SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

Examination of R-loop Regulators Using High-Throughput Genomic Screening and Meta-analysis

by Beáta Kissné Boros-Oláh

Supervisor: Dr. Lóránt Székvölgyi

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF MOLECULAR CELLULAR AND IMMUNE BIOLOGY

DEBRECEN, 2024

Examination of R-loop Regulators Using High-Throughput Genomic Screening and Meta-analysis

By Beáta Kissné Boros-Oláh, MSc

Supervisor: Dr. Lóránt Székvölgyi, PhD, DSc

Doctoral School of Molecular Cellular and Immune Biology, University of Debrecen

Head of the **Defense Committee**:	Prof. Dr. Éva Csősz, PhD, DSc
Reviewers:	Prof. Dr. Margit Balázs, PhD, DSc
	Dr. Tibor Pankotai, PhD
Members of the Defense Committee:	Dr. Katalin Kormosné Goda, PhD
	Dr. Péter Burkovics, PhD

The PhD Defense takes place at the Lecture Hall of Building A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen at 1 p.m., 11th of December, 2024.

1. Introduction

1.1. Introduction to R-loop Biology

An R-loop is a three-stranded nucleic acid structure consisting of a DNA-RNA hybrid (formed by the complementary base pairing of an RNA molecule with DNA) and a displaced single-stranded DNA. R-loops occur relatively frequently in the genome (genome coverage: 8% in yeast, 10% in Arabidopsis thaliana, 5% in humans) and have significant physiological importance. They fundamentally carry out positive functions necessary for important biological processes; however, their excessive accumulation can lead to DNA damage and genome instability. The accumulation of R-loops has been linked to numerous human diseases, including playing roles in various neurodegenerative and cancerous diseases.

1.1.1. Discovery of R-loops

The R-loop structure was discovered over 45 years ago in an in vitro experiment. A hybrid DNA molecule consisting of yeast DNA and bacteriophage DNA formed an R-loop in the presence of 70% formamide with the yeast 26S rRNA, which was complementary to the yeast DNA. The presence of 70% formamide was necessary because the formation of the DNA-RNA hybrid structure is more likely near the melting point of the double-stranded DNA. It was determined that the R-loop structure is very stable and could only be disrupted by RNA strand degradation. Therefore, at that time, it was thought that R-loops were biologically irrelevant.

In vivo, R-loops were successfully detected 20 years later (Drolet et al., 1995). The formation of R-loops in vivo was demonstrated using modified Escherichia coli strains in which either the topA gene (encoding topoisomerase I) was deleted or the gyrB gene (encoding the DNA gyrase B subunit) had a temperature-sensitive mutation. The researchers found that the DNA gyrase B subunit facilitates the formation of R-loops by resolving positive supercoiling within the DNA, while topoisomerase I inhibits the formation of R-loops by resolving negative supercoiling. The excessive expression of RNase H, an endoribonuclease that degrades RNA in the DNA-RNA hybrid part of the R-loop, resolved the growth abnormalities in the topA mutant. Furthermore, growth defects caused by the deletion of RNase H were compensated by a reduction in the activity of the DNA gyrase B subunit. These observations indicated that the accumulation of R-loops caused the growth defects. Subsequently, R-loops became a focal point of many researchers' attention, turning into a popular area of study

1.1.2. Formation of R-loops

R-loops generally form near active genes, especially prone to arise in promoter regions. Although a DNA-RNA hybrid continuously forms within the transcription bubble of RNA polymerase, it is unlikely that the R-loop is simply an extension of this few-nucleotide hybrid; rather, it may arise through the re-hybridization of the RNA 5'-end. This is supported by the structure of the RNA polymerase complex, where the RNA and DNA strands exit through different channels, which prevents the transcription bubble's R-loop from elongating (Westover et al., 2004). This type of R-loop is referred to as a cis R-loop. The other type occurs in trans, meaning the RNA hybridizes to a complementary DNA segment distant from the site of transcription. For example, non-coding RNAs and guide RNAs associated with the CRISPR/Cas9 protein can form such R-loops.

1.1.2.1. Factors promoting R-loops

Several factors are known to increase the likelihood of R-loop formation. For example, the formation of guanine-rich transcripts from a cytosine-rich template DNA, due to the increased stability of the resulting DNA-RNA hybrid. The secondary structure of DNA on the displaced single strand, such as guanine quadruplexes or DNA-binding proteins, can also stabilize the R-loop structure. Negative supercoiling can also promote R-loop formation. In actively transcribing genes, the presence of a nick in the non-template DNA strand, if formed after the passage of RNA polymerase, can also increase the likelihood of R-loop formation. The presence of a nick reduces the possibility of re-hybridization of the DNA strands, thus increasing the likelihood that the template strand will hybridize with the newly synthesized RNA. Damage to proteins that inhibit R-loops (e.g., Ribonuclease H and RNA-DNA helicases) can lead to the accumulation of R-loops (Hegazy et al., 2020).

1.1.2.2. Factors preventing R-loops

In addition to the factors promoting R-loops, several factors are known to prevent or disrupt R-loop structures. For example, the involvement of RNA molecules with RNA-processing or exporting proteins prevents the formation of DNA-RNA hybrids (Li & Manley, 2005). Topoisomerase I also inhibits R-loop formation by resolving negative supercoiling behind RNA polymerase II. Less accessible heterochromatin also does not favor the formation of R-loops. Heterochromatic marks, such as histone H3 lysine 9 di- and trimethylation (H3K9me2/3), are associated with low occurrences of DNA-RNA hybrids in C. elegans (Zeller et al., 2016). DNA-RNA hybrid helicases, such as senataxin, can dissolve R-loops by separating the two strands,

while Ribonuclease H can eliminate R-loops by digesting the RNA strand of the DNA-RNA hybrid

1.2. Physiological roles of R-loops

R-loops participate in numerous physiological processes, such as mitochondrial DNA replication (B. Xu & Clayton, 1996; Pohjoismäki et al., 2010), class-switch recombination of immunoglobulin genes (Yu et al., 2003), or the functioning of the CRISPR-Cas9 system in bacteria (Jinek et al., 2012). R-loops play a role in centromere function, essential for accurate chromosome segregation, which is crucial for maintaining stable genome integrity. Centromere loci also contain R-loops (CEN R-loops), which are evolutionarily conserved and regulated in a cell cycle-dependent manner. In human cells, the displaced single-stranded DNA in CEN R-loops is bound by RPA, which recruits the DNA damage-responsive ATR kinase, leading to the activation of Aurora B, facilitating proper microtubule-kinetochore attachment (Niehrs & Luke, 2020). R-loops also form at telomeres, where the non-coding RNA, TERRA, hybridizes with the telomere DNA. The formation of R-loops in cells lacking telomerase enzyme leads to recombinational events, resulting in telomere elongation and thereby delaying cellular aging (Santos-Pereira & Aguilera, 2015).

1.2.1. The role of R-loops in gene expressison and chromatin structure regulation

R-loops can play a role in gene expression, as evidenced by their enrichment at the promoters and termination regions of genes (Skourti-Stathaki & Proudfoot, 2014). R-loops located at promoters can inhibit the binding of transcription factors, leading to transcriptional repression. Conversely, they can facilitate the binding of other transcription factors and inhibit the binding of repressors, thus also participating in transcriptional activation. The enrichment of R-loops at guanine-rich terminator elements facilitates the pausing of RNA polymerase II before the efficient completion of transcription (García-Muse & Aguilera, 2019). DNA-RNA hybrids are also associated with the unmethylated state of CpG islands, which are characteristic of many gene promoters in mammalian cells. R-loops formed at CpG islands protect these regions against DNA methylation, avoiding transcriptional inhibition. This protection can be explained by the preference of DNA methyltransferase 1 for binding and methylating double-stranded DNA over a DNA-RNA hybrid (Grunseich et al., 2018) and by R-loops attracting the DNA demethylating TET enzyme (Arab et al., 2019). A good example of R-loop regulation mediated through long non-coding RNAs and DNA-binding proteins is observed in the model plant Arabidopsis thaliana. The antisense non-coding RNA COOLAIR negatively regulates the

transcription of the FLOWERING LOCUS C (FLC) gene. However, the R-loop formed at the promoter region of COOLAIR is stabilized by the NODULIN HOMEOBOX (AtNDX) single-stranded DNA-binding protein, which inhibits the transcription of COOLAIR and thereby activates the expression of the FLC gene through this inhibition (Sun et al., 2013; Csorba et al., 2014). Furthermore, R-loops can induce chromatin decondensation, but they can also facilitate the establishment of a heterochromatic state, leading to chromatin compaction (Skourti-Stathaki et al., 2014; Castellano-Pozo et al., 2013).

1.2.2. The Role of R-loops in DNA repair

DNA-RNA hybrids can promote DNA repair in actively transcribed genomic regions. When DNA single-strand breaks (nicks) or double-strand breaks (DSBs) form in the vicinity of active transcription, several repair pathways converge to ensure that transcription in the immediate vicinity of the break is inhibited. These include removing Pol II, pausing transcription; which in turn could lead to the stabilization of the Rhurok-. Local chromatin structure changes to a repressive state around damage sites to ensure transcriptional inhibition around break sites. Interestingly, both PRC1 and PRC2, which localize in an R-loop-dependent manner to some promoters, are also characteristic of DNA breaks and induce transcriptional silencing through ubiquitination at Lys119 histone H2A (H2AK119) (Niehrs & Luke, 2020). The proteins involved may differ depending on the nature of the repair, but the sequence of events appears to be consistent: at the DSB, following 5′-end resection, a DNA-RNA hybrid is formed at the exposed 3′-end, which through proteins helps repair like RAD52, XPG1, BRCA1 and BRCA2. The RNA is then removed by senataxin, RNase H1, RNase H2, or DDX1, allowing RAD51 to bind to DNA and undergo homologous recombination. The presence of DNA-RNA hybrids at DSBs exemplifies the importance of a transient hybrid that has an important function but is later removed to avoid potential problems (Niehrs & Luke, 2020).

