The role of the RGG/RG motif of DDX21 in the DNA damage response

by

Samantha Daley

Department of Biochemistry

McGill University

April 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science.

Abstract

R-loops are three stranded nucleic acid structures made up of an RNA:DNA hybrid and a displaced ssDNA. Unscheduled, stable R-loops are known to cause genomic instability due to defective transcription, DNA replication, and DNA damage repair. Enzymes that remove R-loops, including DEAD-box RNA helicases (DDX), are essential for cell survival. DDX helicases are known to play critical roles in DNA damage repair processes. Previous work in our lab has shown DDX5 to be excluded from sites of DNA double strand breaks, where it functions to clear adjacent R-loops and prevent defects in homologous recombination. In order to discover additional DDX helicases similarly involved in DNA damage repair, we tested 6 DEAD-box family helicases using the LacI-FokI-mCherry reporter system. This study identified DDX21 to be excluded from sites of DNA damage, similar to DDX5. Our previous research found the RGG/RG motif of DDX5 to be required for its function in DNA damage response and arginine methylation is necessary for R-loop resolution. DDX21 has a previously uncharacterized C-terminal RGG/RG motif. We therefore examined the methylation status and functional relevance of the RGG/RG motif of DDX21. To determine whether DDX21 is arginine methylated, we performed in vitro methylation assays using PRMT1 and PRMT5. We determined that the C-terminal RGG/RG motif of DDX21 is arginine methylated. This observation was confirmed in cells by immunoprecipitation and western blot analysis. In order to study the functional role of the RGG/RG motif of DDX21, we assessed the localization of DDX21, with or without its RGG/RG motif, by immunofluorescence. We found that the C-terminal domain of DDX21 is essential for its nucleolar localization. Moreover, we showed that the exclusion of DDX21 from sites of DNA damage using the LacI-FokI-mCherry reporter system was abolished upon the mutation of the C-terminal arginine residues to lysine. Additionally, we performed DRIP-qPCR to examine the role of DDX21 in the

regulation of R-loops. Upon the knockdown of DDX21, we found that R-loop levels significantly increased at distinct ribosomal genes. These findings identify DDX21 to be a regulator of R-loop homeostasis, a substrate of arginine methylation and a potential player in the DNA damage response.

Résumé

Les boucles R, pour « ribonucléique », sont des structures d'acide nucléique à trois brins composés d'une boucle d'ADN simple brin formée d'une hybridation avec un ARN. Les boucles R stables et non programmées sont connues à déclencher une instabilité génomique à cause des problèmes de transcription, de réplication de l'ADN et de réparation des dommages à l'ADN. Les enzymes qui dégradent les boucles R, y compris les ARN hélicases à boite DEAD (DDX), sont essentielles pour la survie des cellules. Il est déjà bien connu que les hélicases DDX jouent un rôle essentiel dans les processus de réparation des dommages à l'ADN. Notre laboratoire a démontré auparavant que DDX5 était exclu des sites de cassures double brin de l'ADN, où il fonctionne pour éliminer les boucles R adjacentes et prévenir les défauts de recombinaison homologue. Afin de découvrir d'autres hélicases DDX impliquées de manière similaire dans la réparation des dommages à l'ADN, nous avons testé 6 hélicases à boite DEAD à l'aide du système rapporteur LacI-FokI-mCherry. Cette étude a identifié DDX21 comme étant exclu des sites de dommages à l'ADN. Nos recherches précédentes ont révélé que le motif RGG/RG de DDX5 était nécessaire pour sa fonction dans la réponse aux dommages à l'ADN et que la méthylation de l'arginine est nécessaire pour la résolution des boucles R. DDX21 a un motif RGG/RG C-terminal non caractérisé précédemment. Nous avons donc examiné l'état de méthylation et la pertinence fonctionnelle du motif RGG/RG de DDX21. Pour déterminer si DDX21 est méthylée pas les protéines méthyltransférases, nous avons d'abord effectué des tests de méthylation *in vitro* en utilisant les deux protéines PRMT1 et PRMT5. Nous avons déterminé que le domaine C-terminal de DDX21 est méthylé au niveau des résidus arginine. Cette observation a été confirmée dans les cellules par immunoprécipitation suivie d'un western blot. Afin d'étudier le rôle fonctionnel du motif RGG/RG de DDX21, nous avons évalué la localisation des mutants DDX21 par immunofluorescence. Nous avons constaté que le domaine

C-terminal de DDX21 est essentiel pour la localisation nucléolaire. De plus, nous avons montré que l'exclusion de DDX21 des sites de dommages à l'ADN était abolie lors de la substitution des résidus arginine (R) C-terminaux en lysine (K). De plus, pour examiner le rôle de DDX21 dans la régulation des boucles R, nous avons réalisé une expérience de DRIP-qPCR. Lorsque DDX21 est supprimé, les niveaux de boucle R au niveau de gènes ribosomiques distincts augmentaient d'une façon significative. Ces résultats identifient DDX21 comme un régulateur de l'homéostasie des boucles R, un substrat de la méthylation de l'arginine et un acteur potentiel dans la réponse aux dommages à l'ADN.

Acknowledgments

I would like to thank my supervisor, Dr. Stéphane Richard, for his continued support and guidance during my graduate studies. His drive and commitment to research has pushed me to learn valuable skills and has shaped the scientist I am today. He has helped me to think critically and to pursue experiments with dedication and focus. I am grateful for the opportunity to have worked in his lab and will always appreciate his mentorship.

I would also like to thank my Research Advisory Committee, Dr. William Pastor and Dr. Marc Fabian, for their valuable insights and advice for my research. Their feedback helped to concentrate my efforts and guide my future directions.

Many thanks are deserving to my fellow colleagues in the Richard lab. The encouragement and assistance from the other members of this lab has been an essential part of my studies. I would especially like to thank Dr. Zhenbao Yu for teaching me several scientific techniques, sharing his experimental expertise and his consistent help with troubleshooting. Special thanks to the past members of the lab, Dr. Ting Cai, Dr. Jeesan Lee and Dr. Claudia Dominici, for always making time to discuss and demonstrating how to be an excellent scientist. I would also like to thank current lab members, Dr. Nivine Srour, Dr. Otto Kletzien, Alexandra Farkouh, Helen Han and Sarah Khan for their constant inspiration. I am so grateful for their support and for the caring and thoughtful work environment.

Lastly, I would like to sincerely thank my friends and family for the love and reassurance throughout my studies. To my sister Sabrina, I want to thank her for always encouraging me and taking care of me whenever I needed it. To my Mom and Dad, thank you for motivating me and inspiring me to work hard towards achieving my goals. Special thanks to my cat Tony and all my Montreal friends for always being there for me throughout my journey.

Contribution of Authors

S. Daley performed all experiments and analysis of data. Dr. S. Richard supervised the work and obtained the funding for the project. Dr. Z. Yu and Dr. S. Richard helped to design experiments. Dr. Z. Yu assisted and reviewed the analysis of data. Figure 1B was taken from Zhenbao et al. in 2020 as the basis for this thesis.

Table of Contents

List of Tables and Figures

List of Abbreviations

53BP1	p53-binding protein 1
aDMA	Asymmetric dimethylarginine
AKT	protein kinase B
AQR	RNA helicase aquarius
ASF/SF2	alternative splicing factor/splicing factor 2
ATM	Ataxia-telangiectasia mutated
ATP	adenosine triphosphate
ATR	ataxia telangiectasia and Rad3-related protein
BLM	Bloom syndrome protein
BRCA1	breast cancer type 1 susceptibility protein
BRCA2	breast cancer type 2 susceptibility protein
CARM1	coactivator-associated arginine methyltransferase 1
CBP	CREB-binding protein
ChIP	chromatin immunoprecipitation
CRISPR	clustered regularly interspaced short palindromic repeats
CSR	class switch recombination
DAPI	4',6-diamidino-2-phenylindole
DDR	DNA damage response
DDX	DEAD-box RNA helicase
DEAD-box	Asp-Glu-Ala-Asp motif
DNA	deoxyribonucleic acid
DNMT3	DNA methyltransferase 3
DRIP	DNA:RNA immunoprecipitation
DSB	double-strand break
E2F-1	transcription factor E2F1
EGR1	early growth response protein 1
ERK1	extracellular signal-regulated kinase 1
FGFR3	fibroblast growth factor receptor 3
FL	full-length

FUS	fused in sarcoma
GST	glutathione S-transferase
HRP	horseradish peroxidase
eIF4AIII	Eukaryotic initiation factor 4A-III
IGS	Intergenic spacer
IR	Ionizing radiation
IRES	Internal ribosome entry site
JMJD6	Jumonji domain-containing protein 6
KDM	Lysine-specific demethylase
MEP50	Methylosome protein 50
MMA	Monomethylarginine
MRE11	Meiotic recombination 11
NSB1	Nibrin
OD_{260}	Optical density at 260 nm
OHT	4-hydroxytamoxifen
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PFA	Paraformaldehyde
PMSF	Phenylmethylsulfonyl fluoride
PRMT	Protein arginine methyltransferase
RAD51	DNA repair protein RAD51 homolog 1
RBMX	RNA binding motif protein, X-linked
RECQ	RecQ helicase
RGG/RG	Arginine glycine rich region
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RPA	Replication protein A
RPPH1	Ribonuclease P RNA component H1
RPS21	Ribosomal protein S21
SAM	S-adenosyl methionine
SDS	Sodium dodecyl sulfate

sDMA	Symmetric dimethylarginine
SEM	Standard error of mean
SETX	Senataxin
SIN3A	Histone deacetylase complex subunit Sin3a
SIRT7	Sirtuin 7
SMN	Survival motor neuron protein
SRSF1	Serine/arginine-rich splicing factor 1
TAF15	TATA-binding protein-associated factor 15
TDRD3	Tudor domain-containing protein 3
TOP3B	DNA topoisomerase III beta
TRC	Transcription-Replication conflict
TREX	Transcription-export complex
TSS	Transcription start site
TTS	Transcription termination site
WDR46	WD repeat-containing protein 46
WT	Wild type
XLID	X-linked intellectual disability
XPG	Xeroderma pigmentosum group G
XRN2	5'-3' exoribonuclease 2

Introduction

1. R-loops

R-loops are three stranded nucleic acid structures that form in the nucleus when an RNA strand hybridizes with DNA. The R-loop structure consists of an RNA:DNA hybrid and a displaced single stranded DNA. They are most often formed during transcription, where the transcribed nascent RNA can easily hybridize with its template DNA (called *cis* R-loops) (Hegazy et al., 2020). R-loop formation has been shown to occur most frequently at highly transcribed regions (Sanz et al., 2016). Studies to map the locations of R-loops in the genome have observed high densities at open chromatin and RNA polymerase II binding sites, with peaks detected at promoter and termination regions (Sanz et al., 2016). Additional studies also showed that R-loops can occur at RNA polymerase I and III targets such as ribosomal DNA (rDNA), transfer RNA genes (tRNA), retrotransposons and mitochondrial DNA (El Hage et al., 2014). Approximately 50% of genes prone to forming R-loops were located at ribosomal DNA loci with accumulation detected specifically at RNA polymerase transcribed regions, in a study performed in *S. cerevisiae* (Wahba et al., 2016). This demonstrates that the majority of R-loops form in a transcription-dependent manner with enrichment at highly transcribed or repetitive loci.

The R-loop structure can be favoured under certain conditions. Some examples include; negative supercoiling, G-quadraplex formation, ssDNA breaks, and G-rich transcripts (Figure I) (Hegazy et al., 2020). DNA negative supercoiling due to active transcription or replication will

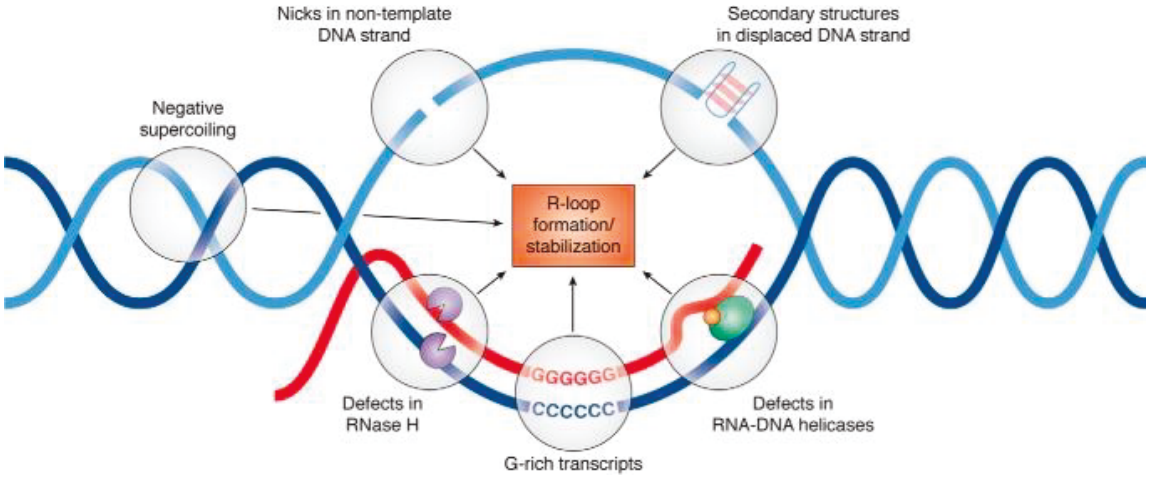


Figure I. Factors that promote R-loop formation or stability (Hegazy et al. 2020)

cause the DNA double helix structure to be more open, making it easier for the RNA strand to hybridize (Stolz et al., 2019). The formation of G-quadraplexes or breaks on the displaced ssDNA will make DNA reannealing less favourable, therefore increasing R-loop stability (Roy et al., 2010). Additionally, G-rich transcripts will stabilize the RNA:DNA interaction. G-C interactions between RNA and DNA strands creates a stronger bond making removal less energetically favourable (Roy and Lieber, 2009). In fact, it has been shown that R-loops occur more frequently at GC-rich regions in the genome (Niehrs and Luke, 2020). Although most R-loops form transiently during transcription, some remain stable in the genome, leading to DNA damage, while others have defined functions in the cell.

1.1 Methods to detect R-loops

Many techniques to detect R-loops have been developed. Several of these methods make use of an antibody, called S9.6, that recognizes RNA:DNA hybrids. The most common technique is DNA:RNA immunoprecipitation (DRIP) which relies on the S9.6 antibody for R-loop detection (Halasz et al., 2017). Many DRIP protocols have been developed and can differ slightly in the readout or the processes used to isolate DNA. qPCR analysis can be performed after DRIP to examine R-loop levels at specific loci. Alternatively, sequencing can be used to map sites of RNA:DNA hybrids throughout the genome. DRIP techniques are well established and often employed due to the ability to compare data with previously published works.

Other methods utilize the specificity of RNaseH for R-loop detection. RNaseH is an exoribonuclease that specifically recognizes and degrades RNA:DNA hybrids (Lee et al., 2022). RNaseH R-loop detection methods will use a catalytically inactive form to bind, but not degrade, RNA:DNA hybrids. Tagged forms of dead RNaseH (dRNaseH) can be used in chromatin immunoprecipitation (ChIP) or Cut&Run protocols. These techniques eliminate the bias presented by non-specific binding of S9.6 in DRIP (Wang et al., 2021). However, protocols using dRNaseH for R-loop detection can affect levels of native R-loops in the cell due to competition with endogenous RNaseH (Wang et al., 2021).

While S9.6 and dRNaseH techniques recognize the RNA:DNA hybrid portion of the R-loop structure, methods such as bisulfite sequencing will map areas of single stranded DNA within the genome. This technique can identify strand-specific R-loops with high resolution and eliminates non-specific recognition which can occur in antibody-based approaches (Dumelie and Jaffrey, 2017). Interestingly, various R-loop detection methods will give rise to the identification of diverse subsets of R-loops in the cell (Chedin et al., 2021). With the development of new techniques and

bioinformatic analyses, a cohesive map of R-loop forming regions should be established to improve discretions between methods.

1.2 Functions in the cell

R-loops play both functional and pathological roles in the cell. Stable R-loops in the genome present an obstacle to transcription, DNA replication and repair. However, the R-loop structure is essential for certain cell processes to occur, such as immunoglobin class switch recombination, mitochondrial and nuclear DNA replication (García-Muse and Aguilera, 2019). The various functions of R-loops will be discussed.

1.2.1 Defined physiological roles of R-loops

Although R-loops are mainly understood to cause conflict in the genome, they also play functional roles in certain cell processes. They have been well known to be involved in immunoglobin class switch recombination (Ig CSR) (Yu et al., 2003) (Figure IIE). CSR is a

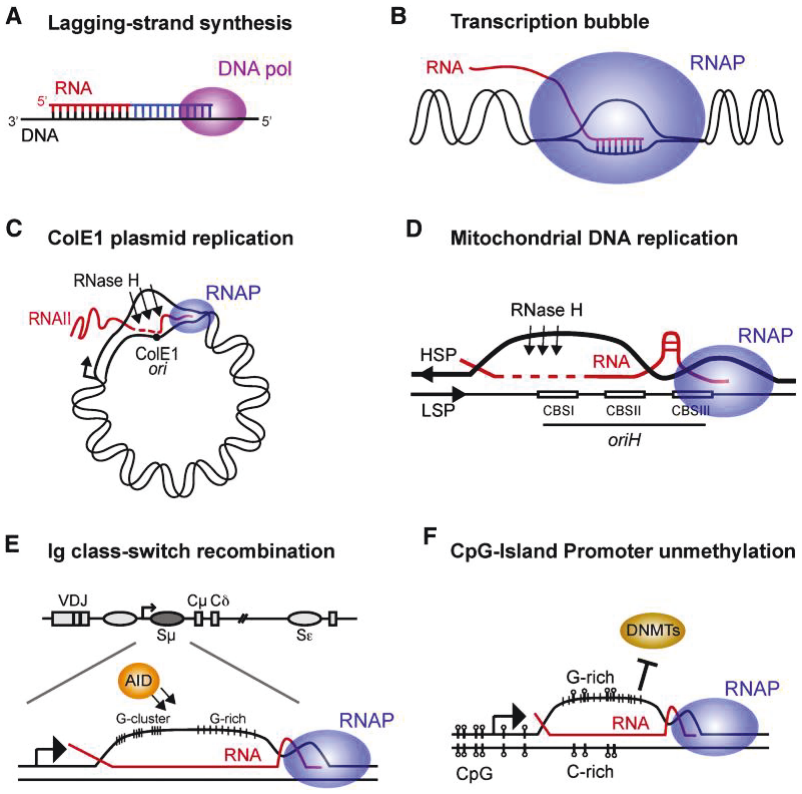


Figure II. Physiological roles of R-loops (Aguilera and Garcia-Muse, 2012)

process that occurs in B cells to switch the immunoglobin constant region isotype in response to antigen binding. Yu et al. found that stable >1kb R-loops form at switch regions and suggests that these structures may be an essential step for CSR to occur (Yu et al., 2003). *In vivo* studies have found that class switch recombination is significantly affected by the transcription of the $S_{\gamma 1}$ region and does not specifically depend on sequence, indicating that formation of RNA:DNA structures is important for CSR (Shinkura et al., 2003).

Also, R-loops can affect transcription termination. RNA polymerase stalling downstream of the 3' transcription termination site (TTS) due to R-loop formation has been found to promote efficient transcription termination (Aguilera and Garcia-Muse, 2012). Previous studies have identified Senataxin (SETX) RNA helicase and XRN2 exoribonuclease to prevent the formation of readthrough transcripts (Skourti-Stathaki et al., 2011). SETX unwinds the RNA:DNA hybrid and XRN2 will cleave the nascent RNA transcript to allow for effective release from the RNA polymerase (Skourti-Stathaki et al., 2011).

Additionally, R-loops serve a functional role in mitochondrial DNA (mtDNA) replication (Figure IID). The initiation of mtDNA replication has been previously examined due to its closed circular structure of ~16.5 kb (Pohjoismaki et al., 2010). mtDNA intermediates were detected to contain traces of RNA molecules suggesting that initiation may employ RNA primers, similar to lagging strand synthesis in chromosomal DNA replication (Yasukawa et al., 2006) (Figure IIA). It was shown that these RNA intermediates precede maturation to DNA strands and are essential for mtDNA replication (Yasukawa et al., 2006).

