were reported to be found in the C-terminal helicase domain, which may lead to a predisposition to hereditary CRC<sup>98</sup>.

**Mutations in Zn subdomain.** In the Zn subdomain, highly conserved residues C1036, C1055, C1063, and C1066 form interactions with the Zn atom in the BLM protein. C1036F mutation influenced the RQC domain. In the native structure, C1055 interacts with the Zn atom, along with C1063. For the C1055R mutation, when the conformations of mutants after simulation are visualized with respect to the native structure, arginine being positively charged shows repulsion with positively charged Zn and destabilization of the domain (Figs. 5, 6). This destabilization disrupts the C-terminus of ATPase domain 1A (Figs. 4, 5, 6). The destabilizing effect is indicated by the large variations observed in the rmsd and rmsf plots. We further calculated the domain-wise rmsd values for C1055R and observed that the mutation resulted in larger deviations within the Zn subdomain and the following RQC domain. However, the tyrosine mutant of C1055 is oriented towards the solvent and forms favourable hydrogen bonds with N858 and stacking interactions with I1039 and F1050, thus leading to less structural disruption than the arginine mutation, which completely disrupts the helical structure (Fig. 7). Therefore, we observe higher rmsd variations for C1055R than for C1055Y. Nonetheless, the loss of Zn ion binding is observed for both mutations. C1055 may form disulfide bridges with one of the cysteines present in close proximity to stabilize the helix structure of the BLM protein. This model predicts that the residues in the helix are important for ssDNA binding, contributing to the formation of the BLM–DNA complex. Consistent with this study, Guo et al. used a gel-shift assay to show that similar mutant C901Y leads to a significant reduction in dsDNA binding and an even greater loss of ssDNA binding<sup>99</sup>. This lower binding of BLM with counterparts of the DNA region may lead to disruption in the functional outcome. This may result in a predisposition to many unwanted mutations and thus to Bloom syndrome. A comprehensive site-directed mutational analysis by Guo et al.<sup>100</sup>, suggests that there is a significant reduction in zinc binding ability, which further indicates that these highly conserved cysteine residues are essential for Zn<sup>2+</sup> ion binding and thus for BLM activity. The effect of cysteine residue mutation on DNA binding ability was shown by Guo et al.<sup>100</sup>, using a gel shift assay, which indicates that there was a significant reduction in DNA binding. This result suggests that conserved cysteine residues are key to the function of zinc binding domains, which are essential for the DNA helicase and ATPase activity of BLM. Therefore, the study suggests that DNA-binding ability is compromised due to this mutation<sup>100</sup>. Another study also showed, using the chimera mutational approach, that mutating C1055 leads to hypersensitivity to hydroxyurea (HU)<sup>85,100</sup>.

D1046 is present within the Zn subdomain and forms hydrogen-bonding interactions with the nearby R1037. This interaction is maintained intermittently during the simulations of the native structure. Though hydrophilic and negatively charged residues have been mutated to neutral and hydrophobic valines, the mutation occurred at the periphery of the protein, thus inducing local perturbations in the structure, and did not significantly affect the overall structural dynamics of the helicase except for the RQC domain (Fig. S3). The loss of the Zn-binding ability of such a mutant is observed as the Zn ion diffuses out of its pocket. Although our computational results suggest that its properties are similar to those of the WT, other studies suggest its crucial role in BS development<sup>85</sup>. C1066 is one of the four cysteines binding the Zn atom in the Zn subdomain. Our *in silico* results suggest that there is no effect because of the C1066Y mutation, but another report using site-directed mutational analysis suggests (Guo et al.) that there is a significant reduction in the zinc-binding ability, which further indicates that these highly conserved cysteine residues are essential for Zn<sup>+2</sup> ion binding for subsequent BLM activity<sup>100</sup>. A report from Shashtri et al. noted that C1066Y is crucial for Zn coordination, which is required for helicase activity and consequently for the development of BS<sup>5</sup>.

For these 18 nsSNP mutations, we observed that few mutations disrupt the overall BLM structure within a sub-nanosecond timescale and thus disrupt the functioning of BLM helicases. These mutations are specifically present within the cleft of two ATPase subdomains, 1A and 1B, and thus hamper the binding of ADP to this site. These mutations were also supported by other experimental studies mentioned in earlier sections<sup>5,85,100</sup>. For some mutations, however, local disruptions are observed in subdomains but can result in a complete loss of function. For instance, the mutation of cysteines (C1036F, C1055Y, C1066Y) and glutamate (D1064V) in the Zn binding subdomain, which results in the loss of Zn binding upon mutation and thus hinders the functioning of BLM helicase, was also supported by other reports<sup>100</sup>. However, in all the mutations, we did not find considerable changes in the binding of DNA. This result may be attributed to the fact that the mutations are performed in the ATPase domain, and the simulation timescales in this study are not long enough to observe large allosteric effects. For the mutations G891V, G952V and G978V, however, we did find larger deviations in RCQ domain binding and in the C-terminal HRDC domain, as indicated by rmsd values, thus suggesting the possibility of allosteric effects of mutations in the ATPase domain in nucleic acid binding. This conclusion was further strengthened by reports suggesting a role for cysteine residues in the Zn-binding domain in helicase activity. The mutation of cysteine hampers BLM helicase activity, which may be responsible for the development of BS<sup>100</sup>. The overall secondary structure remained consistent throughout the simulations of mutants as well as of the native structure within the simulated timescale (Fig. S4).

**Relation between BLM dysregulation and survival analysis in cancer.** To infer the functional consequences of BLM deregulation, we studied the relationship between the dysregulation of BLM and the clinical database of cancer patients. Using Kaplan–Meier plot analysis, we found that BLM dysregulation has distinct consequences in different types of cancers. High expression of the BLM gene (Affy probe id 205733\_at) is associated with lower overall survival for lung and gastric cancer patients, whereas we did not find a significant association between expression of the BLM gene and overall survival for breast and ovarian cancer patients (Fig. 8).

Substitutions	IMutant2		PoPMuSiC	
	DDG	Stability of variants	$\Delta\Delta G$	Stability of variants
P690L	0.74	Increase	0.51	Destabilizing
<b>P702L</b>	0.70	Increase	- 1.68	Stabilizing
W803R	-	-	-	-
W803L	-	-	-	-
H805L	-	-	-	-
<b>Y811C</b>	- 0.74	Decrease	2.37	Destabilizing
P825L	0.50	Increase	1.43	Destabilizing
G891V	- 3.71	Decrease	- 1.09	Stabilizing
<b>G952V</b>	- 2.83	Decrease	2.20	Destabilizing
<b>G972V</b>	- 0.09	Decrease	0.56	Destabilizing
<b>Y974C</b>	- 1.56	Decrease	2.63	Destabilizing
<b>G978V</b>	- 1.53	Decrease	0.90	Destabilizing
<b>C1036F</b>	- 1.48	Decrease	1.32	Destabilizing
<b>Y1044C</b>	- 0.68	Decrease	1.36	Destabilizing
<b>C1055R</b>	- 2.29	Decrease	1.70	Destabilizing
<b>C1055Y</b>	- 1.49	Decrease	0.45	Destabilizing
<b>D1064V</b>	- 1.65	Decrease	0.16	Destabilizing
<b>C1066Y</b>	- 1.50	Decrease	1.09	Destabilizing

**Table 2.** Effect of amino acid substitutions on the stability of mutant BLM protein using IMutant2 and PoPMuSiC v3.1 tools. IMutant2: Gibbs free energy change value (DDG) in Kcal/mol, where  $DDG < 0$  is decrease stability, while  $DDG > 0$  is increase stability of mutant protein. PoPMuSiC: The  $\Delta\Delta G$  indicates change in folding free energy in Kcal/mol where  $\Delta\Delta G < 0$  is stabilizing mutation. Consensus effect on amino acid substitutions between IMutant2 and PoPMuSiC are in bold. The DDG/ $\Delta\Delta G$  score for W803R, W803L, and H805L are absent because the amino acid residues from 799 to 807 are missing in the BLM helicase crystal structure (PDB id: 4O3M).

SNP	Phenotype	Environment	Area buried	Frac polar	Secondstr
P690L	Disease	P1C	0.472	0.458	C
P702L	Disease	B2H	0.526	0.417	H
W803R	Disease	-	-	-	-
W803L	Disease	-	-	-	-
H805L	Disease	-	-	-	-
Y811C	Disease	B2H	0.68	0.417	H
P825L	Neutral	P2C	0.432	0.708	C
G891V	Disease	ES	0.171	0.292	S
G952V	Disease	EC	0.165	0.333	C
G972V	Disease	EH	0.105	0.844	H
Y974C	Disease	B3H	0.713	0.448	H
G978V	Disease	EH	0.138	0.687	H
C1036F	Disease	P1C	0.221	0.24	C
Y1044C	Disease	B3H	0.615	0.615	H
C1055R	Disease	P1H	0.239	0.354	H
C1055Y	Disease	P1H	0.239	0.354	H
D1064V	Disease	P2H	0.385	0.594	H
C1066Y	Disease	P1H	0.227	0.458	H

**Table 3.** Predicting disease-associated amino acid substitution and phenotypic effect using nsSNPAnalyzer. Phenotype: Phenotype annotated by the Swiss-Prot DB. "Disease" or "Neutral". Environment: The structural environment of the SNP B1, B2, B3, P1, P2 and E\* AreaBuried: Solvent accessibility score. FracPolar: Environmental polarity score. Secondstr: Secondary structure. H: alpha-helix, S: beta-sheet, C: coil. \*The first character denotes the solvent accessibility B: buried, P: partially buried, E: exposed. The second number (if exists) denotes different environmental polarity provided the solvent accessibility is the same, with a larger number corresponding to a larger polarity. The structural features of last four columns for W803R, W803L, and H805L are absent because the amino acid residues from 799 to 807 are missing in the BLM helicase crystal structure (PDB id: 4O3M).