1.3. R-loops can lead to genome instability

R-loops can cause DNA damage through several mechanisms. R-loops result in exposed ssDNA stretches that are more chemically unstable than dsDNA and more prone to transcriptional mutagenesis (TAM), recombination (TAR) and DSBs. Members of mammalian cell-expressed activation-induced cytidine deaminase (AID) and Apolipoprotein B mRNA-editing catalytic polypeptide (APOBEC) can deaminate ssDNA cytosines to uracil. Subsequently, uracil DNA glycosylase cleaves the uracil base, creating an abasic site that can lead to base substitutions or ssDNA nicks, resulting in DSB formation at the replication fork

(Costantino & Koshland, 2015). Julie Solier and colleagues demonstrated in human cells that R-loops induced by the absence of various RNA processing factors, including the RNA/DNA helicases Aquarius (AQR) and Senataxin (SETX) or inhibition of topoisomerase I, cause DNA double-strand breaks through cleavage by the nucleotide excision repair (NER) endonucleases XPF and XPG. Furthermore, the transcription-coupled nucleotide excision repair (TC-NER) factor, Cockayne syndrome group B (CSB), is required for DSB formation, but not the global genome repair protein, XPC. Shutdown of RNA polymerase allows CSB to recruit XPF and XPG endonucleases. XPF and XPG create a single-stranded DNA gap that can be converted to a DSB by DNA replication, and/or XPF and XPG cleave the R-loop on both strands, creating a DSB. These results indicate that TC-NER factors are potentially involved in R-loop-induced DNA damage and genome instability (Sollier et al., 2014). DSBs can also arise as a result of physical collisions between the R-loop and the replication machinery. This is also proven by the fact that R-loop-induced DSB formation can be prevented by inhibiting DNA replication. R-loops can be a huge physical barrier to the replication fork due to their relatively stable secondary DNA structure and the presence of the transcription apparatus and other RNA processing enzymes. R-loops can directly cause fork collapse and DSB formation, even without the intervention of nucleases such as XPF and XPG (Sollier & Cimprich, 2015).

1.4. R-loops in human disease

R-loops are implicated in many human diseases, including tumors, neurological disorders, and autoimmune diseases. Genomic instability mediated by R-loops may contribute to mutations and chromosomal rearrangements characteristic of cancer cells (Negrini et al., 2010). Common fragile sites (CFS - common fragile sites) are regions of the human genome that contain long genes and are slowly replicated in the S-phase, hot spots for DNA breaks (Glover, 2006). R-loops are formed in CFSs, and the resulting collisions between the DNA replication fork and the transcription machinery contribute to the instability of CFS loci. Removal of R-loops by RNaseH overexpression reduces genomic instability associated with CFS (Helmrich et al., 2011). Depletion of the tumor suppressor BRCA2 causes R-loop accumulation and activation of the DNA repair checkpoint. BRCA2 may be involved in the processing/removal of RNA-DNA hybrids, otherwise a source of replication stress and genomic instability (Bhatia et al., 2014). BRCA2 is frequently mutated and inactivated in tumors, especially breast cancer (Roy et al., 2012).

1.5. Methods for studying R-loops

Several techniques are now available to identify and characterize the R-loop structure. Until recently, most studies used the S9.6 antibody to detect and isolate R-loops using a method called DNA:RNA immunoprecipitation (DRIP), immunofluorescence labeling, or dot blot techniques.

The S9.6 monoclonal antibody was produced in mice in 1986 by immunization with a synthetic DNA-RNA hybrid antigen derived from bacteriophage ΦX174. The antibody was characterized by high DNA-RNA hybrid specificity and affinity (Boguslawski et al., 1986). The affinity of S9.6 scFv is not significantly affected by different buffer conditions or ionic strength below 500 mM NaCl. The smallest epitope it binds strongly is a fragment containing a DNA-RNA hybrid of six base pairs. S9.6 shows a dissociation constant of about 0.6 nM for the DNA–RNA hybrid and 2.7 nM for the AU-rich double-stranded RNA (Philips et al. 2013). To prevent the relatively strong double-stranded RNA binding, it is recommended to perform RNase digestion on the nucleic acid before immunoprecipitation or labeling, in a condition in which the R-loops are not damaged, but the free RNAs can be digested. We can use Rnase T1, Rnase III or RNase A at high salt (300 mM NaCl) (Chédin et al. 2021, Halász et al., 2017). Since the structures of dsRNA and R-loops are significantly different, it is unlikely that the antibody recognizes only the phosphate-sugar backbone of the R-loop, which would be required for sequence-independent binding. Fabian König and colleagues showed that the affinity of the antibody is sequence specific (König et al., 2017). For this reason, it is worthwhile to examine R-loops independently of S9.6. The most widely used method for R-loop mapping is DRIPseq (Ginno et al., 2012), which uses next-generation sequencing to map R-loops isolated by S9.6 immunoprecipitation. In DRIP-seq, nucleic acid is isolated from cells and then fragmented with restriction enzymes or sonication. Fragmentation of the genome with restriction enzymes can cause distortion and limit resolution, especially at the 5' end of genes (Halász et al., 2017). For this reason, ultrasound treatment of DNA is recommended for fragmentation, instead of using restriction enzymes (El Hage et al., 2014; Halász et al., 2017). After immunoprecipitation, sequencing libraries are generated using standard dsDNA-based library preparation. Sequencing validation can be performed on some selected genomic regions by QPCR on the IP sample, which allows us to achieve more accurate quantification. DRIP-seq has proven to be a consistent, reproducible and popular method for sequencing R-loops. There have been several further developments of the method, with which, for example, the mapping has been made thread-specific. For example, this can be addressed by strand-specific DNA library preparation as in ssDRIP-seq (Xu et al., 2017). Furthermore, by sequencing the R-loop RNA strand after

cDNA synthesis (DRIPc-Seq) (Sanz et al., 2016). And the method called S1-DRIP-Seq includes the use of S1-nuclease to degrade the displaced ssDNA of the R-loop, so sequencing the immunoprecipitated DNA also enables the identification of the strand orientation of the R-loop (Wahba et al., 2016) . Finally, bis-DRIP-seq combines in situ ssDNA bisulfite treatment with S9.6 immunoprecipitation, thereby improving the specificity of hybrids and ssDNA identification. During bisulfite treatment, unmethylated cytosines on single-stranded DNA are deaminated and converted to uracil (and during PCR amplification, uracil is converted to thymine), this sequence modification enables strand specificity (Dumelie & Jaffrey, 2017).

Another approach is the transfection of catalytically inactive, fragmented RNase H1 into cells and immunoprecipitation using this modified RNase H1, which can recognize DNA-RNA hybrids but cannot degrade the hybrid RNA strand. This method was named R-ChIP (Chen et al., 2017). R-ChIP identifies fewer R-loop regions than S9.6-based DRIP-seq or DRIPc-seq. With R-ChIP, it is mostly possible to identify R-loops at promoters, but not at transcriptional terminator regions. R-ChIP identifies smaller R-loop forming regions than DRIPc-seq. Overexpression of the catalytically inactive RNase H1 enzyme risks disrupting the dynamics of R-loops in vivo. For the implementation of R-ChIP on mammalian cells, it is therefore important that the catalytically inactive RNase H1 is expressed at an appropriate level: in case of too high expression, there is a risk of dominant-negative effects; in the case of too low expression, even though there is a risk that the endogenous, catalytically active enzyme may interfere with the binding of the catalytically inactive mutant and the efficiency of R-ChIP. There are probably different types of R-loops recognized by different types of proteins. This may be the reason why R-ChIP did not identify terminator regions as R-loop hotspots in human cells (Chen et al., 2017), where R-loops are recognized and cleaved by DNA-RNA helicases such as Senataxin, while RNase H1 does not bind (Skourti-Stathaki et al. 2011). This may explain why R-ChIP identified fewer R-loop fragments than DRIP-seq or DRIPc-seq. It is conceivable that DRIP is more suitable for detecting long-lived R-loops, while R-ChIP may be better for detecting highly dynamic R-loops degraded by RNase H1 (Vanoosthuyse, 2018). One of the latest S9.6 antibody-independent methods, the further development of the bisulfite treatment technique, SMRF-seq. After gentle DNA isolation and restriction enzyme digestion, the bisulfite treatment is performed in a non-denatured environment. PacBio sequencing, which sequences single molecules with long reads, enables the identification of R-loop “footprints” with near-nucleotide resolution, in a strand-specific manner on long single DNA molecules, with ultra-deep coverage (Malig et al., 2020).

1.6. Identifying R-loop regulatory genes

In recent years, several studies have been published in which high-throughput screenings were performed in order to identify R-loop regulatory and binding proteins.