RNA:DNA hybrids also exist during transcription in the RNA polymerase active site (Figure IIB) (Westover et al., 2004). The presence of R-loops has also been studied to affect epigenetic marks like DNA methylation, where R-loop structures at CpG islands were shown to block de

novo DNA methylation by DNMT3B1 (Ginno et al., 2012). Overall, these examples show that R-loops are essential structures for a variety of cellular processes and clearly require significant regulation to avoid unnecessary accumulation.

1.2.2 R-loops as a source of genomic instability

Despite the roles that R-loops may play to encourage certain cell functions, they are more commonly understood to cause genomic instability. Persistent R-loops are capable of causing a variety of problems within the nucleus, such as mutagenesis, transcriptional stress, replication and DNA damage repair conflicts (Figure III).

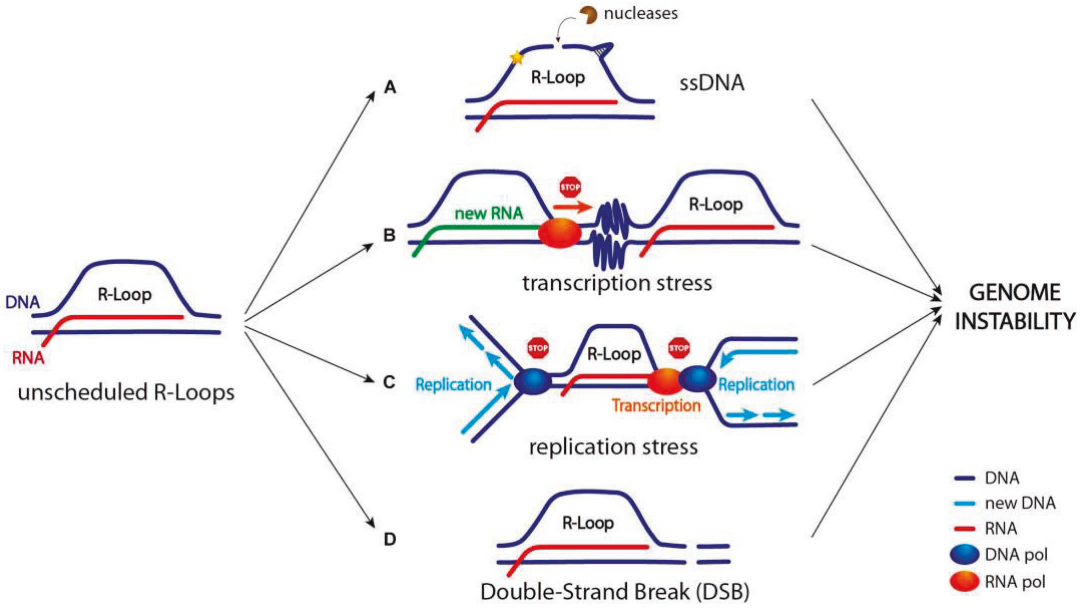


Figure III. R-loops potential for genomic instability in the cell (Rinaldi et al., 2021)

DNA damage can occur on the exposed ssDNA portion of the R-loop structure (Figure IIIA). Although this represents a minor percentage of R-loop associated damage, the unprotected strand is susceptible to mutagenesis and single strand breaks (Rinaldi et al., 2021). Nicks in the ssDNA caused by DNA deaminases (AID) or cleavage by endonucleases (XPG for example) can eventually lead to the activation of DNA damage repair pathways (Cristini et al., 2019; Gomez-Gonzalez and Aguilera, 2007). Activation of these pathways can result in additional mutations or deletions through the repair process.

R-loops occur at sites of active transcription as high transcription rates increase the chances for an RNA to hybridize with its template DNA. Transcriptional stress occurs when RNA polymerases are stalled due to obstructions in the DNA (Rinaldi et al., 2021) (Figure IIIB). The presence of RNA:DNA hybrids in the genome can cause transcriptional stress by blocking the path for new mRNA to be synthesized. Also, the formation of an R-loop co-transcriptionally can impair proper elongation of RNA polymerases. The RNA transcript hybridized to the DNA template will cause the polymerase to stall which inhibits processive transcript elongation (Huertas and Aguilera, 2003). The transcriptional stress caused by R-loop formation due to stalled RNA polymerase is known to activate ATM, initiating the DNA damage response (DDR) pathway (Tresini et al., 2015). Therefore, the removal of stalled RNA polymerases and R-loop obstacles is an essential process to maintain genomic integrity.

The major source of R-loop associated DNA damage occurs during transcription-replication conflicts (TRC) (García-Muse and Aguilera, 2019). The presence of stagnant RNA polymerase, with RNA hybridized to DNA, will not allow for the DNA replication machinery to pass through. The resulting conflict between RNA polymerase and DNA replication proteins can lead to genomic instability due to the formation of double strand breaks and deleterious repair events (Alexander and Orr-Weaver, 2016). Conflicts between R-loops and replication machinery can occur co-directionally or as a head-on collision with RNA polymerase (Figure IIIC). Head-on collisions, however, are those which have been studied to produce the most significant DNA damage. These conflicts can result in replication fork pausing or collapse (Prado and Aguilera, 2005). Studies in yeast and *E. coli* have demonstrated that head-on collisions led to inhibition of replication fork progression, increased cell senescence and overall loss of genomic integrity (Garcia-Rubio et al.,

2018; Mirkin and Mirkin, 2005). Therefore, the presence of stable R-loops during DNA replication can cause significant DNA damage and affect cell fate.

As outlined above, the presence of R-loops in the genome can activate DNA damage repair pathways due to transcription or replication conflicts. However, once activated, R-loops can prevent the proper functioning of DNA damage proteins. Studies have presented both positive and negative roles for R-loops in the DNA damage response pathways. Some examples will be described below.

The presence of an R-loop around sites of DNA damage can impair the cells ability to properly repair DNA breaks (Figure IIID). One of the first steps in homologous recombination (HR) repair of double strand breaks (DSBs) is end resection of the DNA (Zhao et al., 2020). Nucleases are responsible for degrading a section of DNA to leave a 3' ssDNA overhang. Exposed single stranded DNA is protected by RPA and RAD51 binding (Zhao et al., 2020). RNA:DNA hybrids can impair the HR resection process, as well as block the ability for RPA to bind DNA (Zhao et al., 2020). Studies in yeast have shown that RNaseH mutants cause sensitivity to DNA damage due to R-loop accumulation blocking end resection and inhibiting repair (Amon and Koshland, 2016; Costantino and Koshland, 2018). Similarly in human cells, the knockdown of AQR, an RNA helicase, resulted in increased DNA damage due to R-loop accumulation. This was accompanied with a decrease in Rad51-RPA foci, demonstrating that R-loops block HR protein binding and disturb the proper repair of DNA (Sakasai et al., 2017). Removal of R-loops from sites of DNA damage has been proven to be an essential step to ensure functional repair processes. Li et al. showed that ionizing radiation (IR) triggered the formation of DDX1 foci, an RNA helicase. They further showed that foci formation was dependent on the presence of R-loops (Li et al., 2016). Using DRIP-qPCR and an inducible DSB system, they found that DDX1 was required to remove

RNA:DNA hybrids from sites of damage to allow for Rad51 binding and the progression of HR (Li et al., 2016). Our lab described a similar role for the DDX5 RNA helicase, showing that DDX5 resolved R-loops adjacent to the sites of DNA damage so as to facilitate DNA damage repair (Yu et al., 2020). Studying R-loop related proteins can elucidate a function in the DNA damage repair pathway.

It has also been established that R-loops may support repair pathways and encourage protection of exposed DNA ends. Breaks that occur on DNA adjacent to actively transcribing RNA polymerase can easily favor the formation of R-loops. In vitro experiments have demonstrated that the formation of single stranded DNA breaks in proximity to active transcription strongly favors the formation of RNA:DNA hybrids (Roy et al., 2010). This process is energetically favourable and can potentially protect DNA from mutagenesis. Moreover, some studies have shown that RNA:DNA hybrids form at DSBs and recruit HR repair proteins, like BRCA1 and Rad52 (D'Alessandro et al., 2018; Yasuhara et al., 2018). D'Alessandro et al. reported that damage induced long noncoding RNAs (dilncRNA) form RNA:DNA hybrids at resected ssDNA. These hybrids were shown to be recognized by BRCA1 and resulted in the recruitment of BRCA2 and RNaseH (D'Alessandro et al., 2018). In a similar study, it was shown that RNA:DNA hybrids at laser irradiated sites encouraged Rad52 recruitment and the formation of BRCA1 foci (Yasuhara et al., 2018). Furthermore, Rad52 functioned to clear these hybrids through interaction with XPG endonuclease, giving rise to a 3' ssDNA overhang (Yasuhara et al., 2018). While both studies identify a positive role for R-loops in the HR process, the interacting factors differ. Also, D'Alessandro et al. demonstrate that RNA:DNA hybrid formation occurred after DNA was resected while Yasuhara et al. show that XPG cleavage of hybrids will initiate HR and DNA resection would follow (D'Alessandro et al., 2018; Yasuhara et al., 2018). The role of R-loops in

DNA damage repair is evidently complex and multifaceted. The question remains whether they have negative or positive functions during DDR, if their formation is due to ongoing transcription or through *de novo* RNA synthesis, and whether R-loop formation is essential for repair of all DSBs. It is possible that a combination of roles is true in different contexts. More about R-loops role in DNA damage repair can be read in the review by Rinaldi et al. (Rinaldi et al., 2021).

1.3 Mechanisms to prevent accumulation

Due to their potential for significant deleterious DNA damage, the cell has certain mechanisms which are able to prevent or inhibit the formation of R-loops. These include splicing of pre-mRNA, the formation of messenger ribonucleo-protein particles (mRNP), and the state of DNA.

Splicing has been studied to play a direct role in R-loop accumulation. Splicing can occur co-transcriptionally and involves the removal of intronic regions from the nascent RNA transcript. This is a cell intrinsic mechanism that prevents the creation of RNA:DNA hybrids. Modifying RNA through splicing will reduce the probability of annealing to its DNA template due to the loss of sequence similarity (Allison and Wang, 2019). Depletion of alternative splicing factor/splicing factor 2 (ASF/SF2) was shown to result in high levels of DNA damage due to the excessive accumulation of R-loops (Li and Manley, 2005). Similar results can be seen upon the knockdown of SRSF1 (Gan et al., 2011). In a more recent study, loss of Splicing Factor Proline and Glutamine rich (SFPQ) caused an increase in R-loop levels and subsequent cell death (Chakraborty et al., 2018). This shows how splicing plays an important role in preventing RNA:DNA hybrid formation.

Additionally, proteins binding to RNA is another method that will help to prevent the formation of R-loops. mRNA transcripts are always covered in RNA binding proteins and are present in the cell as mRNPs (Voronina and Pshennikova, 2021). Spliced, capped and poly-

adenylated RNAs bound by proteins are capable of exiting the nucleus and will proceed to translation. The THO complex functions in mRNP development and nuclear export and has been linked to preventing RNA:DNA hybridization (Huertas and Aguilera, 2003). In yeast, THO mutants gave rise to an increase in R-loop levels, causing hyperrecombination (Huertas and Aguilera, 2003). Similar discoveries were made with the TREX and THSC complexes (Gonzalez-Aguilera et al., 2008). As the mRNA matures to mRNP, it undergoes many changes in sequence and structure (Voronina and Pshennikova, 2021). These alterations discourage the annealing of RNA to DNA and help to prevent R-loop formation. Proteins involved in this process have been identified as key factors in maintaining genomic stability.

DNA negative supercoiling occurs during transcription and leaves the structure more loosely wound. When DNA is in this state, it facilitates RNAs ability to invade and form R-loops. DNA topoisomerases are the enzymes responsible for relieving supercoiling and play important roles in DNA replication, transcription and recombination (Champoux, 2001). Studies have found that the knockdown of Topoisomerase 1 (TOP1) resulted in the accumulation of DNA damage (El Hage et al., 2010; Tuduri et al., 2009). This was caused by a significant increase in R-loop levels, going to show that DNA supercoiling can greatly affect R-loop formation.

DNA chromatin state plays a vital role in the transcriptional status of all genes. Active transcription takes place within open and accessible areas of the genome to allow for binding of transcription factors and RNA polymerase. Since R-loops are characteristically dependent on transcription, the state of chromatin is therefore an important factor that can direct their formation. Genome-wide analysis has mapped R-loops to sections of open-chromatin, such as unmethylated promoter regions (Sanz et al., 2016). Yeast studies have also shown mutations of histones H3 and H4 that prevent the formation of closed chromatin causes an increase in R-loops and subsequent

genomic instability (Garcia-Pichardo et al., 2017). Histone deacetylases have also been shown to affect R-loop homeostasis. Upon the depletion of SIN3A, a histone deacetylase, chromatin enters a more open state and was shown to lead to an increase in RNA:DNA hybrids, DNA damage and defects in replication (Salas-Armenteros et al., 2017). Overall, the state of DNA and epigenetics play important roles in regulating R-loop levels.

1.4 Methods to remove R-loops

Despite the cells effort to maintain R-loop homeostasis through various mechanisms as explained above, unscheduled and stable R-loops must be removed to prevent DNA damage. There are several enzymes that function to remove R-loops from the genome, including RNaseH and RNA helicases.

RNaseH is an endoribonuclease that specifically degrades RNA strands of an RNA:DNA hybrid in a non-sequence-specific manner. It functions to remove R-loops and degrade Okazaki fragments, by-products of DNA replication (Lee et al., 2022). Loss of RNaseH is embryonic lethal in mice and mutations in humans have been linked to Aicardi-Goutières syndrome and impaired development (Rabe, 2013; Reijns et al., 2012). RNaseH1 and RNaseH2 are the two types found in eukaryotes.

RNA helicases are enzymes that can unwind RNA:DNA structures and have been shown to be essential in removing R-loops. Helicases differ significantly from RNaseH as they will simply unwind the RNA strand and will not degrade it. Senataxin was one of the earliest helicases detected to resolve R-loops *in vivo* (Skourti-Stathaki et al., 2011). It was found to remove R-loops and promote the termination of transcription through interaction with XRN2 (Skourti-Stathaki et al., 2011). The major class of helicases responsible for resolving R-loops are called DEAD-box family RNA helicases (DDX). Features of DEAD-box helicases will be discussed further below.

2. DEAD box family RNA helicases

DEAD-box RNA helicases are the largest family of RNA helicases made up of 37 proteins (Linder and Jankowsky, 2011). These proteins depend on ATP to help unwind RNA structures. They have a wide variety in functions, involved in all stages of RNA metabolism including transcription, splicing, RNA export, translation and RNA decay (Linder and Jankowsky, 2011). These helicases often function in a complex of proteins to target RNAs for specific processes. Many DEAD-box RNA helicases have been found to play essential roles in the cell and are required for cell survival. The characteristics of DEAD-box helicases will be discussed.

2.1 Conserved domains

DEAD-box helicases are so termed due to the conserved aspartate-glutamate-alanine-aspartate (DEAD) motif (Linder and Jankowsky, 2011). This motif can be found within the helicase core. The conserved helicase core consists of 2 domains made up of 12 conserved motifs (Linder and Jankowsky, 2011). Conserved regions are involved in ATP binding, hydrolysis and RNA binding. Structural studies have found that the 2 helicase core domains are split, forming a site for ATP binding located at their junction (Linder and Jankowsky, 2011). DEAD-box helicases form a network of interactions with ATP at the binding site, which is essential for enzymatic hydrolysis (Linder and Jankowsky, 2011). Due to this string of interactions between ATP and the helicase, mutations within this helicase core often render the DEAD-box protein as catalytically inactive.

RNA binding of DEAD-box helicases is also conserved within the family and similarly occurs in the helicase core domain, opposite to the ATP binding site. The conformation of RNA upon binding is a bent structure with contacts in both core domains (Linder and Jankowsky, 2011). The helicase interacts with RNA solely through the sugar backbone, in non-sequence-specific manner (Linder and Jankowsky, 2011). The fact that the RNA interaction is non-sequence specific is likely

what allows for the broad range of functions within the DEAD-box family. Since RNA binding has been shown to be non-sequence specific, these enzymes rely on other aspects to direct their specific cellular function.

2.2 Variable regions

In addition to their highly conserved helicase core, the proteins in the DEAD-box family have also been found to have a variety of variable domains within the protein sequence. These regions differ in size, type of domain, and location in the protein sequence (Linder and Jankowsky, 2011). Usually these variable domains are significantly less structurally defined, and the identification of these regions have not been established for all members of the family. The functional relevance of certain domains has been identified for many DEAD-box RNA helicases. The most commonly observed domains are involved in RNA binding. For example, RNA recognition motifs have been discovered in C-terminal domains of DEAD-box bacterial orthologs (Linder and Jankowsky, 2011). Structural analysis of other helicases in the family have also identified Zinc finger and KH domains similarly involved in nucleic acid binding (Cargill et al., 2021). These recent studies suggest that variable regions of DEAD-box helicases can impart specificity in function. The additional domains likely affect localization, post-translational modifications and protein-protein interactions to establish defined functions in RNA metabolism.

RGG/RG motifs have also been identified in several DEAD-box RNA helicases. RGG/RG motifs are regions within the protein sequence that are arginine-glycine rich. Strong RGG/RG motifs are found in DDX5, DHX9, DDX21, DDX36 and DDX54. One example of their importance for DEAD-box helicase function is given with DDX5. The RGG/RG motif of DDX5 was found to be essential for its function in DNA damage response and R-loop resolution

(Mersaoui et al., 2019; Yu et al., 2020). This finding is central to the results that will be presented in this thesis.

2.3 The role of DEAD-box RNA helicases in R-loop resolution

DEAD-box RNA helicases involved in transcription have often been linked to the prevention and resolution of R-loops. DDX1, DDX5, DDX19, DDX21, DDX39B, DDX41, DHX9, DDX17, DDX18 and DDX47 are some examples of helicases that have already been identified to regulate R-loop levels (Boleslavska et al., 2022; Cargill et al., 2021; Chakraborty et al., 2018; Lin et al., 2022; Marchena-Cruz et al., 2023). Our lab has previously studied DDX5 and its role in R-loop resolution. We showed that the knockdown of DDX5 resulted in R-loop accumulation. Methylation of DDX5 by PRMT5 at the RGG/RG motif was found to be essential for association with XRN2 exonuclease, which function together to resolve R-loops (Mersaoui et al., 2019). Additional studies have also identified DDX5 to resolve R-loops close to the sites of DNA double-strand breaks. DDX5 is excluded from FokI-induced DSBs in a transcriptionally dependant manner. This exclusion was found to be dependent on activation of ATM signalling kinase and the RGG/RG motif in the C-terminal domain of DDX5 (Yu et al., 2020). As previously described, R-loops can be obstacles to DNA damage repair and this work defined a new role for DDX5 to clear R-loops for proper homologous recombination to occur (Yu et al., 2020). DDX1 has also been shown to colocalize with γH2AX foci. The localization of DDX1 to sites of induce DNA damage was dependent on ATM and RNA:DNA hybrids (Li et al., 2016). These findings are essential to this thesis and led us to question whether other R-loop resolving helicases may similarly be triggered by DNA double strand breaks. The list of DEAD-box helicases that have been implicated in R-loop resolution continues to grow and their identification will help to further comprehend R-loop regulation.

2.4 DDX21

This thesis focuses on the role of DDX21 in DNA damage and R-loop resolution. DDX21, also known as nucleolar RNA helicase 2, is found mainly in the nucleolus where it has been shown to promote rRNA biogenesis (Zhang et al., 2014). Many roles of DDX21 have been previously characterized. Mainly, it has been shown that DDX21 plays an important role in RNA Pol I and II transcription. ChIP data revealed that DDX21 binds actively transcribed regions, with enrichment at ribosomal loci (Calo et al., 2015). It was found that DDX21 assists in the modification of ribosomal RNA and encourages Pol II transcriptional elongation through 7SK snRNP binding and release of P-TEFb (Calo et al., 2015). Furthermore, Song et al. in 2017 identified a role for DDX21 in R-loop resolution. Depletion of DDX21 resulted in higher levels of R-loops which could not be rescued with the expression of inactive DDX21 (Song et al., 2017). This role in R-loop resolution was linked to regulation by SIRT7 deacetylase. This group found that DDX21 acetylation by CBP caused inhibition of its activity. Upon deacetylation by SIRT7, the helicase activity of DDX21 was increased and allowed for effective R-loop removal (Song et al., 2017). Other studies have outlined a role for DDX21 in viral defense pathways, where it was shown to form a dsRNA sensing complex in the cytoplasm (Zhang et al., 2011). For our purposes, we focused on the roles of DDX21 in the nucleus.

