BLM Aggregation-Prone Ability and Similarity to Intrinsically Disordered Proteins Indicate its Potential Role in Neurodegenerative Diseases

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**

Suresh Singh Yadav, suresh4bhu@gmail.com

Abstract

Bloom's syndrome (BS) is a rare autosomal recessive human genetic disorder characterized by several abnormalities like short stature, immunodeficiency, sun sensitivity, male sterility, and cancer predisposition. BLM role has mostly been studied in DNA repair, replication and recombination; however, recent study has highlighted the RNA binding capability of the BLM protein. Several RNA binding proteins, including TDP43, Tau, and alpha-synuclein, are known to play roles in neurodegenerative diseases. These proteins possess residues prone to parallel aggregation and fibril formation. We utilized various in-silico tools like PASTA 2.0, SIFT, PolyPhen 2, PhD SNP. SNP&GO, Meta-SNP and SNAP etc to identify the molecular signature of the HRDC domain. Our insilico analysis predict the ability of HRDC domain of BLM in parallel aggregation and fibril formation. Moreover, its similarity with proteins like α-synuclein, TDP-43, and Tau, suggest it may have role in neurodegenerative diseases. Additionally, we identified the deleterious mutations that might affect the stability of protein and associations between BLM and PARP1/polyADP-ribosylation (PARylation). Based on these findings, we propose a potential role for BLM in neurodegenerative diseases.

Keywords

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

1. Introduction

RNA-binding proteins play an essential role in diverse cellular functions such as multifunctional regulators, regulating gene expression, alternative splicing, mRNA trafficking, stability, and translation. Around 1,500 catalogued members of these RNA binding proteins (RBPs), having diverse functions and roles have been annotated in the human genome (Corley et al., 2020; Gerstberger et al., 2014). These RBPs are characterized by distinct RNA binding domains (RBDs) including the RNA recognition motif (RRM), the K-homology (KH) domain, the DEAD/DEAH helicase domain, zinc-finger domains, and nearly 30 other less abundant domains with well-defined 3D structures or features (Corley et al., 2020; Kelaini et al., 2021). RBPs' role within the nervous system is very crucial and they exert precise control over fundamental processes such as neurogenesis, neuronal function, and nervous system development (Bryant & Yazdani, 2016; Chan et al., 2022). Alterations in the expression and function of RBPs are associated with a number of neurological disorders, neurodegenerative diseases, and the development of brain tumors (Gebauer et al., 2021; Kavanagh et al., 2022; Marcelino Meliso et al., 2017; Nussbacher et al., 2015; Prashad & Gopal, 2021; Sanya et al., 2023; Velasco et al., 2019).

RBPs form a complex network of multivalent protein-protein, protein-RNA, and RNA–RNA interactions resulting in liquid-liquid phase separation (LLPS). The process of LLPs facilitates the formation of membrane-less organelles such as stress granules, Cajal bodies, nucleoli, Ribonucleoprotein granules, and P-bodies (An et al., 2021; Drino & Schaefer, 2018; Gomes & Shorter, 2019). Most RBPs involved in LLPs mostly possess

intrinsically disordered regions (IDRs) and RGG boxes. IDR-RBPs can interact with RNA without the presence of specific folded RNA-binding domains (Zeke et al., 2022). Studies have reported the association of IDR-RBPs with neurodegenerative diseases 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) (Tsoi et al., 2023).

Many RBPs having a role in neurodegenerative disease also share a binding preference for G-quadruplexes (G4s) and participate in LLPS. G4s are complex DNA or RNA tetrad structures of guanosines stabilized by Hoogsteen base pairing. A number of biological functions such as gene expression, LLPs, DNA repair, and recombination are neurodegenerative diseases, neurological disorders, and prion diseases (Cave & Willis, 2022; Kosiol et al., 2021; Masai & Tanaka, 2020; Nakanishi & Seimiya, 2020). Recent studies have identified the RNA binding function of BLM protein and its preferential ability to bind to G-quadruplexes (G4s). BLM also has an N-terminal disordered region. Historically BLM gene and its mutation are known to result in a rare autosomal recessive disorder, Bloom syndrome. BLM belongs to the RecQ helicase family and carries an evolutionary conserved 3' to 5' DNA helicase which unwinds the double-helix DNA to produce single-stranded DNA as an intermediate required for the processes of replication, repair, and recombination (Mohaghegh et al., 2001).

Protein posttranslational modifications (PTMs) 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 (McRae et al., 2020). 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 (Baudin et al., 2021; Bryant & Yazdani, 2016; Kelaini et al., 2021; Marcelino Meliso et al., 2017; Martini et al., 2021; Prashad & Gopal, 2021; Stark et al., 2006). PARP1 activity is already known to be increased in Alzheimer's disease brains. Several proteins having a role in AD and ALS like α-synuclein, TDP-43, and hnRNP A1 are also regulated by PARylation and PAR-binding. PARylation and PAR-binding result in these prion-like proteins' phase separation properties and aggregation tendency (Nakanishi & Seimiya, 2020; Su et al., 2021; Supek et al., 2011).

In the present study using *in-silico* tools, we have investigated the potential role of BLM in neurodegenerative disease. Using the PASTA 2.0 tool we have identified the residues having the probability of parallel aggregation and fibril formation in the HRDC domain of BLM protein. We have also found many similarities between BLM compared with other IDPs having a role in neurodegenerative disease. The mutation in the HRDC domain identified in our study also affects the stability of the protein. Finally, our study also identified important associations between BLM and PARP1. Hence with this study, we propose the potential role of BLM in neurodegenerative disease.

2. Material and Methods

2.1 Data Mining

The human BLM 1 protein sequence and its basic information were taken from the National Center for Biotechnology Information (NCBI) (Protein ID: NP_000048.1) and the UniProt database (UniProtKB ID: P54132).

2.2 Sequence, structural, and physicochemical properties of BLM 1

ProtParam tool calculated the physicochemical parameters of human BLM 1 (Expasy - ProtParam tool). For the prediction of disorder in BLM 1 protein, the server prediction of amyloid structure aggregation (PASTA 2.0) (Aggregation Prediction with PASTA (un6ipd.it)) (Walsh et al., 2014), and Protein Structure Prediction Server (PSIPRED) (PSIPRED Workbench (ucl.ac.uk)) (Buchan & Jones, 2019).