One of the most significant results of the identification of human R-loop binding proteins so far comes from the research of Wang et al. Using mass spectrometry following pull-down assays with two types of synthetic DNA-RNA hybrids, 803 human proteins that bind to DNA-RNA hybrids were identified, and it was confirmed that the proteins also bind to R-loops in vivo. The 803 protein is characterized by numerous cellular functions, including most steps in RNA processing. Most of the proteins have K-homologous (KH) and helicase domains. (The KH domain recognizes and binds RNA) Of the identified proteins, more than 300 proteins preferred to bind to the DNA-RNA hybrid over double-stranded DNA (Wang et al., 2018). Of course, binding to DNA-RNA hybrids does not necessarily mean that the protein also regulates the R-loops. However, proteins can influence R-loop stability through indirect effects on transcription or DNA topology without binding R-loops (Fournier et al. 2021). In a study published in 2021, Wu et al. performed mass spectrometric analysis of S9.6 immunoprecipitation on mouse embryonic stem cells, identifying 364 R-loop-interacting proteins. Nucleolar proteins, including several DEAD-box family helicase proteins, were highly enriched among the hits. A closer examination of DEAD-box helicases revealed the role of the proteins in the post-transcriptional maturation of rRNA and their direct or indirect role in the regulation of genes related to differentiation. These findings revealed a vast network of R-loop-associated proteins that are crucial for stem cell homeostasis (Wu et al., 2021).

2. Objectives

The uncontrolled accumulation of R-loops poses a threat to genome integrity, thus it is crucial to identify and better understand the proteins involved in the recognition and regulation of R-loop structures. Our first objective was to examine the relationships between known R-loop regulatory genes and cancers. Using genomic databases, we analyzed key DNA-RNA hybrid binding protein gene expression data, comparing gene expression data from cancer patient groups with their survival data, thereby examining the role of R-loops in disease progression. Another goal was to explore the relationship between drug sensitivity in human tumor cell lines and the expression of R-loop regulators within those lines. The second objective was to develop a high-throughput screen for R-loop regulators using a barcoded yeast gene deletion library (DRIP-BC-seq). This screen allows parallel measurement of R-loop levels across numerous mutants on the same genomic test segment. Our goal was to validate findings from the DRIP-BC screen across the whole genome with DRIP-seq, examining the physiology of the mutants with spot assays, and performing further functional gene expression studies with mRNA-seq. These gene expression studies will help us understand how gene deletions have altered the positioning and quantity of R-loops. The third objective was to develop a new spatial genome structure investigation screen (4C-BC-seq) that allows us to simultaneously examine all 3D interactions of a given genomic test segment in a large number of yeast mutants. An additional goal was to identify mutants that affect both R-loops and the spatial structure of the genome using the findings common to DRIP-BC-seq and 4C-BC-seq. To validate the 4C-BC-seq at the whole genome level, the fourth goal was to implement the Hi-C chromatin three-dimensional structure investigation method in our lab. We studied wild-type and Nodulin homeobox (NDX) mutant Arabidopsis thaliana model plants. This also aimed to clarify the role of R-loops in chromatin structure changes, since NDX was the first protein shown to bind to the single-stranded DNA section of an R-loop, presumably stabilizing this structure (Sun Q et al Science 2013).

3. Results

3.1. Pharmacogenomic Investigation of R-loop Regulatory Genes - Focus on Cancer Therapy

During our analysis, we examined whether key R-loop binding proteins are correlated with cancer survival and drug sensitivity. We conducted a systematic pharmacogenomic analysis to identify these correlations. To investigate the connection between R-loop genes and cancer, we first determined how many of the DNA-RNA hybrid binding proteins identified in the study by Wang et al. were present in the cancer gene registry, which contains 2372 genes whose somatic mutations can lead to cancer development. Of the 448 R-loop genes examined, 92 appeared in the cancer-causing gene list, which is a statistically significant enrichment.

3.1.1. Analysis of Gene Expression and Survival Association in Cancer Patients

We downloaded mRNA expression of R-loop genes from The Cancer Genome Atlas (TCGA), compared the gene expression values of healthy tissues and primary tumors. Expression of several genes was found increased in multiple primary tumors compared to normal tissues, while others showed decreased expression. Subsequently, we examined whether the gene expression level of R-loop genes correlates with cancer patient survival. Using TCGA's gene expression (RNA-seq) and survival data, our analysis produced numerous Kaplan-Meier survival curves, identifying several cases that showed significant survival correlation with R-loop gene expression levels..

3.1.2. Analysis of Cell Line Gene Expression and Drug Sensitivity Association

In order to identify those R-loop genes that could be potential drug targets in cancer therapy, we sought synergistic interactions between the mRNA expression status of R-loop genes derived from cancer cell lines and their sensitivity to chemotherapeutic agents. We compared the sensitivity of 1065 cancer cell lines to 297 drugs with the mRNA expression data of the R-loop genes examined in these cell lines. The drugs examined are approved by the US Food and Drug Administration (FDA) and cover a broad spectrum of biological pathways, including protein kinase signaling, cytoskeleton function, DNA replication, DNA repair, and cell cycle regulation. We identified a total of 22,414 drug sensitivity and gene expression associations (22,414 - $p < 0.05$; 508 - FDR < 0.05). The total number of significant interactions showed substantial variations per drug (70–230 interactions per drug) and were also grouped according to the molecular pathways targeted by the drugs. The most broadly represented pathways

included the Ser/Thr protein kinase pathway and the PI3K/RTK/MAPK signaling, while the most common drug targets were MEK1/2, BRAF, PARP1/2, HSP90, AKT1/3, GSKs, PI3Ks, IGF1R, ROCK1/3, and EGFR. We found significant drug sensitivity associations for 80% of the R-loop genes examined (29 $p < 0.05$; except for BRCA2, BUGZ, DDX19, RNASEH1, RTEL1, THOC3, and THOC4; see Table S1_4), with CARM1, EWSR1, DHX9, and THOC1 showing the most drug interactions. Significant variations were observed in the number and direction of drug interactions across different cancer cell lines. For example, small cell lung carcinoma and ovarian cancer cell lines showed negative associations for most compounds and genes examined, while B-cell leukemia, Hodgkin's lymphoma, head and neck cancer, and Ewing's sarcoma cells predominantly showed positive associations. These differences suggest that the effects of RNA-DNA hybrid binding proteins can vary in outcome in response to different chemotherapeutic treatments depending on the genomic context/type of cancer.

3.1.3. Survival and Drug Interaction Associations

Since cancer cell lines derived from natural tumors accurately reflect the genomic context and tissue type of primary cancers, we narrowed down the identified drug interactions to those R-loop genes that showed significant survival associations in the corresponding primary cancers. Thus, we identified 1630 significant survival and drug interaction associations ($p < 0.05$) related to the expression of 29 R-loop genes (seven R-loop genes are not included in the results because there were no survival or drug interaction associations, or the examined cell lines did not include the corresponding primary tumor cell line). We observed the following trends:

i) The most survival interactions were observed with the following 10 genes, which also have significant drug interactions: BLM, RNASEH2A, ATXN1, BRCA1, BUB3, CARM1, GADD45A, FANCD2, THOC1, THOC2.

ii) High expression of THOC2 in sarcoma cell lines was associated with high IC50 values for CX-5461 (G4 ligand / RNA polymerase I inhibitor), suggesting that CX-5461 is less effective in treating sarcomas with high THOC2 expression. Conversely, low IC50 values for CX-5461 were associated with high expression of ATXN1 and TREX1, as well as longer survival times in patients with melanoma and endometrial cancer.

iii) In esophageal carcinoma cell lines, BRCA1 expression was associated with high IC50 values (resistance) to 5-fluorouracil (5-FU), while low BRCA1 levels resulted in better survival in esophageal carcinoma. This suggests that 5-FU treatment might be effective in treating BRCA1(-) esophageal carcinomas. Similarly, BRCA1 expression in glioma cell lines was

associated with decreased efficacy of (52)-7-oxozeaenol, GDC0941, refametinib, and selumetinib, while low BRCA1 levels had a beneficial effect on the survival of low-grade gliomas. Additionally, mesothelioma cell lines are resistant to doxorubicin, OSU-03012, or thapsigargin when expressing BRCA1. These drugs could potentially be effective chemotherapeutic agents for treating BRCA(-) gliomas and mesotheliomas. It is important to note that 5-FU is being tested in a clinical trial for esophageal carcinoma (NCT00052910), selumetinib for low-grade glioma (NCT01089101), and doxorubicin for patients with mesothelioma (NCT05063420506342).

iv) The most drug/gene associations were observed with RDEA119, selumetinib, and olaparib. Most cell lines showed resistance to RDEA119 and selumetinib, but sensitivity to olaparib at high R-loop gene expression levels (Supplementary Figure S3). For example, selumetinib treatment was ineffective in gastric adenocarcinoma cell lines overexpressing AQR (NCT02448290 - Phase 2), sarcoma cell lines overexpressing ATXN1 or BLM (NCT03155620 - Phase 2, NCT01752569 - Phase 2), glioma cell lines overexpressing BRCA1 or RNASEH2A, kidney cancer cell lines overexpressing FANCM or RNASEH2A, and endometrial carcinoma cell lines overexpressing TREX1. Regarding patient survival, low expression of AQR, ATXN1, BLM, BRCA1, and RNASEH2A was associated with longer survival in primary cancer cases, except for FANCM and TREX1, where higher mRNA levels were beneficial for patients with kidney cell carcinoma and endometrial carcinoma. Olaparib treatment was effective in prostate adenocarcinoma cells showing high DDX19A expression and in breast cancer cells overexpressing GADD45A, but mesothelioma cell lines with high ZNF207 (BUGZ) expression were resistant to treatment. Olaparib is undergoing clinical trials in mesothelioma (NCT03531840 – Phase 2), prostate adenocarcinoma (NCT01682772 – Phase 2), and breast cancer (NCT02000622 – Phase 3).

v) Several drugs under clinical trial for breast cancer treatment showed low IC50 values in breast cancer cell lines with high BRCA1 gene expression. For example, rucaparib (NCT02505048 – Phase 2, BRCA1,2 mutant patients), dabrafenib (NCT02465060 – Phase 2, BRAF V600E/R/K/D mutant patients), vismodegib (NCT02465060 – Phase 2 - SMO or PTCH1 mutant, NCT02694224 - Phase 2), BMS-754807 (NCT01225172 - Phase 2), and ruxolitinib (NCT01562873 - Phase 2, NCT02928978 - Phase 2). vi) Esophageal carcinoma cell lines showing high BRCA1 gene expression had low IC50 values for motesanib, cisplatin, and trametinib treatments, which are being tested in various clinical trials (NCT00101907 - Phase 1, NCT00655876 - Phase 3, NCT20246, NCT2020246). vii) In glioma cell lines, low CARM1

expression was associated with better efficacy of SN38 treatment (Type I topoisomerase inhibitor, the active metabolite of irinotecan), while patients with gliomas having low CARM1 levels showed longer survival. It is possible that CARM1(-) cancers are more sensitive to SN38 drug treatment, which leads to inhibition of Top3B and Top1, potentially increasing R-loop levels [32] and inducing cell death.