Substitution	Prediction	Score	HMM ID	HMM description	HMM Pos	HMM prob. W	HMM prob. M	HMM weights D	HMM weights O
P690L	PASSENGER/OTHER	1.79	41,442	P-loop containing nucleoside triphosphate hydrolases	118	0.148	0.011	1	4
P702L	PASSENGER/OTHER	1.49	41,442	P-loop containing nucleoside triphosphate hydrolases	130	0.346	0.070	1	4
W803R	PASSENGER/OTHER	1.92	41,442	P-loop containing nucleoside triphosphate hydrolases	231	0.088	0.033	1	4
W803L	PASSENGER/OTHER	2.01	41,442	P-loop containing nucleoside triphosphate hydrolases	231	0.088	0.092	1	4
H805L	PASSENGER/OTHER	1.99	41,442	P-loop containing nucleoside triphosphate hydrolases	233	0.092	0.086	1	4
Y811C	PASSENGER/OTHER	2.45	DEAD	DEAD/DEAH box helicase	140	0.050	0.026	5	28
P825L	PASSENGER/OTHER	2.41	DEAD	DEAD/DEAH box helicase	154	0.076	0.029	5	28
G891V	PASSENGER/OTHER	2.84	36,155	P-loop containing nucleoside triphosphate hydrolases	11	0.017	0.173	1	6
G952V	CANCER	- 1.49	Helicase_C	Helicase conserved C-terminal domain	46	0.798	0.006	8	14
G972V	PASSENGER/OTHER	3.06	36,155	P-loop containing nucleoside triphosphate hydrolases	83	0.002	0.282	1	6
Y974C	PASSENGER/OTHER	2.68	36,155	P-loop containing nucleoside triphosphate hydrolases	85	0.001	0.067	1	6
G978V	CANCER	- 1.12	41,442	P-loop containing nucleoside triphosphate hydrolases	263	0.885	0.001	1	4
C1036F	PASSENGER/OTHER	2.62	36,155	P-loop containing nucleoside triphosphate hydrolases	167	0.017	0.039	1	6
Y1044C	PASSENGER/OTHER	2.55	36,155	P-loop containing nucleoside triphosphate hydrolases	175	0.033	0.011	1	6
C1055R	PASSENGER/OTHER	2.61	36,155	P-loop containing nucleoside triphosphate hydrolases	207	0.012	0.027	1	6
C1055Y	PASSENGER/OTHER	2.61	36,155	P-loop containing nucleoside triphosphate hydrolases	207	0.012	0.029	1	6
D1064V	PASSENGER/OTHER	2.47	36,155	P-loop containing nucleoside triphosphate hydrolases	215	0.097	0.022	1	6
C1066Y	PASSENGER/OTHER	2.62	36,155	P-loop containing nucleoside triphosphate hydrolases	217	0.008	0.029	1	6

**Table 4.** Predicting cancer and disease-associated amino acid substitution and phenotypic effect using Fathmm cancer.

Mutation in the BLM gene in BL cells showed a reduction in the expression of its mRNA and protein, resulting in excessive chromosome instability and a high frequency of sister chromatid exchanges (SCEs)<sup>101</sup>. However, another study found that BLM mRNA overexpression is associated with poor survival of breast cancer patients<sup>102</sup>. Interestingly, the same study reported that subcellular localization of the BLM protein is found to be high in the cytoplasm compared to the nucleus, although the analysis of normal breast tissue revealed that BLM protein is strongly localized in the nucleus and not the cytoplasm<sup>102</sup>. It might be possible that variants of the BLM gene express a truncated BLM, which lacks signals for protein localization.

Therefore, this study also indicates the importance of BLM gene expression as a better prognostic marker for the detection of gastric and lung cancers. Studies have found that nsSNPs can affect the functional activity of proteins; therefore, we expect that the 18 nsSNPs identified here have functional effects, as in BLM deregulation.

**Expression levels of BLM genes in different cancers.** To understand the role of the BLM gene in cancer, we studied the its expression levels in different cancers. A box plot was generated using GEPIA (|Log2FC| cut-off = 1.5; p-value cut-off = 0.01), which showed that the expression of BLM significantly increased in different cancer patients compared to normal expression (Fig. 9). Therefore, based upon this data, we propose the role of BLM as a diagnostic marker for several cancers.

Substitution	Prediction	Score	HMM ID	HMM Description	HMM pos	HMM prob. W	HMM prob. M	HMM weights D	HMM weights O
C1036F	TOLERATED	- 0.97	36,155	P-loop containing nucleoside triphosphate hydrolases	167	0.017	0.039	12	6
C1055R	TOLERATED	- 0.98	36,155	P-loop containing nucleoside triphosphate hydrolases	207	0.012	0.027	12	6
C1055Y	TOLERATED	- 0.98	36,155	P-loop containing nucleoside triphosphate hydrolases	207	0.012	0.029	12	6
C1066Y	TOLERATED	- 0.97	36,155	P-loop containing nucleoside triphosphate hydrolases	217	0.008	0.029	12	6
D1064V	TOLERATED	- 1.11	36,155	P-loop containing nucleoside triphosphate hydrolases	215	0.097	0.022	12	6
G891V	TOLERATED	- 0.75	36,155	P-loop containing nucleoside triphosphate hydrolases	11	0.017	0.173	12	6
G952V	DAMAGING	- 3.19	Helicase_C	Helicase conserved C-terminal domain	46	0.798	0.006	26	14
G972V	TOLERATED	- 0.52	36,155	P-loop containing nucleoside triphosphate hydrolases	83	0.002	0.282	12	6
G978V	DAMAGING	- 4.12	41,442	P-loop containing nucleoside triphosphate hydrolases	263	0.885	0.001	8	4
H805L	TOLERATED	- 1.01	41,442	P-loop containing nucleoside triphosphate hydrolases	233	0.092	0.086	8	4
P690L	TOLERATED	- 1.21	41,442	P-loop containing nucleoside triphosphate hydrolases	118	0.148	0.011	8	4
P702L	DAMAGING	- 1.51	41,442	P-loop containing nucleoside triphosphate hydrolases	130	0.346	0.070	8	4
P825L	TOLERATED	2.41	DEAD	DEAD/DEAH box helicase	154	0.076	0.029	5	28
W803L	TOLERATED	- 0.99	41,442	P-loop containing nucleoside triphosphate hydrolases	231	0.088	0.092	8	4
W803R	TOLERATED	- 1.08	41,442	P-loop containing nucleoside triphosphate hydrolases	231	0.088	0.033	8	4
Y1044C	TOLERATED	- 1.03	36,155	P-loop containing nucleoside triphosphate hydrolases	175	0.033	0.011	12	6
Y811C	TOLERATED	2.45	DEAD	DEAD/DEAH box helicase	140	0.050	0.026	5	28
Y974C	TOLERATED	- 0.9	36,155	P-loop containing nucleoside triphosphate hydrolases	85	0.001	0.067	12	6

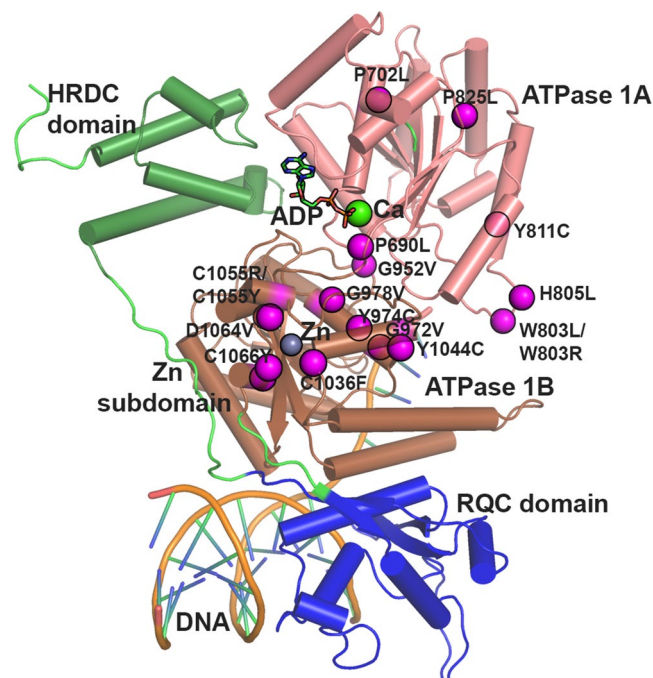
**Table 5.** Predicting cancer and disease-associated amino acid substitution and phenotypic effect using Fathmm Disease Ontology.

