Identification of BLM's Aggregation-Prone Nature and Its Similarity to Intrinsically Disordered Proteins: An In-silico Study

Sourav Sharma[1], Suresh Singh Yadav[2], Rohini Ravindran Nair[1*]

[1] Department of Medical Biotechnology, Gujarat Biotechnology University, Gandhinagar, India

[2] Department of Molecular Biology and Biochemistry, Guru Nanak Dev University, Amritsar, India

***Correspondence: Rohini Ravindran Nair, rohini.nair@gbu.edu.in**

Abstract

BLM helicase is a member of the RecQ family of DNA helicases, enzymes that play a critical role in maintaining genome stability. The BLM protein is named after Bloom syndrome (BS), a rare genetic disorder caused by mutations in the helicase domain of BLM gene. BLM role has mostly been studied in DNA repair, replication, and recombination; however, a recent study has highlighted the RNA binding capability of the BLM protein. Several proteins, including TDP43, Tau, and alpha-synuclein, are known to play key roles in neurodegenerative diseases. These proteins possess residues prone to parallel aggregation and fibril formation, the key contributor to neurodegenerative disease development. We utilized various in-silico tools like PASTA 2.0, SIFT, PolyPhen 2, and PhD SNP, SNP&GO, Meta-SNP, and SNAP, etc., to identify the molecular signature of the BLM's protein. Our in-silico analysis suggests that the BLM's HRDC domain has ability of parallel aggregation and fibril formation. Moreover, structural similarity with proteins like α-synuclein, TDP-43, and Tau and its interactions with PARP1 suggests it may have a role in neurodegenerative diseases. Furthermore, we identified deleterious mutations in the HRDC domain of BLM protein that may compromise its stability and alter its function. Hence, these findings suggest that in addition to BLM's well-known functions, the protein may have ability to form parallel aggregation and fibril formation and its role in neurodegenerative disease need to further be explored.

Author Summary:

This paper higlights the importance of considering the structural and physical characteristics of the BLM protein, which may have been previously overlooked in the context of disease pathology. Through in-silico analyses, we identified that the BLM protein possesses residues within the HRDC domain that are prone to

parallel aggregation and fibril formation. Moreover, the structural similarity of BLM to proteins such as α-synuclein, TDP-43, and Tau indicates a potential role for BLM in neurodegenerative processes. Additionally, we identified deleterious mutations within the HRDC domain that could compromise the stability of the protein. Furthermore, our findings suggest that the association between BLM and PARP1 may involve a regulatory mechanism that could significantly influence BLM's function.

Keywords

BLM; PARP1; *in-silico* analysis; Single nucleotide polymorphism; Intrinsically disordered proteins; neurodegenerative diseases

Introduction

BLM's is well-known for its role in repairing DNA double-strand breaks through homologous recombination. It plays a key part in DNA end resection, D-loop formation, Rad51 filament formation, branch migration, and the dissolution of Holliday junctions [1]. Additionally, BLM efficiently unwinds DNA secondary structures, with a particularly strong ability to unwind G-quadruplexes, making it one of the most effective human DNA helicases in this regard. Recent research has revealed that BLM also binds to RNA, showing a preference for binding to RNA G-quadruplexes (G4s) [2]. RNA binding proteins (RBPs) are known to have role in neurodegenerative disease, and they also share a binding preference for G-quadruplexes (G4s) [3].

Neurodegenerative diseases are associated to a group of proteins referred as intrinsically disordered proteins (IDPs) [4] . IDPs are a normal class of proteins, but unlike most proteins that fold into specific three-dimensional structures, IDPs or proteins with intrinsically disordered regions (IDRs) do not have a fixed structure under physiological conditions. This lack of a defined structure allows them to be highly flexible and versatile, enabling them to interact with multiple different partners and play key roles in various biological processes. Alterations in the cellular environment or mutations within IDPs can interfere with their normal functions, leading to protein misfolding and the formation of aggregates or fibrils [5, 6]. Studies have reported the association of IDPs/IDRs proteins such as TAR DNA-binding protein (TDP-43) in amyotrophic lateral sclerosis (ALS), α-synuclein (α-syn) in Parkinson's disease (PD), and amyloid beta (Aβ) and tau for Alzheimer's disease (AD) [7].

Protein posttranslational modifications (PTMs) of IDPs/IDRs also determine the formation of pathologic aggregation. One of the known PTMs involved in neurodegeneration is poly (ADP-ribosylation) (PARylation). A family of PARP enzymes performs PARylation that attaches multiple NAD-derived ADP-ribose (ADPr) units to the acceptor protein [8]. The number of cellular processes such as gene expression, cell death pathway, DNA repair, mitochondrial biogenesis, and neuroinflammation, are regulated by PARylation and PAR-binding [9-15]. PARP1 activity is known to be elevated in the brains of individuals with neurodegenerative diseases, including Alzheimer's disease. Several proteins involved in Alzheimer's disease (AD) and amyotrophic lateral sclerosis (ALS), such as α-synuclein, TDP-43, and hnRNP A1, are regulated by PARylation and PAR-binding. These modifications influence the phase separation properties and aggregation tendencies of these proteins [16-18].

In this study, we utilized in-silico tools to explore the functional aspects of BLM, particularly its potential to form aggregation, which may contribute to the pathology of neurodegenerative diseases. Using the PASTA 2.0 tool, we identified amino acid residues within the HRDC domain of the BLM protein that are likely to undergo parallel aggregation and fibril formation. Notably, our findings reveal significant similarities between BLM and intrinsically disordered proteins (IDPs) implicated in neurodegenerative diseases. Moreover, the mutation identified in the HRDC domain of BLM impacts its stability, further suggesting it may have a potential role in disease pathology. Additionally, our study uncovered critical associations between BLM and PARP1. Based on these insights, we suggest that the HRDC domain of BLM possesses aggregation and fibril formation capabilities, which may contribute to the pathophysiology of neurodegenerative diseases.

Results

BLM's Expression Across Cell Types and Brain Regions

The human BLM belongs to the RecQ helicase family of proteins consisting of 1417 amino acids encoded by the BLM gene on chromosome 15 (UniProtKB ID: P54132) and has a molecular weight of approximately 159 kDa **(Fig 1A).** This protein has a conserved ATP-dependent DNA helicase domain ranging from 657-865 amino acids. BLM also has RecQ C-terminal (RQC; 1072-1194) and helicase-and-ribonuclease D-C-terminal (HRDC; 1212-1292) domains. HRDC domain has single-stranded DNA (ssDNA) and the DNA Holliday junction binding residues between 1227-1244 amino acids. The unstructured regions of the protein

include the N-terminus (residues 1–657), which is crucial for interaction and oligomerization with partner proteins, and the C-terminus (residues 1293–1417), which has nuclear localization signals and is crucial for the nuclear function of BLM.

Gene Ontology (GO) analyses [16, 19, 20] indicated that BLM participates in diverse regulatory complexes in the cytoplasm and nucleus **Fig 1B**. Mostly its function has been reported on cellular metabolic processes such as DNA repair, and replication. Analysis of its expression pattern revealed its highest expression in serous glandular cells (**Fig 1C**). Oligodendrocyte precursor cells were also among the cells that had the third highest level of BLM gene expression after monocytes. On comparing the expression level in different regions of the brain its expression was found to be highest in the cerebellum. The expression of BLM at the protein level has not been identified in some regions of the brain even after its expression at the transcript level **Fig 1D and E.** Hence from the above finding we can suggest that BLM is expressed in neuronal cells and it may have a role in neuronal development and its function.