3.2. Identification of R-loop Regulators Using a Yeast Gene Deletion Library

Parallel to investigating the role of R-loops in cancer therapy, we have initiated a series of "screening" experiments using a barcoded yeast gene deletion library to identify novel R-loop regulatory genes.

3.2.1. Implementing the DRIP-BC-seq method

For the development of the DRIP-BC-seq method, we performed DNA-RNA hybrid immunoprecipitation (DRIP) on the Epi-ID yeast gene deletion library, then PCR-amplified the barcodes (BC) identifying different gene deletions from the immunoprecipitated samples, and finally quantified the amount of each barcode using next-generation sequencing. This allowed us to measure the differences in R-loop levels in the environment around the barcode among various gene deletion yeast strains. For the DRIP experiment, we only mildly fragmented the isolated genomic DNA with ultrasonic treatment in order to gather information from a broader environment around the barcode. It can be seen in the gel image that the DNA used for DRIP was smaller than 5 kb in fragment size. The barcode PCR following immunoprecipitation was conducted with 50 ng of DNA. We previously tested the barcode PCR on 12.5 ng, 25 ng, and 50 ng of DNA from the Epi-ID yeast library and, by checking the PCR products on an agarose gel, we observed that the amount of PCR product linearly increased between 12.5 - 50 ng, thus performing the PCR with this initial range of DNA amounts using 17 PCR cycles, we obtained quantitative results. We also performed immunoprecipitation on RNaseH pre-treated samples, but the barcode PCR from the RNaseH pre-treated DNA immunoprecipitated samples did not yield any product. Presumably, the immunoprecipitate of the RNaseH pre-treated sample contained only non-specifically bound DNAs. From this, it can be inferred that the immunoprecipitation (IP) performed on untreated samples resulted in specific DNA segments in the IP eluate and the barcode region of at least some of the yeast strains contained R-loops. We also verified the DRIP samples prepared for DRIP-BC-seq by QPCR on a genomic region positive and negative for R-loops. In the RNaseH treated sample, the DRIP signal (IP/input ratio) in the positive genomic region (Pos1 primer) decreased by more than half, while in the

negative genomic region (CEN16 primer), we detected very low QPCR signals in both RNaseH treated and untreated samples.

3.2.2. Significant hits of the DRIP-BC-seq screen

Having conducted the DRIP-BC-seq experiment across all 4200 mutant yeast strains, we found 110 strains in which the DRIP signal in the ~10 kb environment around the barcode significantly increased (IP/input > 1.5), and 89 strains where the DRIP signal significantly decreased (IP/input < 0.67). We performed functional annotation on the two groups of genes using the DAVID database and found that among the deleted genes which absence resulted in an increase in R-loops, there was a significant enrichment of genes related to mitochondrial function and metabolic processes. Among the deleted genes whose absence resulted in a decrease in R-loops based on DRIP-BC-seq, we did not find significant consistency in gene functions. However, it is worth mentioning that some genes related to ubiquitination (7 genes) and autophagy (4 genes) were found among the results with decreased DRIP-BC-seq signals. From the strains with significantly higher DRIP signals, we selected 11 strains with mitochondrial functions, on which we performed whole genome DRIP-seq and DRIP-QPCR measurements. We examined the growth of the selected strains using a growth assay called spot assay on YPD and YPG agar plates. We observed that all mutants were able to grow on YPD rich medium, however, the rpo41Δ, met7Δ, mgm101Δ, mss116Δ strains grew significantly slower compared to the wild type. On YPG agar, where glycerol replaces glucose as the carbon source, these strains showed no growth at all. The gas1Δ and pet20Δ strains grew much slower than the wild-type yeast strain.

3.2.3. Validating the DRIP-BC-seq hits by whole-genome DRIP-seq and QPCR

All DRIP experiments were performed on RNaseH pre-treated DNA as a negative control, and then we examined the samples by QPCR in both R-loop positive and negative genomic regions (El Hage & Tollervey, 2017). In the RNaseH control sample, we detected very low signals in the positive region (18S rDNA), and in the R-loop negative genomic region (CEN16), very low signals were measured in both normal and RNaseH pre-treated samples alike. This indicates that the DRIPs show R-loop-specific signals. The 18S rDNA region is not a suitable choice for comparing DRIP signal strength across different mutants because this DNA section can be present in 100-200 copies in the genome, but it is a very strong positive control region for R-loops, excellent for verifying the method's efficacy. When examining the 10 kb environment around the barcode (summarized as DRIP-seq score), aside from the exo5Δ, mgm1Δ, dss1Δ

strains, the DRIP signal increased compared to the wild type in the strains selected for validation, thus the DRIP-BC-seq validation was successful. However, strains showing an increase in R-loop levels around the barcode, when considering the total number of peaks in the whole-genome DRIP-seq, showed different results than expected in some cases. In ung1Δ, rpo41Δ, oye2Δ, met7Δ, gas1Δ mutants, the total peak number significantly decreased, meaning that R-loops largely disappeared in these strains compared to the wild type. In exo5Δ, mgm1Δ, pet20Δ, dss1Δ, mss116Δ strains, consistent with the DRIP-BC-seq results, the number of R-loops globally increased compared to the wild type. In the mgm101Δ mutant, a slight increase in peak number was observed. Thus, examining the total peak number, the DRIP-seq results of 6 strains matched the results examined among 11 strains, this discrepancy can be attributed to the screening nature of the method. Therefore, validation is very important, however, strains showing a decrease in R-loops identified by whole-genome DRIP-seq are as noteworthy and worthy of further investigation as the strains characterized by an increase in R-loops. We also wanted to validate the whole-genome DRIP-seq results (and indirectly the DRIP-BC-seq results), so we performed QPCR on the DRIP samples with two primers designed near the barcode. One primer pair was designed approximately 2 kb from the barcode, while the other was about 5 kb away. The average fragment size in DRIP-BC-seq was about 3 kb, while in whole-genome DRIP-seq it was only about 1 kb, hence it is advisable to examine whole-genome DRIP-seq at regions about 3 kb apart in the 5 kb environment around the barcode. With DRIP-QPCR, we were able to validate changes covering a broad DNA region or significant changes observed in the ~6 kb environment around the barcode in the DRIP-seq. Such is the significant increase in R-loop level in the exo5Δ and pet20Δ strains, and the significant decrease in RNA-DNA hybrid in the ung1Δ, met7Δ strains compared to the wild type. The most notably contradictory result was seen in the DRIP-QPCR results of the rpo41Δ strain in the two genomic regions examined. While the DRIP signal increased in the PTP1 region, it decreased in the SSB1 region. Based on the DRIP-seq signal strength displayed in the genome browser, a decrease is visible in both regions, but a generally low-level signal is also observable across the entire 14 kb length of the field of view, making it difficult to discern the difference with the naked eye. Examining the 10 kb environment around the barcode, based on the quantification of the total DRIP signal amount, there was an increase in R-loop levels in the environment around the barcode. However, globally examining the rpo41Δ strain, a significant decrease in R-loop level was observed, consistent with the later performed mRNA-seq results. The dynamics of R-loops, as well as genome variability, may cause some strains to come to our attention in the DRIP-BC-seq for showing increased R-loop levels, even though they are globally characterized by R-loop

depletion. In summary, examining the 10 kb environment around the barcode by DRIP-seq, we were able to validate the DRIP-BC-seq results in most of the strains. However, examining the global change in R-loop level, significant discrepancies were detected compared to the DRIP-BC-seq results. In many mutants, the DRIP signal significantly decreased across the entire genome.

3.2.4. Gene expression analysis of top mutants identified by the DRIP-BC-seq screen

We conducted gene expression analysis on the yeast strains selected for validation using mRNA-seq. Globally examining the gene expression levels, there is no major deviation between the various mutant strains and the wild-type yeast. However, looking at the annotation of genes that are underexpressed and overexpressed compared to the wild type, the gene expression changes in 3 gene mutant strains proved very interesting and are consistent with the results obtained from the whole-genome DRIP-seq. In the met7Δ strain, 33 transposon-related genes are highly expressed, of which 29 have RNA-DNA hybrid ribonuclease activity. This correlates well with the fact that in the met7Δ mutant, the quantity of R-loops decreases globally to about one-fifth. In contrast, in dss1Δ yeasts, 12 transposon genes are underexpressed, 11 of which have RNA-DNA hybrid ribonuclease activity. This also aligns with the fact that in dss1Δ, the level of R-loops nearly doubles compared to the wild type. In the met7Δ and rpo41Δ strains, the reduced R-loop level is accompanied by lower expression of genes related to ribosomes and ribonucleoprotein complexes compared to the wild type. In rpo41Δ, similar to met7Δ, there is a significant decrease in the occurrence of R-loops. In dss1Δ, however, the increased level of R-loops is associated with an increase in the expression of genes related to ribosomes and ribonucleoprotein complexes.

