***In silico* study for decoding the correlated role of MCM7 gene in progression of breast cancer and Alzheimer's disorder**

Navneeth Sriram¹, Sunny Mukherjee¹, Mahesh Kumar Sah¹✉

¹Department of Biotechnology, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, Punjab-144011, India

*Corresponding author: Dr. M. K. Sah (sahmk@nitj.ac.in)

ABSTRACT

Breast cancer and Alzheimer's disease (AD) are two of the progressive and detrimental disorders affecting large population around the globe. While the chemotherapy of breast cancer is well established and enriched, the AD still lacks it due to unrecognized peripheral biomarkers for detection and targeted therapy. This study aimed to identify common molecular signature markers in breast cancer (grade 1, 2, and 3) and AD for the diagnosis and prognosis. We used two microarray datasets (GSE42568, GSE33000) respectively for both disorders that led to identification of two common differentially expressed genes (DEGs), namely MCM7 and CD209, as common players in both these two conditions. While the pattern of expression of CD209 gene running upregulated in both disorders, the MCM7 showed unusual contrary in its pattern of expression. The expression of MCM7 is downregulated in breast cancer but upregulated in AD. Gene set and protein overrepresentation analysis, protein-protein interaction (PPI), and protein subcellular localizations analyses of this underrated MCM7 gene was performed with further prediction and validation of its structure. The findings may pave the way in designing therapeutic approaches to ameliorate AD.

Keywords: Breast cancer, Alzheimer's disorder, CD209, MCM7, gene annotation, pathway analysis

1. INTRODUCTION

The modern technologies employed for delivering research and clinical treatment for cancer and neurodegenerative disorders have brought the diagnosis, prognosis, and treatment more effective and

personalized. While earlier, those were based on clinical-pathological observations, the present in silico assisted research revolutionized the efficacy and discovery of diagnostic and therapeutic agents. Among wide range of cancer types, breast cancer dominant *w.r.t.* high mortality and growth rate of new women patients. According to WHO report, in 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally. The subtypes of breast cancer such as ductal and lobular carcinoma, oestrogen receptor-positive (luminal A), HER2⁺ (Human Epidermal Growth Factor Receptor 2), triple negative and basal-like are common [1]. The histological grading of breast cancer indicates the extent of malignancy in a patient are grade 1 (well differentiated cancer cells growing slowly), grade 2 (moderately differentiated cancer cells) and grade 3 (poorly differentiated cancer cells growing faster)[2]. An exhaustive cancer research encompasses the use of microarray and sequencing-based technologies to analyse gene expression, to evaluate pathogenesis, and predict prognostic factors for various tumors[3]. The importance of the DNA replication licensing factors or MCMs in promoting expression coupled with altered phosphorylation of p53 that leads to drug resistance is reported earlier[4]. It is reported that at a cellular level, there is an inverted dynamic that happens between Cancer and Alzheimer's disease (AD), a complex and neurodegenerative disorder. AD is a common form of dementia in elders (>65 years), leading to cognitive impairment and memory loss and occurs due to neuron cells death, atrophy, and/or inflammation in hippocampus region of brain [5]. At a molecular level, the condition is primarily known by detecting the accumulation of amyloid β plaques, NFTs (Neurofibrillary Tangles) and the neurites containing the hyperphosphorylated tau protein [6]. The most common mutation that was found in them was the mutation in ϵ 4 allele of the ApoE gene and it increased the genetic risk factor for sporadic AD by 95% [7]. The most common genes whose mutation greatly increases the susceptibility to AD are APP (Amyloid Precursor protein), PS-1 (Presenilin-1) and PS-2 (Presenilin-2) respectively [8]. In addition, genome-wide association studies (GWASs) have identified several genes as potential risk factors for AD, such as *clusterin (CLU)*, *complement receptor 1 (CR1)*, *phosphatidylinositol binding clathrin assembly protein (PICALM)*, and *sortilin-related receptor (SORL1)* [9]. The recent inclusions are novel genes that might be engaged in triggering late-onset D, such as *triggering receptor expressed on myeloid cells 2 (TREM2)* and *cluster of differentiation 33 (CD33)* [9]. The identification of new AD-related genes

plays vital role in unravelling the pathogenesis mechanisms that may guide the discovery of new treatment strategies.

As reported earlier, in Cancer, there is a state of sustained cell signalling that results in its proliferation, whereas in AD, the proliferative signalling is impaired [10]. The cancer cells tend to evade the growth suppressors but these have a very strong effect on the AD-afflicted neurons [10]. The apoptotic processes are downregulated in Cancer while they are upregulated in AD [11]. Due to this property, studies are being carried out on the repurposing of anti-cancer drugs in order to cover both the conditions [12]. This is further supported by the finding that the patients who underwent chemotherapy had a lower risk of AD than those who hadn't undergone chemotherapy [13]. In the present work, the common genes for both breast cancer and the AD is analysed and their pattern of expression are further studied for possible repurposing of anti-cancer drugs of breast cancer for the treatment of AD. The article focusses on the effect of both the conditions on decoded gene functioning and to establish the regulatory effect of associated protein on AD.

2. EXPERIMENTAL DESIGN

The datasets for Cancer and Alzheimer's Disease were collected from GEO database of NCBI (<https://www.ncbi.nlm.nih.gov/geo/>) [14]. These datasets were compared on the basis of normal controls versus diseased conditions. To statistically analyze the data, the data were normalized and log transformation was applied. Further, Benjamini-Hochberg correction was used to eliminate the false discovery rate. The obtained output and the list of significant differentially expressed genes (DEGs) were downloaded ($P_{adj} < 0.01$). The genes in common were then identified with the help of VENNY 2.1.0, a Venn Diagram generating tool (https://bioinfogp.cnb.csic.es/tools/VENNY_2.1.0/). The gene ontology and gene enrichment studies were conducted using Gene Ontology Resource (<http://geneontology.org/>) [15] and ENRICH [16] respectively. Further, the analysis of protein-protein interaction of the target protein and their respective sub-cellular localization were determined. The domain analysis of the target gene, prediction of its structure followed by the validation of the analysis were performed. The workflow for the present methodology is depicted in **Figure 1**.

Figure 1: The flowsheet of methodology for the identification of potential genes associated with breast cancer and Alzheimer's disorder in common.

2.1. Datasets and identification of DEGs from microarray datasets

For collecting the DEGs associated with Breast Cancer, the gene expression profiles were studied (GSE42568) [17]. The total number of samples were 121, out of which 17 were controls, 11 were of Grade 1, 40 were of Grade 2 and 53 were of Grade 3. For collecting the DEGs associated with AD, the gene expression profile of human prefrontal cortex brain tissues (GSE33000) was analysed [18]. The total number of present samples were 621, out of which 467 from the AD patients while the rest 154 were samples from HD (Huntington's Disease) patients. Among 467, 135 were of Male AD, 123 were Male control, 175 were Female AD, and 34 were Female controls. Those 467 samples were analyzed for the study.

These samples were first selected, assigned a group, and then GEO2R tool was used to analyze the data. Benjamini-Hochberg correction was applied to the analysed data to eliminate the false discovery and further the data was normalized by log transformation. To further analyze the data, the LIMMA precision weights (Linear Models for Microarray Analysis) was also applied ($P < 0.01$). After applying all these modifications, the data was analyzed. To obtain the DEGs for these dataset, the plot displays were also selected, which were Breast cancer control vs Breast Cancer Grade 1; Breast Cancer Control vs Breast Cancer Grade 2 and Breast Cancer Control vs Breast Cancer Grade 3, Female AD vs Female Control and Male AD vs Male Control, respectively.