2.3 Identification and analysis of potent nsSNPs of BLM1

ENSEMBL genome browser 111 (Ensembl genome browser 111) and dbSNP were used to retrieve nsSNPs in BLM1. We used the genomic coordinates of the "NM_000057.4" transcript of the BLM helicase "P54132" isoform 1 (MANE SELECT) for this study. Sequence-based tools such as SIFT (Sorting Intolerant From Tolerant; https://sift.bii.a-star.edu.sg/) (Sim et al., 2012), PolyPhen-2 (Polymorphism phenotyping v2; http://genetics.bwh.harvard.edu/pph2/) (Adzhubei et al., 2013), 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/), Predict SNP Screening of non-acceptable polymorphisms (SNAP) were used to check the effect of missense variants (Capriotti, Altman, et al., 2013). SIFT (score <0.05); PolyPhen-2 (score>0.95 - probably harmful; Meta-SNP, PhD-SNP, SNPs & Go (score>0.5) and with PredictSNP (Bendl et al., 2014) & SNAP (Pers et al., 2015) based on the accuracy of the damaging effect of amino acid substitution.

2.4 Predicting the effect of nsSNPs on the stability of BLM

I-Mutant 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html) (Capriotti et al., 2005) and MuPro (https://mupro.proteomics.ics.uci.edu) (Cheng et al., 2006), was used to predict change in the stability of BLM upon a single nucleotide change through protein sequence as input. The associated change in free energy (Kcal/mol; represented as DDG) and stability for shortlisted SNPs. A negative value shows decrease and positive value shows increase stability respectively. PremPS (https://lilab.jysw.suda.edu.cn/research/PremPS/) (Chen et al., 2020) and PoPMusic 2.1 (http://babylone.ulb.ac.be/popmusic) (Dehouck et al., 2011).uses pdb file or pdb ID as input to predict change in thermodynamic stability of BLM upon a single nucleotide change. They evaluate the effects of single mutations on protein stability by calculating the changes in unfolding and folding Gibbs free energy (ΔΔG) respectively. A positive and negative sign corresponds to destabilizing and stabilizing mutations respectively.

2.5 Structural and functional changes in BLM caused by Deleterious nsSNPs

MutPred2 (http://mutpred.mutdb.org), a popular machine learning method, detects pathogenic variants as well as the structural and functional changes brought on by harmful nsSNPs (Pejaver et al., 2020). 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. As previously mentioned, a P-value cutoff of 0.05 and a prediction score ranging from 0 to 1 were employed.

2.6 Gene ontology and network analysis of BLM-associated proteins

BLM interactors were downloaded from BioGRID (v4.4) (Stark et al., 2006) and the Search Tool for Retrieval of Interacting Genes and Proteins (STRING v12.0) (Szklarczyk et al., 2023) database is mainly used to assess the BLM intercoms, its interacting genes, and the proteins network. Gene Ontology (GO) enrichment analysis was performed using SRPlot (Tang et al., 2023). The Cortecon (van de Leemput et al., 2014) database is used to identify the shared similarities between BLM and PARP1 expression profiling during in vitro cortex development. The gene expression correlation was analyzed using Correlation AnalyzeR (Miller & Bishop, 2021).

3. Results and discussion

3.1 BLM expression in the brain

The BLM protein consists of 1417 amino acids encoded by the BLM1 gene on chromosome 15 (UniProtKB ID: P54132) and has a molecular weight of approximately 159 kDa. It contains conserved domains such as RecA-like domains and a Zn2+ binding domain (residues 657–1143), as well as an HRDC domain (residues 1213–

1292), which is unique to BLM and WRN, two of the five human RecQ helicases (RecQ1, RecQ4, RecQ5, BLM and WRN) found in Bacteria and Yeast. These domains act together as a central catalytic center (Kitano, 2014; Wu et al., 2005). The unstructured regions of the protein, including the N-terminus (residues 1–657), which is responsible for the interaction and oligomerization of the partner protein, and the C-terminus (residues 1293–1417), which contains nuclear localization signals, **Figure 1(A)** represents the features of the BLM1 gene and protein.
Gene Ontology (GO) analyses (Supek et al., 2011; Tang et al., 2023; Zhou et al., 2019) indicate that BLM participates in diverse regulatory complexes in the cytoplasm and nucleus **Figure 1(B)**. Mostly its function has been reported on cellular metabolic processes such as DNA repair, and replication. On analysing its expression pattern, it was found to be in serous glandular cells **Figure 1(C)**. Oligodendrocyte precursor cells were also among the cells that had the highest level of gene expression of BLM (Karlsson et al., 2021). 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 **Figure 1(D, E)** (Sjostedt et al., 2020; Uhlen et al., 2015).

3.2 BLM 1 is a thermostable protein having the ability to form parallel aggregation and fibril

On computing the physical and chemical parameters of BLM using ProtParam it was classified as an unstable protein with the instability index of 44.57. The aliphatic index of BLM1 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)}. PASTA 2.0 predicted that a significant region of BLM 1 (38.96%) contains an alpha-helical, 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, BLM 1 N-terminal 1-21, 39-42, 72-75, 92-107, 142-175, 183-232, 247-288, 290, 291, 313, and 323–339 residues are disordered, and residues 1224–1250 show the probability of parallel aggregation and fibril formation **(Figure 2A)**, and C-terminal 1277-1417 residues are disordered. In neurodegenerative diseases, a significant proportion of proteins with intrinsically disordered regions (IDRs) play crucial roles. Approximately 69% of human disease-associated proteins, including those linked to neurodegenerative disorders, are expected to contain disorder regions of ≥30 consecutive amino acid residues (Raychaudhuri et al., 2009). These IDPs (intrinsically disordered proteins) or IDPRs (intrinsically disordered protein regions) are mixed with few structured domains or can be completely unstructured. These proteins are mainly composed of polar or charged amino acids and lack enough hydrophobic residues to facilitate the proper folding of proteins that are involved in conditions such as Alzheimer’s disease, Parkinson’s disease, amyotrophic lateral sclerosis (ALS), and more (Uversky, 2015). Both structured and unstructured intrinsically disordered regions of the protein interact to produce cytotoxicity and contribute in distinct ways to the selectivity and efficiency of RNP granule condensation and neurodegeneration (Petrauskas et al., 2024). Misfolding and aggregation/fibrillation of proteins are caused by mutations in IDPs, changes in the cellular environment, and regular post-translational modification activities that act as conformational switches to increase cellular stress or damage. Also, misfolded proteins can cause other proteins to aggregate through cross-seeding (Coskuner-Weber et al., 2022; Lim, 2019). Common neurodegenerative diseases are associated with IDPs 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) (Tsoi et al., 2023). IDPs also play important roles in various biological and cellular processes including transcription and translation (Tsafou et al., 2018), cell cycle regulation (p21 & p27) (Yoon et al., 2012), cell signaling, and formation of membrane-less organelles (Bondos et al., 2021).