3.3 The impact of R-loop regulatory gene deletion ont he 3D structure of chromosomes

Beyond the effects of gene deletions on R-loops, we were also curious about their impact on the genome's three-dimensional structure, which led us to develop a technique we called 4C-BC-seq. In this method, we examine the chromatin interactions of the barcode in the yeast strains from the epi-ID library. We hope to identify previously unknown genes that modify chromatin structure, and if we find common hits between the DRIP-BC-seq and 4C-BC-seq analyses, we might also discover potential genes that regulate both R-loops and chromatin structure.

3.3.1 Implementing the 4C-BC-seq method

As a result of numerous optimizations, we were able to set up the method successfully. We demonstrated that in situ proximity ligation yielded fragments nearly the size of intact genomic DNA. The size of the fragments did not change after the second ligation, which was performed on isolated DNA following restriction enzyme digestion. This was expected because at this stage the DNA pieces only need to circularize; we do not want different fragments to ligate together and thus increase in size. The inverse PCR can also be considered successful; the blurred bands indicate that sufficiently diverse fragments were formed during the PCR process. The few stronger bands are the products of fragments that closed into themselves without an intermediate fragment ligating to them. We also performed what is called input PCR on the 4C-BC samples, which is identical to the simple, barcode-amplifying PCR used in DRIP-BC-seq. This allows us to examine the proportion of different strains present in the sample pool and thereby normalize the results of the sequenced inverse PCR samples. Thus, avoiding the detection of false interaction enrichment due to one strain being overrepresented or underrepresented in the sample mix. Based on the results of the sequenced chimeric fragments, among the nearly 400 strains examined so far, we found 6 gene mutants where the frequency and/or strength of intra-chromosomal interactions were higher than average and 3 genes where we observed more and/or stronger inter-chromosomal interactions.

The promising experiment is ongoing with other members of the epi-ID library. Based on the results of the entire yeast library, we would like to examine the chromatin interaction network at the whole genome level in a few promising strains, using the Hi-C method.

3.4. Implementing the Hi-C method for studying 3D chromosome conformation in the whole genome

We plan to later examine at the whole genome level the gene mutant strains identified by the 4C-BC-seq experiment as potentially causing changes in genome conformation, using the Hi-C method. Initially, we set up a method for examining the three-dimensional structure of chromatin that requires high-complexity bioinformatics analysis as part of another project. Our goal was to map all intra- and interchromosomal connections.We studied the effect of the

absence of the NDX (Nodulin homeobox) protein, which is presumed to have a stabilizing role in R-loops, on the three-dimensional structure of chromatin in the model plant Arabidopsis thaliana. For the study, we used nuclei isolated from 10-day-old ndx1-4 mutants and Col-0 wild-type plants. We performed in situ Hi-C on these nuclei and sequenced the samples with next-generation sequencing, achieving 200 million reads. The normalized contact matrices, examined at a resolution of 25 kb, showed very similar patterns in the ndx1-4 and Col-0 plants. This suggests that the absence of the NDX protein does not lead to extensive restructuring of the three-dimensional chromosome architecture in Arabidopsis. However, noteworthy quantitative changes were found between the two Hi-C maps when performing differential analysis of the contact matrices. In the ndx1-4 mutant, numerous genomic regions were intensified (red) or reduced (blue), indicating a global rearrangement of spatial relationships in the absence of NDX. The circos diagram of the identified differential contacts, marking the decreased (blue) and increased (red) Hi-C interaction relationships, highlights that intrachromosomal contacts mostly decrease, while interchromosomal contacts generally increase in the ndx1-4 mutant. Particularly, the number of connections between centromeres and regions characterized by clustering - knot engaged elements (KEE) - has increased. Based on this, it appears that the absence of NDX mediates distant chromatin interactions through chromocenters and transcriptionally silent chromatin (KNOT regions).

4. Discussion

R-loops are dynamic structures located in a significant part of the human genome (Sanz et al., 2016). Our new understanding of tumor genomes suggests that the genomic position, lifespan and function of R-loops are disrupted in most cancers. Experimental data demonstrate that alteration of R-loop homeostasis leads to an abnormal epigenetic phenotype that can disrupt telomere function or induce DNA replication stress and mutagenesis. However, little causal information is available regarding cancer-specific changes in which R-loops directly influence mutagenesis and chromosomal instability leading to oncogenesis. Methods suitable for the detection and genetic or chemical perturbation of R-loops have been continuously developed in recent times, which may also provide an opportunity to regulate R-loops in cancer cells. It is important to know whether the change in the R-loop level of cancer cells is merely a symptom of cell transformation or the actual driver of the cancer phenotype, so that we can decide which approach to use for their therapeutic targeting. If the accumulation of R-loops in cancer reflects a sensitive cell state, then increasing R-loops may lead to therapeutic benefits. At the same time, global inhibition of R-loops or certain types of R-loops (e.g. telomeric R-loops) can also be a therapeutic tool (Wells et al., 2019). It is currently an open question which path will be used in the field of cancer therapies. Our pharmacogenomic meta-analysis examining human RNA-DNA hybrid binding proteins confirms the association of R-loop genes with tumorigenesis and directs attention towards possible tumor therapies. In our pharmacogenomic studies with human-specific R-loop regulatory proteins, we showed that the expression level of R-loop regulators can be used to sensitize cancer cells to chemotherapy treatments. Thus, modulating the expression levels of R-loop genes can influence clinical responses to anticancer drug treatments, and these expression changes can be used to identify patient groups most likely to respond well to therapy. In addition, we identified several tumor types that showed a significant R-loop regulator-dependent survival association, i.e. the survival chances of the patient groups with high and low expression of R-loop regulators were significantly different from each other.

To expand the potential tumor therapeutic targets, we developed the DRIP-BC-seq method, which enabled us to identify previously unknown R-loop regulatory genes. Among the ~4,300 mutant strains of the epi-ID gene deletion yeast library, an increase in R-loop level in ~10 kb of the examined genomic region - barcode was observed in 110 strains, while in 89 mutants a decrease was observed. Due to the high degree of uncertainty of the screening methods, confirmation of the results requires validation experiments. For this task, we introduced the spike-in DRIP-seq technique, in which fragmented E. coli DNA is used as an internal control

during immunoprecipitation, in order to compare the samples more precisely. We selected 11 yeast mutants for validation, in which, based on DRIP-BC-seq, the amount of R-loops increased around the examined genomic region and the genes have a mitochondrial function. The results of spike-in DRIP-seq of 11 mutants, in 8 strains, confirmed that the R-loop level around the region examined by DRIP-BC-seq is higher compared to the wild-type strain in the mutant strains (ung1Δ, rpo41Δ, oye2Δ, pet20Δ, met7Δ, gas1Δ, mgm101Δ, mss116Δ), opposite results were obtained for 3 strains (exo5Δ, mgm1Δ, dss1Δ) (Figure 29). Compared to this, considering the entire genome, a significant enrichment of R-loops can be observed in the exo5Δ and dss1Δ mutants, and although to a lesser extent, the number of R-loops also increased in the mgm1Δ strain (Table 3). Furthermore, in mutants in which the level of R-loops increased around the barcode region examined by DRIP-BC-seq, we observed a large decrease globally (ung1Δ, rpo41Δ, oye2Δ, met7Δ, gas1Δ). and met7Δ mutants (Table 3). Based on the mRNA-seq data, we found a clear correlation between the expression of transposon-related genes and the global R-loop level in met7Δ and dss1Δ mutants. Similarly, in the dss1Δ and rpo41Δ mutants, the expression of genes related to the ribosome and ribonucleoprotein complex was significantly consistent with the change in the R-loop level. Based on our results, all 11 investigated gene deletions have an effect on the occurrence of R-loops. We found the results of the MET7, RPO41 and DSS1 gene mutants to be the most interesting. The MET7 and RPO41 genes seem to promote the formation and/or stabilization of R-loops, while the DSS1 gene is assumed to reduce the level of R-loops based on DRIP-seq and mRNA-seq results.

5. Summary

My PhD work focused on the examination of R-loops and chromatin loops. Within this framework, we used pharmacogenetic meta-analysis to study the role of R-loops in cancer therapy. We developed two new methods (DRIP-BC-seq, 4C-BC-seq) for identifying genes regulating R-loops and chromatin loops, respectively. Additionally, we used the Hi-C method to study the effect of the NDX gene on the spatial conformation of the genome in the model plant Arabidopsis thaliana.

The results of our pharmacogenomic analyses suggest that R-loop formation processes in cancer cells can be utilized as biomarkers, therapeutic targets, and for sensitizing certain tumors to chemotherapy treatments. The uncovered correlations offer new pathways to epigenetic therapies that are based on modifying the R-loop levels in tumors.