2.2. Gene ontology and enrichment study

In order to conduct gene ontology study, Gene Ontology Resource (<http://geneontology.org/>) web tool was used. The genes that were common in both the datasets were firstly identified with VENNY 2.1.0 (https://bioinfogp.cnb.csic.es/tools/VENNY_2.1.0/), a Venn Diagram generating tool. After identification, the common genes were searched and analysed for enrichment using ENRICHR online tool.

2.3. Protein-protein interaction analysis

The Protein-Protein interaction (PPI) network of the target protein unveiled after enrichment analysis, MCM7, was analysed using Cytoscape v3.8.2 [19]. The STRING plug-in of Cytoscape v3.8.2 was used to perform an automatic search for known interacting partners of target protein under study using data from the STRING network database [20]. Search parameters were set as follows: confidence score set at highest (0.99), additional interactors set at maximum (100) and smart delimiters were activated. Nodes represent the proteins and the edges represent the interactions among them. After the network was imported into Cytoscape, the individual interactions of each nodes were observed and image data was exported for further analyses.

2.5 Determination of the subcellular localization of interacting proteins

The WoLF PSORT [21] online server in combination with experimentally validated data from UniProt [22] were used to predict the subcellular localization of all the proteins present in the interaction network of target protein. The prediction of the subcellular localization is useful for determination of protein functions as well as the feasibility of a protein involved in the interaction network (direct/via cascade) to serve as a potential target during the drug discovery process.

2.6 In-silico domain analysis and structure prediction of target protein

Domain analysis is necessary for the prediction of active or functional sites on a protein. Since these are the regions that pharmaceutical ligands often bind to, those can be potentially served as drug targets. The domain analysis of target protein was performed using European Bio-Informatics Institute (EBI) INTERPRO [23] server after extraction of the FASTA sequence from UniProt. The secondary structure prediction of a protein, in order to understand its folded state as well as the presence of any loops or turns, was made using UCLCS Bio-Informatics group's PSIPRED (PSI-BLAST based secondary structure PREDiction) [24] online tool. 3D homology modelling was performed using SWISSMODEL [25] in order to validate pre-existing structure from RCSB PDB [26] and for the observation of various reactive amino acid groups. The predicted model was validated using the SAVES v6.0 utilizing the VERIFY3D [27] and PROCHECK [28]. Finally, all the predicted structures and published structures were compared for sequence similarity using PROMALS3D [29] server, identical sequences would conclusively validate the PDB structure for target protein.

3. RESULTS AND DISCUSSION

3.1. Identification of DEGs

After conducting the statistical analyses using GEO2R tools of NCBI, the Volcano plots of the four conditions against their respective controls was obtained (*Figure 2*). After obtaining the volcano plots, the significant genes were downloaded and analysed using VENN to find out the common genes that were present in the four conditions. After getting the common genes, the network was designed using

CYTOSCAPEv3.8.2 where the networks files were imported from STRING database. A high confidence (>90%) was set as a parameter to obtain the network.

After analysing the microarray data, it was determined that there were 4091 significant DEGs in Breast Cancer Grade 1, 6548 in Breast Cancer Grade 2 and 7081 in Breast Cancer Grade 3 and in Alzheimer's Disease dataset, there were 5 DEGs that were significant. Volcano plot was made where x-axis represented $\log_2(\text{fold change})$ and the y-axis represented the negative $\log(\text{fold change})$. The red points denoted the genes whose fold change was greater than 0 and the blue points denoted the genes whose fold change was less than 0. There were no significant genes for Male AD vs Male Control data. Upon determining the DEGs that were common in both Alzheimer's Disease and Breast Cancer datasets, it was found that two genes, namely, CD209 and MCM7 were the only ones present in both of them with CD209 showing a consistent downregulation pattern across the three grades of breast cancer and Alzheimer's Disease and MCM7 showing a consistent downregulation pattern against the three grades of breast cancer and an upregulation in case of Alzheimer's Disease. The gene MCM7 is chosen for further analyses.

Figure 2: Volcano plots of DEGs between breast cancer Grade 1 (A), Grade 2 (B), Grade 3 (C) and normal breast cancer tissue. Volcano plot of DEGs between Alzheimer's disordered and normal brain tissues (d). Red dots, significantly upregulated genes in breast cancer tissues; blue dots, significantly downregulated genes in breast cancer tissues; grey dots, not significantly expressed genes. Adj. $P < 0.01$ was considered to indicate a statistically significant difference.

Figure 3: Venn diagram of overlapping DEGs from the GSE42568 (breast cancer grade 1, grade 2, and grade 3) and GSE330000 datasets (Alzheimer's disorder). DEGs, differentially expressed genes.

Table 1: Fold change and F-value of common genes

Gene Symbol	Gene Title	Log2(fold change)				F-value			
		BCG1	BCG2	BCG3	AD	BCG1	BCG2	BCG3	AD
MCM7	Minichromosome Maintenance Complex Component 7	-0.176	-0.149	-0.207	0.399	3.471	3.908	7.155	6.407
CD209	CD209 molecule	0.917	0.862	0.844	0.366	16.766	23.925	24.647	9.523

Besides, Venny 2.1.0 was used to identify overlapping DEGs across the 4 datasets (**Figure 3**, **Table 1**).

Upon analysing the annotations obtained for MCM7, it is found that in case of biological process, it is involved in cell division(GO:0000278) and DNA replication(GO:0006260); in case of molecular function, it is involved in the initial stages of DNA replication (GO:1990518) and in case of cellular

compartment, this gene is present as a component of different complexes, most notably being protein-DNA complex(GO:0032993), DNA replication preinitiation complex(GO:0031261) and also is present as a component of nuclear chromosomes (GO:0000228).

Table 2: Gene Annotation for Biological Process of MCM7 (Gene Ontology Resource)

Category	GO ID	Term	P-value
Biological Process	GO:1902299	pre-replicative complex assembly involved in cell cycle DNA replication	0.000437
	GO:0006267	pre-replicative complex assembly involved in nuclear cell cycle DNA replication	0.000437
	GO:0000727	double-strand break repair via break-induced replication	0.000631
	GO:0006268	DNA unwinding involved in DNA replication	0.000825
	GO:0006271	DNA strand elongation involved in DNA replication	0.000971
	GO:0022616	DNA strand elongation	0.00121
	GO:0006270	DNA replication initiation	0.00214
	GO:0071364	cellular response to epidermal growth factor stimulus	0.00223
	GO:0070849	response to epidermal growth factor	0.00243
	GO:0033260	nuclear DNA replication	0.00286
	GO:0044786	cell cycle DNA replication	0.00291
	GO:0036388	pre-replicative complex assembly	0.0032
	GO:0000724	double-strand break repair via homologous recombination	0.00495
	GO:0000725	recombinational repair	0.00505
	GO:0032508	DNA duplex unwinding	0.0052
	GO:0032392	DNA geometric change	0.00554
	GO:0000082	G1/S transition of mitotic cell cycle	0.00612
	GO:0044843	cell cycle G1/S phase transition	0.00621
	GO:0071466	cellular response to xenobiotic stimulus	0.00631
	GO:0009410	response to xenobiotic stimulus	0.00665
	GO:0006261	DNA-dependent DNA replication	0.00947
	GO:0006302	double-strand break repair	0.00986
	GO:0006310	DNA recombination	0.0109
	GO:0065004	protein-DNA complex assembly	0.0124
	GO:0006260	DNA replication	0.0132
	GO:0044772	mitotic cell cycle phase transition	0.0139
	GO:0044770	cell cycle phase transition	0.0143
	GO:0071824	protein-DNA complex subunit organization	0.0143
	GO:0071103	DNA conformation change	0.0152