DDX21 has been implicated in many diseases such as cancer and autoimmune conditions. DDX21 overexpression in breast cancer cell lines has been identified as a requirement for cell proliferation and was found to promote tumorigenicity *in vivo* (Zhang et al., 2014). Studies in colorectal cancer and lymphomas have also shown overexpression of DDX21, indicating its potential for therapeutic targeting in a variety of cancer types (Bonzheim et al., 2013; Tanaka et al., 2020). Additionally, autoantibodies for DDX21 have been identified in patients with

autoimmune disease, like gastric antral vascular ectasia and connective tissue disease (Tanaka et al., 2020). Studying the various functions of DDX21 is therefore valuable to understand the mechanisms that may be targeted in the treatment of diseases.

3. Protein arginine methylation

Post translational modifications (PTMs) of proteins are essential for several processes in the cell. Phosphorylation of proteins in the most common example of this. However, arginine methylation is a post translational modification that has been identified to play important roles in gene expression, splicing, DNA damage and R-loop resolution (Blanc and Richard, 2017). Arginine methylation involves the addition of a methyl group to the nitrogen atom within the arginine side chain.

3.1 Protein arginine methyltransferases (PRMTs)

PRMTs are the enzymes that catalyze the addition of the methyl group to the arginine residue from a donor molecule, S-adenosylmethionine (SAM). Arginines can be monomethylated or dimethylated, due to the presence of two guanidino nitrogen atoms within the arginine molecule

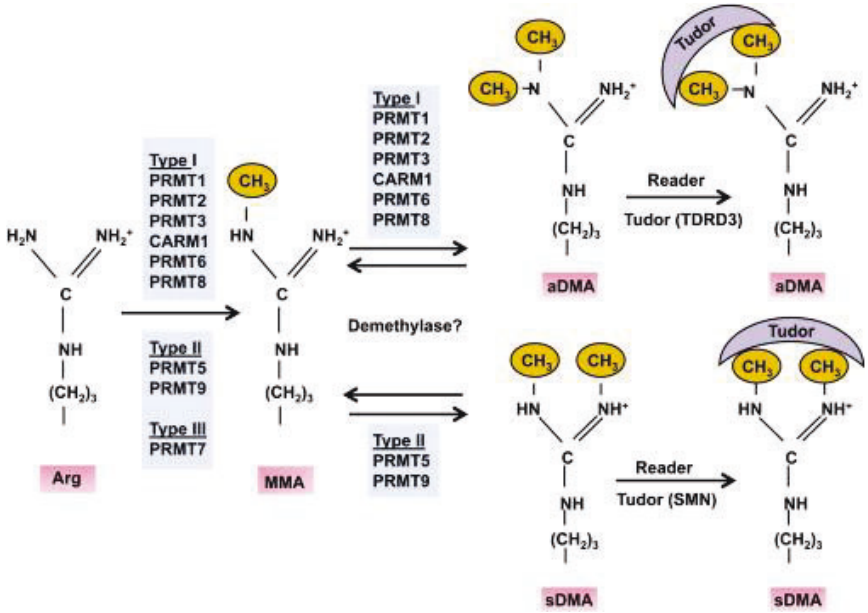


Figure IV. Arginine methylation by PRMTs (Thandapani et al., 2013)

(Figure IV.) (Xu and Richard, 2021). This gives rise to three forms of methylarginines, monomethylated (MMA), and two forms of dimethylated; asymmetric (aDMA) and symmetric (sDMA). Protein arginine methyltransferases are categorized based on which type of methylarginine they will produce. Type 1 consists of PRMT1, PRMT2, PRMT3, PRMT4, PRMT6, PRMT8 and will form MMA as an intermediate in the creation of asymmetric dimethylarginines. Type 2 PRMTs include PRMT5 and PRMT9 which will also synthesize MMA as an intermediate to symmetric dimethylarginine. The only identified Type 3 enzyme is PRMT7, which is only capable of catalyzing monomethylation of arginines (Figure IV). It has been shown that PRMTs will preferentially methylate arginines located within an arginine and glycine-rich sequences. These regions are called RGG/RG motifs. Different enzyme types have been found to vary in their specificity for arginine motifs, for example PRMT4 is shown to favour arginines within PGM-rich regions (proline, glycine, methionine) (Yang and Bedford, 2013). PRMT1 and PRMT5, the two major type 1 and type 2 methyltransferases, have a preference for RGG/RG motifs.

Methylated arginines in proteins can be recognized by Tudor domain-containing proteins (Blanc and Richard, 2017) (Figure IV). These interactions are biologically significant and allow for a variety of cell processes which will be briefly mentioned in following sections.

Other post-translational modifications have been found to be reversible, like phosphorylation or acetylation (Blanc and Richard, 2017). Currently, there is no evidence to support the existence of specific arginine demethylases *in vivo*. However, it has been shown that known lysine demethylases (KMD3A, KMD4E, KDM5C) can act on methylarginines *in vitro* (Walport et al., 2016). JMJD6 was thought to act on a methylarginines but was later shown to demethylate only lysine residues (Webby et al., 2009). Additional studies are required to determine whether *in vivo* arginine demethylation occurs and what role it plays in the cell.

3.2 RGG/RG motif containing proteins

RGG/RG motifs are the preferred substrate for PRMT arginine methylation. Over 1,000 proteins contain RGG/RG motifs which explains why they regulate a diverse set of biological functions (Thandapani et al., 2013). They are important low-complexity domains and are involved in nucleic acid binding, protein-protein interactions and liquid-liquid phase separation (Blanc and Richard, 2017; Roy et al., 2022). RGG/RG motifs have been shown to bind RNA in several proteins, including hnRNPU and nucleolin (Thandapani et al., 2013). Several RGG/RG motif-containing proteins have additional RNA binding domains, such as RNA recognition motifs (RRM) or Zinc fingers (Thandapani et al., 2013). RGG/RG motifs have also been found to facilitate protein interactions, for example the interaction between RNA helicase A and β-actin was identified to be RGG/RG-dependent (Tang et al., 2009). Furthermore, these motifs have been found to be important in protein localization and the formation of stress granules. Arginine-glycine rich regions in nucleolin are essential for its distinct nucleolar localization (Pellar and DiMario, 2003). The intrinsically disordered nature of RGG/RG motifs facilitate liquid-liquid phase separation, an important process for stress granule or P-body formation. For example, RGG/RG motifs in FUS and TAF15 RNA binding proteins, have been shown to direct phase separation into stress granules (Marko et al., 2012; Murthy et al., 2021). Additionally, RBMX forms nuclear granules but upon the deletion of the RGG/RG motif, these membraneless organelles were abolished (Cai et al., 2021). This demonstrates the importance of the RGG/RG motif in a range of cellular processes.

3.3 Functions of arginine methylation

While the RGG/RG motif alone can affect nucleic acid binding, protein-protein interactions and liquid-liquid phase separation, the methylation of arginines within these motifs have distinct

affects on gene expression, splicing, DNA damage and R-loop resolution (Blanc and Richard, 2017). The methylation of histones by PRMTs plays an important role in gene expression. PRMT1 and CARM1, both type 1 PRMTs, deposit H4R3me2a histone modifications which act as transcriptional activators (Blanc and Richard, 2017). Recruitment of p300/CBP complex to these methylated histones results in increase histone acetylation which is a well-known transcriptional activation mark (Bedford and Clarke, 2009). Histone arginine methylation can also repress transcription. PRMT5 methylates H4R3me2s and H3R8me2s, and this methylation was found to recruit DNA methyltransferase DNMT3a (Zhao et al., 2009). Binding of DNMT3a to these modifications resulted in methylation of DNA and subsequent gene inactivation (Zhao et al., 2009). Additionally, PRMTs can also affect gene expression through methylation of non-histone targets. The E2F-1 transcription factor was found to be methylated by PRMT5. E2F-1 activates the transcription of genes involved in cell cycle progression. Cho et al. found that arginine methylation of E2F-1 decreases the protein stability. Thus, the depletion of PRMT5 increased cell death due to higher E2F-1 protein levels (Cho et al., 2012).

Splicing is another process that is affected by PRMT5 methylation. Sm proteins are key components of the splicing machinery and methylation of their RGG/RG motifs is essential for recognition by the tudor domain of SMN (Brahms et al., 2001). The mature snRNP complex is formed only upon binding of Survival motor neuron protein (SMN) to symmetric dimethylated Sm proteins (Brahms et al., 2001). Furthermore, it has been shown that knockdown of PRMT5 results in increased aberrant splicing events which causes cell cycle defects (Sachamitr et al., 2021).

Several proteins involved in the DNA damage response are also known to be regulated by arginine methylation. Studies employing conditional knockouts of PRMT1 in mice revealed

defects in cell cycle progression and genomic instability (Yu et al., 2009). This phenotype was linked to a decrease in RAD51 foci formation and hypomethylation of MRE11 (Yu et al., 2009). Asymmetric dimethylation of MRE11 by PRMT1 was found to be required for its role in DNA damage checkpoint control (Boisvert et al., 2005a). Mutation of arginine residues within the glycine arginine-rich (GAR) domain of MRE11 proved to impair its exonuclease function, resulting in defective S-phase DNA damage checkpoints (Boisvert et al., 2005a). Additionally, 53BP1 was identified to be methylated by PRMT1. Treatment with PRMT inhibitors resulted in decreased recruitment of 53BP1 to DSBs (Boisvert et al., 2005b). BRCA1, a major factor involved in the regulation of DNA damage repair pathways, has also been identified to be methylated by PRMT1. Inhibiting methylation of BRCA1 was shown to effect its function as a transcriptional activator (Auclair and Richard, 2013). Also, PRMT1 methylation of hnRNPUL1 was found to be required for interaction with NSB1 and recruitment to DNA DSBs (Polo et al., 2012). Type 2 PRMT5 is also linked to the regulation of the DNA damage response. Induction of double strand breaks has been shown to cause methylation of p53 by PRMT5. PRMT5 knockdown resulted in increased p53 binding to apoptotic gene promoters and cell death (Auclair and Richard, 2013). In summary, PRMTs have a wide range of functions as arginine methylation has the ability to affect gene expression, splicing and DNA damage pathways.

3.3.1 R-loops

Our lab has recently focused on studying arginine methylation in the context of R-loop resolution. Several PRMT targets have been shown to impact the levels of R-loops in the genome. For example, arginine methylation of histones can affect chromatin state as well as recruit protein complexes that regulate R-loop levels, one example is given by the TDRD3+TOP3B complex. Asymmetric dimethylation of histones H3R17 and H4R3 can bind the tudor domain-containing

protein 3 (TDRD3) in complex with topoisomerase TOP3B (Yang et al., 2014). Binding of this complex to the c-Myc promoter region was linked to an increase in gene expression and a decrease in R-loop levels, thereby reducing R-loop associated DNA damage (Yang et al., 2014). The TDRD3-TOP3B complex was later found to recruit DHX9, an RNA helicase. The tudor domain of TDRD3 is required for DHX9 interaction, which functions with TOP3B at target genes to suppress R-loop formation (Yuan et al., 2021).

Moreover, the carboxy-terminal domain (CTD) of RNA polymerase II is another substrate of PRMT5 arginine methylation. R1810 methylation of RNA Pol II is detected at sites of transcription termination. This methylarginine was found to bind the tudor domain of SMN protein. The interaction between SMN and Senataxin, an RNA helicase, has been proven to resolve R-loops formed at transcriptional termination sites (Zhao et al., 2016).

Arginine methylation of DEAD-box helicases has been detected to affect their function in R-loop resolution. DRIP-seq data from our lab revealed R-loop resolving capabilities of DDX5 RNA helicase. Further studies showed that PRMT5 methylation of the RGG/RG motif in DDX5 allowed for interaction with the exoribonuclease XRN2. DDX5 and XRN2 function together at transcription termination sites to unwind and degrade R-loops, respectively (Mersaoui et al., 2019). Interaction between DDX5 and XRN2 is dependent on arginine methylation and the depletion of PRMT5 resulted in increased γH2AX foci due to the accumulation of R-loops (Mersaoui et al., 2019). This goes to show how arginine methylation of DDX5 can greatly impact its role in R-loop homeostasis. Our lab further demonstrated that DDX5 resolution of R-loops at sites of DNA is important for proper DSB repair. DDX5 was identified to remove R-loop structures adjacent to sites of DNA damage so that repair pathways were not impaired (Yu et al., 2020). This discovery

was found to be dependent on the C-terminal RGG/RG motif of DDX5. These examples demonstrate how arginine methylation can significantly regulate R-loop levels in the cell.

3.4 Implication in disease

Dysregulation of PRMTs has been linked to many diseases, including neurodegenerative and metabolic diseases. PRMTs are often studied within the context of cancer since overexpression of PRMTs is observed in many cancer types. For example, the upregulation of PRMT1 in lung cancer has been shown to stimulate metastasis through the decrease in E-cadherin expression (Avasarala et al., 2015). Similarly, the overexpression of PRMT5 was vital for lung cancer cell survival in tissue culture and *in vivo* mouse xenografts (Gu et al., 2012). In addition, the depletion of PRMT5 was shown to inhibit lung cancer cell growth through modulation of histone modifications that regulate expression of fibroblast growth factor receptor 3 (FGFR3). Lung cancer cells with overexpressed PRMT5 were correspondingly found to activate ERK1/2 and AKT signalling through the FGFR3 to promote cell proliferation (Jing et al., 2018). These are simply two examples of PRMT1 and PRMT5 have been identified as important markers in cancer progression.

The RGG/RG motif of RBMX has also been characterized to play an important physiological role in X-linked intellectual disability (XLID). Deletion of the RGG/RG motif in Shashi-XLID syndrome was shown to cause an increase in apoptotic markers due to aberrant p53 signalling (Cai et al., 2021). This phenotype was caused by the regulation of the RGG/RG motif of RBMX in splicing of the *MDM4* transcript. Loss of this critical motif in RBMX resulted in decreased MDM4 protein and a corresponding increase in p53 protein levels (Cai et al., 2021). This goes to show the importance of arginine methylation in a variety of disorders.

Several PRMT inhibitors are currently undergoing clinical trials for the treatment of cancers (Wu et al., 2021). Type 1 PRMT inhibitor, MS023, specifically inhibits asymmetric dimethylation with no affinity for Type 1 and Type 3 PRMTs or lysine methyltransferases (Eram et al., 2016). EPZ015666 is a specific Type 2 PRMT inhibitor currently undergoing Phase 1 clinical trials for solid tumour cancers and non-Hodgkin lymphoma patients (Clinical trial: NCT02783300) (Chan-Penebre et al., 2015). In conclusion, PRMTs play many essential roles in the cell which can be advantages for cancer cells to promote oncogenesis and metastasis. This makes them excellent targets for cancer therapeutics.

Aim

Previous studies in our lab identified a new role for DDX5 RNA helicase in DNA damage repair (Yu et al., 2020). It was found that DDX5 is excluded from sites of DNA double strand breaks to resolve adjacent R-loops for efficient homologous recombination to occur, a process that was established to be dependent on the DDX5 RGG/RG motif. We therefore sought to identify other DEAD-box helicases containing RGG/RG motifs with a role in DNA damage repair and R-loop resolution.

Materials and Methods

1. *Cell culture and cell lines*

U2OS, HEK293 and U2OS-265 cell lines, purchased from ATCC, were cultured in Dulbecco's Modified Eagle Medium (Wisent Bioproducts) supplemented with 10% fetal bovine serum and 100 U/ml penicillin, 100 μg/ml streptomycin (Wisent Bioproducts). All cell lines were kept at 37°C in 5% CO_2 in a certified cell culture incubator.

2. *Plasmids and subcloning*

The following plasmids were purchased from Addgene:

- Flag-DDX3X, pMXs mammalian expression vector (70648)
- Flag-DDX11, pcDNA3 mammalian expression vector (120727)
- Flag-DDX21, pHAGE-EF1a-IRES-ZsGreen mammalian expression vector (128803)
- Flag HA-DDX36 isoform 1, pFRT mammalian expression vector (159585)
- EGFP-DDX41, pcDNA3.1 mammalian expression vector (175494)
- Flag HA-DDX54, pFRT/TO/FLAG/HA-DEST mammalian expression vector (97060)

Human Flag-DDX21 truncation and mutation plasmids (DDX21 ΔN, DDX21 ΔC, and DDX21 R-K) were synthesized through subcloning with BamHI, XhoI and AfeI restriction enzymes (NEB). GST-DDX21 constructs were similarly synthesized (N domain, DDX domain and C domain). Portions of the Flag-DDX21 full length (FL) plasmid was PCR amplified using primers listed in Table 2. Primers used for PCR amplification are listed below:

- Flag-DDX21 FL: BamF1 and XhoR1
- Flag-DDX21 ΔN: BamF2 and XhoR1
- Flag-DDX21 ΔC: BamF1 and XhoR3
- GST-N domain: BamF1 and XhoR2

- GST-DDX domain: BamF2 and XhoR3
- GST-C domain: BamF3 and XhoR1

Primers were used to amplify the desired DDX21 sequence and PCR products were purified according to the manufacturers protocol (QIAGEN). IDT gene synthesis was used to synthesize the DDX21 R-K mutated sequence, listed in supplemental. Purified PCR amplicons and the empty expression vectors were digested with BamHI-HF and XhoI for 2-4 hours at 37°C (NEB). pcDNA3 and pGEX6P.1 expression vectors were used for Flag and GST construct synthesis, respectively. Digests were run on 1% agarose gels with ethidium bromide then gel purified using QIAGEN gel purification kit. Purified plasmid vectors were ligated with the corresponding gene insert using T4 DNA ligase (NEB) overnight at 8-20°C in a water bath. Flag-DDX21 FL, Flag-DDX21 ΔN, Flag-DDX21 ΔC digested inserts were ligated with digested pcDNA3 empty vector. DDX21-N domain, DDX21-DDX domain and DDX21-C domain digested inserts were ligated with digested pGEX6P.1 empty vector. Ligated plasmids were transformed into DH5α competent cells (ThermoFisher) and incubated overnight at 37°C on LB-Ampicillin agar plates. Clones were selected for colony PCR. Primers used for colony PCR and sequencing are listed in Table 3. Clones with the correct insert size were used to grow large colonies for MaxiPrep plasmid extraction (QIAGEN). Purified plasmids were sent for sequencing (Nanuq, Sanger Sequencing).

3. Transfections and drug treatments

siRNAs were purchased from Dharmacon. Sequences are as follows:

siDDX21 5' GCAUGUAUCUGCCUAUACUUUdTdT 3', siPRMT1 5′ CGUCAAAGCCAA CAAGUUAUU 3′, siPRMT5 5’ UGGCACAACUUCCGGACUUdTdT 3’, and siLuciferase 5′ CGUACGCGGAAUACUUCGAdTdT 3′ was used as a control.

Cells were transfected with Lipofectamine 3000 reagent for plasmids and RNAiMax for siRNAs according to manufacturer's protocol (Invitrogen). Cells were seeded at 60-90% confluency the day before transfection depending on the transfection. Transfection reagents were diluted in incomplete DMEM (Multicell). 25 nM siRNA and 1 μg/mL plasmid were used for transfections, sequences shown above. Cells treated with siRNAs were left to grow for 72 hours post transfection, and 48 hours following plasmid transfection.

PRMT inhibitors, EPZ015666 and MS023, were diluted in DMEM and left to incubate with cells for 48 hours. 5 μM and 1 μM final concentrations were used for EPZ015666 and MS023 treatments, respectively.

4. Immunoprecipitation (IP)

Cells were lysed with non-denaturing lysis buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% TritonX-100 with 1X protease inhibitors (PMSF, aprotinin, leupeptin, pepstatin and E-64 from Sigma)) for 30 minutes gently rocking on ice. Lysates were collected and spun down at 16,000 rcf for 10 min at 4°C. Anti-FLAG M2 Affinity gel beads (Sigma-Aldrich) were washed 3 times with 1 mL cold lysis buffer and 50 μL bead slurry was aliquoted for each IP. 50 μL of cell lysate was saved as input, the rest was added to Flag beads and incubated for 2-4 hours while inverting at 4°C. After incubation, beads were washed three times with 1 mL of lysis buffer (beads were pelleted at 2,000 rcf for 1 minute at 4°C). After the last wash, supernatant was discarded and 2X Laemmli buffer was added to the beads as well as the input lysate. Samples were stored at -20°C until used for SDS-PAGE. Lysates were used for western blot analysis as described in section 1.5.