**Impact of 3'UTR SNPs in BLM gene regulation.** Earlier studies have shown the role of SNP in the 3'UTR of mRNA, which may lead to the partial or complete attenuation of complementary binding of miRNAs to the 3'UTR region<sup>19</sup>. On the other hand, SNPs in the 3'UTR region of mRNA can introduce new binding sites for other new miRNAs. We analysed the impact of gene regulation by miRNA due to SNPs in BLM mRNA using PolymiRTS v3.0. A list of 17 #rsIDs mapped to the 3'UTR and was fed into the PolymiRTS server, which yielded a list of binding miRNAs affected by these polymorphisms. The results showed that the binding of hsa-miR-6507-5p and hsa-miR-3976 to BLM mRNA was abolished due to polymorphisms rs116293756 and rs28363374, respectively (Table 6). Therefore, due to these SNPs, the BLM genes were not under the control of hsa-miR-6507-5p and hsa-miR-3976, which might result in higher expression of the BLM protein. In past decades, small interfering RNAs (siRNAs) have been widely used to silence the expression of target genes<sup>103,104</sup>. Designing allele-specific siRNAs using desiRm webserver (<https://crdd.osdd.net/raghava/desirm/>) against disease-causing SNPs would be a very promising therapeutic approach to suppress the overexpression of the BLM mutant gene<sup>105</sup>.

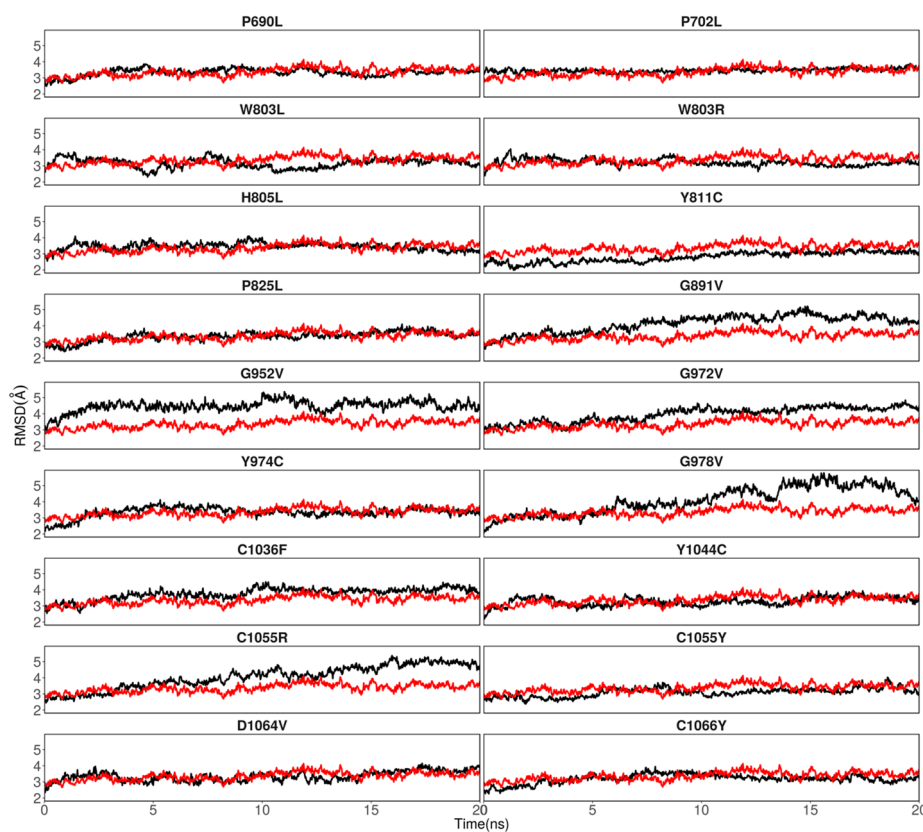
**The possible mechanism of defective BLM and cancer predisposition.** DNA damage during the replication process is corrected by the HR pathway. The HR starts with the degradation of 5'-terminal of DSBs, which generates 3'-overhang ssDNA<sup>106</sup>. Subsequently, RAD51 recombinase assembles at the ssDNA breakpoint and catalyzes the HR and DNA repair. During the HR process, invasion of ssDNA into the homologous sequence, typically a sister chromatid that works as a template, synthesizes the damaged DNA. At the last stage of HR, BLM forms a complex along with other enzymes and helps to separate the repaired DNA from the template DNA strand, and in this way BLM helps to maintain the genomic integrity<sup>107</sup>.

Biochemical studies showed that the deletion of BLM enhances the assembly of RAD51 at the ssDNA break site, while overexpression of BLM disrupts the RAD51 assembly through its helicase activity<sup>108,109</sup>. Therefore, BLM works as a 'anti-recombinase' at early and late stages of HR by: (i) preventing HR by disrupting the RAD51 assembly on the ssDNA of DSBs to make sure that HR only occurs between sequences with high homology;

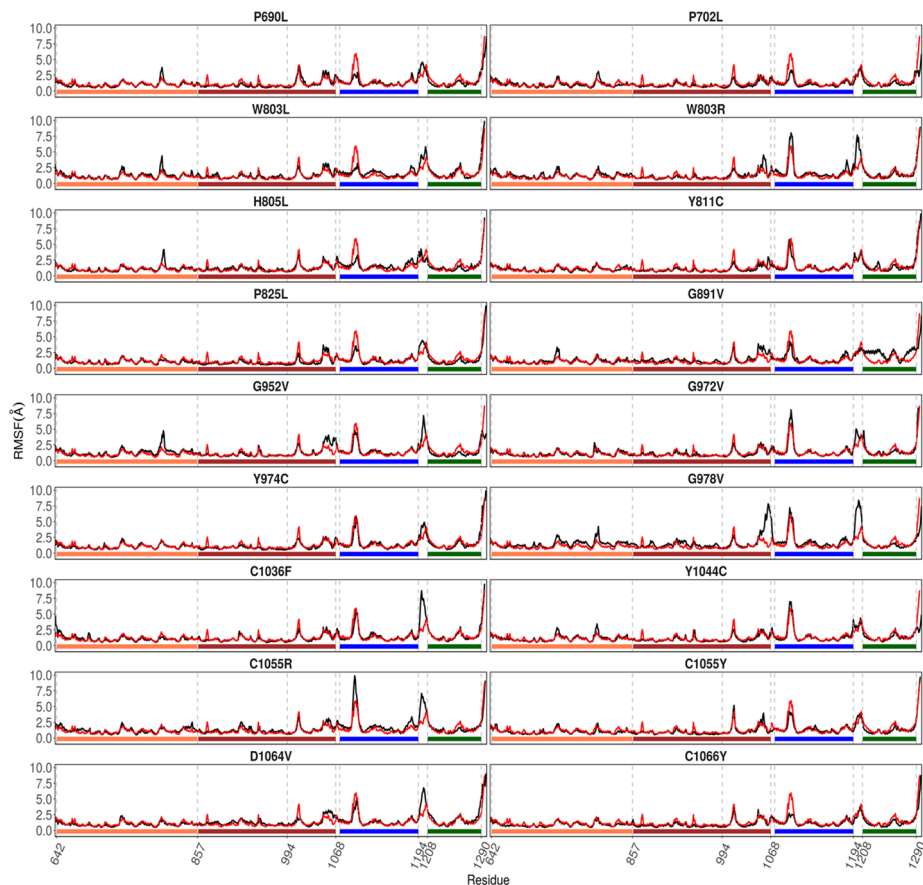




**Figure 3.** Various domains were observed in the BLM helicase crystal structure. (PDB id: 4O3M). The eighteen SNPs discussed in this study are shown by magenta spheres for Ca atoms. The bound ADP (in stick representation), Zn ion (grey sphere), Ca ion (green sphere) and DNA duplex with 3' overhang are also shown.



**Figure 4.** Variation in rmsd values with respect to time during simulations. Black lines correspond to mutant structures, and red lines correspond to WT or native structures.