Using a barcoded yeast gene deletion library with a method we developed for identifying R-loop regulating genes, named DRIP-BC-seq, we identified several new R-loop regulator genes. Among the 11 genes selected for validation, five mutants - exo5Δ, pet20Δ, dss1Δ, mgm1Δ, mss116Δ - which showed R-loop increase based on DRIP-BC-seq, also showed a significant global increase in R-loop formation in terms of total DRIP-seq peak numbers. mRNA-seq analysis showed a clear correlation between gene deletion and R-loop quantity differences for three genes - MET7, DSS1, RPO41.

The 4C-BC-seq method developed to map chromatin conformation regulating genes is expected to reveal previously unknown genes controlling to 3D genome conformation. Furthermore, if there are common hits between DRIP-BC-seq and 4C-BC-seq, we may also get closer to understanding the impact of R-loops on 3D chromatin structure regulation.

With the introduction of the Hi-C method into our lab, we had the opportunity to examine the three-dimensional structure of the genome at the whole-genome level. Hi-C experiments on the ndx1-4 mutant and wild-type Arabidopsis thaliana cells showed that NDX causes global genomic architecture changes. Intrachromosomal contacts decrease in the absence of NDX, while interchromosomal contacts generally increase. We found that NDX mediates distant chromatin interactions through chromocenters and transcriptionally silent chromatin.

6. References

Aguilera, A., & Gómez-González, B. (2017). DNA–RNA hybrids: The risks of DNA breakage during transcription. *Nature Structural & Molecular Biology*, *24*(5), 439–443. https://doi.org/10.1038/nsmb.3395

Arab, K., Karaulanov, E., Musheev, M., Trnka, P., Schäfer, A., Grummt, I., & Niehrs, C. (2019). GADD45A binds R-loops and recruits TET1 to CpG island promoters. *Nature Genetics*, *51*(2), 217–223. https://doi.org/10.1038/s41588-018-0306-6

Arora, R., Lee, Y., Wischnewski, H., Brun, C. M., Schwarz, T., & Azzalin, C. M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms6220

Aymard, F., Bugler, B., Schmidt, C. K., Guillou, E., Caron, P., Briois, S., Iacovoni, J. S., Daburon, V., Miller, K. M., Jackson, S. P., & Legube, G. (2014). Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nature Structural & Molecular Biology*, *21*(4), 366–374. https://doi.org/10.1038/nsmb.2796

Barroso, S., Herrera-Moyano, E., Muñoz, S., García-Rubio, M., Gómez-González, B., & Aguilera, A. (2019). The DNA damage response acts as a safeguard against harmful DNA–RNA hybrids of different origins. *EMBO Reports*, *20*(9). https://doi.org/10.15252/embr.201847250

Becherel, O. J., Sun, J., Yeo, A. J., Nayler, S., Fogel, B. L., Gao, F., Coppola, G., Criscuolo, C., De Michele, G., Wolvetang, E., & Lavin, M. F. (2015). A new model to study neurodegeneration in ataxia oculomotor apraxia type 2. *Human Molecular Genetics*, *24*(20), 5759–5774. https://doi.org/10.1093/hmg/ddv296

Becherel, O. J., Yeo, A. J., Stellati, A., Heng, E. Y. H., Luff, J., Suraweera, A. M., Woods, R., Fleming, J., Carrie, D., McKinney, K., Xu, X., Deng, C., & Lavin, M. F. (2013). Senataxin plays an essential role with DNA damage response proteins in meiotic recombination and gene silencing. *PLoS Genetics*, *9*(4), e1003435. https://doi.org/10.1371/journal.pgen.1003435

Bhatia, V., Barroso, S. I., García-Rubio, M. L., Tumini, E., Herrera-Moyano, E., & Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature*, *511*(7509), 362–365. https://doi.org/10.1038/nature13374

Boguslawski, S. J., Smith, D. E., Michalak, M. A., Mickelson, K. E., Yehle, C. O., Patterson, W. L., & Carrico, R. J. (1986). Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. *Journal of Immunological Methods*, *89*(1), 123–130. https://doi.org/10.1016/0022-1759(86)90040-2

Chédin, F., Hartono, S. R., Sanz, L. A., & Vanoosthuyse, V. (2021). Best practices for the visualization, mapping, and manipulation of R-loops. *The EMBO Journal*, *40*(4). https://doi.org/10.15252/embj.2020106394

Chen, L., Chen, J.-Y., Huang, Y.-J., Gu, Y., Qiu, J., Qian, H., Shao, C., Zhang, X., Hu, J., Li, H., He, S., Zhou, Y., Abdel-Wahab, O., Zhang, D.-E., & Fu, X.-D. (2018). The augmented r-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Molecular Cell*, *69*(3), 412-425.e6. https://doi.org/10.1016/j.molcel.2017.12.029

Chen, L., Chen, J.-Y., Zhang, X., Gu, Y., Xiao, R., Shao, C., Tang, P., Qian, H., Luo, D., Li, H., Zhou, Y., Zhang, D.-E., & Fu, X.-D. (2017). R-ChIP using inactive rnase H reveals dynamic coupling of r-loops with transcriptional pausing at gene promoters. *Molecular Cell*, *68*(4), 745-757.e5. https://doi.org/10.1016/j.molcel.2017.10.008

Colak, D., Zaninovic, N., Cohen, M. S., Rosenwaks, Z., Yang, W.-Y., Gerhardt, J., Disney, M. D., & Jaffrey, S. R. (2014). Promoter-Bound Trinucleotide Repeat mRNA Drives Epigenetic Silencing in Fragile X Syndrome. *Science*, *343*(6174), 1002–1005. https://doi.org/10.1126/science.1245831

Costantino, L., & Koshland, D. (2015). The Yin and Yang of R-loop biology. *Current Opinion in Cell Biology*, *34*, 39–45. https://doi.org/10.1016/j.ceb.2015.04.008

Crossley, M. P., Bocek, M., & Cimprich, K. A. (2019). R-Loops as cellular regulators and genomic threats. *Molecular Cell*, *73*(3), 398–411. https://doi.org/10.1016/j.molcel.2019.01.024

Csorba, T., Questa, J. I., Sun, Q., & Dean, C. (2014). Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences*, *111*(45),

16160–16165. https://doi.org/10.1073/pnas.1419030111

Drolet, M., Phoenix, P., Menzel, R., Masse, E., Liu, L. F., & Crouch, R. J. (1995). Overexpression of RNase H partially complements the growth defect of an Escherichia coli delta topA mutant: R-loop formation is a major problem in the absence of DNA topoisomerase I. *Proceedings of the National Academy of Sciences*, *92*(8), 3526–3530. https://doi.org/10.1073/pnas.92.8.3526

Dumelie, J. G., & Jaffrey, S. R. (2017). Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *ELife*, *6*. https://doi.org/10.7554/elife.28306

eGTEx Project. (2017). Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. *Nature Genetics*, *49*(12), 1664–1670. https://doi.org/10.1038/ng.3969

El Hage, A., Webb, S., Kerr, A., & Tollervey, D. (2014). Genome-Wide Distribution of RNA-DNA Hybrids Identifies RNase H Targets in tRNA Genes, Retrotransposons and Mitochondria. *PLoS Genetics*, *10*(10), e1004716. https://doi.org/10.1371/journal.pgen.1004716

Fournier, L.-A., Kumar, A., Smith, T., Su, E., Moksa, M., Hirst, M., & Stirling, P. C. (2021). *Global prediction of candidate R-loop binding and R-loop regulatory proteins*. Cold Spring Harbor Laboratory. http://dx.doi.org/10.1101/2021.08.09.454968

García-Muse, T., & Aguilera, A. (2019). R loops: From physiological to pathological roles. *Cell*, *179*(3), 604–618. https://doi.org/10.1016/j.cell.2019.08.055

García-Rubio, M. L., Pérez-Calero, C., Barroso, S. I., Tumini, E., Herrera-Moyano, E., Rosado, I. V., & Aguilera, A. (2015). The fanconi anemia pathway protects genome integrity from r-loops. *PLOS Genetics*, *11*(11), e1005674. https://doi.org/10.1371/journal.pgen.1005674

Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., & Chédin, F. (2012). R-Loop formation is a distinctive characteristic of unmethylated human cpg island promoters. *Molecular Cell*, *45*(6), 814–825. https://doi.org/10.1016/j.molcel.2012.01.017

Glover, T. W. (2006). Common fragile sites. *Cancer Letters*, *232*(1), 4–12. https://doi.org/10.1016/j.canlet.2005.08.032

Gorthi, A., Romero, J. C., Loranc, E., Cao, L., Lawrence, L. A., Goodale, E., Iniguez, A. B., Bernard, X., Masamsetti, V. P., Roston, S., Lawlor, E. R., Toretsky, J. A., Stegmaier, K., Lessnick, S. L., Chen, Y., & Bishop, A. J. R. (2018). EWS–FLI1 increases transcription to cause R-loops and block BRCA1 repair in Ewing sarcoma. *Nature*, *555*(7696), 387–391. https://doi.org/10.1038/nature25748

Groh, M., & Gromak, N. (2014). Out of Balance: R-loops in human disease. *PLoS Genetics*, *10*(9), e1004630. https://doi.org/10.1371/journal.pgen.1004630

Groh, M., Lufino, M. M. P., Wade-Martins, R., & Gromak, N. (2014). R-loops associated with triplet repeat expansions promote gene silencing in friedreich ataxia and fragile X syndrome. *PLoS Genetics*, *10*(5), e1004318. https://doi.org/10.1371/journal.pgen.1004318

Grossman, R. L., Heath, A., Murphy, M., Patterson, M., & Wells, W. (2016). A case for data commons: Toward data science as a service. *Computing in Science & Engineering*, *18*(5), 10–20. https://doi.org/10.1109/mcse.2016.92