5. Protein extraction and immunoblotting

Protein extracts were separated by SDS-PAGE using Bio-Rad Mini-PROTEAN Tetra Cell electrophoresis chamber. Proteins were transferred to nitrocellulose membranes using Semi-Dry

TurboTransfer system (Bio Rad). Membranes were blocked with 5% skim milk in TBS-T for 1 hour at room temperature with slow rocking. Primary antibodies were incubated with membranes overnight at 4°C, washed three times in TBS-T (5 minutes each) then incubated with the appropriate horse radish peroxidase (HRP) conjugated secondary antibody for 1 hour at room temperature with in 1% skim milk and TBS-T. After washing as previously described, ECL reagent (Bio Rad) was added and chemiluminescence signals were detected using Blu-Lite UHC western blotting film and developer. Primary antibodies are listed in Table 4.

6. In vitro methylation assay

1 μg of GST tagged proteins were incubated with 1 μL S-adenosyl-L-[methyl3H] methionine (15 Ci/mmol; Perkin-Elmer) and 0.5 μg of recombinant PRMT5:MEP50 active complex (Sigma SRP0145) in a 100 μL reaction for 1 h at 30°C. The methylation reaction was stopped by adding 2X Laemmli buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl pH 6.8, 10% β-mercaptoethanol, 0.02% bromophenol blue) then run on a 10% SDS gel. Gels were stained with Coomassie Blue then destained overnight (50% methanol, 10% acetic acid). Dyed gels were scanned, then incubated with ENH_3ANCE (Perkin Elmer) solution for 1 hour followed by 30-minute incubation in water. Gels were dried for 4 hours with vacuum and 2 hours of heating at 70°C, then exposed to Blu-Lite film for 1-7 days in a film cassette at -80°C. Film was developed as described in section 1.4.

7. DNA:RNA immunoprecipitation (DRIP)

U2OS cells were pelleted, washed in 1X PBS and lysed with 20% SDS and 5 μL of 20 mg/ml Proteinase K overnight at 37°C while inverting. DNA was extracted using 1 volume Phenol:Chloroform:isoamylalcohol (pH 8), 1/10 volume 3M NaOAc pH 5.2 and 2.4 volumes 100% ethanol. DNA was washed three times with 70% ethanol and air dried completely, then

gently resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0) on ice. Extracted genomic DNA was digested with EcoRI-HF, XbaI, HindIII-HF, SspI-HF and BsrGI-HF restriction enzymes (NEB) with 1 mM spermidine and 1X Bovine Serum Albumin (BSA) overnight at 37°C while inverting. Digested DNA was extracted as previously described, with 20 μg glycogen added to the aqueous DNA as a coprecipitant. To help precipitate the DNA, ethanol complexes were incubated at -20°C for 1 hour and spun at 16,000 rcf for 35 minutes at 4°C, then washed once with 70% ethanol. The purified digested DNA was air dried then resuspended (without vortexing) in TE buffer on ice for 1 hour. The concentration was measured using the NanoDrop (OD_{260}). A negative control of 4.4 μg of digested DNA was treated with 3 μl of Ribonuclease H (NEB, cat. #M0297) overnight at 37°C. 12.12 μg of digested DNA from each condition was diluted to 1.4 mL with TE buffer. A sample of 50 μL was saved to be used as input for performing qPCR analysis. 150 μL of 10X binding buffer (100mM $NaPO_4$ pH 7.0, 1.4 M NaCl, 0.5% Triton X-100) was added to the diluted DNA then divided into 3 tubes to perform immunoprecipitation in triplicate. 10 μg of S9.6 antibody (purified from HB-8730 cells from ATCC) was added and incubated for 16 hours at 4°C while inverting. Protein A agarose beads (ThermoFisher) were washed in 1X binding buffer (10X diluted in TE buffer) three times for 10 minutes with gentle shaking. 50 μL of bead slurry was added to the DNA:Antibody complexes and incubated for 2 hours at 4°C while inverting. Beads were spun down at 1,000 rcf for 1 min at 4°C. Supernatant was discarded, and beads were washed three times as described above. 250 μL elution buffer (50 mM Tris pH 8.0, 10 mM EDTA, 0.5% SDS) and 7 μL of 20 mg/ml Proteinase K was added to beads, then incubated at 55°C for 45 minutes while inverting. Eluted DNA was purified through the same phenol:chloroform and

ethanol precipitation used throughout. DNA pellet was resuspended in 100 µL 10 mM Tris-HCl pH 8.0 and left on ice for 15 minutes.

1.7.1 qPCR analysis

Input DNA was diluted to 150 µL with 10 mM Tris-HCl pH 8.0. Input DNA and the extracted DRIP DNA was used with PowerUp SYBR Green Mastermix (Applied Biosystems) and run on 7500 Fast Real-Time PCR System (Applied Biosystem). DNA concentrations were calculated using the ΔΔCt method and adjusted for the different concentrations of input and eluted DRIP DNA (adjusting factor = 0.07407). Results are shown as percent input. All PCR reactions were performed in triplicate. Primers used for DRIP-qPCR experiments are listed in Table 1.

8. Immunofluorescence, LacI-FokI-mCherry double strand break reporter system

U2OS-265 cells were used for immunofluorescent analysis. Cells were seeded at 30% confluency onto 12mm round coverslips (Fisher) in 24-well plates. The next day, cells were transfected with plasmids according to Lipofectamine 3000 protocol (Invitrogen). After 48 hours, cells were treated with 500 nM 4-OHT (Sigma T176) and Shield-1 (1:1000 dilution, Takara 632189) for 2-4 hours then washed once with PBS and fixed for 10 minutes with 4% paraformaldehyde. Coverslips were washed three times in PBS, incubated with 0.5% Triton for 5 minutes to permeabilize cells then blocked overnight at 4°C in 10% FBS. Primary antibodies (1:200 dilution in 5% FBS) were incubated with coverslips for 30 minutes, listed in Table 4. Following 3 washes with PBS for 5 minutes while rocking, the appropriate Alexa488-conjugated secondary antibodies were incubated as previously mentioned. Coverslips were washed then mounted onto slides using Immuno-mount (ThermoFisher) mounting medium containing 1 µg/ml of 4′,6-diamidino-2-phenylindole (DAPI). Cells were imaged using a Zeiss M1 fluorescence microscope at 63X magnification. ~100 cells were imaged for each condition and analyzed using ImageJ. Intensity of the transfected protein of

interest at the mCherry foci and the nucleus (excluding the nucleolus) was measured and used to calculate DNA damage exclusion pattern ((FokI colocalized – nuclear intensity)/nuclear intensity). The average was taken for all cells and standard was calculated.

Table 1. Primers used for DRIP-qPCR

Name	Forward primer sequence	Reverse primer sequence
rDNA 1.0 kb	GGCGGTTTGAGTGAGACGAGA	ACGTGCGCTCACCGAGAGCAG
rDNA 4.0 kb	CGACGACCCATTCGAACGTCT	CTCTCCGGAATCGAACCCTGA
rDNA 8.0 kb	AGTCGGGTTGCTTGGGAATGC	CCCTTACGGTACTTGTTGACT
rDNA 13 kb	ACCTGGCGCTAAACCATTCGT	GGACAAACCCTTGTGTCGAGG
rDNA 18 kb	GTTGACGTACAGGGTGGACTG	GGAAGTTGTCTTCACGCCTGA
rDNA 27 kb	CCTTCCACGAGAGTGAGAAGCG	CTCGACCTCCCGAAATCGTACA
rDNA 32 kb	GGAGTGCGATGGTGTGATCT	TAAAGATTAGCTGGGCGTGG
rDNA 42 kb	GCTTCTCGACTCACGGTTTC	CCGAGAGCACGATCTCAAA
EGR1	TTCGGATTCCCGCAGTGT	TCACTTTCCCCCCTTTATCCA
ActG	CTTTCGCTGTTCCAGGCTCT	AACGCAGGCAGAAACCAAAT
RPPH1	GTGCGTCCTGTCACTCCACT	TTCCAAGCTCCGGCAAAGGA
RPS21	ACTTCGGGACATCGTGGACT	TCTCAGAGCACCCCTCTTCC

Table 2. Primers used for subcloning

Name	**Sequence**
BamF1	gggGGATCC ccgggaaaactccgtagtgac
BamF2	gggGGATCC atatctgaagaaactattaaacttctc
BamF3	gggGGATCC acgggggtgtgcatctgcttttat
XhoR1	gggCTCGAG ttattgaccaaatgctttactgaaactc
XhoR2	gggCTCGAG ttagcctttgagtttaatagtttcttcagatat
XhoR3	gggCTCGAG ttattcttccttgtgctgataaaagcagat

Table 3. Primers used for sequencing

Name	**Sequence**
DDX21 S1	accccagtga agctgccagt
DDX21 S2	tggaccagatgtt ggatatggga
DDX21 S3	ttcatcgatc cgggcggaca

Table 4. Primary antibodies

Target	Species	Company	Dilution
Flag	Mouse	Sigma Aldrich F1804	1:5000
Mono-methyl arginine	Rabbit	Cell Signaling Technology 8015S	1:1000
Symmetric dimethyl arginine (Sym10)	Rabbit	EpiCypher 13-0012	1:1000
Asymmetric dimethyl arginine (Asym26)	Rabbit	EpiCypher 13-0011	1:1000
DDX21	Rabbit	Novus NB100-1718	1:1000
Tubulin	Mouse	Sigma Aldrich T6199	1:1000
γH2AX	Mouse	Millipore 05636	1:1000 *for IF

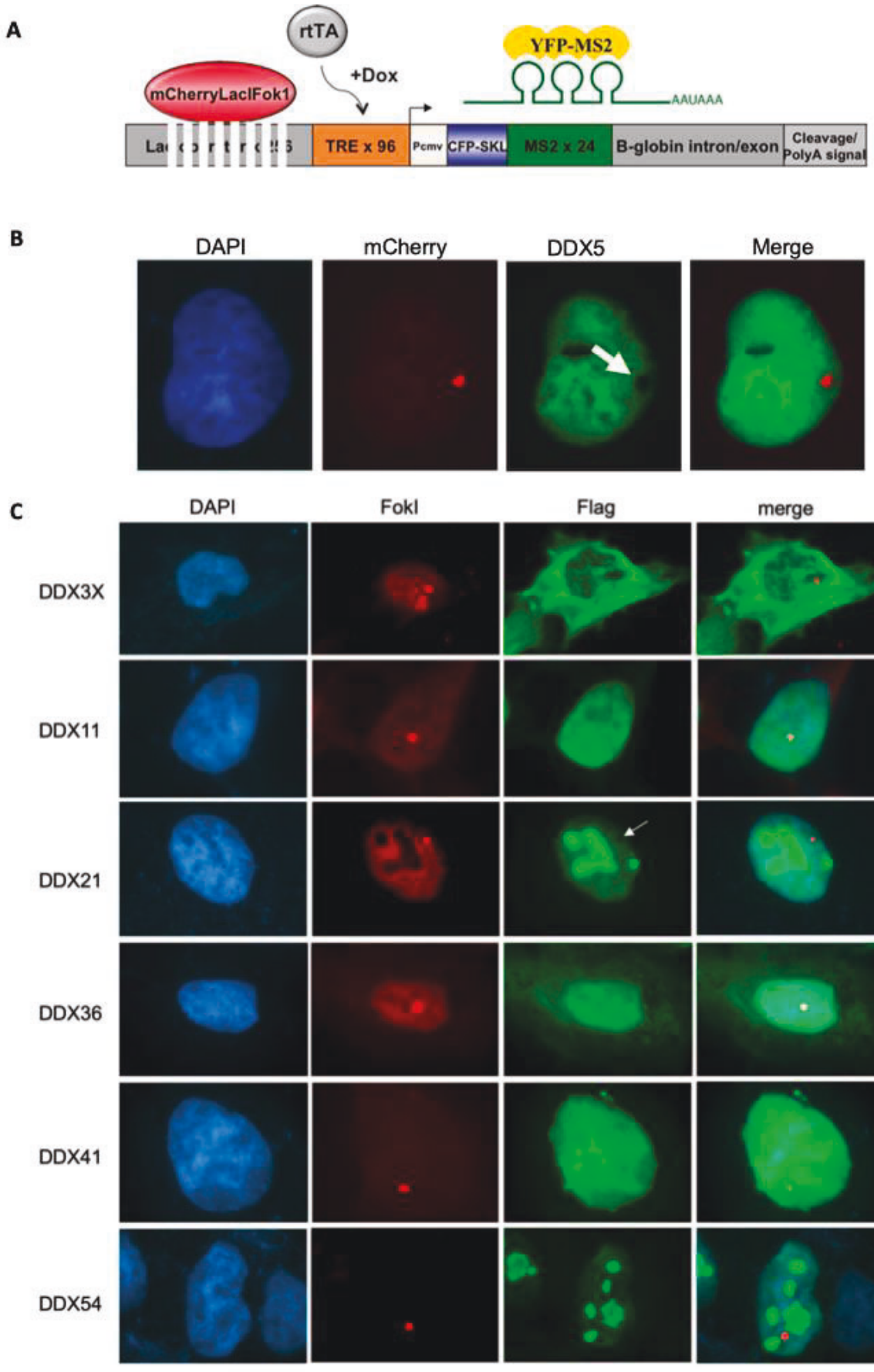
Results

1. DDX21 is excluded from sites of FokI-induced double strand breaks

Our lab has previously studied the RNA helicase DDX5 and its role in DNA damage repair. The LacI-FokI-mCherry double strand break reporter system was used to visualize the recruitment or exclusion of proteins to sites of induced DNA damage. This system was developed to study the effect of DNA damage on local transcription (Shanbhag and Greenberg, 2013). U2OS-265 cells have genomically integrated Lac operator repeats which bind the stably expressed mCherryLacI-FokI fusion protein (figure 1A). The LacI subunit specifically targets the fusion protein to the Lac operator repeats, resulting in the formation of double strand breaks by the FokI non-specific nuclease domain. The mCherry tag of this protein allows for the direct visualization of induced DNA double strand breaks. As previously mentioned, this system was developed to study the effect of DNA damage on local transcription. Treatment with doxycycline induces transcription of the reporter gene, which contains MS2 repeats. The transcripts can be visualized through the binding of YFP-MS2 coat proteins to the stem-loop structured RNA. Furthermore, the RNA encodes a CFP gene allowing for the visualization of the reporter protein following translation. A diagram illustrating the entire system is shown in Figure 1A. For our purposes, this system will only be used to study the induction of DSBs and how this affects localization of proteins of interest. The role of DDX5 in DNA damage repair was identified using this system. As illustrated in figure 1B, DDX5 is excluded from the site of FokI-induced DNA damage. The findings from Zhenbao et al. in 2020 prompted this study to identify additional DEAD-box helicases similarly involved in DNA damage repair (Yu et al., 2020).

Six DDX helicases were tested using the LacI-FokI-mCherry DSB system. Transfected U2OS-265 cells were treated with 4-OHT and Shield-1 for 2 hours to induce DNA damage. Flag-tagged

proteins of interest were visualized through immunofluorescence. Figure 1C shows the result from this screen which includes the following proteins; Flag-DDX3X, Flag-DDX11, Flag-DDX21, Flag-DDX36, Flag-DDX41, and Flag-DDX54. Localization of DDX3X, DDX11, DDX21, DDX36, DDX41 and DDX54 is consistent with published immunofluorescence data (Iwamoto et al., 2008; Lee et al., 2008; Parish et al., 2006; Singh et al., 2022; Yu et al., 2021; Zhang et al., 2014). Looking for recruitment or exclusion at the mCherry foci, DDX21 was the only protein identified to show exclusion. Exclusion is defined as a considerable loss of immunofluorescence signal at the mCherry foci compared to adjacent area. The site of exclusion is highlighted with a white arrow. No considerable change in localization was observed at the mCherry foci for DDX3X, DDX11, DDX36, DDX41 and DDX54. More examples of DDX21 exclusion are shown in Figure 1D. Figure 1E illustrates the co-localization of γ-H2AX with the mCherry foci confirming the formation of DNA double strand breaks. This experiment led to the identification of DDX21 as a potential player in DNA damage repair.


A
rtTA
+Dox
YFP-MS2
mCherryLacIFok1
AAUAAA
TRE x 96
Pcmv
CFP-SKL
MS2 x 24
B-globin intron/exon
Cleavage/
PolyA signal
B
DAPI
mCherry
DDX5
Merge
C
DAPI
FokI
Flag
merge
DDX3X
DDX11
DDX21
DDX36
DDX41
DDX54

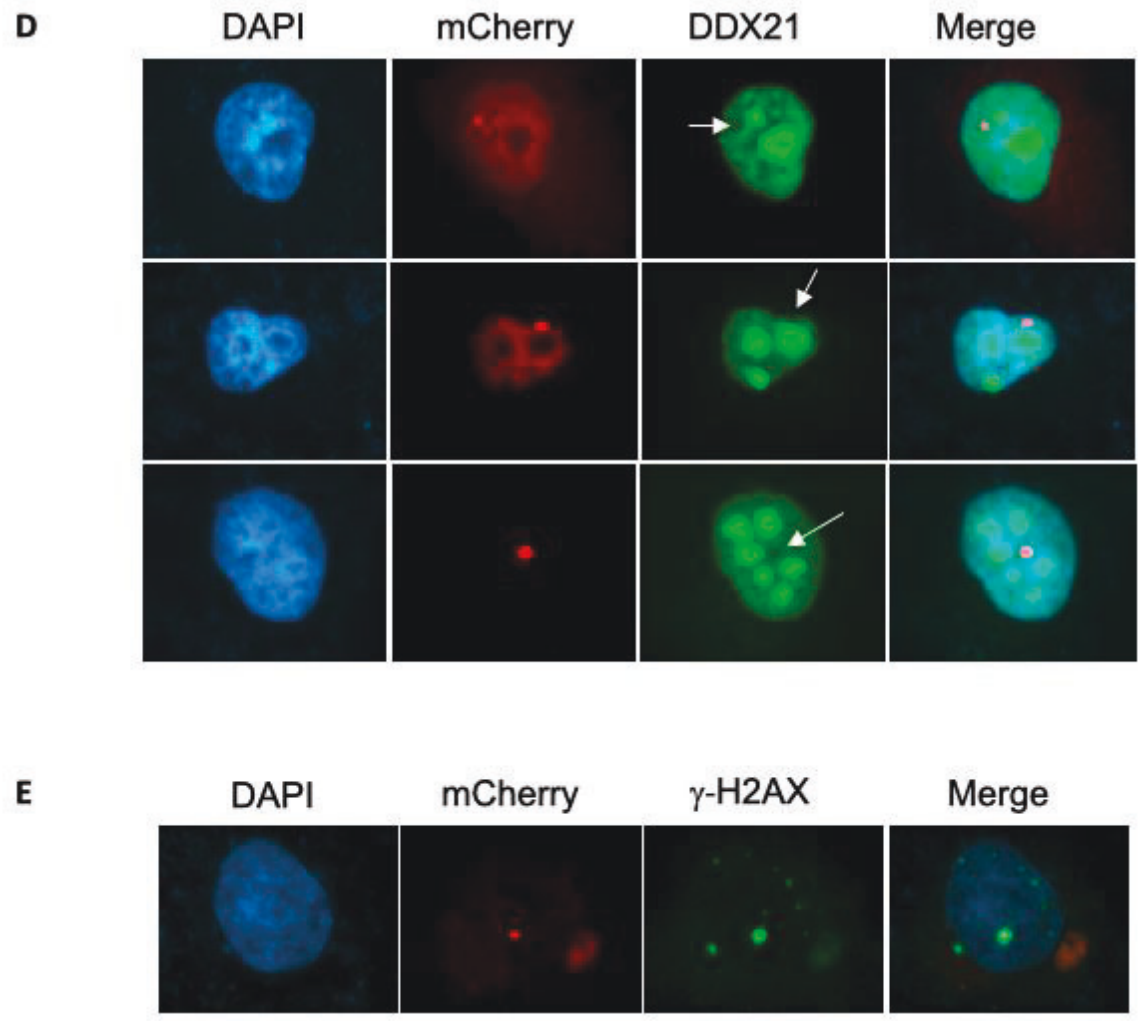


Figure 1. DDX21 RNA helicase excluded from FokI-induced site of DNA damage.

A. Diagram of genomic sequence used in the LacI-FokI-mCherry DNA double strand break reporter system of U2OS-265 cells. Pictured is the mCherryLacIFokI fusion protein bound to Lac operon where it will create double strand breaks. Additional elements of this system are illustrated but were not used in this study. Adapted from (Shanbhag and Greenberg, 2013).

