Roles for transcription factor RPRD proteins in mRNA synthesis and R-loop formation

by

Hürmüz CEYLAN

UNIVERSITY OF BIRMINGHAM

School of Bioscience

The University of Birmingham

2023

ABSTRACT

Transcription is a process of copying DNA into RNA. The transcription pattern determines the identity of the cell, and transcriptional regulation plays an essential role in this process. The primary protein driving this process is RNA polymerase II (RNAPII), which interacts with other regulatory proteins via its C-terminal domain (CTD). The RPRD1A, RPRD1B, and RPRD2 proteins have a CTD-interacting domain (CID), although their specific roles and functions in transcriptional processes remain inadequately understood.

My research demonstrates that depletion of RPRD1A, RPRD1B, and RPRD2 leads to an increase in nascent RNA levels, while their overexpression has the opposite effect. However, these proteins did not significantly affect total RNA levels. I found that deregulation of RPRD2 affects the stability of RNAs inversely with RPRD2 levels. Furthermore, RPRD protein levels correlated with the level of R-loops, and increased R-loops in RPRD up-regulated cells induced the DNA damage response. As a byproduct of transcription, the alteration in transcription rate and the ensuing DNA damage caused by the accumulation of R-loops lead to the inhibition of cell cycle progression. I analyzed the glioma dataset and found that higher expression of RPRD1B proteins increased glioma aggressiveness, while low expression of RPRD1A and RPRD2 decreased patient survival times.

My results suggest that RPRD proteins negatively impact newly transcribed RNA by interacting and recruiting other transcription factors to the site. As a response, cells change the stability of existing mRNAs in order to maintain cellular homeostasis. However, cells with transcription control problems may not respond appropriately to metabolic conditions, internal or external signals, or stress, despite adjusted mRNA levels. Accumulated errors could lead to the transformation of healthy cells into cancerous ones.

Acknowledgements

I would like to express my deepest appreciation to my esteemed thesis advisor, Dr. Pawel Grzechnik, for his invaluable support and guidance throughout my research. Additionally, I am grateful to my second supervisor, Dr. Saverio Brogna, for his support and contributions to my work. I extend special thanks to Dr. Kinga Winczura for generously sharing her knowledge and experience, which greatly contributed to the success of my project.

I cannot adequately express my gratitude to my professor and committee chair, as their patience and feedback have been instrumental in shaping my research. I am also deeply grateful to my defense committee, whose knowledge and expertise were indispensable to my research.

I would also like to express my heartfelt appreciation to my labmates, Dr. Monika Sledziowska, Onofrio Zanin, and Aytaj Nabiyeva, for their valuable collaboration, support, and the wonderful memories I have shared over the last four years. I am also grateful to my dear friends, whose constant presence, encouragement, and support have made the research and writing process less arduous.

I am thankful to the Turkish Ministry of National Education for providing me with a generous scholarship that made my studies possible.

I would not have been able to pursue my research without the unwavering support and encouragement of my parents, Elif and Adem Ceylan. Throughout my studies and research, they have been a constant source of inspiration and motivation. A special thanks also goes to my siblings, Sinem Ceylan, Selda Ceylan, and Furkan Ceylan, for their support and for always being there for me.

LIST OF CONTENT

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

4sU	4-thiouridine
APC	Anaphase promoting complex
AP-MS	Affinity purification-mass spectrometry
BRE	B-recognition element
CC	Coil-coil domain
CDK	Cyclin- dependent kinase
CDK7	Cyclin- dependent kinase-7
CDK9	Cyclin- dependent kinase-9
cDNA	Complementary deoxyribonucleic acid
ChIP	Chromatin immunoprecipitation
CID	CTD-interacting domain
CKI	Cyclin-dependent kinase inhibitor
COSMIC	Catalogue of Somatic Mutations in Cancer
CPSF	Cleavage and polyadenylation specificity factor
CREPT	Cell cycle-related and expression elevated protein in tumor
CstF	Cleavage stimulation factor
CTD	C-terminal domain
CTDP1	CTD Phosphatase Subunit 1
DCE	Downstream core element
DMEM	Dulbecco's Modified Eagle Medium
DPE	Downstream promoter element
DRB	5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole

DRIP	DNA-RNA immunoprecipitation
DSB	Double-strand breaks
dsDNA	Double-stranded
DSIF	DRB Sensitivity Inducing Factor
DTB	Double thymidine block
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FBS	Foetal Bovine Serum
GFP	Green fluorescent protein
GMP	Guanosine monophosphate
HDAC2	Histone Deacetylase 2
HIV	Human immunodeficiency virus
IF	Immunofluorescence
Inr	Initiator element
KD	Knockdown
Kub5-	Ku70-binding protein 5- Hera
lncRNA	Long noncoding ribonucleic acid
ncRNA	Non-coding ribonucleic acid
NELF	Negative elongation factor
OE	Overexpression
pA	Polyadenylation
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

PIC Preinitiation complex

p-TEFb Positive transcription elongation factor

q-DRIP Quantitative differential DNA–RNA immunoprecipitation

qPRC Quantitative polymerase chain reaction

REAF RNA-associated early-stage antiviral factor

RECQL5 ATP-dependent DNA helicase Q5

RNA Ribonucleic acid

RNAPII RNA polymerase II

RNAseH Ribonuclease H

Rpb1 RNA polymerase II subunits 1

RPAP2 RNA Polymerase II Associated Protein 2

RPRDs Regulation of Nuclear Pre-mRNA Domain Containing proteins

RPRD1A Regulation of Nuclear Pre-mRNA Domain Containing 1A

RPRD1B Regulation of Nuclear Pre-mRNA Domain Containing 1B

RPRD2 Regulation of nuclear pre-mRNA domain containing 2

RT Reverse transcriptase

SDS Sodium dodecyl sulfate

ssDNA single-stranded Deoxyribonucleic acid

STRING Search Tool for the Retrieval of Interacting Genes/Proteins

TAF TBP-associated factors

TBP TATA-binding protein

TCF4 Transcription Factor 4

TFIIB Transcription factor II B

TFIID	Transcription factor II D
TFIIF	Transcription factor II F
TFIIH	Transcription factor II H
TFs	Transcription factors
TFSII	Transcription elongation factor II S
tRNA	Transfer ribonucleic acid
TSS	Transcription start site
TTS	Transcription termination site
XBP	X-box binding protein 1

PART 1: INTRODUCTION

Chapter 1: GENE EXPRESSION

The genetic information that defines an organism's structure and function is entirely contained in DNA, with functional units called genes composing the genomic DNA. Each gene produces a functional product that is essential for performing a specific task within the cell. The inheritance of genetic information and its conversion from one form to another is facilitated by three different processes: replication, transcription, and translation. This genetic information flow is referred to as the central dogma (Crick, 1970).

Although the majority of cells within an organism contain identical copies of the genome, they differentiate into several types to organize the entire system. The usage of different portions of the genome, the transcription of different genes, and changes in the expression level lead to the creation of different types of cells, including cancer cells (Weinberg, 2013).

In response to environmental changes, cellular protein levels must be dynamically controlled. Controlling the abundance of proteins plays a vital role in maintaining cellular structure and function, as well as cellular development and differentiation. The expression of proteins is a complex and multi-step process that involves transcription, mRNA splicing, translation, and post-transcriptional modifications. Gene expression is responsible for linking information encoded in a gene to a final functional gene product, such as a protein-coding or non-coding RNA (ncRNA). The capability of controlling gene expression allows cells to produce functional proteins whenever they require them for normal functioning or survival. Transcription is the first step in gene expression (Gökbuget & Blelloch, 2019; Watson et al., 2015).

The mRNA turnover is one of the important steps in this regulation. The abundance of RNA in the cell is arranged through the synthesis of new RNAs from DNA templates, degradation of pre-existing RNA, and changes in the stability of produced RNA. In order to respond quickly, cells regulate the level of RNA in the cytoplasm. This mechanism is relied upon by numerous physiological and pathological processes, including cellular adaptations to novel environments, maintenance of homeostasis, and recovery from damage (Alberts et al., 2014; Watson et al., 2015).

1. TRANSCRIPTION

Transcription is a complex, multistep, and highly regulated process that converts DNA information into RNA. The most important protein that drives transcription is RNA polymerase enzyme which synthesizes RNA using a DNA template. In eukaryotic cells, DNA is packaged into nucleosomes and higher-order chromatin structures within the nucleus. Both bacteria and eukaryotic organisms utilize RNA polymerase enzymes to transcribe RNA from a DNA template. However, while bacteria and archaea possess only one RNA polymerase, eukaryotic cells contain three distinct multi-subunit RNA polymerases (I, II, III). These three RNA polymerases transcribe the three primary classes of genes in eukaryotic cells (Cramer, 2019; Cramer et al., 2008).

- RNA Polymerase I transcribes larger ribosomal RNA precursors.
- RNA Polymerase II transcribes mRNAs and various non-coding RNAs.
- RNA Polymerase III transcribes transfer RNAs and small ribosomal RNA.

Eukaryotic RNA polymerase I, II, and III are complex protein assemblies and have 14, 12, and 17 subunits, respectively. All three RNA polymerases in eukaryotic cells require other transcription factors to recognize promoter regions. Basal transcription factors bind to promoter regions to provide a binding site for the appropriate RNA polymerase. These basal factors are less complex for RNA polymerase I and III than for RNA polymerase II (RNAPII). Thanks to its numerous accessory factors, RNAPII has a much wider recognition range for various promoter regions (Cramer et al., 2008; Turowski & Boguta, 2021).

The C-terminal domain (CTD) RNAPII is a flexible, long structure that is composed of repeating units of a heptapeptide motif Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7. The CTD functions as a platform for the recruitment of a diverse range of factors that are involved in various stages of transcription such as initiation, elongation, and termination thanks to dynamic post-transcriptional events. The process of transcription by RNAPII is a highly regulated and intricate process facilitated by the CTD (Buratowski, 2003, 2009; Egloff et al., 2012; Eick & Geyer, 2013). In order to initiate transcription, RNAPII needs to bind a special region of a gene called promoter. RNAPII initiates transcription at a specific region of the gene and then progresses along the gene until it reaches the terminator region, which signifies the end of the transcription process. During this movement, lots of regulatory factors are required for robust and accurate transcription (Casamassimi & Ciccodicola, 2019; Cramer, 2019).

The promoter region may exist in three distinct states in terms of its activity level within a cell. The gene promoter can be surrounded by a compact and tightly packed chromatin structure, known as closed chromatin, which can render the promoter inaccessible to transcription factors and RNA polymerase, resulting in gene inactivation or silencing. Secondly, it is a gene promoter that is opened chromatin structure and bound to RNA polymerase however, it cannot start to transcription. This type of promoter needs some second class of signalling factors to initiates transcription such as heat shock gene promoter. Lastly, it is an active gene promoter that is found in opened chromatin structure and active state like housekeeping genes (Cramer, 2019).

A canonical protein-coding gene is comprised of several discrete regions, each with a specialized function during the processes of transcription and subsequent mRNA translation. These distinct regions are commonly referred to as the promoter region, 5' untranslated region (5' UTR), coding sequence (CDS), 3' untranslated region (3' UTR), terminator region, and polyadenylation signal. The promoter region function as a initiation site of transcription and located at the 5’ upstream of a protein coding gene.

The 5’UTR is coming following the promoter region and has role in regulation of translation which is a process to convert mRNA into functional protein. This region consists specific sequences including a ribosome binding and the formation of translation initiation complexes. Within the coding sequence (CDS), there are intronic (introns) and exonic (exons) regions . Introns are found in the gene after transcription and are removed during splicing process. Introns have crutial role in gene regulation and alternative splicing. Exons are the regions which contain essential information of protein sequence and remains after splicing in the mature mRNA. The 3’ UTR are located the downstream of the coding sequence and involves regulatory function for mRNA stability and localization within the cell. Following the 3’ UTR region, the terminator region is situated at the 3’ end of the gene and serves as a signal for releasing of RNA polymerase to terminate transcription. The polyadenylation signal creates a signal for addition of a polyadenine (poly-A) tail to the mRNA during post-transcriptional processing. The poly-A tail plays a crucial role in mRNA stability, transport, and translation (Barrett et al., 2012).

A promoter region is a DNA sequence where the RNAPII binds to start transcription of a gene. In eukaryotic cells, the core promoter region typically comprises approximately 40-60 nucleotides, which extend both upstream and downstream of the transcription start site. All promoters have conserved and characteristic sequences that are recognized by different type of basal transcription factors. These DNA sequence elements are BRE (TFIIB recognition element), TATA box, Inr (initiator element), DCE (downstream core element), and DPE (downstream promoter element). TATA box and DPE regions are not found at the same time in a gene. Inr is the most prevalent element for promoters and can be found either with a TATA box or a DPE. In order to maintain an efficient transcription, some other regulatory sequences are required, for instance, promoter-proximal elements, enhancers, silencers, upstream activator sequences, and insulators (Andersson & Sandelin, 2020; Haberle & Stark, 2018) (**Figure 1**). These regulatory sequences can be located near the promoter or thousand base pairs upstream or downstream of the gene. Recognition of these regulatory regions by some regulatory proteins is important for the progression or inhibition of transcription (Core & Adelman, 2019; Fuda et al., 2009; Haberle & Stark, 2018). As previously stated, transcription is a multifaceted biological process characterized by three primary sequential phases: initiation, elongation, and termination.

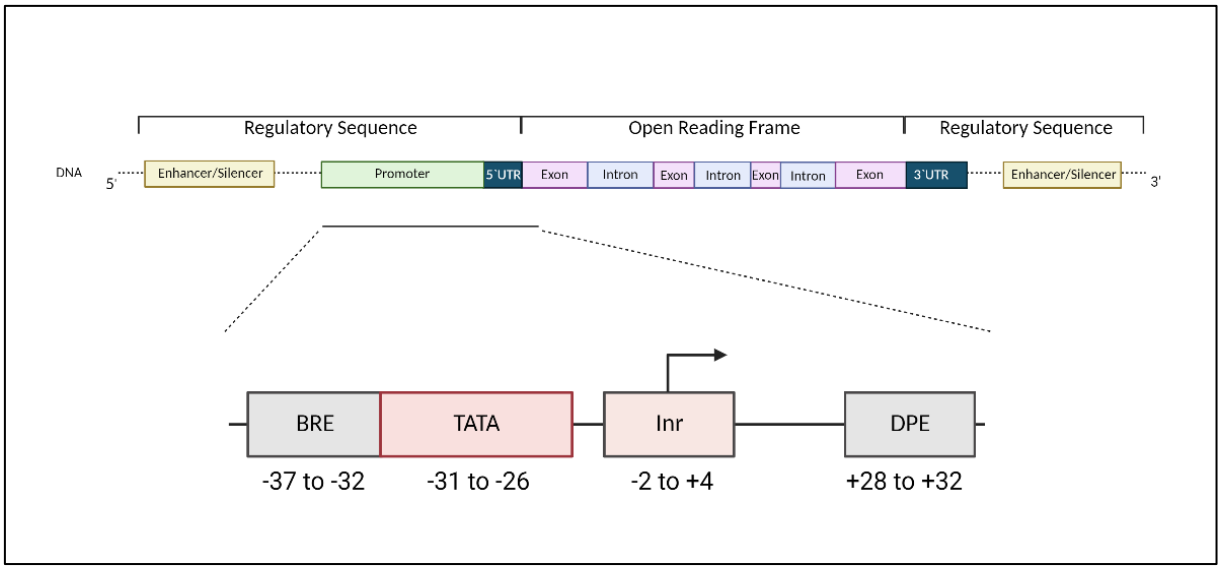


Figure 1: **Eukaryotic gene structure.**

The schematic of the eukaryotic gene structure, including the regulatory elements and open reading frame. The expression of the protein coding region is governed by various regulatory sequences. Transcription of the gene into mRNA is regulated by promoters (represented in green) and enhancers/silencers (represented in yellow). The untranslated regions (UTRs) located at the 5' and 3' ends of the mRNA molecule (represented in dark green) regulate translation into final proteins. The core promoter elements, namely B recognition element (BRE), TATA box, initiator motif (Inr), and downstream promoter element (DPE), are shown along with their proximity to the transcription start site

1.3.1 TRANSCRIPTION INITIATION

To initiate protein coding gene transcription, RNA polymerases II and general transcription factors bind to the promoter region of the gene to form the preinitiation complex (PIC) (Roeder, 1996). Upon transitioning from the closed to the open state, the two DNA strands are melted, and the transcription start site in the template strand is positioned at the active site of RNAPII which can initiate the synthesis of RNA using the template DNA strand to guide ribonucleotide selection and polymerization chemistry. Initiation steps include specific sub-steps: separation of DNA template, formation of the PIC on the promoter, abortive initiation, and promoter escape (Cramer, 2004, Cramer et al. 2019).

Eukaryotic DNA is compacted into chromatin structures which poses a significant obstacle for RNAPII to access the promoter region and initiate transcription. Chromatin remodelers and remodeling complexes play a crucial role in the regulation of gene expression by modulating chromatin structure to promote or inhibit access to the promoter region by RNAPII. These complexes alter the structure of chromatin by repositioning, evicting or modifying the histones around which DNA is wound. This allows RNAPII to gain access to the promoter region and initiate transcription efficiently (Li et al., 2007).

The TATA box containing promoters are well-studied promoters that are common for RNAPII expression. It is located approximately 25 bp upstream of the start point and contains an A-T rich region surrounded by a G-C rich sequence (Struhl, 1995). The protein called TATA-binding protein (TBP) that can recognize and directly bind to the TATA box region. TBPs help position RNAPII to the promoter for all types of RNA polymerase in eukaryotes. While most DNA binding proteins bind to DNA via major groove, TBP interacts with DNA via the minor groove. This binding bends DNA and creates a surface for association with other transcription factors and RNAPII. TFIID is a multi-subunit positioning factor for RNAPII and contains TBP and TBP-associating factors (TAFs) (Burley & Roeder, 1996). TAFs can recognize some other promoter elements such as the Inr, DPE (Hahn, 1998). TFIID is an important protein for establishing the PIC on the core promoter region. Once TFIID binds to the TATA box, TFIIB is recruited downstream of the TATA box and near the TBP that is called the BRE region. The binding of TFIIB creates a surface for recruiting RNAPII. Then, TFIIF joins the complex to stabilize TFIIB (Buratowski & Zhou, 1993). In order to initiate transcription, DNA strands should be opened, but RNAPII does not have this function. TFIIH has ATPase activity to unwind DNA with the XPB subunit. TFIIH also has kinase activity to phosphorylate CTD of RNAPII on serine 5 residues (Svejstrup et al., 1996). Recruitment of RNAPII to the promoter and melting of DNA initiate transcription. At that stage, RNAPII stalls after 30-60 nucleotides are transcribed and does not move further. Short RNA transcripts are degraded. This abortive process is called promoter-proximal pausing. It is suggested that this is required for promoter proofreading and regulatory point for a signal response (Gilmour & Lis, 1986;Adelman & Lis, 2012; Nechaev & Adelman, 2011).

The escape of RNAPII from promoters in eukaryotic cells is a complex process that involves multiple steps. Two crucial steps involve the phosphorylation of the DSIF-NELF complex and the CTD of RNAPII. DSIF (DRB-sensitivity inducing factor) and NELF (negative elongation factor) are separate proteins that contribute to the stabilization of stalled RNAPII (Yamaguchi et al., 1999). During the transition from initiation to active elongation phase, a positive elongation factor-b (p-TEFb) is required, which consists of a cyclin-dependent kinase, CDK9, as its catalytic subunit. The release of RNAPII from the promoter is achieved by the phosphorylation of the DSIF-NELF complex and the serine 2 residues of the heptapeptide repeats of RNAPII CTD by p-TEFb. After the phosphorylation of the DSIF-NELF complex, NELF dissociates from the complex, while DSIF remains associated with RNAPII throughout the gene. DSIF subsequently facilitates the recruitment of the elongation machinery to the transcription site, subsequently assuming the role of a positive elongation factor (Erickson et al., 2018; Gilmour & Lis, 1986; Krebs et al., 2017; Nechaev & Adelman, 2011).

mRNA 5' capping is a modification that occurs shortly after the initiation of transcription and is essential for the stability, translation, and processing of mRNA molecules. The 5' cap is a structure consisting of a 7-methylguanosine residue (m7G) linked to the 5' end of the mRNA via a triphosphate bridge, followed by a series of methylated nucleotides. The cap structure is added co-transcriptionally and is essential for the recruitment of the ribosome to the mRNA and for protecting the mRNA from degradation by exonucleases. The process of mRNA 5' capping involves the activity of several enzymes and proteins, which are recruited to the transcription initiation complex. The first step involves the binding of the RNAPII CTD to the cap-binding

complex (CBC), which consists of the cap-binding protein (CBP) and the small nuclear ribonucleoprotein particle (snRNP). The CBC recognizes the nascent mRNA as it emerges from the RNAPIIand binds to the 5' end of the mRNA. Next, the guanosine triphosphate (GTP) molecule is enzymatically hydrolyzed to form GMP and pyrophosphate, which is then transferred to the 5' end of the mRNA by the enzyme guanylyltransferase (GTase) to form the m7G cap structure (Furuichi, 2015; Ramanathan et al., 2016). The final step involves the addition of 2'-O-methylated nucleotides to the 5' end of the mRNA by the enzyme methyltransferase, which is essential for the stability of the mRNA and for its proper recognition by the ribosome. The mRNA 5' capping is a critical modification that occurs co-transcriptionally and is essential for the stability, translation, and regulation of mRNA molecules (Alberts et al., 2014; Watson et al., 2015).

1.1. TRANSCRIPTION ELONGATION

Elongation is a crucial stage during transcription where the RNA molecule is synthesized and elongated by the addition of nucleotides. The RNA polymerase complex moves along the DNA template strand in a 3' to 5' direction, adding a complementary nucleotide to the 3' end of the RNA strand for every nucleotide in the template. The double-stranded DNA enters the enzyme, which unwinds the DNA to allow for RNA synthesis. As the polymerase separates two DNA base pairs, it forms a hybrid RNA:DNA base pair, and the DNA and RNA molecules exit the transcription bubble separately.

The RNAPII -DSIF-NELF transcription inhibitory complex transforms into an active elongation complex, RNAPII -DSIF-PAF-SPT6, after the dissociation of NELF from the complex (Peterlin & Price, 2006). This transformation allows RNAPII to shift into the active elongation phase. Following this stage, the C-terminal domain (CTD) of RNAPII plays a crucial role in recruiting regulatory proteins to the transcription site, enabling accurate transcription. The phosphorylation of serine 2 residues by PTEFb creates a signal for the recruitment of positive elongation factors, such as splicing factors and histone modifying enzymes, to ongoing transcription(Peterlin & Price, 2006, Watson et al., 2015). TFIIS is another positive elongation factor that aids in proofreading RNAPII by backtracking and cleaving any misincorporated bases (Kim et al., 2007). The processes of elongation, termination of transcription, and splicing of intronic regions are interconnected. The process of mRNA splicing is a critical step in the expression of genes in eukaryotic cells, involving the removal of introns and the joining of exons to form mature mRNA molecules. This process is facilitated by a large

ribonucleoprotein complex called the spliceosome, which recognizes specific sequences at the junctions between introns and exons and cleaves the RNA to remove the non-coding intron segments. The spliceosome then joins the remaining coding exons together to form a continuous coding sequence that can be translated into protein (Wahl et al., 2009). Recent research has uncovered the dynamic and intricate nature of mRNA splicing, including alternative splicing events that can generate multiple isoforms from a single gene. This process can produce different protein variants with distinct functional properties, leading to an increased complexity and diversity of the proteome (Kalsotra & Cooper, 2011).

1.3. TRANSCRIPTION TERMINATION

RNAPII transcribes continuously until it encounters a DNA sequence called a terminator, which signals the termination of transcription. When RNAPII reaches a specific terminator region, three important steps occur subsequently: cleavage of mRNA, polyadenylation of 3' of cleaved mRNA, and degradation of RNA remaining associated with the polymerase. For the protein-coding genes, RNAPII reaches to poly-A signal sequence in DNA. The poly-A signal sequence is recognised by two CTD binging protein complexes, CPSF (cleavage and polyadenylation specificity factor) and CstF (cleavage stimulation factor) immediately after is transcribed into RNA. Poly-adenylation is performed by poly-A polymerase which adds around 200 adenine nucleotides to the 3' end of the transcript after releasing from the RNAPII. In yeast, RNA, which is uncapped and still associated with polymerase, is degraded by a RNase Rat1 (XRN2 human homolog) loaded by Rtt103 that is CTD interacting protein. Reaching terminator by RNAPII, the elongation complex does not dissolve immediately after the synthesized RNA is released from the complex (Cramer, 2004, 2019; Fuda et al., 2009; Haberle & Stark, 2018b; Rosonina et al., 2006).

In recent years, a hybrid model of transcription termination of protein coding gene has emerged, which combines elements of the two previously proposed models: the torpedo model and the allosteric model (Bentley, 2014; Eaton & West, 2020). The termination process in the transcription complex is explained differently by two models. The allosteric model suggests that it occurs due to changes in the complex's shape, whereas the torpedo model proposes that the degradation of the downstream product, which results from processing poly(A) signal (PAS), is a critical factor for the

termination (Connelly & Manley, 1988; Logan et al., 1987). Once the PAS is recognized, the RNA molecule is cleaved and polyadenylated by the respective machinery. The elongation rate of RNAPII slows down due to the dephosphorylation of DSIF by protein phosphatase 1 (PP1). Additionally, there is an increase in threonine-4 phosphorylation at the termination site, which may represent an allosteric change. The termination factors bind to RNAPII, triggering an allosteric conformational change that ultimately leads to transcription termination. The hybrid model suggests that the XRN2 exonuclease is involved in degrading the RNA that remains after cleavage of the RNA at the PAS. The conformational changes occurring at RNAPII enable the termination of transcription by XRN2. (Eaton et al., 2020). Several lines of evidence support the hybrid model of transcription termination (Bentley, 2014). Furthermore, structural studies of the RNAPII have revealed the interaction of the termination factors with the polymerase, providing further support for the hybrid model (Glover-Cutter et al., 2008; Lunde et al., 2010). This integrated model merges components from both the allosteric and torpedo models, providing a more comprehensive understanding of the intricate mechanisms that take place during transcriptional termination.

2. R-LOOPS

During the transcription process, R-loops represent deleterious nucleic acid structures characterized by a three-stranded configuration, comprising two strands of DNA and one strand of RNA, which arises following the hybridization of nascent RNA with the transcribed DNA template.In the process of transcription, double-stranded DNA is physically separated, and nascent RNA transiently anneals 8-10 base pairs with template DNA in RNA polymerase active site. The formation of the RNA-DNA hybrid leads to the displacement of the non-transcribed DNA strand (Chédin, 2016) (**Figure 2**).

It is of significant importance to note that R-loops and RNA:DNA hybrids are distinct structures. R-loops arise from the displacement of one of the DNA strands by an RNA molecule, resulting in a three-stranded structure. In contrast, RNA:DNA hybrids form a double-stranded structure that consists of one strand of RNA and one strand of DNA molecules. When the RNA molecule dissociates from the DNA strand, the displaced DNA strand can form a single-stranded structure, which is more susceptible to damage and mutations. If not properly resolved, these ssDNA regions can be accessible to the mutagens and DNA breaks, leading to DNA damage and replication stress.

RNA processing throughout transcription has an important role in resolving prolonged RNA:DNA hybrids. The formation of R-loops occurs with the trade-back model which supposes newly synthesized RNA invades double-stranded DNA immediately after it is produced by RNAPII (Aguilera & García-Muse, 2012). Negative DNA supercoiling, presence of guanine nucleotide on the non-template strand and formation of guanine-quadruplex secondary structures enhances the formation of R-loops. Also, pausing of

the RNA polymerase, lncRNA transcription, modifications at RNA or DNA, and malfunction of mature RNA synthesis increase the accumulation of R-loops (Chédin, 2016; A. Kim & Wang, 2021). Negative supercoiling causes melting double stranded DNA and facilitates invasion probability of nascent RNA to DNA. As RNA:DNA duplex is thermodynamically more stable than DNA-DNA duplex, G-rich regions enhance the initiation of R-loops (**Figure 2**) (De Magis et al., 2019; Roberts & Crothers, 1992).

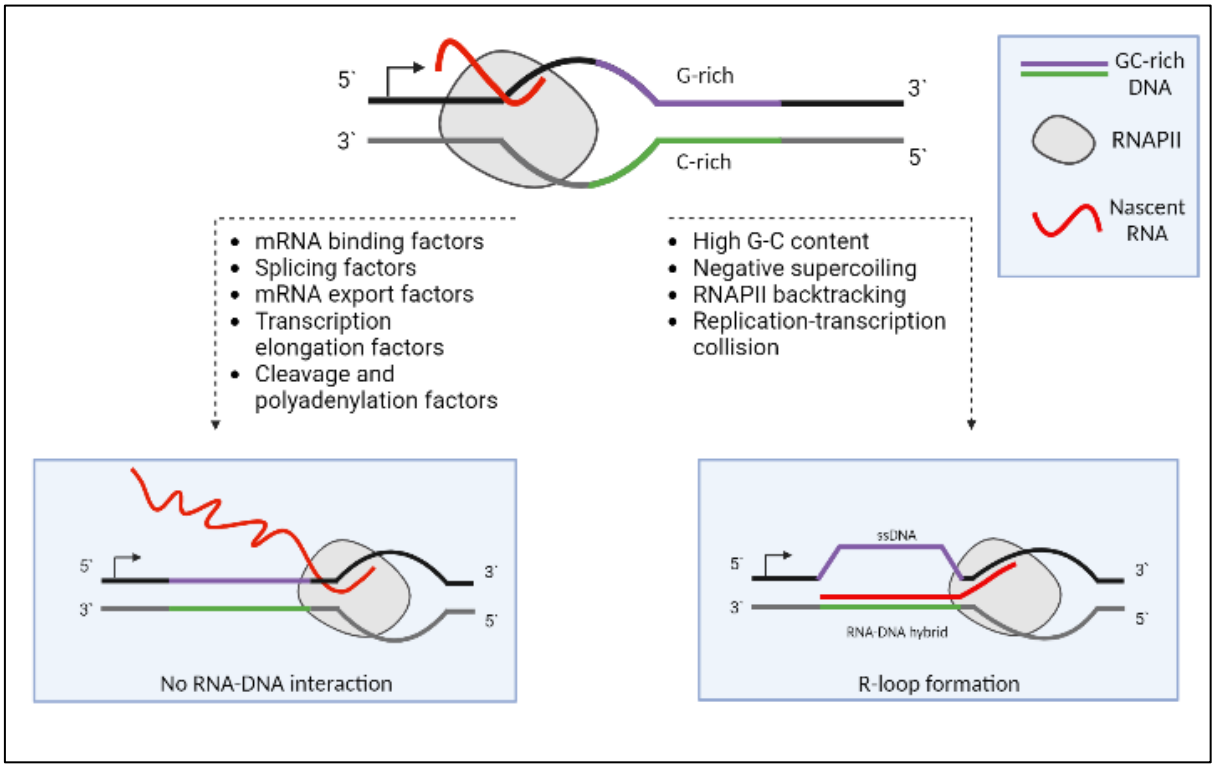


Figure 2: Co-transcriptional formation of R-loops.

Overview of the formation of R-loops during transcription. In most cases, R-loop does not form while transcription is proceeding thanks to the R-loop resolving factor (left). The nascent RNA may reanneal in certain cases because of GC strand asymmetries in DNA regions, negative supercoiling, RNAPII backtracking and replication-transcription collision (right). Factors that prevent or facilitate the formation of R-loops are listed.

For a long time, it was thought that generation of R-loops is only a byproducts of transcription. However, it was discovered that R-loops have a crucial role in transcription termination, DNA repair and gene regulation (Hegazy et al., 2020). Most of human promoters contain CpG islands which can be methylated to silence transcription, so staying unmethylated of these CpG promoter regions has important role in initiation of transcription for these genes. It has been suggested that formation of R-loop at promoter regions can protect this region from methylation since DNA methyl-transferase enzymes have a low affinity to bind DNA: RNA hybrid regions It also prevents or enhances the recruitment of chromatin remodelling complexes which activate or inactivate transcription (Ginno et al., 2012). Most of the studies showed that R-loop formation also increases at 3'of the genes. RNAPII starts to oscillation between elongation and backtracking when it reaches to termination site of the gene. R-Loop formation is occurred at downstream of the polyadenylation site, but RNAPII still elongates nascent RNA. Due to the generated torsional stress, R-loops then help in transcription termination with XRN2 dependent termination. It was also suggested that R-loop formation at gene termination regions prevents starting to transcribe adjacent gene (Niehrs & Luke, 2020; Skourti-Stathaki et al., 2011; Sollier & Cimprich, 2015). It was observed that RNA polymerase accumulation is correlated with R-loop forming regions. There are two possible reasons causing that accumulation. R-loops can be reason for stalling of RNA polymerase or stalled RNA polymerase leads formation of R-loops (Chédin, 2016; Sanz et al., 2016).

R-loops naturally occurs during the transcription and are removed if necessary. Any deficiency at detection or resolving pathways of R-loops can be dangerous (Chédin, 2016; Wahba et al., 2011). Some factors can resolve or prevent R-loop formation such as RNaseH, RNA binding proteins, splicing factors, topoisomerases, transcription elongation factors and cleavage and polyadenylation factors (Kumar et al., 2022; Mischo et al., 2011; Mosler et al., 2021; Skourti-Stathaki & Proudfoot, 2014; Stein & Hausen, 1969). Genomic or epigenomic changes are associated with oncogene activation or tumour repressor gene inactivation. Excessive accumulation of R-loops can lead to genomic instability in different ways. One of the DNA strands, which is unhybridized with RNA, remains as single-stranded DNA that is more susceptible to targeting by mutagens or DNA modifying enzymes than double-stranded DNA (Aguilera & García-Muse, 2012). While transcription is processing, it can come across with a replication fork. This causes collisions between transcription and replication machineries co-directionally or head-on. R-loop is an impediment to the progression of the replication fork. If it does not remove on the way of replication fork properly, this can cause the DNA double-strand breaks (Sollier and Cimprich, 2015).

Genome-wide R-loops profiling techniques are based on bisulfide foot-printing, RNaseH dependent approaches, and S9.6 antibody-based immunoprecipitation. Bisulfide foot-printing relies on the deamination of the unmethylated cytosine residues in ssDNA so that deamination by bisulfide treatment can be used as a marker of R-loop sites. RNaseH enzyme specifically recognizes and solves R-loop structures in the cell. Catalytically inactive RNaseH enzyme can recognize R-loops but not resolve hybrid structures. Using tagged catalytically RNaseH, immunoprecipitation is performed to detect RNA: DNA hybrids (R-ChIP). MapR technique is based on

catalytically inactive RNaseH coupled with micrococcal nuclease digestion. Some of these methods use the S9.6 antibody which was produced in mice using synthetic RNA:DNA antigen. S9,6 antibody has specific affinity to detect the DNA: RNA hybrids and can recognize the 6 bp region of the R-loop structure. Different modified techniques for RNA: DNA immunoprecipitation using S9,6 antibody was developed to get robust results. the most recent one is q-DRIP, which is ground on S9,6 antibody coupling with using synthetic R-loops as spike-in. In this technique, synthetic R-loops are used as internal controls to quantify R-loops in high resolutio (Crossley et al., 2020). It has been demonstrated that the S9.6 antibody exhibits a relatively lower binding affinity to AU-rich RNA-RNA hybrids, which has led to some considerations regarding its specificity. Nevertheless, it demonstrates a distinct kinetic preference for binding DNA-RNA hybrids (Phillips et al., 2013).

3. RNA POLYMERASE II

The enzyme RNA polymerase II (RNAPII), which synthesizes mRNAs as well as most of ncRNAs, is composed of 12 polypeptides. The largest subunit is Rpb1 which carries the catalytic core of the enzyme. The C-terminal domain (CTD) of Rpb1 is a unique feature that set it apart from all other polymerases. The absence of one-half or more than one-half of the repeats cause the growth defects in *Saccharomyces cerevisiae* or mammalian cells. CTD contains different numbers of tandem heptapeptide repeats (Corden et al., 1985). The number of these tandem repeats varies from organism to organism. There are 26 repeats in *Saccharomyces cerevisiae* CTD, 45 repeats in *Drosophila melanogaster* and 52 repeats in *Homo sapiens*. It seems that there is a correlation between the number of repeats and the complexity of the organism (Liu et al., 2010). Truncation experiments revealed that shortening the CTD within a limited range is compatible with cell viability and growth. The CTD displays a strong periodicity of heptad-repeats, this periodicity is dispensable and not essential for growth. Second, the CTD can be divided into functional units and does not act as an entire structure. Third, the presence and spacing of the phosphor-serin motifs is the most conserved feature of a minimal functional unit. Finally, the model that the CTD acts as a platform for the recruitment and dissociation of cellular factors is compatible with the presence of functional units (Buratowski, 2003; Eick & Geyer, 2013; Hsin & Manley, 2012).

All repeats of the CTD undergo a dynamic cycle of post-translational modifications. Similarly, modifications of histones, these changes are performed by enzymes known as "writers" and "erasers" (Pineda et al., 2015). Readers, which recognize these sequence marks, interact with the modification patterns. A transient "code" emerges

as a result of the interaction between modifying enzymes and recognizing factors. These conserved sequences can be modified at different sites throughout transcription. These modifications serve as docking site for other transcription factors to regulate transcription at different stages. Different post-transcriptional modifications at CTD such as phosphorylation, acetylation, methylation, cis-trans isomerisation and ubiquitination create a group of signals for the proceeding of transcription (Erick & Geyer, 2013). These specific modifications are called CTD code (Egloff & Murphy, 2008). The post-translational modification of CTD also alters its intrinsic properties. When the sequence is phosphorylated, negative charges are introduced, which may induce an extended conformation. In the same way, proline cis*trans isomerization creates conformational changes on the CTD, and O-glycosylation can affect its properties by increasing its size and mass (Egloff et al., 2012; Eick & Geyer, 2013; Garrido-Godino et al., 2022; Harlen & Churchman, 2017; Mosley et al., 2009).

From the initiation to the termination of transcription, CTD phosphorylation patterns undergo significant changes. Different sets of antibodies specific for CTD phosphorylation have been used to analyse CTD phosphorylation genome-wide in *S. cerevisiae* as well as in mammalian cells and plant cells. (Egloff et al., 2012; Hsin & Manley, 2012). The phosphorylation pattern of CTD across the protein-coding genes is well studied by using chromatin immunoprecipitation. All CTD phospho-marks upstream of the TSS have very low signals according to the ChIP analysis. There was a strong increase in the Ser5-P and Ser7-P signals at the TSS. The signal for Ser7-P remains high throughout the transcription cycle, whereas the signal for Ser5-P steadily decreases toward the poly-A (pA) site. Tyr1-P, Ser2-P and Thr4-P signals are low at TSS, but downstream signals are increased (Eick & Geyer, 2013). Ser2-P signaling is

highest at poly-A and its downstream, consistent with the recruitment of 3' RNA processing factors generated by Ser2 phosphorylation of CTD (Suh et al., 2016). High levels of Tyr1-P in the body of genes promote the binding of elongation factors and prevent binding of termination factors to the CTD. In addition, factors binding to Thr4-P have yet to be characterized. During termination, all CTD phospho-marks are removed by phosphatases (Buratowski, 2009) (**Figure 3**).

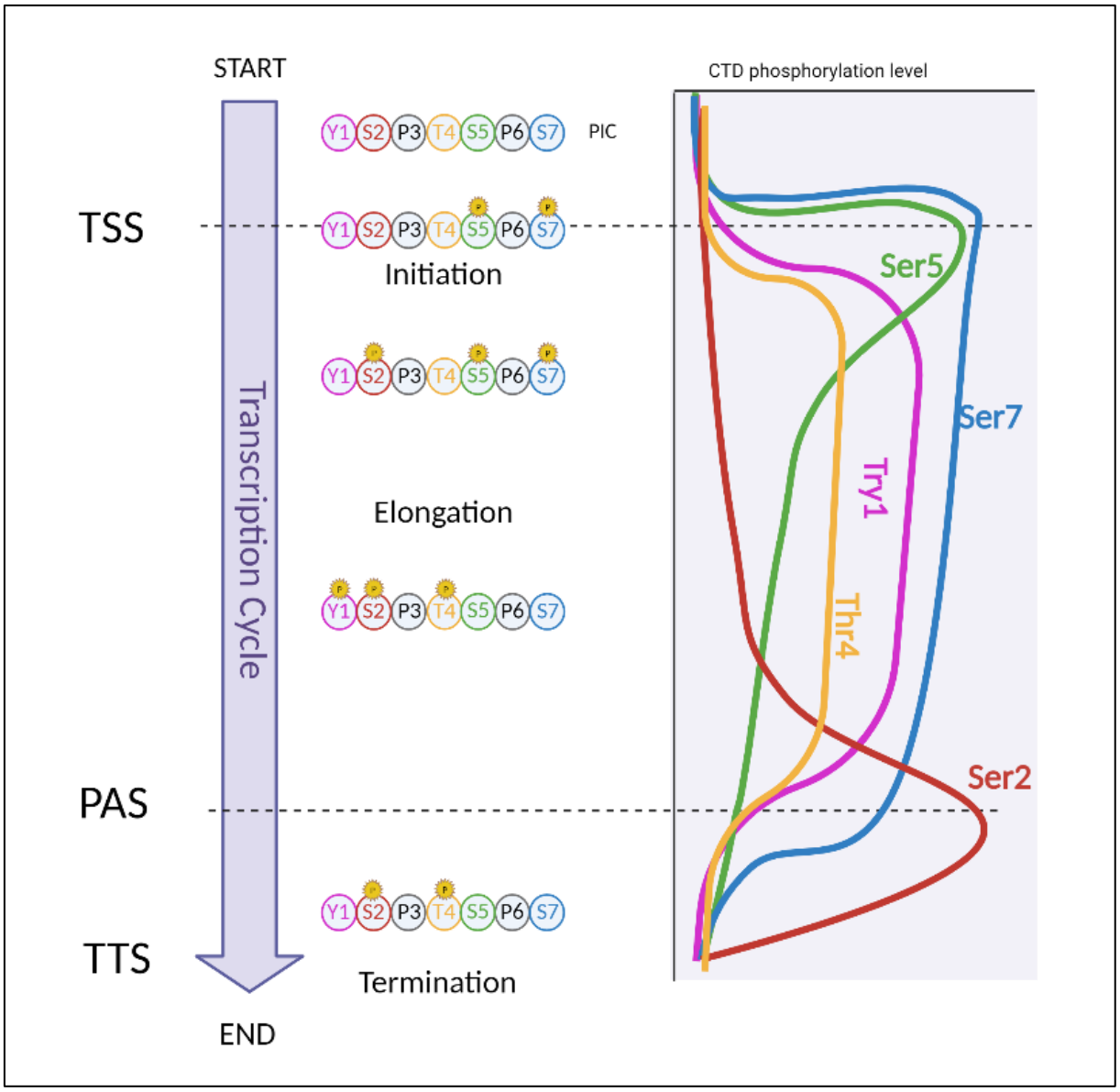


Figure 3: CTD phosphorylation throughout transcription cycle in protein coding genes.

Illustration depicting the dynamic process of CTD phosphorylation at various stages throughout the transcription cycle in protein-coding genes. The figure provides a visual representation of the changes in CTD phosphorylation patterns as RNA polymerase progresses through transcription initiation, elongation, and termination. The CTD phosphorylation pattern is plotted genome-wide. A peak in Ser5 and Ser7 signals is observed at the transcription start site (TSS) of genes, whereas Tyr1-P, Ser2, and Thr4-P signals are present near the 3′ end of the gene (PAS: polyadenylation site; TTS: transcription termination site).

3.1 CELLULAR PROCESSES DRIVEN BY CTD

The role of CTD in the regulation of all steps of transcription makes it a crucial component of the dynamic and highly regulated transcription process. The assembly of the preinitiation complex (PIC) is the first step in transcription, which requires general transcription factors and the mediator complex. The hypo-phosphorylated CTD is necessary to establish the pre-initiation complex, and it cooperates with the mediator complex to activate transcription through the distal promoter elements of genes (Eick & Geyer, 2013; Hsin & Manley, 2012; Petrenko et al., 2016). A hypo-phosphorylated CTD facilitates assembly of the initiation complex, but afterwards it functions as a strong repressor by preventing RNAPII from entering the elongation phase (Usheva et al., 1992). As a result of its phosphorylation, the CTD can be dissociated from the mediator complex, resulting in the abolishment of its negative function (Petrenko et al., 2016). The transcription start site (TSS) or downstream thereof, or both, has extremely high RNAPII density in higher eukaryotes. This suggests that RNAPII pauses at the promoter proximal pausing site and requires activating signals in order to continue transcription (Eick & Geyer, 2013; Hsin & Manley, 2012).

3.2 CTD INTERACTORS

In order to give a proper response to CTD code, some writer, reader, and eraser proteins have roles throughout transcription. Writers modify CTD by adding chemical groups, erasers remove these chemicals added marks, readers bind specifically to these marks to recognize the modification pattern to interpret those modifications. The enzymes responsible for adding phosphate groups to other molecules, known as kinases, are categorized separately for mammals, budding yeast, and fission yeast. They can be considered as the writer enzymes. (Eick & Geyer, 2013). Readers are divided into general transcription factors, histone and chromatin modifying factors, RNA processing factors, and interactions retro-acted to the CTD. One of the general transcription factors, TFIIH, has a serine kinase function thanks to the cdk7 subunit to phosphorylate Ser5 residue. The P-TEFb (CDK9) kinase phosphorylates both negative elongation factor (DSIF-NELF complex) to switch to active elongation state, and Ser2 to recruit several elongation and chromatin-modifying factors. In addition to serine residues phosphorylation, Tyr1 and Thr4 residues are also phosphorylated by kinases (cAlb1 and Plk3 respectively) to regulate the termination (Eick & Geyer, 2013).

These phosphate residues are removed by phosphatases which are specific to each modification; Ssu72, RPAP2 for Ser5; Ssu72 for both Ser5 and Ser7; SCP1–3 for Ser2 dephosphorylation. As CTD reader, some histone-modifying, mRNA capping, and mRNA splicing enzymes recognize these varieties of CTD codes to progress transcription, accurately (Eick & Geyer, 2013; Hsin & Manley, 2012). One class of CTD interacting proteins is RPRD proteins and their role in transcription has not known.

Chapter 2: RPRD PROTEINS

Many proteins regulate transcription by binding to the CTD of RNAPII. It has been predicted that there are 40 proteins containing the CID domain from yeast to humans (Doerks et al., 2002). The Regulation of Nuclear Pre-mRNA Domain-containing (RPRD) proteins interact with the CTD through their CTD-interaction domains (CID) to participate in the of newly synthesized RNA (Ni et al., 2011). The RPRD protein family, recently discovered as interactors with phosphorylated CTD of RNAPII and RNAPII -associated proteins, consists of three proteins in humans: RPRD1A, RPRD1B, and RPRD2 (Ni et al., 2011). While RPRD1A and RPRD1B have similar protein sequences and domains, RPRD2 is longer than these two proteins and contains serine-rich and proline-rich regions in addition to the CID domain **(Figure 4).** Both RPRD1A and RPRD1B have two similar CID and CC domains. The CID is responsible for interacting with the CTD, while the CC domain has a role in forming a homodimer or heterodimer with other RPRD proteins. RPRD2 shares similar CID and CC domains with RPRD1A and RPRD1B but has a much longer amino-acid sequence and extra serine/proline-rich regions (Li et al., 2021; Ni et al., 2014; Winczura et al., 2021).

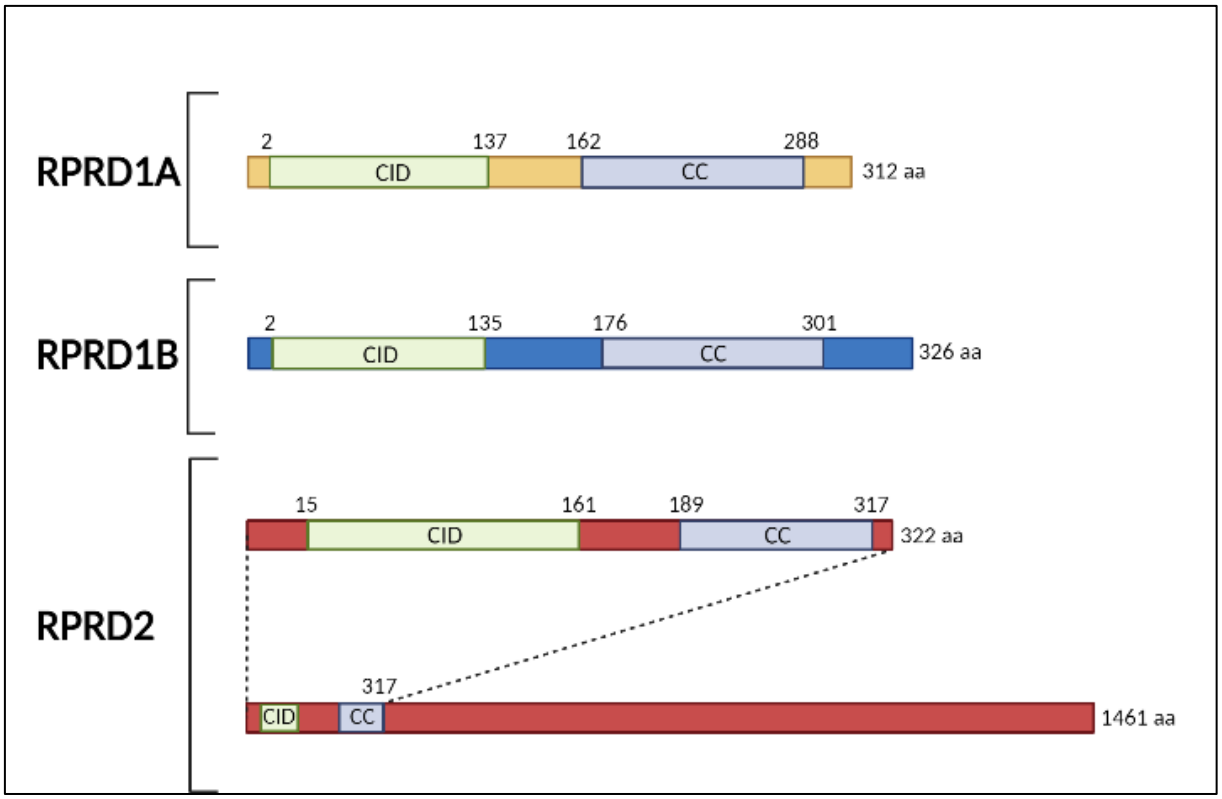


Figure 4: Structure of RPRD proteins.

Demonstration of the protein domains of RPRD protein family consisting of RPRD1A (312 aa), RPRD1B (326 aa), and RPRD2 (1461 aa). All of the three members of this family have CID (C-terminal interaction domain, light green) and CC (coiled-coil domain, light blue). In addition to these domains RPRD2 also has long serine-proline rich region.