3.3 Similarity of BLM with TDP43, Tau, and α-synuclein

In a biological context, BLM, TDP43, Tau, and α-synuclein are different from each other and perform some sort of different biological functions. Similar to TDP-43, α-syn, Aβ, and Tau, BLM (Bloom syndrome protein), also carries intrinsically disordered, and is critically involved in DNA repair, recombination processes, and maintenance of genomic stability. By comparing the physical and chemical parameters of BLM with TDP43,

Tau, and α-synuclein all are unstable except α-synuclein with the instability index of 44.57, 56.76, 56.02, and 25.47 respectively. 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 (Buchan & Jones, 2019) to predict the secondary structure and intrinsic disordered regions of BLM, TDP43, Tau, and α-synuclein showing the position of structure domains like α-helix, beta-strands present in order region of protein having threshold <0.5 are involved to form parallel aggregation mainly beta-strands as shown in **Figure 2**, identified using PASTA 2.0 tool. As we noticed BLM shows the probability of parallel aggregation and fibril formation like TDP43, Tau, and α-synuclein which are commonly related to neurodegenerative diseases like amyotrophic lateral sclerosis (ALS), Parkinson's disease (PD), and Alzheimer's disease (AD) (Tsoi et al., 2023; Walsh et al., 2014). It is supposed that increased expression of BLM and may result in formation of pathological aggregates resulting in neurodegenerative diseases.

3.4 Predicting deleterious and damaging nsSNPs

Mutations in BLM can affect the stability of protein hence in order 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 111). Furthermore, these SNPs were further validated by the NCBI Variation Viewer dbSNP database (BLM: Chr15 - Variation Viewer (nih.gov). Using both tools we identified 40919 single nucleotide variants of the BLM gene. Later we used the genomic coordinates of the "NM_000057.4" transcript of the BLM helicase "P54132" isoform 1 (MANE SELECT) and found 2647 SNPs in the coding region sequence that were identified as missense variants (non-synonymous SNPs) with 123 stop-gained and 2 start-lost nsSNPs were identified **Figure 3**. The remaining nsSNPs were filtered based on the standard scoring of various sequence-based consensus tools: SIFT (score <0.05) (Sim et al., 2012); PolyPhen-2 (score>0.95 - probably harmful, score>0.5 – possibly damaging) (Adzhubei et al., 2013); Mutation Assessor (score>0.5); Meta-SNP, PhD-SNP, SNPs & Go (score>0.5) (Capriotti, Calabrese, et al., 2013) and with PredictSNP (Bendl et al., 2014) & SNAP (Pers et al., 2015) based on the accuracy of the harmful effect of amino acid substitution.

In this study, we examined the highly conserved HRDC domain (helicase and RNase D C-terminal), which we found to form parallel aggregation (1224–1250 amino acid residues) **Figure 2(A)** and is important for binding of ssDNA and DNA Holliday junction is necessary (1227-1244 aa residue). A total of 64 nsSNPs were found in the HRDC domain and based on the score predicted by the above sequence-based tools, 8 nsSNPS (H1236Y, F1238L, L1253S, T1267I, E1268V, E1272G, V1278M, S1286) were classified as predominantly deleterious in at least 5 of the above algorithmic tools through manual concordance. All 8 SNPs were found to be effective variants and further investigated for structural and functional consequences using various tools.

3.5 Predicting the effect of 8 deleterious nsSNPs

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) (Capriotti et al., 2005), MuPro (https://mupro.proteomics.ics.uci.edu) (Cheng et al., 2006), PremPS (https://lilab.jysw.suda.edu.cn/research/PremPS/) (Chen et al., 2020) and PoPMusic 2.1 (http://babylone.ulb.ac.be/popmusic) (Dehouck et al., 2011).

Both I-Mutant 2.0 and MuPro predict the effect of the mutation based on the input sequence, while PremPS and PoPMusic use the 3D structure of proteins as the whole protein 3D structure of BLM gene is not available yet, 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 found that only three amino acid substitutions (H1236Y, E1268V) and (T1267I) from 8 nsSNPS (H1236Y, F1238L, L1253S, T1267I, E1268V, E1272G, V1278M, S1286) increases

protein stability and remaining decreases by using I-Mutant 2.-0 and MuPro, respectively. In contrast, all substitutions destabilize the protein structure by PremPS and PoPMusic. To infer the structural, molecular, and phenotypic effects of amino acid variations, MutPred2 (Pejaver et al., 2020) was used.

The MutPred 2 webserver predicted 6 out of the 8 shortlisted deleterious nsSNPs to cause structural and functional changes as they scored higher than 0.50. These changes affected properties like Altered ordered interface, altered transmembrane protein, protein stability, metal binding sites, DNA binding site, and disordered interface; loss/gain of allosteric sites, catalytic, and proteolytic cleavage site; loss/gain of the loop, Helix, B-factor, intrinsic disorder, and posttranslational modifications at different residues, etc. (as shown in table 4).

3.6 BLM association with PAPR1

The STRING analysis provided the protein-protein interaction profiling for BLM and its interacting proteins **Figure 4(A)** BLM interacts with PARP1, 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 and controls DNA repair processes (Huang et al., 2023).

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 cortex development **Figure 4(B)** (van de Leemput et al., 2014). We also found the correlated co-expression and divergent expression of BLM and PARP1 in normal brain tissue indicating their important existence together **Figure (D)**. Also, BLM contains a PARPylation site (1303-1333) (Gibson et al., 2016) between HRDC and NLS which was present on intrinsically disordered region. By comparing the result of corGSEA between BLM and PARP1, we found differential enrichment of several pathways implicated in neurodegenerative diseases, such as “KEGG Alzheimers Disease”, “Lu aging brain down”, “Kim all disorders duration corr down”, and “Mody Hippocampus postnatal”. This suggests the possible association of these two pathways related to neurodegenerative diseases (Miller & Bishop, 2021).

4. Conclusion

This is the first study to investigate the elucidation of the involvement of BLM in neurodegenerative diseases. BLM contains two intrinsically disordered regions at the N-terminal and -C-terminal ends and various tools predicted the presence of residue with a high propensity to form IDR throughout the protein. Our study provides a basic knowledge about the HRDC domain of BLM having the ability to form parallel aggregation which mainly helps to regulate the helicase and is responsible for DNA binding. Furthermore, the HRDC domain of BLM harbors potent nsSNPs altering either the structure or stability of BLM. This is the first study regarding nsSNPs of the HRDC domain of BLM, using various in-silico tools we suggest F1238L, L1253S, E1272G, V1278M, and S1286 destabilizing the protein based on Gibbs free energy change value and change in folding energy. Moreover, the deleterious nsSNPs found in the HRDC domain can affect and alter its structure and functions and aggregation-forming ability. Additionally, the BLM and PARP1 show a similar expression pattern in cortex development, and the association of BLM and PARP1 suggests that several BLM complexes or related pathways may be modulated by PARylation. Further studies of the molecular mechanism of the role of PARPylation in the IDR region of BLM and aggregation formation by the HRDC domain affecting neurodegenerative diseases are needed.