B. DDX5 is excluded from the mCherry site of DNA damage.

C, D, E. U2OS-265 cells transfected with the indicated Flag-tagged plasmid were treated with 4-OHT and Shield-1 for 2 hours. Cells were fixed with 4% PFA, stained with specific antibodies and mounted with DAPI, as described in Materials and Methods. White arrows highlight the sites of exclusion. Images taken at 63X magnification

D. DDX21 is excluded from the FokI loci. Shows further examples of exclusion.

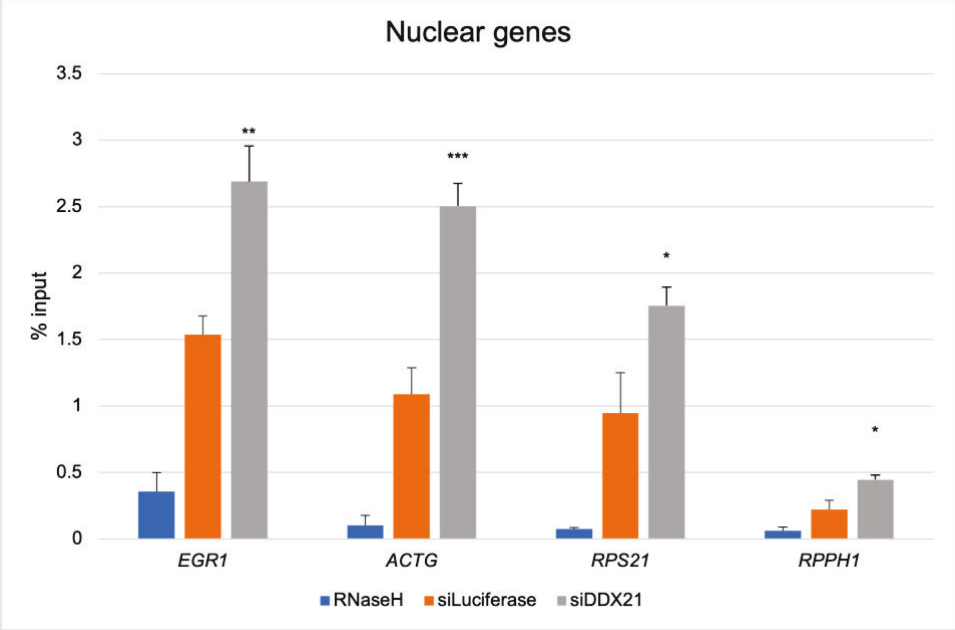
E. γ-H2AX staining of U2OS-265 cells transfected with Flag-DDX21.

2. Depletion of DDX21 leads to accumulation of R-loops at ribosomal genes

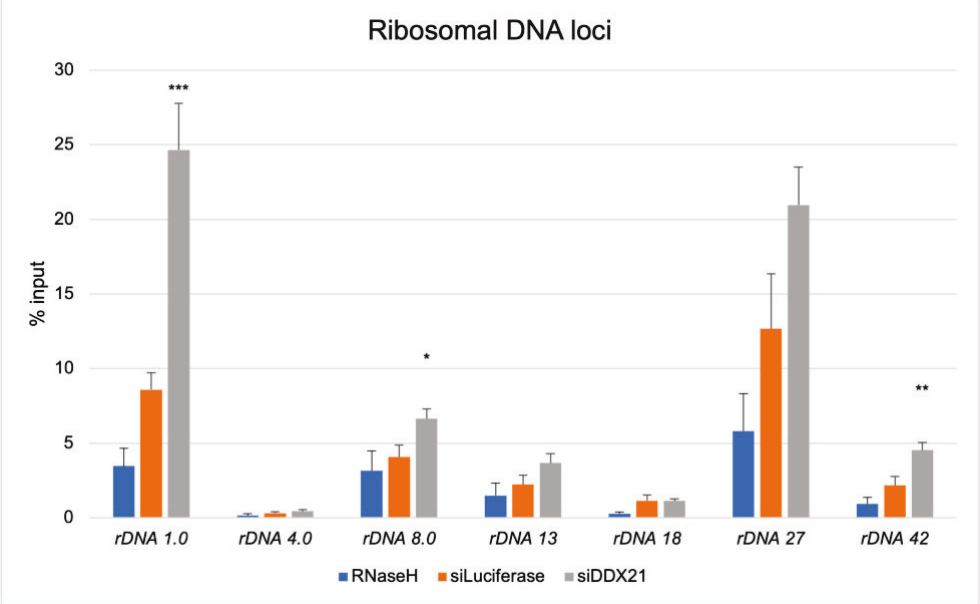
DDX5 was previously identified to be excluded from DSBs in a transcription dependent-manner and it was found to resolve RNA:DNA hybrids near the sites of induced DNA damage (Yu et al., 2020). It was hypothesized that the observed exclusion of DDX21 from figure 1 may be occurring through a similar mechanism. DDX21 has been identified to affect R-loop formation (Song et al., 2017). In order to confirm the role of DDX21 in R-loop resolution, we performed DRIP-qPCR. Figure 2 shows a substantial increase in R-loop levels with knockdown of DDX21 when compared to siLuciferase treatment. DRIP-qPCR on RNaseH treated genomic DNA is not expected to detect elevated levels of RNA:DNA hybrids and therefore serves as a negative control in these experiments. Nuclear gene loci established to form R-loops, such as *EGR1*, and DDX21 known target loci, *RPS21*, were tested in figure 2A. Significant increases in R-loop levels were detected when examining these nuclear genes. Since DDX21 is a well-known nucleolar RNA helicase involved in facilitating ribosomal RNA biogenesis (Calo et al., 2015), we were interested in studying whether it regulates R-loop formation at these loci. Ribosomal DNA (rDNA) encodes for ribosomal RNA and exists as repetitive sequences within nucleolus (Wang and Lemos, 2019). Calo et al. identified DDX21 binding at rDNA loci at various distances from the transcription start site (TSS). Peak binding was detected at 1.0 to 13 kb from the TSS, which spans the rRNA coding region (Calo et al., 2015). DRIP-qPCR was performed at the loci identified by Calo et al. and is presented in figure 2B. DDX21 plays a significant role in resolving R-loops at rDNA but the extent of regulation varies throughout the region. DRIP-qPCR levels follow a similar trend to the ChIP-qPCR data presented by Calo et al. with the exception of the 27 kb loci. Figure 2C shows the efficient knockdown of DDX21 in cells used for DRIP-qPCR analysis. Overall, this confirms that

DDX21 regulates R-loop levels and specifically identifies an important role for DDX21 in controlling R-loop formation at ribosomal DNA.

A


Nuclear genes
3.5
3
2.5
2
1.5
1
0.5
0
% input
**
***
*
*
EGR1
ACTG
RPS21
RPPH1
RNaseH
siLuciferase
siDDX21

B


Ribosomal DNA loci
30
25
20
15
10
5
0
% input
***
*
**
rDNA 1.0
rDNA 4.0
rDNA 8.0
rDNA 13
rDNA 18
rDNA 27
rDNA 42
RNaseH
siLuciferase
siDDX21

C

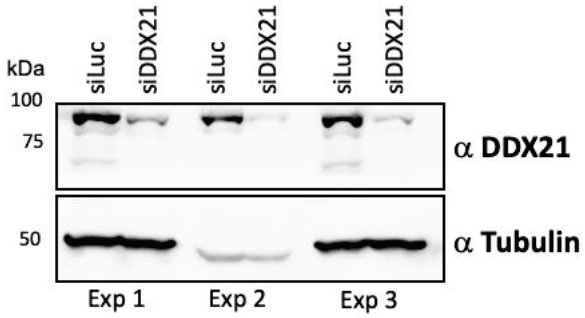

kDa
siLuc
siDDX21
siLuc
siDDX21
siLuc
siDDX21
100
75
50
α DDX21
α Tubulin
Exp 1
Exp 2
Exp 3

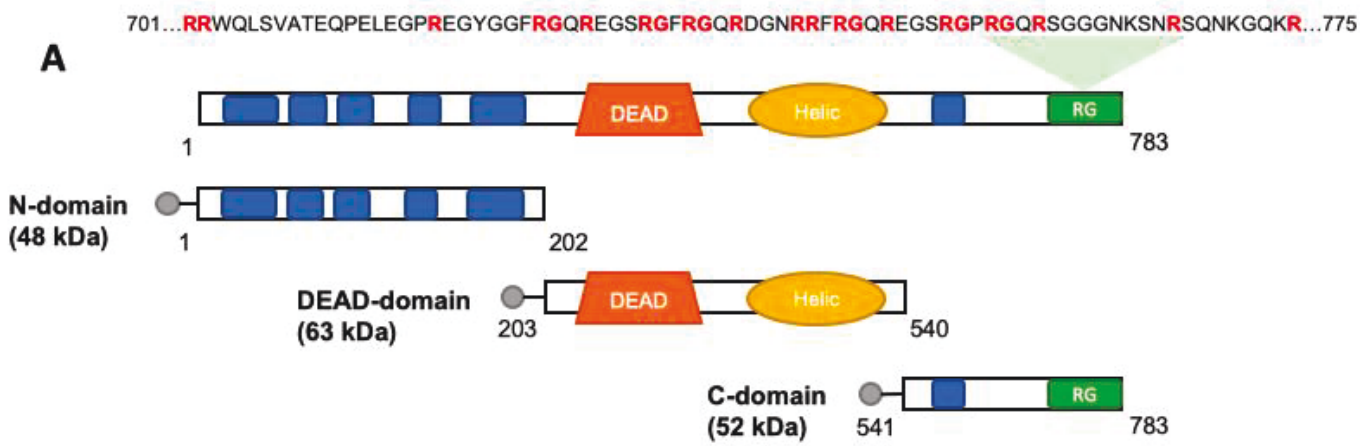
Figure 2. DDX21-deficient cells accumulate R-loops at target loci.

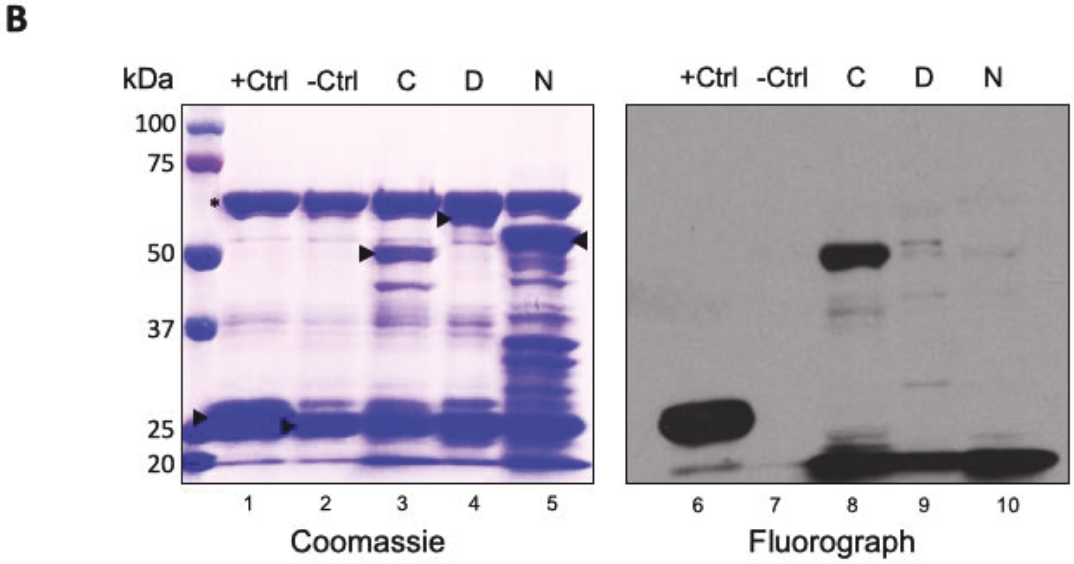
A, B. U2OS cells treated with siDDX21 or siLuciferase were used for DRIP-qPCR analysis. The average and SEMs of 6 independent IPs, made from 2 biological replicates, are shown. Statistical analysis was performed using Student's t-test. *P<0.05, **P<0.01, and ***P<0.001.

C. Western-blot of total cell lysates from siDDX21 and siLuciferase transfected cells.

3. C-domain of DDX21 is arginine methylated *in vitro*

DDX21 has an extensive C-terminal RGG/RG motif from amino acids 699 to 741, as illustrated in Figure 3A. We were therefore interested in determining whether DDX21 is a substrate of PRMT1 and PRMT5. To test this hypothesis, we first performed *in vitro* methylation assays using GST-DDX21 fusion proteins and purified PRMT1 and PRMT5. We created three human DDX21 constructs consisting of; the N-terminal region (aa 1-202, N domain), the catalytic core (aa 203-540, DDX domain), and the C-terminal region (aa 541-762, C domain) (Figure 3A). These truncated DDX21 proteins were incubated with either GST-PRMT1 or recombinant PRMT5/MEP50 complex (Sigma) and radioactively labeled (^{3}H)-SAM for the identification of methylated residues. A positive control (GST-RBMX aa 366-391 (Cai et al., 2021)) and negative control (GST) were used in both PRMT1 and PRMT5 assays. We found that only the C-domain of DDX21 was a substrate of PRMT1 (Figure 3B, lane 8). Similarly, assays with PRMT5 showed very minor methylation only for the C-domain (Figure 3C, lane 8). The other domains lacking RGG/RG motifs were not methylated. These findings confirm that DDX21 can be methylated by either PRMT1 and PRMT5 *in vitro*.


701…RRWQLSVATEQPELEGPREGYGGFRGQREGSRGFRGQRDGNRRFRGQREGSRGPRGQRSGGGNKSNRSQNKGQKR…775
A
DEAD
Helic
RG
1
783
N-domain
(48 kDa)
1
202
DEAD-domain
(63 kDa)
203
540
C-domain
(52 kDa)
541
783


B
kDa
+Ctrl
-Ctrl
C
D
N
100
75
50
37
25
20
1
2
3
4
5
Coomassie
+Ctrl
-Ctrl
C
D
N
6
7
8
9
10
Fluorograph

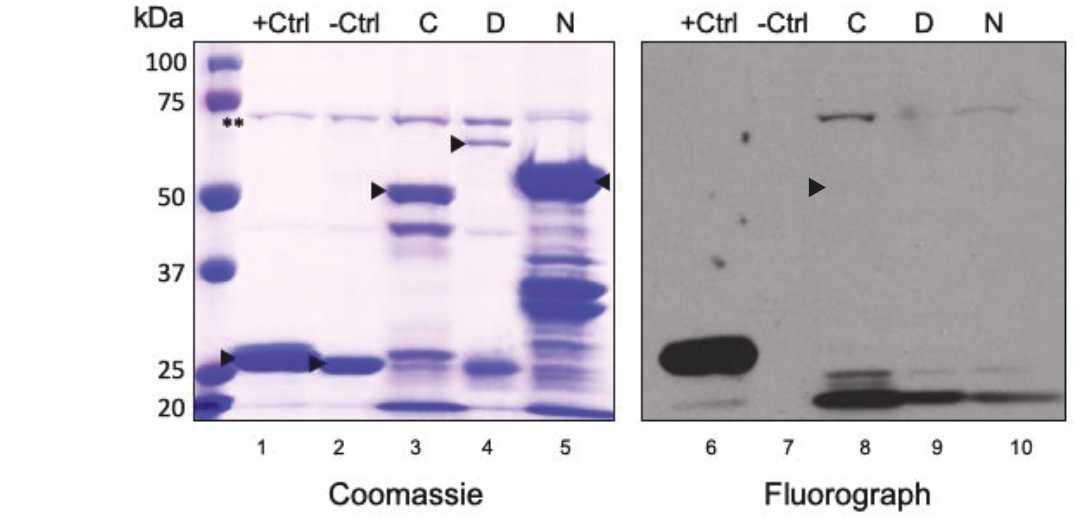

C
kDa
+Ctrl
-Ctrl
C
D
N
100
75
50
37
25
20
1
2
3
4
5
Coomassie
+Ctrl
-Ctrl
C
D
N
6
7
8
9
10
Fluorograph

Figure 3. C-terminal domain of DDX21 is arginine methylated in vitro.

A. Diagram of DDX21 full length and GST domain constructs. The C-terminal RGG/RG motif is highlighted.

B, C. Coomassie stain and *in vitro* methylation assays using the indicated GST-DDX21 constructs with either GST-PRMT1 (B) or PRMT5/MEP50 (C). Arrowheads indicate GST-constructs of interest. *indicates GST-PRMT1 (66 kDa) and **indicates PRMT5/MEP50 (72 kDa).

4. DDX21 is arginine methylated in HEK293T cells

Due to the discovery of DDX21 arginine methylation *in vitro*, we next wanted to test the status of DDX21 arginine methylation in cells. Flag-DDX21 truncations plasmids were synthesized to confirm methylation in the C-domain. Additionally, a DDX21 mutant plasmid was created where the arginine residues in the C-terminal RGG/RG motif were mutated to lysines (R701, R702, R718, R725, R728, R732, R735, R738, R742, R743, R745, R748, R752, R755, R758, R767, R775 replaced with K) (Figure 4A). We then immunoprecipitated the overexpressed Flag-DDX21 plasmids and performed western blot analysis. Symmetric dimethylation was detected in the full length DDX21 protein in cells, as well as asymmetric and monomethylation (Figure 4B, lane 7, Sym10 panel). Furthermore, the symmetric methylation was still detected in Flag-DDX21ΔN, whereas the deletion of the C-domain resulted in a considerable loss in methylation (Figure 4B, lane 8 and 9, Sym10 panel). This confirms that DDX21 is in fact arginine methylated in cells and demonstrates that methylation occurs in the C-terminal RGG/RG motif. The band at 100 kDa in lane 6 is not caused from underloading as the IgG heavy chain at 50 kDa from the Flag antibody can be used as a loading control. The band at 100 kDa in lanes 6, 8 and 9 are likely non-specific proteins and not caused by spill-over from lanes 7/10.

Surprisingly, the methyl-antibodies also detected Flag-DDX21 R-K suggesting that the full length DDX21 is likely also arginine methylated at other regions (Figure 4B, lane 10). Taken together, these data show that DDX21 is arginine methylated in cells.