BLM's Parallel Aggregation and Fibril Formation Ability

On computing the physical and chemical parameters of BLM using ProtParam it was classified as an unstable protein with an instability index of 44.57 (smaller than 40 is considered as stable protein). The aliphatic index of BLM is 73.83, implying that the protein is thermostable. The GRAVY score calculated by the server was -0.602, indicating that the protein is hydrophilic. The estimated half-life is 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), >10 hours (Escherichia coli, in vivo) in the respective cell types. PASTA 2.0 predicted that a significant region of BLM contains 38.96% alpha-helical structure, 12.07% forms the beta-strands secondary structure, 48.98% forms coils and about 23.42% of the region remains disordered. According to PASTA 2.0, the BLM protein's N-terminal residues 1-21, 39-42, 72-75, 92-107, 142-175, 183-232, 247-288, 290, 291, 313, and 323–339 was found to be disordered, while residues 1224–1250 show a probability of parallel aggregation and fibril formation (**Fig 2A**). Additionally, the C-terminal residues 1277-1417 were found to be disordered. This suggests that BLM has intrinsically disordered regions and has the ability of parallel aggregation and fibril formation.

BLM's Similarity with TDP43, Tau, and A-Synuclein

Similar to TDP-43, α-syn, Aβ, and Tau, BLM (Bloom syndrome protein), also carries intrinsically disordered regions. By comparing the physical and chemical parameters of BLM with TDP43, Tau, and α-

synuclein, we demonstrated that all are unstable with an instability index of 44.57, 56.76, and 56.02 respectively except α-synuclein which has an instability index of 25.47. Also, BLM, TDP43, Tau, and α-synuclein are all hydrophilic proteins with GRAVY scores of -0.602, -0.535, -0.882, and -0.403 respectively.

We used PSIPRED and DISOPRED3 webservers [21] to predict the secondary structure and intrinsically disordered regions of BLM, TDP43, Tau, and α-synuclein. The prediction revealed the presence of structural domains like α-helix and beta-strands strands in the ordered region of these proteins. Interestingly we noticed all had a disordered C-terminus region. Beta-strands were identified as regions in BLM protein with a threshold <0.5 to form parallel aggregation using the PASTA 2.0 tool **(Fig 2).** Hence the ability of BLM protein to form parallel aggregation is similar to TDP43, Tau, and α-synuclein which are commonly involved in neurodegenerative diseases like amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), and Alzheimer's disease (AD) **(Table 1**).

Sr. No.	**Protein**	**Disordered region (%)**	**α-helix (%)**	**beta-strands (%)**	**Coil region (%)**
1.	BLM	23.42	38.96	12.07	48.98
2.	TDP43	34.05	16.18	21.5	61.32
3.	TAU	77.83	8.31	13.32	78.36
4.	α-	34.28	26.43	22.14	51.43

Table 1. Structural Composition and Secondary Structure Elements of BLM, TDP43, Tau, and α-Synuclein Proteins.

	Synuclein				

Identification Of Deleterious and Damaging nsSNPs in BLM

Mutations in BLM can affect the stability of protein hence to identify the most potentially harmful nsSNPs in BLM, we used the ENSEMBL browser displaying 44201 genetic variations (Transcript: ENST00000355112.8 (BLM-201) - Variant table - Homo sapiens - Ensembl genome browser 112). Furthermore, these SNPs were further validated by the NCBI Variation Viewer and dbSNP database (BLM: Chr15 - Variation Viewer (nih.gov). Using both of these tools we identified 40919 single nucleotide variants of the BLM gene. Using the genomic coordinates of the "NM_000057.4" transcript of the BLM helicase isoform 1, we identified 2647 SNPs in the coding region sequence. These included missense variants (non-synonymous SNPs), 123 stop-gained variants, and 2 start-lost nsSNPs, as illustrated in **Fig 3**. The remaining nsSNPs were filtered based on the standard scoring of various sequence-based consensus tools: SIFT (score <0.05) [22]; PolyPhen-2 (score>0.95 - probably harmful, score>0.5 – possibly damaging) [23]; Mutation Assessor (score>0.5); Meta-SNP, PhD-SNP, SNPs & Go (score>0.5) [24]. The filtered nsSNPs using the above tools were further validated using PredictSNP [25] & SNAP [26] to determine the accuracy of deleterious mutations.

We further examined the highly conserved HRDC (helicase and RNase D C-terminal) domain that was found to form parallel aggregations (**Fig 2A**). A total of 64 nsSNPs was found in the HRDC domain and based on the score predicted by the above sequence-based tools, 8 nsSNPS (**Table 2**) (H1236Y, F1238L, L1253S, T1267I, E1268V, E1272G, V1278M, S1286) was classified as predominantly deleterious.

Table 2. List of deleterious missense SNPs in the HRDC domain of BLM gene using different sequence-based tools.

Sr. No.	**#rsID**	**AA substitution**	**SIFT**	**PolyPhen 2**	**Mutation Accesso**	**PhD-SNP**	**SNPs & GO**	**Meta SNP**	**Predict SNP**	**SNAP**

					r					
1	rs1897237440	H1236Y	Deleterious 0	Possibly Damaging 0.851	Medium 0.62	Disease 0.733	Pathogenic 0.719	Disease 0.739	Disease	Disease
2	rs1897237692	F1238L	Deleterious 0	Benign 0.093	Medium 0.811	Disease 0.673	Pathogenic 0.654	Disease 0.668	-	Disease
3	rs2151198266	L1253S	Deleterious 0.01	Possibly Damaging 0.662	Medium 0.863	Neutral 0.342	Pathogenic 0.654	Neutral 0.367	-	Disease
4	rs2151198306	T1267I	Deleterious 0	Probably Damaging 0.972	Medium 0.671	Disease 0.657	Pathogenic 0.624	Disease 0.73	Disease	Disease
5	rs1897348799	E1268V	Deleterious 0	Possibly Damaging 0.583	Medium 0.757	Disease 0.576	-	Disease 0.529	Disease	Disease
6	rs2151198318	E1272G	Deleterious 0	Possibly Damaging 0.861	Medium 0.781	Neutral 0.491	-	Disease 0.605	Disease	Disease
7	rs753563391	V1278M	Deleterious 0.02	Possibly Damaging 0.662	Low 0.448	Disease 0.504	-	Disease 0.544	Disease	Disease
8	rs1897350617	S1286P	Deleterious 0	Possibly Damaging 0.525	Medium 0.732	Neutral 0.46	Pathogenic 0.463	Disease 0.501	-	Disease

Various studies have shown that most disease-related missense mutations alter the stability of proteins. We analyzed these nsSNPs for amino acid substitutions and their impact on the stability of mutant BLM protein using I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) [27], MuPro

(https://mupro.proteomics.ics.uci.edu) [28], PremPS (https://lilab.jysw.suda.edu.cn/research/PremPS/) [29] and PoPMuSiC v3.1 2.1 (http://babylone.ulb.ac.be/PoPMuSiC v3.1) [30].