The alignment of the CID domains from all three RPRD proteins in a binary combination elucidates the identity and similarity percentages shared among these protein sequences. Utilizing the ClustalW protein alignment tool, I conducted sequence alignments for the CID sequences from three distinct proteins, employing binary combinations of these sequences. Through the utilization of this tool, it becomes possible to visually discern proteins that exhibit both identical and analogous chemical properties. In practical application, asterisks (*) denote positions characterized by the

highest degree of conservation, colons (:) signify positions demonstrating intermediate levels of conservation, while dots (.) indicate positions exhibiting comparatively lower levels of conservation (**Figure 5).**

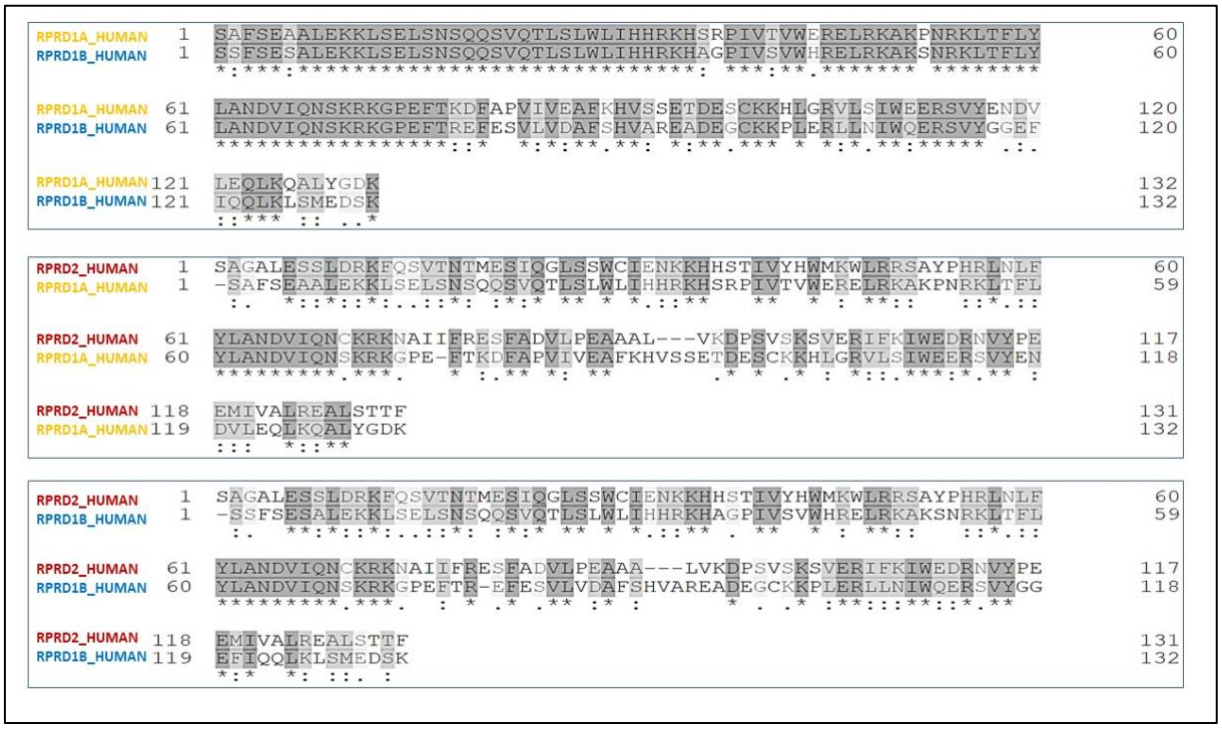


Figure 5: Alingnment of the CID domain of RPRD proteins in binary combinations.

The symbols asterisk (*), colon (:), and dot (.) are used to denote different levels of similarity between amino acid residues in aligned sequences. An asterisk (*) signifies identical residues at a specific position, indicating a high degree of conservation. A colon (:) indicates strong similarity between residues, suggesting shared chemical properties, while a dot (.) represents weaker similarity.

Similarity refers to the extent of analogous chemical attributes shared among aligned amino acids, whereas identity pertains to the measure of completely identical residues in the comparison of two protein sequences. Following the binary alignment of protein sequences, identity and similarity scores were computed for the comparison of CID

sequences derived from RPRD proteins. Subsequently, these scores were employed to generate a heatmap, illustrating the proportion of shared residues that are either similar or identical within the compared sequences (**Figure 6**). The CID domains of RPRD1A and RPRD1B exhibit a remarkable degree of sequence conservation, displaying 92.4% identity and 73.5% similarity. These two proteins, in terms of both identity and similarity, manifest the highest percentage of conservation among the compared sequences. Notably, the levels of identity and similarity observed between the CID domain of RPRD2 and those of RPRD1A and RPRD1B closely approximate one another (**Figure 6**).

Figure 6: CID identity and similarity percentages of RPRD proteins.
Heatmap representation of CID identity and similarity percentages of three RPRD proteins. The red colour indicates lowest level of identity or similarity, yellow is midpoint (50 %), green is the highest percentage of identity and similarity.

1. RPRD Proteins in transcription

It was demonstrated that RPRD1A and RPRD1B exhibit physical interactions with three RPRD proteins, RNAPII, and some RNAPII-associated proteins, RNA polymerase I, and III subunits. RPRD1A and RPRD1B were co-purified with phosphorylated Ser2, Ser5, and Ser7 isoforms of CTD, with unmodified CTD showing low levels of RPRD1B recruitment but not RPRD1A (Ni et al., 2011)

As mentioned earlier, the CTD of human RNAPII consists of 52 repeats of heptapeptide sequences. Among these repeats, the 21 proximal ones, close to the core region, are predominantly conserved, while the 31 distal repeats, located further away from the enzyme's core, can contain residues of asparagine, threonine, and lysine at position serine-7. These residues can undergo diverse post-translational modifications, to regulate transcription and RNA processing (Ali et al., 2019; Eick & Geyer, 2013; Harlen & Churchman, 2017). In addition, RPRD1A and RPRD1B respond to acetylated lysine residues at CTD via their CID domains, with lysine residues at the 7 positions of CTD repeats playing a crucial role in the productive elongation of polymerase. The acetylation of lysine is conserved in higher eukaryotes and enriched at +500 bp downstream of the TSS, indicating its significance in the transition to the elongation phase. Knockdown of RPRD1B does not affect the unmodified polymerase II occupancy at the TSS site; however, it leads to the enhancement of occupancy of acetylated lysin-7 at TSS site and global accumulation of S5P CTD due to a failure in recruiting the phosphatase (Ali et al., 2019).

It was demontrated that RPAP2, though to be a phosphatase responsible for dephosphorylating Ser5 of CTD, is one of the proteins that interacts with RPRD1A and RPRD1B (Ni et al., 2011). Depletion of RPAP2 results in a significant increase in Ser5P, but not in Ser2P and Ser7P (Egloff et al., 2012). It was also demonstrated that RPRD1B recruits RPAP2 to CTD (Ni et al., 2014). Recent studies indicate that RPAP2 lacks phosphatase activity, contrary to the previous assumption. Instead, RPAP2's negative impact on transcription is due to its capability to hinder the association between RNAPII and TFIIF (Wang et al., 2022).

Therefore, RPRD proteins also serve as intermediate players for the K7 deacetylase enzyme and phosphate-eraser proteins in (Ali et al., 2019). *In vitro* experiments indicate that RPRD1A and RPRD1B bind to unphosphorylated and Ser7 phospho-form of CTD, but exhibit ~3-6 times stronger affinity for binding Ser2 phosphorylation (Mei et al., 2014). These findings collectively suggest that RPRD proteins may play a significant role in various aspects of transcription, including initiation, elongation, and termination, although further research is needed to comprehensively elucidate their precise role.

Furthermore, RPRD1B may play a role in transcription termination by mediating XRN2-RNAPII interaction, which is essential for transcription termination. In the absence of RPRD1B, RNAPII accumulates at the 3’ end of the genes (Ali et al., 2019).

HIV-1 is an RNA virus that necessitates reverse transcription by viral reverse transcriptase (RT) to produce viral hybrid RNA: DNA intermediates. Studies have shown that after infection, RPRD2, also known as RNA-associated early-stage antiviral factor (REAF), associates with viral nucleic acids. Proteasomal degradation of RPRD2

is required for initiating reverse transcription (Marno et al., 2014, 2017). Vpr, one of four non-structural accessory genes of HIV-1, is essential for efficient virus replication. Nuclear localization of Vpr is necessary for RPRD2 degradation, as it interacts with DCAF1 E3 ubiquitin ligase (Gibbons et al., 2020). Thus, it has been suggested that RPRD2 plays a crucial role in the anti-viral surveillance system, which inhibits early and late viral transcript levels by degrading incoming retroviruses.

It has been demonstrated that loss of RPRD1B results in the formation of R-loops, as shown by immunofluorescence experiments. Excessive R-loop formation can lead to the accumulation of mutations on the displaced DNA strand, ultimately leading to genomic instability. RPRD1B knockdown activates the DNA-damage response, resulting in increased R-loop formation and double-strand breaks (DSBs). A study has shown that the CID domain of RPRD1B is required for stabilizing Artemis, a crucial protein for DNA DSB repair. Thus, the absence of RPRD1B leads to a deficiency in DNA damage repair (Morales et al., 2014). However, the exact mechanism between RPRD1B and R-loops is not yet fully understood, and no clear relationship between R-loop and other RPRD proteins has been reported in the literature.

Overall, all these three proteins, RPRD1B, RPRD1A and RPRD2, may have important roles in transcription initiation, elongation, and termination, however their roles have not been elucidated so far.

2. RPRD Proteins Interactome

The RPRD proteins possess a CID domain that enables them to bind directly to the CTD domain of RNAPII. However, no enzymatic activity has been demonstrated for all RPRD proteins. Studies have demonstrated that RPRD proteins interact with a diverse range of proteins in addition to RNAPII subunits. Among these proteins, RPRD1B is a studied protein in terms of protein-protein interactions, high-throupout affinity purification followed by mass spectrometry (AP-MS) revealing its association with 41 different proteins (Hein et al., 2015). This suggests that RPRD1B is likely to be involved in multiple distinct protein complexes within the cell. Recent research has indicated that RPRD1B might functions in RNA termination through its interaction with XRN2 and in DNA double-strand break repair by interacting with Ku70 and Ku86, which are required for the non-homologous end joining (Morales et al., 2014).

The experimentally proven results shown in the STRING database were obtained using database option filters (**Figure 7**). All three RPRD proteins are interconnected and interact with several proteins that are important during transcription, such as RECQL5, RPAP2, LEO1, CTDP1, and POLR2M.

RECQL5, a helicase family member associated with RNAPII, plays a crucial role in transcription, as its depletion causes significant increases in the elongation rate, RNAPII stalling, pausing, arresting, and/or backtracking. Moreover, RECQL5 deficiency results in chromosomal instability due to increased numbers of sister chromatid exchange events and DNA damage from double-strand breaks (Hamadeh & Lansdorp, 2020; Saponaro et al., 2014).

All three RPRD proteins interact with Gdown1 (also known as GRINL1A or POLR2M), a novel subunit of RNAPII and a responsive element for mediator to RNAPII. Gdown1 has an inhibitory effect on elongation by competing with the PIC stabilizer protein TFIIF, blocking TFIIF's interaction with the PIC, and affecting the DSIF/NELF complex's function. Additionally, Gdown1 inhibits the termination role of TTF2 (Transcription termination factor 2)**,** and depletion of Gdown1 leads to increased RNAPII numbers on the gene body (Cheng et al., 2012; Wu et al., 2012). Recent studies have demonstrated that Gdown1 is primarily located in the cytoplasm and translocates to the nucleus under stress conditions to modulate transcriptional inhibition and adapt to the stressful environment (Zhu et al., 2022). These findings suggest that RPRD proteins may have a role in recruiting the negative elongation factor Gdown1 to the transcription site in stress conditions.

RPAP2, an analog of yeast Rtr1, was initially identified as an RNAPII transporter protein from the cytoplasm to the nucleus. RPAP2 also has a serine 5 dephosphorylation function for CDT, and depletion of RPAP2 causes an accumulation of CTD Ser5. This is correlated with a decrease in transcription and defects in transcription termination. Recently, it was suggested that RPAP2 inhibits the formation of the RNAPII-TFIIF complex, and depletion of RPAP2 leads to an elevated number of PIC formations on promoters by increasing the recruitment of TFIIF on PIC (Chen et al., 2021). However, recent findings reveal that RPAP2 is enzymatically inactive as a phosphatase. Instead, the negative effect of RPAP2 on transcription is attributed to its ability to prevent the interaction between RNAPII and TFIIF (Wang et al., 2022).

The PAF (or polymerase-associated factor) complex is a protein complex that plays a crucial role in regulating transcription elongation by RNAPII. It consists of several subunits, including PAF1, CDC73, LEO1, CTR9, and RTF1. The PAF complex interacts with the C-terminal domain (CTD) of RNAPII and promotes the recruitment of factors involved in histone modification, chromatin remodeling, and RNA processing (Jaehning, 2010)). As a member PAF complex, LEO1 exhibits the interaction with RPRD proteins according to STRING database. PAF1 is a conserved protein that is associated with RNAPII and acts as a docking site for histone modifier enzymes and RNA processing factors. Loss of PAF1 leads to an increase in RNA polymerase released from the promoter, as well as an elevation in Ser2 phosphorylation and newly synthesized RNA (Chen et al., 2015).

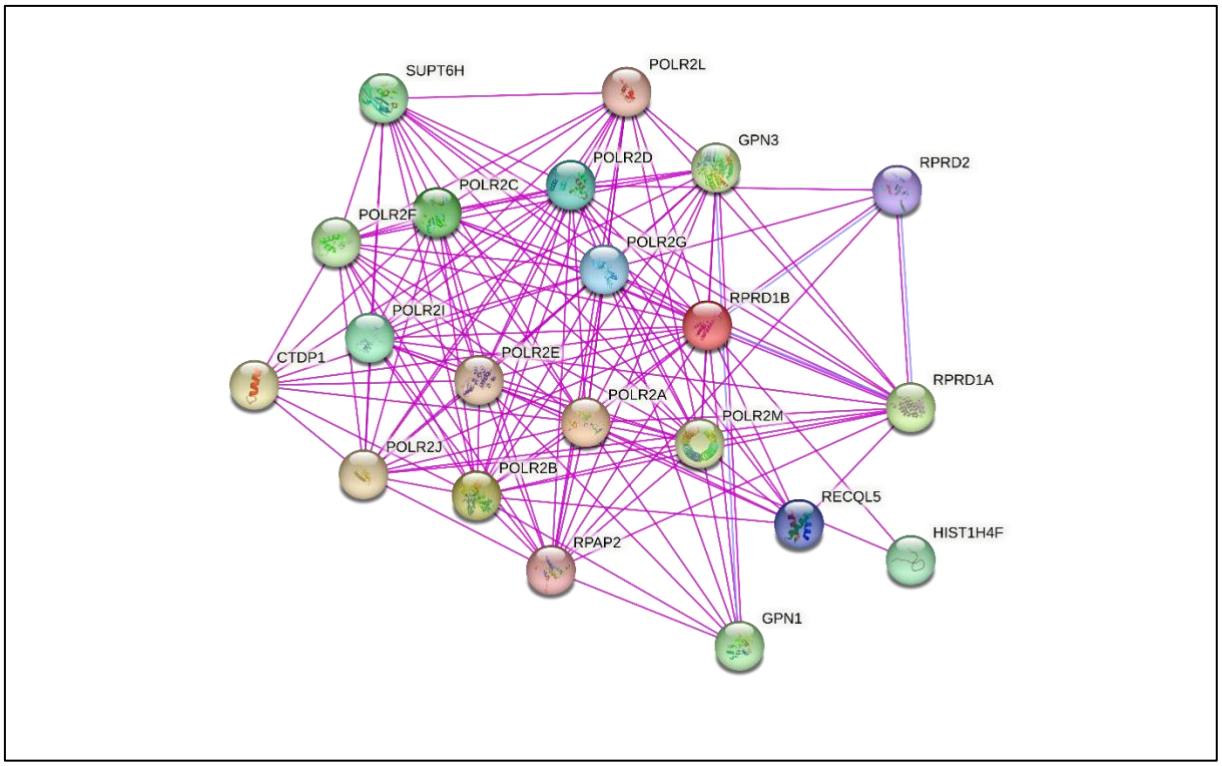


Figure 7: RPRD interactome created by using STRING database.

The RPRD interactome, constructed through data obtained from the STRING database, illustrates the network of protein-protein interactions involving RPRD proteins(pink lines). Purple lines between protein nodes indicate protein homology. 3D structure of each proteins is represented in each node.

The venn diagram generated based on data from the STRING database (**Figure 8**). I updated the data settings to only show experimentally detected protein-protein interactions with medium confidence (0.4). I individually searched each RPRD protein interactome and identified common proteins at the intersection region of the venn diagram. RPRD1B and RPRD1A demonstrate a higher degree of overlap in their protein-protein interaction profiles when compared with the two other intersecting datasets (**Figure 8**).

POLR2D and POLR2G are homolog proteins of the Rpb4 and Rpb7 in yeast, respectively. The yeast RNAPII core enzyme consists of ten subunits, including the Rpb4/Rpb7 heterodimer (Choder, 2004, 2011). This heterodimer remains associated with the RNAPII transcripts throughout their lifespan and plays a crucial role in mRNA imprinting to regulate mRNA export, translation, and decay. The association of this heterodimer is facilitated by Rtr1, whose human homolog is RPAP2, a CTD phosphatase (Garrido-Godino et al., 2022). According to the STRING database data, all three RPRD proteins interact with POL2D, POLR2G, and RPAP2.

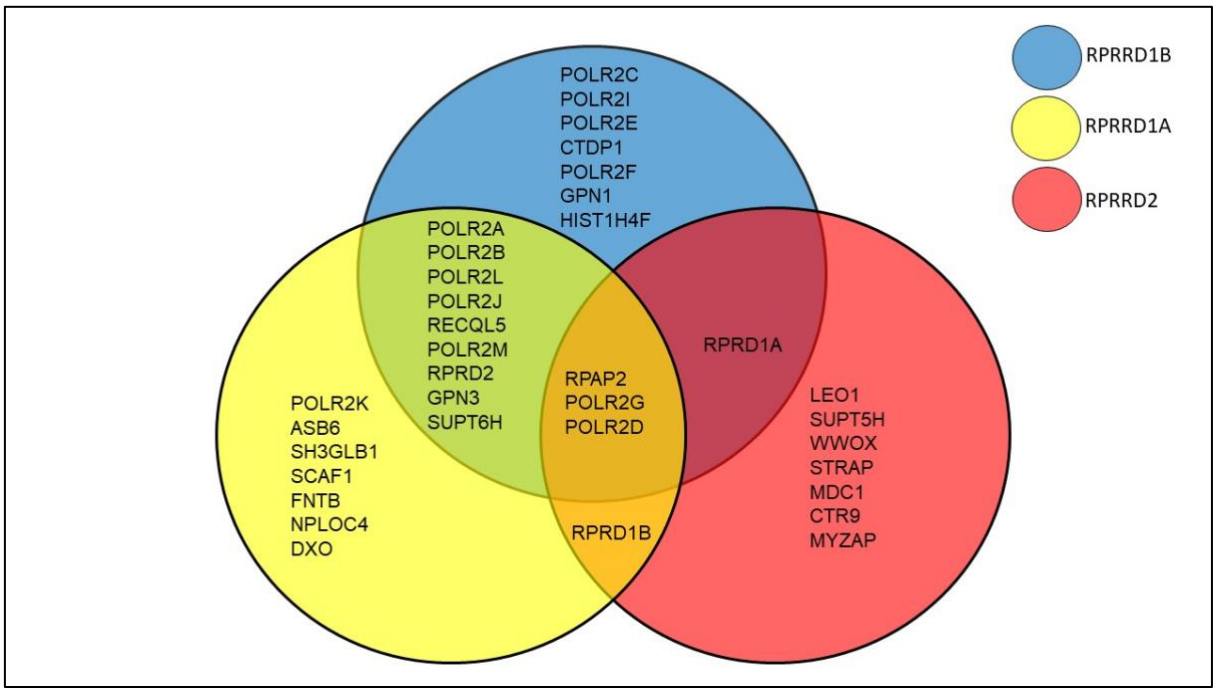


Figure 8: Illustration of RPRD interacting proteins by using venn diagram.
This graph was generated based on the data at STRING database. The each differently colored circle represents individual RPRD protein interactome cluster (blue circle RPRD1B: yellow circle RPRD1A; red circle RPRD2). Protein names in each colored circle represent the protein which are interacting with relevant RPRDs. Protein names at intersection region of two or three protein show proteins that have an interaction capability both or all.

3. Cell Cycle and RPRD Proteins

Eukaryotic cells follow a life cycle that encompasses the stages of cellular birth, growth, division, or programmed cell death. This sequence of events, from the cell's formation (birth) to its eventual division, is collectively referred to as the cell cycle. This intricate process is rigorously regulated to maintain cellular health and ensure the faithful transmission of genetic material to progeny cells. The cell cycle comprises two main steps: interphase and M-phase (mitotic phase). Interphase consists of three phases: G1, S, and G2 phases. While some cells in an organism do not divide, others divide only occasionally to replace damaged or dead cells. Cells that do not divide further enter an inactive phase of the cell cycle called the G0 stage (quiescent phase), where they remain metabolically active but do not proliferate unless stimulated to divide. Interphase serves as a preparation step for mitosis (M phase), with G1 and G2 phases acting as gap phases between other steps. S phase, on the other hand, is a DNA synthesis step for mitosis, where genomic material is duplicated (**Figure 9**) (Barnum & O'Connell, 2014; Vermeulen et al., 2003).

The cell cycle is regulated by a complex set of proteins including cyclins, cyclin-dependent serine/threonine protein kinases (CDKs), and cyclin-dependent kinase inhibitors (CKIs). CDKs drive cell cycle progression, while cyclins manage the transition between phases. During adverse conditions, CKIs inhibit the activity of CDK/cyclin complexes (Lim & Kaldis, 2013; Rhind & Russell, 2012; Vermeulen et al., 2003). The cell cycle contains various internal quality control points that regulate the transition between its phases. These checkpoints serve to identify and repair any damages or faults that could cause significant harm to the daughter cells. Checkpoints

can also modulate cell cycle progression in response to internal or external signals. The G1, intra-S, G2/M, and mitotic spindle checkpoints are crucial for ensuring accurate cell division (**Figure 9**). At the G1 checkpoint, cells monitor various features, such as cell size, DNA integrity, and the presence of growth factors and nutrients, to determine whether they should divide or not. The intra-S checkpoint is responsible for detecting errors in DNA replication, while the G2/M checkpoint monitors the accuracy and completion of DNA replication. If everything is in order, cells enter the mitosis phase, where the attachment of chromosomes to the mitotic spindles is detected by the anaphase promoting complex (APC) to ensure an equal distribution of chromosomes to the daughter cells (Barnum & O'Connell, 2014).

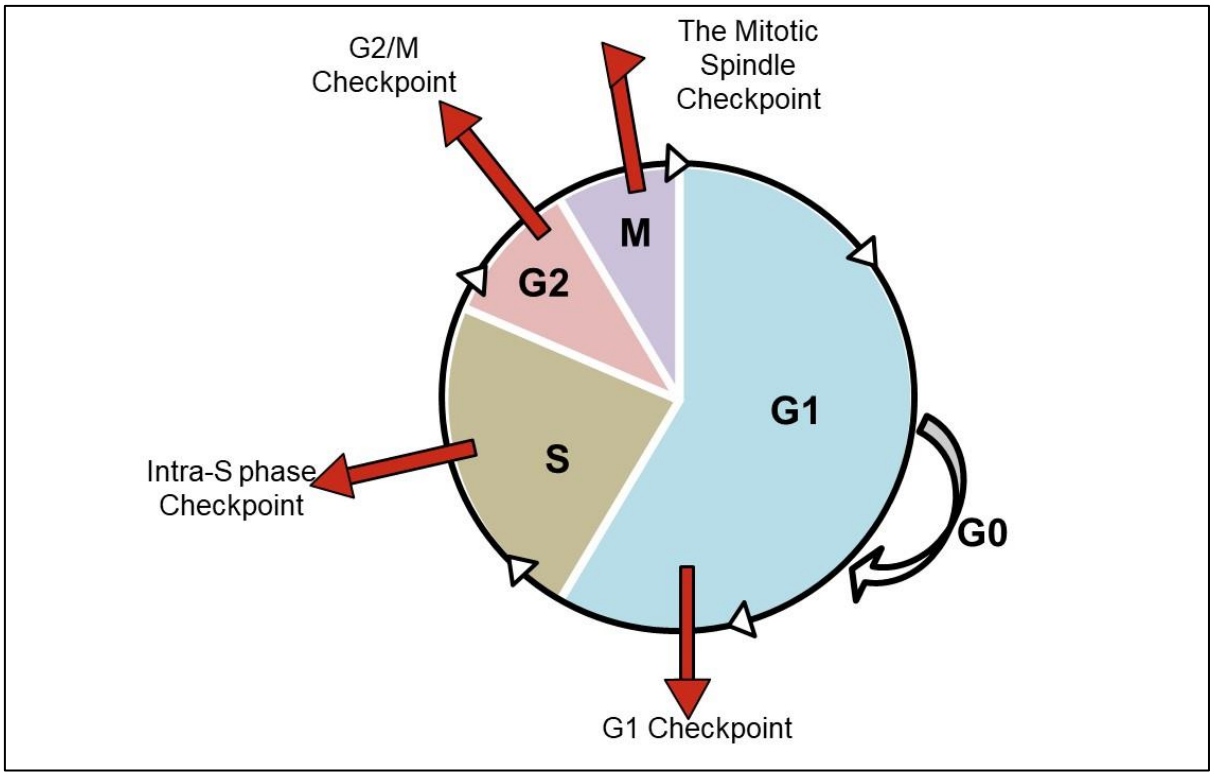


Figure 9: The cell cycle and check points representation.

The diagram depicts the distinct phases of the cell cycle, comprising G1 (Gap 1), S (Synthesis), G2 (Gap 2), and M (Mitosis), in conjunction with pivotal regulatory checkpoints, specifically the G1 checkpoint, intra-s phase checkpoint,G2/M checkpoint, and mitotic spindle checkpoint. G0 (quiescent) is a phase where cells temporarily or permanently exit the active cell cycle and enter a quiescent state.

One of the CID-containing and RNAPII -associated protein is RPRD1B which is also known as cell-cycle related and expression-elevated protein in tumour (CREPT) (Morales et al., 2014; Ni et al., 2011). Identification of RPRD1A protein initially started with the screening of genes involved in the regulation of one of the cell cycle inhibitory proteins, P15INK4b. RPRD1A protein is also named as p15-related sequence (p15RS) (Liu et al., 2002). Thus, these two proteins were identified as cell cycle related proteins

The accumulating evidence was demonstrated that RPRD1B and RPRD1A proteins have role in the cell cycle. First of all, these two proteins affected the transcription of several proteins which are cell cycle regulatory proteins. It was also observed that when expression of RPRD1A was inhibited using antisense sequence, CYCLIN-D1 and CYCLIN-E protein levels increased 2-3 times. (Liu et al., 2002). As a cell growth inhibitor, RPRD1A decreased expression the Wnt signalling pathway targeted gene including CYCLIN D1 and c-MYC in in MCF-7 cells (Wu et al., 2010). The upregulation of CREPT promotes an accelerated rate of tumor growth, while its depletion yields an opposing, decelerating effect on tumorigenesis (Lu et al., 2012). RPRD1B was defined as oncogene. Cancer cells proliferate uncontrolled way due to defect on cell signalling pathways such as cell-cycle regulation. Regulating role of RPRD1B on cell-cycle related protein genes is also demonstrated especially for CYCLIN-D1 gene. It has been suggested that RNAPII occupancy on the CYCLIN-D1 gene promoter facilitated by RPRD1B (Lu et al., 2012). Not only CYCLIN-D1 was affected by RPRD1B expression, but also RPRD1B overexpression enhanced the expression of activator genes for cell cycle such as cyclin D1, CDK6, CDK4, CYCLIN-E, CDK2, the reverse trend on these genes expression was observed when RPRD1B was knockdown (Lu et al., 2012; Zheng et al., 2016). These deregulated gene products by RPPD proteins are master regulators for cell cycle progression.

Depletion of RPRD1B resulted in the arrest of SW620 (colorectal cancer), Ishikawa (endometrial adenocarcinoma), Calu-1 (non-small-cell lung cancer), BCG-823, and MGC-803 (gastric cancer) cells at the G1 stage, indicating that it impedes cancer cell proliferation (Li et al., 2018; Sun et al., 2018; Wang et al., 2014; Zheng et al., 2016). RPRD1B regulates CYCLIN B1 expression to enhance the G2/M transition in gastric

cancer by interacting with Aurora B which is a kinase. They explained this mechanism by stimulating effect of phosphorylated RPRD1B on CYCLIN-B1 expression. The G2 phase and metaphase were prolonged under the RPRD1B depleted condition (Ding et al., 2018). In DF-1 cells, overexpression or inhibition of RPRD1A and RPRD1B had opposing effects on cell cycle regulation. While RPRD1A depletion decreased the proportion of cells in G1/G0 and S phase and increased the G2/M phase, overexpressed RPRD1B had the opposite effect. The opposite effect was observed under opposite conditions RPRD1A overexpression and RPRD1B knockdown (Jin et al., 2018). Taken together, these results suggest that RPRD1A and RPRD1B may have opposing roles in cell proliferation and cycle regulation.

The role of RPRD2 in cell cycle regulation remains poorly understood. However, one study suggests that depletion of RPRD2 in PM1 cells resulted in G2/M accumulation. RPRD2 protein levels vary during the cell cycle, with the lowest levels observed during G0/G1 phase, increasing through S phase, and reaching a peak at G2/M. (Gibbons et al., 2020)

Thus, the available evidence indicates that RPRD1A and RPRD1B play crucial roles in cell cycle regulation, while the role of RPRD2 in this process requires further investigation.

4. RPRD Proteins and Cancer

In a complex multicellular organism, cells possess the same unique genome, yet there exists a wide range of distinct cell types throughout the body. Gene expression patterns are a critical determinant of cellular morphology, function, and fate. This process, by which cells with identical DNA give rise to diverse cell types, is known as cell differentiation. Transcription factors, which recognize specific DNA sequences and bind to them, play a major role in regulating gene expression. While approximately half of all transcription factors are expressed in every cell, a small subset is vital for defining cell identity. The typical differentiated cells may transform into cancer cells due to genetic alterations that disrupt gene expression patterns (**Figure 10**). To safeguard the genome and detect mutations, cells employ numerous control mechanisms (Bradner et al., 2017).

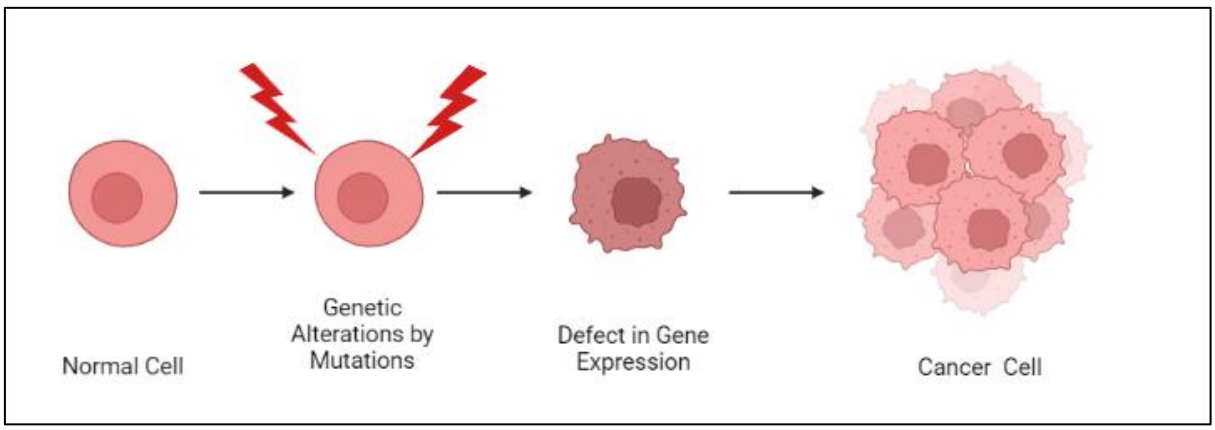


Figure 10: Cancer cell formation by genetic alteration and defect in gene expression.

Schematic representation of cancer cell formation through genetic alterations and dysregulated gene expression. The accumulation of genetic changes and dysregulated gene expression can transform a healthy cell into a cancer cell.

RPRD1B is overexpressed in approximately 85% of isolated tumour tissues from several cancer types, including lung, liver, breast, prostate, colon, stomach, cervical and endometrial. The overexpression of RPRD1B was correlated with enhanced cell proliferation, tumour formation, and shorter survival times of cancer patients (Lu et al., 2012).

The Wnt/β-catenin signalling pathway plays a role in important biological events such as cell cycle regulation, cell proliferation, differentiation and adhesion, tumorigenesis, synapse formation, adipogenesis and angiogenesis (Clevers & Nusse, 2012). In response to Wnt signalling pathway, while RPRD1A disrupted formation of β-catenin-TCF4 complex by recruiting HDAC2 and competing with β-catenin to bind TCF4, RPRD1B enhances the formation of β-catenin-TCF4 complex on β-catenin regulating genes (Liu et al., 2015; Zhang et al., 2014). It has been suggested that RPRD1A dimerized via its coil-coil domain to bind TCF4 to inhibit the binding of β. catenin. Upon activation of Wnt signaling, β-catenin separates dimerized RPRD1A upon interaction with monomer RPRD1A. Thus, dimerized RPRD1A is competed with β-catenin to forms a complex with TCF4 to regulate transcription (Fan et al., 2018). RPRD1A was defined as a tumour suppressor due to the antagonist effect for the Wnt/β-catenin signalling pathway. A downregulation of RPRD1A correlates with breast cancer metastasis and is directly mediated by miR-454-3p which has a role in breast cancer metastasis (Ren et al., 2019).

MicroRNAs (miRNAs) play a part in regulation of gene transcription via binding 3' UTR of the target miRNAs to inhibit translation by mRNA degradation. RPRD1B expression in glioma cells is inhibited by miR-596, but miR-449b-5p inhibits this expression in breast cell lines, resulting in suppression of the Wnt signalling pathway (Jiang et al., 2019; Wei et al., 2019). In colorectal cancer cell lines, it was also observed that RPRD1B interacted with p300 which is an activator for β-catenin-TF4 complex and RPRD1B helped the formation of these association (Zhang et al., 2018). Malignant cervical tissues showed correlations between RPRD1B expression and TCF4 expression (Wen et al., 2020). However, RPRD1A was defined as tumor suppressor protein which inhibit the formation of β-catenin.TF4.

AIMS OF THE PROJECT

Transcription converts information in DNA to RNA. Although every cell has the same DNA, differences in the transcription defines the cell identity to make different cell types. Many proteins and protein complexes can regulate mRNA synthesis in either positive or negative way. Transcription is a complex, multistep and highly regulated dynamic process that consists of three main steps: initiation, elongation and termination coupled with mRNA 3' end processing (Core & Adelman, 2019; Cramer, 2019). Transcription can result in the formation of R-loops, which occur when newly synthesized RNA hybridizes with the template DNA, resulting in a three-stranded nucleic acid structure. However, if R-loops accumulate excessively on the genome, it can cause genomic instability (Crossley et al., 2019).

During mRNA synthesis and maturation, regulatory proteins can interact directly with DNA or can bind to the CTD of RNAPII which coordinates transcription by recruiting these regulatory proteins. Proteins with a CTD-interaction domain (CID) can interact with RNAPII and control its activity (Harlen & Churchman, 2017). One group of the CID-containing and RNAPII -associated protein is RPRD proteins including RPRD1A, RPRD1B and RPRD2. Enhanced expression of RPRD proteins was shown several tumour samples. Numerous investigations have sought to elucidate the involvement of RPRD1B and RPRD1A in transcription, cellular proliferation, cell cycle regulation, and tumorigenesis; however, certain gaps persist in our understanding. Moreover, the functions of RPRD2 in the fields of transcriptional regulation, cell cycle modulation, and carcinogenesis remain relatively unexplored.(Li et al., 2021).

To address these knowledge gaps, this research endeavors to comprehensively investigate the activities of RPRD proteins, with a particular focus on theirrole in transcription and the involvement in R-loop formation—a nucleic acid structure that can result from transcription. Moreover, this project aims to assess the potential consequences of RPRD protein dysregulation in the context of cancer and investigate the mechanisms underlying how RPRD proteins modulate transcription and influence DNA damage repair processes. By achieving a deeper understanding of these processes, we seek to contribute valuable insights into the molecular basis of gene expression and its perturbations, particularly in cancer-related scenarios.

PART 2: MATERIAL and METHODS

1. Cell Systems and Cell Culture

The experiments were conducted using two cell lines, namely Hek293T and Flp-In T-REx 293. The HEK293 cells were originally derived from an aborted female fetus in 1973 and cultured by transfection with sheared adenovirus 5 DNA (Graham et al., 1977). HEK293T cells are a derivative of the HEK293 cell line and stably express the SV40 large T antigen, which confers an increased transient transfection efficiency(Lin et al., 2014). These cells are commonly used as a model organism in cellular analysis.

Flp-In TREx 293 cells another derivative of HEK293 cells and stably express the pFRT/lacZeo and the pcDNA™6/TR plasmids independently. To induce expression of the gene of interest, co-transfection with pcDNA5/FTR/TO (which contains the gene of interest sequence) and pOG44 (which expresses Flp recombinase for homologous recombination) is required. This cell line allows for tetracycline-inducible expression of the gene of interest. In our laboratory, we established a cell line by overexpressing RPRD within the Flp-In TREx 293 cellular system. To serve as a negative control in our experiments, we also developed a cell line that solely overexpresses the green fluorescent protein (GFP) within the Flp-In TREx 293 cells. Notably, all exogen RPRD proteins includes three GFP-tagged proteins for experimental purposes.

All cell lines were cultured in Dulbecco's Modified Eagle Medium (Corning, 10-013-CV) containing 10% heat-inactivated fetal bovine serum (LabTech) and 1% Penicillin/Streptomycin (Gibco, 15140) at 37°C with 5% CO2.

2. RPRD Proteins Knockdown and overexpression

To knockdown endogenous RPRD proteins, a siRNA transfection protocol was used. For this purpose, ON-TARGETplus SMARTpool siRNAs specific to RPRD1A (Horizon, L-007734-00-020), RPRD1B (Horizon, L-013787-01-0020), RPRD2 (Horizon, L-021712-01-0020), and an off-target negative control (Horizon, D-001810-10-50) were employed. HEK293T cells were seeded into p10 plates 24 hours before transfection, and after replacing the cell media with fresh DMEM, individual siRNAs were mixed with the siLentFect lipid transfection reagent (BioRad, 1703362) at a final concentration of 20nM in 1 ml serum-free OPTI-MEM (Thermo Scientific, 31985062). The mixture was left to stand at room temperature for 15-20 minutes before being directly added to plates for 48 hours of knockdown. Flp-In TREx 293 cells were transfected individually with the full-length RPRD sequences containing 3-flag tags at their C-terminals. After confirmation of transfection, the cells were grown as described above. To induce overexpression, tetracycline was directly added to the cell media at a concentration of 50 ng/ml and cells were incubated in tetracycline-containing media for 48 hours (**Figure 11**).

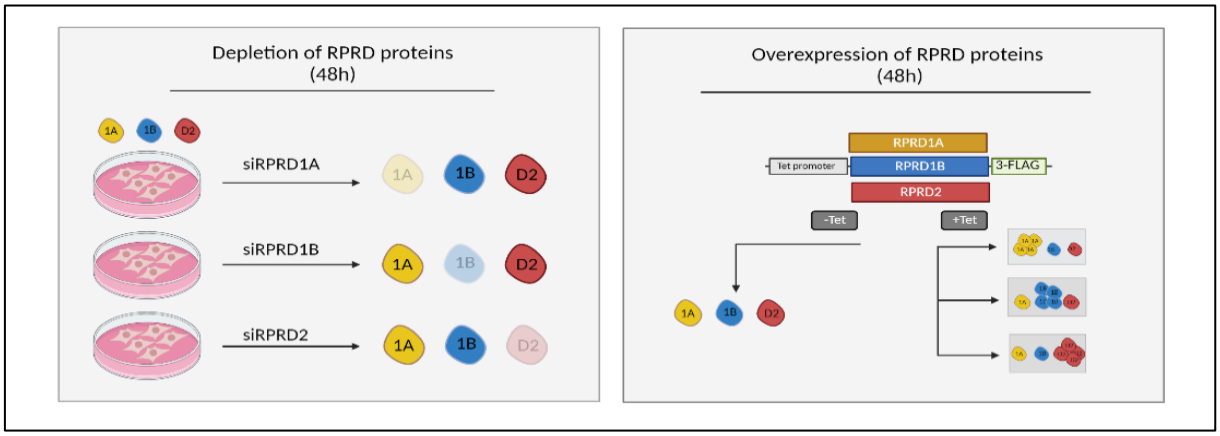


Figure 11: A representation of the cell systems that are being used

3. RNA isolation

The isolation of total RNA was performed using TRIzol reagent (Thermo Fisher, 15596026), following the manufacturer's instructions. When the cells in the p10 plate reached 70-80% confluency, 1 mL of TRIzol was directly added to the plate. The cells were then scraped and collected into a 1.5 mL tube, followed by incubation at room temperature for 5 minutes. After adding 200 µL of chloroform, the tube was vortexed for 15 seconds and incubated at room temperature for 3 minutes. Subsequently, the tube was centrifuged at 12,000 g for 15 minutes at 4°C, and the upper transparent phase was transferred to an RNase-free 1.5 mL tube. To precipitate the RNA, 600 µL of absolute isopropanol was added to the upper phase by inverting several times, and the tubes were left at room temperature for 10 minutes. The total RNA pellet was obtained by centrifuging at 12,000 g for 10 minutes at 4°C and washed with 75% ethanol, followed by air-drying at room temperature. The RNA concentration was determined using a Nanodrop after resuspending the RNA pellets in 50 µL of nuclease-free water.

4. RT-qPCR analysis

cDNA was synthesized using SuperScript III Reverse Transcriptase (Thermo Fisher, 18080044) and random hexamers (Thermo Fisher, SO142). A total of 2 µg of RNA was used for the reaction, following the manufacturer's instructions. The following oligonucleotides were used for quantitative PCR analysis (FD - forward primer, RE - reverse primer). The ΔΔCt method was utilized to analyse qPCR data and results were normalized using appropriate control gene and condition.

5. Chromatin Immunoprecipitation (ChIP)

To investigate potential interactions between chromatin and RPRD proteins, chromatin immunoprecipitation (ChIP) analysis was employed. ChIP is a reliable method for assessing the binding of specific proteins to chromatin, offering insights into their functional roles. Elucidating the binding profile of a protein across the entire gene locus can provide valuable insights into its functional role during transcriptional processes. Antibodies against RPRD1A, RPRD1B, and RPRD2 were used in this analysis to examine their binding patterns within the MYC gene. The SimpleChIP Enzymatic Chromatin IP Kit (CST, 9003) was used to perform ChIP experiement. HEK293T cells were grown to 70%-80% confluency, and the medium was removed. The cells were washed with ice-cold PBS and then fixed with 1% formaldehyde for 10 min at room temperature with rotation. The crosslinking reaction was stopped by incubation with glycine (125 mM) for 5 min at room temperature with rotation. The plate was washed twice with ice-cold PBS containing protease inhibitor cocktail (PIC) before the cells were scrapped. For nuclei preparation and chromatin digestion, cells were resuspended in 1 mL of ice-cold 1X Buffer A (250 µL 4X Buffer A + 0.5 µL 1M DTT + 5 µL 200x PIC) and incubated on ice for 10 minutes by inverting the tube every 3 minutes. The cell nuclei were pelleted by spinning down at 2000 g for 5 min at 4°C. The pellet was then resuspended in 1 mL of ice-cold 1X Buffer B (250 µL 4x Buffer B + 0.5 µL 1M DTT) following the centrifugation step. Nuclei pellets were resuspended in 100 µL 1x Buffer, and 0.5 µL of Micrococcal Nuclease was added for 20 minutes at 37°C to digest the DNA to a length of approximately 150-900 bp. Next, 10 µL of 0.5 M EDTA was added to inactivate the nucleases. The pelleted nuclei were then resuspended in 100 µL of 1x ChIP Buffer (10 µL 10x ChIP Buffer + 0.5 µL 200x PIC),

and the nuclei were sheared by sonication (high power, 30s on/off mode, 6 cycles). After precipitation of the sheared chromatin by centrifugation (at 9400 g for 10 min), the chromatin extract was diluted in 1x ChIP buffer. A mixture of 5 µg RPRD1A, 5 µg RPRD1B, and 20 µg RPRD2 antibodies (**Table 1**) and digested chromatin extract was incubated overnight at 4°C with rotation. For the negative control, one sample was incubated without antibody. An equal amount of ChIP-grade Protein A and Protein G beads were mixed and washed twice with 1x ChIP buffer. The antibody-conjugated chromatin extract was incubated with the beads for 2 hours with rotations. At the end of the incubation period, the beads were washed three times with 1 mL low salt (1x ChIP buffer) and once with 1 mL high salt (1x ChIP Buffer + 70 µL 5M NaCl) buffers, respectively. To elute the chromatin that interacts with antibodies, the beads were incubated in 150 µL 1x ChIP Elution Buffer (75 µL 2X ChIP Elution buffer) supplemented with RNase A for 3 hours at 65°C, then incubated with ProteinaseK for 1 hour at 45°C. The eluted sample was cleared with the QIAquick PCR Purification Kit (QIAGEN). q-PCR was performed with MYC gene primers shown in **Table 2**. Percent input (Input %) normalization was calculated for each sample using the following formula.

$$\text{Input \%} = \left(2^{((Ct(input) - Log_2^{Dilution\ Factor} - Ct(IP))}\right) * 100$$

Table 1: List of MYC gene primes used in the ChIP experiment

Primer	Sequence Forward (5’ to 3’)	Sequence Reverse (5’ to 3’)
PROM	CCCTCCCATATTCTCCCGTC	TCCCAATTTCTCAGCCAGGT
TSS	TCATAACGCGCTCTCCAAGT	GAAGGGTATTAATGGGCGCG
+150	GGAGGGATCGCGCTGAGTA	TCTGCCTCTCGCTGGAATTAC
+1400	GTTCCAGAACAGCTGCTAC	ACTCAATACGGAGATGCAA
+5300	AAGTACATTTTGCTTTTTAAAGT	GGCTCAATGATATATTTGCCAGTT
F1	GGAATTCTGCCCAGTTGATG	TCTGCGTGGCTACAGATAAGTT
F2	TGGCTGCTTGTGAGTACAGG	AACTGGCTTCTTCCCAGGAG
pA+1.7	CTGGTTGGAGACGACTGTAAA	GGAAATGGGCACTGGAGATAG
IFRG2	GCTCTTTATCTCTCTCTCAGCAA(	CAGTTTCGGTGTTCGGTTCA

Table 2:List of antibodies used in the experiments

Antibody Name	Provider	Catolog number	Used experiment
Anti-RPRD1A antibody	scbt	sc-81849	ChIP, Western blot
Anti-RPRD1B antibody	abcam	ab137246	ChIP, Western blot
Anti-RPRD2 antibody	abcam	ab17996	ChIP, Western blot
Anti- S9.6 antiboy	Sigma	MABE1095	DRIP
Anti-gamma H2A.X (phospho S139)	Abcam	ab81299	Western Blot
Anti-Mouse IgG Secondary Antibody	li-cor	925-68072	Western blot
anti-Rabbit IgG Secondary Antibody	li-cor	925-32213	Western Blot
Anti B-actin	Sigma	A2228	Western Blot

6. 4sU Labelling

To analyze the newly synthesized RNA, the 4sU labelling technique, a metabolic labelling approach for nascent RNA, was employed. 4sU, an uridine analogue commonly utilized for nascent RNA isolation, was employed in this study. HEK293T or Flp-In™ T-REx™ 293 cells were seeded into p10 plates. After 24 hours of seeding, Hek293T cells were transfected with appropriate siRNA to deplete RPRD proteins for 48 hours, while Flp-In™ T-REx™ 293 cells were induced with 50 ng final concentration of tetracycline to overexpress individual RPRD proteins for 48 hours. A 50 µM stock concentration of 4-thiouridine (4sU, Sigma T4509) was dissolved in pure water. Each differential RPRD expressed cell was treated with a final concentration of 500 µM 4sU for 5 minutes in the incubator in the dark. Upon completion of labelling, 4sU-containing medium was immediately discarded, and the plate was washed twice with ice-cold PBS. Subsequently, 1 ml of TRIzol (Fisher Scientific, 15596026) was directly added onto the cells to isolate RNA. After scrapping, cells were collected into 1.5 ml tubes, and total RNA extraction was performed with TRIzol following the manufacturers' instructions. To normalize the data, 4sU-labelled total RNAs were mixed with 4-thiouracil (4tU, Sigma, 440736)-labelled total *S.Cerevisiae* RNA after their concentrations were measured. A spike-in of 5 µg *S. Cerevisiae* RNA was added to 95 µg total RNA, resulting in a 19:1 ratio. The mixed RNAs were incubated at 65°C for 10 minutes, followed by immediate cooling on ice for 5 minutes. To biotinylate the 4sU-incorporated residues, RNAs were incubated with 200 µg EZ-Link™ HPDP-Biotin at 65°C for 30 minutes. To remove unbound biotins, 500 µl of chloroform was added, vortexed, and spun down at 4°C, maximum RPM for 5 minutes. The upper phases were carefully transferred to 500 µl of chloroform-containing phase-maker tube,

inverted gently 5 or 10 times, and let stand for 5 minutes at RT. After spinning down (at 4°C, maximum RPM for 5 minutes), the upper phases were moved into a new tube (~800 µl). To precipitate RNA, 2 volumes of absolute isopropanol and 1/10 volume of 5M NaCl were added to the upper phase. The mixture was incubated at -20°C for 2 hours or overnight. Tubes were spun down at maximum RPM for 10 minutes to precipitate the biotinylated RNA, followed by washing with 70% ethanol. RNAs were dissolved into 200 µl of water, and 30 µl of streptavidin beads were added for each sample. After washing with 200 µl of buffer A and 200 µl of buffer B, respectively, RNAs were incubated at 65°C for 5 minutes to open secondary structures. The washed beads were mixed with RNA in 400 µl of B&W buffer and incubated at room temperature for 1 hour with rotation. At the end of the period, beads were washed once with pre-heated (65°C). 1x B&W buffer, followed by four washes with 1x streptavidin B&W buffer at room temperature for 5 minutes each with rotation. The elution of 4sU-labelled RNA was performed by washing the beads twice with 100 mL 100 mM DTT, followed by precipitation with a mixture of 2 volumes of 400 µl of absolute isopropanol, 1/10 volume of 7.5 M ammonium acetate (20 µl) and 1 µl of glycogen. This mixture was incubated at -20°C for 2 hours or overnight. Subsequently, newly synthesized RNA was isolated through centrifugation (**Figure12**).

To synthesize cDNA, 5 µl of the eluted RNA was used. The confirmation of nascent RNA enrichment was achieved through qPCR analysis, employing exon-intron junction primers specific to the KPNB1 gene which is studied in an article (Laitem et al., 2015). The levels of nascent RNA were measured using q-PCR (**Table 3**).