	GO:0042493	response to drug	0.0186
	GO:0008283	cell population proliferation	0.0232
	GO:0071363	cellular response to growth factor stimulus	0.0245
	GO:0006281	DNA repair	0.0258
	GO:0070848	response to growth factor	0.0259
	GO:1903047	mitotic cell cycle process	0.0333
	GO:0000278	mitotic cell cycle	0.037
	GO:0006259	DNA metabolic process	0.038
	GO:0006974	cellular response to DNA damage stimulus	0.0383
	GO:0034622	cellular protein-containing complex assembly	0.0418
Molecular Function	GO:1990518	single-stranded 3'-5' DNA helicase activity	0.000243
	GO:0043138	3'-5' DNA helicase activity	0.000874
	GO:0017116	single-stranded DNA helicase activity	0.000971
	GO:0003688	DNA replication origin binding	0.00117
	GO:0003678	DNA helicase activity	0.00369
	GO:0003697	single-stranded DNA binding	0.00578
	GO:0004386	helicase activity	0.00777
	GO:0140097	catalytic activity, acting on DNA	0.0103
	GO:0016887	ATPase activity	0.0237
	GO:0017111	nucleoside-triphosphatase activity	0.042
	GO:0016462	pyrophosphatase activity	0.0447
	GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	0.0448
	GO:0016817	hydrolase activity, acting on acid anhydrides	0.0448
	GO:0071162	CMG complex	0.000534
Cellular Component	GO:0031261	DNA replication preinitiation complex	0.000631
	GO:0042555	MCM complex	0.000631
	GO:0000781	chromosome, telomeric region	0.00728
	GO:0032993	protein-DNA complex	0.00869
	GO:0000228	nuclear chromosome	0.0111
	GO:0098687	chromosomal region	0.0163

Upon conducting the enrichment analysis, it is found that the fold enrichment of MCM7 exceeds 100 in cases of initiation of DNA replication (R-HSA-176974) and DNA replication (R-HSA-69306) processes. This also happens in G1/S transitions (R-HSA-69206), G2/M checkpoints (R-HSA-69481) and S phase (R-HSA-69242). The fold enrichment is less than 100 in case of progression of cell cycle (R-HSA-69278), which is consistent with the involvement of MCM7 in replication process.

Table 3: Gene Enrichment Analysis of MCM7

Term	Adjusted p-value	Associated genes
DNA Replication (hsa03030)	5.68×10^{-43}	RFC5;PRIM2;FEN1;RFC3;PCNA;RFC4;MCM7;PRIM1;RPA1;RPA2;POLA1;POLA2;RPA3;POLE2;MCM3;MCM4;MCM5;MCM6;POLE;MCM2
Cell Cycle (hsa04110)	4.31×10^{-39}	PCNA;MCM7;CDC7;CDC6;CCNA2;ORC5;CCNA1;ORC4;DBF4;ORC6;CDC45;ORC1;CCNE1;ORC3;CHEK1;ORC2;CDK2;MCM3;E2F1;CDK1;MCM4;MCM5;MCM6;MCM2
Nucleotide excision repair (hsa03420)	4.02×10^{-14}	RFC5;RFC3;PCNA;RFC4;RPA3;POLE2;RPA1;RPA2;POLE
Mismatch repair (hsa03430)	1.19×10^{-12}	RFC5;RFC3;PCNA;RFC4;RPA3;RPA1;RPA2
Cellular Senescence (hsa04218)	1.39×10^{-6}	CCNA2;CCNA1;CCNE1;CHEK1;CDK2;E2F1;CDK1
Base Excision Repair (hsa03410)	1.07×10^{-5}	FEN1;PCNA;POLE2;POLE
Homologous Recombination (hsa03440)	1.95×10^{-5}	RPA3;RPA1;RPA2;TOPBP1
Hepatitis B (hsa05161)	1.95×10^{-5}	CCNA2;CCNA1;PCNA;CCNE1;CDK2;E2F1
Viral Carcinogenesis (hsa05203)	5.78×10^{-5}	CCNA2;CCNA1;CCNE1;CHEK1;CDK2;CDK1
Human T-Cell Leukemia Virus 1 infection (hsa05166)	8.45×10^{-5}	CCNA2;CCNA1;CCNE1;CHEK1;CDK2;E2F1
DNA Replication R-HSA-69306	6.98×10^{-89}	PRIM2;FEN1;PCNA;MCM7;MCM8;PRIM1;GMNN;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;RFC5;GINS1;GINS2;CDT1;RFC3;RFC4;GINS3;GINS4;RPA1;RPA2;CDC7;CDC6;CCNA2;CCNA1;POLA1;POLA2;DBF4;POLE2;RPA3;CDK2;MCM3;MCM4;MCM5;MCM6;MCM2
Activation of the prereplicative complex R-HSA-68962	7.20×10^{-77}	PRIM2;MCM7;MCM8;PRIM1;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;POLE;CDT1;RPA1;RPA2;CDC7;CDC6;POLA1;POLA2;DBF4;RPA3;POLE2;CDK2;MCM3;MCM4;MCM5;MCM6;MCM2
Synthesis of DNA R-HSA-69239	6.59×10^{-76}	PRIM2;FEN1;PCNA;MCM7;MCM8;PRIM1;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;POLE;RFC5;GI

		NS1;GINS2;CDT1;RFC3;RFC4;GINS3;GINS4;RPA1;RPA2;CDC6;CCNA2;CCNA1;POLA1;POLA2;POLE2;RPA3;CDK2;MCM3;MCM4;MCM5;MCM6;MCM2
S-phase R-HSA-69242	4.84×10^{-74}	PRIM2;FEN1;PCNA;MCM7;MCM8;PRIM1;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;POLE;RFC5;GINS1;GINS2;CDT1;RFC3;RFC4;GINS3;GINS4;RPA1;RPA2;CDC6;CCNA2;CCNA1;POLA1;POLA2;CCNE1;POLE2;RPA3;CDK2;MCM3;MCM4;MCM5;MCM6;MCM2
G1/S Transition R-HSA-69206	1.01×10^{-67}	PRIM2;PCNA;MCM7;MCM8;PRIM1;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;CDT1;RPA1;RPA2;CDC7;CDC6;CCNA2;CCNA1;POLA1;POLA2;DBF4;CCNE1;POLE2;RPA3;CDK2;MCM3;CDK1;MCM4;MCM5;MCM6;MCM2
Cell Cycle R-HSA-1640170	6.16×10^{-65}	PRIM2;FEN1;PCNA;MCM7;MCM8;PRIM1;GMNN;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;CHEK1;ORC2;E2F1;TOPBP1;POLE;RFC5;GINS1;GINS2;CDT1;RFC3;RFC4;GINS3;GINS4;RPA1;RPA2;CDC7;CDC6;CCNA2;CCNA1;POLA1;POLA2;DBF4;CCNE1;POLE2;RPA3;CDK2;MCM3;CDK1;MCM4;MCM5;MCM6;MCM2
Mitotic G1-G1/S phases R-HSA-453579	2.03×10^{-64}	PRIM2;PCNA;MCM7;MCM8;PRIM1;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;CDT1;RPA1;RPA2;CDC7;CDC6;CCNA2;CCNA1;POLA1;POLA2;DBF4;CCNE1;POLE2;RPA3;CDK2;MCM3;CDK1;MCM4;MCM5;MCM6;MCM2
Cell cycle, Mitotic R-HSA-69278	2.91×10^{-64}	PRIM2;FEN1;PCNA;MCM7;MCM8;PRIM1;GMNN;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;RFC5;GINS1;GINS2;CDT1;RFC3;RFC4;GINS3;GINS4;RPA1;RPA2;CDC7;CDC6;CCNA2;CCNA1;POLA1;POLA2;DBF4;CCNE1;POLE2;RPA3;CDK2;MCM3;CDK1;MCM4;MCM5;MCM6;MCM2
DNA Replication pre-initiation R-HSA-69002	1.98×10^{-62}	PRIM2;MCM7;MCM8;PRIM1;GMNN;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;CDT1;RPA1;RPA2;CDC7;CDC6;POLA1;POLA2;DBF4;POLE2;RPA3;CDK2;MCM3;MCM4;MCM5;MCM6;MCM2
M/G1 Transition R-HSA-68874	1.98×10^{-62}	PRIM2;MCM7;MCM8;PRIM1;GMNN;MCM10;ORC5;ORC4;ORC6;CDC45;ORC1;ORC3;ORC2;E2F1;POLE;CDT1;RPA1;RPA2;CDC7;CDC6;POLA1;POLA2;DB