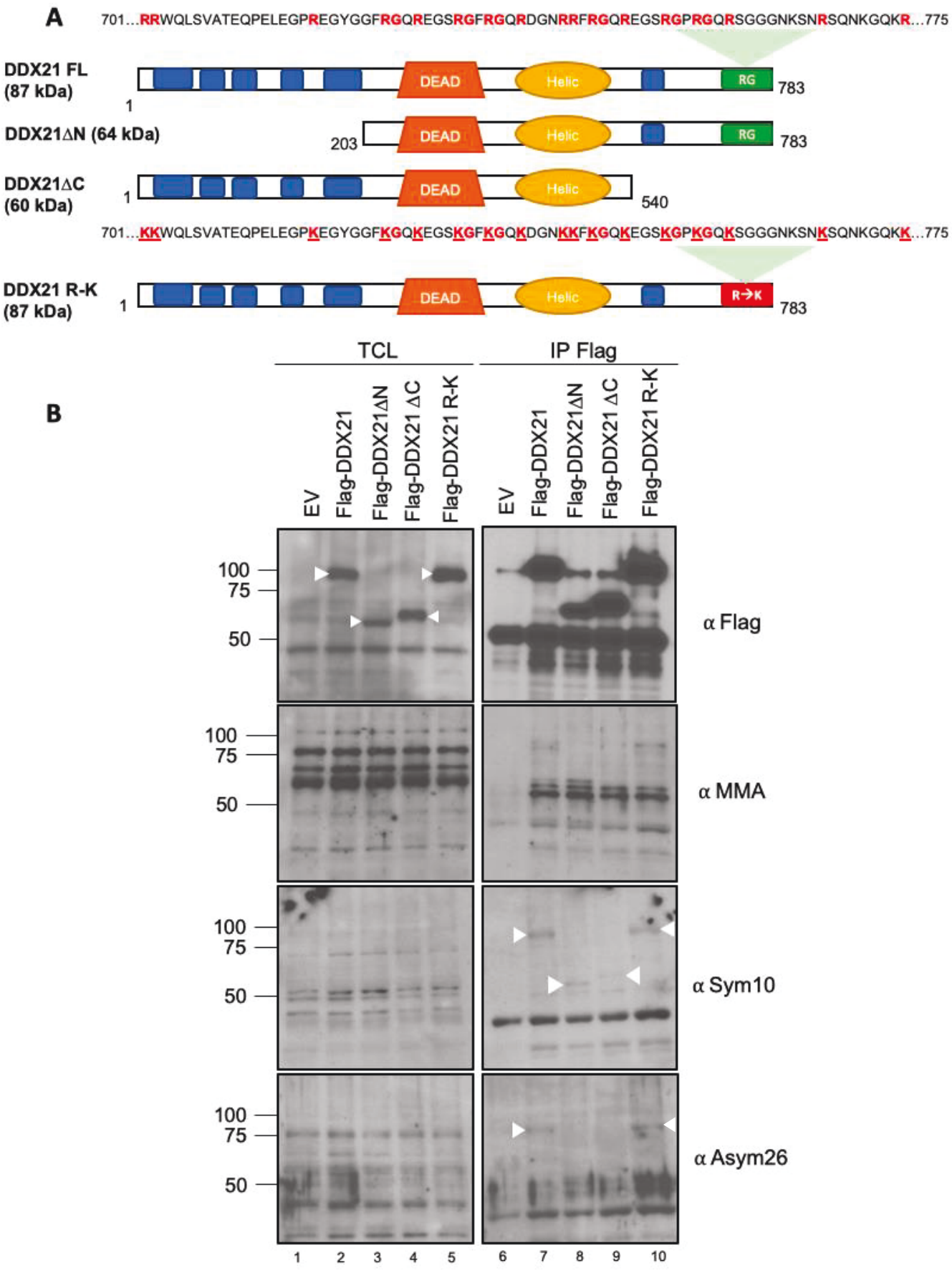

A
701…RRWQLSVATEQPELEGPREGYGGFRGQREGSRGFRGQRDGNRRFRGQREGSRGPRGQRSGGGNKSNRSQNKGQKR…775
DDX21 FL
(87 kDa)
1
DEAD
Helic
RG
783
DDX21ΔN (64 kDa)
203
DEAD
Helic
RG
783
DDX21ΔC
(60 kDa)
1
DEAD
Helic
540
701…KKWQLSVATEQPELEGPKEGYGGFKGQKEGSKGFKGQKDGNKKFKGQKEGSKGPKGQKSGGGNKSNKSQNKGQKK…775
DDX21 R-K
(87 kDa)
1
DEAD
Helic
R→K
783
B
TCL
IP Flag
EV
Flag-DDX21
Flag-DDX21ΔN
Flag-DDX21 ΔC
Flag-DDX21 R-K
100
75
50
α Flag
α MMA
α Sym10
α Asym26
1
2
3
4
5
6
7
8
9
10

Figure 4. DDX21 is arginine methylated in HEK293T cells.

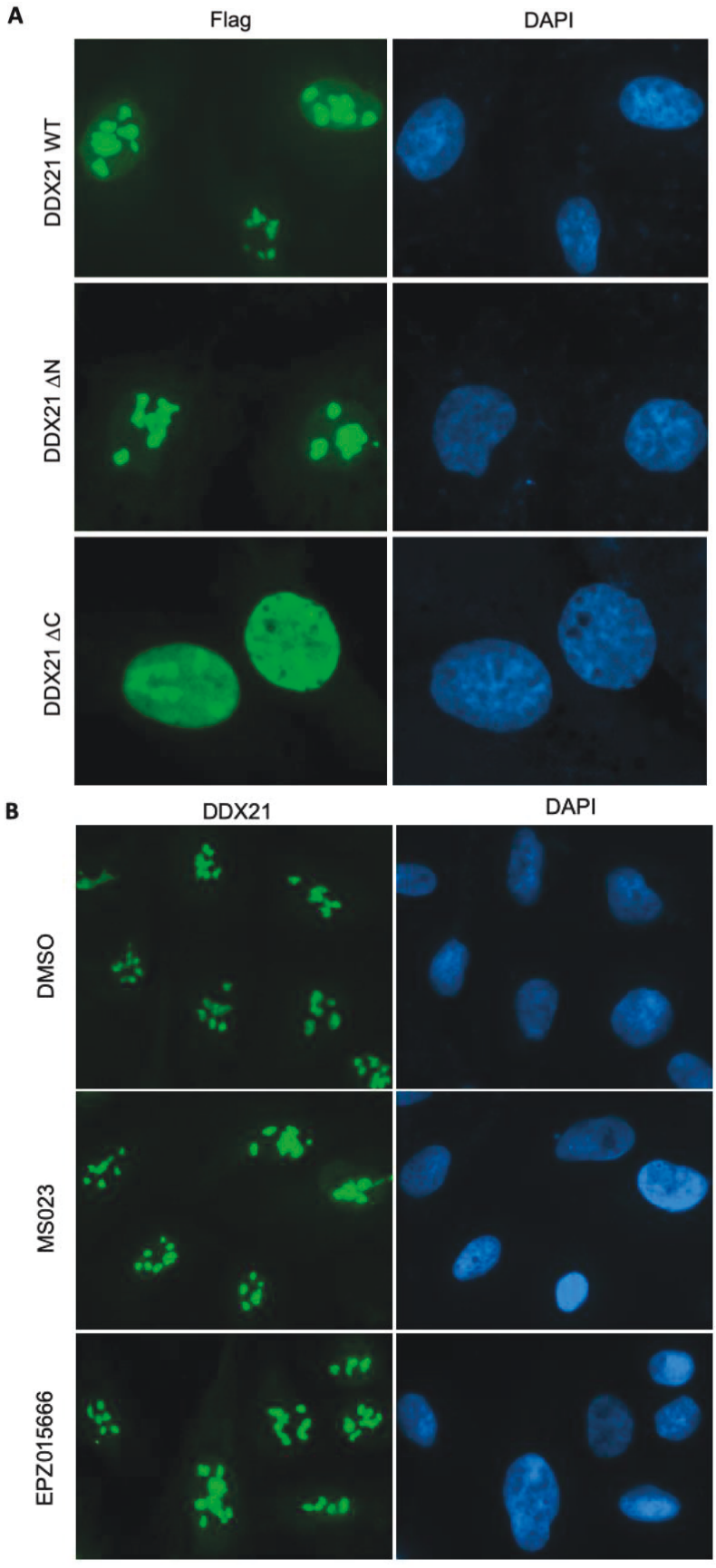
A. Schematic diagram of Flag-DDX21 proteins used in immunoprecipitation experiments.

B. HEK293T cells were transfected with Flag-DDX21 plasmids illustrated in panel A. Cell lysates were immunoprecipitated with anti-Flag beads. Total cell lysates and purified proteins were analyzed by western blot using anti-Flag, anti-monomethylarginine, anti-symmetrical dimethylarginine or anti-asymmetrical dimethylarginine antibodies. Arrowheads indicate bands of interest.

5. The C-domain of DDX21 is essential for nucleolar localization

We next decided to test the effect of the N-terminal and C-terminal domains on cellular localization of DDX21. Immunofluorescence was performed using the Flag-DDX21 truncations plasmids, DDX21ΔN and DDX21ΔC transfected in U2OS cells. We observed normal nucleolar localization with the full length DDX21 (Figure 5A). A similar localization pattern was observed when the N-terminal domain was deleted (DDX21ΔN). However, a distinct nucleolar localization pattern was absent when the C-terminal domain was deleted (DDX21ΔC). DDX21ΔC was diffused throughout the nucleoplasm with only partial nucleolar localization (Figure 5A). These findings reveal that the C-domain is essential for the strong nucleolar localization of DDX21.

We next wished to test whether the methylation of the RGG/RG motif within the C-domain regulates its nucleolar localization. These experiments were done before the subcloning of the DDX21 R-K construct, therefore this section makes use of arginine methylation inhibitors and the siRNA knockdown of PRMT1/PRMT5 and their effect on endogenously expressed DDX21. U2OS cells were treated with the type 1 PRMT inhibitor MS023 or the type 2 PRMT inhibitor EPZ015666 to examine the localization of DDX21 (Chan-Penebre et al., 2015; Eram et al., 2016). Immunofluorescence revealed that treatment with methylation inhibitors did not affect the localization of DDX21 (Figure 5B). To confirm that methylation of the RGG/RG motif in the C-domain of DDX21 is not playing a role in its nucleolar localization, we depleted cells of PRMT1 and PRMT5 with siRNAs then examined the cellular distribution of DDX21 in the cell. We observed no significant change in DDX21 localization with siPRMT1 or siPRMT5 compared to siLuciferase (Figure 5C). Overall, this reveals the importance of the C-domain sequence for proper localization of DDX21 and it is independent of PRMT1 and PRMT5.


A
Flag
DAPI
DDX21 WT
DDX21 ΔN
DDX21 ΔC
B
DDX21
DAPI
DMSO
MS023
EPZ015666

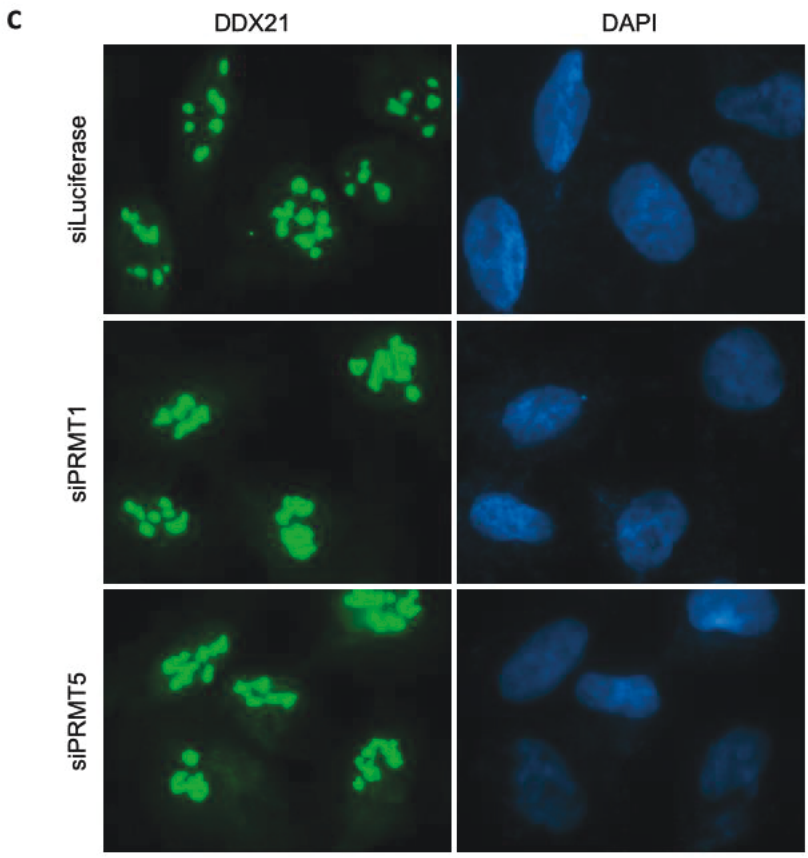


Figure 5. C-domain of DDX21 is required for nucleolar localization.

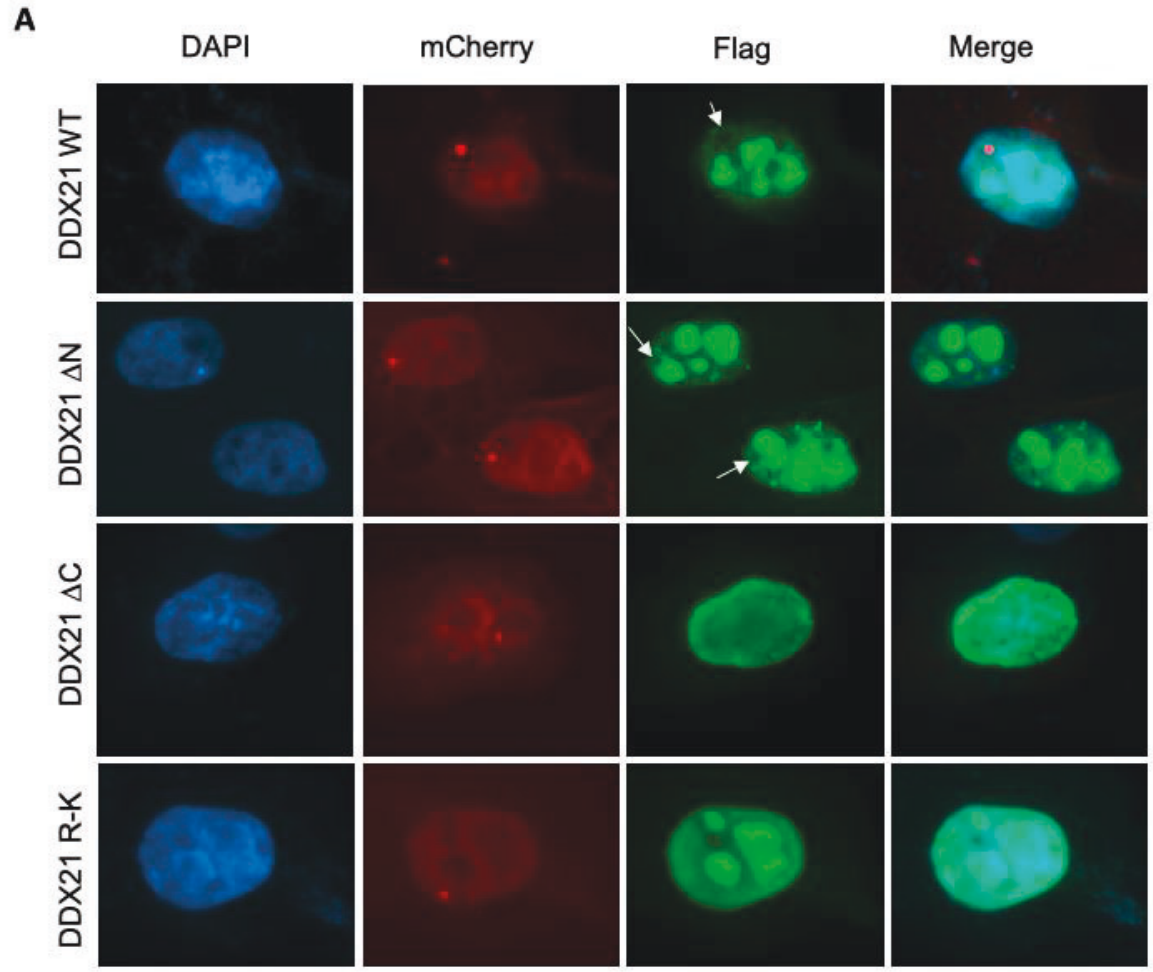
A. U2OS-265 cells were transfected with the indicated Flag-DDX21 plasmid. Cells were fixed with PFA, stained with anti-Flag antibody, and mounted with DAPI. Representative images are shown taken at 63X magnification.

B. U2OS cells were treated with MS023 or EPZ015666 for 48 hours. Cells were fixed with PFA, stained with anti-DDX21 antibody and mounted with DAPI. Representative images are shown taken at 63X magnification.

C. U2OS cells were transfected with siLuciferase, siPRMT1 or siPRMT5. Cells were fixed with PFA, stained with anti-DDX21 antibody and mounted with DAPI. Representative images are shown taken at 63X magnification.

6. DDX21 RGG/RG motif is required for exclusion from FokI-induced DSBs

Previously, we showed that the exclusion of DDX5 from DSBs was dependent on its C-terminal RGG/RG motif (Yu et al., 2020). We therefore wanted to test whether the same was true for the exclusion observed with DDX21. Using the LacI-FokI-mCherry DSB inducible system, we tested the DDX21 mutants for their ability to be excluded following DNA damage. DDX21 was excluded from the mCherry induced sites of DNA damage, as expected (figure 6A). Similarly, Flag-DDX21ΔN was also excluded from the FokI loci, revealing that the N domain does not affect its role in DNA damage (Figure 6A, indicated with white arrows). The deletion of the C-domain however abolished the exclusion pattern observed with DDX21 and DDX21ΔN. Finally, we observed that the DDX21 R-K mutant also lost the exclusion pattern seen with DDX21 (Figure 6A). Intensity analysis showed that the DNA damage exclusion pattern of DDX21ΔC and DDX21 R-K was significantly lost when compared to DDX21 WT and DDX21 ΔN (Figure 6B). These findings reveal that the exclusion of DDX21 from the sites of DNA damage is dependent on the arginine residues within the C-terminal RGG/RG motif.


A
DAPI
mCherry
Flag
Merge
DDX21 WT
DDX21 ΔN
DDX21 ΔC
DDX21 R-K

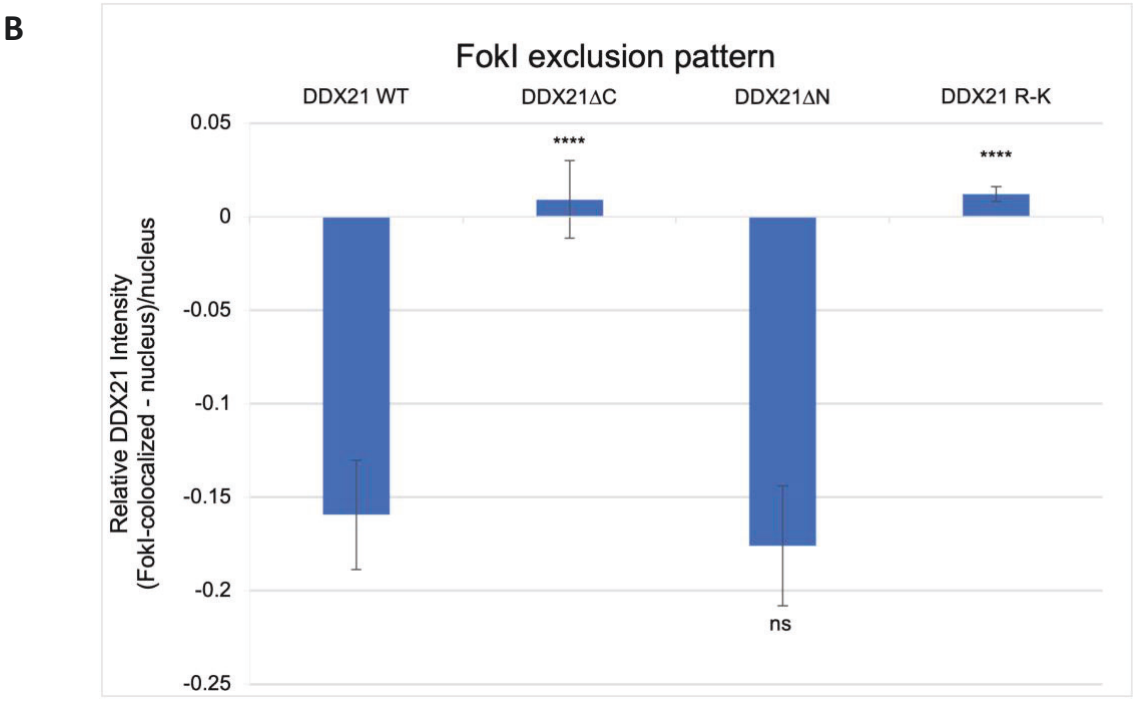

B
FokI exclusion pattern
DDX21 WT
DDX21ΔC
DDX21ΔN
DDX21 R-K
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
****
****
ns
Relative DDX21 Intensity
(FokI-colocalized - nucleus)/nucleus

Figure 6. C-terminal RGG/RG motif is required for exclusion from mCherryFokI foci.

A. U2OS-265 cells transfected with the indicated Flag-DDX21 plasmid were treated with 4-OHT and Shield-1 for 2 hours. Cells were fixed with 4% PFA, stained with anti-Flag antibody and mounted with DAPI, as described in Materials and Methods. White arrows highlight the sites of exclusion. Representative images are shown taken at 63X magnification.

B. DDX21 exclusion from the FokI-induced DNA damage was analyzed as described in Materials and Methods. The fluorescence intensity at the site of FokI-induced DSBs was analyzed relative to the rest of the nucleus (without the nucleolus). Data presented is the mean relative intensity at the mCherry foci from ~70 cells per transfection. Statistical significance was calculated using the Student's t-test. *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Discussion

In this study, we identified a new role for the nucleolar RNA helicase DDX21. We showed that DDX21 was excluded from sites of DSBs, suggesting a possible role in DNA damage repair (Figure 1). We also confirmed the role of DDX21 in regulating levels of R-loops, investigating both RNA polymerase I and II target genes (Figure 2). It was shown that the C-terminal RGG/RG motif is arginine methylated *in vitro* by PRMT1, with very minor methylation detected by PRMT5 (Figure 3). Furthermore, we confirmed that arginine methylation of DDX21 occurs in cells and interestingly observed a strong symmetric dimethylation band through western blotting (Figure 4). The intensity of the methylation band is comparable to that of DDX5 (Mersaoui et al., 2019). Immunofluorescence revealed that the C-domain of DDX21 is essential for proper nucleolar localization and this observation was independent on arginine methylation (Figure 5). Lastly, the mutation of the RGG/RG motif in the C-domain of DDX21 lead to the loss of DSB exclusion that can be seen in DDX21 and DDX21ΔN proteins (Figure 6). Based on previous work with DDX5, we propose that DDX21 is excluded from DNA damage sites to resolve adjacent R-loops in order to promote efficient DNA damage repair processes.