Currently, the whole protein 3D structure of the BLM gene is not available, hence, 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. Using I-Mutant 2.-0 and MuPro, respectively we found that out of eight nsSNPs only three amino acid substitutions (H1236Y, E1268V, and T1267I) increased protein stability and remaining (F1238L, L1253S, E1272G, V1278M, S1286) decreased the stability of the protein. In contrast, the PremPS and PoPMuSiC v3.1 predicted that all nsSNPs destabilize the protein structure **(Table 3)**.

MutPred2 was further utilized to predict the structural, molecular, and phenotypic impacts of amino acid variations. The MutPred2 [31] webserver identified 6 out of the 8 shortlisted deleterious nsSNPs (H1236Y, F1238L, L1253S, T1267I, E1268V, and E1272G) as likely to cause structural and functional changes, with prediction scores exceeding 0.50. These changes impacted various properties, including alterations in ordered interfaces, transmembrane proteins, protein stability, metal-binding sites, and DNA-binding sites. Additionally, they led to the loss or gain of allosteric sites, catalytic and proteolytic cleavage sites, as well as alterations in post-translational modifications at different residues. (as shown in **Table 4**).

Table 3: Impact of amino acid substitutions on BLM protein stability

AA substitution	**I-Mutant 2.0**		**MuPro**		**PremPS**		**PoPMuSiC v3.1**	
	DDG (kcal/mol)	**Stability**	**DDG (kcal/mol)**	**Stability**	**ΔΔG**	**Stability**	**ΔΔG**	**Stability**
H1236Y	0.65	Increase	-0.128	Decrease	0.23	Destabilizing	1.13	Destabilizing
F1238L	-3.24	Decrease	-0.67	Decrease	1.52	Destabilizing	1.65	Destabilizing
L1253S	-1.92	Decrease	-1.79	Decrease	3.39	Destabilizing	2.29	Destabilizing
T1267I	-0.48	Decrease	0.209	Increase	0.49	Destabilizing	1.8	Destabilizing
E1268V	0.19	Increase	-0.365	Decrease	0.2	Destabilizing	0.56	Destabilizing
E1272G	-0.78	Decrease	-1.562	Decrease	0.31	Destabilizing	0.31	Destabilizing
V1278M	-0.67	Decrease	-0.727	Decrease	0.51	Destabilizing	0.47	Destabilizing
S1286P	-0.79	Decrease	-0.813	Decrease	0.64	Destabilizing	0.41	Destabilizing

Table 4: List of deleterious nsSNPs and the resulting structural and functional alterations in BLM protein.

Sr. No.	Variant ID	Substitutions	MutPred2 Score	Molecular Mechanism
1.	rs1897237440	H1236Y	0.8	Altered Ordered interface
2.	rs1897237692	F1238L	0.681	Loss of Loop
3.	rs2151198266	L1253S	0.628	Altered Metal-binding, Altered Ordered interface, Altered Transmembrane protein, Gain of B-factor, Gain of Intrinsic disorder, Gain of Ubiquitylation at K1248
4.	rs2151198306	T1267I	0.783	Altered Coiled coil, Altered Metal-binding, Altered Transmembrane protein, Gain of Acetylation at K1270, Gain of Methylation at K1270, Loss of Relative solvent accessibility
5.	rs1897348799	E1268V	0.767	Altered Coiled coil, Altered DNA binding, Altered Metal-binding, Altered Transmembrane protein, Gain of Acetylation at K1273, Gain of Relative solvent accessibility, Loss of Methylation at K1270, Loss of SUMOylation at K1270
6.	rs2151198318	E1272G	0.841	Altered Coiled coil, Altered DNA binding, Altered Metal-binding, Altered Transmembrane protein, Gain of Acetylation at K1273, Gain of Methylation at K1270, Loss of Relative solvent accessibility, Loss of SUMOylation at K1270

Association of BLM with PAPR1

STRING analysis was used to profile the protein-protein interactions involving BLM and its associated interacting proteins **Fig 4A.** Among the total 236 interactomes, PARP1 was noticed as its activity is known

to be elevated in the brains of individuals with neurodegenerative diseases. We identified important associations between BLM and PARP1, an enzyme that catalyzes the addition of poly (ADP-ribose) (PAR) on target proteins. BLM and PARP1 showed similar expression profiles during cerebral cortex development **Fig 4B**. We observed both correlated co-expression and divergent expression patterns of BLM and PARP1 in normal brain tissue, underscoring their significant association, as illustrated in **Fig 4C**. Also, BLM contains a PARPylation site (1303-1333) [32] in the intrinsically disordered regions, between the HRDC and NLS sequence. By comparing the result of corGSEA between BLM and PARP1, we found differential enrichment of several pathways implicated in neurodegenerative diseases, including “Alzheimer's Disease”, “Lu aging brain down”, “Kim all disorders duration corr down”, and “Mody Hippocampus postnatal” **Fig 4D**. This suggests the possible association of BLM and PARP1 protein in relation to neurodegenerative diseases.

Discussion

BLM belongs to the RecQ helicase family and carries an evolutionary conserved 3’ to 5’ DNA helicase activity which unwinds the double-helix DNA to produce single-stranded DNA as an intermediate required for the processes of DNA replication, repair, and recombination [33]. Immunohistochemical analysis of BLM expression across human developmental stages reveals its presence in various cell types, including Purkinje cells, neurons, thymic Hassall’s corpuscles, pancreatic beta-cells, and testicular sperm cells, with patterns varying by age and tissue. In a BS patient, BLM expression was absent in pancreatic and testicular tissues, suggesting that BLM plays a critical role in neuronal development, immune function, insulin secretion, and spermatogenesis, aligning with the major symptoms of BS [34].

Diseases associated with proteins that contain intrinsically disordered regions (IDRs) include, Neurodegenerative Diseases (Alzheimer's Disease due to Tau protein and Amyloid-beta), Parkinson's Disease (due to α-Synuclein), Huntington's Disease (due to Huntingtin protein), Amyotrophic Lateral Sclerosis due to TDP-43 and FUS) [4]. These diseases are often linked to the misfolding, aggregation, or aberrant interactions of these intrinsically disordered proteins, contributing to the pathogenesis of the disease [35]. It has been shown that the proteins that have IDRs have a strong tendency to form protein aggregation, although IDR itself is not the only cause of aggregate formation [36]. Proteins such as TAR DNA-binding protein (TDP-43) in amyotrophic lateral sclerosis (ALS), α-synuclein (α-syn) in Parkinson's disease (PD),

and amyloid beta (Aβ) and tau for Alzheimer's disease (AD) form aggregates [7]. BLM has both N-terminal and C-terminal intrinsically disordered regions [37] and HRDC domain was found to have the ability of aggregation and fibril formation similar to IDPs/IDRs proteins (**Fig 2**) (**Table 1**). Furthermore, the HRDC domain of BLM contains several non-synonymous single nucleotide polymorphisms (nsSNPs) that alter the structure or stability of the protein. Our analysis suggests various SNPs of BLM including F1238L, L1253S, and E1272G mutations might destabilize the protein based on changes in Gibbs free energy and folding energy. Additionally, these deleterious nsSNPs found in the HRDC domain may affect its structure, function, and aggregation-forming ability.