Grunseich, C., Wang, I. X., Watts, J. A., Burdick, J. T., Guber, R. D., Zhu, Z., Bruzel, A., Lanman, T., Chen, K., Schindler, A. B., Edwards, N., Ray-Chaudhury, A., Yao, J., Lehky, T., Piszczek, G., Crain, B., Fischbeck, K. H., & Cheung, V. G. (2018). Senataxin mutation reveals how r-loops promote transcription by blocking DNA methylation at gene promoters. *Molecular Cell*, *69*(3), 426-437.e7. https://doi.org/10.1016/j.molcel.2017.12.030

Haeusler, A. R., Donnelly, C. J., Periz, G., Simko, E. A. J., Shaw, P. G., Kim, M.-S., Maragakis, N. J., Troncoso, J. C., Pandey, A., Sattler, R., Rothstein, J. D., & Wang, J. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature*, *507*(7491), 195–200. https://doi.org/10.1038/nature13124

Halász, L., Karányi, Z., Boros-Oláh, B., Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó-L, Á., Mázló, A., Rajnavölgyi, É., Halmos, G., & Székvölgyi, L. (2017). RNA-DNA hybrid (R-loop) immunoprecipitation mapping: An analytical workflow to evaluate inherent biases. *Genome Research*, *27*(6), 1063–1073. https://doi.org/10.1101/gr.219394.116

Hatchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., Dimitrov, S., Pathania, S., McKinney, K. M., Eaton, M. L., Kellis, M., Hill, S. J., Parmigiani, G., Proudfoot, N. J., & Livingston, D. M. (2015). BRCA1 recruitment to transcriptional pause sites is required for r-loop-driven DNA damage repair. *Molecular Cell*, *57*(4), 636–647. https://doi.org/10.1016/j.molcel.2015.01.011

Hegazy, Y. A., Fernando, C. M., & Tran, E. J. (2020). The balancing act of R-loop biology: The good, the bad, and the ugly. *Journal of Biological Chemistry*, *295*(4), 905–913. https://doi.org/10.1016/s0021-9258(17)49903-0

Helmrich, A., Ballarino, M., & Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. *Molecular Cell*, *44*(6), 966–977. https://doi.org/10.1016/j.molcel.2011.10.013

Jackson, B. R., Noerenberg, M., & Whitehouse, A. (2014). A novel mechanism inducing genome instability in kaposi's sarcoma-associated herpesvirus infected cells. *PLoS Pathogens*, *10*(5), e1004098. https://doi.org/10.1371/journal.ppat.1004098

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-rna–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

König, F., Schubert, T., & Längst, G. (2017). The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLOS ONE*, *12*(6), e0178875. https://doi.org/10.1371/journal.pone.0178875

Li, X., & Manley, J. L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. *Cell*, *122*(3), 365–378. https://doi.org/10.1016/j.cell.2005.06.008

Lim, Y. W., Sanz, L. A., Xu, X., Hartono, S. R., & Chédin, F. (2015). Genome-wide DNA hypomethylation and RNA:DNA Hybrid accumulation in Aicardi–Goutières syndrome. *ELife*, *4*. https://doi.org/10.7554/elife.08007

Loomis, E. W., Sanz, L. A., Chédin, F., & Hagerman, P. J. (2014). Transcription-Associated r-loop formation across the human FMR1 cgg-repeat region. *PLoS Genetics*, *10*(4), e1004294. https://doi.org/10.1371/journal.pgen.1004294

Malig, M., Hartono, S. R., Giafaglione, J. M., Sanz, L. A., & Chedin, F. (2020). Ultra-deep coverage single-molecule r-loop footprinting reveals principles of r-loop formation. *Journal of Molecular Biology*, *432*(7), 2271–2288. https://doi.org/10.1016/j.jmb.2020.02.014

Negrini, S., Gorgoulis, V. G., & Halazonetis, T. D. (2010). Genomic instability — an evolving hallmark of cancer. *Nature Reviews Molecular Cell Biology*, *11*(3), 220–228. https://doi.org/10.1038/nrm2858

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

Petropavlovskiy, A. A., Tauro, M. G., Lajoie, P., & Duennwald, M. L. (2020). A quantitative imaging-based protocol for yeast growth and survival on agar plates. *STAR Protocols*, *1*(3), 100182. https://doi.org/10.1016/j.xpro.2020.100182

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

Pohjoismäki, J. L. O., Holmes, J. B., Wood, S. R., Yang, M.-Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., Rigby, R. E., Jackson, A. P., Spelbrink, J. N., Griffith, J. D., Crouch, R. J., Jacobs, H. T., & Holt, I. J. (2010). Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *Journal of Molecular Biology*, *397*(5), 1144–1155. https://doi.org/10.1016/j.jmb.2010.02.029

Powell, W. T., Coulson, R. L., Gonzales, M. L., Crary, F. K., Wong, S. S., Adams, S., Ach, R. A., Tsang, P., Yamada, N. A., Yasui, D. H., Chédin, F., & LaSalle, J. M. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. *Proceedings of the National Academy of Sciences*, *110*(34), 13938–13943. https://doi.org/10.1073/pnas.1305426110

Qin, Y., Conley, A. P., Grimm, E. A., & Roszik, J. (2017). A tool for discovering drug sensitivity and gene expression associations in cancer cells. *PLOS ONE*, *12*(4), e0176763. https://doi.org/10.1371/journal.pone.0176763

Reddy, K., Schmidt, M. H. M., Geist, J. M., Thakkar, N. P., Panigrahi, G. B., Wang, Y.-H., & Pearson, C. E. (2014). Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic Acids Research*, *42*(16), 10473–10487. https://doi.org/10.1093/nar/gku658

Repana, D., Nulsen, J., Dressler, L., Bortolomeazzi, M., Venkata, S. K., Tourna, A., Yakovleva, A., Palmieri, T., & Ciccarelli, F. D. (2019). The Network of Cancer Genes (NCG): A comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-018-1612-0

Roy, R., Chun, J., & Powell, S. N. (2012). BRCA1 and BRCA2: Different roles in a common pathway of genome protection. *Nature Reviews Cancer*, *12*(1), 68–78. https://doi.org/10.1038/nrc3181

Sagie, S., Toubiana, S., Hartono, S. R., Katzir, H., Tzur-Gilat, A., Havazelet, S., Francastel, C., Velasco, G., Chédin, F., & Selig, S. (2017). Telomeres in ICF syndrome cells are vulnerable to DNA damage due to elevated DNA:RNA Hybrids. *Nature Communications*, *8*(1). https://doi.org/10.1038/ncomms14015

Santos-Pereira, J. M., & Aguilera, A. (2015). R loops: New modulators of genome dynamics and function. *Nature Reviews Genetics*, *16*(10), 583–597. https://doi.org/10.1038/nrg3961

Sanz, L. A., Hartono, S. R., Lim, Y. W., Steyaert, S., Rajpurkar, A., Ginno, P. A., Xu, X., & Chédin, F. (2016). Prevalent, dynamic, and conserved r-loop structures associate with specific epigenomic signatures in mammals. *Molecular Cell*, *63*(1), 167–178. https://doi.org/10.1016/j.molcel.2016.05.032

Schwab, R. A., Nieminuszczy, J., Shah, F., Langton, J., Lopez Martinez, D., Liang, C.-C., Cohn, M. A., Gibbons, R. J., Deans, A. J., & Niedzwiedz, W. (2015). The fanconi anemia pathway maintains genome stability by coordinating replication and transcription. *Molecular Cell*, *60*(3), 351–361. https://doi.org/10.1016/j.molcel.2015.09.012

Shivji, M. K. K., Renaudin, X., Williams, Ç. H., & Venkitaraman, A. R. (2018). BRCA2 regulates transcription elongation by RNA polymerase II to prevent r-loop accumulation. *Cell Reports*, *22*(4), 1031–1039. https://doi.org/10.1016/j.celrep.2017.12.086

Skourti-Stathaki, K., Kamieniarz-Gdula, K., & Proudfoot, N. J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature*, *516*(7531), 436–439. https://doi.org/10.1038/nature13787

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

Skourti-Stathaki, K., Proudfoot, N. J., & Gromak, N. (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote xrn2-dependent termination. *Molecular Cell*, *42*(6), 794–805. https://doi.org/10.1016/j.molcel.2011.04.026

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

Sollier, J., Stork, C. T., García-Rubio, M. L., Paulsen, R. D., Aguilera, A., & Cimprich, K. A. (2014). Transcription-Coupled nucleotide excision repair factors promote r-loop-induced genome instability. *Molecular Cell*, *56*(6), 777–785. https://doi.org/10.1016/j.molcel.2014.10.020

Stork, C. T., Bocek, M., Crossley, M. P., Sollier, J., Sanz, L. A., Chédin, F., Swigut, T., & Cimprich, K. A. (2016). Co-transcriptional R-loops are the main cause of estrogen-induced DNA damage. *ELife*, *5*. https://doi.org/10.7554/elife.17548

Sun, Q., Csorba, T., Skourti-Stathaki, K., Proudfoot, N. J., & Dean, C. (2013). R-Loop Stabilization Represses Antisense Transcription at the Arabidopsis FLC Locus. *Science*, *340*(6132), 619–621. https://doi.org/10.1126/science.1234848

Tan, S. L. W., Chadha, S., Liu, Y., Gabasova, E., Perera, D., Ahmed, K., Constantinou, S., Renaudin, X., Lee, M., Aebersold, R., & Venkitaraman, A. R. (2017). A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability. *Cell*, *169*(6), 1105-1118.e15.