Conflict of interest statement

The authors disclose no conflict of interest.

Acknowledgement

The authors would like to thank Guru Nanak Dev University, Amritsar Punjab, India, and Gujrat Biotechnology University, Gandhinagar, Gujrat, India, for providing a space to work and an internet facility to complete this manuscript. This work was supported by DBT-RLS grant to support Mr Saurav Sharma.

Bibliography:

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*, *Chapter 7*, Unit7 20. https://doi.org/10.1002/0471142905.hg0720s76

An, H., de Meritens, C. R., & Shelkovnikova, T. A. (2021). Connecting the "dots": RNP granule network in health and disease. *Biochim Biophys Acta Mol Cell Res*, *1868*(8), 119058. https://doi.org/10.1016/j.bbamcr.2021.119058

Baudin, A., Moreno-Romero, A. K., Xu, X., Selig, E. E., Penalva, L. O. F., & Libich, D. S. (2021). Structural Characterization of the RNA-Binding Protein SERBP1 Reveals Intrinsic Disorder and Atypical RNA Binding Modes. *Front Mol Biosci*, *8*, 744707. https://doi.org/10.3389/fmolb.2021.744707

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E. D., Zendulka, J., Brezovsky, J., & Damborsky, J. (2014). PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*, *10*(1), e1003440. https://doi.org/10.1371/journal.pcbi.1003440

Bondos, S. E., Dunker, A. K., & Uversky, V. N. (2021). On the roles of intrinsically disordered proteins and regions in cell communication and signaling. *Cell Commun Signal*, *19*(1), 88. https://doi.org/10.1186/s12964-021-00774-3

Bryant, C. D., & Yazdani, N. (2016). RNA-binding proteins, neural development and the addictions. *Genes Brain Behav*, *15*(1), 169-186. https://doi.org/10.1111/gbb.12273

Buchan, D. W. A., & Jones, D. T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res*, *47*(W1), W402-W407. https://doi.org/10.1093/nar/gkz297

Capriotti, E., Altman, R. B., & Bromberg, Y. (2013). Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics*, *14 Suppl 3*(Suppl 3), S2. https://doi.org/10.1186/1471-2164-14-S3-S2

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, *14 Suppl 3*(Suppl 3), S6. https://doi.org/10.1186/1471-2164-14-S3-S6

Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res*, *33*(Web Server issue), W306-310. https://doi.org/10.1093/nar/gki375

Cave, J. W., & Willis, D. E. (2022). G-quadruplex regulation of neural gene expression. *FEBS J*, *289*(12), 3284-3303. https://doi.org/10.1111/febs.15900

Chan, J. N., Sanchez-Vidana, D. I., Anoopkumar-Dukie, S., Li, Y., & Benson Wui-Man, L. (2022). RNA-binding protein signaling in adult neurogenesis. *Front Cell Dev Biol*, *10*, 982549. https://doi.org/10.3389/fcell.2022.982549

Chen, Y., Lu, H., Zhang, N., Zhu, Z., Wang, S., & Li, M. (2020). PremPS: Predicting the impact of missense mutations on protein stability. *PLoS Comput Biol*, *16*(12), e1008543. https://doi.org/10.1371/journal.pcbi.1008543

Cheng, J., Randall, A., & Baldi, P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, *62*(4), 1125-1132. https://doi.org/10.1002/prot.20810

Corley, M., Burns, M. C., & Yeo, G. W. (2020). How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Mol Cell*, *78*(1), 9-29. https://doi.org/10.1016/j.molcel.2020.03.011

Coskuner-Weber, O., Mirzanli, O., & Uversky, V. N. (2022). Intrinsically disordered proteins and proteins with intrinsically disordered regions in neurodegenerative diseases. *Biophys Rev*, *14*(3), 679-707. https://doi.org/10.1007/s12551-022-00968-0

Dehouck, Y., Kwasigroch, J. M., Gilis, D., & Rooman, M. (2011). PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, *12*, 151. https://doi.org/10.1186/1471-2105-12-151

Drino, A., & Schaefer, M. R. (2018). RNAs, Phase Separation, and Membrane-Less Organelles: Are Post-Transcriptional Modifications Modulating Organelle Dynamics? *Bioessays*, *40*(12), e1800085. https://doi.org/10.1002/bies.201800085

Gebauer, F., Schwarzl, T., Valcarcel, J., & Hentze, M. W. (2021). RNA-binding proteins in human genetic disease. *Nat Rev Genet*, *22*(3), 185-198. https://doi.org/10.1038/s41576-020-00302-y

Gerstberger, S., Hafner, M., & Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat Rev Genet*, *15*(12), 829-845. https://doi.org/10.1038/nrg3813

Gibson, B. A., Zhang, Y., Jiang, H., Hussey, K. M., Shrimp, J. H., Lin, H., Schwede, F., Yu, Y., & Kraus, W. L. (2016). Chemical genetic discovery of PARP targets reveals a role for PARP-1 in transcription elongation. *Science*, *353*(6294), 45-50. https://doi.org/10.1126/science.aaf7865

Gomes, E., & Shorter, J. (2019). The molecular language of membraneless organelles. *J Biol Chem*, *294*(18), 7115-7127. https://doi.org/10.1074/jbc.TM118.001192

Huang, M., Chen, L., Guo, Y., Ruan, Y., & Xu, H. (2023). PARP1 negatively regulates transcription of BLM through its interaction with HSP90AB1 in prostate cancer. *J Transl Med*, *21*(1), 445. https://doi.org/10.1186/s12967-023-04288-z

Karlsson, M., Zhang, C., Mear, L., Zhong, W., Digre, A., Katona, B., Sjostedt, E., Butler, L., Odeberg, J., Dusart, P., Edfors, F., Oksvold, P., von Feilitzen, K., Zwahlen, M., Arif, M., Altay, O., Li, X., Ozcan, M., Mardinoglu, A., . . . Lindskog, C. (2021). A single-cell type transcriptomics map of human tissues. *Sci Adv*, *7*(31). https://doi.org/10.1126/sciadv.abh2169

Kavanagh, T., Halder, A., & Drummond, E. (2022). Tau interactome and RNA binding proteins in neurodegenerative diseases. *Mol Neurodegener*, *17*(1), 66. https://doi.org/10.1186/s13024-022-00572-6

Kelaini, S., Chan, C., Cornelius, V. A., & Margariti, A. (2021). RNA-Binding Proteins Hold Key Roles in Function, Dysfunction, and Disease. *Biology (Basel)*, *10*(5). https://doi.org/10.3390/biology10050366