BLM interacts with PARP1, and both are involved in DNA repair mechanisms e.g. repair of double-strand break through homologous recombination and resolution of DNA replication stress. PARP1 has been identified as a BLM promoter-binding protein that negatively regulates the transcription of BLM in prostate cancer [38]. Moreover, our findings also indicate that BLM and PARP1 have a similar expression pattern in cortical development (**Fig 4**), suggesting that BLM complexes or related pathways may be modulated by PARylation. Further studies are needed to elucidate the molecular mechanisms of PARylation's role in the intrinsically disordered region (IDR) of BLM and its impact on aggregation formation by the HRDC domain, which may have a role in neurodegenerative diseases. In brief, this is the first study that provides foundational knowledge about the HRDC domain of the BLM protein, highlighting its ability to form parallel aggregations and fibril formation. This new dimension could provide insights into its unexplored role in the pathology of neurodegenerative diseases.

Material and Methods

The application of computational methods to gain biological insights is widely recognized and highly valued. Previous studies have demonstrated that utilizing an array of advanced tools and algorithms can significantly improve prediction accuracy. To ensure maximum precision, we employed several computational algorithms specifically designed for predicting aggregation propensity, identifying similarities to intrinsically disordered proteins (IDPs), and analyzing non-synonymous single nucleotide polymorphisms (nsSNPs) in BLM helicases, given their disease relevance. The in-silico tools utilized in this study include ProtParam, Expasy server, PASTA 2.0, PSIPRED, SIFT, PROVEAN, PolyPhen-2, SNAP2, SNP&GO, nsSNPAnalyzer, and MutPred-2. The Cortecon [39] database was used to identify the shared

similarities between BLM and PARP1 expression profiling during in vitro cortex development. This comprehensive approach allowed to analyze the structural features of BLM protein and to identify disease-associated mutations in BLM genes.

Amino Acid Sequence, Expression, Structural, and Physicochemical Properties Of BLM

The human BLM protein sequence and its other basic information were acquired from the National Center for Biotechnology Information (NCBI) (Protein ID: NP_000048.1) and the UniProt database (UniProtKB ID: P54132). We extracted RNA and protein expression data for BLM specifically in brain tissue from The Human Protein Atlas (HPA- The Human Protein Atlas) database. Comparative analysis of RNA expression (normalized transcripts per million, nTPM) and protein expression levels across various brain tissues was conducted and side-by-side bar graphs were generated to visualize the RNA and protein expression levels of BLM in these brain tissues. We used a computational tool PASTA 2.0 (Aggregation Prediction with PASTA (un6ipd.it) for the prediction of disorder and aggregation propensity in BLM protein [40]. We also used the PSIPRED and DISOPRED3 server (PSIPRED Workbench (ucl.ac.uk) to predict the secondary structure of BLM proteins based on their amino acid sequences. PSIPRED server employs neural networks and position-specific scoring matrices (PSSMs) to predict regions of the protein that are likely to form secondary structure [21]. The physicochemical parameters of the human BLM protein were calculated using the ProtParam tool available on the Expasy server (Expasy - ProtParam tool).

Identification and Analysis of Potent nsSNPs of BLM

ENSEMBL genome browser 112 (Ensembl genome browser 112) and dbSNP database was used to retrieve nsSNPs in BLM gene using genomic coordinates “NM_000057.4” for transcript of the BLM helicase. We used the sequence-based tools such as SIFT (Sorting Intolerant From Tolerant; https://sift.bii.a-star.edu.sg/) [22], PolyPhen-2 (Polymorphism phenotyping v2; http://genetics.bwh.harvard.edu/pph2/) [23], Mutation Assessor, PhD-SNP (Predictor of human Deleterious Single Nucleotide Polymorphisms; http://snps.biofold.org/phd-snp/phd-snp), SNPs & GO (https://snps.biofold.org/snps-and-go/), Meta-SNP (http://snps.biofold.org/meta-snp/), and Screening of non-acceptable polymorphisms (SNAP) to predict the function of nonsynonymous single nucleotide polymorphism (nsSNP) [41].

SIFT (score <0.05); PolyPhen-2 (score>0.95 - probably harmful); Meta-SNP, PhD-SNP, SNPs & Go (score>0.5), PredictSNP [25] and SNAP [26] was used to determine the accuracy of the damaging effect of amino acid substitution.

Predicting the Effect of nsSNPs on the Stability of BLM

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) [27] and MuPro (https://mupro.proteomics.ics.uci.edu) [28], were used to predict change in the stability of BLM upon a single nucleotide change using protein sequence as input. The associated change in free energy (ΔΔG, in kcal/mol) and stability were analyzed for the shortlisted SNPs. A negative ΔΔG value indicates a decrease in stability, while a positive ΔΔG value signifies an increase in stability. PremPS (https://lilab.jysw.suda.edu.cn/research/PremPS/) [29] and PoPMuSiC v3.1 2.1 (http://babylone.ulb.ac.be/PoPMuSiC v3.1) [30] was also used to predict change in thermodynamic stability of BLM upon a single nucleotide change. Pdb file or pdb ID was used as input file respectively by PremPS and PoPMuSiC to evaluate the effects of single mutations on protein stability by calculating the changes in unfolding and folding Gibbs free energy (ΔΔG). A positive and negative sign corresponds to destabilizing and stabilizing mutations respectively.

Structural and Functional Changes in BLM Caused by Deleterious nsSNPs

MutPred2 (http://mutpred.mutdb.org), a popular machine learning method, is used to detect pathogenic variants as well as the structural and functional changes brought on by harmful nsSNPs [31]. It predicts their effects on protein features, enabling it to categorize them as pathogenic or benign. A list of amino acid substitutions in the relevant FASTA headers was utilized as input along with a protein sequence in FASTA format. A P-value cutoff of 0.5 and a prediction score ranging from 0.5 to 1 was employed to identify the structure and functional alterations caused by deleterious nsSNPs of BLM.

Gene Ontology and Network Analysis of BLM-associated Proteins

BLM interactors were downloaded from BioGRID (v4.4) [12], and STRING v12.0 (the Search Tool for Retrieval of Interacting Genes and Proteins) [42] database and was used to assess its interacting genes and the protein network. Gene Ontology (GO) enrichment analysis was performed using SRPlot [19] to identify and interpret the biological functions, processes, and cellular components that are significantly associated with BLM. Finally, the gene expression correlation was analyzed using Correlation AnalyzeR [43].

Financial Disclosure Statement

This work was supported by a Department of Biotechnology, India Ramalingaswami Re-Entry fellowship grant (BT/RLF/Re-entry/37/2021).