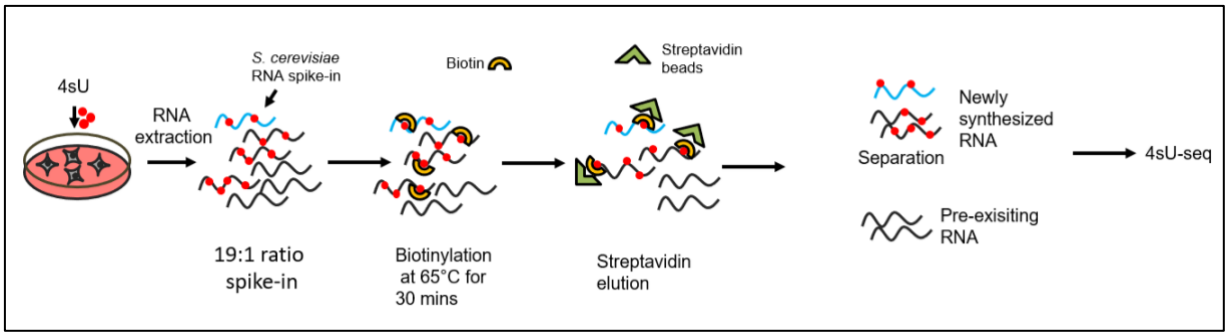


Figure 12: Representative image of the 4sU labelling experiment

Schematic of 4sU labelling experiment. Red dots indicate 4sU, blue lines are *S. cerevisiae* RNA black lines are RNAs which are extracted from cells. Yellow hemisphere represents biotin and green triangles are streptavidin beads.

Table 3: List of KPNB1 gene primers used in evaluating 4sU labelled RNAs

Primer Name	Sequence Forward (5' to 3')	Sequence Reverse (5' to 3')
Ex3	AGTCAGGTTGCCAGAGTTG	AGCATCAATAGCAAGCCAC
Int2	GTCCCCATCTCTCTCCTTTG	TCAACACGAGGTGGGAAAC
INT/EX3	CTCTGGTGGTTCAGTGCTCT	TCTGGTGGTTCAGTGCTCT
INT/EX13	TGGCAACCTGACTGTTTCCT	GTCGTTTTCTGGACAGCAG
INT/EX20	GCAGAACGCTTCATGTCCT	GTCTTCGATCTCCGCCCTT
ADH1	CTTGTGTGCTGGTATCACCGTC	GCAGCACCGGAGATAGCAAC

7. Library Preparation and Sequencing of Nascent RNA

The 4sU-seq libraries were prepared following the manufacturer's instructions using the NEBNext Ultra II RNA Library Preparation Kit for Illumina (E7770S). I initiated the library preparation process with 50 ng of nascent RNA for each experimental condition. To facilitate RNA molecule fragmentation, the RNA samples were subjected to incubation at 94°C for 15 minutes in a fragmentation buffer. Subsequently, first and second strand cDNA synthesis reactions were initiated using the appropriate buffers as prescribed in the kit protocol. The resulting double-stranded cDNAs were subjected to purification utilizing Ampure beads. An end repair process was executed to render the cDNA fragments amenable for adaptor ligation. Diluted adaptors, with a five-fold dilution factor, were ligated to the end-prepared cDNA molecules. To clear cDNAs from the ligation reaction buffer, AMPure XP Beads (Beckman Coulter, A63880) were employed for purification. PCR enrichment was subsequently performed to selectively amplify adaptor-ligated fragments, with distinct primers being employed for each experimental condition to introduce barcoding for distinguishing RNA samples originating from different conditions.

The library concentrations and sizes were determined using TapeStation (Agilent), and sequencing was conducted at the Genomic Facility at the University of Birmingham. An individual library was diluted and mixed to generate a library pool containing 8nM of 4sU-labelled RNA libraries. The library pool was sequenced on an Illumina NEXTseq sequencer.

8. Quantification of 4Su seq data

RNA-Seq (RNA sequencing) stands as a robust methodology employed for the comprehensive exploration of an organism's transcriptome. It facilitates the elucidation of gene expression profiles and the discrimination of various RNA species, encompassing messenger RNAs (mRNAs), non-coding RNAs, and more. Analyzing 4sU labelled RNA-Seq data adheres to a standardized pipeline. Initiating this workflow, quality control (QC) procedures were rigorously applied to the raw sequencing data, typically utilizing tools, FastQC, to ensure data integrity and reliability.

Subsequently, the high-quality reads were aligned to a reference genome, by using STAR. Due to 4sU data contains the yeast spike-in RNA , threads was aligned to yeast and human genome. Following alignment, the estimation of gene expression levels wasundertaken using tools HTSeq. These tools facilitated the quantification of RNA abundances and the generation of expression profiles. This step was pivotal for understanding the transcriptomic landscape under specific experimental conditions. Briefly,the raw gene counts, after sequencing and alignment can vary widely between genes and samples due to various factors, including gene expression levels and sequencing depth. In order to make the data more suitable for analysis and comparisons, log2 normalization was applied. After the log2 normalized counts, this made it easier to compare gene expression levels between samples and genes. The normalized counts were typically used in downstream analyses such as differential gene expression analysis, where we compared the fold change in expression between samples and visualization of the data.

Furthermore, to identify genes that exhibit differential expression between distinct experimental conditions, the application of statistical methods was imperative. DESeq2 is a widely recognized tool employed for this purpose, enabling the robust detection of genes whose expression levels vary significantly. This step was crucial for unraveling the molecular mechanisms underlying the observed biological differences.

To enhance the interpretability of the results and provide meaningful insights, the visualization of data was integral. Various plots and graphs were constructed to present the outcomes in a comprehensible manner. For this purpose, R libraries were utilized, allowing for the creation of informative visual representations that aid in the interpretation and communication of the findings. 4su seq analysis was done by Dr. Monika Sledziowska.

9. Preparation of synthetic R-loops

Genomic DNA was isolated from *Escherichia coli* using the following procedure. An overnight bacterial culture in 15 ml LB was transferred into 1.5 ml tubes and centrifuged at maximum rpm for 1 minute to pellet the cells. The supernatant was discarded and the cell pellet was resuspended in 600 µl of lysis buffer (0.6% SDS, 0.12mg/ml Proteinase K, and 1µg/ml RNase A in TE buffer) and vortexed. The cells in the buffer were then incubated at 37°C for 1 hour. Afterward, an equal volume (~600 µl) of phenol/chloroform was added to the cells and mixed well by inverting the tube (without vortexing) until the phases were completely mixed. The tubes were then spun down at maximum rpm for 5 minutes, and the upper phase was transferred into new 1.5 ml tubes. To remove all phenol from the mixture, an equal volume of chloroform was added again, mixed, and centrifuged at maximum speed for 5 minutes. The top aqueous layer was carefully moved into new 1.5 ml tubes.

To precipitate the genomic DNA, 2.5 volumes of ice-cold proof ethanol were added and mixed by inverting the tubes 8-10 times. The tubes were then incubated at -20°C for 30 minutes or more. At the end of the incubation period, the tubes were centrifuged at maximum rpm for 15 minutes, and the DNA pellet was washed with 70% ethanol and dried. The dried DNA was dissolved in TE buffer (10mM Tris-HCl pH:8, 1mM EDTA pH:8). The primer for spike-ins was used as previously described.

Table 4: List of primers used in production of synthetic R-loops.

Name	Sequence Forward (5’ to 3’)	Sequence Reverse (5’ to 3’)
L286	GGATCCTGCAGTAATACGACTCACTATAGGG CCGTGATGGTTTTCGCCTTT	ACTGCGACAATAATCCCGGA
H281	GGATCCTGCAGTAATACGACTCACTATAGGG AATGTCGCTGTGTTGTTGCT	CCAGAACACCATCAACACCC

Primers have a T7 promoter region for later use in-vitro transcription. For PCR reaction, mix was prepared using reagents and volumes as shown in Table 3.

Table 5: List of reagents and volumes used in PCR reaction

Reagent Name	Volume
5X Reaction Buffer	10 µL
10mM dNTP	1 µL
Forward primer (T7 promoter contains)	2.5 µL
Reverse primer	2.5 µL
Template DNA	300 ng
Polymerase (Phusion)	0.5 µL
Water	Up to 50 µL

In order to synthesize RNA from these DNA templates, Invitrogen™ Ambion™ MEGAshortscript™ T7 Transcription Kit was used. To perform in- vitro transcription, 200 ng DNA templates was used for in- vitro transcription. Following mix was prepared and incubated at 37 °C for 3 hours.

Table 6: List of reagents and volumes used in *in-vitro* transcription

Reagent Name	Volume
T7 10x reaction buffer	2 µL
T7 ATP solution	2 µL
T7 CTP solution	2 µL
T7 GTP solution	2 µL
T7 UTP solution	2 µL
Template DNA	<8 µL
T7 Enzyme Mix	2 µL

Upon completion of the incubation period for in vitro transcription, RNA precipitation was performed by mixing 115 µL of water, 15 µL of 7.5 M NH4AC, and 150 µL of phenol/chloroform (pH: 4.5). Tubes were centrifuged at 1200 g and 4°C for 25 minutes, and the upper aqueous phase was transferred to a new 1.5 mL tube. Two volumes of isopropanol were mixed with the aqueous phase and incubated overnight at -20°C to pellet RNA. Tubes were then centrifuged at maximum rpm, and the RNA pellets were washed with 70% ethanol, dried, and resuspended in 50 µL of pure water. Concentrations were measured using nanodrop spectrophotometry.

For hybridization reactions between DNA and RNA templates, a 1:20 ratio was used with a total volume of 50 µL, and 2.1 buffer (New England Biolabs) was used as the reaction buffer. The annealing reaction was performed at 95°C for 10 minutes for denaturation, followed by 70 cycles at 1°C decreasing intervals of temperature until it reached 25°C, holding at each temperature for 90 seconds. The RNA:DNA hybrids were run on a 0.9% agarose gel for 1 hour, and the hybrid bands were purified from

the gel by excising the band with a clean razor blade. Gel fragments were frozen for 1 hour or more at -20°C.

To prepare the elution column, a hole was created in the end of a 650 µL tube from the inside, and the cap was removed. A small piece of cotton was placed at the bottom of the tube using a 1 mL pipette tip to pack the cotton. The tube with the cotton was then placed in a 1.8 mL tube. The frozen gel fragment was placed into the 650 µL tube, and the elution tubes were centrifuged at maximum rpm and room temperature for 15 minutes. The final concentration of RNA:DNA hybrids was measured using the Qubit HS RNA assay kit.

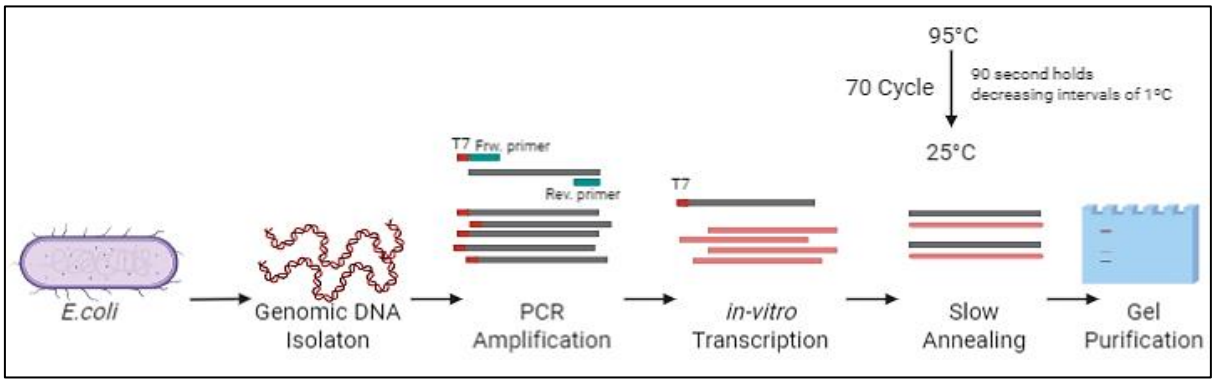


Figure 13: Schematic representation of synthetic R-loop generation.
Genomic DNA was isolated from *E. coli*, followed by PCR amplification utilizing primers containing the T7 promoter sequence. Subsequent to in vitro transcription, a controlled temperature decrease facilitated the annealing of DNA and RNA strands. Synthetic R-loops were subsequently excised and isolated from the gel.

I produced two different types of synthetic R-Loops. First one is named as L286 which is 286 base pair long and has 43% GC content. Second one is, H281, 281 base pair long and contains 58.4% GC content. I also use single and double stranded DNA fragments as negative controls.

10. DNA-RNA hybrid Immunoprecipitation

The DRIP experiment was conducted using buffers from the SimpleChIP® Enzymatic Chromatin IP Kit. Following 48 hours of protein depletion or overexpression, cells were incubated on ice for 10 minutes in 1 ml of 1X Buffer A (250 µl 4X Buffer A + 0.5 µl DTT). The pelleted nuclei were resuspended in 1 ml of 1X Buffer B (250 µl 4X Buffer B + 0.5 µl DTT). After a spinning step, the pellet was resuspended in 500 µl of 1X ChIP buffer. Samples were sonicated (high power, 30s on/off mode, 6 cycles), and then incubated with Proteinase K for 3 hours at 65 °C. DNA was extracted using the phenol/chloroform purification method as described briefly below. Phenol/chloroform (500 µl or 1V) was added to the samples and vortexed thoroughly. Samples were then centrifuged at 4 °C for 15 minutes at 16,000 × g, and the upper aqueous phase was carefully transferred to a fresh tube. The aqueous phase was mixed with 1.5 volumes of absolute ethanol and placed at –80 °C for at least 1 hour to pellet the DNA. After pelleting, the DNA was washed with 70% ethanol and resuspended in 200 µl of pure water. To shear the DNA, sonication was performed at high power, 30s on/off mode, with 6 cycles. Immunoprecipitation was performed using the same amount of DNA and synthetic R-loop spike-in in 1X ChIP buffer at 4 °C with overnight incubation. Then, 25 µl of Protein A and 25 µl of Protein G beads were mixed with the samples and incubated at 4 °C for 2 hours. After washing the beads, R-loops were eluted in 150 µl of 1X elution buffer for 30 minutes at 65 °C before a 2-hour Proteinase K incubation. The eluted sample was cleared using the QIAquick PCR Purification Kit (QIAGEN). Based on the primers shown in Table 6, q-PCR was conducted on the 5’ and 3’ ends of the genes.

Table 7: List of primers used in q-DRIP

Primer name	Sequence Forward (5’ to 3’)	Sequence Reverse (5’ to 3’)
NEAT1	GGAGAGGGTTGGTTAGAGAT	CCTTCAACCTGCATTTCCTA
TFPT	TCTGGGAGTCCAAGCAGACT	AAGGAGCCACTGAAGGGTTT
TRIM33-5'	ATGCCCAGCTTTCCCTAACT	GGAAAGTGGACTGCATGGTT
TRIM33-3'	TATGGCCACCATGCACTAGA	GGCTGGAGATAGAGCCTGTG
RPL13A-5'	GCTTCCAGCACAGGACAGGTA	CACCCACTACCCGAGTTCAAG
RPL13A-3'	AGGTGCCTTGCTCACAGAGT	GGTTGCATTGCCCTCATTAC

11. Rate of *in situ* Transcription

To explore the functions of RPRD1B, RPRD1A, and RPRD2 proteins on in situ transcription rate, the DRB stop-chase experiment was conducted. DRB (5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole) is a chemical compound that inhibits the pTEFb dependent phosphor-Ser2 of the CTD of RNAPII, thereby preventing it from entering the elongation phase. This inhibition can be reversed by removing DRB from the medium. The DRB stop-chase experiment was carried out following the protocol previously described (**Figure 14**) (Saponaro et al., 2014; Singh & Padgett, 2009).

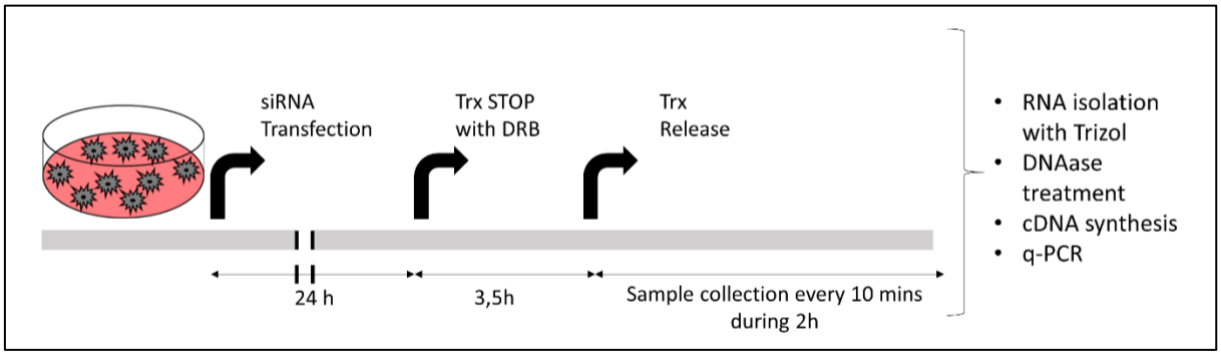


Figure 14: Representative image of the DRB treatment experiment

HEK293T cells were seeded into six-well plates and subjected to siRNA transfection using Ctrl, RPRD1B, RPRD1A or RPRD2 siRNAs for 48 hours. After transfection, cells were subjected to treatment with 100 mM of DRB (Sigma, D1916) for a duration of 3.5 hours.The wells were then washed with pre-warmed PBS and fresh medium (without DRB) was added to induce transcription. Samples were collected by adding TRIzol directly into the wells every 10 minutes. Total RNA was isolated using the manufacturer's instructions, and DNase treatment was performed using 5 µg RNA to eliminate background signal. RT-PCR was carried out using random hexamers, and

cDNA levels were quantified by qPCR using the same primers as previously reported (Singh & Padgett, 2009). All samples were normalized with 18S primer and untreated sample (no DRB).

Table 8: List of KIFAP3 gene primers used in DRB stop-chase experiment

Name	Sequence Forward (5’ to 3’)	Sequence Reverse (5’ to 3’)
Ex/In1	AAA TAA CCG CGC CTG CCT CTC AA	AAA CTA GCG TTG CCC AGT GAC A
Ex/In20	CCC TGC TAG GAA GAG AAT CTT GGT	TGG TTG GCC AAA GCC ATC CAT T
18s	TGT GCC GCT AGA GGT GAA ATT	TGG CAA ATG CTT TCG CTT T

12. Evaluation of RNA stabilization

To assess the stabilization of mRNA molecules, a prolonged treatment with 5,6-Dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB) was employed to inhibit transcription. Following either the depletion or overexpression of the RPRD1A RPRD1B and RPRD2 over a 48-hour period, 50 mM of DRB was introduced into the cell culture medium. An untreated control sample (referred to as the 0-hour sample) was collected immediately prior to the treatment of DRB. Subsequently, samples were collected at 2-hour intervals following the initiation of DRB treatment.

Total RNA was extracted from these collected samples using Trizol RNA isolation protocol. cDNA was synthesized from the extracted total RNA using random hexamer primers. q-PCR was performed using primers designed to target specific regions: exon 3 KPNB1 gene and the junction region between intron-exon 3 of the same gene, as previously described in the literature. Normalization of gene expression data was conducted using the 18S ribosomal RNA as a reference, and comparisons were made with the untreated control samples.

13. Western Blot

Protein isolation was carried out using lysis buffer containing 20 mM HEPES-Na, pH 7.4, 0.5% Triton, and 600 mM NaCl supplemented with a 1x cocktail of protease and phosphatase inhibitors (Thermo Scientific, A32961). The concentration of isolated proteins was determined using the Bradford assay. The 5X Bradford reagent was diluted with an equal volume of protein, and the same amount of protein was loaded onto NuPAGE gels, including 4-12% Bis-Tris gels (Thermo Fisher, 10247002) for small-sized proteins and 3-8% Tris-Acetate gels (Thermo Fisher, 12095655) for higher molecular weight proteins. A pre-stained protein marker (Thermo Fisher, 26619) was used to determine protein size. Proteins were run at 200V for approximately 1.5 hours until the loading buffer reached the bottom of the gel. The negatively charged proteins were then separated according to their molecular weight and transferred from the gel to a PVDF membrane (Merck, IPFL00010) using a semi-dry transferring system (Bio-Rad, 1704150). The efficiency of transferring was confirmed by staining the membrane with Ponceau. The membranes were then blocked with 5% milk in PBST for 1 hour at room temperature, followed by 3 washes of 5 minutes each with PBS-T to remove any residual milk. The primary antibody (was incubated with the membrane at room temperature for 1 hour or overnight at 4 °C. Following primary antibody staining, the membranes were washed 3 times for 5 minutes each and incubated with an appropriate IRDye secondary antibody for 1 hour at room temperature. The fluorescently labeled probe was detected using a Li-Cor Odyssey Classic Imager .

14. Cell cycle analysis

The effect of RPRD proteins throughout the cell cycle was assessed using the double thymidine block method, which enables the uniform synchronization of cells at the G1/S transition. Thymidine (Merck, T1895) treatment at a high concentration competitively inhibits deoxynucleoside metabolism, thereby inhibiting DNA replication. The first step of thymidine treatment induces cell arrest throughout the S phase, while the second treatment, after releasing, ensures uniform stacking of cells at the G1-S boundary (Harper, 2005).

The DTB protocol was followed as previously described for siRNA transfection (Pettinati et al., 2018). Tetracycline at a final concentration of 2 mM was added to the cell medium 6 hours after inducing Flp-In TREx 293 cells with tetracycline and transfecting HEK293T cells with siRNA. Cells were incubated with thymidine for 18 hours, after which the thymidine-containing medium was replaced with fresh medium and cells were released for 9 hours. Thymidine was then added to the cell media for 15 hours to achieve uniform synchronization at the G1/S boundary. Cells were washed with pre-warmed PBS and harvested at different time points (0, 4 hours, and 8 hours) as soon as the thymidine-containing medium was removed (time 0) by trypsinization. After pelleting, cells were washed three times with ice-cold PBS to remove debris. To fix cells, 70% ethanol incubation was performed overnight at 4°C. Then, cells were washed three times with ice-cold PBS and stained in 500 µl of PI (Thermo Fisher, BMS500PI) working solution (0.1 mg/ml RNase A, 25 µg/ml Propidium iodide) for 20 minutes at 37°C. Cells were pelleted and resuspended in PBS to run on an Attune NxT Flow Cytometer (Thermo Fisher). FlowJO software was used for data analysis.

15. CbioPortal data analysis

The cBioPortal is a web-based database utilized for the analysis of genomic and clinical data pertinent to cancer research. To investigate the impact of RPRD proteins on cancer progression, we selected a lower-grade glioblastoma dataset encompassing mRNA expression, mutation, putative copy number alterations, protein expression, and more, sourced from 514 patient samples.

To assess the relative mRNA expression of a specific gene within tumor samples, we calculated it relative to the gene's expression distribution across a reference population of samples. We estimated the gene's expression distribution by evaluating the mean and variance of expression values across the entire sample cohort. Utilizing the Z-score method, the degree of deviation of expressions within the reference population from the mean IS quantified. For the analysis of RPRD gene expression levels within tumor samples, a comparative evaluation was conducted, comparing each individual RPRD gene's expression distribution against a reference population comprised of all profiled samples. Expression levels were dichotomized into high or low categories using a Z-score threshold, defined as lower than-1 for low expression and higher than +1 for high expression, subsequently facilitating the analysis of the queried genes. Subsequently, we conducted a survival analysis by comparing survival data between high and low expression groups. The analyses pertaining to my inquiries can be accessed through the following URLs: For RPRD1A, please visit *https://bit.ly/3iL6B3N*. For RPRD1B, please visit *https://bit.ly/3XqmxqW*. And for RPRD2, please follow this link: *https://bit.ly/3Xj4qDM*. Subsequently, I conducted comparisons between survival data, protein expression data, and mRNA expression data among groups categorized based on high and low RPRD expression levels.

The protein expression data were acquired using the Reverse Phase Protein Array (RPPA) method, involving the immobilization of proteins from lysed cell or tissue samples, followed by probing with specific antibodies tagged with fluorescent dyes. The signal intensity generated was proportional to the target protein or modification levels within the sample. In this dataset, 206 proteins were analyzed using RPPA. Comparative analyses of protein expression were visualized through box plots, offering a summary of protein expression level distributions across distinct samples or patient groups, displaying median values, quartiles, and identifying potential outliers. Significance testing between compared groups was conducted using the t-test.

In addition, Kaplan-Meier survival curves were employed for survival analysis, estimating the probability of an event occurring over time, particularly patient survival. These curves featured time on the x-axis and the estimated probability of survival on the y-axis. To discern significant differences in survival between groups, Log-rank tests were utilized for the comparison of survival curves.

16. Gene Ontology Analysis

After obtaining a list of differentially expressed genes from glioma samples with RPRD deregulation, a Gene Ontology (GO) enrichment analysis was conducted to elucidate the underlying biological implications. GO enrichment analysis is a well-established method in molecular biology that associates biological functions, processes, and cellular locations with gene products to elucidate their roles.

To facilitate this analysis, GO terms were generated using the ShinyGO tool, a GO enrichment tool. These GO terms serve as standardized vocabularies widely used in molecular biology and bioinformatics to systematically categorize and annotate gene products based on their functions, processes, and subcellular localizations. Within the GO analysis, the 'Biological Process' category delineates the events and processes in which the differentially expressed gene products are involved. The 'Molecular Function' category provides insights into the biochemical and molecular activities that these gene products engage in. The 'Cellular Component' terms describe the specific cellular locations where these gene products are active. To address the challenge of multiple testing and to determine the significance of GO terms, a statistical method known as the False Discovery Rate (FDR) was employed. By applying a stringent criterion with a p-value cutoff of 0.05 (FDR), the dataset was filtered to focus on highly significant findings. The top 20 pathways were selected for examination, with data sorting based on the -log10(FDR) values. Generated (-log10(FDR)) values by using differentially expressed gene in RPRD deregulated glioma samples indicates the level of biologically relevant pathways and processes. Higher -log10(FDR) values correspond to more significant result.

PART 3: RESULTS

Chapter 1: Evaluating the role of RPRD proteins on transcription

1.1 RPRD proteins bind to 5’ and 3’ ends of the genes

To explore the potential interaction between chromatin and RPRD proteins, a chromatin immunoprecipitation (ChIP) analysis was employed. This approach was chosen due to the fact that identification of the genomic location of transcription factors can offer valuable insights into the functional role of the protein of interest. ChIP is a reliable method to test a protein of interest interactions with the chromatin. I employed this analysis using antibodies against RPRD1A, RPRD1B, and RPRD2 to investigate the binding pattern of these proteins throughout the MYC gene. Quantitative polymerase chain reaction (qPCR) was used to amplify different amplicons of MYC gene, and the percent input (Input %) was calculated from the input samples as outlined in the methods section. The percent input normalization represents the ratio of IP isolate target sequences to input isolate target sequences. The input sample measures the amount of chromatin used in ChIP, and 2% of starting chromatin was used as input during these experiments.

In a ChIP assay, several negative controls can be used, including a no antibody control and an isotype control. Mock samples represent the no antibody-added IP samples and pull down only the material that binds to the IP beads, which were typically Protein A/G conjugated magnetic beads. I also used normal rabbit IgG as a negative control antibody, which targeted isotype to pull down anything that might bind non-specifically to the antibody or the immunoprecipitation beads used throughout the ChIP experiment. If the amount of product in the negative control sample is the same as the

amount of product in the target-specific sample, it is possible to conclude that the target-specific antibody is showing non-specific binding or low levels of signal.

Primers were tested as a negative control locus (IFRG gene and downstream of the MYC gene (pA+1.7)) where the probability of the presence of the protein of interest is low. This will tell us whether the observed enrichment is specific or not. Including these controls is crucial since some antibodies result in non-specific enrichment. The qPCR following the ChIP experiment was performed using 8 different amplicons throughout the MYC gene. The distribution of these amplicons on the gene provided an idea of the binding pattern of the three RPRD proteins throughout the MYC gene.

The qPCR results following the ChIP experiments using 8 different primers across the MYC gene revealed that the anti-RPRD1B antibody significantly precipitated down primers found in the promoter region (PROM), 5' end (TSS and +150), and 3' end (+5300, F1, and F2) of the MYC gene compared to the Mock samples. However, primers found in the gene body (+1400 and +5300) of the MYC gene were not precipitated with the RPRD1B antibody. My results demonstrate that RPRD1B binds strongly to the 5' and 3' ends of the MYC gene. Previous research with different genes on the RPRD1B gene supports this findings (Lu et al., 2012; Ni et al., 2011). As expected, enrichment at negative control amplicons (pA+1.7 AND IFRG) was statistically insignificant (**Figure 15**).

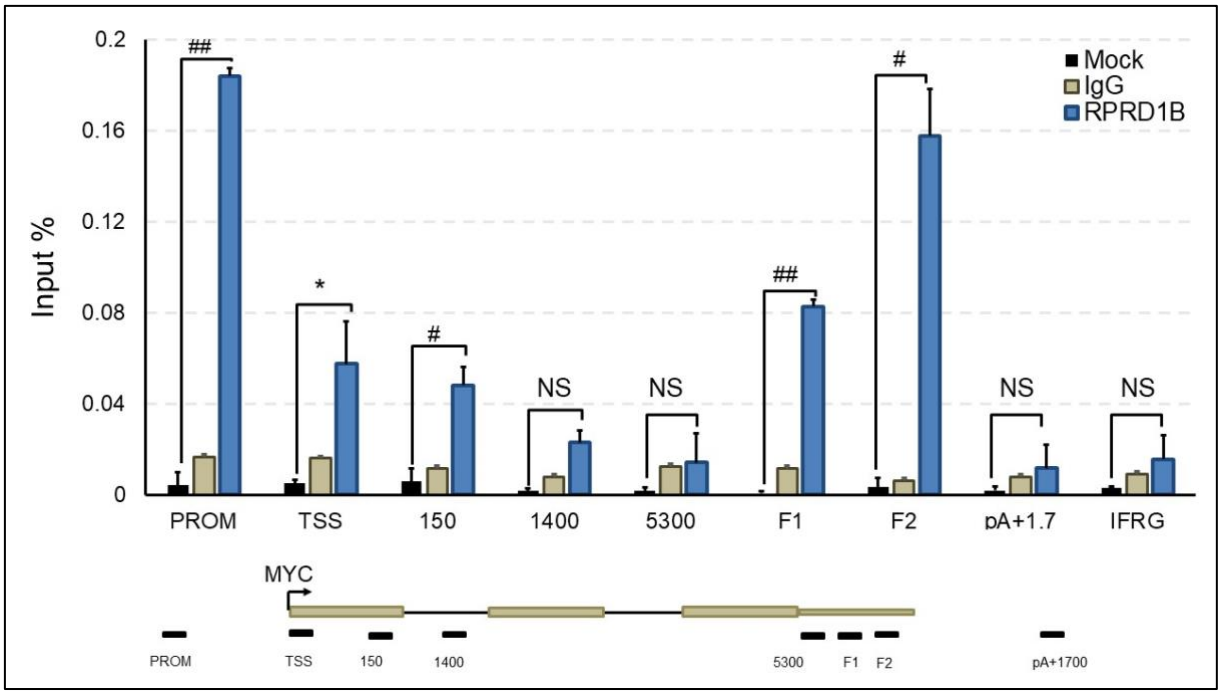


Figure 15: qPCR results following ChIP for RPRD1B enrichment at the MYC gene's promoter, gene body and downstream of pA site in HEK293T cells.

ChIP enrichment is presented as % input. The mean with SD of at least three replicates of IP is shown for all analyzed MYC gene regions. Asterisks and hashtags indicate significant values compared to Mock group. (*p < 0.05, **p < 0.01, # p < 0.005 and ## p < 0.001 (calculated by two-sided unpaired t test)) NS non-significant.

To investigate the occupancy pattern of RPRD1A protein on the MYC gene, a ChIP experiment was conducted using anti-RPRD1A antibody. The ChIP assay revealed that the anti-RPRD1A antibody significantly precipitated chromatin at the promoter (PROM), transcription start site (TSS), +150, and +1400 sites of the MYC gene, while the occupancy of RPRD1A at the poly(A) site (+5300), downstream of the MYC gene (pA+1.4), and negative control (IFRG) site was not significant. Despite the majority of the regions being precipitated with RPRD1A, the level of precipitation was not as high as that of RPRD1B, which may be attributed to the antibody specificity. These results suggested that RPRD1A exhibits a somewhat uniform binding pattern throughout the

MYC gene, with a slightly higher enrichment observed near the promoter region (**Figure 16**).

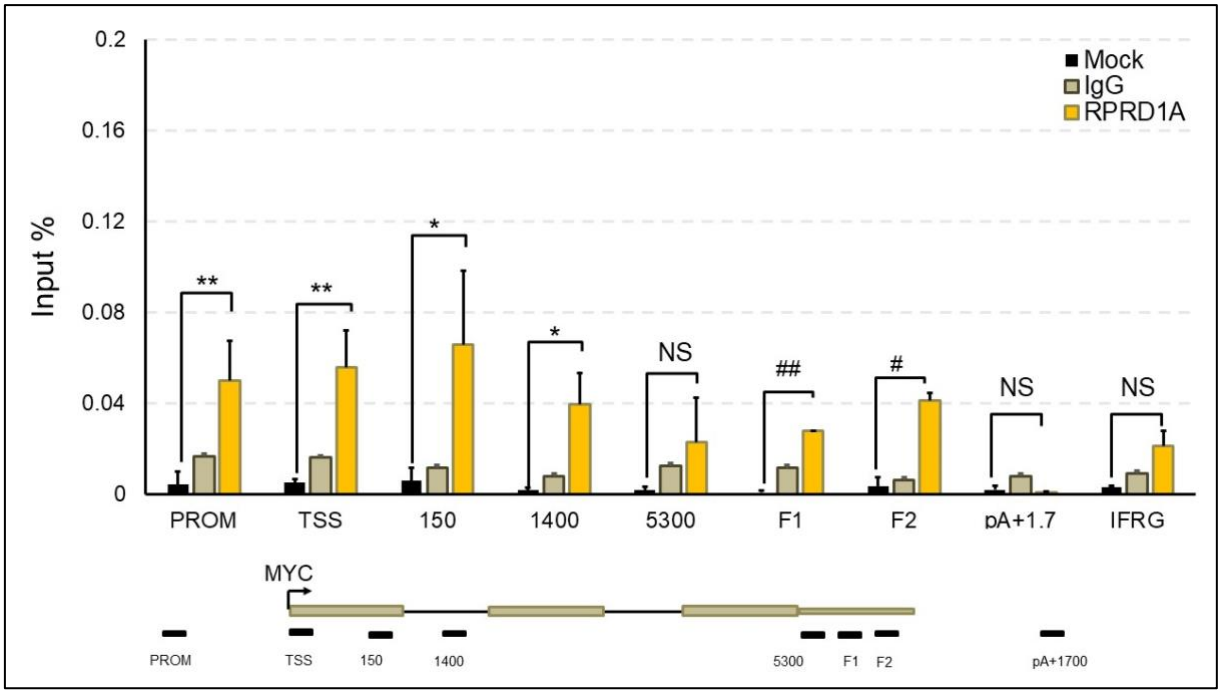


Figure 16: qPCR results following ChIP for RPRD1A enrichment at the MYC gene's promoter, gene body and downstream of the pA site in HEK293T cells. ChIP enrichment is presented as % input. The mean with SD of at least three replicates of IP is shown for all analyzed MYC gene regions. Asterisks and hashtags indicate significant values compared to Mock group. (*p < 0.05, **p < 0.01, # p < 0.005 and ## p < 0.001 (calculated by two-sided unpaired t test)) NS non-significant.

I also conducted the ChIP assay using the anti-RPRD2 antibody, the third member of the RPRD family. My findings revealed that the promoter site and termination site of MYC gene were significantly enriched with the anti-RPRD2 antibody, as shown in (**Figure 17**).

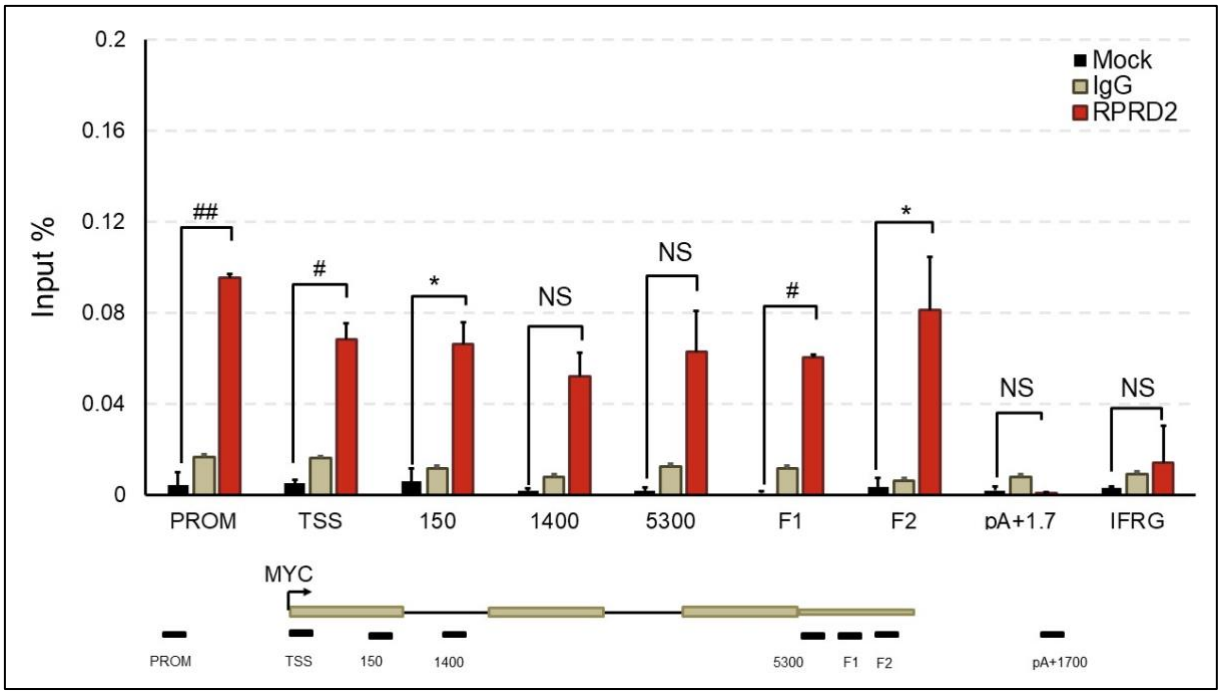


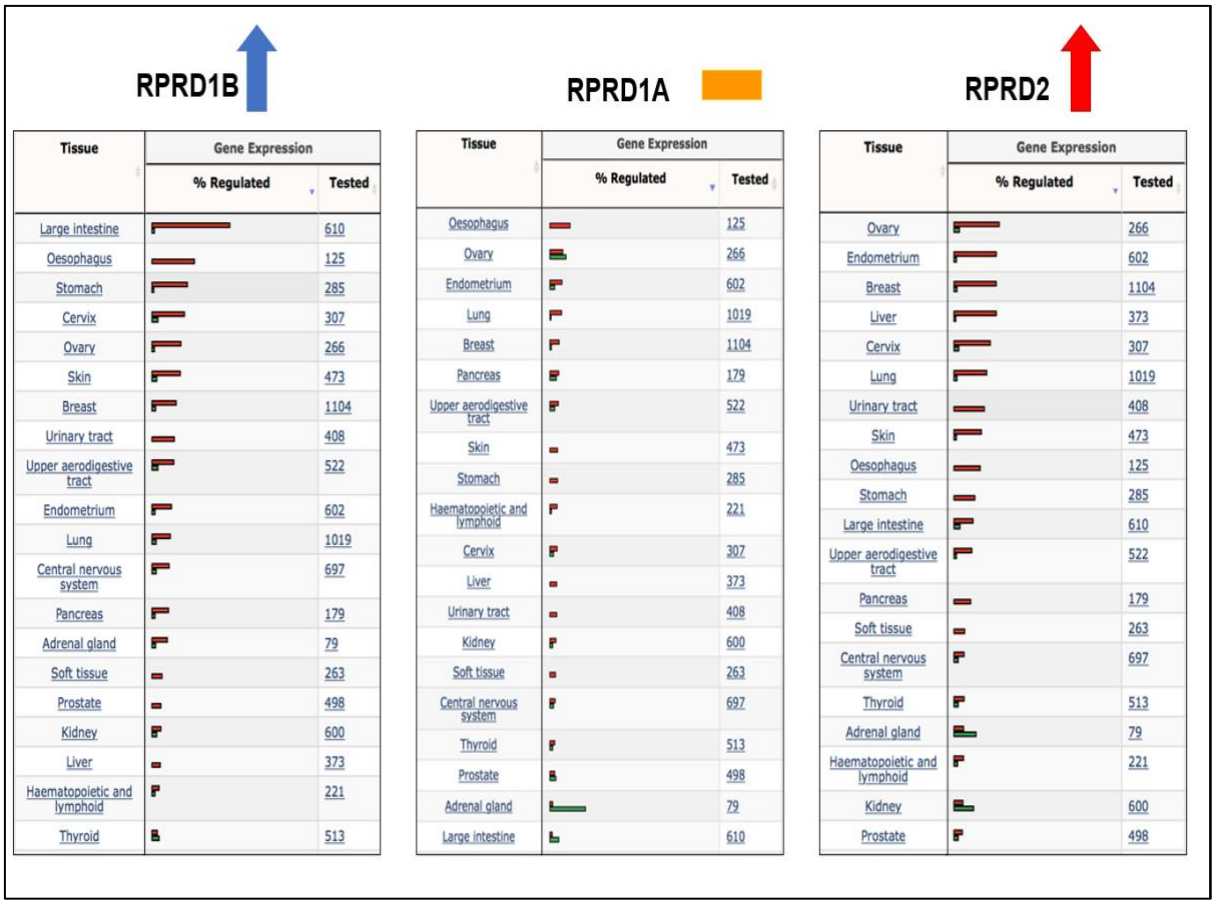
Figure 17: qPCR results following ChIP for RPRD2 enrichment at the MYC gene's promoter, gene body and downstream of the pA site in HEK293T cells.

ChIP enrichment is presented as % input. The mean with SD of at least three replicates of IP is shown for all analyzed MYC gene regions. Asterisks and hashtags indicate significant values compared to Mock group. (*p < 0.05, **p < 0.01, # p < 0.005 and ## p < 0.001 (calculated by two-sided unpaired t test)) NS non-significant.

The results of my study indicated that all three RPRD proteins exhibit a strong binding pattern at the 5' ends of the MYC gene. While there is no information in the literature regarding the binding pattern of RPRD2 protein, the occupancy pattern of RPRD1B and RPRD1A proteins in my study is consistent with previous researches (Ni, et al. 2011). The results indicate that the RPRD1B proteins exhibit a more robust cross-linking pattern with the MYC promoter region and a region linked to transcription termination in comparison to RPRD1A and RPRD2.

1.2 Evaluation of protein depletion and overexpression

To investigate the precise function of RPRD proteins in transcription, the initial step involved selecting an appropriate cell system for subsequent experimentation. Using the Catalogue of Somatic Mutations in Cancer (COSMIC, https://cancer.sanger.ac.uk) database, I visualized the gene expression level of individual RPRD proteins (**Figure 18**). The COSMIC database is the source of information on somatic mutations related to human cancers (Tate et al., 2019). The yearly updated data were collected here, and this data provided a comprehensive overview of the cancer genomic landscape. While RPRD2 is also an RNAPII interacting protein, alongside RPRD1A and RPRD1B, its functions are largely unknown, and no study has shown its involvement in cancer. The most of analysed different kind of cancer tissues highly overexpressed RPRD1B and RPRD2, however RPRD1A protein expression level did not change significantly. RPRD1B and RPRD2, not RPRD1A, probably have a significant role in the transition from normal to cancerous state for a cell (**Figure 18**). Based on these results, I utilized two distinct cell systems to investigate the role of RPRD proteins in cells. Overexpression and depletion of a specific protein in a cell are commonly used techniques to understand the function of that protein. Overexpression can help to determine the role of a protein by observing the effects of an excess amount of that protein in the cell. Depletion of a protein, on the other hand, can reveal the effects of the absence of that protein in the cell.



Tissue	Gene Expression	
	% Regulated	Tested
Large intestine		610
Oesophagus		125
Stomach		285
Cervix		307
Ovary		266
Skin		473
Breast		1104
Urinary tract		408
Upper aerodigestive tract		522
Endometrium		602
Lung		1019
Central nervous system		697
Pancreas		179
Adrenal gland		79
Soft tissue		263
Prostate		498
Kidney		600
Liver		373
Haematopoietic and lymphoid		221
Thyroid		513

Tissue	Gene Expression	
	% Regulated	Tested
Oesophagus		125
Ovary		266
Endometrium		602
Lung		1019
Breast		1104
Pancreas		179
Upper aerodigestive tract		522
Skin		473
Stomach		285
Haematopoietic and lymphoid		221
Cervix		307
Liver		373
Urinary tract		408
Kidney		600
Soft tissue		263
Central nervous system		697
Thyroid		513
Prostate		498
Adrenal gland		79
Large intestine		610

Tissue	Gene Expression	
	% Regulated	Tested
Ovary		266
Endometrium		602
Breast		1104
Liver		373
Cervix		307
Lung		1019
Urinary tract		408
Skin		473
Oesophagus		125
Stomach		285
Large intestine		610
Upper aerodigestive tract		522
Pancreas		179
Soft tissue		263
Central nervous system		697
Thyroid		513
Adrenal gland		79
Haematopoietic and lymphoid		221
Kidney		600
Prostate		498

Figure 18: Gene expression level of RPRD proteins in different cancer tissues (generated by using COSMIC database)

In the generated histogram tables, Gene Expression data was shown in the total analyzed samples as upregulated and downregulated. In the '% Regulated' column, the red line demonstrated the overexpressed percentage of samples, and the green line demonstrated the downregulated percentage of samples. Columns that were labelled as 'Tested' indicated the total number of samples analyzed. Data were sorted by overexpression of samples from high to low for related proteins in different cancer tissues.

To deplete RPRD proteins, I employed a specific and effective siRNA knockdown method. HEK293T cells were transfected with a siRNA targeting each RPRD proteins for 48 hours, and the RPRD levels were quantified using western blotting. Although I initially used a 24-hour knockdown, I found that a 48-hour knockdown was more effective in terms of depletion. As shown in **Figure 19** , the protein band intensities in the depleted samples were significantly weaker than those in the control sample after 48 hours knockdown. Interestingly, I observed that RPRD1B knockdown affected the level of other RPRD proteins, whereas RPRD1A and RPRD2 knockdown did not show this effect, which may be due to the presence of RPRD1B protein in different protein complexes. To overexpress RPRD proteins, RPRD proteins were clonned of interest in a Flp-In expression vector and generated Flp-In TREx 293 cell lines. These cells contain one stably integrated FRT site at a transcriptionally active genomic locus and overexpress RPRD proteins under a tetracycline-inducible promoter upon tetracycline induction. I used various concentrations of tetracycline to achieve the optimal level of overexpression of RPRD proteins and chose 50 ng/ml for tetracycline induction. I used GFP overexpressing Flp-In TREx 293 cells as a control (negative control). To assess the levels of endogenous and exogenous RPRD proteins, I conducted western blot experiments using Flp-In TREx 293 cells. Flag-tagged RPRD proteins were observed over the original size of the RPRD protein bands, indicating that the thick bands over the RPRD protein bands were endogenously expressed RPRD proteins. The generated cells overexpressed RPRD protein of interest as shown in **Figure 19**, and I used these cells as a model of RPRD overexpressing cells throughout the project.

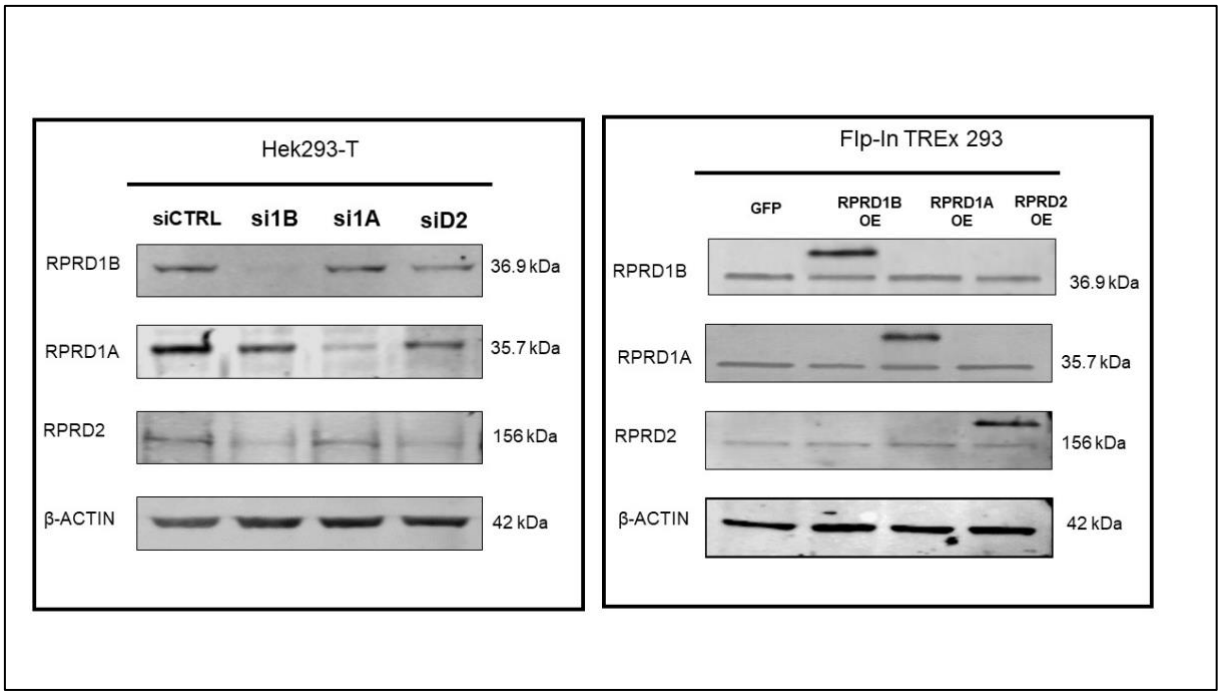


Figure 19: Evaluation of RPRD protein levels after siRNA knockdown to HEK293T cells and tetracycline induction to Flp-In TREx 293 cells.

HEK293T cells were transfected with specific siRNA. The siCTRL represents the non-specific siRNA transfected cells (negative control). It is marked as si1B, si1A, and siD2 in order to represent the cells which transfected with RPRD1B, RPRD1A, and RPRD2 siRNAs respectively. Flp-In TREx 293 cells were induced with 50 ng/ml tetracyline for 48 hours. The GFP line demonstrates the GFP overexpressing cells (negative control). It is marked as RPRD1B OE, RPRD1A OE, and RPRD2 OE in order to indicate the cells which overexpress the labelled RPRD protein after tetracyline induction.

1.4 RPRD proteins have a role in nascent RNA synthesis

After inducing deregulation of RPRD proteins through depletion and overexpression, I conducted qPCR analysis using total RNA to assess the impact of RPRD proteins on transcripts. My study investigated changes in steady-state mRNA levels in cells with altered RPRD expression. Specifically, I examined the ex3 and ex21 regions of the KPNB1 gene. The relative mRNA levels in cells with downregulated or upregulated RPRD expression was illustrated (**Figure 20**). My results did not reveal any significant differences between the control samples (siCTRL or GFP overexpression) and the samples with altered RPRD expression. Next, I performed total RNA sequencing on siCTRL and RPRD1B depleted cells (**see the Appendix Figure A4**). However, I did not observe any significant difference between the two groups. I also performed 3'-sequencing to detect changes at the total RNA level after 48 hours of RPRD depletion or overexpression (Winczura et al., 2021). Although a few genes showed differential expression, the effect of RPRD on total transcription may have been buffered by the RNA stability pathway of the cell. It is important to note that these experiments were performed after 48 hours of depletion or overexpression, and the accumulation of nascent RNA in the total RNA level may require more time than 48 hours.