Kitano, K. (2014). Structural mechanisms of human RecQ helicases WRN and BLM. *Front Genet*, *5*, 366. https://doi.org/10.3389/fgene.2014.00366

Kosiol, N., Juranek, S., Brossart, P., Heine, A., & Paeschke, K. (2021). G-quadruplexes: a promising target for cancer therapy. *Mol Cancer*, *20*(1), 40. https://doi.org/10.1186/s12943-021-01328-4

Lim, K. H. (2019). Diverse Misfolded Conformational Strains and Cross-seeding of Misfolded Proteins Implicated in Neurodegenerative Diseases. *Front Mol Neurosci*, *12*, 158. https://doi.org/10.3389/fnmol.2019.00158

Marcelino Meliso, F., Hubert, C. G., Favoretto Galante, P. A., & Penalva, L. O. (2017). RNA processing as an alternative route to attack glioblastoma. *Hum Genet*, *136*(9), 1129-1141. https://doi.org/10.1007/s00439-017-1819-2

Martini, S., Davis, K., Faraway, R., Elze, L., Lockwood, N., Jones, A., Xie, X., McDonald, N. Q., Mann, D. J., Armstrong, A., Ule, J., & Parker, P. J. (2021). A genetically-encoded crosslinker screen identifies SERBP1 as a PKCepsilon substrate influencing translation and cell division. *Nat Commun*, *12*(1), 6934. https://doi.org/10.1038/s41467-021-27189-5

Masai, H., & Tanaka, T. (2020). G-quadruplex DNA and RNA: Their roles in regulation of DNA replication and other biological functions. *Biochem Biophys Res Commun*, *531*(1), 25-38. https://doi.org/10.1016/j.bbrc.2020.05.132

McRae, E. K. S., Dupas, S. J., Booy, E. P., Piragasam, R. S., Fahlman, R. P., & McKenna, S. A. (2020). An RNA guanine quadruplex regulated pathway to TRAIL-sensitization by DDX21. *RNA*, *26*(1), 44-57. https://doi.org/10.1261/rna.072199.119

Miller, H. E., & Bishop, A. J. R. (2021). Correlation AnalyzeR: functional predictions from gene co-expression correlations. *BMC Bioinformatics*, *22*(1), 206. https://doi.org/10.1186/s12859-021-04130-7

Mohaghegh, P., Karow, J. K., Brosh, R. M., Jr., Bohr, V. A., & Hickson, I. D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res*, *29*(13), 2843-2849. https://doi.org/10.1093/nar/29.13.2843

Nakanishi, C., & Seimiya, H. (2020). G-quadruplex in cancer biology and drug discovery. *Biochem Biophys Res Commun*, *531*(1), 45-50. https://doi.org/10.1016/j.bbrc.2020.03.178

Nussbacher, J. K., Batra, R., Lagier-Tourenne, C., & Yeo, G. W. (2015). RNA-binding proteins in neurodegeneration: Seq and you shall receive. *Trends Neurosci*, *38*(4), 226-236. https://doi.org/10.1016/j.tins.2015.02.003

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K. A., Lin, G. N., Nam, H. J., Mort, M., Cooper, D. N., Sebat, J., Iakoucheva, L. M., Mooney, S. D., & Radivojac, P. (2020). Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat Commun*, *11*(1), 5918. https://doi.org/10.1038/s41467-020-19669-x

Pers, T. H., Timshel, P., & Hirschhorn, J. N. (2015). SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics*, *31*(3), 418-420. https://doi.org/10.1093/bioinformatics/btu655

Petrauskas, A., Fortunati, D. L., Kandi, A. R., Pothapragada, S. S., Agrawal, K., Singh, A., Huelsmeier, J., Hillebrand, J., Brown, G., Chaturvedi, D., Lee, J., Lim, C., Auburger, G., VijayRaghavan, K., Ramaswami, M., & Bakthavachalu, B. (2024). Structured and disordered regions of Ataxin-2 contribute differently to the specificity and efficiency of mRNP granule formation. *PLoS Genet*, *20*(5), e1011251. https://doi.org/10.1371/journal.pgen.1011251

Prashad, S., & Gopal, P. P. (2021). RNA-binding proteins in neurological development and disease. *RNA Biol*, *18*(7), 972-987. https://doi.org/10.1080/15476286.2020.1809186

Raychaudhuri, S., Dey, S., Bhattacharyya, N. P., & Mukhopadhyay, D. (2009). The role of intrinsically unstructured proteins in neurodegenerative diseases. *PLoS One*, *4*(5), e5566. https://doi.org/10.1371/journal.pone.0005566

Sanya, D. R. A., Cava, C., & Onesime, D. (2023). Roles of RNA-binding proteins in neurological disorders, COVID-19, and cancer. *Hum Cell*, *36*(2), 493-514. https://doi.org/10.1007/s13577-022-00843-w

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, *40*(Web Server issue), W452-457. https://doi.org/10.1093/nar/gks539

Sjostedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., Huang, J., Du, Y., Lin, L., Dong, Z., Yang, L., Liu, X., Jiang, H., Xu, X., Wang, J., . . . Mulder, J. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*, *367*(6482). https://doi.org/10.1126/science.aay5947

Stark, C., Breitkreutz, B. J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res*, *34*(Database issue), D535-539. https://doi.org/10.1093/nar/gkj109

Su, Q., Mehta, S., & Zhang, J. (2021). Liquid-liquid phase separation: Orchestrating cell signaling through time and space. *Mol Cell*, *81*(20), 4137-4146. https://doi.org/10.1016/j.molcel.2021.09.010

Supek, F., Bosnjak, M., Skunca, N., & Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A. L., Fang, T., Doncheva, N. T., Pyysalo, S., Bork, P., Jensen, L. J., & von Mering, C. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*, *51*(D1), D638-D646. https://doi.org/10.1093/nar/gkac1000

Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023). SRplot: A free online platform for data visualization and graphing. *PLoS One*, *18*(11), e0294236. https://doi.org/10.1371/journal.pone.0294236

Tsafou, K., Tiwari, P. B., Forman-Kay, J. D., Metallo, S. J., & Toretsky, J. A. (2018). Targeting Intrinsically Disordered Transcription Factors: Changing the Paradigm. *J Mol Biol*, *430*(16), 2321-2341. https://doi.org/10.1016/j.jmb.2018.04.008

Tsoi, P. S., Quan, M. D., Ferreon, J. C., & Ferreon, A. C. M. (2023). Aggregation of Disordered Proteins Associated with Neurodegeneration. *Int J Mol Sci*, *24*(4). https://doi.org/10.3390/ijms24043380

Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., . . . Ponten, F. (2015). Proteomics. Tissue-based map of the human proteome. *Science*, *347*(6220), 1260419. https://doi.org/10.1126/science.1260419

Uversky, V. N. (2015). Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Front Aging Neurosci*, *7*, 18. https://doi.org/10.3389/fnagi.2015.00018

van de Leemput, J., Boles, N. C., Kiehl, T. R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R. A., Levi, B. P., Thompson, C. L., Yao, S., Kaykas, A., Temple, S., & Fasano, C. A. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. *Neuron*, *83*(1), 51-68. https://doi.org/10.1016/j.neuron.2014.05.013

Velasco, M. X., Kosti, A., Penalva, L. O. F., & Hernandez, G. (2019). The Diverse Roles of RNA-Binding Proteins in Glioma Development. *Adv Exp Med Biol*, *1157*, 29-39. https://doi.org/10.1007/978-3-030-19966-1_2

Walsh, I., Seno, F., Tosatto, S. C., & Trovato, A. (2014). PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res*, *42*(Web Server issue), W301-307. https://doi.org/10.1093/nar/gku399

Wu, L., Chan, K. L., Ralf, C., Bernstein, D. A., Garcia, P. L., Bohr, V. A., Vindigni, A., Janscak, P., Keck, J. L., & Hickson, I. D. (2005). The HRDC domain of BLM is required for the dissolution of double Holliday junctions. *EMBO J*, *24*(14), 2679-2687. https://doi.org/10.1038/sj.emboj.7600740

Yoon, M. K., Mitrea, D. M., Ou, L., & Kriwacki, R. W. (2012). Cell cycle regulation by the intrinsically disordered proteins p21 and p27. *Biochem Soc Trans*, *40*(5), 981-988. https://doi.org/10.1042/BST20120092

Zeke, A., Schad, E., Horvath, T., Abukhairan, R., Szabo, B., & Tantos, A. (2022). Deep structural insights into RNA-binding disordered protein regions. *Wiley Interdiscip Rev RNA*, *13*(5), e1714. https://doi.org/10.1002/wrna.1714

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., & Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6

Table 1: A table represents the structural composition of BLM, TDP43, Tau, and α-synuclein proteins, showing the percentage of different secondary structure elements: Disordered region, alpha-helix, beta-strand, and coil regions.

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-Synuclein	34.28	26.43	22.14	51.43

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	PolyPhen2	Mutation Accessor	PhD-SNP	SNPs & GO	Meta SNP	Predict SNP	SNAP
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

Table 3: Effect of amino acid substitutions on the stability of mutant BLM protein using IMutant2, MuPro, PremPS and PoPMuSiC v3.1 tools. IMutant2 & MuPro: Gibbs free energy change value (DDG) in Kcal/mol, where DDG < 0 decreases stability, while DDG > 0 increases stability of mutant protein. PremPS & PoPMuSiC: The ΔΔG indicates the change in folding free energy in Kcal/mol where ΔΔG < 0 is stabilizing mutation.

AA substitution	**I-Mutant 2.0**		**MuPro**		**PremPS**		**PoPMuSiC**	
	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 six deleterious nsSNPs and the resulting structural and functional alterations in BLM protein, as predicted by MutPred2

Sr. No.	Variant ID	Substitutions	ConSurf Grade	MutPred2 Score	Molecular Mechanism
1.	rs1897237440	H1236Y	9	0.8	Altered Ordered interface
2.	rs1897237692	F1238L	9	0.681	Loss of Loop
3.	rs2151198266	L1253S	9	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	9	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	9	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	8	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