Bibliography

1. Chu WK, Hickson ID. RecQ helicases: multifunctional genome caretakers. Nat Rev Cancer. 2009;9(9):644-54. Epub 20090806. doi: 10.1038/nrc2682. PubMed PMID: 19657341.
2. Danino YM, Molitor L, Rosenbaum-Cohen T, Kaiser S, Cohen Y, Porat Z, et al. BLM helicase protein negatively regulates stress granule formation through unwinding RNA G-quadruplex structures. Nucleic Acids Res. 2023;51(17):9369-84. doi: 10.1093/nar/gkad613. PubMed PMID: 37503837; PubMed Central PMCID: PMCPMC10516661.
3. Cave JW, Willis DE. G-quadruplex regulation of neural gene expression. FEBS J. 2022;289(12):3284-303. Epub 20210510. doi: 10.1111/febs.15900. PubMed PMID: 33905176.
4. Uversky VN. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. Front Aging Neurosci. 2015;7:18. Epub 20150302. doi: 10.3389/fnagi.2015.00018. PubMed PMID: 25784874; PubMed Central PMCID: PMCPMC4345837.
5. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. Biophys J. 2007;92(5):1439-56. Epub 20061208. doi: 10.1529/biophysj.106.094045. PubMed PMID: 17158572; PubMed Central PMCID: PMCPMC1796814.
6. Oldfield CJ, Dunker AK. Intrinsically disordered proteins and intrinsically disordered protein regions. Annu Rev Biochem. 2014;83:553-84. Epub 20140305. doi: 10.1146/annurev-biochem-072711-164947. PubMed PMID: 24606139.
7. Tsoi PS, Quan MD, Ferreon JC, Ferreon ACM. Aggregation of Disordered Proteins Associated with Neurodegeneration. Int J Mol Sci. 2023;24(4). Epub 20230208. doi: 10.3390/ijms24043380. PubMed PMID: 36834792; PubMed Central PMCID: PMCPMC9966039.
8. McRae EKS, Dupas SJ, Booy EP, Piragasam RS, Fahlman RP, McKenna SA. An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21. RNA. 2020;26(1):44-57. Epub 20191025. doi: 10.1261/rna.072199.119. PubMed PMID: 31653714; PubMed Central PMCID: PMCPMC6913123.
9. Kelaini S, Chan C, Cornelius VA, Margariti A. RNA-Binding Proteins Hold Key Roles in Function, Dysfunction, and Disease. Biology (Basel). 2021;10(5). Epub 20210424. doi: 10.3390/biology10050366. PubMed PMID: 33923168; PubMed Central PMCID: PMCPMC8146904.
10. Bryant CD, Yazdani N. RNA-binding proteins, neural development and the addictions. Genes Brain Behav. 2016;15(1):169-86. doi: 10.1111/gbb.12273. PubMed PMID: 26643147; PubMed Central PMCID: PMCPMC4944654.
11. Baudin A, Moreno-Romero AK, Xu X, Selig EE, Penalva LOF, Libich DS. Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes. Front Mol Biosci. 2021;8:744707. Epub 20210924. doi: 10.3389/fmolb.2021.744707. PubMed PMID: 34631798; PubMed Central PMCID: PMCPMC8497785.
12. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34(Database issue):D535-9. doi: 10.1093/nar/gkj109. PubMed PMID: 16381927; PubMed Central PMCID: PMCPMC1347471.
13. Prashad S, Gopal PP. RNA-binding proteins in neurological development and disease. RNA Biol. 2021;18(7):972-87. Epub 20200830. doi: 10.1080/15476286.2020.1809186. PubMed PMID: 32865115; PubMed Central PMCID: PMCPMC8216196.
14. Martini S, Davis K, Faraway R, Elze L, Lockwood N, Jones A, et al. A genetically-encoded crosslinker screen identifies SERBP1 as a PKCepsilon substrate influencing translation and cell division. Nat Commun. 2021;12(1):6934. Epub 20211126. doi: 10.1038/s41467-021-27189-5. PubMed PMID: 34836941; PubMed Central PMCID: PMCPMC8626422.
15. Marcelino Meliso F, Hubert CG, Favoretto Galante PA, Penalva LO. RNA processing as an alternative route to attack glioblastoma. Hum Genet. 2017;136(9):1129-41. Epub 20170612. doi: 10.1007/s00439-017-1819-2. PubMed PMID: 28608251; PubMed Central PMCID: PMCPMC6685205.
16. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7):e21800. Epub 20110718. doi: 10.1371/journal.pone.0021800. PubMed PMID: 21789182; PubMed Central PMCID: PMCPMC3138752.
17. Su Q, Mehta S, Zhang J. Liquid-liquid phase separation: Orchestrating cell signaling through time and space. Mol Cell. 2021;81(20):4137-46. Epub 20211006. doi: 10.1016/j.molcel.2021.09.010. PubMed PMID: 34619090; PubMed Central PMCID: PMCPMC8541918.
18. Nakanishi C, Seimiya H. G-quadruplex in cancer biology and drug discovery. Biochem Biophys Res Commun. 2020;531(1):45-50. Epub 20200417. doi: 10.1016/j.bbrc.2020.03.178. PubMed PMID: 32312519.
19. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: A free online platform for data visualization and graphing. PLoS One. 2023;18(11):e0294236. Epub 20231109. doi: 10.1371/journal.pone.0294236. PubMed PMID: 37943830; PubMed Central PMCID: PMCPMC10635526.
20. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. Epub 20190403. doi: 10.1038/s41467-019-09234-6. PubMed PMID: 30944313; PubMed Central PMCID: PMCPMC6447622.

21. Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 2019;47(W1):W402-W7. doi: 10.1093/nar/gkz297. PubMed PMID: 31251384; PubMed Central PMCID: PMCPMC6602445.
22. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40(Web Server issue):W452-7. Epub 20120611. doi: 10.1093/nar/gks539. PubMed PMID: 22689647; PubMed Central PMCID: PMCPMC3394338.
23. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20. doi: 10.1002/0471142905.hg0720s76. PubMed PMID: 23315928; PubMed Central PMCID: PMCPMC4480630.
24. Capriotti E, Calabrese R, Fariselli P, Martelli PL, Altman RB, Casadio R. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. BMC Genomics. 2013;14 Suppl 3(Suppl 3):S6. Epub 20130528. doi: 10.1186/1471-2164-14-S3-S6. PubMed PMID: 23819482; PubMed Central PMCID: PMCPMC3665478.
25. Bendl J, Stourac J, Salanda O, Pavelka A, Wieben ED, Zendulka J, et al. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS Comput Biol. 2014;10(1):e1003440. Epub 20140116. doi: 10.1371/journal.pcbi.1003440. PubMed PMID: 24453961; PubMed Central PMCID: PMCPMC3894168.
26. Pers TH, Timshel P, Hirschhorn JN. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. Bioinformatics. 2015;31(3):418-20. Epub 20141013. doi: 10.1093/bioinformatics/btu655. PubMed PMID: 25316677; PubMed Central PMCID: PMCPMC4308663.
27. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33(Web Server issue):W306-10. doi: 10.1093/nar/gki375. PubMed PMID: 15980478; PubMed Central PMCID: PMCPMC1160136.
28. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins. 2006;62(4):1125-32. doi: 10.1002/prot.20810. PubMed PMID: 16372356.
29. Chen Y, Lu H, Zhang N, Zhu Z, Wang S, Li M. PremPS: Predicting the impact of missense mutations on protein stability. PLoS Comput Biol. 2020;16(12):e1008543. Epub 20201230. doi: 10.1371/journal.pcbi.1008543. PubMed PMID: 33378330; PubMed Central PMCID: PMCPMC7802934.
30. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. BMC Bioinformatics. 2011;12:151. Epub 20110513. doi: 10.1186/1471-2105-12-151. PubMed PMID: 21569468; PubMed Central PMCID: PMCPMC3113940.
31. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam HJ, et al. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. Nat Commun. 2020;11(1):5918. Epub 20201120. doi: 10.1038/s41467-020-19669-x. PubMed PMID: 33219223; PubMed Central PMCID: PMCPMC7680112.
32. Gibson BA, Zhang Y, Jiang H, Hussey KM, Shrimp JH, Lin H, et al. Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. Science. 2016;353(6294):45-50. Epub 20160602. doi: 10.1126/science.aaf7865. PubMed PMID: 27256882; PubMed Central PMCID: PMCPMC5540732.
33. Mohaghegh P, Karow JK, Brosh RM, Jr., Bohr VA, Hickson ID. The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. Nucleic Acids Res. 2001;29(13):2843-9. doi: 10.1093/nar/29.13.2843. PubMed PMID: 11433031; PubMed Central PMCID: PMCPMC55766.
34. Hachiya Y, Motonaga K, Itoh M, Masuko T, Enomoto T, Sonobe H, et al. Immunohistochemical expression and pathogenesis of BLM in the human brain and visceral organs. Neuropathology. 2001;21(2):123-8. doi: 10.1046/j.1440-1789.2001.00379.x. PubMed PMID: 11396677.
35. Babu MM. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. Biochem Soc Trans. 2016;44(5):1185-200. doi: 10.1042/BST20160172. PubMed PMID: 27911701; PubMed Central PMCID: PMCPMC5095923.
36. Uemura E, Niwa T, Minami S, Takemoto K, Fukuchi S, Machida K, et al. Large-scale aggregation analysis of eukaryotic proteins reveals an involvement of intrinsically disordered regions in protein folding. Sci Rep. 2018;8(1):678. Epub 20180112. doi: 10.1038/s41598-017-18977-5. PubMed PMID: 29330519; PubMed Central PMCID: PMCPMC5766493.
37. Bereda CC, Dewey EB, Nasr MA, Sekelsky J. Functions of the Bloom Syndrome Helicase N-terminal Intrinsically Disordered Region. bioRxiv. 2024. Epub 20240415. doi: 10.1101/2024.04.12.589165. PubMed PMID: 38659896; PubMed Central PMCID: PMCPMC11042211.
38. Huang M, Chen L, Guo Y, Ruan Y, Xu H. PARP1 negatively regulates transcription of BLM through its interaction with HSP90AB1 in prostate cancer. J Transl Med. 2023;21(1):445. Epub 20230706. doi: 10.1186/s12967-023-04288-z. PubMed PMID: 37415147; PubMed Central PMCID: PMCPMC10324254.
39. van de Leemput J, Boles NC, Kiehl TR, Corneo B, Lederman P, Menon V, et al. CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron. 2014;83(1):51-68. doi: 10.1016/j.neuron.2014.05.013. PubMed PMID: 24991954.
40. Walsh I, Seno F, Tosatto SC, Trovato A. PASTA 2.0: an improved server for protein aggregation prediction. Nucleic Acids Res. 2014;42(Web Server issue):W301-7. Epub 20140521. doi: 10.1093/nar/gku399. PubMed PMID: 24848016; PubMed Central PMCID: PMCPMC4086119.
41. Capriotti E, Altman RB, Bromberg Y. Collective judgment predicts disease-associated single nucleotide variants. BMC Genomics. 2013;14 Suppl 3(Suppl 3):S2. Epub 20130528. doi: 10.1186/1471-2164-14-S3-S2. PubMed PMID: 23819846; PubMed Central PMCID: PMCPMC3839641.

05 42. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein
06 association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res.
07 2023;51(D1):D638-D46. doi: 10.1093/nar/gkac1000. PubMed PMID: 36370105; PubMed Central PMCID: PMCPMC9825434.
08 43. Miller HE, Bishop AJR. Correlation AnalyzeR: functional predictions from gene co-expression correlations. BMC
09 Bioinformatics. 2021;22(1):206. Epub 20210420. doi: 10.1186/s12859-021-04130-7. PubMed PMID: 33879054; PubMed Central
10 PMCID: PMCPMC8056587.