DDX21 has been previously identified to resolve R-loops in the cell (Song et al., 2017). However, our study specifically examined the role of DDX21 in resolving R-loops at the ribosomal DNA loci. We found that DDX21 significantly regulated RNA:DNA hybrids at rDNA specifically at 1.0 kb from the TSS (Figure 2B). It is known that the ribosomal DNA loci accumulates RNA:DNA hybrids with highest detected level within the coding region and a peak detected at the 5' region of the 18S subunit (El Hage et al., 2010). This correlates with the DRIP-qPCR data that was produced in this study, where we observe a peak in DRIP signal at the 1.0 kb loci. However, we do observe strong R-loop formation at the 27 kb loci, which represents the intergenic spacer

repeat region of rDNA (Calo et al., 2015). R-loop signal within IGS non-coding ribosomal DNA regions have been detected previously (Abraham et al., 2020). This region is not significantly regulated by DDX21 which correlates with low DDX21 ChIP signal at this locus in the study done by Calo et al. 2015. Our data demonstrates a new role for DDX21 in nucleolar DNA, where it functions to resolve R-loops during the synthesis of ribosomal RNA.

Figure 3B shows that the C terminal domain of DDX21 is methylated by PRMT1, which catalyzes asymmetric dimethylation (lane 8). However, experiments in cells provide evidence that DDX21 is mainly symmetrically dimethylated (Figure 4B, lane 7). Though our in vitro methylation assay with PRMT5 and DDX21 domains did not show a significant level of methylation (Figure 3C, lane 8). This discrepancy may be due to the lack of cofactors or binding partners in the *in vitro* experiment. PRMTs are often regulated through protein interactions to provide target specificity and localization to various organelles (Stopa et al., 2015). For example, PRMT5 is known to interact with pICIn in the cytoplasm to stimulate the methylation of the Sm proteins, essential for spliceosome complex formation (Pesiridis et al., 2009). RioK1 competes with pICIn for PRMT5 binding and similarly stimulates methylation specifically for nucleolin (Guderian et al., 2011). It was shown that nucleolin methylation occurs due to the recruitment by RioK1, as it was found to only bind RioK1 and not PRMT5 in pull down assays (Guderian et al., 2011). Additionally, symmetric dimethylation of nucleolin was significantly reduced upon the depletion of RioK1 (Guderian et al., 2011). This data demonstrates how binding of other proteins can be essential for PRMT5 methylation. Using PRMT5 complexes purified from cell extracts as opposed to recombinant PRMT5/MEP50 may increase the observed in vitro methylation of the DDX21 C-domain. Mass spectrometry and Co-IP experiments can be performed to identify new components of a PRMT5-DDX21 complex in cells.

Western blot analysis of DDX21 methylation in cells shows strong symmetric dimethylation (Figure 4B, Sym10 panel, lane 7). The expression of a DDX21 R-K mutant was used to confirm arginine methylation occurs in the RG-rich region within the C domain. However, symmetric dimethylation was still observed in the Flag-DDX21 R-K mutant (Figure 4B, Sym10 panel, lane 10). Although the band appears slightly less intense than the Flag-DDX21 WT, we expected to see a more significant loss of methylation. It is possible that other arginines within the full-length protein are also arginine methylated. The sequence of DDX21 contains 26 other arginine residues located outside of the 17 C-terminal arginines mutated in DDX21 R-K. These have the potential to be methylated by PRMTs and could be the explanation for the band seen in Figure 4B, Sym10 panel, lane 10. Previous studies showed that symmetric dimethylation was still observed in RBMX-RK and RBMX-ΔRGG proteins in cells whereas only the C-terminal 366-391 aa sequence was found to be methylated in vitro (Cai et al., 2021). This example is similar to the observations made with DDX21. However, the deletion of the DDX21 C domain resulted in a significant decrease in methylation when compared to both DDX21 WT and R-K (Figure 4B, Sym10 panel, lane 9). It is possible that the full C domain structure is required for contact with PRMT5 and its deletion abolishes their interaction and subsequent arginine methylation. This theory is supported by the symmetric dimethylation that is still detected with the DDX21 ΔN protein (Figure 4B, Sym10 panel, lane 8). More DDX21 plasmid constructs should be created with various smaller deletions so as to map the PRMT5 binding site and the sites of DDX21 arginine methylation in cells. Furthermore, arginine methylation of DDX21 should be examined by immunoprecipitating endogenous DDX21 in the presence or absence of methylation inhibitors or siPRMT1/5. This would provide a better understanding of the physiological level of DDX21 arginine methylation.

DEAD-box family RNA helicases are crucial for efficient biogenesis of all RNAs at all stages of metabolism. Structural analysis of DEAD-box RNA helicases has shown that this family binds RNA in an ATP dependent manner but does not require binding to a specific RNA sequence. Since they are identified to bind RNA non-specifically, DEAD-box helicases are regulated through binding of other factors. Specific binding regions of DEAD-box helicases can be mapped, and these regions are often essential for complex formation and therefore functional RNA binding (Ballut et al., 2005). Examining binding partners of DDX21 specifically, it has been found to interact with the nucleolar scaffold protein WDR46 (Hirai et al., 2013). Upon the depletion of WDR46, the nucleolar localization of DDX21 was affected (Hirai et al., 2013). In our study, we found that the C-terminal domain of DDX21 is essential for its distinct nucleolar localization (Figure 5A). We can speculate that the intrinsically disordered RGG/RG motif in the C-domain of DDX21 may be involved in the WDR46 interaction. Hirai et al. identified WDR46 to interact with DDX21 through its intrinsically disordered regions (N and C-terminal domains) for proper organization of nucleoli (Hirai et al., 2013). It is well known that the nucleolus is formed through liquid-liquid phase separation (Guillen-Chable et al., 2021). Importantly, disordered regions in proteins are found in many nucleolar proteins (Guillen-Chable et al., 2021). Therefore, it is likely that the intrinsically disordered C-domain of DDX21 would regulate its separation into the nucleolus, possibly through interaction with WDR46. Pull down experiments with DDX21 FL, DDX21ΔC and DDX21 R-K constructs could be performed to confirm WDR46 binding to the DDX21 C domain and whether this interaction is arginine dependent.

The importance of the RGG/RG motif and how it may affect cellular localization is discernible. However, the role it plays in the exclusion of DDX21 from DNA damage is still uncertain. We found that the mutation of arginine residues in the RGG/RG motif to lysines resulted

in the significant loss of exclusion pattern (Figure 6). Studies performed in our lab previously with DDX5 similarly identified the requirement of the RGG/RG motif in directing exclusion (Yu et al., 2020). The reason for this is still unclear. It is possible that exclusion is mediated through interactions with other proteins through the RGG/RG motif. DDX5 was shown to bind XRN2 in an RGG/RG motif dependent manner, although this interaction was not studied in the context of DNA damage exclusion (Mersaoui et al., 2019). A specific RGG/RG motif interaction with DDX21 may direct its exclusion from DSBs, but this requires further study. Co-IP experiments should be conducted to discern whether DDX21 interacts with XRN2 or other notable DNA damage proteins. This would help to solidify DDX21 as a player in the DNA damage response.

It was identified that exclusion of DDX5 was also dependent on ATM activation, a central DNA damage signalling kinase (Marechal and Zou, 2013; Yu et al., 2020). Inhibition of ATM resulted in the significant loss of exclusion. It is possible that DDX21 exclusion is also dependent on ATM activation, but these experiments were not conducted. DDX21 harbours three serine-glutamine (SQ) and two threonine-glutamine (TQ) sites which are the substrates for ATM or ATR phosphorylation (Marechal and Zou, 2013). PhosphositePlus® has reported several sites of phosphorylation on DDX21, however only one low-confidence SQ site. Therefore, ATM phosphorylation may be the signal that allows for DDX21 to be excluded from the site of DNA breaks. Examination of DDX21 exclusion patterns while using ATM or ATR inhibitors would be required to determine whether activation is necessary. Similarly, studying DDX21 exclusion patterns upon the knock down of well-known DNA damage response proteins, such as BRCA1/2, MRE11 or 53BP1, would help to identify factors directing DDX21 exclusion. In summary, Co-IP and additional exclusion experiments using inhibitors or siRNA knockdowns should be performed to understand the mechanism behind exclusion from DNA damage.

DDX21 has been found to be overexpressed in several cancer types including colorectal cancer, breast cancer and lymphoma (Bonzheim et al., 2013; Tanaka et al., 2020; Zhang et al., 2014). It is also a known autoantigen that causes connective tissue and watermelon stomach diseases (Arnett et al., 1997; Garcia et al., 2000). Considering the implications of DDX21 in disease, it may be an effective therapeutic target. Further studies in its role in DNA damage repair and ribosomal R-loop regulation will help to understand the impact of DDX21 in disease progression and how they might be targeted. Currently, there are no inhibitors for DDX21 in development. However, there are inhibitors which target the ATPase domain of DDX3X that have been shown to impair stress granule formation (Cui et al., 2020). Inhibitors for the catalytic activity of DDX21 may prove to be effective. Although, this study shows that arginine methylation plays a significant role in the DNA damage function of DDX21. Therefore, targeting post-translational modifications of DDX21 is a worthy area of study that may lead to clinical applications of this research.

In conclusion, the results from this study identify a new role for DDX21 in DNA damage response. DDX21 was shown to be excluded from sites of DNA double strand breaks. We define this function to be dependent on its C-terminal RGG/RG motif. Additionally, it was found that the C-terminal domain of DDX21 is arginine methylated. The exclusion of DDX21 from sites of DNA damage is likely due to its role in resolving R-loops, as seen previously with DDX5 (Yu et al., 2020). Further studies using the CRISPR-Cas LMNA HDR system would be necessary to confirm a link between the exclusion of DDX21 from DNA damage sites and the removal of R-loops (Yu et al., 2020).

References

Abraham, K.J., N. Khosraviani, J.N.Y. Chan, A. Gorthi, A. Samman, D.Y. Zhao, M. Wang, M. Bokros, E. Vidya, L.A. Ostrowski, R. Oshidari, V. Pietrobon, P.S. Patel, A. Algouneh, R. Singhania, Y. Liu, V.T. Yerlici, D.D. De Carvalho, M. Ohh, B.C. Dickson, R. Hakem, J.F. Greenblatt, S. Lee, A.J.R. Bishop, and K. Mekhail. 2020. Nucleolar RNA polymerase II drives ribosome biogenesis. *Nature*. 585:298-302.

Aguilera, A., and T. Garcia-Muse. 2012. R loops: from transcription byproducts to threats to genome stability. *Mol Cell*. 46:115-124.

Alexander, J.L., and T.L. Orr-Weaver. 2016. Replication fork instability and the consequences of fork collisions from rereplication. *Genes Dev*. 30:2241-2252.

Allison, D.F., and G.G. Wang. 2019. R-loops: formation, function, and relevance to cell stress. *Cell Stress*. 3:38-46.

Amon, J.D., and D. Koshland. 2016. RNase H enables efficient repair of R-loop induced DNA damage. *Elife*. 5.

Arnett, F.C., J.D. Reveille, and B.C. Valdez. 1997. Autoantibodies to a nucleolar RNA helicase protein in patients with connective tissue diseases. *Arthritis Rheum*. 40:1487-1492.

Auclair, Y., and S. Richard. 2013. The role of arginine methylation in the DNA damage response. *DNA Repair (Amst)*. 12:459-465.

Avasarala, S., M. Van Scoyk, M.K. Karuppusamy Rathinam, S. Zerayesus, X. Zhao, W. Zhang, M.R. Pergande, J.A. Borgia, J. DeGregori, J.D. Port, R.A. Winn, and R.K. Bikkavilli. 2015. PRMT1 Is a Novel Regulator of Epithelial-Mesenchymal-Transition in Non-small Cell Lung Cancer. *J Biol Chem*. 290:13479-13489.

Ballut, L., B. Marchadier, A. Baguet, C. Tomasetto, B. Seraphin, and H. Le Hir. 2005. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. *Nat Struct Mol Biol*. 12:861-869.

Bedford, M.T., and S.G. Clarke. 2009. Protein arginine methylation in mammals: who, what, and why. *Mol Cell*. 33:1-13.

Blanc, R.S., and S. Richard. 2017. Arginine Methylation: The Coming of Age. *Mol Cell*. 65:8-24.

Boisvert, F.M., U. Dery, J.Y. Masson, and S. Richard. 2005a. Arginine methylation of MRE11 by PRMT1 is required for DNA damage checkpoint control. *Genes Dev*. 19:671-676.

Boisvert, F.M., A. Rhie, S. Richard, and A.J. Doherty. 2005b. The GAR motif of 53BP1 is arginine methylated by PRMT1 and is necessary for 53BP1 DNA binding activity. *Cell Cycle*. 4:1834-1841.

Boleslavska, B., A. Oravetzova, K. Shukla, Z. Nascakova, O.N. Ibini, Z. Hasanova, M. Andrs, R. Kanagaraj, J. Dobrovolna, and P. Janscak. 2022. DDX17 helicase promotes resolution of R-loop-mediated transcription-replication conflicts in human cells. *Nucleic Acids Res*. 50:12274-12290.

Bonzheim, I., M. Irmler, M. Klier-Richter, J. Steinhilber, N. Anastasov, S. Schafer, P. Adam, J. Beckers, M. Raffeld, F. Fend, and L. Quintanilla-Martinez. 2013. Identification of C/EBPbeta target genes in ALK+ anaplastic large cell lymphoma (ALCL) by gene expression profiling and chromatin immunoprecipitation. *PLoS One*. 8:e64544.

Brahms, H., L. Meheus, V. de Brabandere, U. Fischer, and R. Luhrmann. 2001. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B' and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA*. 7:1531-1542.

Cai, T., J.K. Cinkornpumin, Z. Yu, O.D. Villarreal, W.A. Pastor, and S. Richard. 2021. Deletion of RBMX RGG/RG motif in Shashi-XLID syndrome leads to aberrant p53 activation and neuronal differentiation defects. *Cell Rep*. 36:109337.

Calo, E., R.A. Flynn, L. Martin, R.C. Spitale, H.Y. Chang, and J. Wysocka. 2015. RNA helicase DDX21 coordinates transcription and ribosomal RNA processing. *Nature*. 518:249-253.

Cargill, M., R. Venkataraman, and S. Lee. 2021. DEAD-Box RNA Helicases and Genome Stability. *Genes (Basel)*. 12.

Chakraborty, P., J.T.J. Huang, and K. Hiom. 2018. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat Commun*. 9:4346.

Champoux, J.J. 2001. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*. 70:369-413.

Chan-Penebre, E., K.G. Kuplast, C.R. Majer, P.A. Boriack-Sjodin, T.J. Wigle, L.D. Johnston, N. Rioux, M.J. Munchhof, L. Jin, S.L. Jacques, K.A. West, T. Lingaraj, K. Stickland, S.A. Ribich, A. Raimondi, M.P. Scott, N.J. Waters, R.M. Pollock, J.J. Smith, O. Barbash, M. Pappalardi, T.F. Ho, K. Nurse, K.P. Oza, K.T. Gallagher, R. Kruger, M.P. Moyer, R.A. Copeland, R. Chesworth, and K.W. Duncan. 2015. A selective inhibitor of PRMT5 with in vivo and in vitro potency in MCL models. *Nat Chem Biol*. 11:432-437.

Chedin, F., S.R. Hartono, L.A. Sanz, and V. Vanoosthuyse. 2021. Best practices for the visualization, mapping, and manipulation of R-loops. *EMBO J*. 40:e106394.

Cho, E.C., S. Zheng, S. Munro, G. Liu, S.M. Carr, J. Moehlenbrink, Y.C. Lu, L. Stimson, O. Khan, R. Konietzny, J. McGouran, A.S. Coutts, B. Kessler, D.J. Kerr, and N.B. Thangue. 2012. Arginine methylation controls growth regulation by E2F-1. *EMBO J*. 31:1785-1797.

Costantino, L., and D. Koshland. 2018. Genome-wide Map of R-Loop-Induced Damage Reveals How a Subset of R-Loops Contributes to Genomic Instability. *Mol Cell*. 71:487-497 e483.

Cristini, A., G. Ricci, S. Britton, S. Salimbeni, S.N. Huang, J. Marinello, P. Calsou, Y. Pommier, G. Favre, G. Capranico, N. Gromak, and O. Sordet. 2019. Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Rep*. 28:3167-3181 e3166.

Cui, B.C., V. Sikirzhytski, M. Aksenova, M.D. Lucius, G.H. Levon, Z.T. Mack, C. Pollack, D. Odhiambo, E. Broude, S.B. Lizarraga, M.D. Wyatt, and M. Shtutman. 2020. Pharmacological inhibition of DEAD-Box RNA Helicase 3 attenuates stress granule assembly. *Biochem Pharmacol*. 182:114280.

D'Alessandro, G., D.R. Whelan, S.M. Howard, V. Vitelli, X. Renaudin, M. Adamowicz, F. Iannelli, C.W. Jones-Weinert, M. Lee, V. Matti, W.T.C. Lee, M.J. Morten, A.R. Venkitaraman, P. Cejka, E. Rothenberg, and F. d'Adda di Fagagna. 2018. BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nat Commun*. 9:5376.

Dumelie, J.G., and S.R. Jaffrey. 2017. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife*. 6.

El Hage, A., S.L. French, A.L. Beyer, and D. Tollervey. 2010. Loss of Topoisomerase I leads to R-loop-mediated transcriptional blocks during ribosomal RNA synthesis. *Genes Dev*. 24:1546-1558.

El Hage, A., S. Webb, A. Kerr, and D. Tollervey. 2014. Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet*. 10:e1004716.

Eram, M.S., Y. Shen, M. Szewczyk, H. Wu, G. Senisterra, F. Li, K.V. Butler, H.U. Kaniskan, B.A. Speed, C. Dela Sena, A. Dong, H. Zeng, M. Schapira, P.J. Brown, C.H. Arrowsmith, D. Barsyte-Lovejoy, J. Liu, M. Vedadi, and J. Jin. 2016. A Potent, Selective, and Cell-Active Inhibitor of Human Type I Protein Arginine Methyltransferases. *ACS Chem Biol*. 11:772-781.

Gan, W., Z. Guan, J. Liu, T. Gui, K. Shen, J.L. Manley, and X. Li. 2011. R-loop-mediated genomic instability is caused by impairment of replication fork progression. *Genes Dev*. 25:2041-2056.

Garcia, M.C., J. Zhou, D. Henning, F.C. Arnett, and B.C. Valdez. 2000. Unique epitopes in RNA helicase II/Gu protein recognized by serum from a watermelon stomach patient. *Mol Immunol*. 37:351-359.

García-Muse, T., and A. Aguilera. 2019. R Loops: From Physiological to Pathological Roles. *Cell*. 179:604-618.

Garcia-Pichardo, D., J.C. Canas, M.L. Garcia-Rubio, B. Gomez-Gonzalez, A.G. Rondon, and A. Aguilera. 2017. Histone Mutants Separate R Loop Formation from Genome Instability Induction. *Mol Cell*. 66:597-609 e595.

Garcia-Rubio, M., P. Aguilera, J. Lafuente-Barquero, J.F. Ruiz, M.N. Simon, V. Geli, A.G. Rondon, and A. Aguilera. 2018. Yra1-bound RNA-DNA hybrids cause orientation-independent transcription-replication collisions and telomere instability. *Genes Dev*. 32:965-977.

Ginno, P.A., P.L. Lott, H.C. Christensen, I. Korf, and F. Chedin. 2012. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol Cell*. 45:814-825.

Gomez-Gonzalez, B., and A. Aguilera. 2007. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A*. 104:8409-8414.

Gonzalez-Aguilera, C., C. Tous, B. Gomez-Gonzalez, P. Huertas, R. Luna, and A. Aguilera. 2008. The THP1-SAC3-SUS1-CDC31 complex works in transcription elongation-mRNA export preventing RNA-mediated genome instability. *Mol Biol Cell*. 19:4310-4318.