Figure 4: (A) Protein-protein interaction (PPI) network of MCM7 obtained using Cytoscape v3.8.2. (B) PPI Network between MCM7 and its first neighbours (undirected)

Figure 4 represents the Protein-protein interaction (PPI) network of MCM7 obtained using Cytoscape v3.8.2 (Figure 4A). Nodes represent the proteins and the edges represent the interactions among them. Yellow octagonal node illustrates the protein of interest which has been selected as the central hub. Red octagonal nodes portray the first neighbours (undirected) of MCM7 protein. Blue elliptical nodes depict those proteins that are interacting with MCM7 via another protein using a cascade mechanism. Figure 4B displays the PPI Network between MCM7 and its first neighbours (undirected) (isolated from Fig 4(A) network). Highest confidence score (0.99), maximum additional interactors (100) and smart delimiters were the parameters used for the construction of this PPI network.

Table 4: Description of cellular functions of all the proteins associated with the PPI network of MCM7

Gene	Function	Reference
GINS complex (GINS1,GINS2, GINS3,GINS4)	Composed of four genes, GINS1-4 involved in initiation of DNA replication and progression of replication forks.	[30]
CDT1 (DNA replication factor Cdt1)	Cooperates with CDC6 and origin recognition complexes during G1 phase to load MCMs onto DNA to give pre-replication complex.	[31]
Replication Factor C (RFC3,RFC4, RFC5)	Replication factor C (RFC), a pentameric AAA+ ATPase complex made up of RFC1–5, transports PCNA to primer-template junctions.	[32]
Replication protein A (RPA1,RPA2, RPA3)	Binds and stabilizes single-stranded DNA intermediates formed during DNA replication or stress as part of the heterotrimeric replication protein A complex (RPA/RP-A). Prevents reannealing while also recruiting and activating various proteins and complexes involved in DNA metabolism.	[33]
Minichromosome Maintenance Complex (MCM2,MCM3,MC M4,MCM5,MCM6,M CM7,MCM8,MCM10 ,MCMBP)	Hexameric helicase MCM2-7 required to unwind dsDNA to generate the ssDNA template for DNA replication. MCM8 and MCM10 not a part of the hexameric complex and MCMBP is the binding protein	[34]
Origin Recognition Complex (ORC1,ORC2,ORC3, ORC4,ORC5,ORC6)	In eukaryotes, the six-subunit origin recognition complex (ORC) binds to DNA to mark the site for replication initiation.	[35]
Cyclin-A1 and Cyclin-A2 (CCNA1,CCNA2)	The canonical cyclin A, known as cyclin A2, required for both DNA replication and mitosis, found in almost all cell types. Cyclin A1, the second A cyclin family member, is only expressed in germ line cells.	[36]
Cell Division Control proteins (CDC6, CDC7,CDC45)	Code for the proteins required for initiation of DNA replication.	[37]
Cyclin dependent Kinases (CDK1, CDK2)	Modulating the centrosome cycle and mitotic onset, controls the eukaryotic cell cycle; promotes G2-M transition and regulates G1 progress and G1-S transition via association with multiple interphase cyclins. Higher cells require it to enter S-phase and mitosis. CDK1 acts at the G1-S transition to promote the E2F transcriptional program and	[38,39]

	the start of DNA synthesis, and modulates G2 progression; controls the timing of mitosis/meiosis entry by controlling the subsequent activation of cyclin B/CDK1 by phosphorylation, and coordinates cyclin B/CDK1 activation at the centrosome and in the nucleus.	
DNA Primase (PRIM1, PRIM2)	Catalytic PRIM1 and regulatory PRIM2 subunits of DNA primase complex and also part of DNA polymerase Alpha complex. Key role in initiation of synthesis of DNA.	[40]
DNA Polymerase (POLA1, POLA2, POLE, POLE2)	POLA1 and POLA2 are the catalytic and accessory subunits of DNA polymerase alpha complex whereas POLE and POLE2 are the catalytic and accessory components of DNA polymerase epsilon. POLA1 and POLA2 involved in DNA replication; POLE and POLE2 involved in DNA repair.	[41,42]
Protein DBF4 homolog A (DBF4)	Involved in S phase progression and regulation of DNA replication initiation.	[43]
Proliferating Cell Nuclear Antigen (PCNA)	Codes for auxiliary protein of DNA polymerase delta that regulates eukaryotic DNA replication by increasing the polymerase's processability during leading strand elongation.	[44]
TIMELESS-interacting protein (TIPIN)	Plays an important role in the regulation of DNA replication and the stability of replication forks. It is critical for cell survival following DNA damage or replication stress.	[45]
G1/S-specific Cyclin E1(CCNE1)	Involved in control of cell cycle during G1/S transition.	[46]
Microtubule associated protein RP/EB family member 2(MAPRE2)	Key role in progression of mitosis and regulation of cell adhesion.	[47]
Leucine-rich repeat and WD repeat-containing protein 1(LRWD1)	Involved in G1/S transition. Recruits and stabilizes the ORC onto chromatin during G1 to establish the pre-replication complex (preRC) and to heterochromatic sites in post-replicated cells.	[48]
Transcription Factor E2F1 (E2F1)	Found in promoter regions of most of the genes involved in regulation of cell cycle or DNA replication.	[49]
Geminin (GMNN)	Inhibits DNA replication by not allowing the MCM complex to incorporate into pre-replication complex.	[50]
DNA topoisomerase 2-binding protein 1(TOPBP1)	Role in the resumption of stalled replication forks and the control of checkpoints. Binds both dsDNA and ssDNA breaks and nicks.	[51]
Serine/threonine-protein kinase Chk1(CHEK1)	Involved in checkpoint mediated cell cycle arrest and activating DNA repair in response to DNA damage or unreplicated DNA.	[52]

Cyclin-L1(CCNL1)	Involved in splicing of pre-mRNA and acts in association with cyclin dependent kinases.	[53]
Flap endonuclease 1 (FEN1)	Possesses both exonuclease and endonuclease activity and involved in DNA replication and DNA repair.	[54]

3.2 Subcellular localization of interacting proteins

Determination of subcellular localization of the target protein and its interacting partners was done using WoLF-PSORT server and a thorough literature survey using UniProt as a source. The results of the analysis have been discussed in the following Table 5 and a Venn diagram, prepared using VENN 2.1.0 online tool, representing the major intracellular regions of localization of these proteins have been enclosed under *Figure 5*.