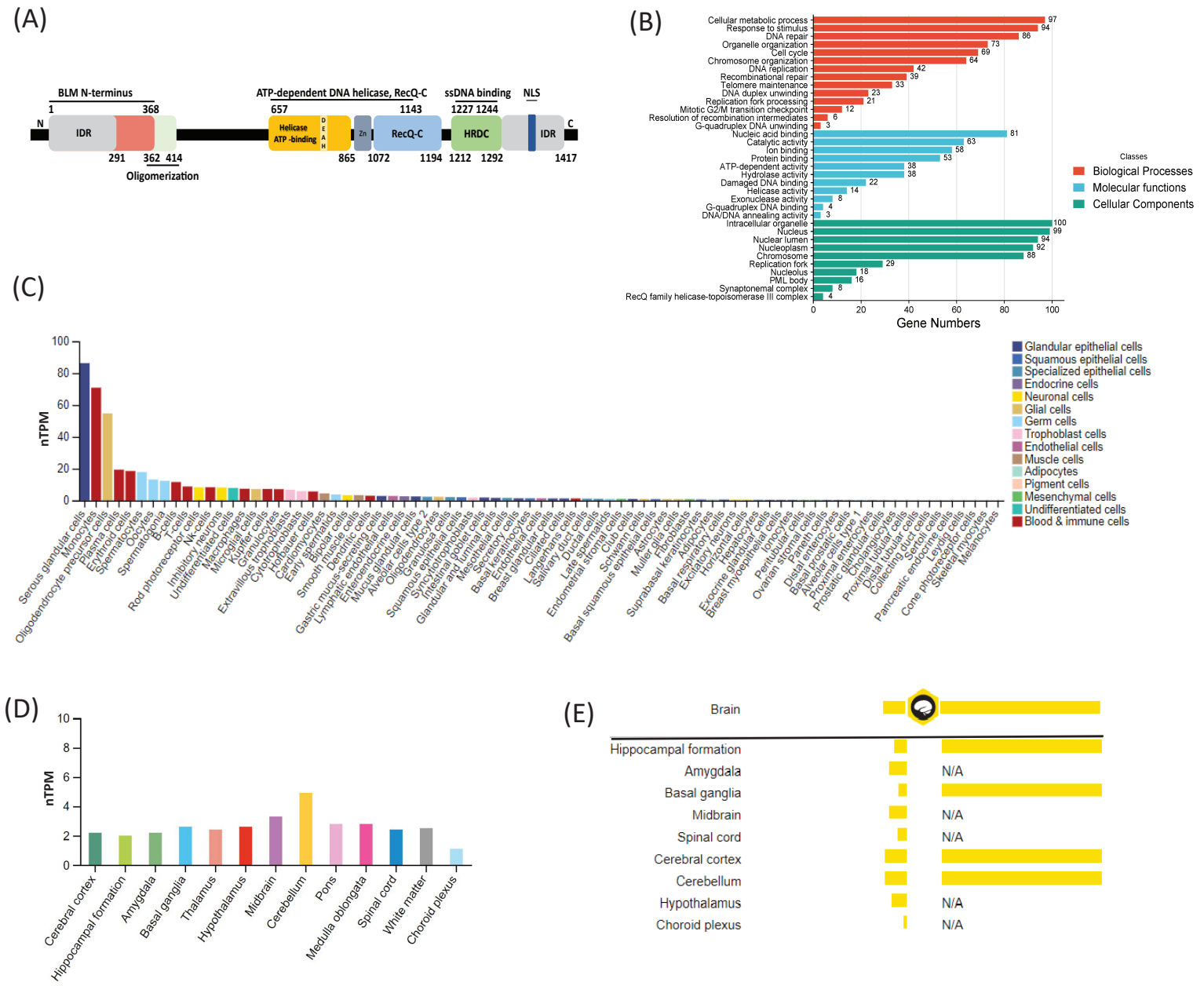

Figure 1: (A) Representation of BLM gene showing IDR (Intrinsically Disordered Regions), and Evolutionary conserved regions like Helicase-ATP binding domain, RecQ-C, and HRDC domain. **B)** Gene ontology (analysis) showing selected GO terms in Biological Processes, Molecular Functions, and Cellular Components 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.

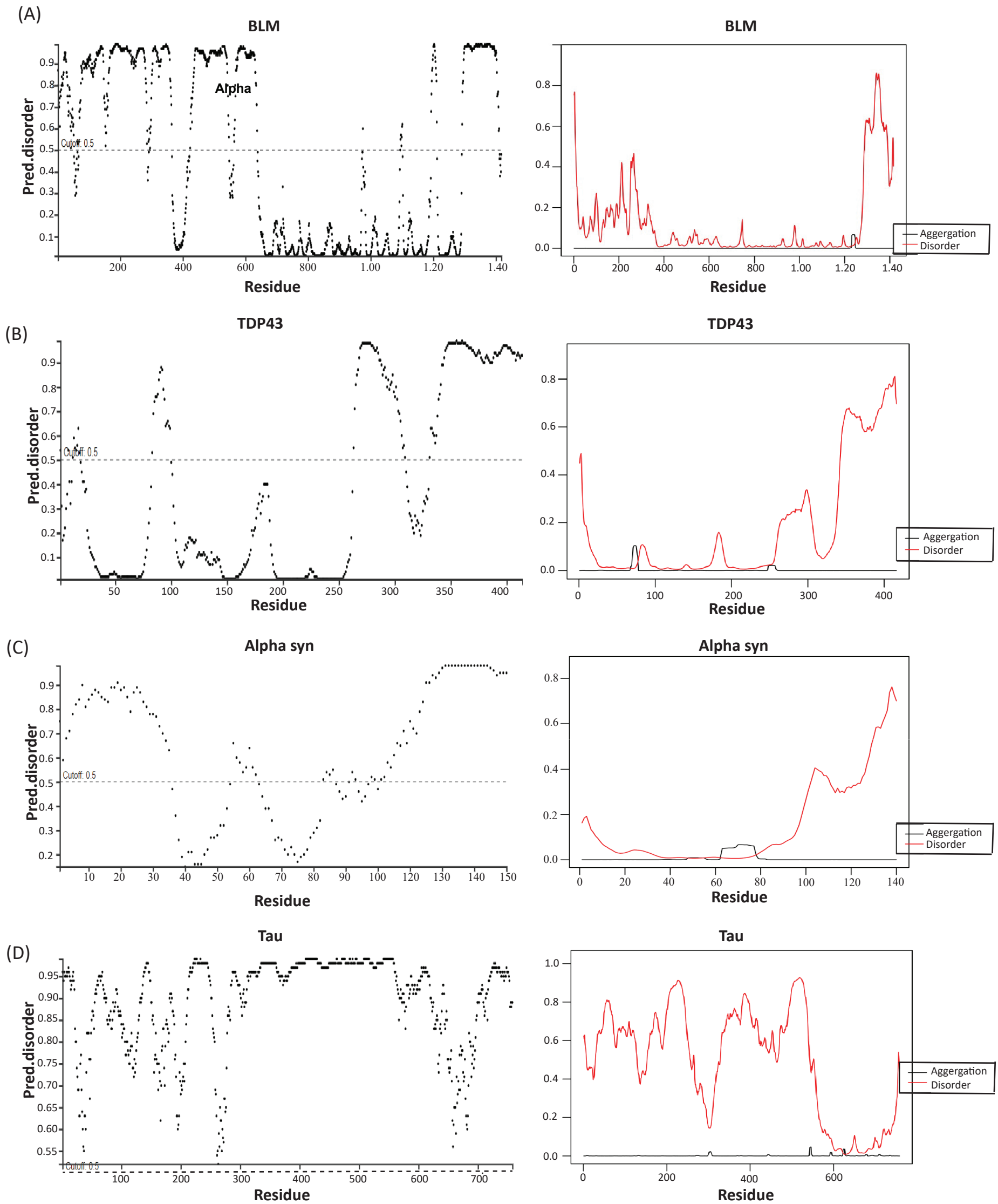

Figure 2: Depiction of Disordered and parallel aggregation formation plots: **A)** BLM (UniProt ID: P54132), **B)** TDP43 (UniProt ID: Q13148), **C)** α-Synuclein (UniProt ID: P37840), and **D)** Tau (UniProt ID: P10636) using DISOPRED3 and PASTA 2.0.