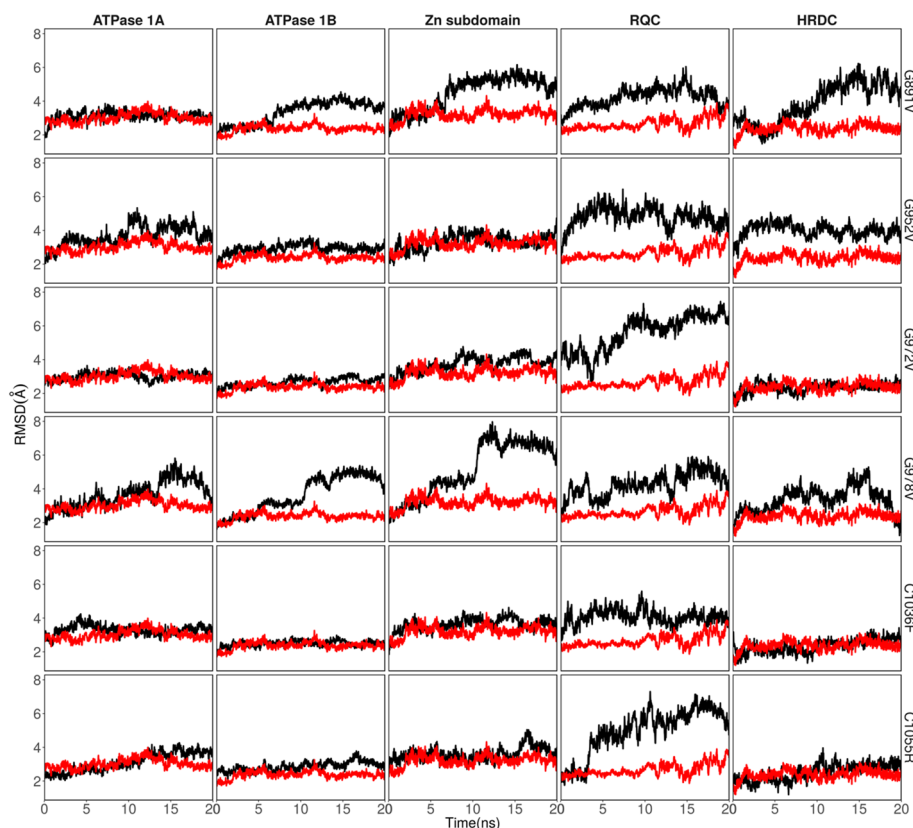
**Figure 5.** RMSF values calculated for each Ca atom with respect to residue. Black lines correspond to mutant and red lines correspond to WT or native structures. The bars below depict the domains, with coral for ATPase 1A, brown for ATPase 1B, blue for RQC and green for the HRDC domain.

and (ii) preventing SCEs by separating the repaired DNA from the template DNA strand that arises at the final stages of HR<sup>110,111</sup>.

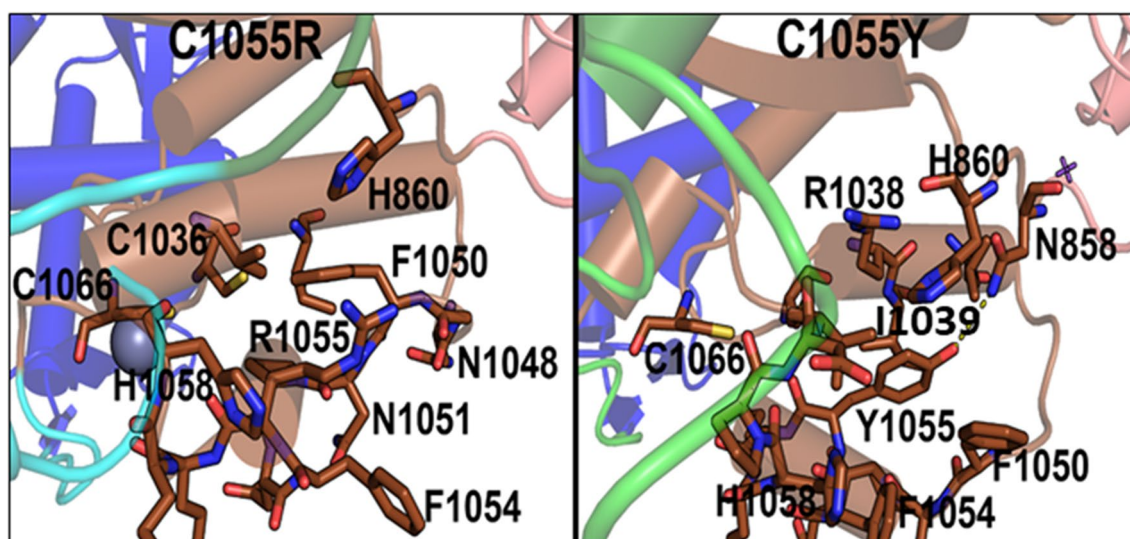
Previous studies showed that overexpression of BLM displaces the localization of RAD51 to the sites of DNA damage<sup>109,112</sup>. The loss of RAD51 leads to insufficient HR resulting in genomic instability and DNA damage. However, defective BLM increases the hyperaccumulation of RAD51 at DNA damage site resulting in a high level of HR and high frequency of SCEs<sup>106,111,112</sup>. Therefore, our study supports previous findings that an aberration in the activity of BLM through pathogenic nsSNPs or high expression of BLM (Figs. 8 and 9) could lead to genomic instability and predisposition to different cancer types.

## Conclusion

This study identifies eighteen highly deleterious and damaging nsSNPs for BLM inactivation using various powerful bioinformatic tools. Four of these mutations (P702L, W805L, P825L, and Y974C) have been reported in the literature to have a partial effect on helicase activity<sup>76</sup>. In support of this hypothesis, our MD studies of these mutations suggest that they maintained structural integrity during the simulations. Combining computational approaches with literature mining, we identified 14 more pathogenic mutations: P690L, W803R, W803L, Y811C, G891V, G952V, G972V, G978V, C1036F, Y1044C, C1055R, D1064V, and C1066Y. Interestingly, we observed that only six of these mutations strongly destabilize the helicase structure within the simulation time: G891V, G952V, G972V, G978V, C1036F, and C1055R. The remaining mutations either showed weak destabilizing effects and maintained the structural integrity of the helicase structure or showed similar structural deviations to those observed in the WT/native structure. However, one cannot rule out the possibility that these mutations may contribute to the likely allosteric effect, which the limited time scales of our simulation studies could not capture. Visualizing the time-evolution of domain-wise rmsd values, as a general trend, we observed that the mutations observed in the Zn subdomain not only destabilized the Zn subdomain, but also destabilized the nearby RQC domain as observed from their higher rmsd values. This suggests the allosteric effect of these mutations. One probable reason can be that in the crystal structure, Zn subdomain interacts with the DNA 3' overhang, while RQC domain interacts with the upstream duplex DNA, and thus, the nucleic acid may induce changes in RQC domain when Zn subdomain is mutated. Nonetheless, these observations warrant the detailed comparative studies in the future with longer sampling time to ascertain the allosteric effect of mutations.

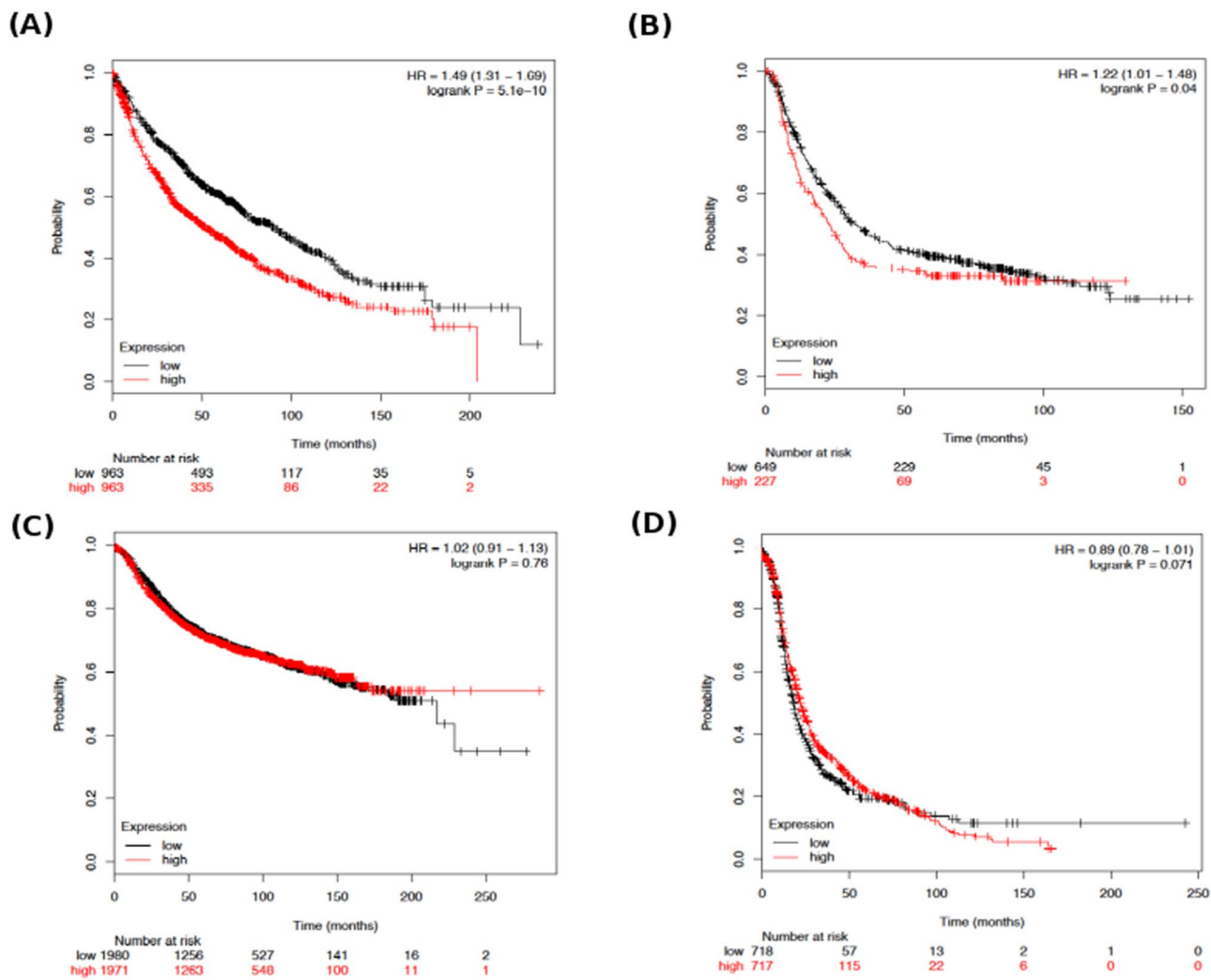


**Figure 6.** Domain wise rmsd plots for mutants destabilizing the complete BLM structure. Black line corresponds for mutants and red line corresponds for WT/native structure.