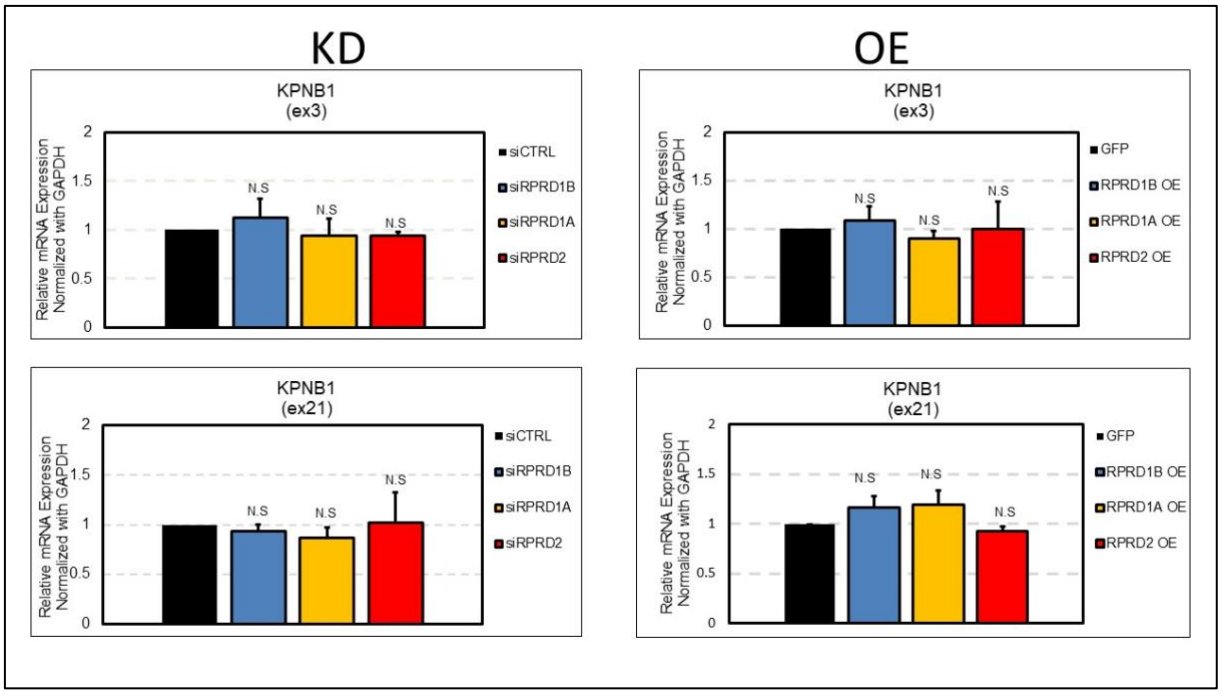


Figure 20: The comparison of the total mRNA levels of KPNB1 gene in RPRD proteins depleted and overexpressed cells.

qPCR was employed to evaluate the total mRNA levels in cells where RPRD proteins were depleted and overexpressed. Normalization of mRNA levels was conducted relative to the expression of GAPDH mRNA. Two distinct exonic regions within the KPNB1 gene were examined in both RPRD proteins-depleted and overexpressed cell conditions. The presented bar graphs depict the average results derived from three independent experiments (n=3), with error bars representing the standard deviation. p-value<0.05, N.S not significant, for a two-tailed T test.

After observing no significant alterations in total RNA levels in the RPRD protein deregulated samples, I intended to assess the levels of newly synthesized RNA in these samples. Given that RPRD proteins contain a CID domain and interact with the CTD, it is reasonable to assume that they play a role in transcription. To isolate newly synthesized RNA molecules, a 4sU labeling experiment was performed. This experiment aims to investigate RNA dynamics and transcriptional regulation by tracing the incorporation of 4sU into RNA molecules. 4sU (4-thiouridine) is a nucleoside

analog that can be incorporated into RNA during transcription, and can be subsequently labeled with a biotin moiety for purification and sequencing. This technique can provide insights into the mechanisms underlying gene expression and regulation, and can be applied to a variety of biological systems and experimental conditions.

To investigate the role of RPRD proteins in nascent RNA synthesis, a 4sU labelling experiment was performed. Newly synthesized RNA was labelled with 4sU for five minutes. The total RNA was then mixed with 4tU-labelled yeast total RNA as a spike-in to normalize nascent RNA and eliminate differences between sample purification processes. The nucleoside analog, 4-thiouridine (4sU), presents a notable restriction in terms of cellular uptake within unmodified yeast cells. This restriction needs the expression of a specific transporter to facilitate its uptake. In yeast, a distinct enzymatic process involving uracil phosphoribosyltransferase enzymatic activity results in the conversion of the base analog 4-thiouracil (4tU) to the nucleotide analog, 4sU monophosphate, allowing for its incorporation into nascent RNA molecules in yeast. Importantly, this metabolic conversion remains distinct from mammalian cells, where it does not occur. Consequently, the selective use of 4sU labeling in human cell lines and 4tU labeling in yeast cell experiments is predicated upon these differential metabolic pathways.

The efficiency of elution and levels of transcription were evaluated with qPCR prior to sequencing of samples. The nascent RNAs from 4sU-labelled (black bars) and non-labelled mock (green bars) samples were compared between siCTRL transfected HEK293T and Flp-In TREx GFP overexpressing cells. Both cell types were assumed to have unaltered transcription levels. RT-qPCR was performed using intron-exon

junction primers for the KPNB1 gene to eliminate the steady-state RNA contamination on nascet RNA. Significant enrichment of eluted RNA was observed in 4sU-labelled samples compared to non-labelled samples for all three intron/exon junction regions for the KPNB1 gene (**Figure 21**)**.** Additionally, as expected for nascent RNA, there was a significant enrichment at the intronic region (intron 2). The enrichment of nascent RNA for the exon-3 region of the KPNB1 gene and 18S RNA was observed to be lower compared to intronic and intron/exon regions, despite being higher than the mock samples. This could be attributed to possible contamination of total RNA during nascent RNA elution.

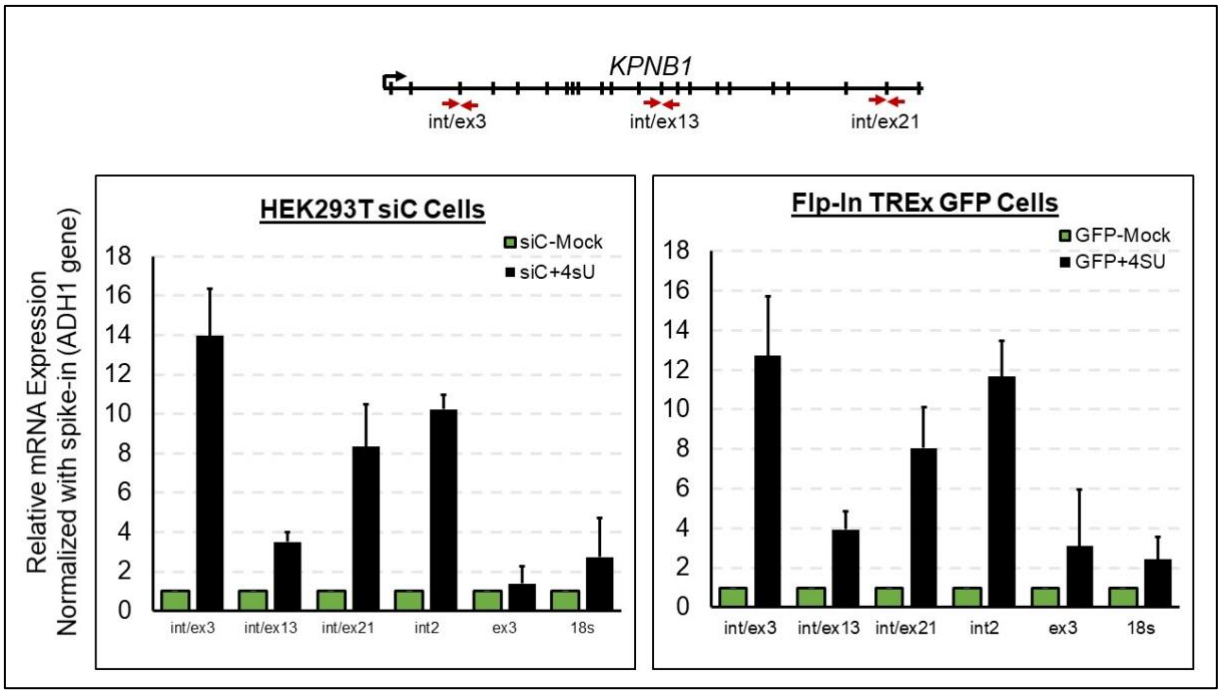


Figure 21: Comparison of precipitated nascent RNA between 4sU-labelled and Mock sample in HEK2893T and Flp-In TREx-GFP cells.

The graphs present the mean values obtained from three distinct and independent experimental. The green bars correspond to samples not subjected to 4sU treatment (Mock), while the black bars represent samples subjected to 4sU treatment. The error bars accompanying the data points indicate the standard deviation. the graphs include illistarion of the KPNB1 gene representation and the respective primer regions positioned above.

To assess the impact of RPRD proteins on nascent transcription, I performed q-PCR with the nascent RNAs after elution. RPRD depletion was evaluated by comparing with control siRNA transfected samples, while RPRD overexpression was compared with GFP overexpression samples. In both knockdown and overexpression graphs, blue bars indicated RPRD1B, yellow bars indicated RPRD1A, and red bars represented RPRD2 (**Figure 22**). The black bar in the knockdown chart indicated control siRNA transfection, while the black bar in the overexpression chart represented GFP overexpression samples. q-PCR results were normalized with co-purified yeast spike-

in RNAs using ADH1 gene primers (**Figure 22**). RPRD proteins had an effect on the nascent transcription of the KPNB1 gene. The qPCR result was normalized using yeast spike-ins. While RPRD protein depletion increased nascent transcription of the KPNB1 gene, the reverse effect was observed in RPRD overexpressing cells. For both knockdown and overexpression conditions, RPRD1B and RPRD2 proteins had more pronounced effects on nascent transcription, while RPRD1A had a slightly lower impact on transcription of the KPNB1 gene **(Figure 22)**.

Next, I compared the enrichment of newly synthesized RNA and total RNA over 18S for the KPNB1 int-ex3 junction and exon 3 regions. Higher enrichment was expected in newly synthesized RNA than in total RNA. After confirming nascent RNA, I sequenced 4sU labelled RNAs for both individual knockdowns and overexpression to observe the genome-wide effect of RPRD proteins on the nascent transcriptome. Libraries for sequencing were prepared using the NEBNext Ultra II RNA Library Preparation Kit

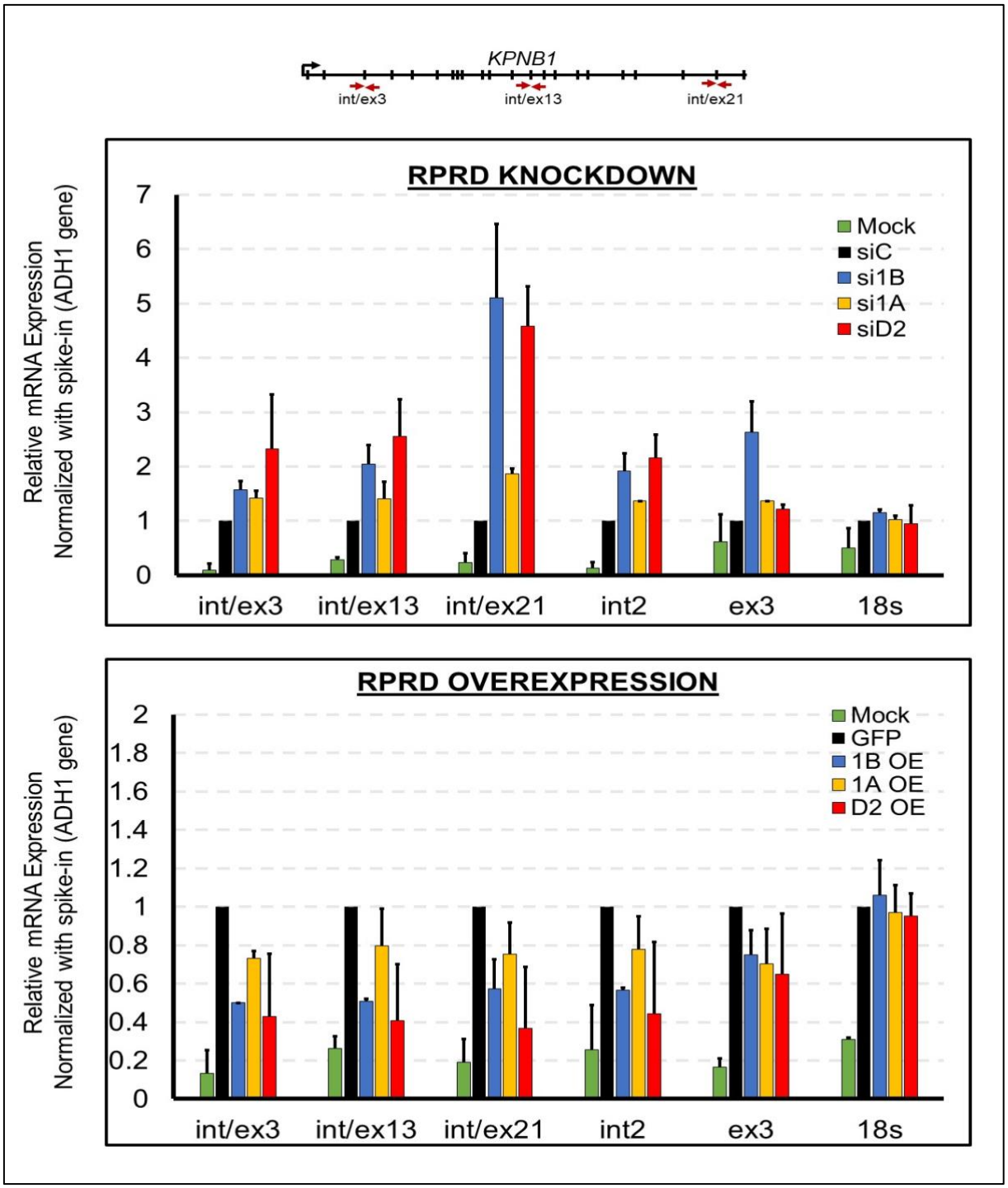


Figure 22: Effect of RPRD proteins on the newly synthesized RNA level of the KPNB1 gene.

The graphs shows the avarege of three independent experiments. The bars represent the Mock (green) siCTRL/GFP (black), 1B KD/OE (blue) 1A KD/OE (yellow), D2KD/OE (red). The error bars indicates the standart deviation. The KPNB1 gene representation and primers regions are demonstrated above the graphs.

1.5 RPRD proteins are negative transcription regulators

In recent years, advances in RNA sequencing techniques have enabled researchers to examine global differences in nascent RNA production across the genome. In this particular study, I investigated the impact of RPRD protein depletion and overexpression on the accumulation of nascent RNA using 4sU RNA-seq.

Normalization of sequenced reads was performed using co-purified yeast spike-in RNAs, and approximately 60,000 transcription units were assessed for analysis. The DEseq2 median ratio method was preferred for normalization to consider the depth of sequencing and RNA composition. GFP overexpression or non-targeting siRNA transfection were used as negative controls.

The scatter plots in illustrate log2 normalized counts of nascent RNA for individual genes, when comparing RPRD KD or RPRD OE samples with control samples (**Figures 23 and 24**). Each dot on the plot represents an individual gene, with the dashed red diagonal line representing the zero difference line, indicating equally expressed genes in both control and KD samples. The x-axis (horizontal) of the scatter plots shows normalized counts of control samples, while the y-axis (vertical) indicates KD of the related RPRD protein, which is labelled on each graph. The blue line on the plots represents the regression trend between control and KD samples. The proximity or distance between the regression line (blue) and zero difference line (red) indicates the level of differential expression between compared samples. If the regression line is close to the zero difference line, this points to similarity in expression levels between compared samples. If not, there is variation in global transcription levels between the two samples.

The number of dots representing newly synthesized genes above the zero-difference line (red) indicates a high number of highly expressed genes in vertical line samples. Compared to control samples, depletion of indicated RPRD with siRNA caused an accumulation of newly synthesized RNA levels. The most evident global accumulation of newly synthesized RNA was observed in RPRD2 KD samples. While loss of RPRD1B protein induced a moderate accumulation of newly synthesized transcriptome, RPRD1A KD had a minimal effect on the accumulation of nascent RNA transcription (**Figure 23A**). The box plot in represents the log2 fold enrichment for the level of nascent RNA differential expression of RPRD proteins over appropriate control (**Figure 23B**). The line at zero (log2(kd/control)=0) indicates that the expression level of the gene is equal to that of the control samples (kd/control=1). I observed that RPRD2 and RPRD1B have positive log2 fold enrichments, which show an increased amount of transcription compared to control.

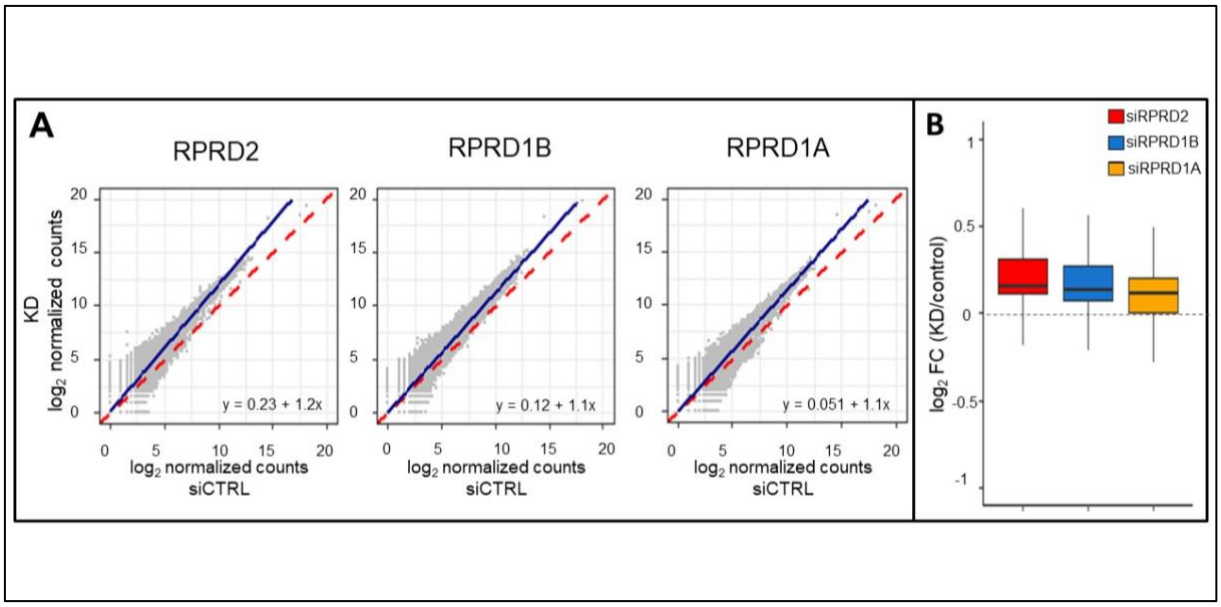


Figure 23: The global effect of RPRD depletion on nascent RNA transcriptome. The scatter plots (A) illustrate log2 normalized gene counts, facilitating a comparative analysis between RPRD-depleted cells (y-axis) and siCTRL samples (x-axis). The red dashed diagonal line represents the zero difference line, while the blue line signifies the regression line. The box plots (B) depict the log2 fold change values between RPRD-depleted and control samples.

The same methodology employed for 4sU-seq of RPRD KDs was used to evaluate the nascent transcriptome in overexpressing cells. Flp-In TREx293 cell systems were utilized to separately overexpress RPRD proteins. Under the control of a tetracycline-regulated promoter, the protein sequences of RPRD1A, RPRD1B, and RPRD2, each tagged with a C-terminal 3-Flag, were integrated into the FTR locus of these engineered cells. Upon induction with a final concentration of 50 ng/ml tetracycline, cells began expressing both exogenous and endogenous copies simultaneously. GFP overexpressing cells were used as a negative control for the analysis of RPRD overexpressing cells. The scatter plot depicts the genes as dots, which accumulate below the zero-difference line and indicate higher expression in GFP samples (x-axis) than RPRD overexpressing samples (y-axis). Compared to RPRD depleted cells,

RPRD overexpressing cells exhibited a completely opposite effect. RPRD2 and RPRD1B overexpression reduced nascent transcription when compared to GFP overexpression. Nevertheless, overexpressing RPRD1A slightly decreased nascent transcription (**Figure 24A,B**).

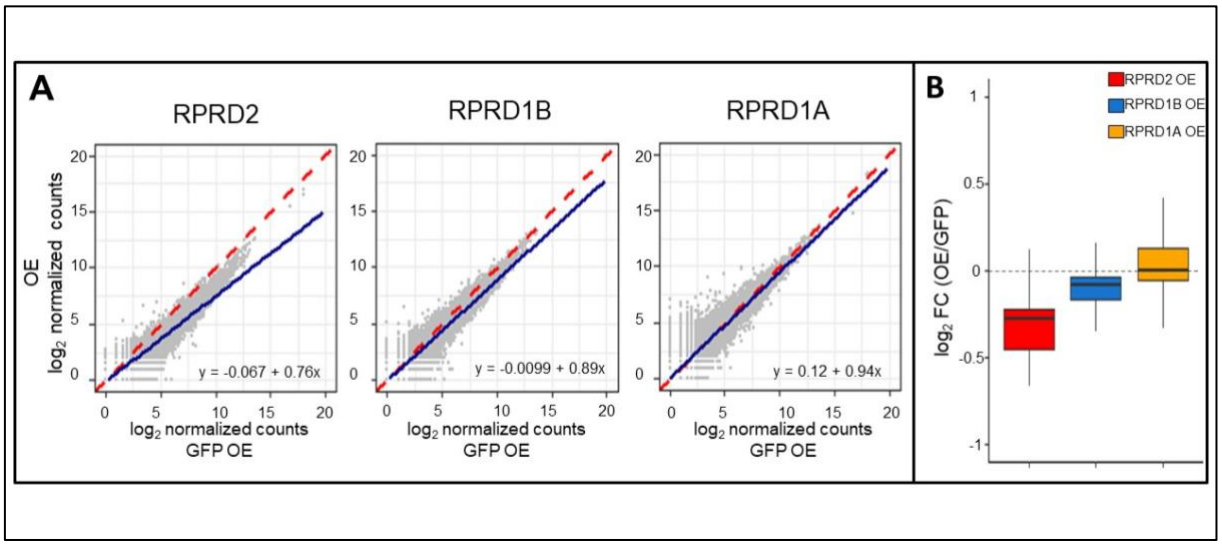


Figure 24: The global effect of RPRD overexpression on nascent RNA transcriptome.

The scatter plots (A) illustrate log2 normalized gene counts, facilitating a comparative analysis between RPRD-upregulated cells (y-axis) and GFP samples (x-axis). The red dashed diagonal line represents the zero difference line, while the blue line signifies the regression line. The box plots (B) depict the log2 fold change values between RPRD-upregulated and GFP samples.

There is a small proportion of the mammalian genome that is transcribed into protein-coding mRNA, with the majority producing long noncoding RNAs (lncRNAs). A lncRNA is generally defined as being longer than 100–200 nucleotides, and has been shown to play a role in regulating chromosome architecture, recruiting chromatin remodelers, forming R-loops to regulate transcription, forming nuclear speckles, and regulating translation (Yao et al., 2019). I conducted an analysis to determine how RPRD proteins affected mRNA and lncRNA expression on a genome-wide scale (**Figure 25, 26**).

To identify the most affected mRNA and lncRNA genes, a filter was applied such that log2 fold change (FC) was greater than 0.5 for upregulated genes and less than -0.5 for downregulated genes. The ratio of upregulated to downregulated lncRNA genes was calculated and labeled on bars for each individual condition. Compared to control samples, I found that 1417 mRNA genes were highly expressed in RPRD2-depleted cells, while only 245 mRNA genes were downregulated (ratio: 5.8). Across all three knockdown conditions, the number of upregulated genes (both mRNA and lncRNA) was greater than the number of downregulated genes. However, the ratio between the upregulated and downregulated genes in RPRD1A-depleted cells, for both mRNA genes and lncRNAs, was smaller than the ratios observed for RPRD2 and RPRD1B (**Figure 25 A, B**). Taken together, the data suggests that depletion of RPRD2 and RPRD1B results in increased RNA synthesis, whereas knockdown of RPRD1A has an uncertain impact on transcription.

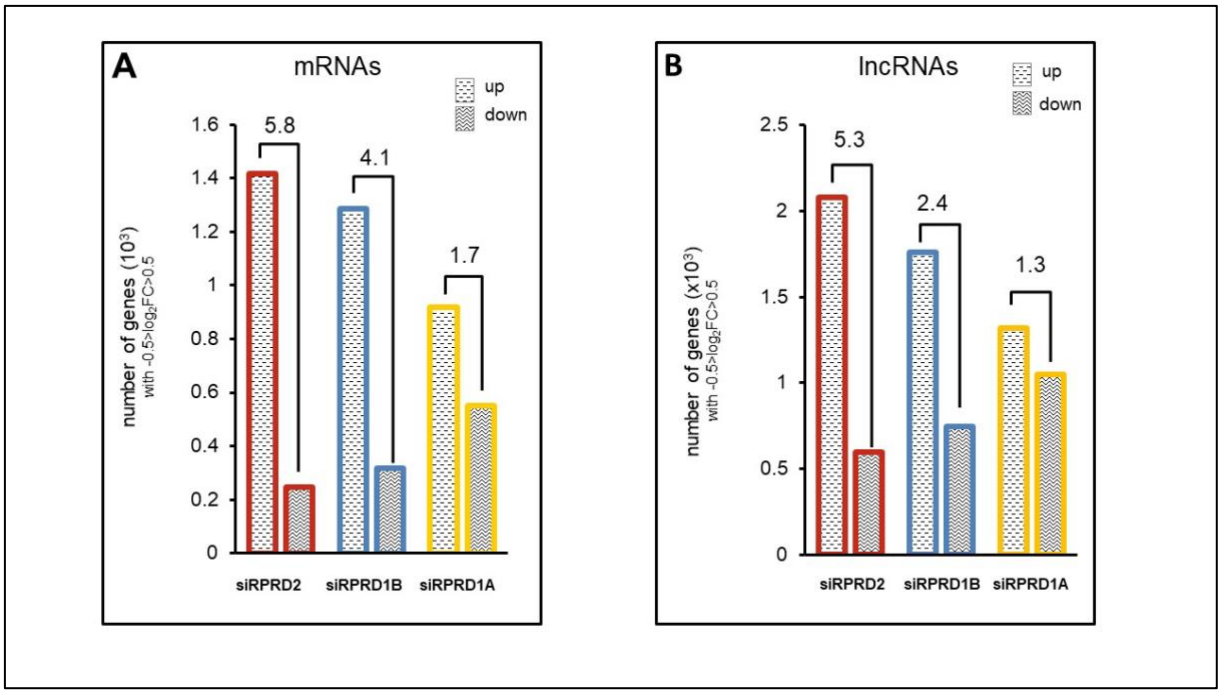


Figure 25: Changes at the number of mostly effected mRNA (A) and lncRNA (B) genes upon depletion of RPRD proteins.

The bar graphs illustrate the quantity (in units of x10^3) of mRNA (A) and lncRNA (B) genes that exhibit upregulation (indicated by a horizontal stripe pattern) and downregulation (denoted by a zigzag pattern) following RPRD depletion. In this representation, bars outlined in red correspond to RPRD2 overexpression, those outlined in blue signify RPRD1B, and bars outlined in yellow indicate RPRD1A overexpression conditions.

I also investigated the effect of RPRD overexpression on protein-coding genes (mRNA) and lncRNAs, to determine whether they were affected positively or negatively. The upregulated and downregulated mRNA and lncRNA genes was compared in RPRD overexpressing cells. I observed that the number of upregulated mRNA and lncRNA genes was lower in cells overexpressing RPRD2 (up/down ratio: 0.3) and RPRD1B (up/down ratio: 0.6) compared to downregulated mRNA and lncRNA genes. However, RPRD1A overexpression resulted in an increase in the number of mostly affected mRNA and lncRNA genes compared to GFP samples (**Figure 26 A,B**).

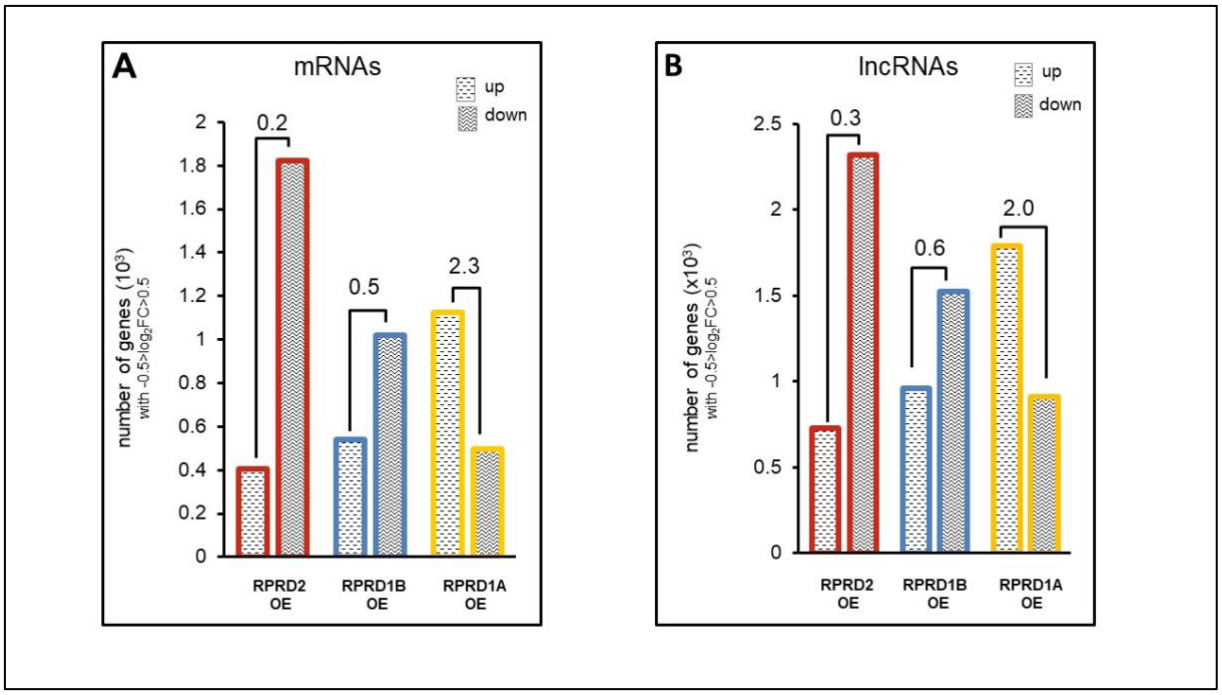


Figure 26: Changes at the number of mostly effected mRNA (A) and lncRNA (B) genes upon overexpression of RPRD proteins.

The bar graphs illustrate the quantity (in units of x10^3) of mRNA (A) and lncRNA (B) genes that exhibit upregulation (indicated by a horizontal stripe pattern) and downregulation (denoted by a zigzag pattern) following RPRD overexpression. In this representation, bars outlined in red correspond to RPRD2 overexpression, those outlined in blue signify RPRD1B, and bars outlined in yellow indicate RPRD1A overexpression conditions.

The single gene examples were displayed in Integrated Genome Viewer (IGV) for KPNB1, MYC, and RNA18SN1 genes (**Figure 27**). The thick lines represent exons, and the thin lines represent introns on the gene representation. There was almost equal coverage of exons and introns as these nascent RNAs were not completely processed. For the MYC and KPNB1 genes, RPRD depletion led to the accumulation of nascent RNA for all three RPRD proteins, while overexpression of them decreased nascent RNA levels. RNA18SN1 gene, which belongs to the rRNA gene family transcribed by RNA polymerase I, was not affected by either RPRD downregulation or

RPRD upregulation. This indicates that RPRD proteins have an effect on RNAPII transcription. Individual RPRD2 and RPRD1B overexpression had a stronger effect in terms of depletion of nascent RNA than RPRD1A overexpression for both genes.

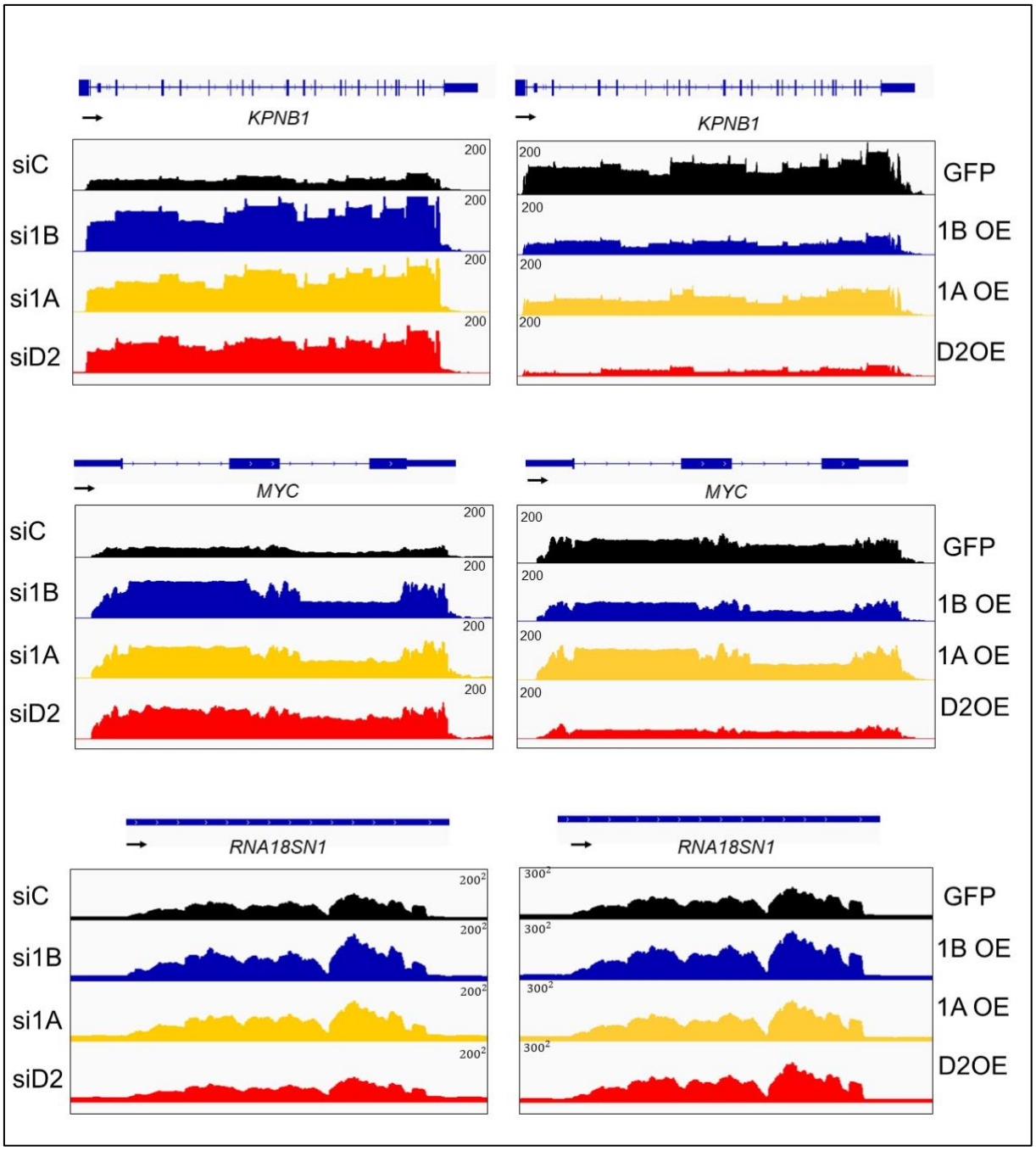


Figure 27: Visualization of MYC, KPNB1 and RNA18SN1 genes for RPRD depletion and overexpression on IGV.

Visualization of MYC and KPNB1 genes expressed by RNAPII, as well as the RNAPI-expressed RNA18SN1 gene, within the sequencing dataset derived from RPRD-deregulated 4sU samples using the IGV.

1.6 The elongation rate of RNA polymerase II is decreased by RPRD proteins

The observed alterations in the abundance of newly synthesized RNA, resulting from changes in the expression of RPRD proteins, prompted us to hypothesize that these variations might be attributed to modifications in transcription rates. To investigate this hypothesis, we employed 5,6-dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) for transcription synchronization, a method previously detailed (Saponaro et al., 2014). DRB functions by impeding the progression of RNAPII into the elongation phase of transcription, achieved through the inhibition of p-TEFb-mediated Ser2 phosphorylation within the CTD of RNAPII. Following synchronization of the transcription cycle with DRB at the initiation step, transcription was resumed by removing DRB from the cell medium. After releasing RNAPII form initiation, I collected samples at 10-minute intervals over a span of two hours. Subsequently, I compared the RNA levels obtained after DRB removal with those representing steady-state RNAs in the absence of DRB treatment. A 153-kilobase pair (kb) genomic distance exists between the exon-intron junctions encompassing the first and 20th exons of the KIFAP3 gene. This difference allowed comparative assessment of newly synthesized RNA levels between siCONTROL and RPRD-deregulated samples. The aim of this analytical approach was to probe the elongation rate of RNAPII through the synchronization of the transcription process at its initiation step, followed by the controlled resumption of transcriptional activity.

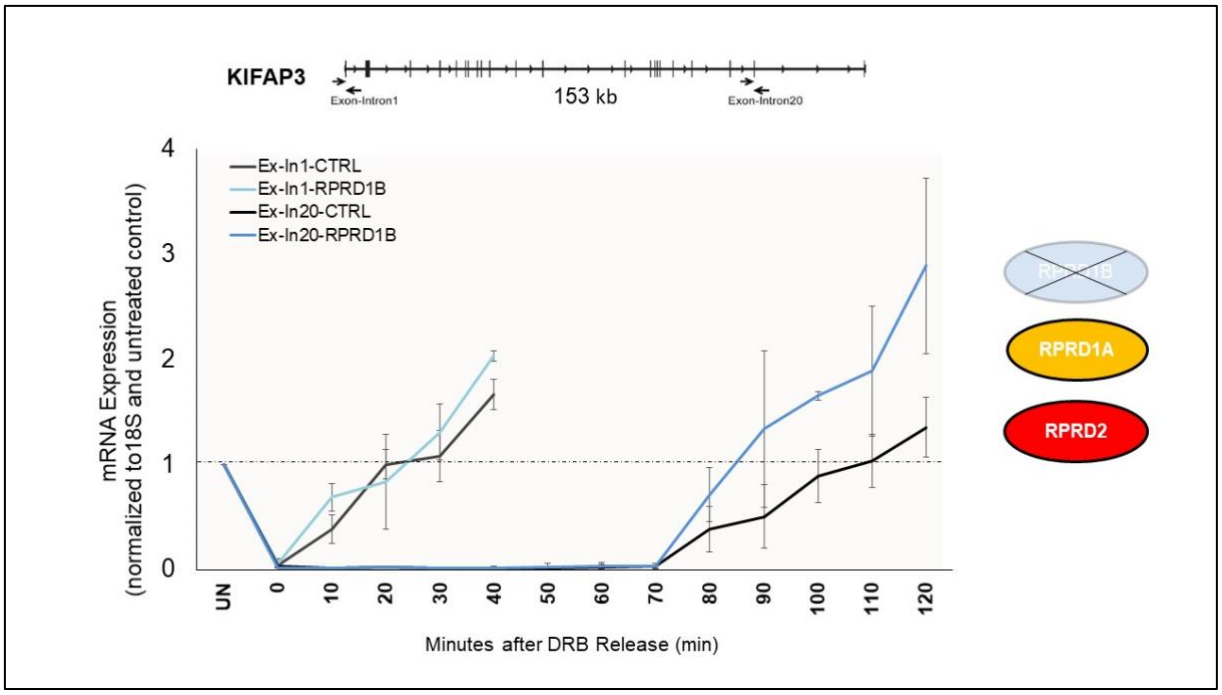


Figure 28: Comparison nascent mRNA production between control and RPRD1B knockdown cells of the KIFAP3 gene after DRB release.

Analysis of nascent mRNA synthesis within distinct segments of the KIFAP3 gene following release from DRB inhibition. An illustration of the KIFAP3 gene's introns and exons and arrows indicated the place of primers. Graph denotes the relative mRNA levels of ex-int1 and ex-int20 at each time point in control (grey / black) and RPRD1B (light blue/ dark blue) depleted samples. The dashed line across the value 1 indicates the normalized steady-state mRNA in untreated sample. The error bars depict the standard deviation, which has been calculated based on data obtained from three independent experiments (n=3).

I conducted an investigation on the long gene KIFAP3, which has been previously reported in literature (Saponaro et al., 2014; Singh & Padgett, 2009). I used the exon-intron junction regions of the KIFAP3 gene, which were separated by a distance of 153 kb (**Figure 28, 29, and 30**). The experiment involved seeding cells onto 14 different plates for each knockdown condition simultaneously. The untreated samples (UN) represented non-DRB treated samples for knockdowns. Initially, Ct values for each time point were normalized with 18s rRNA. Subsequently, the RPRD knockdown was

compared to the level of untreated samples. The line indicating the value 1 represents the normalized steady-state mRNA accumulation level in untreated samples. The time point at which the mRNA level crosses the 1 level for the first time indicates that the released RNAPII has reached steady-state levels of transcription for individual KD samples.

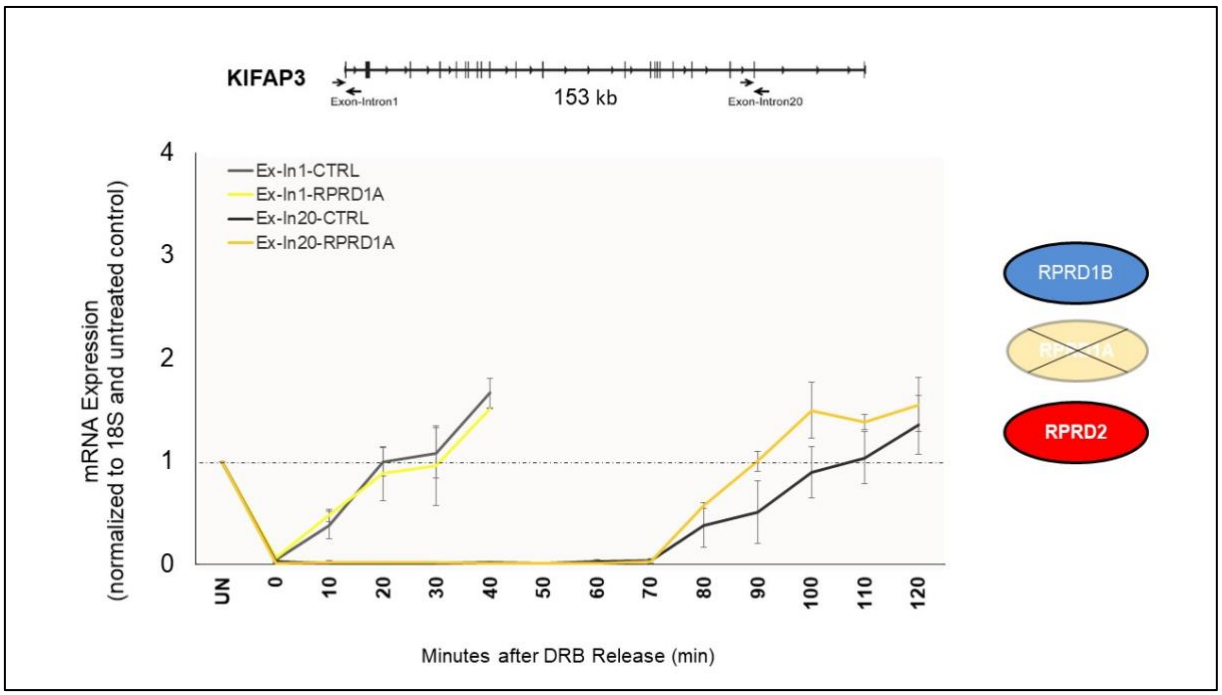


Figure 29: Comparison of nascent mRNA production between control and RPRD1A knockdown cells of the KIFAP3 gene after DRB release.

Analysis of nascent mRNA synthesis within distinct segments of the KIFAP3 gene following release from DRB inhibition. An illustration of the KIFAP3 gene's introns and exons and arrows indicated the place of primers. Graph denotes the relative mRNA levels of ex-int1 and ex-int20 at each time point in control (grey / black) and RPRD1A (light yellow/ dark yellow) depleted samples. The dashed line across the value 1 indicates the normalized steady-state mRNA in untreated sample. The error bars depict the standard deviation, which has been calculated based on data obtained from three independent experiments (n=3).

In RPRD1B knockdown samples, the mRNA levels reached steady state earlier than those in the control cells for both amplicons **(Figure 28**). For the first amplicon (exon-intron 1), RNAPII arrived approximately 5 minutes earlier in RPRD1B knockdown samples than in control samples. The difference was more pronounced for the second amplicon, where RNAPII arrived approximately 20 minutes earlier in RPRD1B knockdown samples.

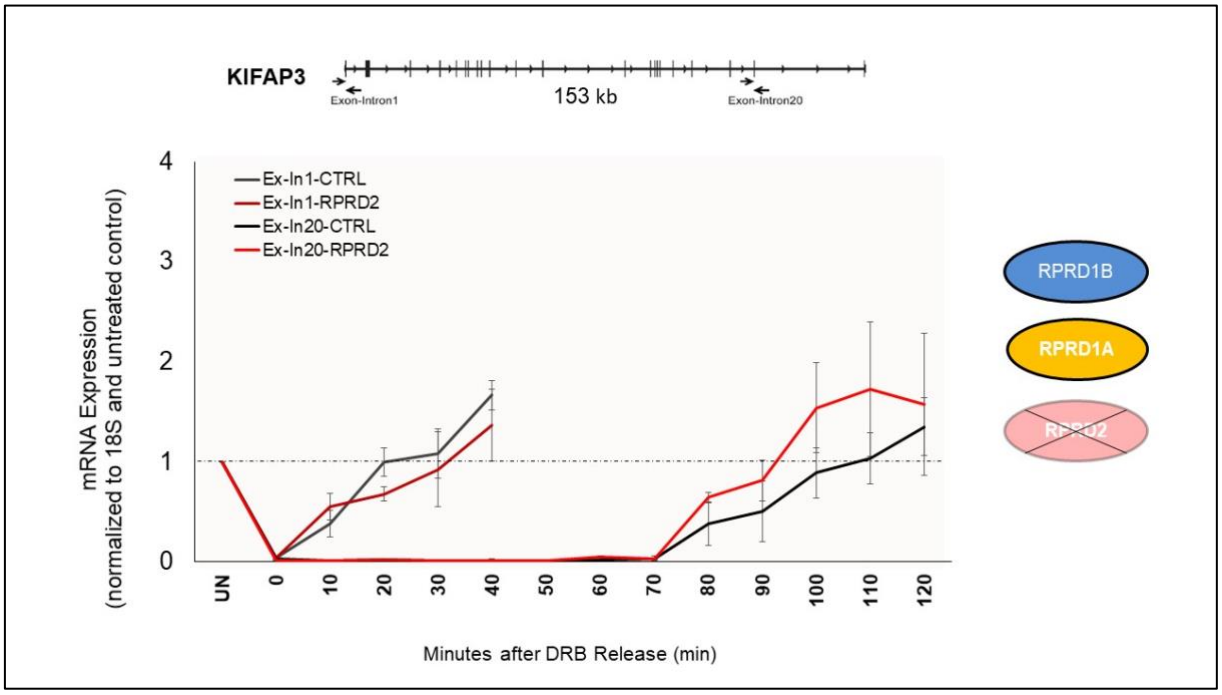


Figure 30: Comparison of nascent mRNA production between control and RPRD2 knockdown cells of the KIFAP3 gene after DRB release.

Analysis of nascent mRNA synthesis within distinct segments of the KIFAP3 gene following release from DRB inhibition. An illustration of the KIFAP3 gene's introns and exons and arrows indicated the place of primers. Graph denotes the relative mRNA levels of ex-int1 and ex-int20 at each time point in control (grey / black) and RPRD2 (light red/ dark red) depleted samples. The dashed line across the value 1 indicates the normalized steady-state mRNA in untreated sample. The error bars depict the standard deviation, which has been calculated based on data obtained from three independent experiments (n=3).

In the case of RPRD1A knockdown samples, mRNA transcription was faster than in the control cells for both amplicons (**Figure 29**). Specifically, RNAPII in the RPRD1A KD sample reached the first amplicon 5 minutes earlier than in the control sample and reached the second amplicon 15 minutes earlier than in the control sample.

The same effect was observed in RPRD2 knockdown samples, wherein RNAPII reached the related amplicon faster by 5 minutes for intron-exon 1 and 15 minutes for intron-exon 20 (**Figure 30**).

Elongation rates were subsequently calculated using the following formulation.

$$\frac{\text{distance between two amplicon(kb)}}{\text{the difference of time points (second amplicon} - \text{first amplicon)}}$$
$$\text{where mRNA expression reachs to 1}$$

Based on the calculation, the transcription rate increased by approximately 30% for all three protein knockdowns individually. This indicates that the presence of RPRD proteins slows down transcription (**Figure 31**).

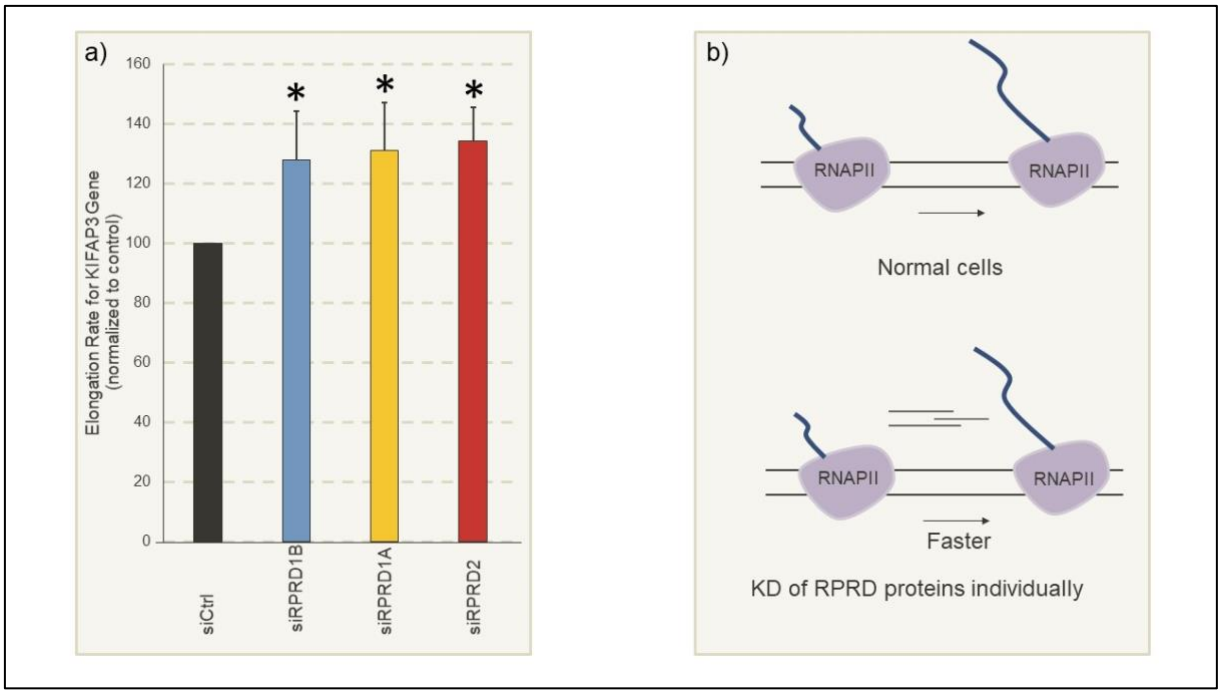


Figure 31: Elongation rate comparison for KIFAP3 gene of RPRD protein knockdown.