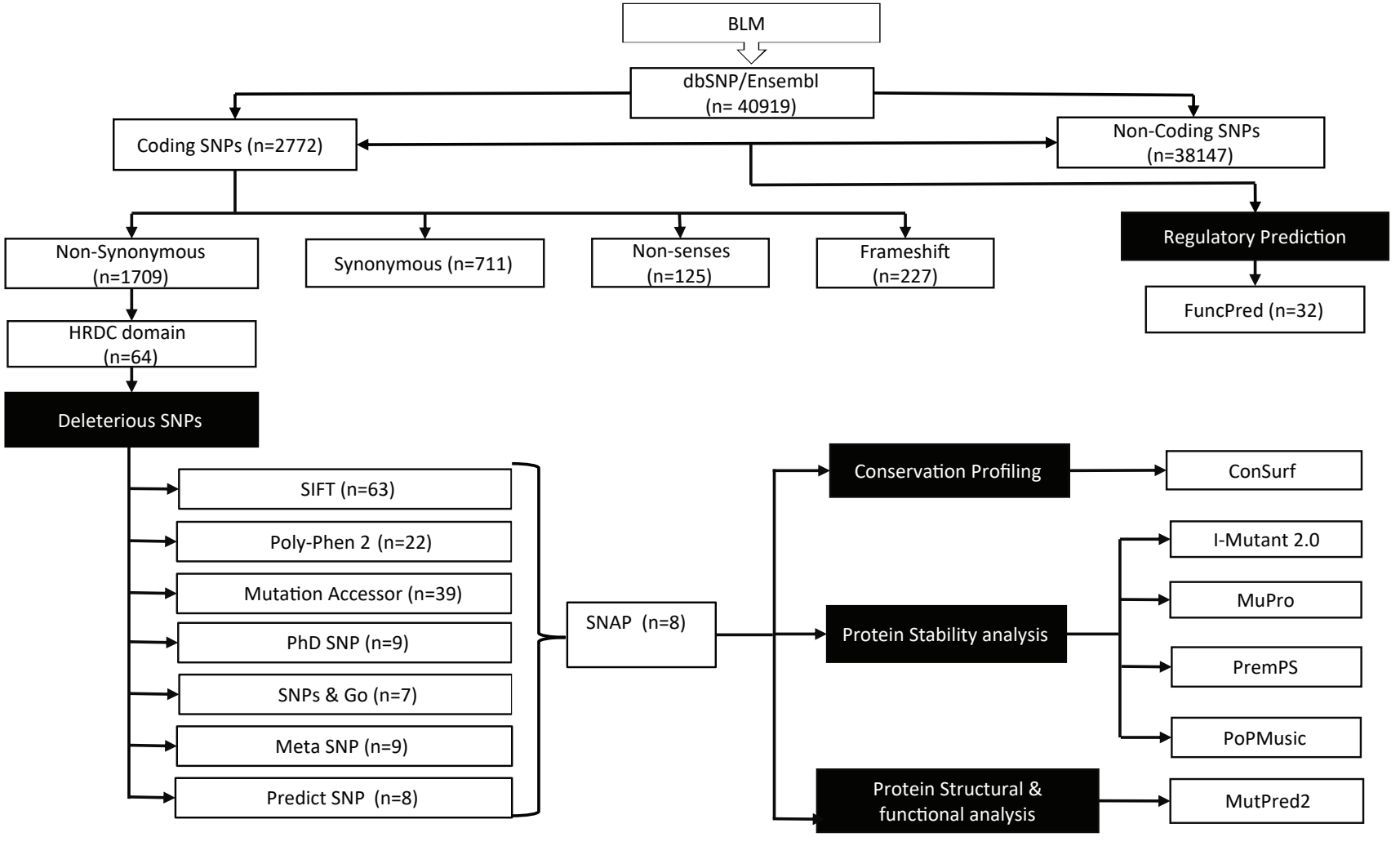

Figure 3: A schematic representation of nsSNPs selection and shortlisting procedure by different tools.

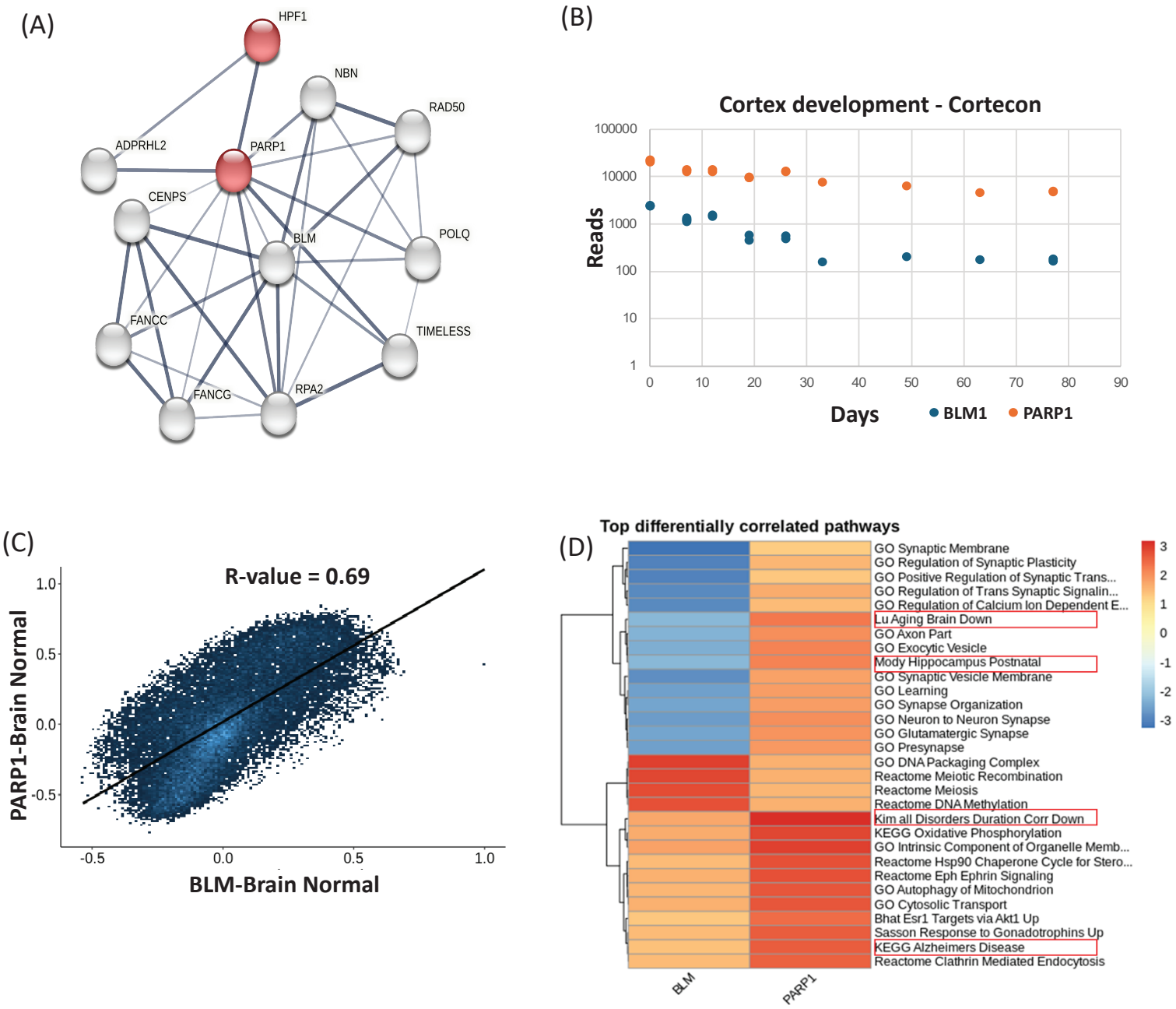

Figure 4: A) Network showing shared BLM and PARP implicated in repair mechanism. **B)** Similar pattern of BLM and PARP1 expression profiles during cortex development according to Cortecon. **D)** Scatter plot showing the relationship between genome-wide co-expression correlations for BLM and PARP1 in 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.