**Figure 7.** Interactions observed in the final conformations of mutants C1055R and C1055Y.

Since the MD simulation-based results suggest that G978V is pathogenic but no experimental evidence suggests its role in BS, future studies may sequence this pathogenic mutation site, especially in BS patients and other related diseases, and/or study it using mutational approaches using various models. Most of the pathogenic nsSNPs listed in the BLM protein suggest hampering the activity, such as ATP hydrolysis, DNA binding, and DNA unwinding, which is associated with various cellular procedures, such as replication, recombination, and DNA repair. BLM protein interacts with various factors for proper DNA replication and repair of DNA damage using the HR pathway. Turbulence during such a procedure always leads to the incorporation of erroneous copying of genomic information, giving rise to genomic instability and consequently leading to neoplastic transformation.



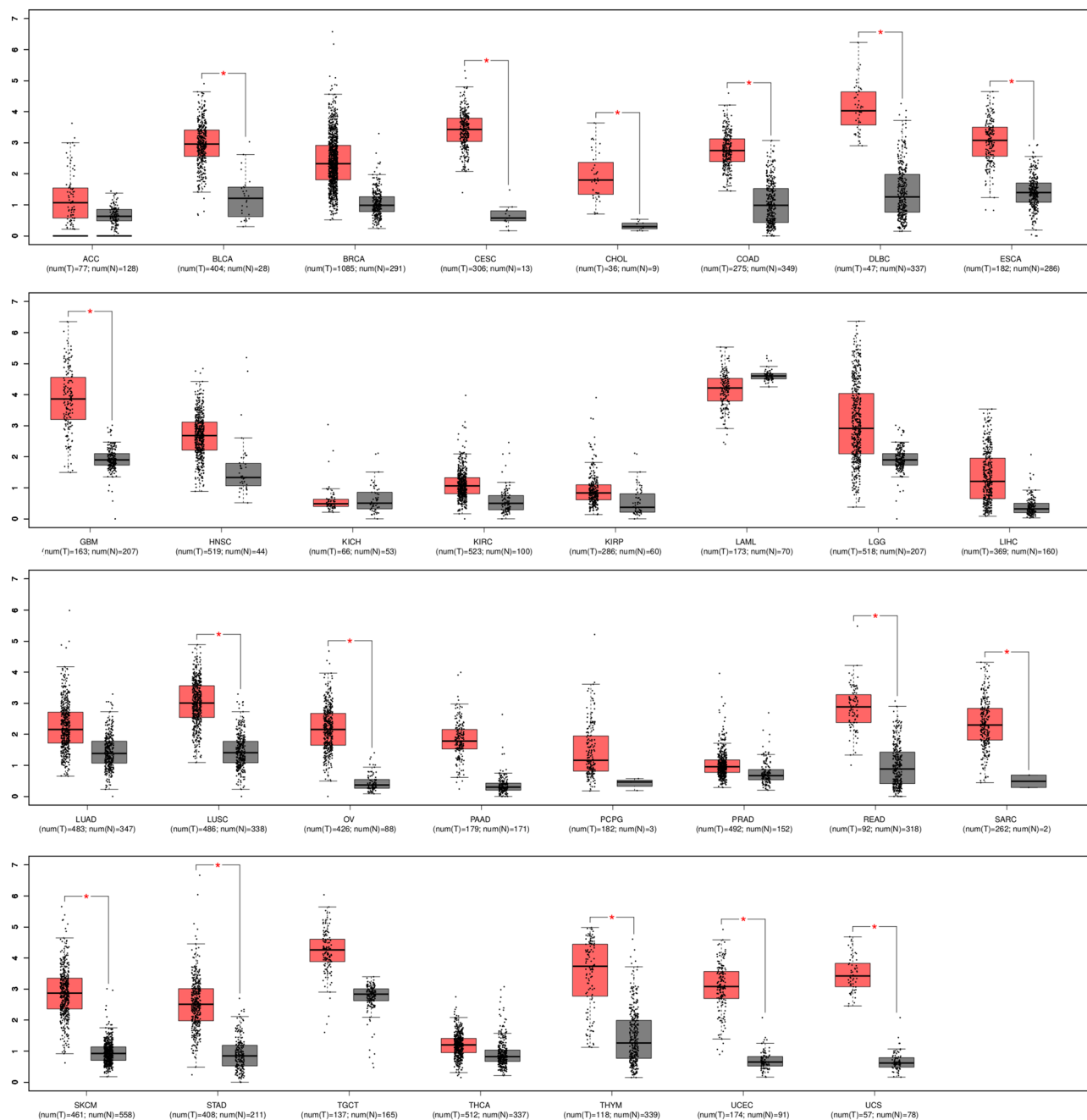
**Figure 8.** Kaplan–Meier curves showing the association of BLM mRNA expression and survival of patients in four cancers: (A) lung, (B) gastric, (C) breast, and (D) ovarian cancer.

Further future studies of BLM protein-related pathological nsSNPs will allow us to manage BS and its associated diseases. Figure 10 summarises our finding on pathogenic nsSNPs in BLM helicase.

Our study found 28 nsSNPs as “stop gained” and one nsSNP as “start lost” mutations, which can express a truncated form of BLM with the loss of important domains and nuclear signals and thus strongly localized BLM in the cytoplasm instead of the nucleus. As other studies have found, cancer cells have high cytoplasmic BLM, and it would be interesting to experimentally validate the nsSNPs with BLM protein localization in cancer cells. Furthermore, our investigation found that two SNPs, rs116293756 and rs28363374, at the 3'UTR of BLM abolished the binding affinity with the miRNAs hsa-miR-6507-5p and hsa-miR-3976, respectively, and thus enhanced the expression of BLM. Interestingly, we also found that BLM expression was significantly higher in some cancers including BLCA (Bladder Urothelial Carcinoma), COAD (Colon adenocarcinoma) and LUSC (Lung squamous cell carcinoma) (Fig. 9). Furthermore, Kaplan–Meier survival analysis suggested that high expression of the BLM gene is one of the reasons for the reduced survival of patients with lung or gastric cancer. Therefore, a deep understanding of how SNPs affect BLM transcription regulation and expression in cancer might be highly useful for the diagnosis and prognosis of disease.

In the future, our selected nsSNPs in the BLM gene can be further studied in different populations to explore and validate the contribution of these variants in BS and cancer, which may further lead to the design and development of potential drugs for the better management of BS and other associated diseases. Furthermore, our study provides key support to investigators to conduct future studies on pathological mutations and their structural consequences on BLM. The SNPs predicted from our study can be further validated by wet-lab scientists to investigate the evidence of BLM protein mutations in association to BS and develop a potential drug target for BS. These findings may enrich the available SNPs databases and then can be utilized for further research.





**Figure 9.** BLM mRNA expression level in normal tissues (N) and different cancer (T) samples. In each plot, y-axis indicates gene expression score calculated by mean value of  $\log_2(\text{TPM} + 1)$ . The red box indicates the tumor samples while the gray box represents the normal samples, and the number of samples is given in brackets. The significant differential expression of mRNA between tumor and normal samples indicates with symbol “\*” with p-value < 0.01. ACC Adrenocortical, BLCA Bladder Urothelial Carcinoma, BRCA Breast invasive carcinoma, CESC Cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL Cholangio carcinoma, COAD Colon adenocarcinoma, DLBC Lymphoid Neoplasm Diffuse Large B-cell Lymphoma, ESCA Esophageal carcinoma, GBM Glioblastoma multiforme, HNSC Head and Neck squamous cell carcinoma, KICH Kidney Chromophobe, KIRC Kidney renal clear cell carcinoma, KIRP Kidney renal papillary cell carcinoma, LAML Acute Myeloid Leukemia, LGG Brain Lower Grade Glioma, LIHC Liver hepatocellular carcinoma, LUAD Lung adenocarcinoma, LUSC Lung squamous cell carcinoma, OV Ovarian serous cystadenocarcinoma, PAAD Pancreatic adenocarcinoma, PCPG Pheochromocytoma and Paraganglioma, PRAD Prostate adenocarcinoma, READ Rectum adenocarcinoma, SARC Sarcoma, SKCM Skin Cutaneous Melanoma, STAD Stomach adenocarcinoma, TGCT Testicular Germ Cell Tumors, THCA (Thyroid carcinoma, THYM Thymoma, UCEC Uterine Corpus Endometrial Carcinoma, UCS Uterine Carcinosarcoma.

dbSNP ID	Variant type	Wobble base pair	Ancestral Allele	Allele	miR ID	Conservation	miR Site	Function class	Exp support	Context + score change
rs116293756	SNP	N	T	T	hsa-miR-6507-5p	2	tgtTAT TCTTggt	D	N	-0.094
rs28363374	SNP	N	C	C	hsa-miR-3976	2	actcgTCTC- TATt	D	N	-0.19

**Table 6.** Predicting miRNA-mRNA binding disruption due to SNPs in the 3'UTR of the BLM gene using PolymiRTS v3.0. Function Class D: The derived allele disrupts a conserved miRNA site. Exp Support N: Predicted target site with no experimental support.