The determined elongation rates for both the control and RPRD-depleted cells are depicted in bar graphs (a). The error bars on the graphs represent the standard deviation calculated from the results of three independent experiments (n=3). Significance levels were assessed using a t-test (*p-value < 0.05). Faster transcription was illustared in RPRD depleted cells (b).

The initial amplicon arrival time indicates that the RNAPII was released from promoter-proximal pausing earlier. However, to achieve more definitive results, the experiment's resolution could be improved by decreasing the time intervals between sample collections. RPRD1B and RPRD2 knockdown results in an accumulation of newly synthesized RNA by enhancing transcription rate , as confirmed by 4sU labeling and DRB stop-chase experiments. These findings suggest that the inhibitory impact of all three RPRD proteins accumulates throughout prolonged transcription.

.

1.7 Evaluation of the newly synthesized RNA stability

My findings demonstrate a significant increase in newly synthesized RNA accumulation in cells depleted of RPRD1B and RPRD2, with a concurrent decrease in nascent RNA levels in cells overexpressing these proteins. To gain further insight, I conducted DRB stop-chase experiments to examine the speed of RNAPII. My results showed that the depletion of RPRD proteins accelerates RNAPII speed. Interestingly, I did not observe any changes in the total RNA levels after RPRD1B and RPRD2 depletion or overexpression, despite the high levels of nascent RNA accumulation and speedy RNAPII. This led us to hypothesize that the stability of nascent RNA might have been altered by the deregulation of RPRD proteins. To test this hypothesis, I treated cells with DRB for an extended period to measure pre-existing RNA degradation.

Notably, I observed the most significant effect on nascent RNA levels after RPRD2 depletion or overexpression, prompting us to focus solely on RPRD2-depleted and RPRD2-overexpressing cells. To conduct this experiment, I treated cells with DRB for 10 hours to halt transcription and collected cells every two hours to perform q-PCR following total RNA isolation. The exonic region primer enabled me to measure the total RNA level at each time point, while the intron-exon junction region allowed us to measure the level of newly synthesized RNA. All data were normalized to 18S rRNA.

Comparing the results with non-specific siRNA knockdown (siCTRL) cells, I observed a clear difference in the exonic RNA level of the KPNB1 gene after four hours of RPRD2 knockdown (**Figure 32**). Specifically, RPRD2-depleted cells exhibited relatively less total RNA compared to control cells, while there was more nascent RNA synthesis in this condition. I also performed this experiment with RPRD2-

overexpressing cells, which showed less nascent RNA synthesis according to the 4sU labeling experiment. Clear differences in total RNA level were observed in the four-hour samples **(Figure 32**). These results suggest that cells attempt to compensate for changes in nascent RNA synthesis by modulating the stability of total RNA. Specifically, we observed that reduced nascent RNA synthesis was associated with an enhancement in the stability of total RNA. Conversely, increased nascent RNA accumulation corresponded to a reduction in RNA stability.

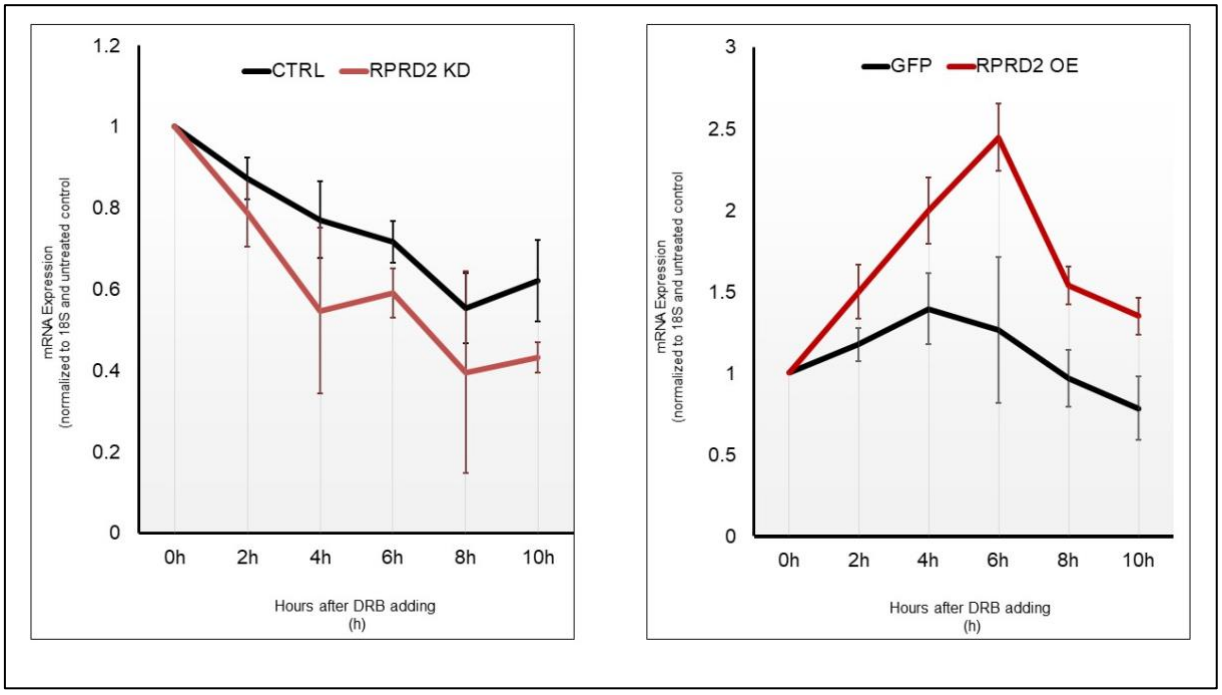


Figure 32: q-PCR analysis after prolong DRB treatment in siCTRL vs siRPRD2 and GFP vs RPRD2 OE.

Graphs show the relative mRNA level of exonic region (exon3) of KIFAP3 gene which is normalize to 18s and untreated control samples. At first graph shows CTRL (black) and RPRD2 KD (red) samples. Graph at right demonstrates GFP (black) and RPRD2 OE (red). Average values of a minimum of three independent experiments are plotted. The error bars represented standard deviation of three independent experiments.

1.8 RPRD proteins have an effect on formation of R-Loops

Changes in transcription rate can affect the accumulation of R-loops in cells, either by increasing or decreasing their levels. By conducting 4sU labeling and DRB elongation rate analyses, I found altered levels of nascent RNA and transcription elongation rates in cells with deregulated RPRD proteins. Based on these findings, I evaluated the levels of R-loops in conditions of RPRD protein deregulation to investigate the impact of RPRDs on R-loop formation.

It has been observed that the absence of the RPRD1B protein causes the accumulation of R-loop foci in nuclei, which can be detected using the S9.6 antibody and immuno-fluorescence (IF) (Morales et al., 2014). Although the S9.6 antibody is widely used to detect RNA:DNA hybrids both *in vitro* and *in situ*, the antibody is not specific for IF due to indiscriminate binding to cellular RNAs. Since this non-specific binding cannot be controlled in cells, it renders the S9.6 antibody-based immunofluorescence unreliable (Smolka et al., 2021). Therefore, I investigated the impact of RPRD expression on R-loop formation by performing q-DRIP experiment. My findings, obtained through DRIP analysis, represent the first evidence of a correlation between R-loop formation and RPRD proteins.

1.8.1 Validation of spike-in R-loops

DRIP is a method used to isolate and identify R-loops in cells. The use of qDRIP enables a more sensitive detection of R-Loops, using sythetic RNA:DNA hybrid makes the approach a more reliable for comparing R-loops levels across different biological conditions (Crossley et al., 2020). Briefly, I synthesized the R-loops in vitro by isolating genomic DNA from *E. coli*, and DNA fragments were synthesized using a T7 promoter-containing primer. The RNA moiety of R-loops was produced by in vitro transcription using a T7 transcription kit. The annealing step between RNA and DNA was achieved through a denaturation step at 95◦C, followed by slow cooling until the temperature reached 25◦C. I synthesized two different R-loops with varying amounts of GC content (L286 - low, H281 - high). The synthesized R-loops were run on a 0.9% agarose gel to observe the shift in band size after hybridization.

The size-shifting was clearly visible after the formation of L286 and H281 hybrids (red triangle) (**Figure 33**). A stronger signal was observed from the H281 R-loop band than from the L286 band, as the RNA:DNA hybrid for L286 was relatively less stable due to the low amount of GC content. To determine their specificity, synthetic R-loops were treated with RNaseH, a ribonuclease that recognizes R-loops specifically and dissolves hybrid formation. The second gel image in **Figure 33** represents the RNaseH-treated spike-in R-loops. The first two lines indicate non-treated synthetic R-loops, the second two lines represent RNaseH-treated R-loops, and the last two lines were indicative of the setting of RNaseH treatment reaction without enzyme addition to control the reaction conditions for the destruction of synthetic R-loops. After RNaseH treatment, the band representing R-loops disappeared; however, the R-loops band

was still present in the non-enzyme treatment conditions. This indicated that the band I observed after hybridization was from in vitro synthesized R-loops.

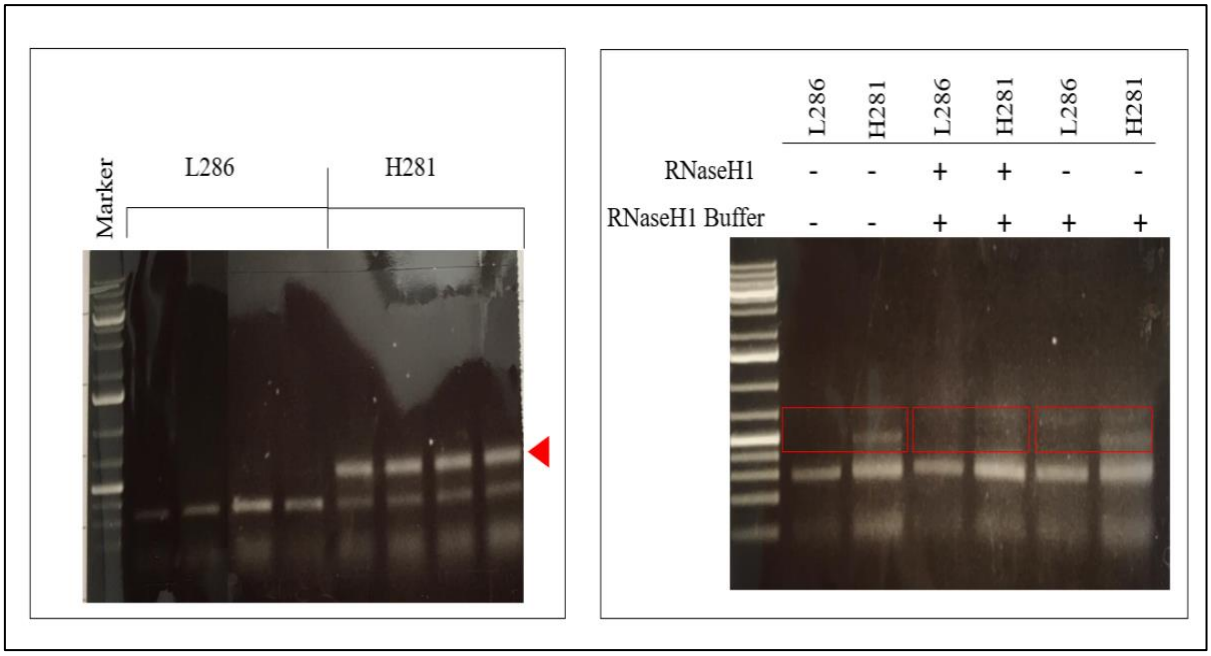


Figure 33: Determination size and location of synthetic R-loops on agarose gel. Samples loaded at 1% agarose gel and run under electric field at 50V for 30 minutes. GeneRuler 1 kb Plus DNA ladder (Thermo Sci) was used as a marker. Red triangle at first image (left) represented the size-shift after RNA-DNA hybridization. The red squares at second image indicated the probable place of R-loops on the gel.

After determining the bands on the gel, I excised the R-loop bands and used the freeze and squeeze method to isolate the R-loops from the gel slice. Next, I analyzed the synthetic R-loops using immunoprecipitation (DRIP) with a specific S9.6 antibody that targets R-loops. Briefly, I incubated the synthetic R-loops with the antibody overnight and then eluted the antibody-bound R-loops using protein A and protein G magnetic beads. As a negative control, I treated the synthetic R-loops with RNaseH for 1 hour at 37°C just before the antibody incubation. The mock group represented the non-S9.6 antibody incubation.

As demonstrated in bar graph, I observed significant enrichment over the mock group for both synthetic R-loops incubated with S9.6 antibody; however, the enrichment for H281 was higher than that for L286 (**Figure 34)**. This result suggested that the experimental conditions probably dissociated the R-loops with low GC content.

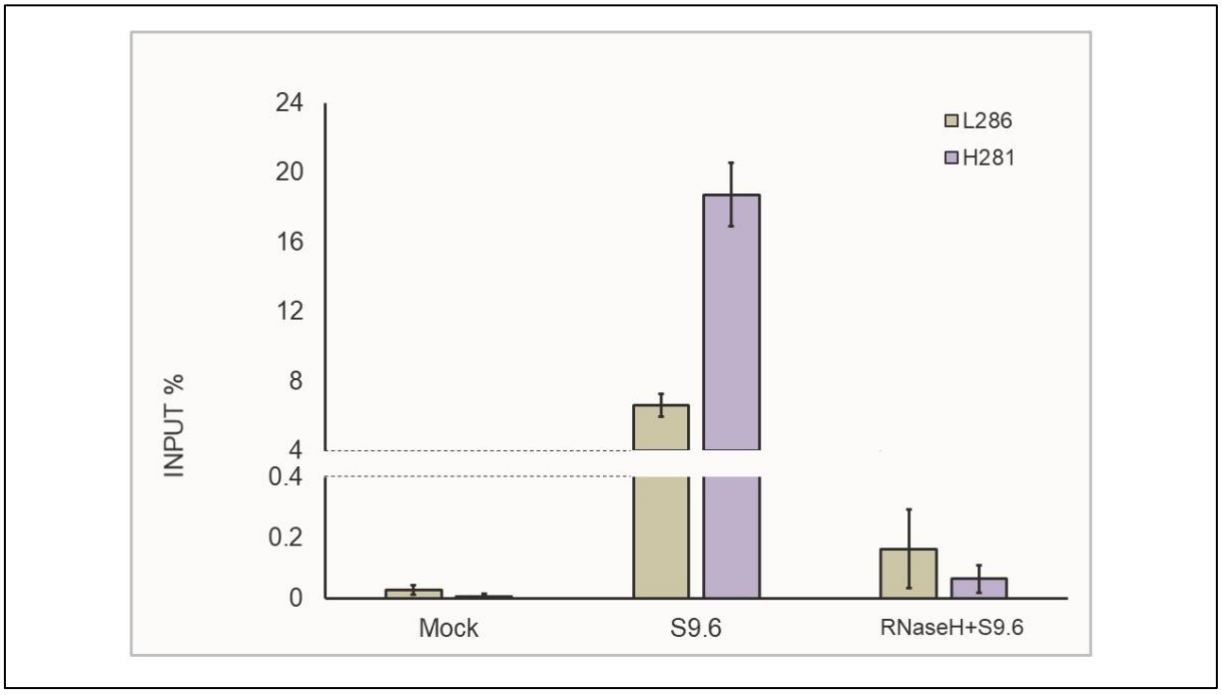


Figure 34: Validation of R-loop spike-in with DRIP experiment.
The graph illustrates the input(%) of the DRIP with the synthetic R-loops, with three distinct conditions depicted on the x-axis. The 'Mock' condition, serving as the negative control, represents conditions without antibody incubation. The 'S9.6' demonstrates S9.6 antibody incubated conditions, while 'RhNaseH+S9.6' shows theRnaseh treatment following S9.6 antibody incubation and is employed to validate the specificity of S9.6 antibody binding. Each experiment was conducted on a minimum of three independent experiments, and the error bars denote the standard deviations.

1.8.2 RRPD proteins caused the accumulation of R-loops.

To investigate the effect of RPRD proteins on R-loop formation, I employed the q-DRIP method, which utilizes synthetic R-loops to normalize the results. In this study, in vitro synthesized R-loops were mixed with an equal amount of isolated DNA from RPRD upregulated and downregulated cells to assess the effect of RPRDs on R-loops.

For DRIP experiment, I used four genes and six amplicons, (**Figure 35**). TFPT, RPL13A, and TRIM33 are genes that are frequently tested for R-loops (Sanz et.al., 2016, Garcia-Rubio et.al 2015). R-loops tend to accumulate at the 5' and 3' ends of genes to regulate transcription initiation and termination processes. Therefore, I tested the 5' ends of TFPT, TRIM33, and RPL13A genes, the 3' ends of TRIM33 and RPL13A genes, as well as the exonic region of the NEAT1 gene. Results were normalized by co-eluted synthetic R-loops.

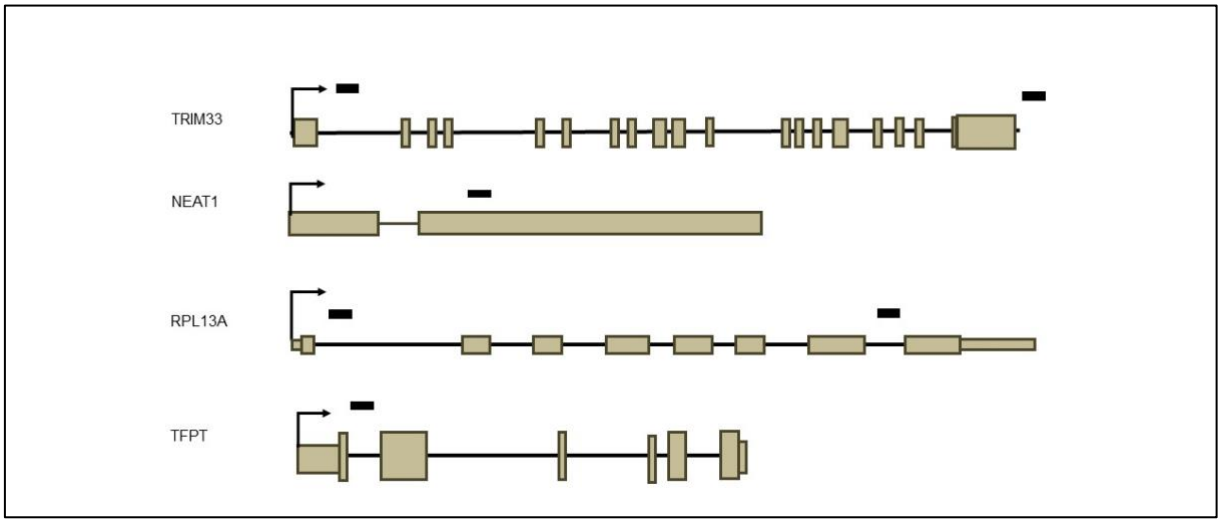


Figure 35: Illustration of the primers on the genes using for q-DRIP analysis.
A schematic representation of four genes subjected to the DRIP experiment is displayed.The exonic regions of the genes are delineated by yellow rectangles. Above this gene representation, the primer locations are depicted as black rectangles.

Following knockdown or overexpression of RPRD proteins, I performed DRIP experiments to assess the level of R-loops at selected gene regions, using synthetic R-loops for normalization. I compared and normalized the results with siCTRL and used mock samples as antibody controls. To test the significance of the differences, I employed a t-test (“#” p-value<0.005; “**” p-value<0.01; “*” p-value<0.05).

As shown, RPRD1B downregulation in HEK293T cells led to a significant decrease in R-loop levels for all tested regions (**Figure 36**). In contrast, RPRD1B overexpression in cells led to the accumulation of R-loops at the gene regions I tested.

The effect of RPRD1A knockdown and overexpression on R-loop formation or accumulation was depicted (**Figure 37**). My results indicate that RPRD1A had no effect on R-loop accumulation at the selected gene regions.

The relationship between RPRD2 levels and R-loop accumulation was shown **(Figure 38**). Knockdown of RPRD2 reduced R-loop levels by almost 50% compared to the control, while RPRD2 overexpression led to an expected increase in R-loop accumulation at the tested gene regions.

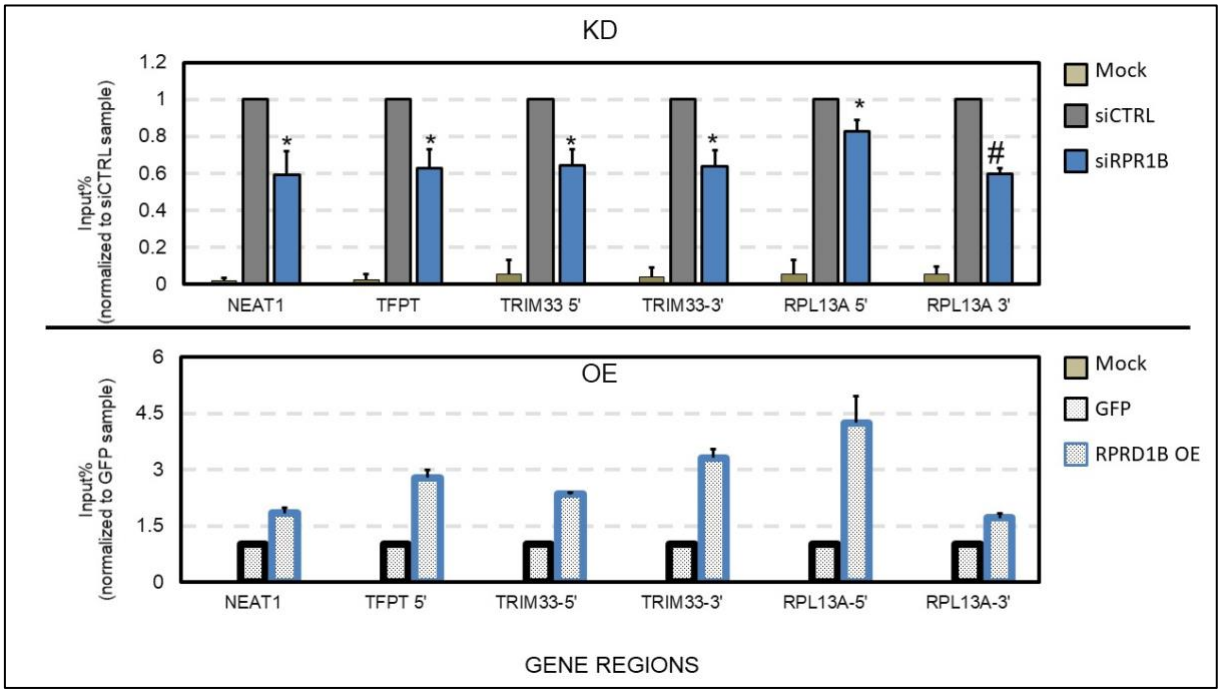


Figure 36: DRIP-qPCR analysis after RPRD1B knockdown and overexpression separately.

In the context of the DRIP-qPCR analysis, the figure shows the relative enrichment of RNA:DNA hybrids at both the 5' and 3' ends of four distinct genes. The bar graphs represent the comparisons of input (%) between the siCtrl vs. siRPRD1B conditions (above) and the GFP vs. RPRD1B OE conditions (below). Each bar within the graphs corresponds to the mean value obtained from a total of n = 3 independent biological replicates for each respective experimental condition. Statistical significance was rigorously assessed using a two-tailed Student's t-test, where statistical significance levels were indicated as follows: (# p-value < 0.005; "**" p-value < 0.01; * p-value < 0.05).

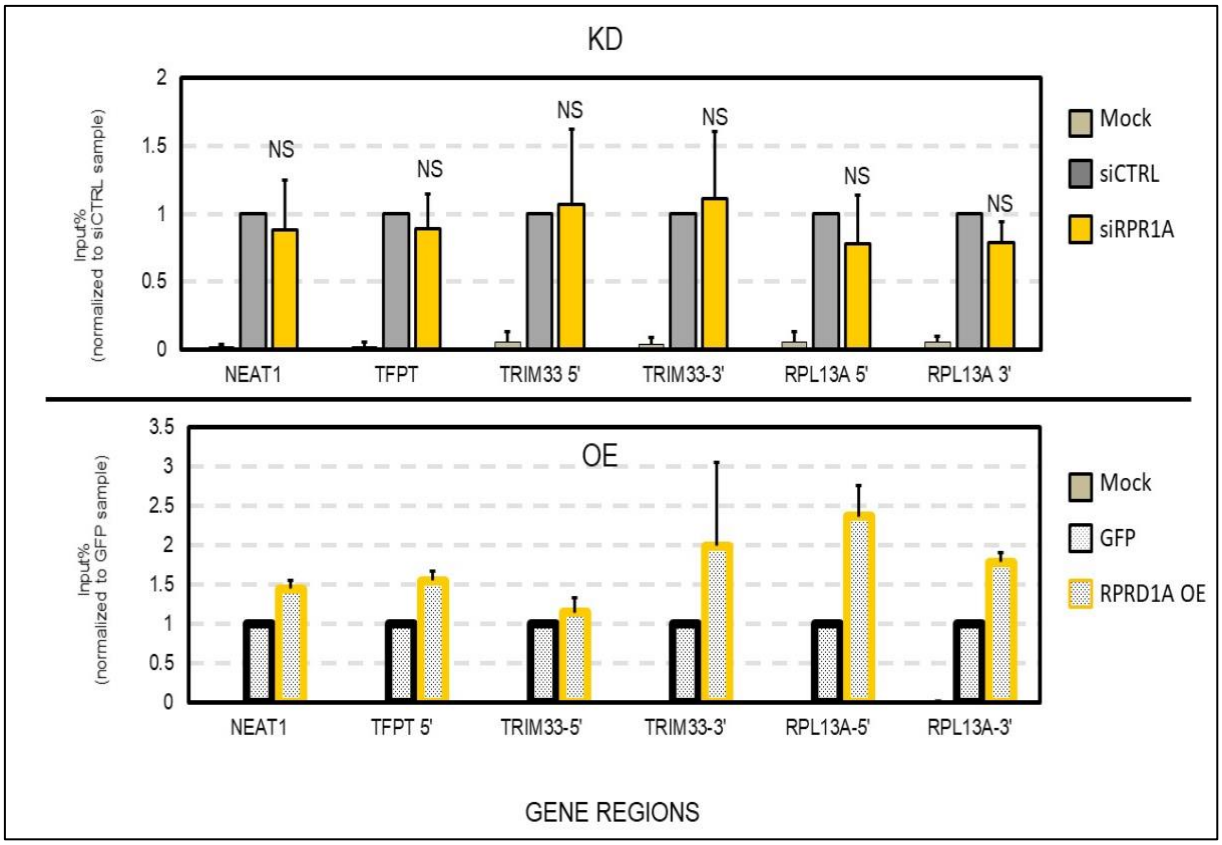


Figure 37: DRIP-qPCR analysis after RPRD1A knockdown and overexpression separately.

In the context of the DRIP-qPCR analysis, the figure shows the relative enrichment of RNA:DNA hybrids at both the 5' and 3' ends of four distinct genes. The bar graphs represent the comparisons of input (%) between the siCtrl vs. siRPRD1A conditions (above) and the GFP vs. RPRD1A OE conditions (below). Each bar within the graphs corresponds to the mean value obtained from a total of n = 3 independent biological replicates for each respective experimental condition. Statistical significance was rigorously assessed using a two-tailed Student's t-test, where statistical significance levels were indicated as follows: (# p-value < 0.005; "**" p-value < 0.01; * p-value < 0.05; NS not significant).

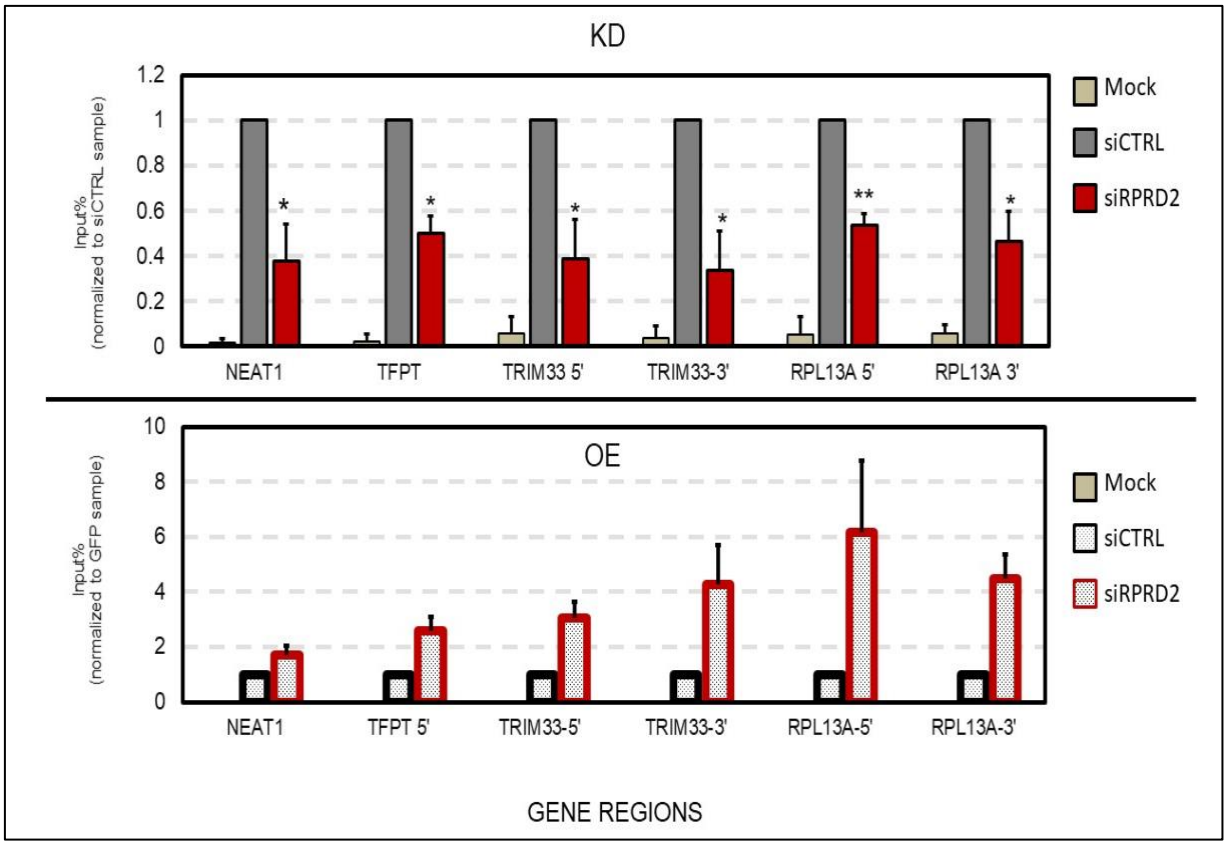


Figure 38: DRIP-qPCR analysis after RPRD2 knockdown and overexpression separately.

In the context of the DRIP-qPCR analysis, the figure shows the relative enrichment of RNA:DNA hybrids at both the 5' and 3' ends of four distinct genes. The bar graphs represent the comparisons of input (%) between the siCtrl vs. siRPRD2 conditions (above) and the GFP vs. RPRD2 OE conditions (below). Each bar within the graphs corresponds to the mean value obtained from a total of n = 3 independent biological replicates for each respective experimental condition. Statistical significance was rigorously assessed using a two-tailed Student's t-test, where statistical significance levels were indicated as follows: (# p-value < 0.005; "**" p-value < 0.01; * p-value < 0.05; NS not significant).

I observed that an excessive amount of RPRD1B and RPRD2 stimulated the formation of R-loops, while their depletions led to a decreased level of R-loops. On the other hand, neither the depletion nor overexpression of RPRD1A affected the accumulation of R-loops. This is likely because RPRD1B and RPRD2 interact with other transcription factors to slow down transcription elongation, giving newly-synthesized RNA sufficient time to hybridize with DNA.

1.8 CONCLUSIONS

The aim of this part was to investigate the impact of RPRD1A, RPRD1B, and RPRD2 proteins on transcription. To achieve this, I conducted a ChIP experiment to observe the binding pattern of RPRD proteins throughout the MYC gene. My results demonstrate that RPRD1B and RPRD2 exhibit high levels of enrichment at the 5' and 3' ends of the gene, while RPRD1A also shows enrichment in these regions, albeit to a lesser extent. This discrepancy could be attributed to the antibody efficiency used during the ChIP analysis. After getting their binding pattern I tried to determine the exact role of RPRDs on transcription, firstly I start to analyse total RNA. however I did not get significant changes at steady state level under RPRD depletion. To investigate changes in nascent RNA levels in RPRD deregulated cells, I conducted a 4sU labeling experiment for 5 minutes to isolate nascent RNAs. The 4sU/q-PCR results revealed an accumulation of nascent RNA in RPRD depleted cells, while less newly transcribed RNA was observed in RPRD overexpressed cells. The most significant effect in the 4sU labeling experiment was observed in RPRD2 and RPRD1B deregulated samples, whereas RPRD1A had a moderate effect on nascent RNA.

To get the genome-wide results. I sequenced these samples to observe the genome-wide impact of RPRD proteins on the nascent transcriptome. My results indicated that the absence of RPRD2 and RPRD1B led to an accumulation of nascent RNAs, whereas RPRD1A depletion had a moderate effect on nascent RNA transcription. Overexpression of RPRD2 and RPRD1B led to a decrease in nascent RNA, but RPRD1A did not exhibit this effect. RPRD1B and RPRD2 proteins negatively impacted newly synthesized RNA. These findings led us to hypothesize that changes in nascent RNA levels were a result of alterations in transcription rate. To test this hypothesis, I

conducted a DRB stop-chase experiment using a long gene (KIFAP3) to measure transcription rate. By releasing transcription after DRB inhibition, I determined the time point at which RNAPII reached the last exon. My results demonstrated that the absence of RPRD proteins accelerated the rate of RNAPII, supporting the hypothesis that RPRDs are negative regulators of RNAPII.

After deregulation of RPRD, I noticed alterations in nascent transcription levels, however, there were no significant changes detected in the levels of steady-state RNA. To investigate the possible effects on mRNA stability, I conducted a long DRB treatment analysis using RPRD2-deregulated cells, which have a greater impact on nascent transcripts. The results revealed decreased stability of nascent RNA synthesized and accumulated under RPRD2 depletion, which did not accumulate at the total RNA level.

Changes in transcription rates could also affect R-loop levels in cells expressing different RPRD proteins. I checked the levels of R-loops in cells expressing RPRD proteins at varying levels and found that RPRD1B and RPRD2 had a positive effect on R-loop formation, but not RPRD1A.

To sum up, the findings suggest that the depletion of RPRD1B and RPRD2 leads to increased transcription rates and higher levels of nascent transcripts, while overexpression of these proteins leads to decrease nascent transcription. Additionally, I observed a high level of R-loops in RPRD-overexpressing cells, possibly due to the slower transcription rates allowing more time for nascent RNA to hybridize with DNA. Increased R-loop induce DNA damage and carcinogenesis.

Chapter 2: Effect of RPRD protein on cellular level

2.1 RRPRD proteins have effect on cancer formation

To investigate molecular aberrations in DNA, RNA, protein, and epigenetic levels of human tumors, the Cancer Genome Atlas (TCGA) Research Network has conducted extensive profiling and analysis of a large number of samples across multiple tumor lineages. The Pan-Cancer Atlas initiative reports that TCGA has profiled 33 types of tumors. The analysis of molecular aberrations and their functions across tumor types has enabled the development of effective therapies for other cancer types with similar genomic profiles (Weinstein et al., 2013).

In order to investigate the impact of RPRD protein on cancer progression, I conducted an analysis using brain lower grade glioma samples obtained from a cohort of 511 patients. This dataset is comprehensive and contains samples of RPRD that are differentially expressed. Lower grade gliomas are a form of brain cancer characterized by slow progression, and are derived from glial cells that provide important support functions for nerve cells in the brain. The aggressiveness of gliomas is graded on a scale of 1 to 4, based on both their histological appearance and genetic makeup, with grades 1 and 2 being considered "low grades". Unlike high-grade gliomas, low-grade gliomas tend to grow locally in the brain, causing symptoms as they expand and potentially disrupting the connections between nearby brain cells by exerting pressure on them. As a result, even small, slowly progressing gliomas can lead to serious brain problems, particularly if they are located in critical parts of the brain (Brat, 2015; Brennan et al., 2013).

I utilized the cBioPortal database, an open-access and open-source platform that enables interactive exploration of cancer genomics datasets, to analyze the cancer data obtained from glioma patients. The portal stores various data types, including non-synonymous mutations, DNA copy-number, mRNA and microRNA expression, protein-level and phosphoprotein level, DNA methylation, and de-identified clinical data. The dataset included mRNA expression data, protein level data obtained from reverse-phase protein array (RPPA), and patient survival rate data (Cerami et al., 2012; Gao et al., 2013).

A gene's relative expression in a tumor sample was typically determined by comparing the gene's expression distribution in a reference population which is all profiled samples. By calculating the mean and variance of all samples with expression values, the expression distribution of the gene was estimated. Using the Z-score method, the returned value indicates how far the expressions in the reference population are from the mean.To determine RPRD gene expression levels in tumor samples, I compared individual RPRD gene's expression distribution in a reference population of all profiled samples. All three RPRD mRNA expression distribution in glioma samples were shown separately (**Figure 39A**). The data exhibited a normal distribution pattern based on RPRD expression. To classify expression as either high or low, a Z-score threshold was established at -1 for low expression and +1 for high expression (**Figure 39B**), subsequently applied to facilitate the analysis of queried genes.

I selected that glioma dataset from the portal and filtered the data based on RPRD expression levels using mRNA expression data. Samples with RPRD expression less than -1 were classified as having low RPRD expression, while samples with RPRD expression greater than 1 were classified as having high RPRD expression. Samples

with Z-scores between -1 and 1 were deemed unaffected (**Figure 39B**).

In this study, a comparative analysis was conducted between glioma samples exhibiting low and high RPRD protein expression levels, facilitated by the generation of a concise hyperlink for streamlined data accessibility, thereby enabling a focused investigation into the differential expression patterns of RPRD proteins within the glioma samples.The analyses related to my queries can be accessed via the following links: *"https://bit.ly/3iL6B3N"* for RPRD1A, *"https://bit.ly/3XqmxqW*" for RPRD1B, and *"https://bit.ly/3Xj4qDM"* for RPRD2.

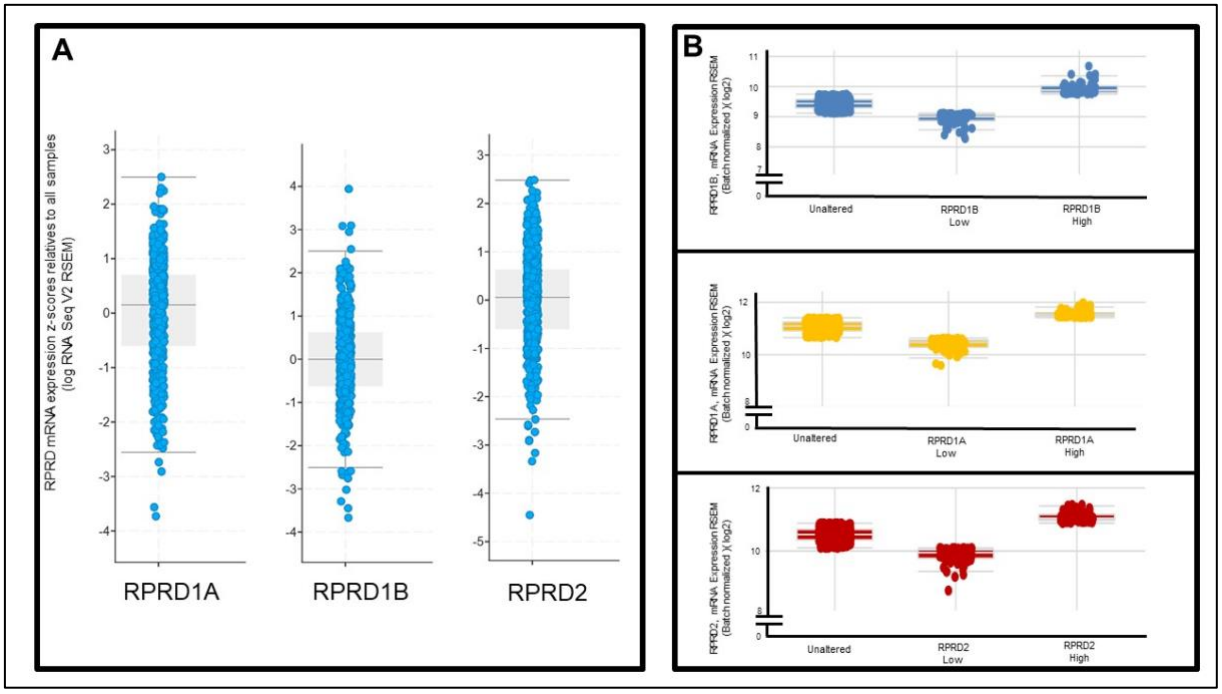


Figure 39: RPRD expression profile of the gliomas.

Graphs illustrating the mRNA expression z-scores relative to all samples of RPRD proteins within glioma samples (A).The plots at (B) shows the level of RPRD proteins after filtering according to their expression level.

My initial analysis focused on assessing the impact of RPRD expression on the survival rates of cancer patients.The number of patients that differentially expressed RPRD proteins and their overall survival months are shown (**Table 9**). The p-values of the survival probability comparing unaltered samples and RPRD downregulation or upregulation are represented under overall survival months values.

Table 9: The number of patients which are used for cancer analysis an their survival rates.

	RPRD EXP <-1		RPRD EXP>1		UNALTERED	
	Number of patients	Median months overall (p-values)	Number of patients	Median months overall (p-values)	Number of patients	Median months overall (p-values)
RPRD1A	91	52.11 (9.261e-5)	73	87.45 (0.828)	346	94.52
RPRD1B	74	146.14 (0.314)	71	50.86 (7.913e-4))	365	93.20
RPRD2	82	33.96 (1.108e-8)	70	75.02 (0.0958)	358	105.2

After filtering the samples based on the downregulation or upregulation of RPRD proteins it was observed that 91, 74, and 82 samples exhibited downregulation in RPRD1A, RPRD1B, and RPRD2, respectively, while 73, 71, and 70 samples displayed overexpression of RPRD1A, RPRD1B, and RPRD2, respectively. Based on the results, patients with deregulated RPRD1B proteins had longer survival times compared to patients with unaltered RPRD1B. On the other hand, patients with deregulated RPRD1A and RPRD2 had shorter survival times compared to patients

with unaltered RPRD1A and RPRD2. For patients with overexpressed RPRD1B and RPRD2, their survival time was found to be lower than that of unaltered patients.

Using the patient survival data, I conducted an investigation into the impact of RPRD protein expression levels on patient survival rates. Kaplan-Meier graphs illustrate the overall survival rate of patients with differential expression of RPRD proteins.(**Figure 40).** I observed a significant reduction in overall survival among patients with lower-grade gliomas expressing low levels of RPRD1A when compared to patients with lower-grade gliomas with unaffected expression of RPRD1A. However, patients suffering from lower-grade gliomas have nearly the same survival rates in RPRD1A overexpressing and unaffected tumors. The low-level expression of RPRD1B did not exert a substantial influence on the overall survival rate of low-grade glioma patients. Conversely, elevated expression of RPRD1B in lower-grade gliomas was associated with increased tumor aggressiveness and decreased overall survival rates among patients. According to the statistical analysis, the downregulation of RPRD2 is significantly associated with a negative impact on the survival rates of low-grade glioma cancer patients. In contrast, the upregulation of RPRD2 is also associated with a negative effect on patients' survival rates, but this association is not statistically significant (p-value: 0.0958) (**Figure 40**).

In summary, our findings provide substantial insights into the role of RPRD proteins in the context of low-grade glioma. Notably, the overexpression of RPRD1B emerges as a pivotal factor contributing to a significant reduction in overall patient survival probability within this patient cohort. Conversely, the downregulation of RPRD1A in glioma cells is associated with heightened cancer progression and diminished patient survival rates. These results are consistent with the literature.

Furthermore, our findings introduce a novel dimension the impact of RPRD2 downregulation on cancer aggressiveness. This downregulation of RPRD2 is correlated with a heightened aggressiveness of gliomas, resulting in earlier mortality among affected patients when compared to those with unaltered RPRD2 expression levels. Our study represents the initial exploration into the relationship between RPRD2 and cancer progression. These findings collectively contribute to a deeper understanding of the intricate roles of RPRD proteins in the context of low-grade glioma, shedding light on potential avenues for therapeutic interventions and further research

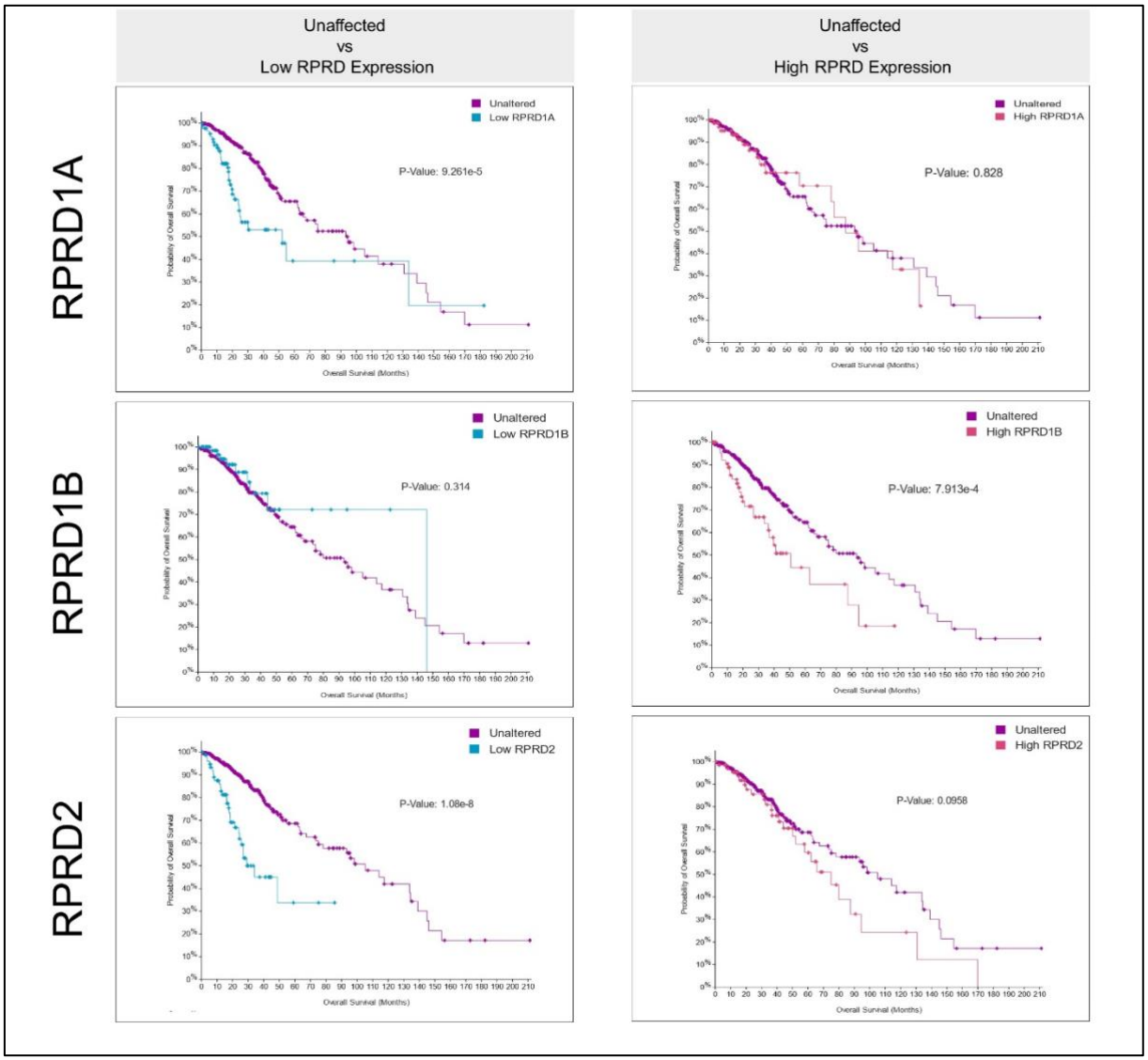


Figure 40: Overall survival rates of cancer patients having RPRD down regulated or up regulated gliomas.

Kaplan-Meier survival plots were generated for a cohort of 511 patients, categorized based on their RPRD protein expression levels. These conditions were compared with those of unaffected patients who exhibit normal RPRD protein expression levels. In the plots, the blue line represents patients with low levels of the respective RPRD expression, the pink lines depict patients with higher RPRD expression levels, and the blue line represents the unaffected patients with normal RPRD protein expression.p-values are calculated by log-ranktest and represented on each graphs.

2.2 Differential expression of RPRD protein in cancer cells affected cellular functions

In my investigation, I aimed to assess the differential gene expression patterns in glioma samples characterized by RPRD deregulation. Among the 19,000 tested genes, it was observed that in comparison to unaffected samples, RPRD1A downregulated gliomas exhibited an upregulation of 5,700 genes, while RPRD1B downregulated gliomas displayed an upregulation of 4,551 genes. Furthermore, RPRD2 downregulation was associated with a positive impact on the transcription of 6,316 genes in gliomas. Conversely, RPRD upregulation in gliomas affected relatively fewer genes, with 3328 genes for RPRD1A, 4739 genes for RPRD1B, and 3021 genes for RPRD2 being upregulated relative to unaffected samples **(Figure 41**). These findings underscore the intricate regulatory roles of RPRD genes in the context of glioma, with their dysregulation significantly influencing the transcriptional landscape of a considerable number of genes.

I observed that RPRD proteins interact with proteins involved in transcription. At glioma samples expressing high or low levels of RPRD, I compared the mRNA levels of POL2RG, POLR2D, POLR2M, RPAP2, and RECQL5, which are RPRD interacting proteins (**Figure 42, 43**). Alterations in RPRD levels in glioma samples affected the mRNA levels of RPRD interacting proteins. It appears that cells probably attempt to maintain homeostasis by adjusting the mRNA levels of proteins that interact with RPRD proteins in response to changes in RPRD protein levels. Interestingly, upregulation or downregulation of RPRD2 protein in glioma cells negatively or positively impacted the expression of all these interacting proteins. RPRD1A and RPRD2 downregulation in glioma cells correlated with an increase in mRNA levels of POLR2G and RECQL5, whereas there were no significant changes in these proteins' mRNA levels in either

RPRD1B high-expressed or RPRD1B low-expressed samples. Higher expression of RPRD1B in glioma cells correlated with higher expression of POLR2D and RPAP2.