Gu, Z., S. Gao, F. Zhang, Z. Wang, W. Ma, R.E. Davis, and Z. Wang. 2012. Protein arginine methyltransferase 5 is essential for growth of lung cancer cells. *Biochem J*. 446:235-241.

Guderian, G., C. Peter, J. Wiesner, A. Sickmann, K. Schulze-Osthoff, U. Fischer, and M. Grimmler. 2011. RioK1, a new interactor of protein arginine methyltransferase 5 (PRMT5), competes with pICln for binding and modulates PRMT5 complex composition and substrate specificity. *J Biol Chem*. 286:1976-1986.

Guillen-Chable, F., A. Bayona, L.C. Rodriguez-Zapata, and E. Castano. 2021. Phase Separation of Intrinsically Disordered Nucleolar Proteins Relate to Localization and Function. *Int J Mol Sci*. 22.

Halasz, L., Z. Karanyi, B. Boros-Olah, T. Kuik-Rozsa, E. Sipos, E. Nagy, L.A. Mosolygo, A. Mazlo, E. Rajnavolgyi, G. Halmos, and L. Szekvolgyi. 2017. RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res*. 27:1063-1073.

Hegazy, Y.A., C.M. Fernando, and E.J. Tran. 2020. The balancing act of R-loop biology: The good, the bad, and the ugly. *J Biol Chem*. 295:905-913.

Hirai, Y., E. Louvet, T. Oda, M. Kumeta, Y. Watanabe, T. Horigome, and K. Takeyasu. 2013. Nucleolar scaffold protein, WDR46, determines the granular compartmental localization of nucleolin and DDX21. *Genes Cells*. 18:780-797.

Huertas, P., and A. Aguilera. 2003. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell*. 12:711-721.

Iwamoto, F., M. Stadler, K. Chalupnikova, E. Oakeley, and Y. Nagamine. 2008. Transcription-dependent nucleolar cap localization and possible nuclear function of DExH RNA helicase RHAU. *Exp Cell Res*. 314:1378-1391.

Jing, P., N. Zhao, M. Ye, Y. Zhang, Z. Zhang, J. Sun, Z. Wang, J. Zhang, and Z. Gu. 2018. Protein arginine methyltransferase 5 promotes lung cancer metastasis via the epigenetic regulation of miR-99 family/FGFR3 signaling. *Cancer Lett*. 427:38-48.

Lee, C.S., A.P. Dias, M. Jedrychowski, A.H. Patel, J.L. Hsu, and R. Reed. 2008. Human DDX3 functions in translation and interacts with the translation initiation factor eIF3. *Nucleic Acids Res*. 36:4708-4718.

Lee, H., H. Cho, J. Kim, S. Lee, J. Yoo, D. Park, and G. Lee. 2022. RNase H is an exo- and endoribonuclease with asymmetric directionality, depending on the binding mode to the structural variants of RNA:DNA hybrids. *Nucleic Acids Res*. 50:1801-1814.

Li, L., D.R. Germain, H.Y. Poon, M.R. Hildebrandt, E.A. Monckton, D. McDonald, M.J. Hendzel, and R. Godbout. 2016. DEAD Box 1 Facilitates Removal of RNA and Homologous Recombination at DNA Double-Strand Breaks. *Mol Cell Biol*. 36:2794-2810.

Li, X., and J.L. Manley. 2005. Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*. 122:365-378.

Lin, W.L., J.K. Chen, X. Wen, W. He, G.A. Zarceno, Y. Chen, S. Chen, T.T. Paull, and H.W. Liu. 2022. DDX18 prevents R-loop-induced DNA damage and genome instability via PARP-1. *Cell Rep*. 40:111089.

Linder, P., and E. Jankowsky. 2011. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol*. 12:505-516.

Marchena-Cruz, E., L.P. Camino, J. Bhandari, S. Silva, J.J. Marqueta-Gracia, S.A. Amdeen, C. Guillen-Mendoza, M.L. Garcia-Rubio, J.M. Calderon-Montano, X. Xue, R. Luna, and A. Aguilera. 2023. DDX47, MeCP2, and other functionally heterogeneous factors protect cells from harmful R loops. *Cell Rep*. 42:112148.

Marechal, A., and L. Zou. 2013. DNA damage sensing by the ATM and ATR kinases. *Cold Spring Harb Perspect Biol*. 5.

Marko, M., A. Vlassis, A. Guialis, and M. Leichter. 2012. Domains involved in TAF15 subcellular localisation: dependence on cell type and ongoing transcription. *Gene*. 506:331-338.

Mersaoui, S.Y., Z. Yu, Y. Coulombe, M. Karam, F.F. Busatto, J.Y. Masson, and S. Richard. 2019. Arginine methylation of the DDX5 helicase RGG/RG motif by PRMT5 regulates resolution of RNA:DNA hybrids. *EMBO J*. 38:e100986.

Mirkin, E.V., and S.M. Mirkin. 2005. Mechanisms of transcription-replication collisions in bacteria. *Mol Cell Biol*. 25:888-895.

Murthy, A.C., W.S. Tang, N. Jovic, A.M. Janke, D.H. Seo, T.M. Perdikari, J. Mittal, and N.L. Fawzi. 2021. Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nat Struct Mol Biol*. 28:923-935.

Niehrs, C., and B. Luke. 2020. Regulatory R-loops as facilitators of gene expression and genome stability. *Nat Rev Mol Cell Biol*. 21:167-178.

Parish, J.L., J. Rosa, X. Wang, J.M. Lahti, S.J. Doxsey, and E.J. Androphy. 2006. The DNA helicase ChlR1 is required for sister chromatid cohesion in mammalian cells. *J Cell Sci*. 119:4857-4865.

Pellar, G.J., and P.J. DiMario. 2003. Deletion and site-specific mutagenesis of nucleolin's carboxy GAR domain. *Chromosoma*. 111:461-469.

Pesiridis, G.S., E. Diamond, and G.D. Van Duyne. 2009. Role of pICLn in methylation of Sm proteins by PRMT5. *J Biol Chem*. 284:21347-21359.

Pohjoismaki, J.L., J.B. Holmes, S.R. Wood, M.Y. Yang, T. Yasukawa, A. Reyes, L.J. Bailey, T.J. Cluett, S. Goffart, S. Willcox, R.E. Rigby, A.P. Jackson, J.N. Spelbrink, J.D. Griffith, R.J. Crouch, H.T. Jacobs, and I.J. Holt. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J Mol Biol*. 397:1144-1155.

Polo, S.E., A.N. Blackford, J.R. Chapman, L. Baskcomb, S. Gravel, A. Rusch, A. Thomas, R. Blundred, P. Smith, J. Kzhyshkowska, T. Dobner, A.M. Taylor, A.S. Turnell, G.S. Stewart, R.J. Grand, and S.P. Jackson. 2012. Regulation of DNA-end resection by hnRNPU-like proteins promotes DNA double-strand break signaling and repair. *Mol Cell*. 45:505-516.

Prado, F., and A. Aguilera. 2005. Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J*. 24:1267-1276.

Rabe, B. 2013. Aicardi-Goutieres syndrome: clues from the RNase H2 knock-out mouse. *J Mol Med (Berl)*. 91:1235-1240.

Reijns, M.A., B. Rabe, R.E. Rigby, P. Mill, K.R. Astell, L.A. Lettice, S. Boyle, A. Leitch, M. Keighren, F. Kilanowski, P.S. Devenney, D. Sexton, G. Grimes, I.J. Holt, R.E. Hill, M.S. Taylor, K.A. Lawson, J.R. Dorin, and A.P. Jackson. 2012. Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell*. 149:1008-1022.

Rinaldi, C., P. Pizzul, M.P. Longhese, and D. Bonetti. 2021. Sensing R-Loop-Associated DNA Damage to Safeguard Genome Stability. *Frontiers in Cell and Developmental Biology*. 8.

Roy, D., and M.R. Lieber. 2009. G clustering is important for the initiation of transcription-induced R-loops in vitro, whereas high G density without clustering is sufficient thereafter. *Mol Cell Biol*. 29:3124-3133.

Roy, D., Z. Zhang, Z. Lu, C.L. Hsieh, and M.R. Lieber. 2010. Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol*. 30:146-159.

Roy, R., G. Das, I.A. Kuttanda, N. Bhatter, and P.I. Rajyaguru. 2022. Low complexity RGG-motif sequence is required for Processing body (P-body) disassembly. *Nat Commun*. 13:2077.

Sachamitr, P., J.C. Ho, F.E. Ciamponi, W. Ba-Alawi, F.J. Coutinho, P. Guilhamon, M.M. Kushida, F.M.G. Cavalli, L. Lee, N. Rastegar, V. Vu, M. Sanchez-Osuna, J. Coulombe-Huntington, E. Kanshin, H. Whetstone, M. Durand, P. Thibault, K. Hart, M. Mangos, J. Veyhl, W. Chen, N. Tran, B.C. Duong, A.M. Aman, X. Che, X. Lan, O. Whitley, O. Zaslaver, D. Barsyte-Lovejoy, L.M. Richards, I. Restall, A. Caudy, H.L. Rost, Z.Q. Bonday, M. Bernstein, S. Das, M.D. Cusimano, J. Spears, G.D. Bader, T.J. Pugh, M. Tyers, M. Lupien, B. Haibe-Kains, H. Artee Luchman, S. Weiss, K.B. Massirer, P. Prinos, C.H. Arrowsmith, and P.B. Dirks. 2021. PRMT5 inhibition disrupts splicing and stemness in glioblastoma. *Nat Commun*. 12:979.

Sakasai, R., M. Isono, M. Wakasugi, M. Hashimoto, Y. Sunatani, T. Matsui, A. Shibata, T. Matsunaga, and K. Iwabuchi. 2017. Aquarius is required for proper CtIP expression and homologous recombination repair. *Sci Rep*. 7:13808.

Salas-Armenteros, I., C. Perez-Calero, A. Bayona-Feliu, E. Tumini, R. Luna, and A. Aguilera. 2017. Human THO-Sin3A interaction reveals new mechanisms to prevent R-loops that cause genome instability. *EMBO J*. 36:3532-3547.

Sanz, L.A., S.R. Hartono, Y.W. Lim, S. Steyaert, A. Rajpurkar, P.A. Ginno, X. Xu, and F. Chedin. 2016. Prevalent, Dynamic, and Conserved R-Loop Structures Associate with Specific Epigenomic Signatures in Mammals. *Mol Cell*. 63:167-178.

Shanbhag, N.M., and R.A. Greenberg. 2013. The dynamics of DNA damage repair and transcription. *Methods Mol Biol*. 1042:227-235.

Shinkura, R., M. Tian, M. Smith, K. Chua, Y. Fujiwara, and F.W. Alt. 2003. The influence of transcriptional orientation on endogenous switch region function. *Nat Immunol*. 4:435-441.

Singh, R.S., V. Vidhyasagar, S. Yang, A.B. Arna, M. Yadav, A. Aggarwal, A.N. Aguilera, S. Shinriki, K.K. Bhanumathy, K. Pandey, A. Xu, N. Rapin, M. Bosch, J. DeCoteau, J. Xiang, F.J. Vizeacoumar, Y. Zhou, V. Misra, H. Matsui, S.R. Ross, and Y. Wu. 2022. DDX41 is required for cGAS-STING activation against DNA virus infection. *Cell Rep*. 39:110856.

Skourti-Stathaki, K., N.J. Proudfoot, and N. Gromak. 2011. Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol Cell*. 42:794-805.

Song, C., A. Hotz-Wagenblatt, R. Voit, and I. Grummt. 2017. SIRT7 and the DEAD-box helicase DDX21 cooperate to resolve genomic R loops and safeguard genome stability. *Genes Dev*. 31:1370-1381.

Stolz, R., S. Sulthana, S.R. Hartono, M. Malig, C.J. Benham, and F. Chedin. 2019. Interplay between DNA sequence and negative superhelicity drives R-loop structures. *Proc Natl Acad Sci U S A*. 116:6260-6269.

Stopa, N., J.E. Krebs, and D. Shechter. 2015. The PRMT5 arginine methyltransferase: many roles in development, cancer and beyond. *Cell Mol Life Sci*. 72:2041-2059.

Tanaka, A., J.Y. Wang, J. Shia, Y. Zhou, M. Ogawa, R.C. Hendrickson, D.S. Klimstra, and M.H. Roehrl. 2020. DEAD-box RNA helicase protein DDX21 as a prognosis marker for early stage colorectal cancer with microsatellite instability. *Sci Rep*. 10:22085.

Tang, W., W. You, F. Shi, T. Qi, L. Wang, Z. Djouder, W. Liu, and X. Zeng. 2009. RNA helicase A acts as a bridging factor linking nuclear beta-actin with RNA polymerase II. *Biochem J*. 420:421-428.

Thandapani, P., T.R. O'Connor, T.L. Bailey, and S. Richard. 2013. Defining the RGG/RG motif. *Mol Cell*. 50:613-623.

Tresini, M., D.O. Warmerdam, P. Kolovos, L. Snijder, M.G. Vrouwe, J.A. Demmers, I.W.F. van, F.G. Grosveld, R.H. Medema, J.H. Hoeijmakers, L.H. Mullenders, W. Vermeulen, and J.A. Marteijn. 2015. The core spliceosome as target and effector of non-canonical ATM signalling. *Nature*. 523:53-58.

Tuduri, S., L. Crabbe, C. Conti, H. Tourriere, H. Holtgreve-Grez, A. Jauch, V. Pantesco, J. De Vos, A. Thomas, C. Theillet, Y. Pommier, J. Tazi, A. Coquelle, and P. Pasero. 2009. Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat Cell Biol*. 11:1315-1324.

Voronina, A.S., and E.S. Pshennikova. 2021. mRNPs: Structure and role in development. *Cell Biochem Funct*. 39:832-843.

Wahba, L., L. Costantino, F.J. Tan, A. Zimmer, and D. Koshland. 2016. S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev*. 30:1327-1338.

Walport, L.J., R.J. Hopkinson, R. Chowdhury, R. Schiller, W. Ge, A. Kawamura, and C.J. Schofield. 2016. Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat Commun*. 7:11974.

Wang, K., H. Wang, C. Li, Z. Yin, R. Xiao, Q. Li, Y. Xiang, W. Wang, J. Huang, L. Chen, P. Fang, and K. Liang. 2021. Genomic profiling of native R loops with a DNA-RNA hybrid recognition sensor. *Sci Adv*. 7.

Wang, M., and B. Lemos. 2019. Ribosomal DNA harbors an evolutionarily conserved clock of biological aging. *Genome Res*. 29:325-333.

Webby, C.J., A. Wolf, N. Gromak, M. Dreger, H. Kramer, B. Kessler, M.L. Nielsen, C. Schmitz, D.S. Butler, J.R. Yates, 3rd, C.M. Delahunty, P. Hahn, A. Lengeling, M. Mann, N.J. Proudfoot, C.J. Schofield, and A. Bottger. 2009. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science*. 325:90-93.

Westover, K.D., D.A. Bushnell, and R.D. Kornberg. 2004. Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*. 303:1014-1016.

Wu, Q., M. Schapira, C.H. Arrowsmith, and D. Barsyte-Lovejoy. 2021. Protein arginine methylation: from enigmatic functions to therapeutic targeting. *Nat Rev Drug Discov*. 20:509-530.

Xu, J., and S. Richard. 2021. Cellular pathways influenced by protein arginine methylation: Implications for cancer. *Mol Cell*. 81:4357-4368.

Yang, Y., and M.T. Bedford. 2013. Protein arginine methyltransferases and cancer. *Nat Rev Cancer*. 13:37-50.

Yang, Y., K.M. McBride, S. Hensley, Y. Lu, F. Chedin, and M.T. Bedford. 2014. Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Mol Cell*. 53:484-497.

Yasuhara, T., R. Kato, Y. Hagiwara, B. Shiotani, M. Yamauchi, S. Nakada, A. Shibata, and K. Miyagawa. 2018. Human Rad52 Promotes XPG-Mediated R-loop Processing to Initiate Transcription-Associated Homologous Recombination Repair. *Cell*. 175:558-570 e511.

Yasukawa, T., A. Reyes, T.J. Cluett, M.Y. Yang, M. Bowmaker, H.T. Jacobs, and I.J. Holt. 2006. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J*. 25:5358-5371.

Yu, K., F. Chedin, C.L. Hsieh, T.E. Wilson, and M.R. Lieber. 2003. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol*. 4:442-451.

Yu, Y., J.L. Wang, L.L. Meng, C.T. Hu, Z.W. Yan, Z.P. He, X.Q. Shi, G.H. Fu, and L.D. Zu. 2021. DDX54 Plays a Cancerous Role Through Activating P65 and AKT Signaling Pathway in Colorectal Cancer. *Front Oncol*. 11:650360.

Yu, Z., T. Chen, J. Hebert, E. Li, and S. Richard. 2009. A mouse PRMT1 null allele defines an essential role for arginine methylation in genome maintenance and cell proliferation. *Mol Cell Biol*. 29:2982-2996.

Yu, Z., S.Y. Mersaoui, L. Guitton-Sert, Y. Coulombe, J. Song, J.Y. Masson, and S. Richard. 2020. DDX5 resolves R-loops at DNA double-strand breaks to promote DNA repair and avoid chromosomal deletions. *NAR Cancer*. 2:zcaa028.

Yuan, W., Q. Al-Hadid, Z. Wang, L. Shen, H. Cho, X. Wu, and Y. Yang. 2021. TDRD3 promotes DHX9 chromatin recruitment and R-loop resolution. *Nucleic Acids Res*. 49:8573-8591.

Zhang, Y., K.C. Baysac, L.F. Yee, A.J. Saporita, and J.D. Weber. 2014. Elevated DDX21 regulates c-Jun activity and rRNA processing in human breast cancers. *Breast Cancer Res*. 16:449.

Zhang, Z., T. Kim, M. Bao, V. Facchinetti, S.Y. Jung, A.A. Ghaffari, J. Qin, G. Cheng, and Y.J. Liu. 2011. DDX1, DDX21, and DHX36 helicases form a complex with the adaptor molecule TRIF to sense dsRNA in dendritic cells. *Immunity*. 34:866-878.

Zhao, D.Y., G. Gish, U. Braunschweig, Y. Li, Z. Ni, F.W. Schmitges, G. Zhong, K. Liu, W. Li, J. Moffat, M. Vedadi, J. Min, T.J. Pawson, B.J. Blencowe, and J.F. Greenblatt. 2016. SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature*. 529:48-53.

Zhao, F., W. Kim, J.A. Kloeber, and Z. Lou. 2020. DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells. *Exp Mol Med*. 52:1705-1714.

Zhao, Q., G. Rank, Y.T. Tan, H. Li, R.L. Moritz, R.J. Simpson, L. Cerruti, D.J. Curtis, D.J. Patel, C.D. Allis, J.M. Cunningham, and S.M. Jane. 2009. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol*. 16:304-311.

Supplementary Data

Supplemental Figure 1. DDX21 R-K sequence from IDT gene synthesis.

AGC GCT cc ttg atc aac tca aat gtg ggt ttt gtg acc atg atc ttg cag tgc tca att gaa atg cca aat att agt tat gct tgg aaa gaa ctt aaa gag cag ctg ggc gag gag att gat tcc aaa gtg aag gga atg gtt ttt ctc aaa gga aag ctg ggt gtt tgc ttt gat gta cct acc gca tca gta aca gaa ata cag gag aaa tgg cat gat tca **aaa aag** tgg cag ctc tct gtg gcc aca gag caa cca gaa ctg gaa gga cca **aag** gaa gga tat gga ggc ttc aag gga cag **aag** gaa ggc agt aaa ggc ttc aag gga cag **aag** gac gga aac **aaa aaa** ttc aaa gga cag **aag** gaa ggc agt aaa ggc ccg aaa gga cag **aaa** tca gga ggt ggc aac aaa agt aac **aaa** tcc caa aac aaa ggc cag aag **aag** agt ttc agt aaa gca ttt ggt caa taa **CTC GAG**

Note:

All raw data used for this thesis can be found in Dropbox with the following link:

https://www.dropbox.com/sh/n6u2fg1a7fqfp3q/AAATveb4VjwEOJeaipzeRNQGa?dl=0

Immunofluorescence files can be found on the LDI Z DRIVE due to file size restrictions.

Path: LDI/UNITS/SRICHARD/Samantha Daley/IF photos