Tools	P690L	P702L	W803R	W803L	H805L	Y811C	P825L	G891V	G952V	G972V	Y974C	G978V	C1036F	Y1044C	C1055R	C1055V	D1064V	C1066Y	Structural & Functional Consequences
SIFT																			Damaging effect on protein structure
PROVEAN																			Damaging effect on protein structure
CONDEL																			Damaging effect on protein structure
PolyPhen2																			Probably damaging
SNAP2																			Damaging effect on protein structure
Weblogo																			Amino acid (wild type) conservation
PopMusic			Not predicted																Protein stability
IMutant2			Not predicted																Protein stability
SNP&GO																			Disease Association
PhD-SNP																			Disease Association
PANTHER																			Disease Association
nsSNPAnalyzer																			Disease Association
MutPred2																			Pathogenic
fathmm																			Disease Association
MD Simulation																			Dynamics
Literature Mining																			Inactivation
Reference	5, 85		86-89, 99			86, 88, 99		7-8, 90-91	92-97	100				85, 100	98	85, 100	99	5, 85	5, 85
Possible reason	DB		DB	Hydrolysis	HM	DB, DW		DW, ATPase	REPL	IAB				IZC	RICRC	IZC	DB	IZC, BLM	IZC

**Figure 10.** Structural functional consequences of nsSNPs on BLM helicase protein based on multiple computational tools prediction and literature mining. The first column represents the tool name, while the last column represents its use for predicting different structural and functional consequences. The mutant effects on W803R, W803L, and H805L was not predicted by IMutant2, and PoPMuSiC because the amino acid residues from 799–807 are missing in the BLM helicase crystal structure (PDB id: 4O3M). Red Box = damaging or disease associated, Green box = Not predicted by software because residue absent in crystal structure, Orange = highly conserved amino acid, Pink box = No damaging effect, DB = DNA Binding, IAB = Impaired ATP binding, DU = DNA Unwinding, IZC = Impaired Zn Coordination, AD = ATPase domain, RICRC = Reported in CRC patient, REPL = Replication, iBLM = deletion cause inactivation of BLM.

Received: 18 April 2018; Accepted: 6 July 2020

Published online: 23 July 2020

## References

- Hall, M. C. & Matson, S. W. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **34**, 867–877. <https://doi.org/10.1046/j.1365-2958.1999.01659.x> (1999).
- Matson, S. W., Bean, D. W. & George, J. W. DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *BioEssays* **16**, 13–22. <https://doi.org/10.1002/bies.950160103> (1994).
- Schmid, S. R. & Linder, P. D-E-A-D protein family of putative RNA helicases. *Mol. Microbiol.* **6**, 283–291. <https://doi.org/10.1111/j.1365-2958.1992.tb01470.x> (1992).
- Hickson, I. D. RecQ helicases: caretakers of the genome. *Nat. Rev. Cancer* **3**, 169–178. <https://doi.org/10.1038/nrc1012> (2003).
- Shastri, V. M. & Schmidt, K. H. Cellular defects caused by hypomorphic variants of the Bloom syndrome helicase gene BLM. *Mol. Genet. Genomic Med.* **4**, 106–119. <https://doi.org/10.1002/mgg3.188> (2016).
- Arora, H. *et al.* Bloom syndrome. *Int. J. Dermatol.* **53**, 798–802. <https://doi.org/10.1111/ijd.12408> (2014).
- Ellis, N. A. *et al.* The Bloom's syndrome gene product is homologous to RecQ helicases. *Cell* **83**, 655–666. [https://doi.org/10.1016/0092-8674\(95\)90105-1](https://doi.org/10.1016/0092-8674(95)90105-1) (1995).
- Foucault, F. *et al.* Characterization of a new BLM mutation associated with a topoisomerase II alpha defect in a patient with Bloom's syndrome. *Hum. Mol. Genet.* **6**, 1427–1434. <https://doi.org/10.1093/hmg/6.9.1427> (1997).

9. German, J., Sanz, M. M., Ciocci, S., Ye, T. Z. & Ellis, N. A. Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Hum. Mutat.* **28**, 743–753. <https://doi.org/10.1002/humu.20501> (2007).
10. German, J. Bloom's syndrome. *Dermatol. Clin.* **13**, 7–18 (1995).
11. Sassi, A., Popielarski, M., Synowiec, E., Morawiec, Z. & Wozniak, K. BLM and RAD51 genes polymorphism and susceptibility to breast cancer. *Pathol. Oncol. Res.* **19**, 451–459. <https://doi.org/10.1007/s12253-013-9602-8> (2013).
12. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
13. Cunniff, C., Bassetti, J. A. & Ellis, N. A. Bloom's syndrome: clinical spectrum, molecular pathogenesis, and cancer predisposition. *Mol. Syndromol.* **8**, 4–23. <https://doi.org/10.1159/000452082> (2017).
14. Yin, Q. K. *et al.* Discovery of Isaingotone derivatives as novel bloom's syndrome protein (BLM) helicase inhibitors that disrupt the BLM/DNA interactions and regulate the homologous recombination repair. *J. Med. Chem.* **62**, 3147–3162. <https://doi.org/10.1021/acs.jmedchem.9b00083> (2019).
15. Sim, N. L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–457. <https://doi.org/10.1093/nar/gks539> (2012).
16. Ramensky, V., Bork, P. & Sunyaev, S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900. <https://doi.org/10.1093/nar/gkf493> (2002).
17. Ahmed, F., Benedito, V. A. & Zhao, P. X. Mining functional elements in messenger RNAs: overview, challenges, and perspectives. *Front. Plant. Sci.* **2**, 84. <https://doi.org/10.3389/fpls.2011.00084> (2011).
18. Ahmed, F., Kumar, M. & Raghava, G. P. Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies. *Silico Biol.* **9**, 135–148 (2009).
19. Kamaraj, B., Gopalakrishnan, C. & Purohit, R. In silico analysis of miRNA-mediated gene regulation in OCA and OA genes. *Cell. Biochem. Biophys.* **70**, 1923–1932. <https://doi.org/10.1007/s12013-014-0152-9> (2014).
20. Ahmed, F. *et al.* Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in *Arabidopsis thaliana*. *RNA Biol.* **11**, 1414–1429. <https://doi.org/10.1080/15476286.2014.996474> (2014).
21. Carvalho, M. A. *et al.* Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis. *Cancer Res.* **67**, 1494–1501. <https://doi.org/10.1158/0008-5472.CAN-06-3297> (2007).
22. Carvalho, M. *et al.* Analysis of a set of missense, frameshift, and in-frame deletion variants of BRCA1. *Mutat. Res.* **660**, 1–11. <https://doi.org/10.1016/j.mrfmmm.2008.09.017> (2009).
23. Kamaraj, B. & Purohit, R. In silico screening and molecular dynamics simulation of disease-associated nsSNP in TYRP1 gene and its structural consequences in OCA3. *Biomed. Res. Int.* **2013**, 697051. <https://doi.org/10.1155/2013/697051> (2013).
24. Kamaraj, B. & Purohit, R. Mutational analysis of oculocutaneous albinism: a compact review. *Biomed. Res. Int.* **2014**, 905472. <https://doi.org/10.1155/2014/905472> (2014).
25. Kamaraj, B. & Bogaerts, A. Structure and function of p53-DNA complexes with inactivation and rescue mutations: a molecular dynamics simulation study. *PLoS ONE* **10**, e0134638. <https://doi.org/10.1371/journal.pone.0134638> (2015).
26. Goldgar, D. E. *et al.* Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am. J. Hum. Genet.* **75**, 535–544. <https://doi.org/10.1086/424388> (2004).
27. Karchin, R. Next generation tools for the annotation of human SNPs. *Brief Bioinform.* **10**, 35–52. <https://doi.org/10.1093/bib/bbn047> (2009).
28. Kamaraj, B. & Purohit, R. Computational screening of disease-associated mutations in OCA2 gene. *Cell. Biochem. Biophys.* **68**, 97–109. <https://doi.org/10.1007/s12013-013-9697-2> (2014).
29. Kamaraj, B. & Purohit, R. Mutational analysis on membrane associated transporter protein (MATP) and their structural consequences in oculocutaneous albinism type 4 (OCA4)-A molecular dynamics approach. *J. Cell. Biochem.* **117**, 2608–2619. <https://doi.org/10.1002/jcb.25555> (2016).
30. Balu, K. & Purohit, R. Mutational analysis of TYR gene and its structural consequences in OCA1A. *Gene* **513**, 184–195. <https://doi.org/10.1016/j.gene.2012.09.128> (2013).
31. Purohit, R., Rajendran, V. & Sethumadhavan, R. Relationship between mutation of serine residue at 315th position in M. tuberculosis catalase-peroxidase enzyme and Isoniazid susceptibility: an in silico analysis. *J. Mol. Model* **17**, 869–877. <https://doi.org/10.1007/s00894-010-0785-6> (2011).
32. Purohit, R., Rajendran, V. & Sethumadhavan, R. Studies on adaptability of binding residues and flap region of TMC-114 resistance HIV-1 protease mutants. *J. Biomol. Struct. Dyn.* **29**, 137–152. <https://doi.org/10.1080/07391102.2011.10507379> (2011).
33. Kumar, A., Rajendran, V., Sethumadhavan, R. & Purohit, R. In silico prediction of a disease-associated STIL mutant and its affect on the recruitment of centromere protein J (CENPJ). *FEBS Open Biol.* **2**, 285–293. <https://doi.org/10.1016/j.fob.2012.09.003> (2012).
34. Kumar, A., Rajendran, V., Sethumadhavan, R. & Purohit, R. Relationship between a point mutation S97C in CK1delta protein and its affect on ATP-binding affinity. *J. Biomol. Struct. Dyn.* **32**, 394–405. <https://doi.org/10.1080/07391102.2013.770373> (2014).
35. Ahmed, F. Integrated network analysis reveals FOXM1 and MYBL2 as key regulators of cell proliferation in non-small cell lung cancer. *Front. Oncol.* **9**, 1011. <https://doi.org/10.3389/fonc.2019.01011> (2019).
36. Wang, B. *et al.* Long-range signaling in MutS and MSH homologs via switching of dynamic communication pathways. *PLoS Comput. Biol.* **12**, e1005159. <https://doi.org/10.1371/journal.pcbi.1005159> (2016).
37. Rehan, M. & Bajouh, O. S. Virtual screening of naphthoquinone analogs for potent inhibitors against the cancer-signaling PI3K/AKT/mTOR pathway. *J. Cell. Biochem.* <https://doi.org/10.1002/jcb.27100> (2018).
38. Jamal, M. S. *et al.* Anticancer compound plumbagin and its molecular targets: a structural insight into the inhibitory mechanisms using computational approaches. *PLoS ONE* **9**, e87309. <https://doi.org/10.1371/journal.pone.0087309> (2014).
39. Sharma, M., Predeus, A. V., Mukherjee, S. & Feig, M. DNA bending propensity in the presence of base mismatches: implications for DNA repair. *J. Phys. Chem. B* **117**, 6194–6205. <https://doi.org/10.1021/jp403127a> (2013).
40. Jamal, M. S. *et al.* Defining the antigen receptor-dependent regulatory network that induces arrest of cycling immature B-lymphocytes. *BMC Syst. Biol.* **4**, 169. <https://doi.org/10.1186/1752-0509-4-169> (2010).
41. Ahmed, F., Kaundal, R. & Raghava, G. P. PHDcleav: a SVM based method for predicting human Dicer cleavage sites using sequence and secondary structure of miRNA precursors. *BMC Bioinform.* **14**(Suppl 14), S9. <https://doi.org/10.1186/1471-2105-14-S14-S9> (2013).
42. Khan, I. & Ansari, I. A. Prediction of a highly deleterious mutation E17K in AKT-1 gene: an in silico approach. *Biochem. Biophys. Rep.* **10**, 260–266. <https://doi.org/10.1016/j.bbrep.2017.04.013> (2017).
43. Grillo, G. *et al.* UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs. *Nucleic Acids Res.* **38**, D75–80. <https://doi.org/10.1093/nar/gkp902> (2010).
44. Yuan, H. Y. *et al.* FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. *Nucleic Acids Res.* **34**, W635–641. <https://doi.org/10.1093/nar/gkl236> (2006).
45. Ahmed, F., Dai, X. & Zhao, P. X. Bioinformatics tools for achieving better gene silencing in plants. *Methods Mol. Biol.* **1287**, 43–60. [https://doi.org/10.1007/978-1-4939-2453-0\\_3](https://doi.org/10.1007/978-1-4939-2453-0_3) (2015).
46. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195> (2015).

47. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249. <https://doi.org/10.1038/nmeth0410-248> (2010).
48. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC Genom.* **16**(Suppl 8), S1. <https://doi.org/10.1186/1471-2164-16-S8-S1> (2015).
49. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797. <https://doi.org/10.1093/nar/gkh340> (2004).
50. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190. <https://doi.org/10.1101/gr.849004> (2004).
51. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100. <https://doi.org/10.1093/nar/18.20.6097> (1990).
52. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **30**, 1237–1244. <https://doi.org/10.1002/humu.21047> (2009).
53. Bao, L., Zhou, M. & Cui, Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res.* **33**, W480–482. <https://doi.org/10.1093/nar/gki372> (2005).
54. Pejaver, V. *et al.* MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv* <https://doi.org/10.1101/134981> (2017).
55. Gonzalez-Perez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score. *Condel. Am. J. Hum. Genet.* **88**, 440–449. <https://doi.org/10.1016/j.ajhg.2011.03.004> (2011).
56. Clifford, R. J., Edmonson, M. N., Nguyen, C. & Buetow, K. H. Large-scale analysis of non-synonymous coding region single nucleotide polymorphisms. *Bioinformatics* **20**, 1006–1014. <https://doi.org/10.1093/bioinformatics/bth029> (2004).
57. Stone, E. A. & Sidow, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986. <https://doi.org/10.1101/gr.3804205> (2005).
58. Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8**, R232. <https://doi.org/10.1186/gb-2007-8-11-r232> (2007).
59. Venselaar, H., Te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases An e-Science approach with life scientist friendly interfaces. *BMC Bioinform.* **11**, 548. <https://doi.org/10.1186/1471-2105-11-548> (2010).
60. Swan, M. K. *et al.* Structure of human Bloom's syndrome helicase in complex with ADP and duplex DNA. *Acta Crystallogr. D* **70**, 1465–1475. <https://doi.org/10.1107/S139900471400501X> (2014).
61. Eswar, N. *et al.* Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* <https://doi.org/10.1002/0471250953.bi0506s15> (2006).
62. Maier, J. A. *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255> (2015).
63. Zgarbova, M. *et al.* Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z- and B-DNA. *J. Chem. Theory Comput.* **11**, 5723–5736. <https://doi.org/10.1021/acs.jctc.5b00716> (2015).
64. Abraham, M. J. *et al.* GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001> (2015).
65. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**(33–38), 27–38 (1996).
66. Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695–2696. <https://doi.org/10.1093/bioinformatics/btl461> (2006).
67. Skjaerven, L., Yao, X. Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinform.* **15**, 399. <https://doi.org/10.1186/s12859-014-0399-6> (2014).
68. Skjaerven, L., Jariwala, S., Yao, X. Q. & Grant, B. J. Online interactive analysis of protein structure ensembles with Bio3D-web. *Bioinformatics* **32**, 3510–3512. <https://doi.org/10.1093/bioinformatics/btw482> (2016).
69. Szasz, A. M. *et al.* Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* **7**, 49322–49333. <https://doi.org/10.18632/oncotarget.10337> (2016).
70. Tang, Z. *et al.* GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res.* **45**, W98–W102. <https://doi.org/10.1093/nar/gkx247> (2017).
71. Bhattacharya, A., Ziebarth, J. D. & Cui, Y. PolymiRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* **42**, D86–91. <https://doi.org/10.1093/nar/gkt1028> (2014).
72. Ahmed, F., Ansari, H. R. & Raghava, G. P. Prediction of guide strand of microRNAs from its sequence and secondary structure. *BMC Bioinform.* **10**, 105. <https://doi.org/10.1186/1471-2105-10-105> (2009).
73. Kamaraj, B., Rajendran, V., Sethumadhavan, R., Kumar, C. V. & Purohit, R. Mutational analysis of FUS gene and its structural and functional role in amyotrophic lateral sclerosis 6. *J. Biomol. Struct. Dyn.* **33**, 834–844. <https://doi.org/10.1080/07391102.2014.915762> (2015).
74. Arshad, M., Bhatti, A. & John, P. Identification and in silico analysis of functional SNPs of human TAGAP protein: a comprehensive study. *PLoS ONE* **13**, e0188143. <https://doi.org/10.1371/journal.pone.0188143> (2018).
75. Singh, D. *et al.* Computational and mutational analysis of TatD DNase of Bacillus anthracis. *J. Cell Biochem.* <https://doi.org/10.1002/jcb.28408> (2019).
76. Wang, Z. & Moul, J. SNPs, protein structure, and disease. *Hum. Mutat.* **17**, 263–270. <https://doi.org/10.1002/humu.22> (2001).
77. Bromberg, Y. & Rost, B. Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinform.* **10**(Suppl 8), S8. <https://doi.org/10.1186/1471-2105-10-S8-S8> (2009).
78. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res* **33**, W306–310. <https://doi.org/10.1093/nar/gki375> (2005).
79. Dehouck, Y., Kwasiogroch, J. M., Gilis, D. & Rooman, M. PopMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinform.* **12**, 151. <https://doi.org/10.1186/1471-2105-12-151> (2011).
80. Shihab, H. A., Gough, J., Cooper, D. N., Day, I. N. & Gaunt, T. R. Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics* **29**, 1504–1510. <https://doi.org/10.1093/bioinformatics/btt182> (2013).
81. Kitano, K. Structural mechanisms of human RecQ helicases WRN and BLM. *Front. Genet.* **5**, 366. <https://doi.org/10.3389/fgene.2014.00366> (2014).
82. Cheok, C. F., Wu, L., Garcia, P. L., Janscak, P. & Hickson, I. D. The Bloom's syndrome helicase promotes the annealing of complementary single-stranded DNA. *Nucleic Acids Res.* **33**, 3932–3941. <https://doi.org/10.1093/nar/gki712> (2005).
83. Wu, L. *et al.* The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *EMBO J* **24**, 2679–2687. <https://doi.org/10.1038/sj.emboj.7600740> (2005).
84. Gyimesi, M. *et al.* Visualization of human Bloom's syndrome helicase molecules bound to homologous recombination intermediates. *FASEB J.* **27**, 4954–4964. <https://doi.org/10.1096/fj.13-234088> (2013).
85. Mirzaei, H. & Schmidt, K. H. Non-Bloom syndrome-associated partial and total loss-of-function variants of BLM helicase. *Proc. Natl. Acad. Sci. USA* **109**, 19357–19362. <https://doi.org/10.1073/pnas.1210304109> (2012).
86. Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S. & Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **97**, 75–84. [https://doi.org/10.1016/s0092-8674\(00\)80716-3](https://doi.org/10.1016/s0092-8674(00)80716-3) (1999).