Table 5: Predicted subcellular localization of MCM7 and all of its protein-protein interaction partners

Protein Name	Predicted Subcellular localization
PCNA, DBF4, POLE2, POLE, POLA2, POLA1, PRIM2, PRIM1, CDC7, CCNA1, GINS3, GINS2, GINS1, CDT1, RFC5, RFC4, RFC3, RPA1, MCM7, RPA2, RPA3, MCM8, MCM4, MCM6, MCM2, MCM3, MCM10, MCMBP, ORC1, ORC2, ORC3, ORC4, ORC5, ORC6, CCNE1, LRWD1, E2F1, TOPBP1, CCNL1	Nucleus
TIPIN, CDC45, CDC6, GINS4, MCM5, GMNN, CHEK1, CCNA2	Nucleus, Cytoplasm
CDK2, CDK1	Nucleus, Cytoplasm, Endosomes, Cytoskeleton
MAPRE2	Cytoskeleton
FEN1	Nucleus, Mitochondria

Figure 5: Total percentage of proteins localised in various subcellular compartments-represented by figure generated using Venny 2.1.0

3.3 In-silico domain analysis and structure prediction of target protein

For a complete in-silico proteomic analysis of the target protein MCM7, the first step involved the extraction of FASTA sequence of MCM7 from UniProt [22]. Then this FASTA sequence was used as a query to search for its known domains, conserved sites, motifs etc using the EBI INTERPRO [23] server. From the data obtained through UniProt, the major domains and their positions in the protein sequence are as follows: MCM_N (MCM-N terminal domain) (10-139), MCM_OB (oligonucleotide/oligosaccharide binding fold) (149-278), MCM_dom (322-541), MCM_2 (332-537), AAA+_ATPase (ATPases associated with diverse cellular activities), (373-526), AAA_5 (373-526), MCM_lid (558-640) [55]. The secondary structure of the protein was analyzed by using PSIPRED [24] from FASTA sequence to observe properties such as α -helix, β -sheet, turns and coils. The results are represented in Figure(6A) and Figure (6B). Results from additional analysis of data obtained through PSIPRED regarding the reactive nature of the constituent amino acids and their specific properties have been included in the Figure S1(A-C) of supplementary datasheet.

Figure 6B: Position dependent feature predictions are mapped onto the sequence schematic shown below. The line height of the Phosphorylation and Glycosylation features reflects the confidence of the residue prediction.

3.4 Protein modelling and validation of predicted structures

For performing the homology modelling, the FASTA sequence of then MCM7 protein was computed in BLASTp by using the Protein Data Bank (PDB) database as the search set. The results obtained through BLAST showed only protein structure (present in humans) with sequence similarity above the 80% threshold required for homology modelling. The protein also showed 100% identity with the input FASTA sequence of our target MCM7 protein-therefore ensuring a high degree of accuracy during modelling procedures. The structure for the protein (PDB ID: 6XTX) was extracted from the PDB database. Homology modelling of MCM7 was done using the automated homology protein modelling server of SWISS-MODEL. The user defined template algorithm was selected for computation and the structure extracted from PDB (PDB ID: 6XTX) was used as a template to the target sequence which is MCM7 protein sequence in FASTA format. The MCM7 protein model was verified using the Ramachandran plot from the MolProbity program and validated because the modelled protein's amino acid residues all fit within the Ramachandran plot's allowed regions (Figure S2). The MCM7 protein showed 1.17% MolProbity score, 92.33% residues were in the favoured residues, 0.78% in the outliers regions; and the Clash score was 0.59%. On a global scale, the QMEAN Z-score provided an estimate of the "degree of nativeness" of the structural features observed in the model. It indicates whether the model's QMEAN score is comparable to the expected score from experimental structures of comparable size. QMEAN Z-score value of approximately zero specifies superior quality between the modelled structure and experimental structures. If -4.0 or lower scores are obtained, then it indicates that the models is of low quality. The QMEAN Z-scores of the MCM7 protein showed -2.46 , and these results indicate that the proposed homology model is reliable and acceptable. The predicted homolog model

has been represented in Figure (7A) and Figure (7B). The representations of the analysis of other properties of the modelled protein has been included in the Figure S3 (A-D) of the supplementary data sheet [56–58].

Figure 7: 3D representation of predicted MCM7 protein model based on QMEAN parameter generated using SWISSMODEL. **Figure 7(A)** represents the molecular surface structure and **Figure 7(B)** represents the thread-ribbon structure. The colour grading is as follows: Orange= low confidence prediction/low QMEAN score/low sequence similarity with template structure, Purple= moderate

confidence prediction/moderate QMEAN score/moderate sequence similarity with template structure,
Blue= high confidence prediction/high QMEAN score/high sequence similarity with template structure.

The results obtained through SWISSMODEL were validated using SAVESv6.0 server through PROCHECK and VERIFY3D algorithms. The results obtained were satisfactory to validate the predicted model. The results are depicted in Figure (8A) and Figure (8B).

Figure 8A: Validation of predicted structure of MCM7 protein using PROCHECK

Figure 8B: Validation of predicted structure of MCM7 protein using VERIFY3D

(6ctx.pdb_chain2_s002:0.309951,model_01.pdb_chainF_s001:0.309951);

Figure 9: PROMALS3D phylogenetic tree prediction (6CTX represents PDB model template; model_01 represents structure predicted by SWISS-MODEL)

For the purpose of performing further validation, the obtained predicted structure was compared against the structure for the MCM7 protein obtained through PDB using the PROMALS3D online server. Although the multiple sequence alignment results in PROMALS3D show the presence of a few minor gaps in the alignment of the two sequences, phylogenetic guide tree results (**Figure 9**) show that these two protein structures are represented on equidistant nodes from the root node, confirming the high degree of similarity or identity between the two sequences.

4. CONCLUSION AND FUTURE PROSPECTS

MCM7 is upregulated in Alzheimer's Disease and downregulated in the three grades of Breast Cancer as per our analysis. This gene forms a key component in the DNA replication process and cell cycle progression. Thus, any mutation in this gene or any structural changes in the subsequent

protein affects the DNA replication process as well as the cell cycle, thereby increasing the probability of occurrence of cellular aberrations such as onset of cancer or changes in protein structure or folding resulting in onset of diseases such as Alzheimer's. In order to assess therapeutic options for breast cancer and/or Alzheimer's Disease in a shorter time-scale without using much capital investment on R&D of novel drugs and toxicological research, drug repurposing studies can be done in order to observe the effect of pre-existing drugs on rectifying expression levels of a mutated MCM7 gene. Based on the above-mentioned findings, novel drugs can also be synthesized which can act on rectifying any erroneous expression of the MCM7 gene.