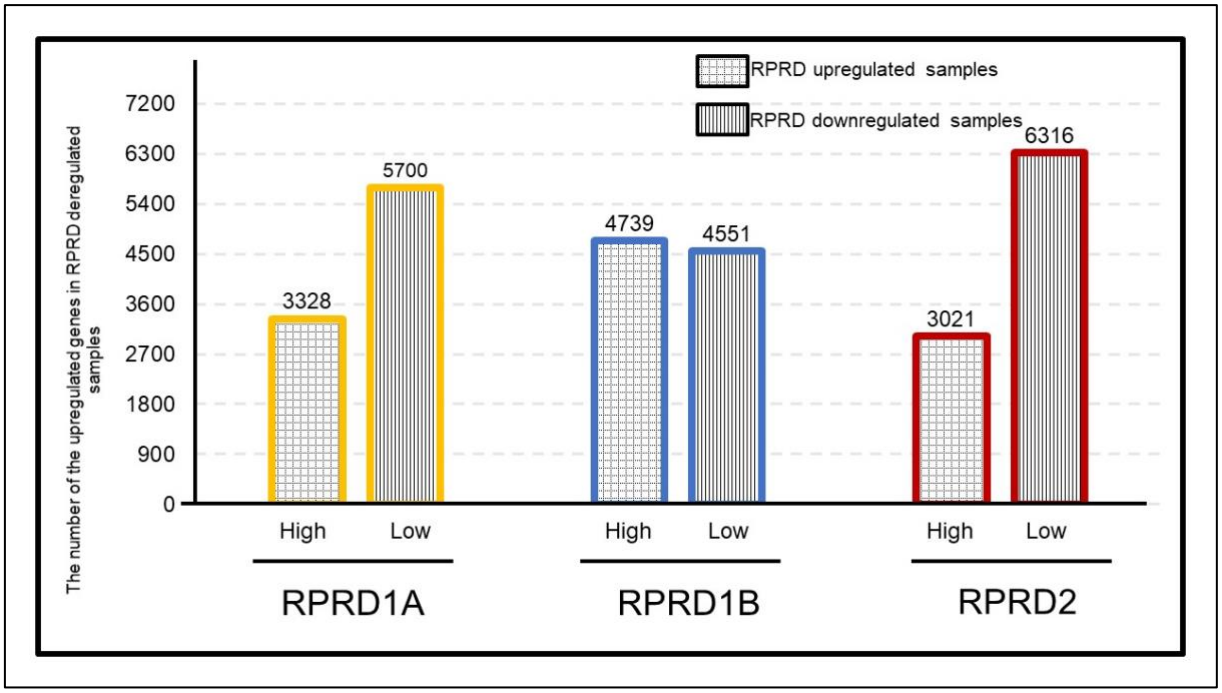


Figure 41: The effect of RPRD protein differential expression on numbers of upregulated gene in glioma cells.

The bar graphs illustrate the gene count of upregulated genes within glioma samples stratified into RPRD upregulated (High) and downregulated (Low) categories, in comparison to unaffected samples. The number of genes displaying alterations in each category is denoted on their respective bars.

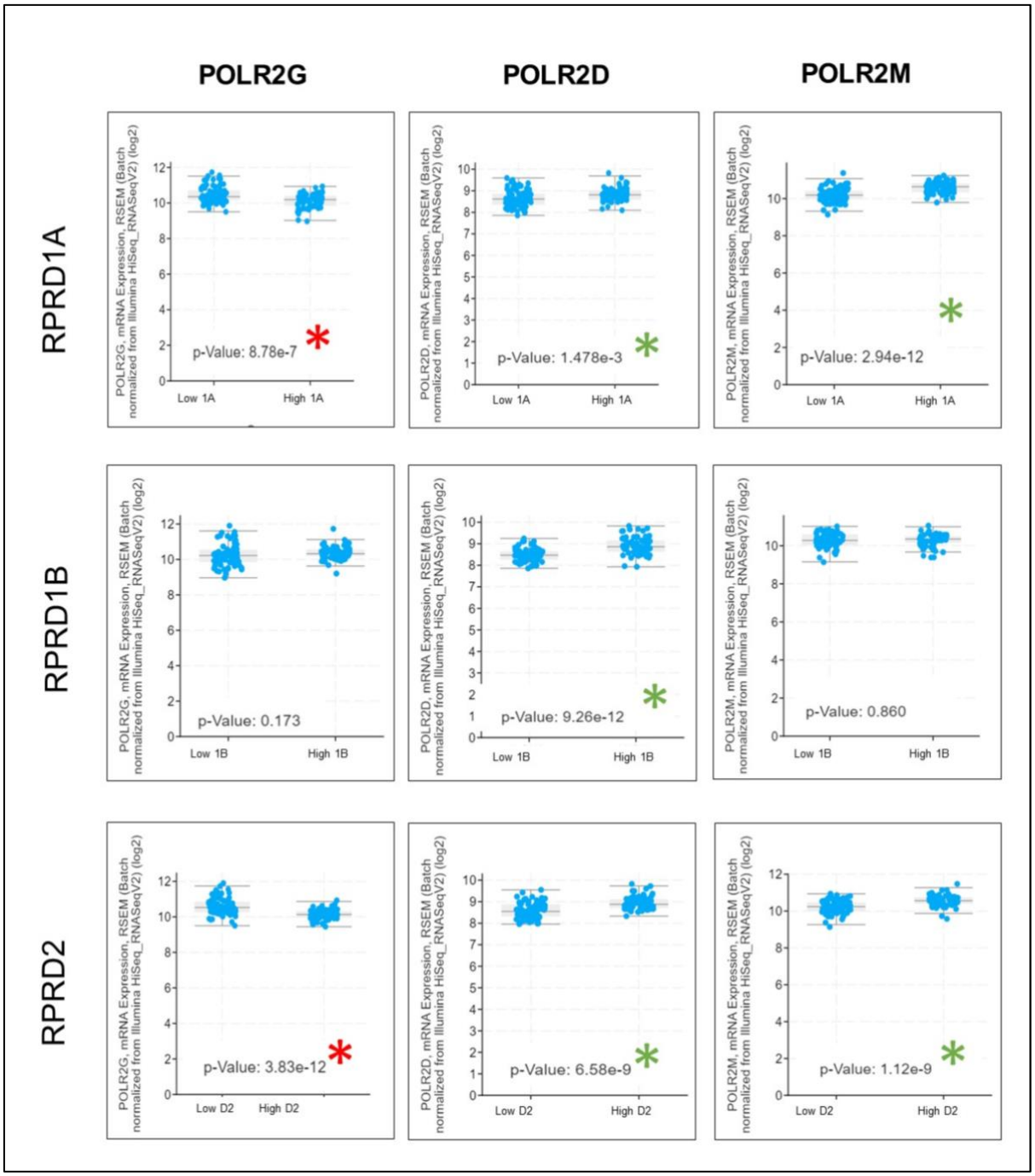


Figure 42:Changes at RPRD interacting RNAPII subunits mRNA expression in glioma samples which were diffrentially expressing RPRDs.

Comparison of mRNA levels for RPRD-interacting RNAPII subunits in glioma cells categorized as either high or low expressing RPRD proteins. The p-values for these comparisons are displayed within each plot(t-test). A red asterisk signifies a statistically significant increase in the queried protein's expression in low RPRD-expressing samples, while a green asterisk indicates a significant increase in the corresponding protein's expression in high RPRD-expressing samples.

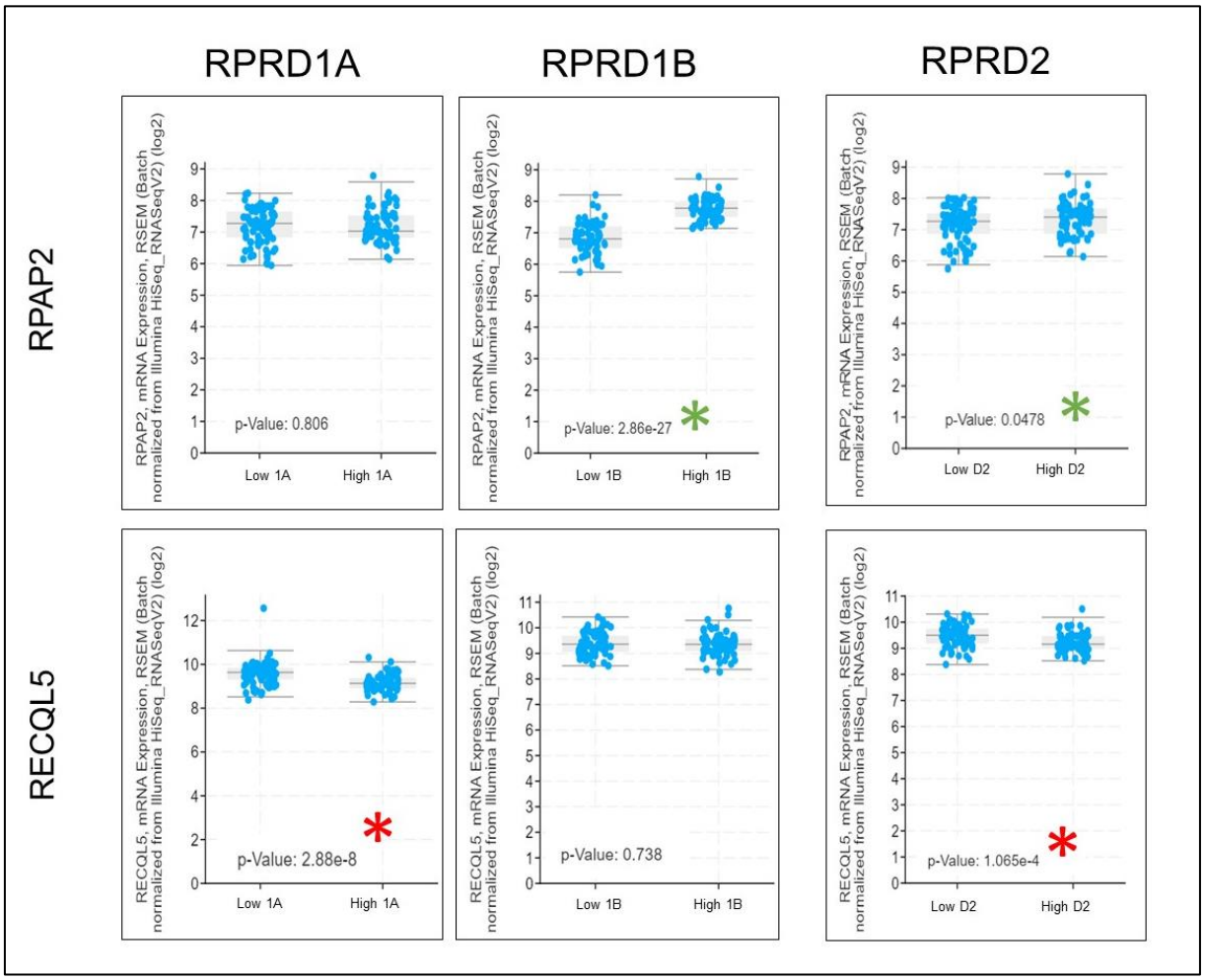


Figure 43: Changes at RPRD interacting proteins, RPAP2 and RECQL5, mRNA expression in glioma samples which were diffrentially expressing RPRDs.

Comparison of mRNA levels for RPRD-interacting proteins RPAP2 and RECQL5 in glioma cells categorized as either high or low expressing RPRD proteins. The p-values for these comparisons are depicted within each plot (t-test). A red asterisk denotes a statistically significant increase in the expression of the queried protein in low RPRD-expressing samples, while a green asterisk signifies a significant increase in the expression of the corresponding protein in high RPRD-expressing samples

I also analysed an array data available for the low-grade glioma dataset. The protein data was obtained using the reverse-phase protein array (RPPA) method, which is a miniature version of dot blot that enables the determination of protein levels in multiple human and animal tissues and cell lines simultaneously (Boellner & Becker, 2015). In this dataset, glioma samples were tested for 206 proteins using RPPA. After distinguishing high and low RPRD-expressed glioma samples, I assessed the differential protein levels under the high and low RPRD expression conditions. It was compared low and high-expressed samples to test how RPRD expression affects protein level changes. I found that 56 proteins were highly expressed in RPRD1A-expressing gliomas, while 36 proteins had high levels in RPRD1A low-expressed samples. Similarly, 60 proteins were highly expressed in RPRD1B low-expressed gliomas, while only 34 proteins were highly expressed in RPRD1B highly-expressed samples. In RPRD2 low-expressed samples, I identified 51 high-level proteins, while 55 proteins had high levels in RPRD2 highly expressed samples. The list of differentially changed proteins is provided in the appendix (**see Appendix Table A1**)

I conducted a comprehensive Gene Ontology (GO) annotation analysis to elucidate alterations in molecular functions within the context of RPRD expression variations in gliomas **(Figure 44, 45, and 46**). Notably, I observed noteworthy modifications in molecular functions related to transcription across all RPRD-deregulated glioma samples. Furthermore, I extended the analysis to assess disparities in cellular component and biological process terms among the RPRD-deregulated gliomas. My findings revealed that changes in RPRD expression significantly impacted molecular functions associated with transcription factor binding and RNA binding.

Specifically, reduced expression of all RPRD genes in glioma samples predominantly influenced molecular functions related to protein kinase binding activities and kinase activities. Of notable significance, I observed considerable modifications in Gene Ontology (GO) terms related to 3’ overhang single-stranded DNA deoxyribonuclease activity within gliomas exhibiting RPRD protein overexpression. Additionally, in gliomas characterized by the overexpression of RPRD1B and RPRD2, there were noteworthy alterations in GO terms associated with DNA damage binding.

Moreover, within the cellular component category, there was prominent enrichment of terms associated with DNA damage-related complexes, including but not limited to the ERCC4-ERCC1 complex, DNA repair complex, and sites of double-strand breaks (as demonstrated in **Appendix Figure A2**).

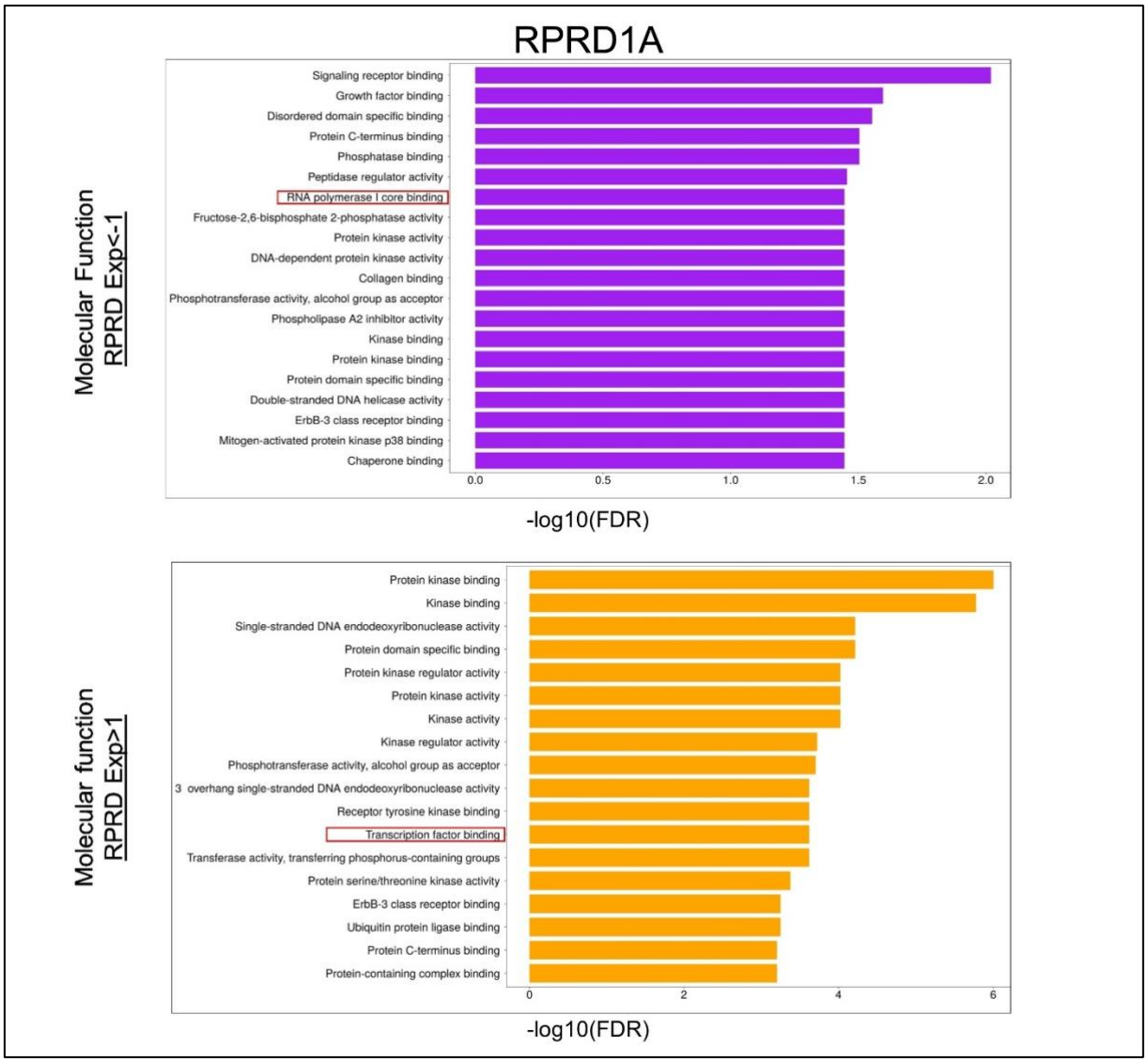


Figure 44: Gene ontology (GO) term enrichment analysis with diffrentially expressed RPRD1A in glioma samples.

The GO annotation results were derived from an analysis of genes that exhibited differential expression patterns in response to low and high expression of RPRD1A within glioma patient samples. These GO categories encompassed molecular functions. The sorting of GO categories within each functional group was performed by arranging them in descending order of statistical evidence, determined by the p-values obtained from the GO enrichment test.

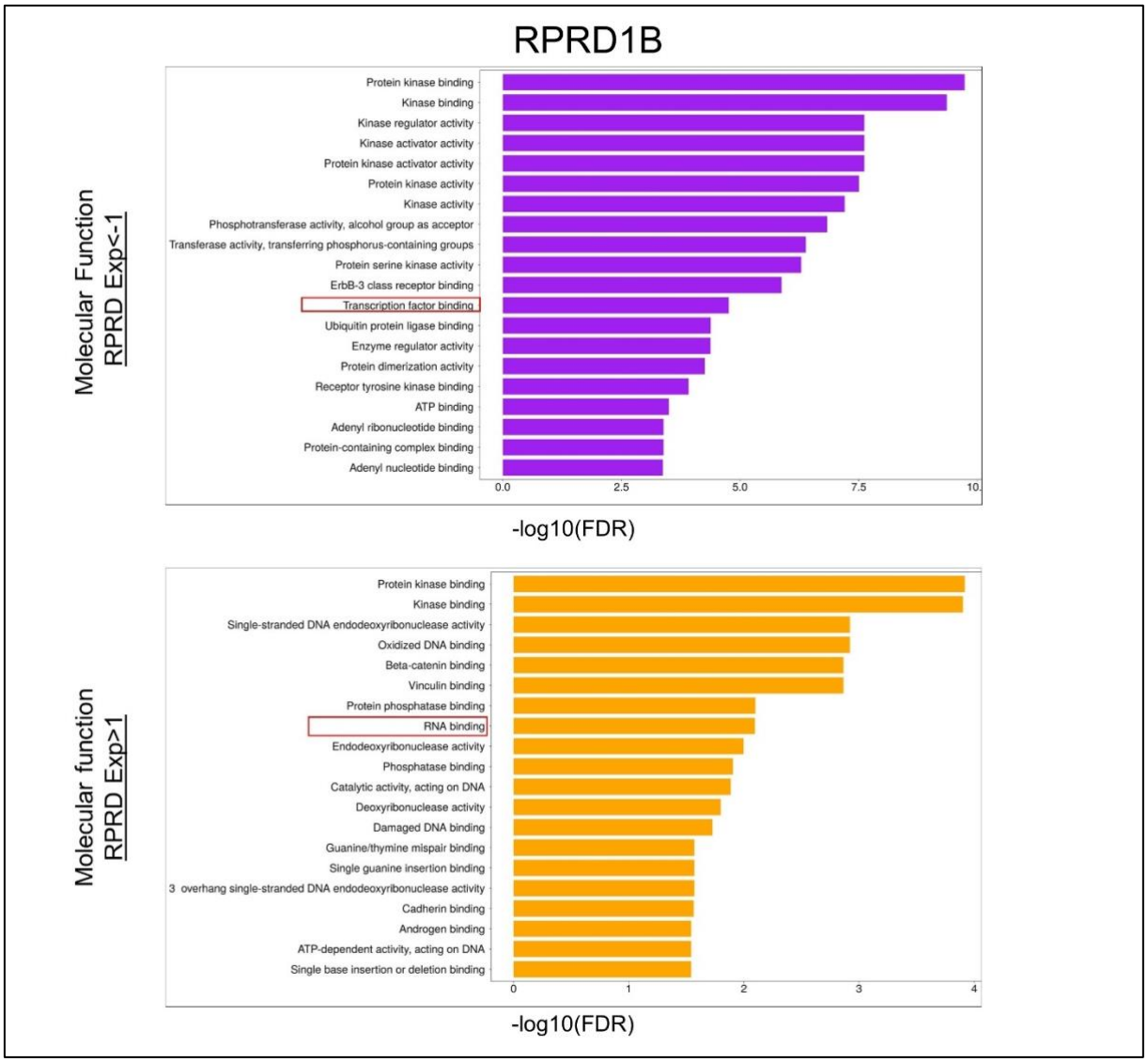


Figure 45: Gene ontology (GO) term enrichment analysis with diffrentially expressed RPRD1B in glioma samples.

The GO annotation results were derived from an analysis of genes that exhibited differential expression patterns in response to low and high expression of RPRD1B within glioma patient samples. These GO categories encompassed molecular functions. The sorting of GO categories within each functional group was performed by arranging them in descending order of statistical evidence, determined by the p-values obtained from the GO enrichment test.

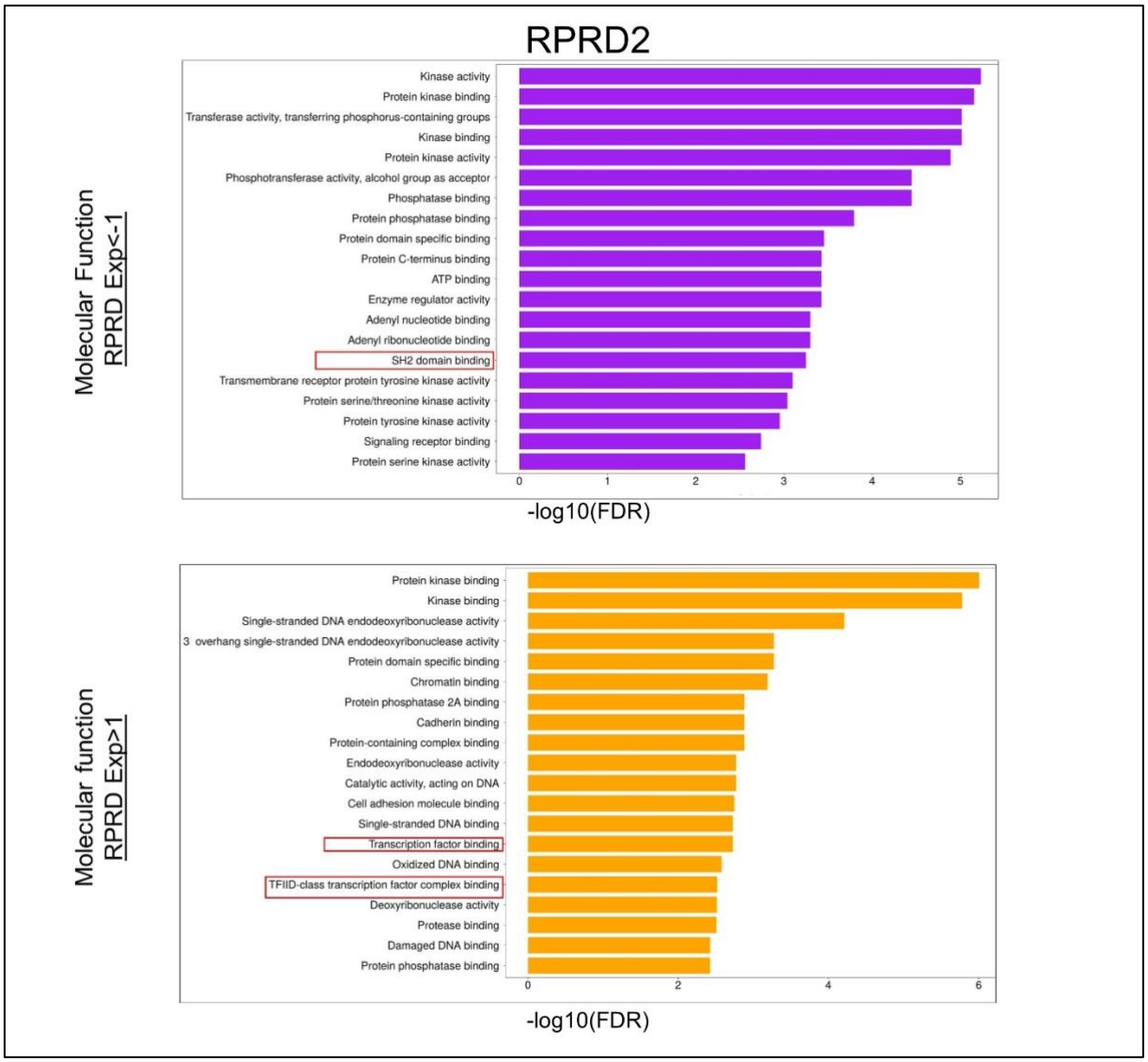


Figure 46: Gene ontology (GO) term enrichment analysis with diffrentially expressed RPRD2 in glioma samples.

The GO annotation results were derived from an analysis of genes that exhibited differential expression patterns in response to low and high expression of RPRD2 within glioma patient samples. These GO categories encompassed molecular functions. The sorting of GO categories within each functional group was performed by arranging them in descending order of statistical evidence, determined by the p-values obtained from the GO enrichment test.

RNaseH is a family of endonuclease enzymes that do not require sequence specificity. A broad classification of RNase H is H1 and H2 subtypes. RNA components of R-loops are degraded by both H1 and H2 enzymes, contributing to the maintenance of genome stability. It has been demonstrated that RNASEH function is induced after stress, which is characterized by the accumulation of R-loops in cells (Lockhart et al., 2019). I have shown that depletion of RPRD1B and RPRD2 in HEK293T cells reduces the relative amount of R-loops in several genes. However, their upregulation in cells leads to the accumulation of R-loops at the 5' and 3' ends of the tested genes. I compared the level of RNASEH1 mRNA levels in low-grade glioma cancer cells expressing differential levels of RPRD proteins (**Figure 47**). For all three RPRD proteins, downregulation of RPRDs in gliomas led to a decreased level of RNASEH1 expression compared to RPRD-unaffected samples. However, RPRD upregulated glioma samples significantly expressed a higher level of RNASEH1. My DRIP results agree with the interpretation that overexpression of RPRD in gliomas leads to the accumulation of R-loops. Increasing the expression of RNASEH1 is probably a response of the cells to accumulated R-loops because the cells try to resolve them.

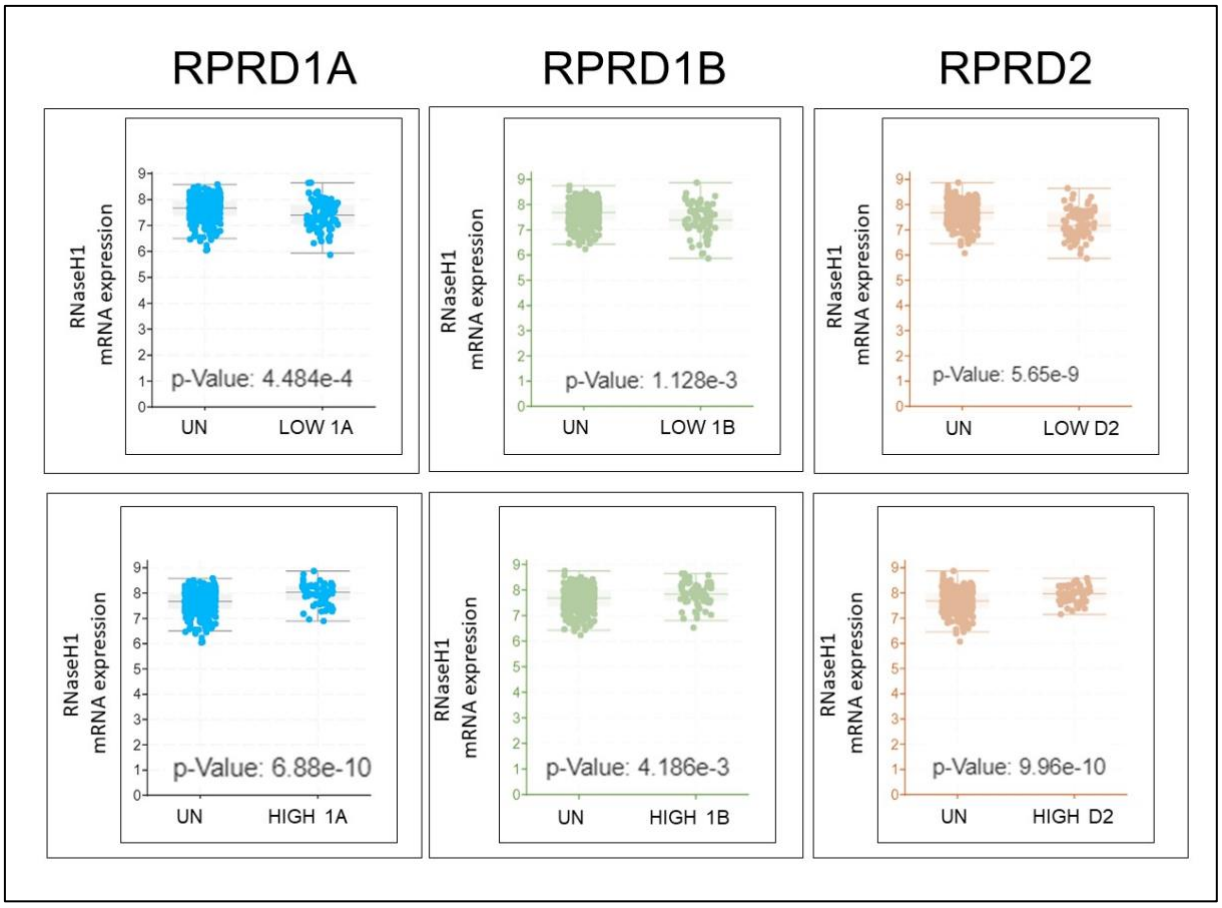


Figure 47: RNaseH1 mRNA expression in differentially RPRD expressed glioma samples.

Comparison of mRNA levels for RNaseH1 in glioma cells categorized as either high or low expressing RPRD proteins. The p-values for these comparisons are displayed within each plot (t-test). *In the graphs, blue dots represent RPRD1A deregulation, while green dots signify RPRD1B deregulation. Samples characterized by RPRD2 deregulation are indicated by orange dots.*

Eukaryotic DNA is organized into chromatin structures consisting of nucleosomes, which are composed of DNA wrapped around an octamer protein complex made up of 8 histone proteins (H2A, H2B, H3, and H4). The modification of these histone proteins can serve as a marker for various conditions. H2AX is a member of the H2A family and plays an important role in the response to double-stranded DNA breaks, specifically through phosphorylation of serine residue 139 (Rogakou et al., 1999). Upon detection of double-stranded DNA breaks, the ATM protein is recruited to the break site to phosphorylate H2AX, which in turn signals for the recruitment of repair enzymes to the

break site(Kinner et al., 2008). R-loops can be a source of double-stranded breaks resulting from transcriptional and replication conflicts or cleavage of the R-loop by nucleases of the TC-NER pathway (XPG and XPF in mammals) (Rinaldi et al., 2021).

In this study, I observed increased levels of R-loops in RPRD overexpressing cells. To assess the level of DNA damage, I measured the levels of γH2AX (pSer139) proteins in RPRD overexpressing cells using western blotting. As shown in **Figure 48**, the levels of γH2AX (pSer139) clearly increased in RPRD overexpressed cells. However, I also detected multiple bands with higher molecular weight than H2AX that were specific to the antibody used. H2AX protein might also be ubiquitinated, which is important for phosphorylation of H2AX and recruitment of ATM to the damage site (Pan et al., 2011). I suggest that the accumulation of R-loops in RPRD overexpressing cells induces double-stranded DNA breaks.

Based on the results obtained from the COSMIC database, RPRD1B and RPRD2 were found to be upregulated in several cancers. My results suggests that overexpression of RPRD1B and RPRD2 may lead to the accumulation of R-loops and double-stranded DNA breaks in the genome. This could potentially trigger the development of cancer.

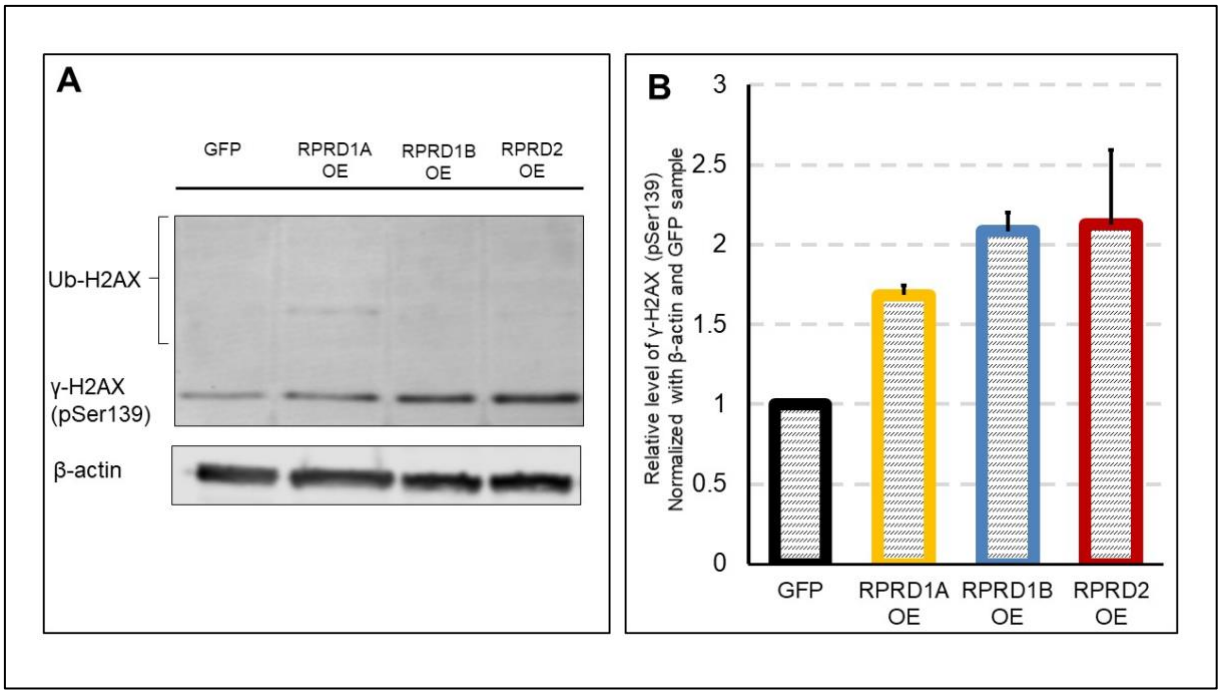


Figure 48: γH2AX (pSer139) level in RPRD overexpressing cells.

Representative Western blot images illustrating the levels of γH2AX (phosphorylated at Ser139) in cells overexpressing RPRD (A). The bar graphs depicting the relative densitometry of the protein bands, which were normalized against β-actin and the control sample (GFP) (B). The error bars correspond to the standard deviation, calculated from two independent experiments (n=2)

2.3 Accumulated effect of simultaneous downregulation or upregulation of all three RPRD proteins in gliomas.

Next I investigated the effects of simultaneous deregulation of all three RPRD proteins on glioma cells. The samples were divided into three groups: 205 samples unaffected by RPRD expression, six samples expressing all three RPRD proteins at high levels (referred to as 3-UP samples), and fifteen samples with downregulated expression of all three RPRD proteins (referred to as 3-DOWN). I observed 508 genes that were differentially overexpressed in 3-UP samples compared to unaffected samples, while 4145 genes were upregulated in 3-DOWN samples when compared to unaffected samples. To analyze these differentially expressed genes, I performed GO analysis using the ShinyGO web tool. The functional terms for biological processes, cellular components, and molecular functions are illustrated (**Figure 49, 50**). The data was sorted according to FDR values and filtered to show the 20 most affected terms. The query can be accessed using the following link:" *https://bit.ly/3HfMJPV".*

Analysis with 3-DOWN samples, the genes in the biological process group were mainly involved in respiratory and cellular and mitochondrial translation. The molecular function terms related to oxireductase, electron, proten transporter activity and rRNA binding. Most of the cellular component genes were located in the mitochondrion, mitochondria and ribosomal subunits (**Figure 49).** As a result of downregulation of all three RPRD proteins in glioma cells, the pathways which are related with respiratory and cellular enegetic pathhways was chanced.

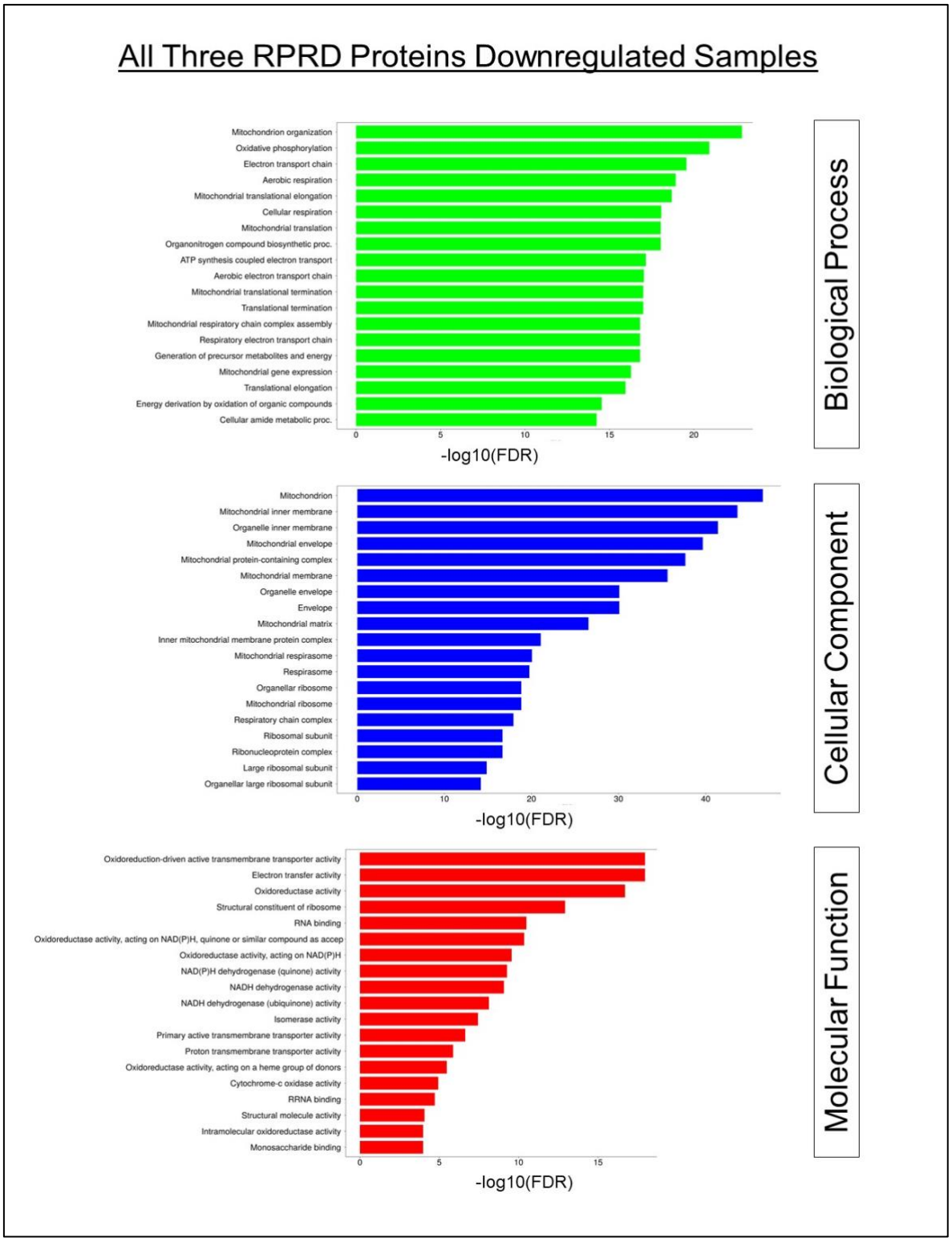


Figure 49: GO annotation results for samples which are downregulating all 3 PRPD protein simultaneously.

Gene Ontology (GO) terms related to biological processes, cellular components, and molecular functions were derived from the upregulated genes within cells simultaneously overexpressing all three RPRD proteins.

The differentially expressed genes between glioma samples overexpressing all three RPRD proteins (6 samples) and unaffected glioma samples (205 samples) were found to be predominantly upregulated (508 genes, p-value: 0.05). I conducted gene ontology (GO) enrichment analysis on the list of upregulated genes using the ShinyGO database, and the results are shown (**Figure 50**). The upregulated genes were mainly associated with RNA biological processes such as RNA metabolic process, positive and negative regulation of transcription, and mRNA processing. Additionally, these genes were predominantly involved in nuclear protein-containing complex, transcription regulatory complex, ERCC4-ERCC1 (nucleotide excision repair complex), and mediator complex. The molecular functions were related to transcription regulatory DNA binding, transcription factor binding chromatin, and RNA binding, all of which are related to transcription. The results suggest that the simultaneous upregulation of all three RPRD proteins in glioma cells has a significant impact on transcriptional pathways.

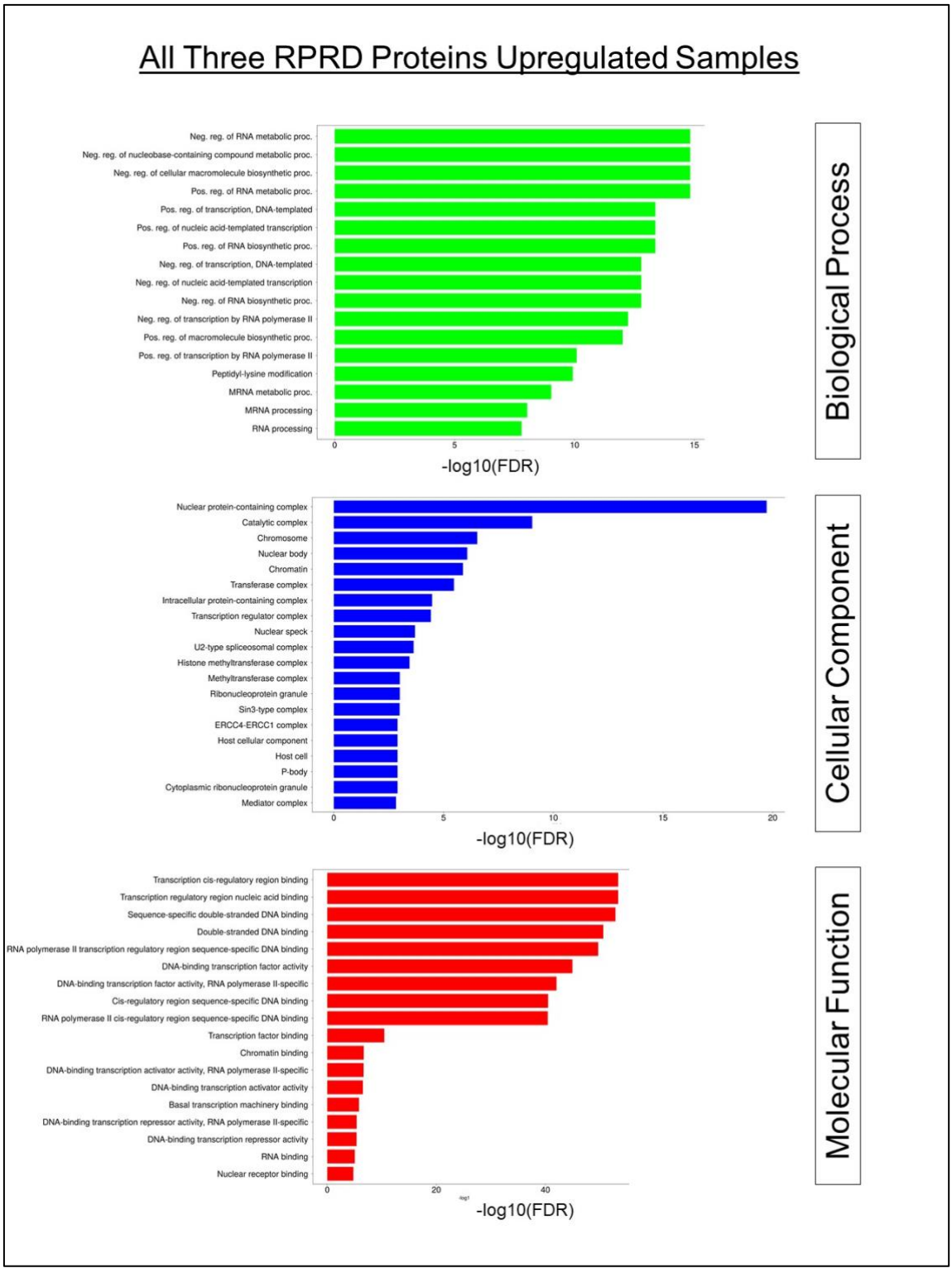


Figure 50: GO annotation results for samples which are overexpressing all 3 RPRD protein simultaneously.

Gene Ontology (GO) terms related to biological processes, cellular components, and molecular functions were derived from the upregulated genes within cells simultaneously overexpressing all three RPRD proteins.

I investigated the changes in mRNA expression levels of R-loop related proteins in gliomas under the conditions of upregulation or downregulation of all three RPRDs (**Figure 51**). Specifically, I measured the expression levels of RNASEH1, SETX, TOP1, and DHX9 mRNAs (Kumar et al., 2022; Skourti-Stathaki & Proudfoot, 2014; Wang et al., 2018). As previously reported, RNASEH1 specifically degrades R-loops and resolves them, while SETX, a homologue of yeast Sen1, functions as an RNA/DNA helicase and resolves R-loop formation (Cohen et al., 2018). Top1 is a DNA topoisomerase that resolves DNA negative supercoiling, and in yeast, the absence of TOP1 results in R-loop accumulation (El Hage et al., 2010). DHX9 is an ATP-dependent DNA helicase that helps prevent R-loop formation. As shown, the expression of all these R-loop related proteins increased in glioma samples where all three RPRDs were upregulated, while their expression levels were lower than in unaffected samples (**Figure 51**). This suggests that cells began to express these proteins under the stress of accumulated R-loops to resolve them.

I further examined the DNA damage markers TP53BP1, ATM, ATR, and Ku80 (XRCC5) in glioma samples (**Figure 52**). Excessive accumulation of R-loops can cause double-strand breaks, which activate ATM and ATR through various mechanisms, including direct interaction with DNA damage, indirect interaction with damage sensor proteins, or both (Marabitti et al., 2019). It has also been demonstrated that R-loops that block transcription are excised by XPF (ERCC4) and XPG (ERCC6) nucleases, which are involved in transcription-coupled nucleotide excision repair pathways, leaving a single-stranded DNA gap that can become a double-strand break over time In conclusion, increased R-loop levels in cells can induce the expression of proteins that are important for R-loop detection and resolution. As a result of R-loop-coupled genome instability, cells begin to express DNA damage marker genes.

Detected R-loop-induced DNA damages can be attempted to be fixed by increasing the expression of DNA repair proteins.

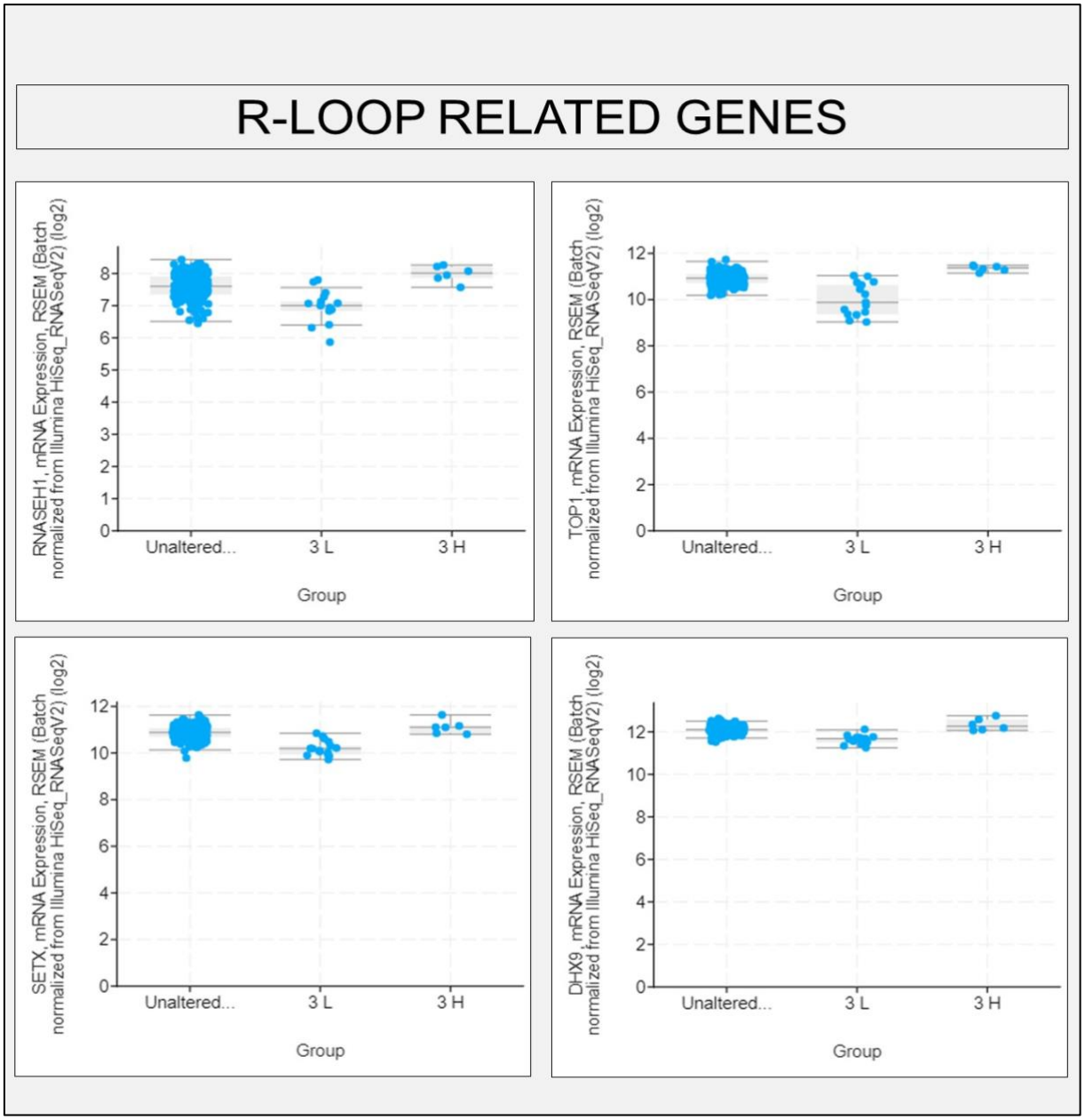


Figure 51: Comparison of R-loop related at 3-UP and 3-DOWN samples

Comparative analysis of mRNA expression levels pertaining to R-loop associated genes, including RNASEH1, TOP1, SETX, and DHX9, was conducted in glioma cells under three distinct conditions: those with unaltered RPRD expression, those with all three RPRD simultaneous downregulated (3L), and those with all three RPRD simultaneous upregulated (3H).