https://doi.org/10.1016/j.cell.2017.05.010

Thomas, M., White, R. L., & Davis, R. W. (1976). Hybridization of RNA to double-stranded DNA: Formation of R-loops. *Proceedings of the National Academy of Sciences*, *73*(7), 2294–2298. https://doi.org/10.1073/pnas.73.7.2294

Vanoosthuyse, V. (2018). Strengths and weaknesses of the current strategies to map and characterize r-loops. *Non-Coding RNA*, *4*(2), 9. https://doi.org/10.3390/ncrna4020009

Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., & Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes & Development*, *30*(11), 1327–1338. https://doi.org/10.1101/gad.280834.116

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

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., & Stuart, J. M. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

Wells, J. P., White, J., & Stirling, P. C. (2019). R loops and their composite cancer connections. Trends in Cancer, 5(10), 619–631. https://doi.org/10.1016/j.trecan.2019.08.006

Westover, K., Bushnell, D., & Kornberg, R. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, *119*(7), 1055. https://doi.org/10.1016/s0092-8674(04)01206-1

Wu, T., Nance, J., Chu, F., & Fazzio, T. G. (2021). Characterization of R-Loop–Interacting Proteins in Embryonic Stem Cells Reveals Roles in rRNA Processing and Gene Expression. *Molecular & Cellular Proteomics*, *20*, 100142. https://doi.org/10.1016/j.mcpro.2021.100142

Xu, B., & Clayton, D. A. (1996). RNA-DNA hybrid formation at the human mitochondrial heavy-strand origin ceases at replication start sites: An implication for RNA-DNA hybrids serving as primers. *The EMBO Journal*, *15*(12), 3135–3143. https://doi.org/10.1002/j.1460-2075.1996.tb00676.x

Xu, W., Xu, H., Li, K., Fan, Y., Liu, Y., Yang, X., & Sun, Q. (2017). The R-loop is a common chromatin feature of the Arabidopsis genome. *Nature Plants*, *3*(9), 704–714. https://doi.org/10.1038/s41477-017-0004-x

Yang, W., Lightfoot, H., Bignell, G., Behan, F., Cokelear, T., Haber, D., Engelman, J., Stratton, M., Benes, C., McDermott, U., & Garnett, M. (2016). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for biomarker discovery in cancer cells. *European Journal of Cancer*, *69*, S82. https://doi.org/10.1016/s0959-8049(16)32839-8

Yang, Y., McBride, K. M., Hensley, S., Lu, Y., Chedin, F., & Bedford, M. T. (2014). Arginine methylation facilitates the recruitment of TOP3B to chromatin to prevent R loop accumulation. *Molecular Cell*, *53*(3), 484–497. https://doi.org/10.1016/j.molcel.2014.01.011

Yu, K., Chedin, F., Hsieh, C.-L., Wilson, T. E., & Lieber, M. R. (2003). R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunology*, *4*(5), 442–451. https://doi.org/10.1038/ni919

Zeller, P., Padeken, J., van Schendel, R., Kalck, V., Tijsterman, M., & Gasser, S. M. (2016). Histone H3K9 methylation is dispensable for Caenorhabditis elegans development but suppresses RNA:DNA Hybrid-associated repeat instability. *Nature Genetics*, *48*(11), 1385–1395. https://doi.org/10.1038/ng.3672

Zhang, X., Chiang, H.-C., Wang, Y., Zhang, C., Smith, S., Zhao, X., Nair, S. J., Michalek, J., Jatoi, I., Lautner, M., Oliver, B., Wang, H., Petit, A., Soler, T., Brunet, J., Mateo, F., Angel Pujana, M., Poggi, E., Chaldekas, K., … Li, R. (2017). Attenuation of RNA polymerase II pausing mitigates BRCA1-associated R-loop accumulation and tumorigenesis. *Nature Communications*, *8*(1). https://doi.org/10.1038/ncomms15908

Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K. H., Ke, E., Verma, S., Soroczynski, J., Yates, J. R., III, Hunter, T., & Verma, I. M. (2018). Heterochromatin-Encoded satellite rnas induce breast cancer. *Molecular Cell*, *70*(5), 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023

7. Acknowledgements

First of all, I would like to express my gratitude to my supervisor, Dr. Lóránt Székvölgyi, for his inexhaustible patience and support during my PhD work. Without his professional and personal support, this work would not have been possible.

I would like to thank Prof. Dr. László Fésüs, Prof. Dr. József Tőzsér and Prof. Dr. István Balogh, former and current heads of the Doctoral School of Molecular Cell and Immunobiology, for allowing me to complete my PhD studies within the framework of the doctoral school.

I thank the former and current members of the Lendület Genome Structure and Recombination Research Group (Éva Nagy, Csaba Fillér, Adrienn Horváth, Dr. Dóra Szabóné Varga, Orsolya Feró, Zsolt Karányi, Dr. Szabolcs Hetey, Dr. László Halász, Dr. Ágnes Mosolygó-Lukács, Dr. Dénes Nagy, Erzsébet Tóth, Zsanett Bukó, Ibolya Fürtös) provided their help during laboratory work and group discussions and made everyday life more colorful.

I am very grateful to Dr. János Roszik (The University of Texas MD Anderson Cancer Center), Zsolt Karányi and Orsolya Feró for performing the bioinformatics analyzes related to the research.

I thank my co-authors for their very high-quality work, thanks to which we managed to publish our joint works in the most scientifically recognized journals.

I would like to thank Dr. Lilla Ozgyin for introducing me to the mysteries of 4C-seq and for sharing her practical experience with the technique.

I am also grateful to Dr. Szilárd Póliska and Erzsébet Mátyás for their help and work in the sequencing.

Last but not least, I am grateful to my family, my husband, my daughter, my parents, and my friends for encouraging me and supporting me along the way.

.

8. Publication list

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/404/2022.PL
Subject: PhD Publication List

Candidate: Beáta Kissné Boros-Oláh
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology
MTMT ID: 10071902

List of publications related to the dissertation

1. Karányi, Z., Mosolygó, Á., Feró, O., Horváth, A., **Boros-Oláh, B.**, Nagy, É., Hetey, S., Holb, I., Szaker, H., Miskei, M., Csorba, T., Székvölgyi, L.: NODULIN HOMEOBOX is required for heterochromatin homeostasis in Arabidopsis.
   Nat Comms. 13, 1-20, 2022.
   DOI: http://dx.doi.org/10.1038/s41467-022-32709-y
   IF: 17.694 (2021)
2. **Boros-Oláh, B.**, Dobos, N., Hornyák, L., Szabó, Z., Karányi, Z., Halmos, G., Roszik, J., Székvölgyi, L.: Drugging the R-loop interactome: RNA-DNA hybrid binding proteins as targets for cancer therapy.
   DNA Repair. 84, 1-10, 2019.
   DOI: http://dx.doi.org/10.1016/j.dnarep.2019.102642
   IF: 3.339

List of other publications

3. Karányi, Z., Halász, L., Acquaviva, L., Jonás, D., Hetey, S., **Boros-Oláh, B.**, Peng, F., Chen, D., Klein, F., Géli, V., Székvölgyi, L.: Nuclear dynamics of the Set1C subunit Spp1 prepares meiotic recombination sites for break formation.
   J. Cell Biol. 217 (10), 3398-3415, 2018.
   DOI: http://dx.doi.org/10.1083/jcb.201712122
   IF: 8.891
4. Hetey, S., **Boros-Oláh, B.**, Kuik-Rózsa, T., Li, Q., Karányi, Z., Szabó, Z., Roszik, J., Szaloki, N., Vámosi, G., Tóth, K. Á., Székvölgyi, L.: Biophysical characterization of histone H3.3 K27M point mutation.
   Biochem. Biophys. Res. Commun. 490 (3), 868-875, 2017.
   DOI: http://dx.doi.org/10.1016/j.bbrc.2017.06.133
   IF: 2.559

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

5. Halász, L., Karányi, Z., **Boros-Oláh, B.**, Kuik-Rózsa, T., Sipos, É., Nagy, É., Mosolygó, Á., Türk-Mázló, A., Rajnavölgyi, É., Halmos, G., Székvölgyi, L.: RNA-DNA hybrid (R-loop) immunoprecipitation mapping: an analytical workflow to evaluate inherent biases. *Genome Res. 27*, 1063-1073, 2017.
DOI: http://dx.doi.org/10.1101/gr.219394.116
IF: 10.101

6. Szénási, T., Kénesi, E., Nagy, A., Molnár, A., Bálint, B. L., Zvara, Á., Csabai, Z., Deák, F., **Boros-Oláh, B.**, Mátés, L., Nagy, L., Puskás, L. G., Kiss, I.: Hmgb1 can facilitate activation of the matrilin-1 gene promoter by Sox9 and L-Sox5/Sox6 in early steps of chondrogenesis. *Biochim. Biophys. Acta. Gene. Regul. Mech. 1829* (10), 1075-1091, 2013.
DOI: http://dx.doi.org/10.1016/j.bbagrm.2013.07.004
IF: 5.44

7. Franyó, D., **Boros-Oláh, B.**, Ozgyin, L., Bálint, B. L.: Befolyásolja-e az életmód génjeink működését?: az epigenetikai kutatások irányvonalai és eredményei. *LAM KID. 2* (1), 37-42, 2012.

Total IF of journals (all publications): 48,024
Total IF of journals (publications related to the dissertation): 21,033

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

30 August, 2022

DEBRECENI EGYETEM
Egyetemi és Nemzeti Könyvtár