ACKNOWLEDGEMENT

The authors acknowledge the COVID-19 pandemic to motivate for switching research work to computational biology and further extend gratitude towards the MHRD (Govt. of India) for financial support to the first author.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] J. Park, R.W. Wysocki, Z. Amoozgar, L. Maiorino, M.R. Fein, J. Jorns, A.F. Schott, Y. Kinugasa-Katayama, Y. Lee, N.H. Won, E.S. Nakasone, S.A. Hearn, V. Küttner, J. Qiu, A.S. Almeida, N. Perurena, K. Kessenbrock, M.S. Goldberg, M. Egeblad, Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps, *Science Translational Medicine*. 8 (2016) 361ra138-361ra138. <https://doi.org/10.1126/scitranslmed.aag1711>.
- [2] N. Olsson, P. Carlsson, P. James, K. Hansson, S. Waldemarson, P. Malmström, M. Fernö, L. Ryden, C. Wingren, C.A. Borrebaeck, Grading breast cancer tissues using molecular portraits, *Molecular & Cellular Proteomics*. 12 (2013) 3612–3623.
- [3] M.J. Van De Vijver, Y.D. He, L.J. Van't Veer, H. Dai, A.A. Hart, D.W. Voskuil, G.J. Schreiber, J.L. Peterse, C. Roberts, M.J. Marton, A gene-expression signature as a predictor of survival in breast cancer, *New England Journal of Medicine*. 347 (2002) 1999–2009.
- [4] S. Løkkegaard, D. Elias, C.L. Alves, M.V. Bennetzen, A.-V. Lænkholm, M. Bak, M.F. Gjerstorff, L.E. Johansen, H. Vever, C. Bjerre, T. Kirkegaard, B. Nordenskjöld, T. Fornander, O. Stål, L.S. Lindström, L.J. Esserman, A.E. Lykkesfeldt, J.S. Andersen, R. Leth-Larsen, H.J. Ditzel, MCM3 upregulation confers endocrine resistance in breast cancer and is a predictive marker of

- diminished tamoxifen benefit, *Npj Breast Cancer*. 7 (2021) 1–15. <https://doi.org/10.1038/s41523-020-00210-8>.
- [5] H. Querfurth, F. LaFerla, Alzheimer's disease., *The New England Journal of Medicine*. (2010). <https://doi.org/10.1056/NEJMra0909142>.
- [6] J.M. Long, D.M. Holtzman, Alzheimer Disease: An Update on Pathobiology and Treatment Strategies, *Cell*. 179 (2019) 312–339. <https://doi.org/10.1016/j.cell.2019.09.001>.
- [7] C.-C. Liu, T. Kanekiyo, H. Xu, G. Bu, Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy, *Nature Reviews Neurology*. 9 (2013) 106–118. <https://doi.org/10.1038/nrneurol.2012.263>.
- [8] B. Boland, A. Kumar, S. Lee, F.M. Platt, J. Wegiel, W.H. Yu, R.A. Nixon, Autophagy Induction and Autophagosome Clearance in Neurons: Relationship to Autophagic Pathology in Alzheimer's Disease, *J. Neurosci*. 28 (2008) 6926–6937. <https://doi.org/10.1523/JNEUROSCI.0800-08.2008>.
- [9] E. Bagyinszky, Y.C. Youn, S.S.A. An, S. Kim, The genetics of Alzheimer's disease, *Clin Interv Aging*. 9 (2014) 535–551. <https://doi.org/10.2147/CIA.S51571>.
- [10] D. Hanahan, R.A. Weinberg, Hallmarks of Cancer: The Next Generation, *Cell*. 144 (2011) 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>.
- [11] K.N.H. Nudelman, B.C. McDonald, D.K. Lahiri, A.J. Saykin, Biological Hallmarks of Cancer in Alzheimer's Disease, *Mol Neurobiol*. 56 (2019) 7173–7187. <https://doi.org/10.1007/s12035-019-1591-5>.
- [12] H.M. Snyder, T. Ahles, S. Calderwood, M.C. Carrillo, H. Chen, C.-C.H. Chang, S. Craft, P. De Jager, J.A. Driver, H. Fillit, D. Knopman, M. Lotze, M.C. Tierney, S. Petanceska, A. Saykin, S. Seshadri, D. Shineman, M. Ganguli, Exploring the nexus of Alzheimer's disease and related dementias with cancer and cancer therapies: A convening of the Alzheimer's Association & Alzheimer's Drug Discovery Foundation, *Alzheimer's & Dementia*. 13 (2017) 267–273. <https://doi.org/10.1016/j.jalz.2016.11.002>.
- [13] L. Frain, D. Swanson, K. Cho, D. Gagnon, K.P. Lu, R.A. Betensky, J. Driver, Association of cancer and Alzheimer's disease risk in a national cohort of veterans, *Alzheimer's & Dementia*. 13 (2017) 1364–1370. <https://doi.org/10.1016/j.jalz.2017.04.012>.
- [14] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall, K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N. Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets—update, *Nucleic Acids Research*. 41 (2013) D991–D995. <https://doi.org/10.1093/nar/gks1193>.
- [15] The Gene Ontology Consortium, The Gene Ontology resource: enriching a Gold mine, *Nucleic Acids Research*. 49 (2021) D325–D334. <https://doi.org/10.1093/nar/gkaa1113>.
- [16] M.V. Kuleshov, M.R. Jones, A.D. Rouillard, N.F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S.L. Jenkins, K.M. Jagodnik, A. Lachmann, M.G. McDermott, C.D. Monteiro, G.W. Gundersen, A. Ma'ayan, Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, *Nucleic Acids Res*. 44 (2016) W90–97. <https://doi.org/10.1093/nar/gkw377>.
- [17] C. Clarke, S.F. Madden, P. Doolan, S.T. Aherne, H. Joyce, L. O'Driscoll, W.M. Gallagher, B.T. Hennessy, M. Moriarty, J. Crown, S. Kennedy, M. Clynes, Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis, *Carcinogenesis*. 34 (2013) 2300–2308. <https://doi.org/10.1093/carcin/bgt208>.
- [18] M. Narayanan, J.L. Huynh, K. Wang, X. Yang, S. Yoo, J. McElwee, B. Zhang, C. Zhang, J.R. Lamb, T. Xie, C. Suver, C. Molony, S. Melquist, A.D. Johnson, G. Fan, D.J. Stone, E.E. Schadt, P. Casaccia, V. Emilsson, J. Zhu, Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases, *Molecular Systems Biology*. 10 (2014) 743. <https://doi.org/10.15252/msb.20145304>.
- [19] D. Otasek, J.H. Morris, J. Bouças, A.R. Pico, B. Demchak, Cytoscape Automation: empowering workflow-based network analysis, *Genome Biology*. 20 (2019) 185. <https://doi.org/10.1186/s13059-019-1758-4>.
- [20] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. von Mering, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, *Nucleic Acids Research*. 47 (2019) D607–D613. <https://doi.org/10.1093/nar/gky1131>.