87. Dillingham, M. S., Soultanas, P. & Wigley, D. B. Site-directed mutagenesis of motif III in PcrA helicase reveals a role in coupling ATP hydrolysis to strand separation. *Nucleic Acids Res.* **27**, 3310–3317. <https://doi.org/10.1093/nar/27.16.3310> (1999).
88. Dillingham, M. S., Soultanas, P., Wiley, P., Webb, M. R. & Wigley, D. B. Defining the roles of individual residues in the single-stranded DNA binding site of PcrA helicase. *Proc. Natl. Acad. Sci. USA* **98**, 8381–8387. <https://doi.org/10.1073/pnas.131009598> (2001).
89. Pike, A. C. *et al.* Structure of the human RECQ1 helicase reveals a putative strand-separation pin. *Proc. Natl. Acad. Sci. USA* **106**, 1039–1044. <https://doi.org/10.1073/pnas.0806908106> (2009).
90. Rong, S. B., Valiaho, J. & Vihinen, M. Structural basis of Bloom syndrome (BS) causing mutations in the BLM helicase domain. *Mol. Med.* **6**, 155–164 (2000).
91. Bahr, A., De Graeve, F., Kedinger, C. & Chatton, B. Point mutations causing Bloom's syndrome abolish ATPase and DNA helicase activities of the BLM protein. *Oncogene* **17**, 2565–2571. <https://doi.org/10.1038/sj.onc.1202389> (1998).
92. Selak, N. *et al.* The Bloom's syndrome helicase (BLM) interacts physically and functionally with p12, the smallest subunit of human DNA polymerase delta. *Nucleic Acids Res.* **36**, 5166–5179. <https://doi.org/10.1093/nar/gkn498> (2008).
93. Bischof, O. *et al.* Regulation and localization of the Bloom syndrome protein in response to DNA damage. *J. Cell. Biol.* **153**, 367–380. <https://doi.org/10.1083/jcb.153.2.367> (2001).
94. Yankiwski, V., Marciniak, R. A., Guarente, L. & Neff, N. F. Nuclear structure in normal and Bloom syndrome cells. *Proc. Natl. Acad. Sci. USA* **97**, 5214–5219. <https://doi.org/10.1073/pnas.090525897> (2000).
95. Dutertre, S. *et al.* Cell cycle regulation of the endogenous wild type Bloom's syndrome DNA helicase. *Oncogene* **19**, 2731–2738. <https://doi.org/10.1038/sj.onc.1203595> (2000).
96. Brosh, R. M. *et al.* Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *J. Biol. Chem.* **275**, 23500–23508. <https://doi.org/10.1074/jbc.M001557200> (2000).
97. Jiao, R. *et al.* Physical and functional interaction between the Bloom's syndrome gene product and the largest subunit of chromatin assembly factor I. *Mol. Cell Biol.* **24**, 4710–4719. <https://doi.org/10.1128/MCB.24.11.4710-4719.2004> (2004).
98. Raskin, L. *et al.* Targeted sequencing of established and candidate colorectal cancer genes in the Colon Cancer Family Registry Cohort. *Oncotarget* **8**, 93450–93463. <https://doi.org/10.18632/oncotarget.18596> (2017).
99. Guo, R. B. *et al.* Structural and functional analyses of disease-causing missense mutations in Bloom syndrome protein. *Nucleic Acids Res.* **35**, 6297–6310. <https://doi.org/10.1093/nar/gkm536> (2007).
100. Guo, R. B., Rigolet, P., Zargarian, L., Fermandjian, S. & Xi, X. G. Structural and functional characterizations reveal the importance of a zinc binding domain in Bloom's syndrome helicase. *Nucleic Acids Res.* **33**, 3109–3124. <https://doi.org/10.1093/nar/gki619> (2005).
101. Croteau, D. L., Popuri, V., Opreko, P. L. & Bohr, V. A. Human RecQ helicases in DNA repair, recombination, and replication. *Annu. Rev. Biochem.* **83**, 519–552. <https://doi.org/10.1146/annurev-biochem-060713-035428> (2014).
102. Arora, A. *et al.* transcriptomic and protein expression analysis reveals clinicopathological significance of bloom syndrome helicase (BLM) in breast cancer. *Mol. Cancer Ther.* **14**, 1057–1065. <https://doi.org/10.1158/1535-7163.MCT-14-0939> (2015).
103. Tyagi, A. *et al.* HIVsirDB: a database of HIV inhibiting siRNAs. *PLoS ONE* **6**, e25917. <https://doi.org/10.1371/journal.pone.0025917> (2011).
104. Dar, S. A., Thakur, A., Qureshi, A. & Kumar, M. siRNAmdb: a database of experimentally validated chemically modified siRNAs. *Sci. Rep.* **6**, 20031. <https://doi.org/10.1038/srep20031> (2016).
105. Ahmed, F. & Raghava, G. P. Designing of highly effective complementary and mismatch siRNAs for silencing a gene. *PLoS ONE* **6**, e23443. <https://doi.org/10.1371/journal.pone.0023443> (2011).
106. Chu, W. K., Hanada, K., Kanaar, R. & Hickson, I. D. BLM has early and late functions in homologous recombination repair in mouse embryonic stem cells. *Oncogene* **29**, 4705–4714. <https://doi.org/10.1038/nc.2010.214> (2010).
107. Wu, L. & Hickson, I. D. The Bloom's syndrome helicase suppresses crossing over during homologous recombination. *Nature* **426**, 870–874. <https://doi.org/10.1038/nature02253> (2003).
108. Bugreev, D. V., Yu, X., Egelman, E. H. & Mazin, A. V. Novel pro- and anti-recombination activities of the Bloom's syndrome helicase. *Genes Dev.* **21**, 3085–3094. <https://doi.org/10.1101/gad.1609007> (2007).
109. Birkbak, N. J. *et al.* Overexpression of BLM promotes DNA damage and increased sensitivity to platinum salts in triple-negative breast and serous ovarian cancers. *Ann. Oncol.* **29**, 903–909. <https://doi.org/10.1093/annonc/mdy049> (2018).
110. Her, J., Ray, C. & Bunting, S. F. The BLM Helicase: Keeping recombination honest?. *Cell Cycle* **17**, 401–402. <https://doi.org/10.1080/15384101.2017.1421045> (2018).
111. Patel, D. S., Misenko, S. M., Her, J. & Bunting, S. F. BLM helicase regulates DNA repair by counteracting RAD51 loading at DNA double-strand break sites. *J. Cell. Biol.* **216**, 3521–3534. <https://doi.org/10.1083/jcb.201703144> (2017).
112. Ouyang, K. J. *et al.* SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Biol.* **7**, e1000252. <https://doi.org/10.1371/journal.pbio.1000252> (2009).

## Acknowledgements

The authors would like to acknowledge the technical support provided by Deanship of Scientific Research, King Abdulaziz University; the Department of Biochemistry, University of Jeddah; and University of Jeddah Center for Scientific and Medical Research (UJCSMR), University of Jeddah; Nature Research Editing Services, Nabila V. Ikram from SciDra Consulting, Inc for manuscript editing assistance.

## Author contributions

F.A.L. conceived the idea, generated data, and wrote the manuscript. F.A. conceived and improved the concept of the presented idea, wrote the computer scripts, generated data, analyzed the results, wrote, edited, guided, supervised and concluded the manuscript. M.S. verified the computational methods, generated data, wrote and edited the manuscript. M.R. verified the computational methods, generated data, wrote and edited the manuscript. M.M. verified the generated data and wrote the manuscript. M.N.B. verified the computational methods and edited the manuscript. Y.H. wrote and edited the manuscript. A.A. wrote and edited the manuscript. S.A.A. wrote and edited the manuscript. M.A.K. verified the computational methods and revised the manuscript. M.K.W. analyzed the data and revised the manuscript. H.C. verified the results and edited the manuscript. M.S.J. conceived the idea, guided to execute the experiments, supervised the project, wrote, analyzed the results, concluded and edited the manuscript. All authors read, revised and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-69033-8>.

**Correspondence** and requests for materials should be addressed to F.A. or M.S.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020