11

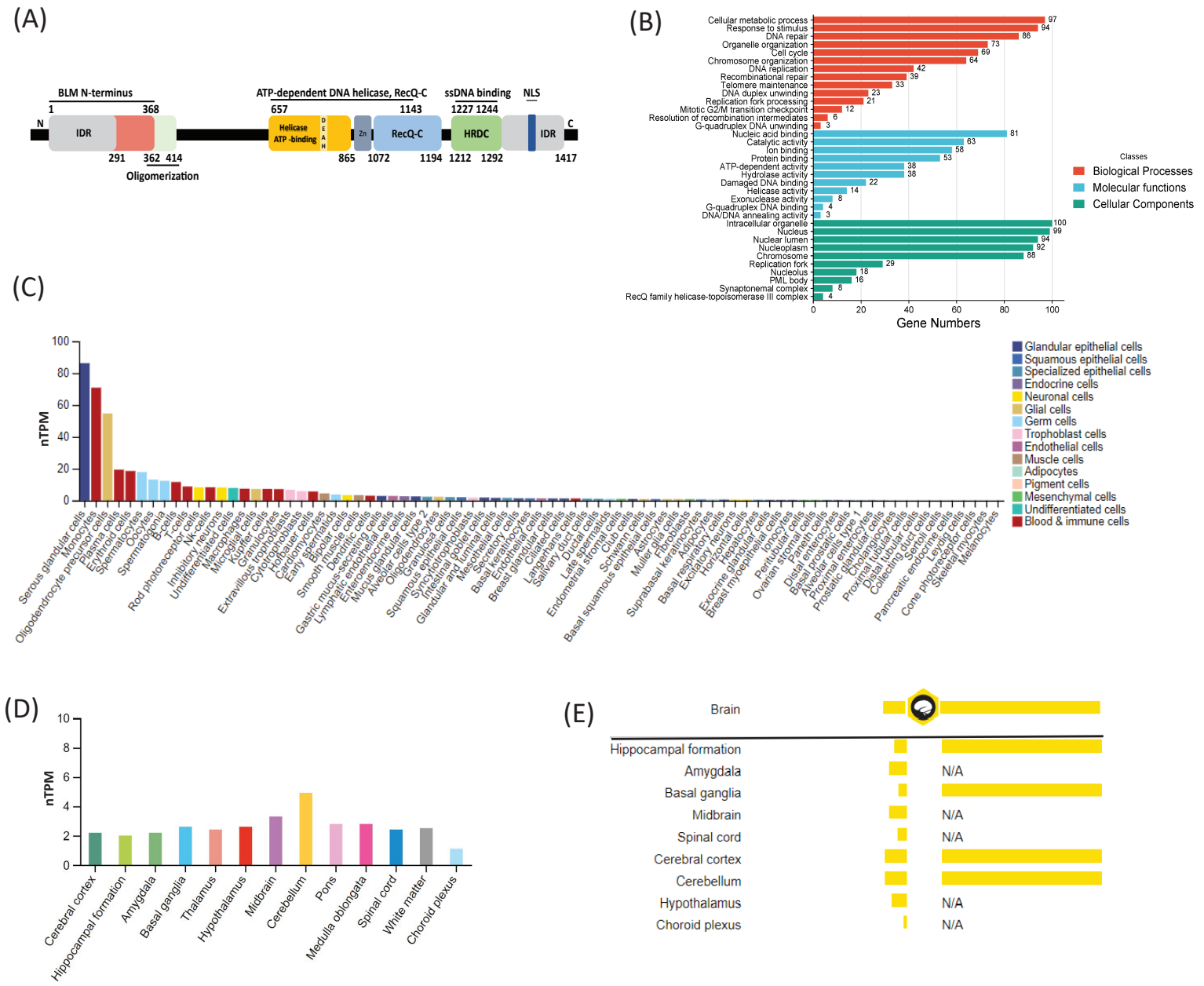

Figure 1: (A) Representation of BLM gene showing IDR (Intrinsically Disordered Regions), and Evolutionary conserved regions like Helicase-ATP binding domain, RecQ-C, and HRDC domain. **B)** Gene ontology (analysis) showing selected GO terms in Biological Processes, Molecular Functions, and Cellular Components for BLM-interacting genes according to SR plot. **C)** A summary of normalized single-cell RNA nTPM for BLM from all single-cell types. Color coding is based on cell type groups, each consisting of cell types with functional features in common. **D)** Normalized RNA expression levels (nTPM) of BLM were shown for the 13 brain regions. Color coding is based on brain region and the bar shows the highest expression among the subregions included. **E)** Bar Graph representation of RNA and Protein expression for Brain tissue according to The Human Protein Atlas.

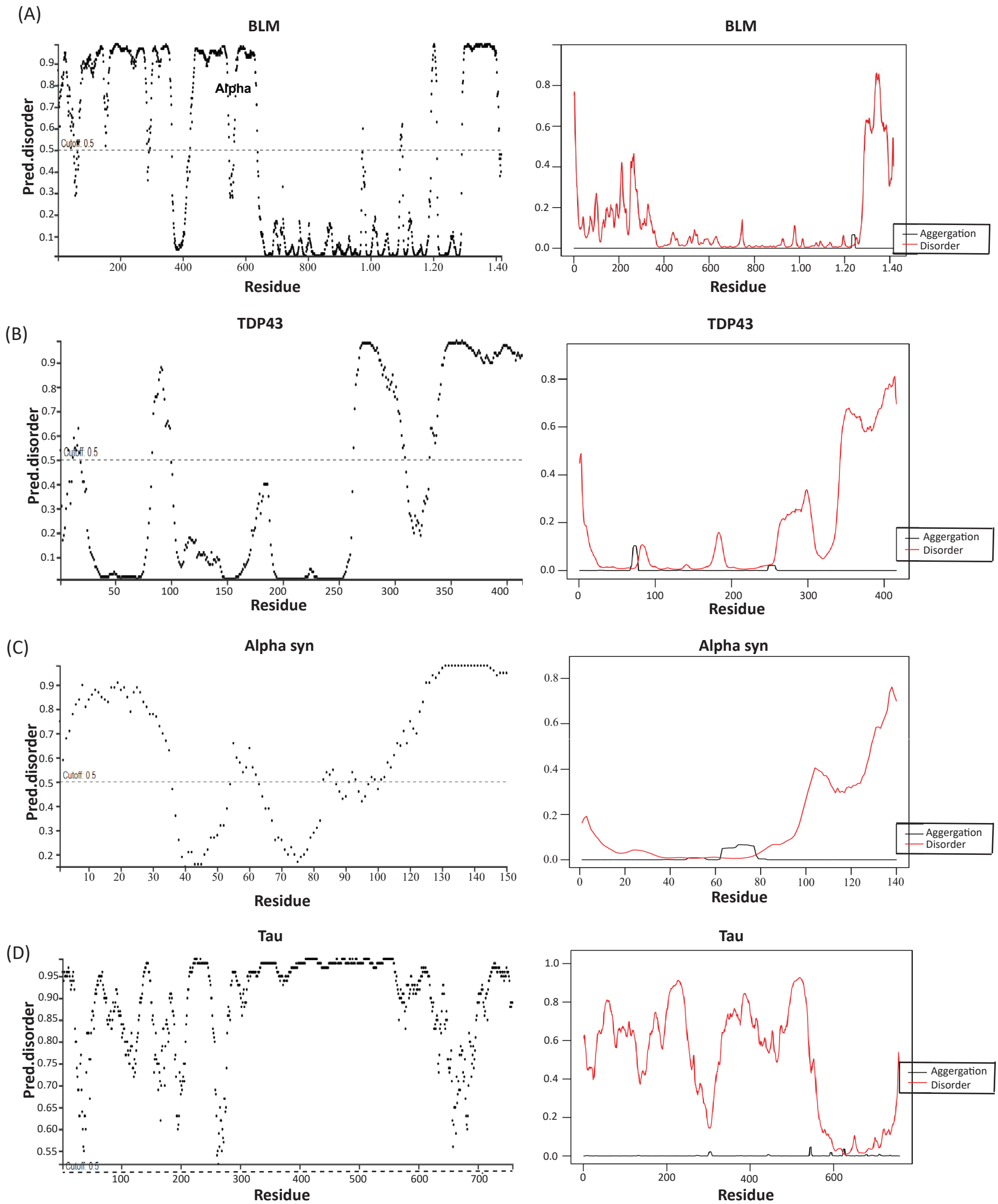

Figure 2: Depiction of Disordered and parallel aggregation formation plots: **A)** BLM (UniProt ID: P54132), **B)** TDP43 (UniProt ID: Q13148), **C)** α-Synuclein (UniProt ID: P37840), and **D)** Tau (UniProt ID: P10636) using DISOPRED3 and PASTA 2.0.