- [21] P. Horton, K.-J. Park, T. Obayashi, K. Nakai, Protein subcellular localization prediction with wolfpsort, in: Proceedings of the 4th Asia-Pacific Bioinformatics Conference, PUBLISHED BY IMPERIAL COLLEGE PRESS AND DISTRIBUTED BY WORLD SCIENTIFIC PUBLISHING CO., 2005: pp. 39–48. https://doi.org/10.1142/9781860947292_0007.
- [22] The UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*. 49 (2021) D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
- [23] M. Blum, H.-Y. Chang, S. Chuguransky, T. Grego, S. Kandasaamy, A. Mitchell, G. Nuka, T. Paysan-Lafosse, M. Qureshi, S. Raj, L. Richardson, G.A. Salazar, L. Williams, P. Bork, A. Bridge, J. Gough, D.H. Haft, I. Letunic, A. Marchler-Bauer, H. Mi, D.A. Natale, M. Necci, C.A. Orengo, A.P. Pandurangan, C. Rivoire, C.J.A. Sigrist, I. Sillitoe, N. Thanki, P.D. Thomas, S.C.E. Tosatto, C.H. Wu, A. Bateman, R.D. Finn, The InterPro protein families and domains database: 20 years on, *Nucleic Acids Research*. 49 (2021) D344–D354. <https://doi.org/10.1093/nar/gkaa977>.
- [24] D.W.A. Buchan, D.T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on, *Nucleic Acids Research*. 47 (2019) W402–W407. <https://doi.org/10.1093/nar/gkz297>.
- [25] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Research*. 46 (2018) W296–W303. <https://doi.org/10.1093/nar/gky427>.
- [26] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, *Nucleic Acids Research*. 28 (2000) 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- [27] R. Lüthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature*. 356 (1992) 83–85. <https://doi.org/10.1038/356083a0>.
- [28] R.A. Laskowski, M.W. MacArthur, J.M. Thornton, PROCHECK: validation of protein-structure coordinates, *Urn:Isbn:978-0-7923-6857-1*. (2006). <https://doi.org/10.1107/97809553602060000724>.
- [29] J. Pei, B.-H. Kim, N.V. Grishin, PROMALS3D: a tool for multiple protein sequence and structure alignments, *Nucleic Acids Research*. 36 (2008) 2295–2300. <https://doi.org/10.1093/nar/gkn072>.
- [30] K. Kamada, Y. Kubota, T. Arata, Y. Shindo, F. Hanaoka, Structure of the human GINS complex and its assembly and functional interface in replication initiation, *Nature Structural & Molecular Biology*. 14 (2007) 388–396. <https://doi.org/10.1038/nsmb1231>.
- [31] J.G. Cook, D.A.D. Chasse, J.R. Nevins, The Regulated Association of Cdt1 with Minichromosome Maintenance Proteins and Cdc6 in Mammalian Cells*, *Journal of Biological Chemistry*. 279 (2004) 9625–9633. <https://doi.org/10.1074/jbc.M311933200>.
- [32] N.Y. Yao, M. O'Donnell, The RFC Clamp Loader: Structure and Function, in: S. MacNeill (Ed.), *The Eukaryotic Replisome: A Guide to Protein Structure and Function*, Springer Netherlands, Dordrecht, 2012: pp. 259–279. https://doi.org/10.1007/978-94-007-4572-8_14.
- [33] T.E. Bass, J.W. Luzwick, G. Kavanaugh, C. Carroll, H. Dungrawala, G.G. Glick, M.D. Feldkamp, R. Putney, W.J. Chazin, D. Cortez, ETAA1 acts at stalled replication forks to maintain genome integrity, *Nature Cell Biology*. 18 (2016) 1185–1195. <https://doi.org/10.1038/ncb3415>.
- [34] C.-W. Chang, M. Li, X. Xu, Y. Liu, Chapter 11 - Helicase Dysfunctions in Human Diseases, in: R. Tuteja (Ed.), *Helicases from All Domains of Life*, Academic Press, 2019: pp. 191–210. <https://doi.org/10.1016/B978-0-12-814685-9.00012-9>.
- [35] N. Li, W.H. Lam, Y. Zhai, J. Cheng, E. Cheng, Y. Zhao, N. Gao, B.-K. Tye, Structure of the origin recognition complex bound to DNA replication origin, *Nature*. 559 (2018) 217–222. <https://doi.org/10.1038/s41586-018-0293-x>.
- [36] L. Radonova, T. Pauerova, D. Jansova, J. Danadova, M. Skultety, M. Kubelka, M. Anger, Cyclin A1 in Oocytes Prevents Chromosome Segregation And Anaphase Entry, *Scientific Reports*. 10 (2020) 7455. <https://doi.org/10.1038/s41598-020-64418-1>.
- [37] S. Carbon, E. Douglass, B.M. Good, D.R. Unni, N.L. Harris, C.J. Mungall, S. Basu, R.L. Chisholm, R.J. Dodson, E. Hartline, P. Fey, P.D. Thomas, L.-P. Albou, D. Ebert, M.J. Kesling, H. Mi, A. Muruganujan, X. Huang, T. Mushayahama, S.A. LaBonte, D.A. Siegele, G. Antonazzo, H. Attrill, N.H. Brown, P. Garapati, S.J. Marygold, V. Trovisco, G. dos Santos, K. Falls, C. Tabone, P. Zhou, J.L. Goodman, V.B. Strelets, J. Thurmond, P. Garmiri, R. Ishtiaq, M. Rodríguez-López, M.L. Acencio, M. Kuiper, A. Lægreid, C. Logie, R.C. Lovering, B. Kramarz, S.C.C.

- Saverimuttu, S.M. Pinheiro, H. Gunn, R. Su, K.E. Thurlow, M. Chibucos, M. Giglio, S. Nadendla, J. Munro, R. Jackson, M.J. Duesbury, N. Del-Toro, B.H.M. Meldal, K. Paneerselvam, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, H.-Y. Chang, R.D. Finn, A.L. Mitchell, N.D. Rawlings, L. Richardson, A. Sangrador-Vegas, J.A. Blake, K.R. Christie, M.E. Dolan, H.J. Drabkin, D.P. Hill, L. Ni, D.M. Sitnikov, M.A. Harris, S.G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bähler, E.R. Bolton, J.L. De Pons, M.R. Dwinell, G.T. Hayman, M.L. Kaldunski, A.E. Kwitek, S.J.F. Laulederkind, C. Plasterer, M.A. Tutaj, M. Vedi, S.-J. Wang, P. D'Eustachio, L. Matthews, J.P. Balhoff, S.A. Aleksander, M.J. Alexander, J.M. Cherry, S.R. Engel, F. Gondwe, K. Karra, S.R. Miyasato, R.S. Nash, M. Simison, M.S. Skrzypek, S. Weng, E.D. Wong, M. Feuermann, P. Gaudet, A. Morgat, E. Bakker, T.Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C.N. Arighi, A. Auchincloss, K. Axelsen, G. Argoud-Puy, A. Bateman, M.-C. Blatter, E. Boutet, E. Bowler, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C.C. Casas, E. Coudert, P. Denny, A. Estreicher, M.L. Famiglietti, G. Georgiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, C. Hulo, A. Ignatchenko, F. Jungo, K. Laiho, P. Le Mercier, D. Lieberherr, A. Lock, Y. Lussi, A. MacDougall, M. Magrane, M.J. Martin, P. Masson, D.A. Natale, N. Hyka-Nouspikel, S. Orchard, I. Pedruzzi, L. Pourcel, S. Poux, S. Pundir, C. Rivoire, E. Speretta, S. Sundaram, N. Tyagi, K. Warner, R. Zaru, C.H. Wu, A.D. Diehl, J.N. Chan, C. Grove, R.Y.N. Lee, H.-M. Muller, D. Raciti, K. Van Auken, P.W. Sternberg, M. Berriman, M. Paulini, K. Howe, S. Gao, A. Wright, L. Stein, D.G. Howe, S. Toro, M. Westerfield, P. Jaiswal, L. Cooper, J. Elser, The Gene Ontology resource: enriching a GOLD mine, *Nucleic Acids Research*. 49 (2021) D325–D334.
- [38] Y. Mori, Y. Inoue, Y. Taniyama, S. Tanaka, Y. Terada, Phosphorylation of the centrosomal protein, Cep169, by Cdk1 promotes its dissociation from centrosomes in mitosis, *Biochemical and Biophysical Research Communications*. 468 (2015) 642–646. <https://doi.org/10.1016/j.bbrc.2015.11.004>.
- [39] B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase, *Nucleic Acids Research*. 48 (2020) D498–D503. <https://doi.org/10.1093/nar/gkz1031>.
- [40] B.E. Weiner, H. Huang, B.M. Dattilo, M.J. Nilges, E. Fanning, W.J. Chazin, An Iron-Sulfur Cluster in the C-terminal Domain of the p58 Subunit of Human DNA Primase*, *Journal of Biological Chemistry*. 282 (2007) 33444–33451. <https://doi.org/10.1074/jbc.M705826200>.
- [41] P. Starokadomskyy, T. Gemelli, J.J. Rios, C. Xing, R.C. Wang, H. Li, V. Pokatayev, I. Dozmorov, S. Khan, N. Miyata, G. Fraile, P. Raj, Z. Xu, Z. Xu, L. Ma, Z. Lin, H. Wang, Y. Yang, D. Ben-Amitai, N. Orenstein, H. Mussaffi, E. Baselga, G. Tadini, E. Grunebaum, A. Sarajlija, K. Krzewski, E.K. Wakeland, N. Yan, M.T. de la Morena, A.R. Zinn, E. Burstein, DNA polymerase- α regulates the activation of type I interferons through cytosolic RNA:DNA synthesis, *Nature Immunology*. 17 (2016) 495–504. <https://doi.org/10.1038/ni.3409>.
- [42] T. Dodd, M. Botto, F. Paul, R. Fernandez-Leiro, M.H. Lamers, I. Ivanov, Polymerization and editing modes of a high-fidelity DNA polymerase are linked by a well-defined path, *Nature Communications*. 11 (2020) 5379. <https://doi.org/10.1038/s41467-020-19165-2>.
- [43] A.W. Almawi, L. Langlois-Lemay, S. Boulton, J. Rodríguez González, G. Melacini, D. D'Amours, A. Guarné, Distinct surfaces on Cdc5/PLK Polo-box domain orchestrate combinatorial substrate recognition during cell division, *Scientific Reports*. 10 (2020) 3379. <https://doi.org/10.1038/s41598-020-60344-4>.
- [44] D.S. Kulkarni, S.N. Owens, M. Honda, M. Ito, Y. Yang, M.W. Corrigan, L. Chen, A.L. Quan, N. Hunter, PCNA activates the MutLy endonuclease to promote meiotic crossing over, *Nature*. 586 (2020) 623–627. <https://doi.org/10.1038/s41586-020-2645-6>.
- [45] J. Rageul, J.J. Park, P.P. Zeng, E.-A. Lee, J. Yang, S. Hwang, N. Lo, A.S. Weinheimer, O.D. Schärer, J.-E. Yeo, H. Kim, SDE2 integrates into the TIMELESS-TIPIN complex to protect stalled replication forks, *Nature Communications*. 11 (2020) 5495. <https://doi.org/10.1038/s41467-020-19162-5>.
- [46] C.C. Lepage, M.C.L. Palmer, A.C. Farrell, N.M. Neudorf, Z. Lichtensztejn, M.W. Nachtigal, K.J. McManus, Reduced SKP1 and CUL1 expression underlies increases in Cyclin E1 and