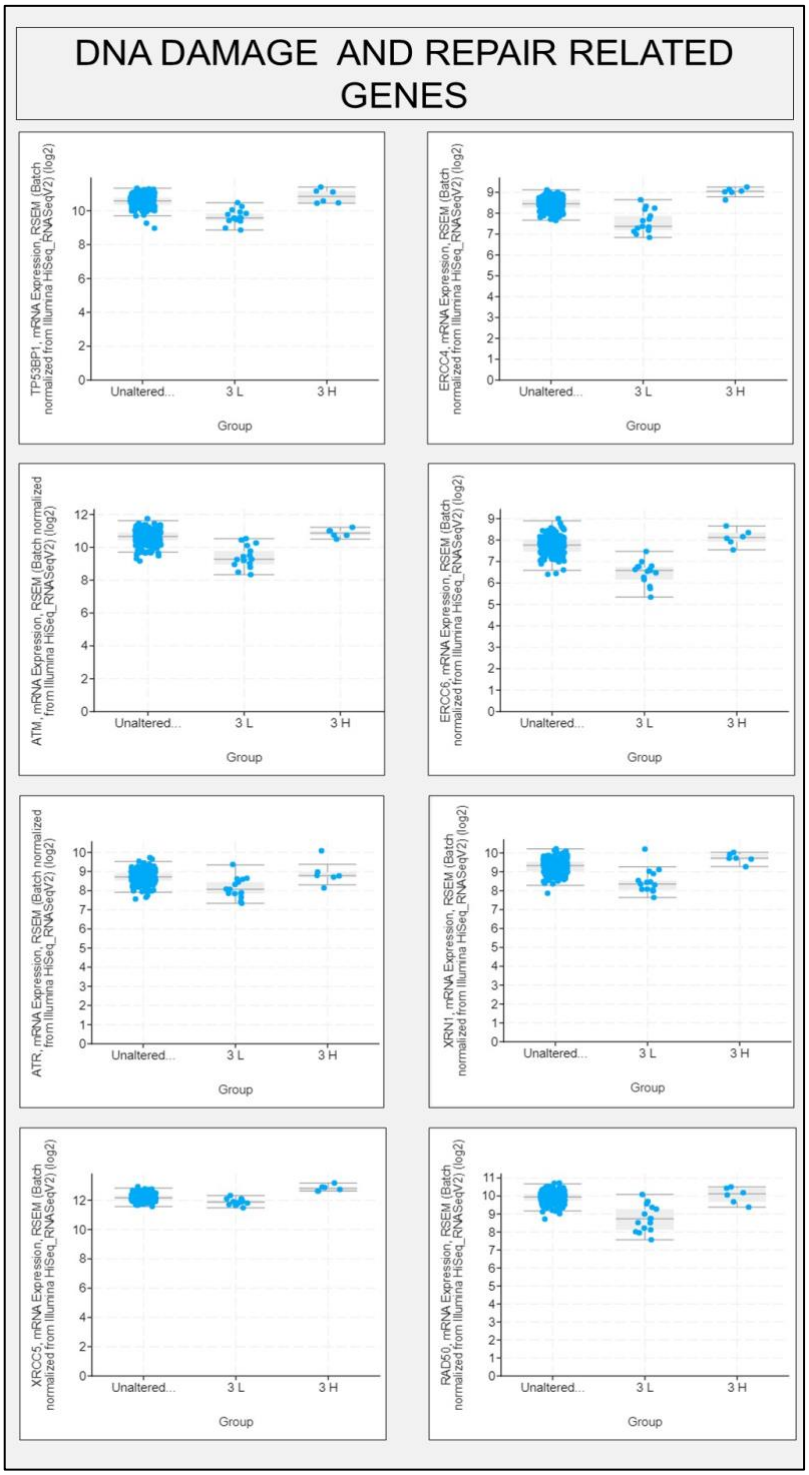


Figure 52:Comparison of DNA damage related and DNA repair related genes at 3-UP and 3-DOWN samples

A comparative assessment of mRNA expression levels pertaining to genes associated with DNA damage and DNA repair pathways was conducted in glioma cells. This evaluation encompassed glioma samples unaltered RPRD expression and samples with all three RPRD similtaneous downregulated (3L) and upregulated (3H).

2.4 Physiological Effect of RPRD proteins

Cells respond to both external and internal stresses by altering cell cycle progression, allowing time for repair the detected error. In this study, I have demonstrated that the absence or abundance of RPRD proteins leads to transcriptional changes in cells. Literature suggests that RPRD1B and RPRD1A have an impact on the cell cycle, and changes in the levels of RPRD proteins may affect the transcription of cell cycle-related genes (Lu et al., 2012; Zheng et al., 2016). Therefore, I aimed to study the impact of RPRD proteins on cell physiology, specifically at the level of cell cycle regulation. Through this experiment, the effects of individual RPRD proteins in the same cell line were elucidated, with detailed analysis performed through both depletion and overexpression of RPRD in cells.

First, I analyzed the cell cycle in asynchronous cells with differential expression of RPRD proteins (**Figure 53**). I first knocked down RPRD proteins in HEK293T cells for 48 hours and then stained cells with PI and performed FACS experiments. My observations indicated that knockdown of RPRD1B, RPRD1A, and RPRD2 led to an accumulation of cells at the G1/S phase. The shortening of the S phase caused this accumulation in RPRD1B and RPRD1A depleted cells, while the G2/M phase was shortened in RPRD2 depleted cells (**Figure 53**; **A,C**). The overexpression of RPRD1B and RPRD2 had a more complex effect on the cell cycle compared to their depletion, but RPRD1A overexpression did not lead to any significant changes in the cell cycle. Overexpression of RPRD1B caused an increase of almost 10% in the number of cells in the G2/M phase, while cells accumulated at the G0/G1 phase under RPRD2 overexpression (**Figure 53; B, D**).

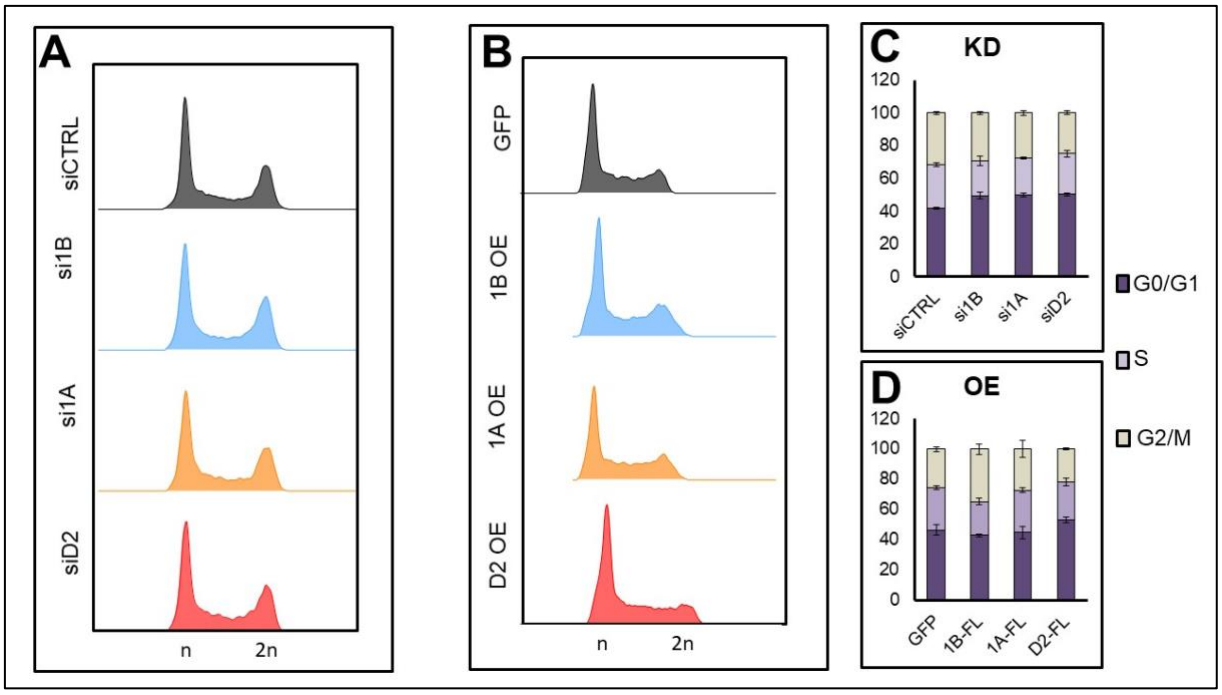


Figure 53: FACS results under RPRD upredulated or downregulated conditions. Illustarion of the pattern of the cell cycle under downregulation (A) and upregulation (B) of RPRD proteins in not synchronized cells, respectively. The bar graphs represent the average of the distribution throughout cell cycle stages in RPRD knockdown (C) and RPRRD overexpressing (D) cells. The error bars indicated standard deviations of three independent expreiments (KD, knockdown; OE, overexpression).

To monitor cell cycle progression in more detail under the deregulation of RPRD protein expression, I utilized the double thymidine block method. Briefly, I added thymidine directly into the cell medium 6 hours after siRNA knockdown or tetracycline induction to disrupt deoxynucleotide metabolism for 18 hours of incubation. This resulted in cells accumulating throughout the S phase. To achieve unique synchronization, I released cells for 9 hours by replacing the media with fresh media,

followed by a second thymidine treatment for 15 hours. Samples were collected at time 0 (the end of synchronization), 4 hours, and 8 hours after thymidine block release.

The synchronization method was applied consistently to all conditions (siCTRL, si1B, si1A, and siD2). After release from the synchronization point (0h), most siCTRL cells were at the S phase after 4 hours of release. At the 8-hour time point, G0/G1 phase cells started to increase because the G2/M phase cells had completed the previous cycle. RPRD1A depleted cells progressed through the G0/G1 to S phase at a similar pace as siCTRL, while RPRD1B and RPRD2 depleted cells progressed slowly. However, cells depleted of RPRD1B, RPRD1A, and RPRD2 were not able to pass through the G2/M phase as quickly as siCTRL, indicating that cells were stalled at the G2/M phase after RPRD knockdown (**Figure 54**).

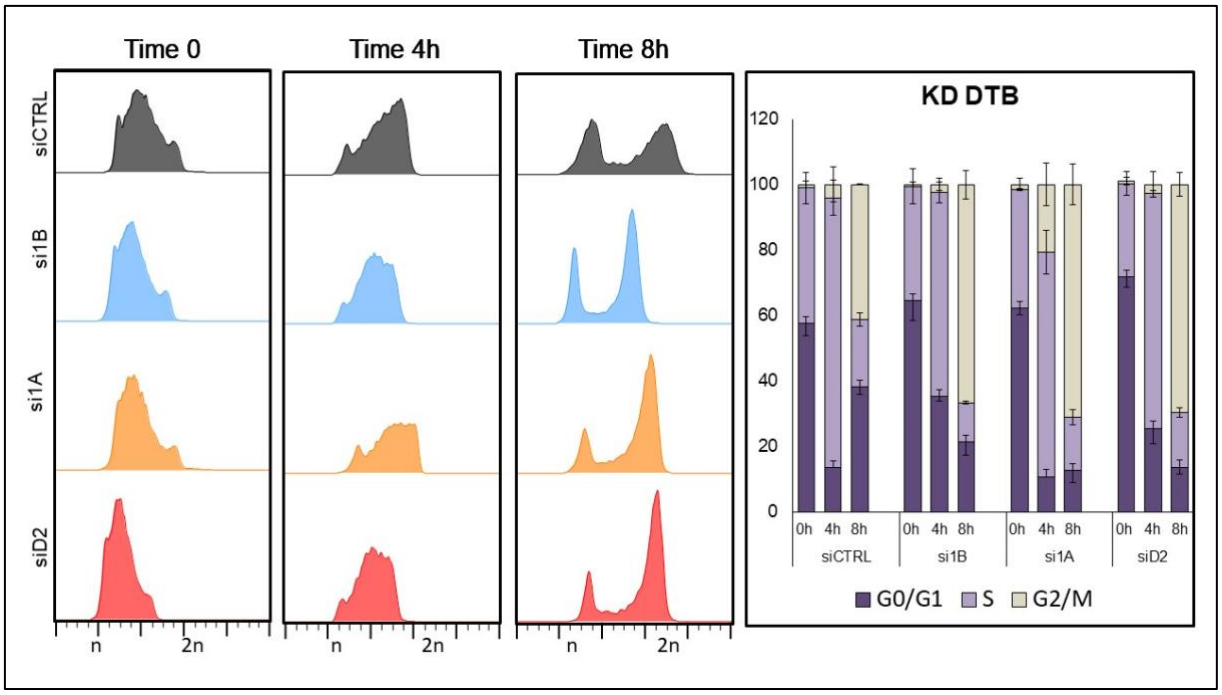


Figure 54: The cell cycle progression in RPRD knockdown cells.

FACS analyses reveal the cell cycle progression at various stages subsequent to release from DTB in RPRD depleted cells (left). A bar graph displaying the percentage of cells in each cell cycle phase following DTB release is presented for both RPRD knockdown (KD) and control conditions (right). This data is derived from two independent experiments, with the error bars indicating standard deviations.

After performing the DTB synchronisation method on both HEK293T and Flp-In T-REx 293 cells, I observed a difference in the accumulation of cells at different phases. HEK293T cells were found to accumulate mostly at the G0/G1 phase, while Flp-In T-REx 293 cells were synchronised mostly at the S phase (**Figure 54 and Figure 55**). This difference is attributed to the cell type. In cells overexpressing RPRD1B and RPRD1A, I observed faster progression from the S phase to the G2/M phase compared to GFP cells. Furthermore, the cells completed the cell cycle as evidenced by the increase in the number of cells in the G0/G1 phase. In contrast, RPRD2

overexpressing cells were found to stack at the G2/M phase while progressing faster through the S phase (**Figure 55**).

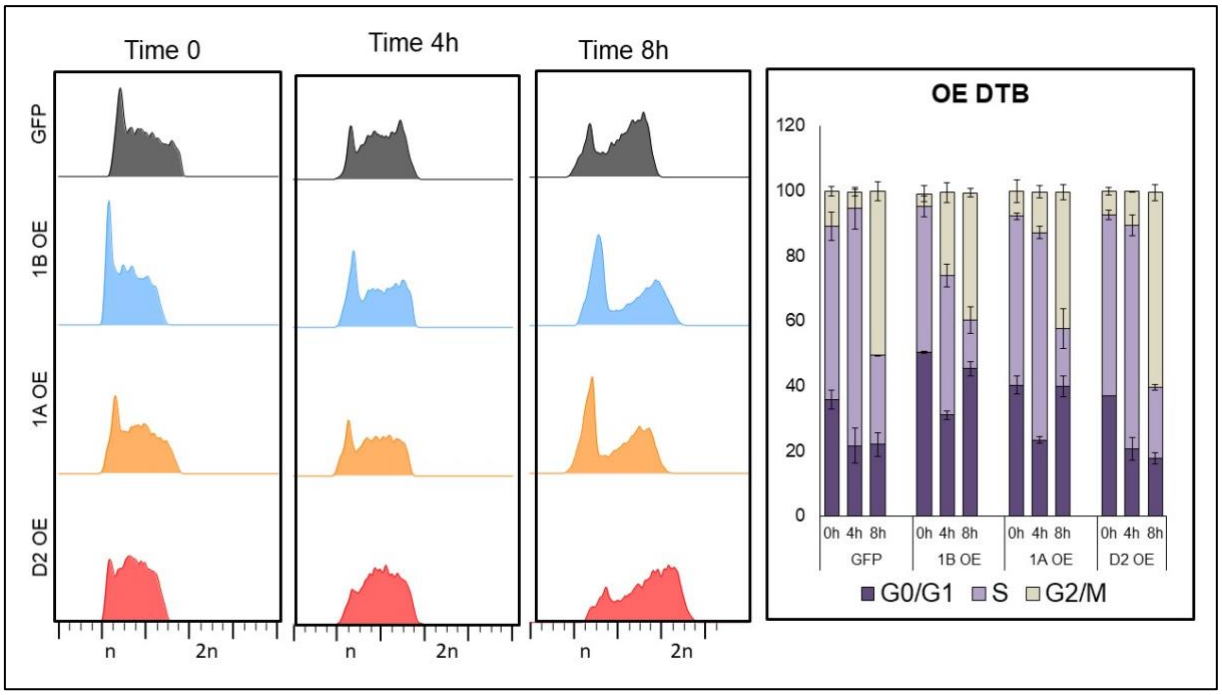


Figure 55:The cell cycle progression in RPRD overexpresing cells.
FACS analyses reveal the cell cycle progression at various stages subsequent to release from DTB in RPRD upregulated cells (left). A bar graph displaying the percentage of cells in each cell cycle phase following DTB release is presented for both RPRD overexpressing (OE) and GFP conditions (right). This data is derived from two independent experiments, with the error bars indicating standard deviations.

As a conclusion, my results indicate that depletion of RPRD proteins leads to G0/G1 arrest in asynchronized cells. RPRD1B overexpression causes G2/M accumulation, while RPRD2 overexpression leads to G0/G1 arrest. Double thymidine block synchronization revealed a more detailed insight into the effect of RPRD on cell cycle progression. Depletion of RPRD proteins slows down the cell cycle, as cells progress more slowly throughout the cell cycle and cannot initiate the next cycle as fast as

control samples. However, RPRD1A and RPRD1B overexpression accelerates cell cycle progression, as evidenced by the accumulation of cells at G0/G1 at 8 hours, while some cells have already started the next cycle. These findings demonstrate a positive effect of RPRD proteins on cell cycle regulation.

2.5 CONCLUSIONS

There is lots of evidences about the RPRD proteins have an effect on multiple cellular processes including transcription, cell cycle, HIV infection as well as cancer formation. In this chapter, I demonstrated that the effect the level of RPRD protein expression on cancer progression. After careful consideration, I chose a dataset from patients with low-grade glioma of the brain as it was the most comprehensive and included samples with varying levels of RPRD deregulation. My findings showed that downregulation of RPRD1A, upregulation of RPRD1B, and either upregulation or downregulation of RPRD2 led to an increase in aggressiveness of gliomas. In this dataset, I also checked the expression of the several genes and run the GO annotation to see the changes in cells. RPRD upregulated glioma samples had more DNA damage response than the low RPRD expression ones.

I also checked the expression level of RNASEH1 in RPRD high expressed gliomas. It was observed that these cells expressed the more RHASEH1 than low RPRD expressing ones. This is probably originates the accumulation of R-loops in gliomas. Changes at transcription rate affect the R-loop levels whose excessive accumulation led to DNA damage. If the cell detect any damage on DNA, the DNA repair mechanisms were activated. During this repair after detection, cells stopped the progression of cell cycle. I assessed the effect of RPRD proteins on cell cycle with FACS experiment. It was found that RPRD depletion led to arrest cells at G0/G1 phase. Overexpression of RPRD1B led to G2/M phase while RPRD2 upregulation caused again G0/G1 arrest. Then, I wondered the how RPRD poteins affected the cell cycle progression. To do this, I performed DTB synchronisation. I observed that RPRD

depletion slowed down the cell cycle progression wheras their overexpression of RPRD proteins led to acceleration of cell cycle.

In conclusion, an abnormal level of RPRD proteins in cells can alter transcription rates, leading to the formation of R-loops and DNA damage. The accumulated DNA damage can affect cell cycle progression and result in cell cycle arrest at different stages. Therefore, the level of RPRD proteins can impact the aggressiveness of cancer cells by altering DNA damage and cell cycle progression.

PART 4: DISCUSSIONS

Transcription regulation is a vital process that enables cells to respond to various intracellular and extracellular signals, define cell identity during development, maintain it throughout life, and coordinate cellular activity and cell fate. This process is governed by numerous nucleoprotein complexes that establish highly dynamic regulation. The transcription cycle encompasses several key levels, including nucleosome disassembly, DNA opening, DNA-protein interactions, recruitment and assembly of the entire transcription machinery, initiation phase, pause release, and elongation phases, as well as the termination phase. Several factors can stimulate or inhibit transcription to control each step (Cramer, 2004, 2019; Lee & Young, 2013).

Proteins that can directly recognize and bind to DNA or interact with the core transcription machinery proteins regulate this dynamic and complex process. RPRD proteins, a group of proteins that can associate with the C-terminal domain (CTD) of Rpb1, the largest subunit of RNA polymerase II, belong to this category (Buratowski, 2009; Ni et al., 2011). RPRD1A and RPRD1B are small proteins containing two main identical CID and CC domains, while RPRD2 is a relatively larger protein consisting of CID and serine-proline rich region. While these proteins do not possess any determined kinase or phosphatase activity, they are capable of binding to RNAPII CTD through their CID domains. Accumulated knowledge about them has suggested that RPRD proteins play roles in several cellular pathways, including transcription, cell cycle, DNA damage response, and carcinogenesis(Li et al., 2021). All three RPRDs were copurified with a wide range of protein complexes (Ding et al., 2018; Morales et al., 2014; Ni et al., 2011). The data suggest that to regulate RNAPII activity, RPRDs might serve as a docking site for proteins. It is important to elucidate their role in transcription. It is crucial to comprehend the regulation of transcription in human cells.

1. RPRD proteins affect the transcription negatively being in different protein complexes

As previously mentioned, all RPRD proteins possess CID domains but lack enzymatic functions. To map RPRD occupancy on genes, a ChIP analysis was performed at the LEO1 locus, which showed significant enrichment of RPRD1A and RPRD1B in both the promoter and downstream sequences, with the promoter showing relatively higher levels (Ni et al., 2011). A study also showed that anti-RPRD1B antibody strongly precipitated the promoter region of CYCLIN-D1 by ChIP using primers specific for the CYCLIN D1 gene. They suggested that RPRD1B enhances the loop formation to recycle RNAPII from the terminator site to the promoter region throughout transcription of the CYCLIN D1 gene (Lu et al., 2012). However, no data is currently available regarding the RPRD2 occupancy on any gene. In this study, I tested the occupancy of all three RPRD proteins on MYC genes by ChIP experiment and obtained significant enrichment over the mock (no antibody), particularly for RPRD1B and RPRD2 ChIP, at the transcription start site and termination site of the gene. The established binding patterns of these proteins on the gene prompted subsequent experiments aimed at elucidating their functional roles in transcription regulation. This finding supports the previous report that RPRD1A and RPRD1B occupy the promoter and termination sites (Ni et al., 2011). Moreover, I showed the RPRD2 binding pattern throughout the gene for the first time in literature. As RPRD proteins precipitate with phospho-forms of CTD and are present in both downstream regions and promoter regions, they may contribute to transcription initiation, elongation, and termination.

Initially, I investigated the total RNA levels in cells after RPRD deregulation but did not observe significant changes at the checked genes. Consequently, I aimed to determine the effect of RPRD proteins on transcription. To gain a deeper understanding of transcription initiation, elongation, and termination, I monitored quantitatively nascent transcription by performing 4sU labeling followed by 4sU-seq. My experimental results showed that RPRD proteins globally negatively affected the level of newly synthesized RNA.

The depletion of individual RPRD proteins resulted in the accumulation of nascent transcripts compared to the control. Notably, the suppression of RPRD1B and RPRD2 had a more significant impact on nascent RNA accumulation than RPRD1A depletion. Overexpression of RPRD proteins led to a decrease in nascent RNA synthesis, with RPRD1B and RPRD2 showing the most significant reduction. Conversely, the overexpression of RPRD1A did not significantly affect the nascent transcriptome. RPRD2 had the most prominent effect on nascent transcription, regardless of overexpression or downregulation, while RPRD1B had a moderate and intermediate effect in both cases, compared to RPRD2 and RPRD1A.

I performed total RNA sequencing on HEK293T cells with siCTRL and siRPRD1B. After ribosomal RNA depletion, I sequenced the samples. However, the RPRD1B depletion had a minimal impact on total RNA levels (**see the Appendix Figure A4**). The RPRD1B knockdown was performed for only 48 hours, and mRNA stability is regulated in response to complex cellular signals, similar to transcription regulation. mRNA decay has become a crucial factor in gene expression regulation (Garneau et al., 2007).

I speculate that the RNA decay mechanism may have been activated to maintain RNA homeostasis.

Subsequently, the accumulation of nascent RNA was enhanced upon RPRD depletion, suggesting that the transcription rate of RNAPII may be altered. To investigate this hypothesis, I conducted a DRB stop-chase experiment, whereby the release of RNAPII from promoter-proximal pausing requires phosphorylation by p-TEFb. DRB is used to block this process, allowing us to monitor nascent RNA synthesis after DRB release and determine the time point at which nascent RNA levels reach that of untreated samples. I found that the nascent RNA recovery was faster in all three individual RPRD depletion samples than in the control. Additionally, the DBR stop-chase experiment revealed that the absence of RPRD proteins in cells increased the speed of RNAPII by approximately 20%. I hypothesized that RPRD2 might have a significant effect on the speed of RNAPII. However, when compared with the control, RPRD1B increased the speed of transcription more than RPRD1A and RPRD2. This discrepancy may be due to the genes that I investigated or the limitations of the experimental resolution. To increase the resolution of this experiment, samples could be collected in every 5 minutes or different gene ampliconscould be examined.

The proteomic analysis of RPRD proteins revealed that RPRD1B exists in two distinct complexes, namely RPRD1B-RPRD2 and RPRD1B-RPRD1A heterodimers. These proteins can form homodimers or heterodimers through their coil-coil domain (Mei et al., 2014; Winczura et al., 2021). I proposed that the moderate effect of RPRD1B on the newly synthesized transcriptome depends on the protein complexes in which RPRD1B is present.

It has been demonstrated that the RPRD1B-RPRD1A heterodimer binds to phospho-Ser2- and/or phospho-Ser7 of CTD, facilitating the recruitment of RPAP2 to the transcription site (Ni et al, 2014). It was belived that RPAP2 is a phosphatase whose presence is crucial for removing Ser5 phosphorylation, thus inducing transcription elongation (Egloff et al., 2012). Nonetheless, RPAP2 failed to exhibit discernible phosphatase activity; instead, it functions as a negative transcription factor by impeding the interaction between RNAPII and transcription initiation factor TFIIF (Chen et al., 2021; Wang et al., 2022). RPRD1B was co-precipitated with with RPRD1A and RPRD2 (Ni et al, 2011). In 4sU-seq experiments, RPRD1B had a moderate effect compared to RPRD2 and RPRD1A on nascent RNA synthesis. The effect of RPRD1B on transcription appears to be influenced by its association with different heterodimer and homodimer formations, which may create distinct docking sites for transcription regulatory proteins.(Mei et al., 2014; Winczura et al., 2021). When Gdown1 is added to early elongation complexes, it inhibits elongation by competing with TFIIF for RNAPII binding. It was suggest that Gdown1 also affects DSIF-NELF functions and stabilizes stalled RNAPII across the human genome (Cheng et al., 2012; Davis et al., 2014; Jishage et al., 2012; Wu et al., 2012). Therefore, the main transcription inhibitory protein associated with RPRD proteins may be Gdown1 due to its interaction with all three proteins. However, other proteins with inhibitory effects on transcription are also associated with RPRD proteins, such as RECQL5 and PAF1. Previous studies have shown that loss of RECQL5 increases the elongation rate of RNAPII across the genome (Saponaro et al., 2014). It was demonstrated that RECQL5 was co-precipitated with RPRD1B and RPRD1A (Ni et al., 2011). In addition, loss of PAF1 leads to an increased phospho-Ser2 form of RNAPII and nascent transcription due to

recruitment of the SEC complex, which is required for faster transcription and releasing from the promoter (Chen et al., 2015) . Its interaction with RPRD1B was shown by AP-MS (Hein et al., 2015). In summary, RPRD proteins act as docking sites for negative elongation factors to recruit to transcription sites.

The regulatory role of RPRD1A and RPRD1B in transcription involves interacting with β-catenin, which coordinates and regulates gene transcription. Studies have suggested that RPRD1A competes with β-catenin to downregulate Wnt-targeted genes. Upon association of RPRD1A with TCF-4 on the promoter, RPRD1A recruits HDAC2 to the promoter to prevent gene expression. When Wnt signalling is activated, β-catenin interacts with RPRD1A and reduces the interaction between RPRD1A and TCF4, allowing for the formation of the β-catenin-TCF4 complex and subsequent transcriptional activation (Wu et al., 2010). On the other hand, RPRD1B increases RNAPII occupancy on promoter regions and assists RNAPII in forming the transcription initiation complex. After Wnt signalling activation, RPRD1B also increases the stabilization of β-catenin by acetylation to induce the transcription of Wnt-targeted genes (Zhang et al., 2018).

The presence of RPRD1A inhibits transcription by competing with β-catenin, while its absence stimulates Wnt-targeted gene expression by recruiting HDAC2 to the promoter and preserving the deacetylated state of histone-3 (Liu et al., 2015; Wu et al., 2010). After the assembly of the transcription initiation complex on the Wnt-targeted gene promoter, the joining of RPRD1B to the complex enhances transcription (Zhang et al., 2014). Thus, while RPRD1A behaves as a transcriptional inhibitory protein for these Wnt-targeted gene transcriptions, RPRD1B acts as an activator. These two

protein shows the opposite effect on transcription. The proteomic analysis which was performed in our lab with RPRD proteins showed that they can exist as homodimers or heterodimers (Winczura et al., 2021). The stoichiometry and formation of different protein complexes with RPRD proteins play a crucial role in cellular processes.

To summarize, my analysis using 4sU-seq and DRB stop-chase experiments indicates that RPRD proteins have a repressive effect on transcription elongation, likely due to their involvement in various protein complexes. They may function as docking sites for transcriptional regulators.

2. RPRD proteins may changes the mRNA stability by interactin with mRNA-decay pathway proteins

As previously mentioned, all three RPRD proteins interact with POLR2D, POLR2G (RNAPII subunits) and RPAP2 proteins. Functional studies with Rpb4, Rpb7, and Rtr1 proteins, which are homologs of human POLR2D, POLR2G, and RPAP2, respectively, have demonstrated that Rtr1 assists in the formation of the Rpb4/Rpb7 heterodimer. RNAPII utilizes the Rpb4/Rpb7 heterodimer to initiate transcription of protein-coding genes under promoter-dependent conditions, and it also plays a role in transcription-coupled DNA repair and pre-mRNA splicing in the nucleus. In addition to these primary functions of the Rpb4/Rpb7 heterodimer, it physically interacts with the transcript (mRNA imprinting) to affect mRNA export, translation, and mRNA decay. These functions suggest that the Rpb4/Rpb7 heterodimer shuttles between the nucleus and cytoplasm (Choder, 2004, 2011; Garrido-Godino et al., 2022). It has been shown that the Rpb4 and Rpb7 proteins mediate mRNA decay in yeast cells (Lotan et al., 2005, 2007). A recent study with these proteins demonstrated that the depletion of Rtr1 in yeast cells caused a decrease in the amount of mRNA imprinting by the Rpb4/Rpb7 protein, and the 3’ and 5’ mRNA degradation machinery was altered by the changes at the level of Rpb4/Rpb7 imprinted mRNA. This resulted in an increase in mRNA stability in the cell (Garrido-Godino et al., 2022). Additionally, depletion of POLR2D in zebrafish caused a reduction in the expression of housekeeping and zygotic genes (Maeta et al., 2020). During my investigation of nascent RNA synthesis, I found that RPRD proteins have a negative impact on transcription, although no changes were detected in the total RNA level. As it was obtained the most prominent effect from

RPRD2 deregulation, I chose to test only the RPRD2-depleted and overexpressed conditions with a long DRB treatment experiment to evaluate mRNA decay. RPRD2 depletion induced mRNA decay, while RPRD2 overexpression somehow blocked the decay pathway to maintain total RNA levels at a steady state. The interaction of RPRD proteins with POLR2D and POLR2G might affect these mRNA decay pathways. I suggest that RPRD proteins probably might inhibit the function of POLR2D and POLR2G by interacting with them in order to ensure mRNA homeostasis in the cell.

3. RPRD proteins triggers the development of cancer cells

It has been previously demonstrated that RPRD1B and RPRD2 are overexpressed in most cancer tissues, whereas RPRD1A is not (COSMIC, https://cancer.sanger.ac.uk) (Tate et al., 2019). However, my experimental findings indicate that the overexpression of RPRD1B, particularly RPRD2, reduces the level of newly synthesized RNA but not the steady-state RNA level in cells. This raises the question of how these proteins contribute to carcinogenesis if they act as negative regulators of transcription.

Cancer cells possess deregulated transcription factors that can cause significant changes in gene expression. Post-transcriptional regulation is important for many oncogenes, including growth factors and cell cycle regulatory proteins. Controlling the rate of mRNA turnover is a critical mechanism under this regulation (Benjamin & Moroni, 2007). When the half-life of a transcript is extended simultaneously with increased transcription, mRNA levels are further increased at a steady-state. Alternatively, stabilization and transcription may not be interconnected processes. In such cases, transcription does not lead to an accumulation of the total RNA level because mRNA is rapidly degraded. Thus, gene expression is controlled to a greater extent by regulating mRNA stability (Pittsburgh & This, 2007). When mRNAs are aberrantly stabilized, important carcinogenesis responses such as overinduction of growth factors and oncogenes may be prolonged in cells.

I have demonstrated that the overexpression of individual RPRD1B and RPRD2, but not RPRD1A, protein inhibits nascent transcription. Cells respond to this nascent RNA depletion by increasing the stability of pre-existing RNA. Non-responsiveness to external signals due to transcription silencing by RPRD proteins, which recruit

inhibitory factors, may accumulate over time in cells. These errors may turn out to be advantageous for cancer cells through the activation of other cancerous pathways.

In this low grade glioma dataset, I have demonstrated that upregulation of RPRD2 and RPRD1B is associated with increased aggressiveness of gliomas compared to unchanged RPRD samples. Furthermore, their upregulation resulted in changes in transcription related pathways, while their downregulation led to changes in mitochondrial pathways. The Warburg effect, first described by Otto Warburg in 1956, suggests that there is a correlation between glycolytic ATP production and tumor cell aggressiveness (Warburg, 1956). However, subsequent studies have challenged this assumption by revealing that tumor mitochondria can produce ATP through respiration (Weinhouse, 1976). In addition to ATP production, mitochondrial activities also produce reactive oxygen species (ROS), which can accumulate in cells and damage cellular physiology by oxidizing proteins, lipids, and nucleic acids. By activating the mitochondrial pathway, cells can revert their energy metabolism to a non-malignant state, which may increase tumor cells' susceptibility to apoptosis (Gogvadze et al., 2008; Solaini et al., 2011). Mitochondria are key players in response to hypoxia, nutrient depletion, or cancer treatments, and are involved in regulating cellular bioenergetics, oxidative response, and cell death.

In conclusion, my results suggest that RPRD proteins are involved in cancer aggressiveness, and that simultaneous upregulation or downregulation of all three RPRD proteins can lead to changes in both transcription and mitochondrial pathways.

4. RPRD protein levels, RPRD1B and RPRD2 but not RPRD1A, correlates with R-loop levels in cells.

R-loops consist of two-stranded DNA attached to one-stranded RNA, formed when nascent RNA hybridizes with transcribed DNA template. The transient annealing of nascent RNA with template DNA in the active site of RNA polymerase during transcription can create R-loops. Any defects in detecting or resolving R-loops can be hazardous. Collisions between transcription and replication machineries, resulting from R-loops, can hinder the progression of the replication fork. If not properly removed, this can cause DNA double-strand breaks (Chédin, 2016; Sollier & Cimprich, 2015). Overexpression of RPRD proteins can lead to slower transcription, resulting in the accumulation of R-loops due to the invading capacity of newly synthesized RNA behind the transcription machinery.

R-loops can indirectly induce the inhibition of repressive chromatin-modifying enzymes and recruitment of activating chromatin-remodelling complexes. However, R-loops can also block transcription-factor binding, although these results have not been generalized beyond a specific promoter locus (Belotserkovskii et al., 2017; Skourti-Stathaki & Proudfoot, 2014). Due to the significant increase of R-loops upon RPRD overexpression, RPRD proteins may interfere with transcription indirectly, and the accumulation of genomic instability can cause carcinogenesis over time.

5. RPRD proteins changes the cell cycle progression rate.

External or internal stimuli can induce or repress cell growth or division by altering the expression of cell cycle-related genes, stabilizing mRNA, and detecting errors in cells. The progression of the cell cycle is regulated by the activity of cyclins, CDKs, and CKI proteins. DNA damage detection leads to cell cycle arrest to allow DNA repair. In tumor development, several oncoproteins promote cyclin and CDK activity while inhibiting CKIs (Matthews et al., 2022).

RPRD1B was initially described as "cell-cycle-related and expression-elevated protein in a tumor (CREPT)." Its high expression in various human tumor samples and its ability to alter the expression level of cell cycle-related genes have been demonstrated conclusively. (Lu et al., 2012). RPRD1A was identified during the screening of P15INK4 regulator genes and was previously named p15RS (Liu et al., 2002). Numerous studies have shown that RPRD1A and RPRD1B affect the expression of several cell cycle-related genes and cell cycle progression (Li et al., 2021; Liu et al., 2016; Li et al., 2018; Lu et al., 2012; Wen et al., 2020; Zhang et al., 2019). However, these two proteins play opposite roles in regulating the cell cycle.

In this study, I have demonstrated that RPRD proteins play a crucial role in regulating nascent transcript and mRNA stability, leading to prolonged signal induction or unresponsiveness to internal or external signals. Moreover, the levels of RPRD1B and RPRD2, but not RPRD1A, have been shown to be correlated with R-loop formation in cells. These accumulated errors affect cell cycle progression over time and determine whether cells will repair or not. I have evaluated cell cycle progression after RPRD deregulation using FACS experiments. While RPRD1A, RPRD1B, and RPRD2

depletion arrested cells at G0/G1 phase, their upregulation did not produce a unique result. RPRD1B overexpression led to G2/M phase accumulation, while RPRD2 upregulation caused G0/G1 arrest. I further investigated the effect of RPRD proteins on cell cycle progression using DTB synchronization and found that RPRD proteins accelerate the cell cycle progression.

In conclusion, the findings of this study suggest that RPRD proteins play a critical role in regulating cell cycle progression by affecting nascent transcript and mRNA stability, which ultimately affects signal induction and response to internal or external stimuli. Furthermore, the levels of RPRD1B and RPRD2 are correlated with R-loop formation in cells, and their deregulation affects the cell cycle progression rate. The findings of this study provide new insights into the mechanisms underlying cell cycle regulation and may have important implications for the development of novel therapeutic strategies for diseases such as cancer.

Concluding Remarks

The goal of my project was to investigate the impact of RPRD (regulator of pre-mRNA-domain-containing) proteins on transcription and their association with cancer. My results suggest that while RPRD proteins do not affect the overall level of RNA, they do negatively regulate nascent transcription. I also found that RPRD protein levels and mRNA stability have opposite effects. Thus, RPRD proteins play a crucial role in balancing newly synthesized RNA with total RNA. I suggested that RPRD proteins serve as a docking site for recruiting other regulatory proteins to the transcription site through protein-protein interactions. Deregulation of RPRD proteins alters transcription rates, which can affect genome-wide R-loop formation, potentially leading to DNA damage accumulation. Abnormal expression of RPRD proteins may contribute to cellular transformation from a healthy state to a cancerous one.

REFERENCES

Adelman, K., & Lis, J. T. (2012). Promoter-proximal pausing of RNA polymerase II : emerging roles in metazoans. *Nature Reviews Genetics*, *13*. https://doi.org/10.1038/nrg3293

Aguilera, A., & García-Muse, T. (2012). R Loops: From Transcription Byproducts to Threats to Genome Stability. *Molecular Cell*, *46*(2), 115–124. https://doi.org/10.1016/J.MOLCEL.2012.04.009

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2014). *Molecular Biology of the Cell* (6th ed.). Garland Science.

Ali, I., Ruiz, D. G., Ni, Z., Johnson, J. R., Zhang, H., Li, P. C., Khalid, M. M., Conrad, R. J., Guo, X., Min, J., Greenblatt, J., Jacobson, M., Krogan, N. J., & Ott, M. (2019). Crosstalk between RNA Pol II C-Terminal Domain Acetylation and Phosphorylation via RPRD Proteins. *Molecular Cell*, *74*(6), 1164-1174.e4. https://doi.org/10.1016/j.molcel.2019.04.008

Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. In *Nature Reviews Genetics* (Vol. 21, Issue 2, pp. 71–87). Nature Research. https://doi.org/10.1038/s41576-019-0173-8

Barnum, K. J., & O'Connell, M. J. (2014). Cell Cycle Regulation by *Checkpoints. Methods Mol Biol., 1170, 29–40. https://doi.org/10.1007/978-1-4939-0888-2*

Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences*, *69*(21), 3613–3634. https://doi.org/10.1007/s00018-012-0990-9

Belotserkovskii, B. P., Shin, J. H. S., & Hanawalt, P. C. (2017). Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Research*, *45*(11), 6589–6599. https://doi.org/10.1093/nar/gkx403

Benjamin, D., & Moroni, C. (2007). mRNA stability and cancer : an emerging link ? *Expert Opin Biol Ther.*, 1515–1530. https://doi.org/doi: 10.1517/14712598.7.10.1515.

Bentley, D. L. (2014). Coupling mRNA processing with transcription in time and space. *Nature Reviews Genetics*, *15*(3), 163–175. https://doi.org/10.1038/nrg3662

Boellner, S., & Becker, K.-F. (2015). Reverse Phase Protein Arrays—Quantitative Assessment of Multiple Biomarkers in Biopsies for Clinical Use. *Microarrays*, *4*(2). https://doi.org/10.3390/microarrays4020098

Bradner, J. E., Hnisz, D., & Young, R. A. (2017). Transcriptional Addiction in Cancer. *Cell*, *168*(4), 629–643. https://doi.org/10.1016/j.cell.2016.12.013

Brat, D. J. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *New England Journal of Medicine*, *372*(26), 2481–2498. https://doi.org/10.1056/NEJMoa1402121

Brennan, C. W., Verhaak, R. G. W., McKenna, A., Campos, B., Noushmehr, H., Salama, S. R., Zheng, S., Chakravarty, D., Sanborn, J. Z., Berman, S. H., Beroukhim, R., Bernard, B., Wu, C. J., Genovese, G., Shmulevich, I., Barnholtz-Sloan, J., Zou, L., Vegesna, R., Shukla, S. A., … McLendon, R. (2013). The somatic genomic landscape of glioblastoma. *Cell*, *155*(2). https://doi.org/10.1016/j.cell.2013.09.034

Buratowski, S. (2003). The CTD code. *Nature Structural Biology*, *10*(9), 679–680.

Buratowski, S. (2009). Progression through the RNA Polymerase II CTD Cycle. *Molecular Cell*, *36*(4), 541–546. https://doi.org/10.1016/j.molcel.2009.10.019

Buratowski, S., & Zhou, H. (1993). Functional domains of transcription factor TFIIB. *Proceedings of the National Academy of Sciences*, *90*(12), 5633–5637. https://doi.org/10.1073/pnas.90.12.5633

Burley, S. K., & Roeder, R. G. (1996). BIOCHEMISTRY AND STRUCTURAL BIOLOGY OF TRANSCRIPTION FACTOR IID (TFIID). *Annual Review of Biochemistry*, *65*(1), 769–799. https://doi.org/10.1146/annurev.bi.65.070196.004005

Casamassimi, A., & Ciccodicola, A. (2019). Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. *International Journal of Molecular Sciences*, *20*(6), 1281. https://doi.org/10.3390/ijms20061281

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery*, *2*(5). https://doi.org/10.1158/2159-8290.CD-12-0095

Chédin, F. (2016). Nascent Connections: R-Loops and Chromatin Patterning. *Trends in Genetics*, *32*(12), 828–838. https://doi.org/10.1016/j.tig.2016.10.002

Chen, F. X., Woodfin, A. R., Gardini, A., Rickels, R. A., Marshall, S. A., Smith, E. R., Shiekhattar, R., & Shilatifard, A. (2015). PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. *Cell*, *162*(5), 1003–1015. https://doi.org/10.1016/j.cell.2015.07.042

Cheng, B., Li, T., Rahl, P. B., Adamson, T. E., Loudas, N. B., Guo, J., Varzavand, K., Cooper, J. J., Hu, X., Gnatt, A., Young, R. A., & Price, D. H. (2012). Functional association of Gdown1 with RNA polymerase II poisedpoised on human genes. *Mol Cell.*, *45*(1), 38–50. https://doi.org/10.1016/j.molcel.2011.10.022.Functional

Chen, X., Qi, Y., Wang, X., Wang, Z., Wang, L., Song, A., Li, J., Zhao, D., Zhang, H., Jin, Q., Chen, F. X., Key, S., Sciences, M., & Xu, Y. (2021). RPAP2 regulates a transcription initiation checkpoint by prohibiting assembly of preinitiation complex. *BioRxiv*, 2021.06.18.448918. http://biorxiv.org/content/early/2021/06/18/2021.06.18.448918.abstract

Choder, M. (2004). Rpb4 and Rpb7: Subunits of RNA polymerase II and beyond. In *Trends in Biochemical Sciences* (Vol. 29, Issue 12). https://doi.org/10.1016/j.tibs.2004.10.007

Choder, M. (2011). mRNA imprinting. *Cellular Logistics*, *1*(1). https://doi.org/10.4161/cl.1.1.14465

Clevers, H., & Nusse, R. (2012). *Review Wnt / b -Catenin Signaling and Disease*. https://doi.org/10.1016/j.cell.2012.05.012

Cohen, S., Puget, N., Lin, Y. L., Clouaire, T., Aguirrebengoa, M., Rocher, V., Pasero, P., Canitrot, Y., & Legube, G. (2018). Senataxin resolves RNA:DNA hybrids forming at DNA double-strand breaks to prevent translocations. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-02894-w

Connelly, S., & Manley, J. L. (1988). A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & Development*, *2*(4), 440–452. https://doi.org/10.1101/gad.2.4.440

Corden, J. L., Cadena, D. L., Ahearn, J. M., & Dahmus, M. E. (1985). A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proceedings of the National Academy of Sciences*, *82*(23), 7934–7938. https://doi.org/10.1073/pnas.82.23.7934

Core, L., & Adelman, K. (2019). *Promoter-proximal pausing of RNA polymerase II : a nexus of gene regulation*. 960–982. https://doi.org/10.1101/gad.325142.119.and

Cramer, P. (2004). *RNA polymerase II structure : from core to functional complexes*. *Pol II*. https://doi.org/10.1016/j.gde.2004.01.003

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*, 45–54. https://doi.org/10.1038/s41586-019-1517-4

Cramer, P., Armache, K.-J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G. E., Dengl, S., Geiger, S. R., Jasiak, A. J., Jawhari, A., Jennebach, S., Kamenski, T., Kettenberger, H., Kuhn, C.-D., Lehmann, E., Leike, K., Sydow, J. F., & Vannini, A. (2008). Structure of Eukaryotic RNA Polymerases. *Annual Review of Biophysics*, *37*(1), 337–352. https://doi.org/10.1146/annurev.biophys.37.032807.130008

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Crossley, M. P., Bocek, M. J., Hamperl, S., Swigut, T., & Cimprich, K. A. (2020). QDRIP: A method to quantitatively assess RNA-DNA hybrid formation genome-wide. *Nucleic Acids Research*, *48*(14), e84. https://doi.org/10.1093/nar/gkaa500

Davis, M. A. M., Guo, J., Price, D. H., & Luse, D. S. (2014). Functional Interactions of the RNA Polymerase II-interacting Proteins Gdown1 and TFIIF *. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, *289*(16), 11143–11152. https://doi.org/10.1074/jbc.M113.544395

De Magis, A., Manzo, S. G., Russo, M., Marinello, J., Morigi, R., Sordet, O., & Capranico, G. (2019). DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proceedings of the National Academy of Sciences*, *116*(3), 816–825. https://doi.org/10.1073/pnas.1810409116

Ding, L., Yang, L., He, Y., Zhu, B., Ren, F., Fan, X., Wang, Y., Li, M., Li, J., Kuang, Y., Liu, S., Zhai, W., Ma, D., Ju, Y., Liu, Q., Jia, B., & Sheng, J. (2018). CREPT / RPRD1B associates with Aurora B to regulate Cyclin B1 expression for accelerating the G2 / M transition in gastric cancer. *Cell Death and Disease*. https://doi.org/10.1038/s41419-018-1211-8

Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P., & Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome Research*, *12*(1), 47–56. https://doi.org/10.1101/gr.203201

Eaton, J. D., Francis, L., Davidson, L., & West, S. (2020). A unified allosteric/torpedo mechanism for transcriptional termination on human protein-coding genes. *Genes & Development*, *34*(1–2), 132–145. https://doi.org/10.1101/gad.332833.119

Eaton, J. D., & West, S. (2020). Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics*, *36*(9), 664–675. https://doi.org/10.1016/j.tig.2020.05.008

Egloff, S., & Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends in Genetics*, *24*(6), 280–288. https://doi.org/10.1016/j.tig.2008.03.008

Egloff, S., Zaborowska, J., Laitem, C., Kiss, T., & Murphy, S. (2012). Ser7 Phosphorylation of the CTD Recruits the RPAP2 Ser5 Phosphatase to snRNA Genes. *Molecular Cell*, *45*, 111–122. https://doi.org/10.1016/j.molcel.2011.11.006

Eick, D., & Geyer, M. (2013). The RNA Polymerase II Carboxy-Terminal Domain (CTD) Code. *Chem. Rev.*, *113*(11), 8456–8490. https://doi.org/dx.doi.org/10.1021/cr400071f

El Hage, A., French, S. L., Beyer, A. L., & Tollervey, D. (2010). Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes and Development*, *24*(14). https://doi.org/10.1101/gad.573310

Erickson, B., Sheridan, R. M., Cortazar, M., & Bentley, D. L. (2018). Dynamic turnover of paused Pol II complexes at human promoters. *Genes & Development*, *32*(17–18), 1215–1225. https://doi.org/10.1101/gad.316810.118

Fan, X., Zhao, J., Ren, F., Wang, Y., Feng, Y., Ding, L., Zhao, L., Shang, Y., Li, J., Ni, J., Jia, B., Liu, Y., & Chang, Z. (2018). Dimerization of p15RS mediated by a leucine zipper–like motif is critical for its inhibitory role on Wnt signaling. *Journal of Biological Chemistry*, *293*(20), 7618–7628. https://doi.org/10.1074/jbc.RA118.001969

Fournier, L.-A., Kumar, A., Smith, T., Su, E., Moksa, M., Hirst, M., & Stirling, P. C. (n.d.). *Predicting R-loop regulators Global prediction of candidate R-loop binding and R-loop regulatory proteins*. https://doi.org/10.1101/2021.08.09.454968

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). *Defining mechanisms that regulate RNA polymerase II transcription in vivo*. *461*(September). https://doi.org/10.1038/nature08449

FURUICHI, Y. (2015). Discovery of m7G-cap in eukaryotic mRNAs. *Proceedings of the Japan Academy, Series B*, *91*(8), 394–409. https://doi.org/10.2183/pjab.91.394

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., & Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling*, *6*(269). https://doi.org/10.1126/scisignal.2004088

Garneau, N. L., Wilusz, J., & Wilusz, C. J. (2007). *The highways and byways of mRNA decay*. *8*(February), 113–126. https://doi.org/10.1038/nrm2104

Garrido-Godino, A. I., Cuevas-Bermúdez, A., Gutiérrez-Santiago, F., Mota-Trujillo, M. D. C., & Navarro, F. (2022). The Association of Rpb4 with RNA Polymerase II Depends on CTD Ser5P Phosphatase Rtr1 and Influences mRNA Decay in Saccharomyces cerevisiae. *International Journal of Molecular Sciences*, *23*(4). https://doi.org/10.3390/ijms23042002

Gibbons, J. M., Marno, K. M., Pike, R., Lee, W. J., Jones, C. E., Ogunkolade, B. W., Pardieu, C., Bryan, A., Fu, M., Warnes, G., Rowley, P. A., & Sloan, R. D. (2020). HIV-1 Accessory Protein Vpr Interacts with REAF/RPRD2 To Mitigate Its Antiviral Activity. *Journal of Virology*, *94*, 1–14.