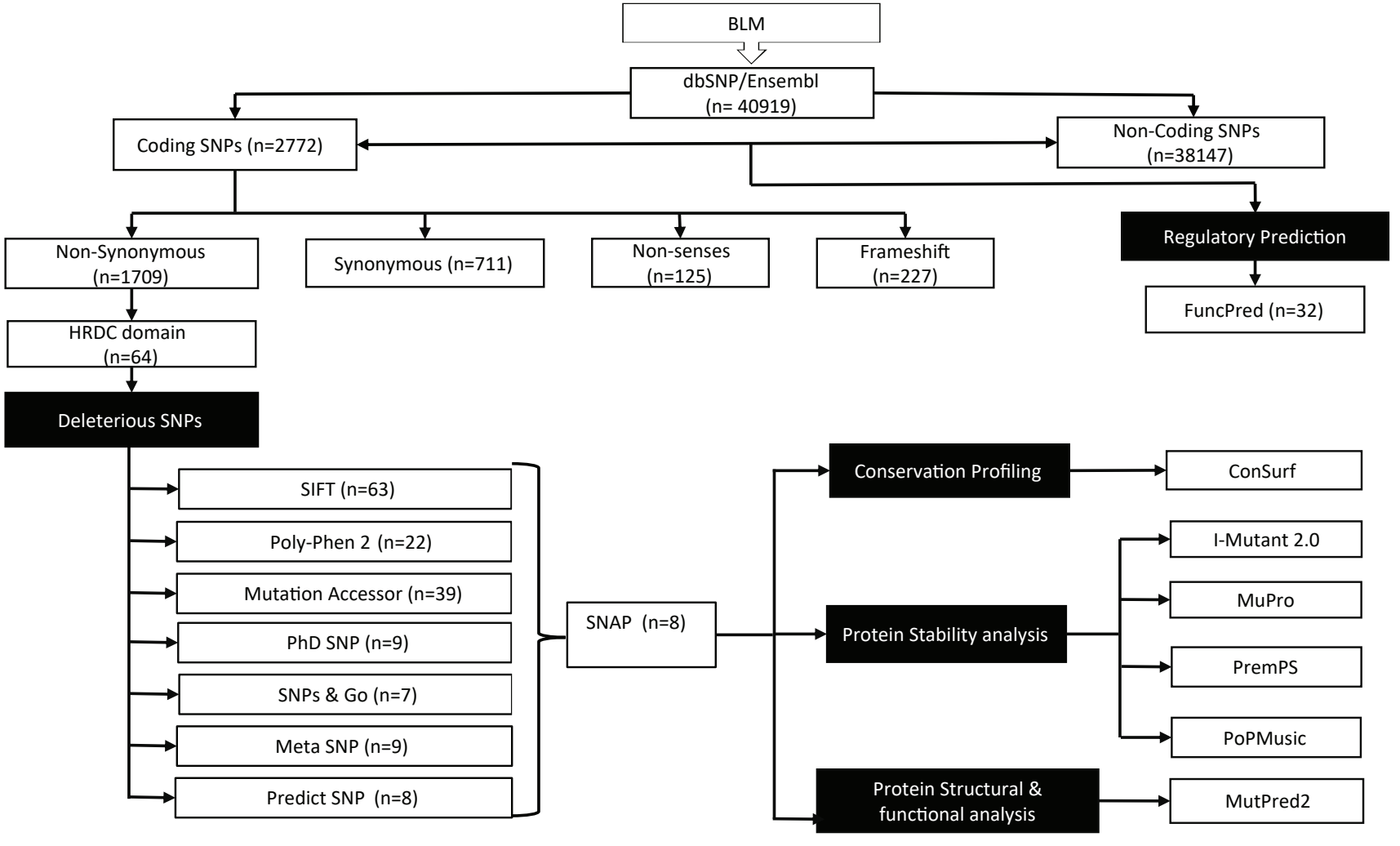

Figure 3: A schematic representation of nsSNPs selection and shortlisting procedure by different tools.

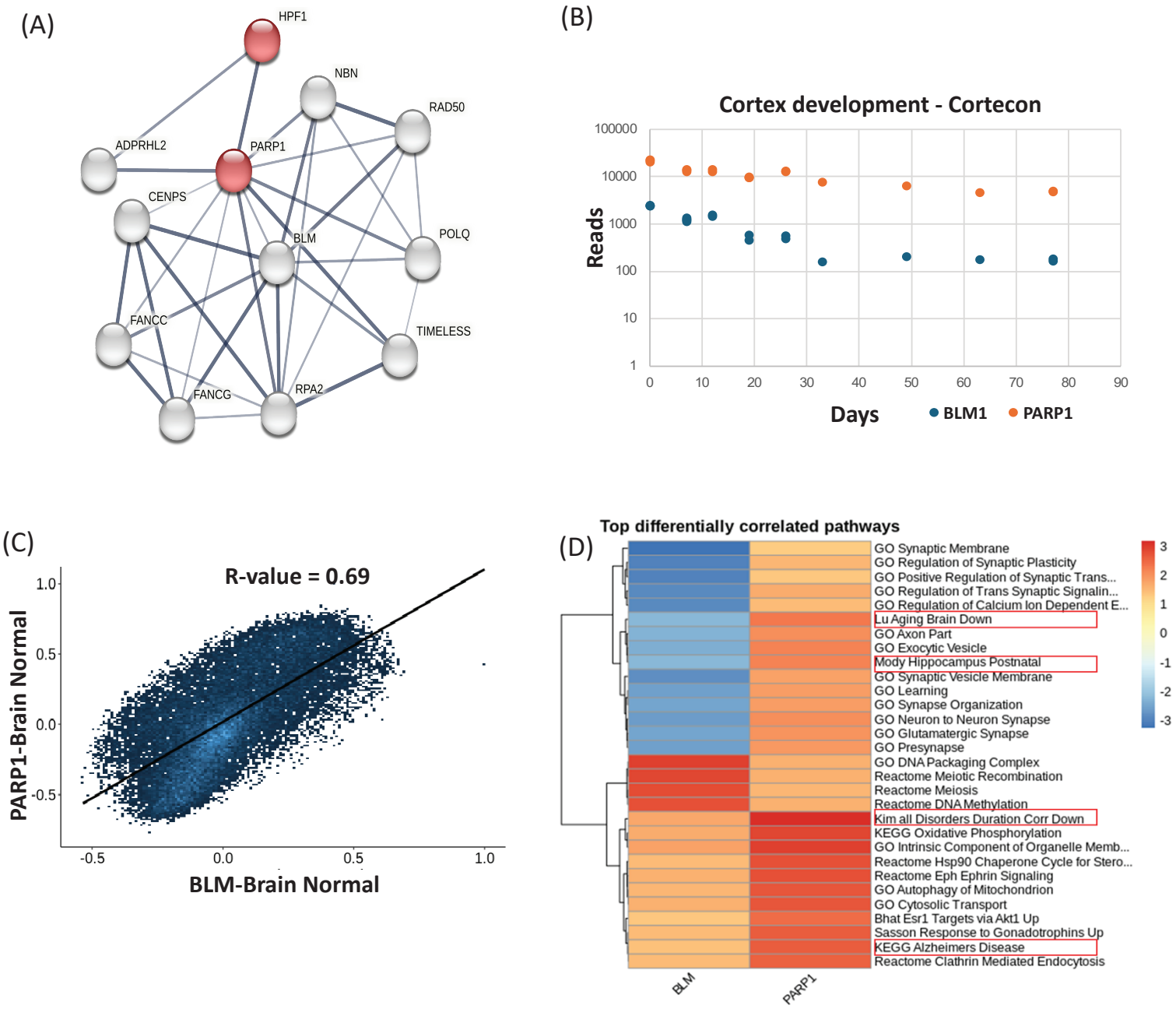

Figure 4: A) Network showing shared BLM and PARP implicated in repair mechanism. **B)** Similar pattern of BLM and PARP1 expression profiles during cortex development according to Cortecon. **D)** Scatter plot showing the relationship between genome-wide co-expression correlations for BLM and PARP1 in normal brain tissues with an R-value of 0.69 determined by Pearson correlation. **E)** Heatmap showing differential corGSEA results for BLM and PARP1 in normal tissues. The color bar displays the normalized enrichment score as determined by GSEA. Four neurodegenerative disease-related pathways are highlighted.