- chromosome instability in cellular precursors of high-grade serous ovarian cancer, *British Journal of Cancer*. 124 (2021) 1699–1710. <https://doi.org/10.1038/s41416-021-01317-w>.
- [47] C. Thues, J.S. Valadas, L. Deaulmerie, A. Geens, A.K. Chouhan, R. Duran-Romaña, J. Schymkowitz, F. Rousseau, M. Bartusel, R. Rehimi, A. Rada-Iglesias, P. Verstreken, H. Van Esch, MAPRE2 mutations result in altered human cranial neural crest migration, underlying craniofacial malformations in CSC-KT syndrome, *Scientific Reports*. 11 (2021) 4976. <https://doi.org/10.1038/s41598-021-83771-3>.
 - [48] K.M. Chan, Z. Zhang, Leucine-rich Repeat and WD Repeat-containing Protein 1 Is Recruited to Pericentric Heterochromatin by Trimethylated Lysine 9 of Histone H3 and Maintains Heterochromatin Silencing*, *Journal of Biological Chemistry*. 287 (2012) 15024–15033. <https://doi.org/10.1074/jbc.M111.337980>.
 - [49] S. Manickavinayam, R. Vélez-Cruz, A.K. Biswas, E. Bedford, B.J. Klein, T.G. Kutateladze, B. Liu, M.T. Bedford, D.G. Johnson, E2F1 acetylation directs p300/CBP-mediated histone acetylation at DNA double-strand breaks to facilitate repair, *Nature Communications*. 10 (2019) 4951. <https://doi.org/10.1038/s41467-019-12861-8>.
 - [50] E. Sami, D. Bogan, A. Molinolo, J. Koziol, W.M. ElShamy, The molecular underpinning of geminin-overexpressing triple-negative breast cancer cells homing specifically to lungs, *Cancer Gene Therapy*. (2021) 1–22. <https://doi.org/10.1038/s41417-021-00311-x>.
 - [51] C. Mooser, I.-E. Symeonidou, P.-A. Leimbacher, A. Ribeiro, A.-M.K. Shorrocks, S. Jungmichel, S.C. Larsen, K. Knechtle, A. Jasrotia, D. Zurbruggen, A. Jeanrenaud, C. Leikauf, D. Fink, M.L. Nielsen, A.N. Blackford, M. Stucki, Treacle controls the nucleolar response to rDNA breaks via TOPBP1 recruitment and ATR activation, *Nature Communications*. 11 (2020) 123. <https://doi.org/10.1038/s41467-019-13981-x>.
 - [52] H.-H. Chao, I.V. Karagounis, C. Thomas, N.B. François, A. Facciabene, C. Koumenis, A. Maity, Combination of CHEK1/2 inhibition and ionizing radiation results in abscopal tumor response through increased micronuclei formation, *Oncogene*. 39 (2020) 4344–4357. <https://doi.org/10.1038/s41388-020-1300-x>.
 - [53] L.B. Ameur, P. Marie, M. Thenoz, G. Giraud, E. Combe, J.-B. Claude, S. Lemaire, N. Fontrodona, H. Polveche, M. Bastien, A. Gessain, E. Wattel, C.F. Bourgeois, D. Auboeuf, F. Mortreux, Intragenic recruitment of NF- κ B drives splicing modifications upon activation by the oncogene Tax of HTLV-1, *Nature Communications*. 11 (2020) 3045. <https://doi.org/10.1038/s41467-020-16853-x>.
 - [54] J.S. Williams, P.P. Tumbale, M.E. Arana, J.A. Rana, R.S. Williams, T.A. Kunkel, High-fidelity DNA ligation enforces accurate Okazaki fragment maturation during DNA replication, *Nature Communications*. 12 (2021) 482. <https://doi.org/10.1038/s41467-020-20800-1>.
 - [55] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: The protein families database in 2021, *Nucleic Acids Research*. 49 (2021) D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
 - [56] E. Lanzarotti, R.R. Biekofsky, D.A. Estrin, M.A. Marti, A.G. Turjanski, Aromatic–Aromatic Interactions in Proteins: Beyond the Dimer, *J. Chem. Inf. Model*. 51 (2011) 1623–1633. <https://doi.org/10.1021/ci200062e>.
 - [57] A.A. Morgan, E. Rubenstein, Proline: The Distribution, Frequency, Positioning, and Common Functional Roles of Proline and Polyproline Sequences in the Human Proteome, *PLoS One*. 8 (2013). <https://doi.org/10.1371/journal.pone.0053785>.
 - [58] C. Wiedemann, A. Kumar, A. Lang, O. Ohlenschläger, Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR, *Front. Chem*. 8 (2020). <https://doi.org/10.3389/fchem.2020.00280>.