Gilmour, D. S., & Lis, J. T. (1986). RNA Polymerase II Interacts with the Promoter Region of the Noninduced *hsp70* Gene in *Drosophila melanogaster* Cells. *Molecular and Cellular Biology*, *6*(11), 3984–3989. https://doi.org/10.1128/mcb.6.11.3984-3989.1986

Ginno, P. A., Lott, P. L., Christensen, H. C., & Korf, I. (2012). *Article R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters*. https://doi.org/10.1016/j.molcel.2012.01.017

Glover-Cutter, K., Kim, S., Espinosa, J., & Bentley, D. L. (2008). RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes. *Nature Structural & Molecular Biology*, *15*(1), 71–78. https://doi.org/10.1038/nsmb1352

Gogvadze, V., Orrenius, S., & Zhivotovsky, B. (2008). Mitochondria in cancer cells: what is so special about them? In *Trends in Cell Biology* (Vol. 18, Issue 4, pp. 165–173). https://doi.org/10.1016/j.tcb.2008.01.006

Gökbuget, D., & Blelloch, R. (2019). Epigenetic control of transcriptional regulation in pluripotency and early differentiation. *Development*, *146*(19). https://doi.org/10.1242/dev.164772

Graham, F. L., Smiley, J., Russell, W. C., & Nairn, R. (1977). Characteristics of a Human Cell Line Transformed by D N A from Human Adenovirus Type 5. *Journal of General Virology*, *36*(1), 59–72.

Haberle, V., & Stark, A. (2018a). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Haberle, V., & Stark, A. (2018b). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, *19*(10), 621–637. https://doi.org/10.1038/s41580-018-0028-8

Hahn, S. (1998). The Role of TAFs in RNA Polymerase II Transcription. *Cell*, *95*(5), 579–582. https://doi.org/10.1016/S0092-8674(00)81625-6

Hamadeh, Z., & Lansdorp, P. (2020). *RECQL5 at the Intersection of Replication and Transcription*. *8*(May), 1–8. https://doi.org/10.3389/fcell.2020.00324

Harlen, K. M., & Churchman, L. S. (2017). The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology*, *18*(4), 263–273. https://doi.org/10.1038/nrm.2017.10

Harper, J. V. (2005). Synchronization of Cell Populations in G 1 / S and G 2 / M Phases of the Cell Cycle. In *Humphrey T., Brooks G. (eds) Cell Cycle Control. Methods in Molecular Biology™* (Vol. 296, Issue 1, pp. 157–166). https://doi.org/https://doi.org/10.1385/1-59259-857-9:157

Hegazy, Y. A., Fernando, C. M., & Tran, E. J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. *Journal of Biological Chemistry*, *295*(4), 905–913. https://doi.org/10.1016/S0021-9258(17)49903-0

Hein, M. Y., Hubner, N. C., Poser, I., Buchholz, F., Hyman, A. A., Hein, M. Y., Hubner, N. C., Poser, I., Buchholz, F., Hyman, A. A., & Mann, M. (2015). A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Article A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell*, *163*, 712–723. https://doi.org/10.1016/j.cell.2015.09.053

Hsin, J., & Manley, J. L. (2012). *The RNA polymerase II CTD coordinates transcription and RNA processing*. 2119–2137. https://doi.org/10.1101/gad.200303.112.Transcription

Jaehning, J. A. (2010). The Paf1 complex: Platform or player in RNA polymerase II transcription? *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1799*(5–6), 379–388. https://doi.org/10.1016/j.bbagrm.2010.01.001

Jiang, J., Yang, X., He, X., Ma, W., Wang, J., Zhou, Q., Li, M., & Yu, S. (2019). MicroRNA-449b-5p suppresses the growth and invasion of breast cancer cells via inhibiting CREPT-mediated Wnt/β-catenin signaling. *Chemico-Biological Interactions*, *302*(February), 74–82. https://doi.org/10.1016/j.cbi.2019.02.004

Jin, K., Chen, H., Zuo, Q., Huang, C., Zhao, R., Yu, X., Wang, Y., Zhang, Y., Chang, Z., & Li, B. (2018). CREPT and p15RS regulate cell proliferation and cycling in chicken DF-1 cells through the Wnt/β-catenin pathway. *Journal of Cellular Biochemistry*, *119*(1), 1083–1092. https://doi.org/10.1002/jcb.26277

Jishage, M., Malik, S., Wagner, U., Uberheide, B., Ishihama, Y., Hu, X., Chait, B. T., Gnatt, A., Ren, B., & Roeder, R. G. (2012). Transcriptional Regulation by Pol II (G) Involving Mediator and Competitive Interactions of Gdown1 and TFIIF with Pol II. *Molecular Cell*, *45*(1), 51–63. https://doi.org/10.1016/j.molcel.2011.12.014

Kalsotra, A., & Cooper, T. A. (2011). Functional consequences of developmentally regulated alternative splicing. *Nature Reviews Genetics*, *12*(10), 715–729. https://doi.org/10.1038/nrg3052

Kim, A., & Wang, G. G. (2021). R-loop and its functions at the regulatory interfaces between transcription and (epi)genome. In *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* (Vol. 1864, Issues 11–12). Elsevier B.V. https://doi.org/10.1016/j.bbagrm.2021.194750

Kim, B., Nesvizhskii, A. I., Rani, P. G., Hahn, S., Aebersold, R., & Ranish, J. A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences*, *104*(41), 16068–16073. https://doi.org/10.1073/pnas.0704573104

Kinner, A., Wu, W., Staudt, C., & Iliakis, G. (2008). Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Research*, *36*(17), 5678–5694. https://doi.org/10.1093/nar/gkn550

Krebs, A. R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., & Schübeler, D. (2017). Genome-wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. *Molecular Cell*, *67*(3), 411-422.e4. https://doi.org/10.1016/j.molcel.2017.06.027

Kumar, A., Fournier, L.-A., & Stirling, P. C. (2022). Integrative analysis and prediction of human R-loop binding proteins. *G3 Genes|Genomes|Genetics*, *12*(8). https://doi.org/10.1093/g3journal/jkac142

Laitem, C., Zaborowska, J., Isa, N. F., Kufs, J., Dienstbier, M., & Murphy, S. (2015). CDK9 inhibitors define elongation checkpoints at both ends of RNA polymerase II-transcribed genes. *Nature Structural and Molecular Biology*, *22*(5), 396–403. https://doi.org/10.1038/nsmb.3000

Lee, T. I., & Young, R. A. (2013). Review Transcriptional Regulation and Its Misregulation in Disease. *Cell*, *152*(6), 1237–1251. https://doi.org/10.1016/j.cell.2013.02.014

Li, B., Carey, M., & Workman, J. L. (2007). The Role of Chromatin during Transcription. *Cell*, *128*(4), 707–719. https://doi.org/10.1016/j.cell.2007.01.015

Li, M., Ma, D., & Chang, Z. (2021). Current understanding of CREPT and p15RS, carboxy-terminal domain (CTD)-interacting proteins, in human cancers. *Oncogene*, *40*(4), 705–716. https://doi.org/10.1038/s41388-020-01544-0

Lim, S., & Kaldis, P. (2013). Cdks, cyclins and CKIs: Roles beyond cell cycle regulation. In *Development (Cambridge)* (Vol. 140, Issue 15, pp. 3079–3093). https://doi.org/10.1242/dev.091744

Lin, Y. C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van De Peer, Y., Tavernier, J., & Callewaert, N. (2014). Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature Communications*, *5*(11). https://doi.org/10.1038/ncomms5767

Liu, C., Zhang, Y., Li, J., Wang, Y., Ren, F., Zhou, Y., Wu, Y., Feng, Y., Zhou, Y., Su, F., Jia, B., Wang, D., & Chang, Z. (2015). P15RS/RPRD1A (p15INK4b-related sequence/regulation of nuclear pre-mRNA domain-containing protein 1A) interacts with HDAC2 in inhibition of the Wnt/β-catenin signaling pathway. *Journal of Biological Chemistry*, *290*(15), 9701–9713. https://doi.org/10.1074/jbc.M114.620872

Liu, J., Liu, H., Zhang, X., Gao, P., Wang, J., & Hu, Z. (2002). Identification and characterization of P15RS, a novel P15. *Biochemical and Biophysical Research Communications*, *299*, 880–885. https://doi.org/https://doi.org/10.1016/S0006-291X(02)02684-0

Liu, P., Kenney, J. M., Stiller, J. W., & Greenleaf, A. L. (2010). Genetic Organization, Length Conservation, and Evolution of RNA Polymerase II Carboxyl-Terminal Domain. *Molecular Biology and Evolution*, *27*(11), 2628–2641. https://doi.org/10.1093/molbev/msq151

Liu, T., Li, W. M., Wang, W. P., Sun, Y., Ni, Y. F., Xing, H., Xia, J. H., Wang, X. J., Zhang, Z. P., & Li, X. F. (2016). Inhibiting CREPT reduces the proliferation and migration of non-small cell lung cancer cells by down-regulating cell cycle related protein. *American Journal of Translational Research*, *8*(5), 2097–2113.

Li, W., Zheng, G., Xia, J., Yang, G., Sun, J., Wang, X., Wen, M., Sun, Y., Zhang, Z., & Jin, F. (2018). Cell cycle-related and expression-elevated protein in tumor overexpression is associated with proliferation behaviors and poor prognosis in non-small-cell lung cancer. *Cancer Science*, *109*(4), 1012–1023. https://doi.org/10.1111/cas.13524

Lockhart, A., Pires, V. B., Bento, F., Kellner, V., Luke-Glaser, S., Yakoub, G., Ulrich, H. D., & Luke, B. (2019). RNase H1 and H2 Are Differentially Regulated to Process RNA-DNA Hybrids. *Cell Reports*, *29*(9), 2890-2900.e5. https://doi.org/10.1016/j.celrep.2019.10.108

Logan, J., Falck-Pedersen, E., Darnell, J. E., & Shenk, T. (1987). A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences*, *84*(23), 8306–8310. https://doi.org/10.1073/pnas.84.23.8306

Lotan, R., Bar-On, V. G., Harel-Sharvit, L., Duek, L., Melamed, D., & Choder, M. (2005). The RNA polymerase II subunit Rpb4p mediates decay of a specific class of mRNAs. *Genes and Development*, *19*(24). https://doi.org/10.1101/gad.353205

Lotan, R., Goler-Baron, V., Duek, L., Haimovich, G., & Choder, M. (2007). The Rpb7p subunit of yeast RNA polymerase II plays roles in the two major cytoplasmic mRNA decay mechanisms. *Journal of Cell Biology*, *178*(7). https://doi.org/10.1083/jcb.200701165

Lu, D., Wu, Y., Wang, Y., Ren, F., Wang, D., Su, F., Zhang, Y., Yang, X., Jin, G., Hao, X., He, D., Zhai, Y., Irwin, D. M., Hu, J., Sung, J. J. Y., Yu, J., Jia, B., & Chang, Z. (2012). Crept accelerates tumorigenesis by regulating the transcription of cell-cycle-related genes. *Cancer Cell*, *21*(1), 92–104. https://doi.org/10.1016/j.ccr.2011.12.016

Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F., Mutschler, H., Buratowski, S., Meinhart, A., & Varani, G. (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nature Structural & Molecular Biology*, *17*(10), 1195–1201. https://doi.org/10.1038/nsmb.1893

Maeta, M., Kataoka, M., Nishiya, Y., Ogino, K., Kashima, M., & Hirata, H. (2020). RNA polymerase II subunit D is essential for zebrafish development. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-70110-1

Marabitti, V., Lillo, G., Malacaria, E., Palermo, V., Sanchez, M., Pichierri, P., & Franchitto, A. (2019). ATM pathway activation limits R-loop-associated genomic instability in Werner syndrome cells. *Nucleic Acids Research*, *47*(7). https://doi.org/10.1093/nar/gkz025

Marno, K. M., Ogunkolade, B. W., Pade, C., Oliveira, N. M. M., Sullivan, E. O., & Mcknight, Á. (2014). *Novel restriction factor RNA-associated early-stage anti-viral factor (REAF) inhibits human and simian immunodeficiency viruses*. *2*, 1–9.

Marno, K. M., Sullivan, E. O., Jones, C. E., Díaz-delfín, J., Pardieu, C., Sloan, R. D., & Mcknight, Á. (2017). RNA-Associated Early-Stage Antiviral Factor Is a Major Component of Lv2 Restriction. *Journal of Virology*, *91*, 1–1.

Matthews, H. K., Bertoli, C., & de Bruin, R. A. M. (2022). Cell cycle control in cancer. In *Nature Reviews Molecular Cell Biology* (Vol. 23, Issue 1, pp. 74–88). Nature Research. https://doi.org/10.1038/s41580-021-00404-3

Mei, K. R., Jin, Z., Ren, F. L., Wang, Y. Y., Chang, Z. J., & Wang, X. Q. (2014). Structural basis for the recognition of RNA polymerase II C-terminal domain by CREPT and p15RS. *Science China Life Sciences*, *57*(1), 97–106. https://doi.org/10.1007/s11427-013-4589-7

Mischo, H. E., Gómez-González, B., Grzechnik, P., Rondón, A. G., Wei, W., Steinmetz, L., Aguilera, A., & Proudfoot, N. J. (2011). Yeast Sen1 Helicase Protects the Genome from Transcription-Associated Instability. *Molecular Cell*, *41*(1), 21–32. https://doi.org/10.1016/j.molcel.2010.12.007

Morales, J. C., Richard, P., Rommel, A., Fattah, F. J., Motea, E. A., Patidar, P. L., Xiao, L., Leskov, K., Wu, S. Y., Hittelman, W. N., Chiang, C. M., Manley, J. L., & Boothman, D. A. (2014). Kub5-Hera, the human Rtt103 homolog, plays dual functional roles in transcription termination and DNA repair. *Nucleic Acids Research*, *42*(8), 4996–5006. https://doi.org/10.1093/nar/gku160

Mosler, T., Conte, F., Longo, G. M. C., Mikicic, I., Kreim, N., Möckel, M. M., Petrosino, G., Flach, J., Barau, J., Luke, B., Roukos, V., & Beli, P. (2021). R-loop proximity proteomics identifies a role of DDX41 in transcription-associated genomic instability. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-27530-y

Mosley, A. L., Pattenden, S. G., Carey, M., Venkatesh, S., Gilmore, J. M., Florens, L., Workman, J. L., & Washburn, M. P. (2009). Rtr1 Is a CTD Phosphatase that Regulates RNA Polymerase II during the Transition from Serine 5 to Serine 2 Phosphorylation. *Molecular Cell*, *34*(2), 168–178. https://doi.org/10.1016/j.molcel.2009.02.025

Nechaev, S., & Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1809*(1), 34–45. https://doi.org/10.1016/j.bbagrm.2010.11.001

Niehrs, C., & Luke, B. (2020). Regulatory R-loops as facilitators of gene expression and genome stability. *Nature Reviews Molecular Cell Biology*, *21*(3), 167–178. https://doi.org/10.1038/s41580-019-0206-3

Ni, Z., Olsen, J. B., Guo, X., Zhong, G., Ruan, E. D., Marcon, E., Young, P., Guo, H., Li, J., Moffat, J., Emili, A., & Greenblatt, J. F. (2011). Control of the RNA polymerase II phosphorylation state in promoter regions by CTD interaction domain-containing proteins RPRD1A and RPRD1B. *Transcription*, *2*(5), 237–242. https://doi.org/10.4161/trns.2.5.17803

Ni, Z., Xu, C., Guo, X., Hunter, G. O., Kuznetsova, O. V., Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W. H. W., Li, J., Young, P., Olsen, J. B., Wan, C., Loppnau, P., El Bakkouri, M., Senisterra, G. A., He, H., Huang, H., … Greenblatt, J. F. (2014). RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation. *Nature Structural and Molecular Biology*, *21*(8), 686–695. https://doi.org/10.1038/nsmb.2853

Ni, Z., Xu, C., Guo, X., Hunter, G. O., Kuznetsova, O. V, Tempel, W., Marcon, E., Zhong, G., Guo, H., Kuo, W. W., Li, J., Young, P., Olsen, J. B., Wan, C., Loppnau, P., Bakkouri, M. El, Senisterra, G. A., He, H., Huang, H., … Greenblatt, J. F. (2014). *RPRD1A and RPRD1B are human RNA polymerase II C-terminal domain scaffolds for Ser5 dephosphorylation*. *21*(8). https://doi.org/10.1038/nsmb.2853

Pan, M. R., Peng, G., Hungs, W. C., & Lin, S. Y. (2011). Monoubiquitination of H2AX protein regulates DNA damage response signaling. *Journal of Biological Chemistry*, *286*(32). https://doi.org/10.1074/jbc.M111.256297

Peterlin, B. M., & Price, D. H. (2006). Controlling the Elongation Phase of Transcription with P-TEFb. *Molecular Cell*, *23*(3), 297–305. https://doi.org/10.1016/j.molcel.2006.06.014

Petrenko, N., Jin, Y., Wong, K. H., & Struhl, K. (2016). Mediator Undergoes a Compositional Change during Transcriptional Activation. *Molecular Cell*, *64*(3), 443–454. https://doi.org/10.1016/j.molcel.2016.09.015

Phillips, D. D., Garboczi, D. N., Singh, K., Hu, Z., Leppla, S. H., & Leysath, C. E. (2013). The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *Journal of Molecular Recognition*, *26*(8), 376–381. https://doi.org/10.1002/jmr.2284

Pineda, G., Shen, Z., de Albuquerque, C. P., Reynoso, E., Chen, J., Tu, C.-C., Tang, W., Briggs, S., Zhou, H., & Wang, J. Y. J. (2015). Proteomics studies of the interactome of RNA polymerase II C-terminal repeated domain. *BMC Research Notes*, *8*(1), 616. https://doi.org/10.1186/s13104-015-1569-y

Pittsburgh, M., & This, I. (2007). mRNA stability control : a clandestine force in normal and malignant hematopoiesis. *Leukemia*, *21*, 1158–1171. https://doi.org/10.1038/sj.leu.2404656

Ramanathan, A., Robb, G. B., & Chan, S.-H. (2016). mRNA capping: biological functions and applications. *Nucleic Acids Research*, *44*(16), 7511–7526. https://doi.org/10.1093/nar/gkw551

Ren, L., Chen, H., Song, J., Chen, X., Lin, C., Zhang, X., & Hou, N. (2019). MiR-454-3p-Mediated Wnt / β -catenin Signaling Antagonists Suppression Promotes Breast Cancer Metastasis. *Theranostics*, *9*(2), 449–465. https://doi.org/10.7150/thno.29055

Rhind, N., & Russell, P. (2012). *Signaling Pathways that Regulate Cell Division*. 1–15.

Rinaldi, C., Pizzul, P., Longhese, M. P., & Bonetti, D. (2021). Sensing R-Loop-Associated DNA Damage to Safeguard Genome Stability. In *Frontiers in Cell and Developmental Biology* (Vol. 8). https://doi.org/10.3389/fcell.2020.618157

Roberts, R. W., & Crothers, D. M. (1992). Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition. *Science*, *258*(5087), 1463–1466. https://doi.org/10.1126/science.1279808

ROEDER, R. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends in Biochemical Sciences*, *21*(9), 327–335. https://doi.org/10.1016/S0968-0004(96)10050-5

Rogakou, E. P., Boon, C., Redon, C., & Bonner, W. M. (1999). Megabase Chromatin Domains Involved in DNA Double-Strand Breaks in Vivo. *Journal of Cell Biology*, *146*(5), 905–916. https://doi.org/10.1083/jcb.146.5.905

Rosonina, E., Kaneko, S., & Manley L., J. (2006). Terminating the transcript: breaking up is hard to do. *GENES & DEVELOPMENT*, *20*, 1050–1056. https://doi.org/10.1101/gad.1431606.2

Sanz, L. A., Hartono, S. R., Lim, Y. W., Ginno, P. A., Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., & Ginno, P. A. (2016). *Structures Associate with Specific Epigenomic Signatures in Mammals Prevalent , Dynamic , and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals*. 167–178. https://doi.org/10.1016/j.molcel.2016.05.032

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G. P., Heron, M., & Williams, H. (2014). RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress. *Cell*, *157*, 1037–1049. https://doi.org/10.1016/j.cell.2014.03.048

Singh, J., & Padgett, R. A. (2009). Rates of in situ transcription and splicing in large human genes. *Nature Structural & Molecular Biology*, *16*(11). https://doi.org/10.1038/nsmb.1666

Skourti-Stathaki, K., & Proudfoot, N. J. (2014). A double-edged sword: R loops as threats to genome integrity and powerful regulators of gene expression. *Genes and Development*, *28*(13), 1384–1396. https://doi.org/10.1101/GAD.242990.114

Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human Senataxin Resolves RNA/DNA Hybrids Formed at Transcriptional Pause Sites to Promote Xrn2-Dependent Termination. *Molecular Cell*, *42*(6), 794–805. https://doi.org/10.1016/j.molcel.2011.04.026

Smolka, J. A., Sanz, L. A., Hartono, S. R., & Chédin, F. (2021). Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *Journal of Cell Biology*, *220*(6). https://doi.org/10.1083/jcb.202004079

Solaini, G., Sgarbi, G., & Baracca, A. (2011). Oxidative phosphorylation in cancer cells. In *Biochimica et Biophysica Acta - Bioenergetics* (Vol. 1807, Issue 6, pp. 534–542). https://doi.org/10.1016/j.bbabio.2010.09.003

Sollier, J., & Cimprich, K. A. (2015). Breaking bad: R-loops and genome integrity. In *Trends in Cell Biology* (Vol. 25, Issue 9, pp. 514–522). Elsevier Ltd. https://doi.org/10.1016/j.tcb.2015.05.003

Stein, H., & Hausen, P. (1969). Enzyme from Calf Thymus Degrading the RNA Moiety of DNA-RNA Hybrids: Effect on DNA-Dependent RNA Polymerase. *Science*, *166*(3903), 393–395. https://doi.org/10.1126/science.166.3903.393

Struhl, K. (1995). YEAST TRANSCRIPTIONAL REGULATORY MECHANISMS. *Annual Review of Genetics*, *29*(1), 651–674. https://doi.org/10.1146/annurev.ge.29.120195.003251

Suh, H., Ficarro, S. B., Kang, U.-B., Chun, Y., Marto, J. A., & Buratowski, S. (2016). Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. *Molecular Cell*, *61*(2), 297–304. https://doi.org/10.1016/j.molcel.2015.12.021

Sun, M., Si, G., Sun, H. S., & Si, F. C. (2018). Inhibition of CREPT restrains gastric cancer growth by regulation of cycle arrest, migration and apoptosis via ROS-regulated p53 pathway. *Biochemical and Biophysical Research Communications*, *496*(4), 1183–1190. https://doi.org/10.1016/j.bbrc.2018.01.167

SVEJSTRUP, J., VICHI, P., & EGLY, J. (1996). The multiple roles of transcription/repair factor TFIIH. *Trends in Biochemical Sciences*, *21*(9), 346–350. https://doi.org/10.1016/S0968-0004(96)10046-3

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., … Forbes, S. A. (2019). COSMIC : the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*, 941–947. https://doi.org/10.1093/nar/gky1015

Turowski, T. W., & Boguta, M. (2021). Specific Features of RNA Polymerases I and III: Structure and Assembly. *Frontiers in Molecular Biosciences*, *8*(May), 1–8. https://doi.org/10.3389/fmolb.2021.680090

Usheva, A., Maldonado, E., Goldring, A., Lu, H., Houbavi, C., Reinberg, D., & Aloni, Y. (1992). Specific interaction between the nonphosphorylated form of RNA polymerase II and the TATA-binding protein. *Cell*, *69*(5), 871–881. https://doi.org/10.1016/0092-8674(92)90297-P

Vermeulen, K., Bockstaele, D. R. Van, & Berneman, Z. N. (2003). The cell cycle : a review of regulation , deregulation and therapeutic targets in cancer. *Cell Prolif*, *36*, 131–149.

Wahba, L., Amon, J. D., Koshland, D., & Vuica-Ross, M. (2011). RNase H and Multiple RNA Biogenesis Factors Cooperate to Prevent RNA:DNA Hybrids from Generating Genome Instability. *Molecular Cell*, *44*(6), 978–988. https://doi.org/10.1016/j.molcel.2011.10.017

Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, *136*(4), 701–718. https://doi.org/10.1016/j.cell.2009.02.009

Wang, I. X., Grunseich, C., Fox, J., Burdick, J., Zhu, Z., Ravazian, N., Hafner, M., & Cheung, V. G. (2018). Human proteins that interact with RNA/DNA hybrids. *Genome Research*, *28*(9), 1405–1414. https://doi.org/10.1101/gr.237362.118

Wang, X., Qi, Y., Wang, Z., Wang, L., Song, A., Tao, B., Li, J., Zhao, D., Zhang, H., Jin, Q., Jiang, Y.-Z., Chen, F. X., Xu, Y., & Chen, X. (2022). RPAP2 regulates a transcription initiation checkpoint by inhibiting assembly of pre-initiation complex. *Cell Reports*, *39*(4), 110732. https://doi.org/10.1016/j.celrep.2022.110732

Wang, Y., Qiu, H., Hu, W., Li, S., & Yu, J. (2014). RPRD1B promotes tumor growth by accelerating the cell cycle in endometrial cancer. *Oncology Reports*, *31*(3), 1389–1395. https://doi.org/10.3892/or.2014.2990

Warburg, O. (1956). On the Origin of Cancer Cells. *Science*, *123*(3191), 309–314. https://doi.org/10.1126/science.123.3191.309

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losick, R. (2015). *Molecular Biology of The Gene* (7th ed.). Pearson/Benjamin Cummings.

Wei, M., Cao, Y., Jia, D., Zhao, H., & Zhang, L. (2019). CREPT promotes glioma cell proliferation and invasion by activating Wnt/β-catenin pathway and is a novel target of microRNA-596. *Biochimie*, *162*, 116–124. https://doi.org/10.1016/j.biochi.2019.04.014

Weinberg, R. A. (2013). *The Biology of Cancer* (2nd ed.). Garland Science.

Weinhouse, S. (1976). The Warburg hypothesis fifty years later. In *Zeitschrift für Krebsforschung und Klinische Onkologie* (Vol. 87, Issue 2). https://doi.org/10.1007/BF00284370

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Sander, C., Stuart, J. M., Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chu, A., … Kling, T. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

Wen, N., Bian, L., Gong, J., & Meng, Y. (2020). Overexpression of cell-cycle related and expression-elevated protein in tumor (CREPT) in malignant cervical cancer. *Journal of International Medical Research*, *48*(1). https://doi.org/10.1177/0300060519895089

Winczura, K., Ceylan, H., Sledziowska, M., Jones, M., Fagarasan, H., Wang, J., Saponaro, M., Arnold, R., Hebenstreit, D., & Grzechnik, P. (2021). Rprd proteins control transcription in human cells. *BioRxiv Preprint*, 1–41. https://doi.org/https://doi.org/10.1101/2021.06.20.449126 ; this

Wu, Y. M., Chang, J. W., Wang, C. H., Lin, Y. C., Wu, P. L., Huang, S. H., Chang, C. C., Hu, X., Gnatt, A., & Chang, W. H. (2012). Regulation of mammalian transcription by Gdown1 through a novel steric crosstalk revealed by cryo-EM. *EMBO Journal*, *31*(17), 3575–3587. https://doi.org/10.1038/emboj.2012.205

Wu, Y., Zhang, Y., Zhang, H., Yang, X., Wang, Y., Ren, F., Liu, H., Zhai, Y., Jia, B., Yu, J., & Chang, Z. (2010). p15RS Attenuates Wnt / β-Catenin Signaling by Disrupting β-Catenin-TCF4 Interaction. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, *285*(45), 34621–34631. https://doi.org/10.1074/jbc.M110.148791

Yamaguchi, Y., Takagi, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J., & Handa, H. (1999). NELF, a Multisubunit Complex Containing RD, Cooperates with DSIF to Repress RNA Polymerase II Elongation. *Cell*, *97*(1), 41–51. https://doi.org/10.1016/S0092-8674(00)80713-8

Yao, R. W., Wang, Y., & Chen, L. L. (2019). Cellular functions of long noncoding RNAs. *Nature Cell Biology*, *21*(5), 542–551. https://doi.org/10.1038/s41556-019-0311-8

Zhang, Y., Liu, C., Duan, X., Ren, F., Li, S., Jin, Z., Wang, Y., Feng, Y., Liu, Z., & Chang, Z. (2014). CREPT/RPRD1B, a recently identified novel protein highly expressed in tumors, enhances the β-catenin·TCF4 transcriptional activity in response to Wnt signaling. *Journal of Biological Chemistry*, *289*(33), 22589–22599. https://doi.org/10.1074/jbc.M114.560979

Zhang, Y., Wang, S., Kang, W., Liu, C., Dong, Y., Ren, F., Wang, Y., Zhang, J., Wang, G., To, K. F., Zhang, X., Sung, J. J., Chang, Z., & Yu, J. (2018). CREPT facilitates colorectal cancer growth through inducing Wnt/β-catenin pathway by enhancing p300-mediated β-catenin acetylation. *Oncogene*, *37*(26), 3485–3500. https://doi.org/10.1038/s41388-018-0161-z

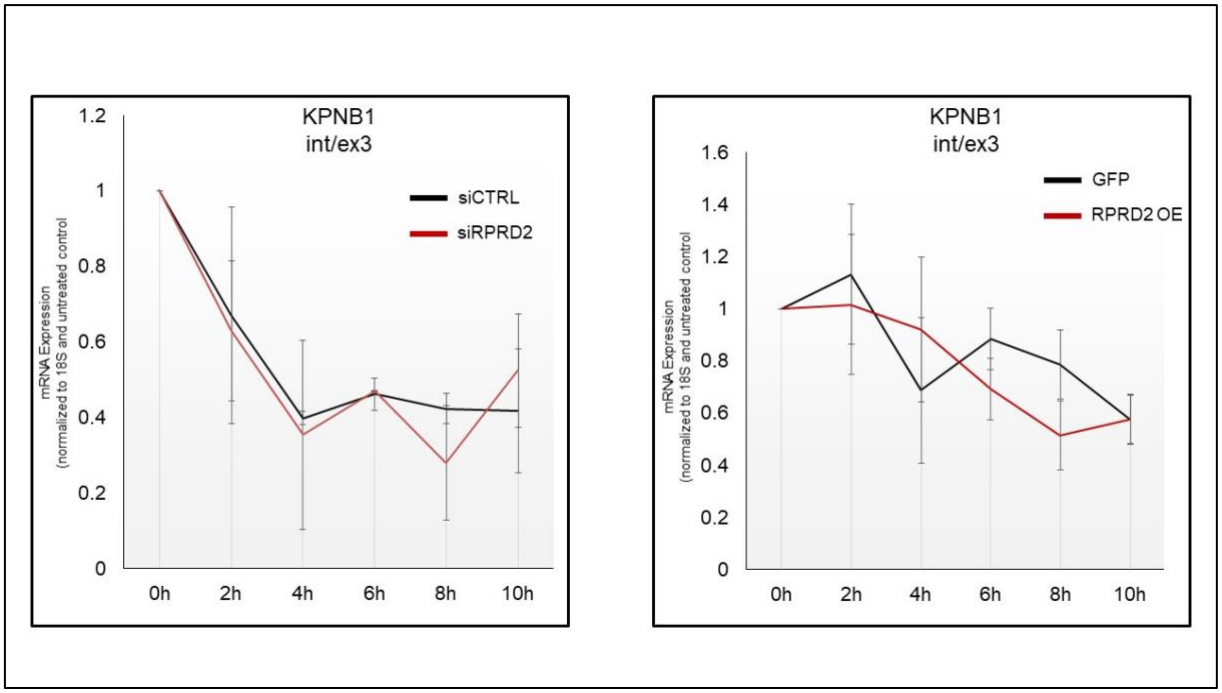
Zhang, Z., Shao, L., Wang, Y., & Luo, X. (2019). MicroRNA-501–3p restricts prostate cancer growth through regulating cell cycle-related and expression-elevated protein in tumor/cyclin D1 signaling. *Biochemical and Biophysical Research Communications*, *509*(3), 746–752. https://doi.org/10.1016/j.bbrc.2018.12.176

Zheng, G., Li, W., Zuo, B., Guo, Z., Xi, W., Wei, M., Chen, P., Wen, W., & Yang, A. G. (2016). High expression of CREPT promotes tumor growth and is correlated with poor prognosis in colorectal cancer. *Biochemical and Biophysical Research Communications*, *480*(3), 436–442. https://doi.org/10.1016/j.bbrc.2016.10.067

Zhu, Z., Liu, J., Feng, H., Zhang, Y., Huang, R., Pan, Q., Nan, J., Miao, R., & Cheng, B. (2022). Overcoming the cytoplasmic retention of GDOWN1 modulates global transcription and facilitates stress adaptation. *ELife*, *11*. https://doi.org/10.7554/eLife.79116

APPENDIX

Figure A1: q-PCR analysis of int/ex3 region of KPNB1 gene after prolong DRB treatment in siCTRL vs siRPRD2 and GFP vs RPRD2 OE.



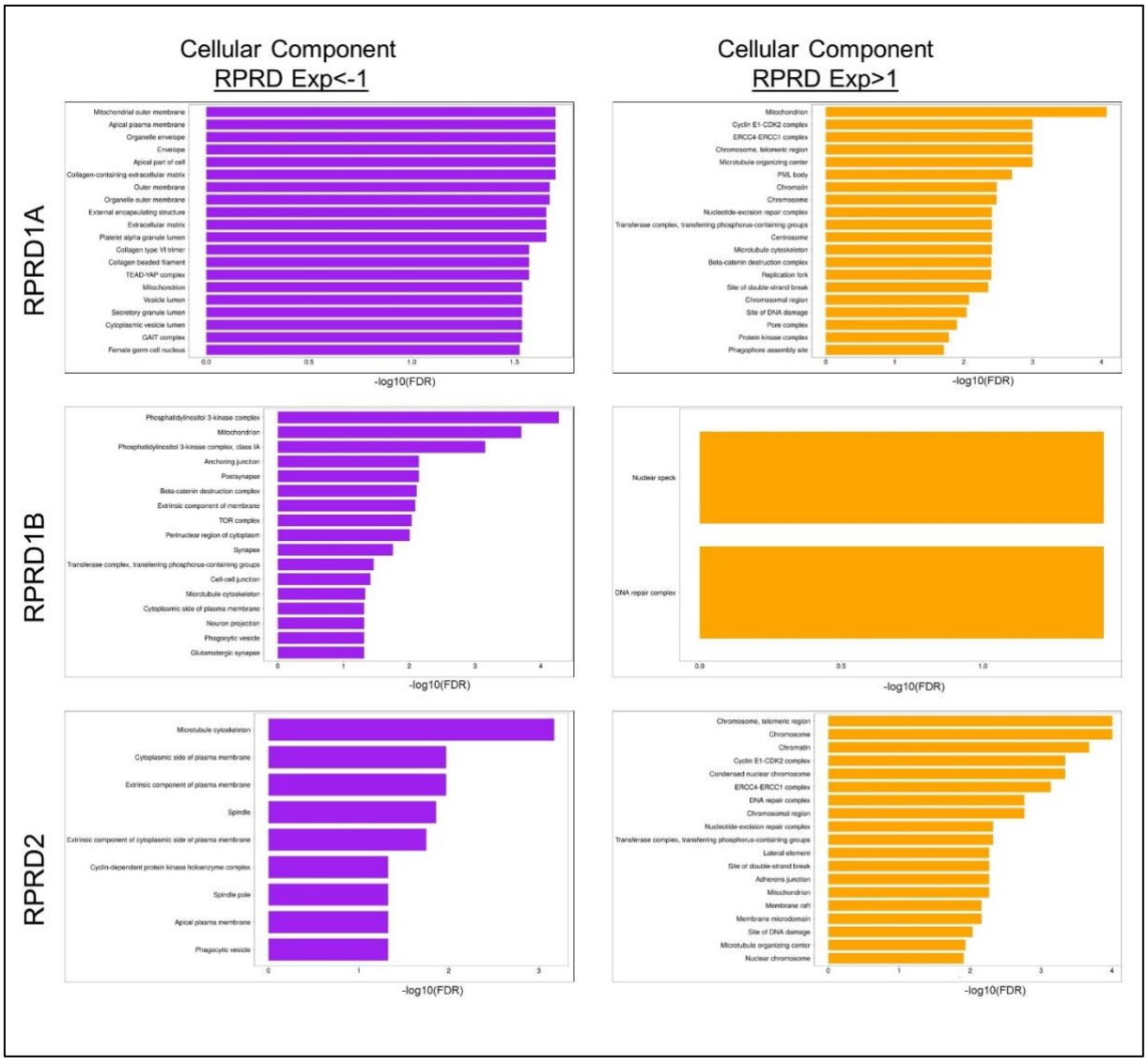
q-PCR was performed to analyze the expression levels within the intronic/exonic region 3 (int/ex3) of the KPNB1 gene, following extended exposure to DRB treatment. Comparisons were made between samples subjected to siCTRL (*black line*) versus siRPRD2 (*red line*) (graph on the left side); as well as samples overexpressing GFP (*black line*) versus overexpressing RPRD2 (*red line*) (graph on the right side). Assessment of the intron-exon region was performed to evaluate the levels of unspliced nascent RNA. The focus on this specific region aimed to demonstrate a reduction in nascent transcription, providing clarity on the overall decay of total RNA throughout the extended DRB treatment.

Table A1. The high level of proteins in RPRD differential expressed gliomas.

RPRD1A		RPRD1B<-1		RPRD2	
LOW	**HIGH**	**LOW**	**HIGH**	**LOW**	**HIGH**
SYK	IRS1	TUBA1B	RAD50	SYK	XRCC1
BAX	SCD	INPP4B	KDR	GAB2	FASN
MAPK14	DIABLO	PRKCB_PS660	AR	CLDN7	SCD
ANXA1	GATA3	RPS6KB1_PT389	EEF2	ANXA1	IRS1
STAT5A	PGR	CLDN7	PXN	IGFBP2	ACACA_PS79
RICTOR_PT1135	RAB25	RPS6KA1	AKT1_PS473	RICTOR_PT1135	ACACB_PS79
SMAD1	ARAF_PS299	BRAF	AKT2_PS473	STAT5A	CHEK1
TIGAR	NRAS	PDK1_PS241	AKT3_PS473	SERPINE1	FOXM1
MAPK1_PT202_Y204	SHC1_PY317	MAP2K1	XRCC1	GAPDH	GATA3
MAPK3_PT202_Y204	CDH3	KIT	JUN_PS73	BAX	RAB25
MAPK14_PT180_Y182	BECN1	PDK1	MSH6	YAP1_PS127	ARAF_PS299
SERPINE1	PECAM1	XBP1	SMAD1	PEA15	DIABLO
ATM	TP53	YWHAE	STAT3_PY705	PCNA	NRAS
PDCD4	FASN	YWHAZ	EGFR_PY1068	EIF4EBP1_PT70	DIRAS3
STAT3_PY705	MAPK9	TSC2	PEA15_PS116	CHEK2_PT68	MAPK9
IGFBP2	YWHAB	TSC1	BAX	LCK	TP53
AKT1_PT308	ANXA7	GSK3A	GSK3A_PS9	EIF4G1	CDH3
AKT2_PT308	PTEN	GSK3B	GSK3B_PS9	ERBB2	RAD50
AKT3_PT308	MAP2K1	BECN1	RPS6_PS240_S244	CASP7	PGR
YAP1_PS127	DIRAS3	PIK3R1	RPS6_PS235_S236	FN1	PEA15_PS116
YAP1	SMAD4	PIK3R2	SRSF1	ASNS	DVL3
CCNB1	ADAR	G6PD	EIF4EBP1	YAP1	SHC1_PY317
AKT1_PS473	ERBB3_PY1298	ADAR	MAPK14	YWHAZ	FOXO3_PS318_S321
AKT2_PS473	CHEK1	STK11	TFRC	RAB11A	BAP1
AKT3_PS473	NRG1	GAB2	RPS6	RAB11B	ERCC1
WWTR1	DVL3	NRG1	SRC_PY416	MAPK14	CCNE1
COL6A1	FOXM1	ESR1_PS118	RBM15	TUBA1B	ANXA7
GAPDH	PDK1	NDRG1_PT346	EEF2K	ESR1_PS118	RAD51
HSPA1A	FOXO3_PS318_S321	EIF4EBP1_PS65	CCNB1	TIGAR	NOTCH1
BAD_PS112	ERRFI1	BAK1	ERBB2_PY1248	COL6A1	PTEN
BID	ACACA_PS79	PIK3CA	MYH9_PS1943	CDKN1B	SMAD4
FN1	ACACB_PS79	PTEN	PREX1	ATM	BECN1
AKT1S1_PT246	XRCC1	PEA15	CTNNA1	STAT3_PY705	EEF2
ERBB2	ERBB3	MS4A1	AKT1S1_PT246	KIT	FOXO3
ASNS	RAD51	TGM2		CCNB1	RPS6
MAP2K1_PS217_S221	NKX2-1	CASP7		SMAD1	ERBB3_PY1298
	BRAF	PRKCA		YBX1_PS102	KDR
	BAP1	ESR1		ERBB2_PY1248	ACACA
	ACACA	NRAS		STK11	PECAM1
	SQSTM1	PECAM1		EGFR_PY1173	RPS6_PS240_S244
	PEA15_PS116	ACACA		MAPK1_PT202_Y204	ERRFI1
	G6PD	PGR		MAPK3_PT202_Y204	PRKCA_PS657
	RAF1_PS338	CHEK2_PT68		WWTR1	JUN_PS73
	BCL2	MYH11		AKT1_PT308	BCL2
	ERCC1	AKT1		AKT2_PT308	NF2
	INPP4B	AKT2		AKT3_PT308	EIF4EBP1
	MRE11	AKT3		AKT1	EEF2K
	CCNE2	ERBB3		AKT2	MSH6
	CHEK1_PS345	MAPK8_PT183_Y185		AKT3	YWHAB
	CCNE1	RPTOR		NFKB1_PS536	CDH2
	PRKCA_PS657	RAB11A		STMN1	TGM2
	GSK3A	RAB11B			BRCA2
	GSK3B	VHL			CCNE2
	PDK1_PS241	RAF1_PS338			RICTOR
	BAK1	NKX2-1			PDK1
	YWHAE	SMAD3			
		EIF4EBP1_PT70			
		CDKN1B_PT157			
		PRKCA_PS657			
		RICTOR			

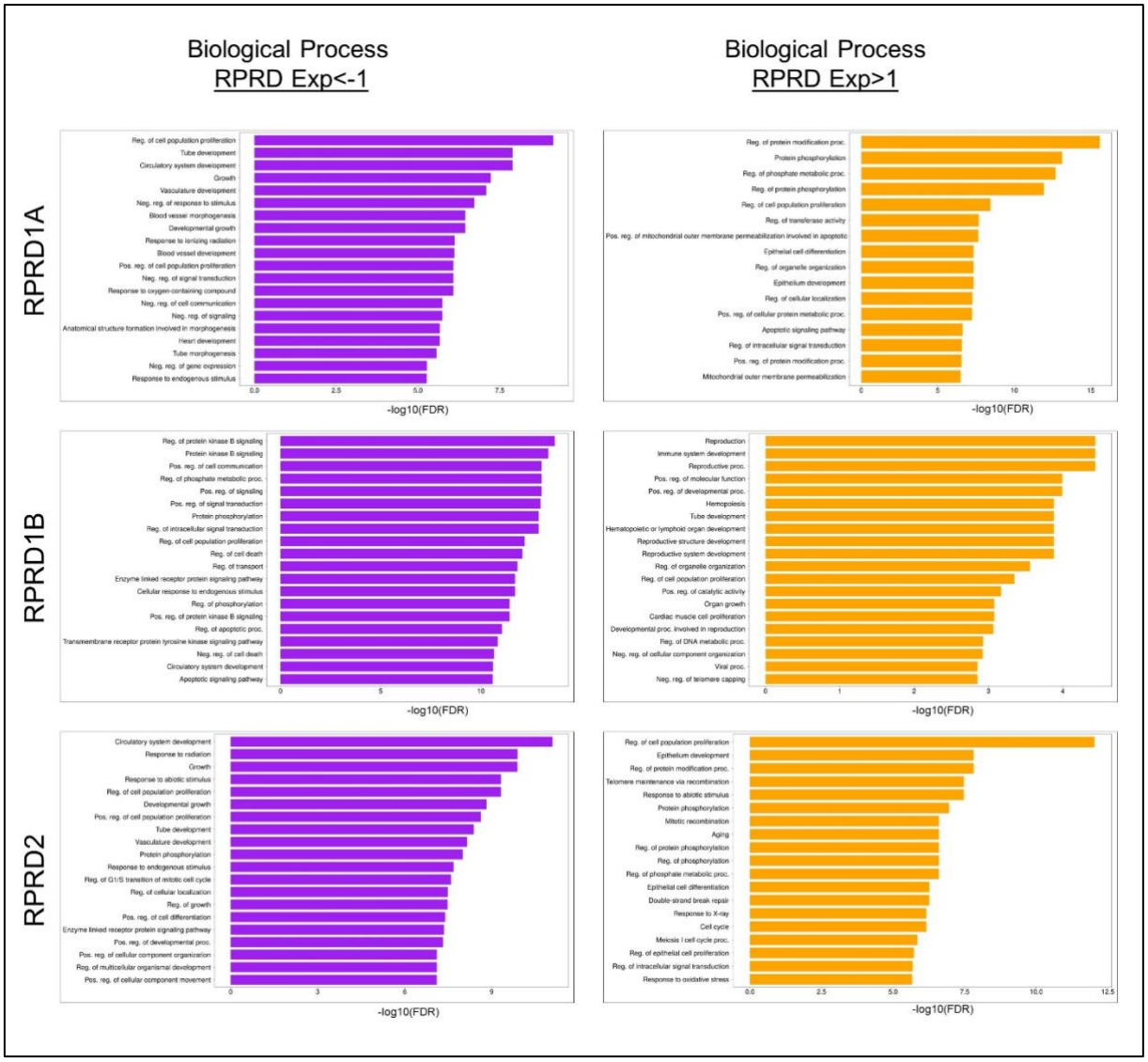
The list of overexpressed proteins identified in gliomas exhibiting differential RPRD expression was acquired through data analysis using CBioPortal. Specifically, this dataset encompasses proteins that elevated RPRD levels is referred to as ***High*** (orange column) and diminished RPRD expression is referred to as ***Low*** (blue column). These protein expression profiles were obtained through the use of the RPPA method.

Figure A2. The cellular component GO terms depending of RPRD low and high expression in gliomas.



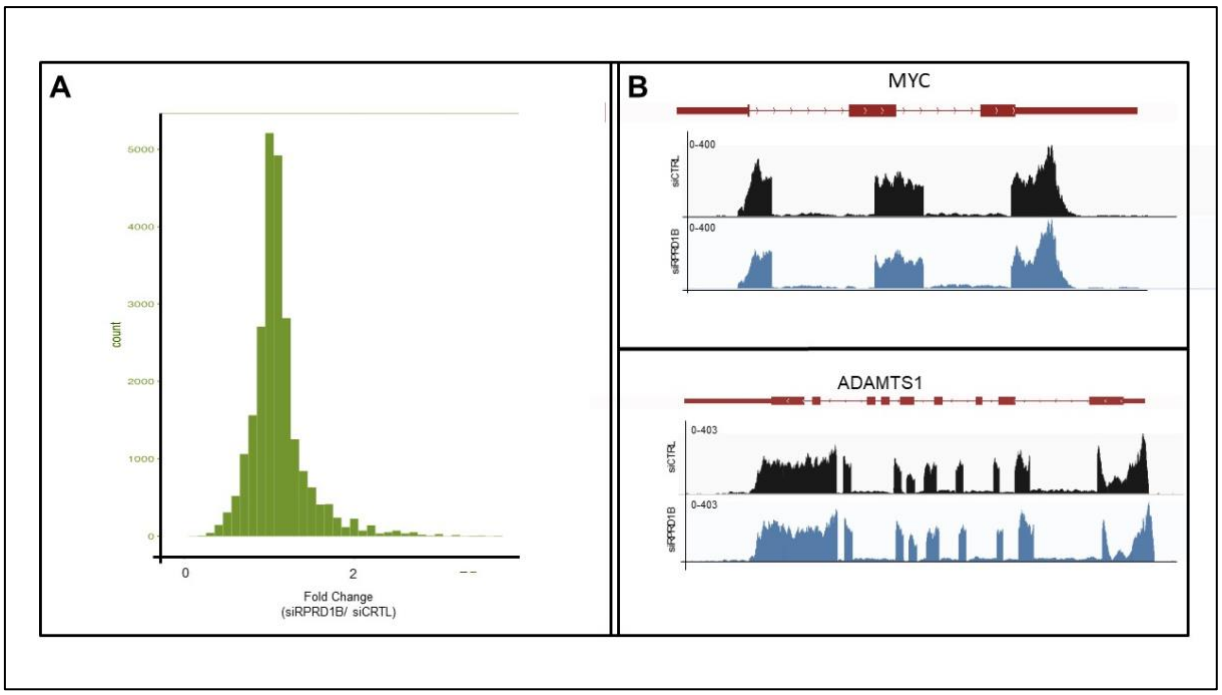
The GO annotation results were derived from an analysis of genes that exhibited differential expression patterns in response to low and high expression of three RPRD proteins within glioma patient samples. These GO terms encompassed cellular component. The sorting of GO categories within each functional group was performed by arranging them in descending order of statistical evidence, determined by the p-values obtained from the GO enrichment test.

Figure A3. The biological process GO terms depending of RPRD low and high expression in gliomas.



The GO annotation results were derived from an analysis of genes that exhibited differential expression patterns in response to low and high expression of three RPRD proteins within glioma patient samples. These GO terms encompassed biological process. The sorting of GO categories within each functional group was performed by arranging them in descending order of statistical evidence, determined by the p-values obtained from the GO enrichment test.

Figure A4. Total RNA sequencing results comparing control and RPRD1B deleted cells.



The findings from total RNA sequencing involving siCTRL and siRPRD1B were examined. The data were graphically represented to illustrate the fold change in counts of siRPRD1B relative to siCTRL (**A**). Notably, a fold change value of 1 was observed for the majority of genes, indicating an absence of significant alterations between the compared groups. The visualization of the total RNA levels of MYC and ADAMTS1 genes in siCTRL and siRPRD1B samples using the Integrative Genomics Viewer (IGV) was represented (**B**). This analysis revealed that there were no appreciable changes in the expression of MYC and ADAMTS1 genes between the compared groups, signifying a lack of significant variations in their transcriptional profiles.