Proximity Labeling as a Tool to Study Transcription-Replication Interference

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften an
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von

Henning Ummethum

September 2023

Diese Dissertation wurde durchgeführt
im Labor von Dr. Stephan Hamperl
am Institut für Epigenetik und Stammzellen
des Helmholtz Zentrum München

Erstgutachterin:	Prof. Dr. Maria-Elena Torres-Padilla
Zweitgutachter:	Prof. Dr. Claude Becker
Tag der Abgabe:	11.09.2023
Tag der mündlichen Prüfung:	27.02.2024

ERKLÄRUNG

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.

Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.

Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 11.09.2023

Henning Ummethum

Table of Contents

Summary

During S phase, thousands of transcription and replication machineries occupy the same DNA template in eukaryotic cells to perform their essential functions in gene expression and genome duplication, respectively. Tight coordination of these two processes is necessary to prevent interference with each other leading to genome-destabilizing transcription-replication conflicts. The exact molecular mechanisms when these conflicts arise and how they are resolved are largely unknown which is in part due to the lack of currently available methods to study them on the genome in a comprehensive way. In this work, I developed a proximity labeling system using a replisome component as a bait to investigate the proteomic composition of active replication forks while inducing unscheduled transcription-replication conflicts in a human breast cancer cell line. As a result, I found a list of 88 candidate proteins that were enriched at transcriptionally challenged replication forks, from which 13 factors were further characterized for their potential role in the regulation of transcription-replication conflicts. Among these, the CGG-triplet repeat binding protein CGGBP1 showed a clear effect to cause transcription-induced replication stress. CGGBP1 is essential and binds to CGG repeats at promoters that are more prone to form unusual DNA secondary structures, such as G-quadruplexes and R-loops. Immuno-fluorescence imaging uncovered a role of CGGBP1 in preventing RNA polymerase II accumulation on chromatin. Intriguingly, CGGBP1 depletion led to an increase in RNA polymerase II – PCNA proximity ligation assay foci, indicating elevated levels of transcription-replication conflicts in CGGBP1-deficient cells. Finally, CGGBP1 depletion led to DNA:RNA hybrid formation at an episomal transcription unit containing CGG repeats, suggesting that this factor counteracts DNA secondary structure formation at CGG repeats, thereby preventing RNAPII accumulation and resulting transcription-replication conflicts. In conclusion, I identified a list of candidate transcription-replication conflict factors and found a role of CGGBP1 in preventing conflicts.

Zusammenfassung

Während der S-Phase besetzen tausende von Transkriptions- und Replikations- maschinerien dieselbe DNA-Vorlage in eukaryotischen Zellen, um ihre wesentlichen Funktionen bei der Genexpression und der Duplikation des Genoms durchzuführen. Eine enge Koordination dieser beiden Prozesse ist notwendig, um Störungen zu vermeiden, die zu genomschädigenden Konflikten zwischen Transkription und Replikation führen könnten. Die genauen molekularen Mechanismen zur Entstehung dieser Konflikte und wie sie gelöst werden, sind größtenteils unbekannt, was zum Teil auf den Mangel an derzeit verfügbaren Methoden zur umfassenden Untersuchung auf genetischer Ebene zurückzuführen ist. In dieser Arbeit habe ich eine Proximity-labeling Methode entwickelt, bei der ich eine Replisomkomponente als Köder verwendete, um die proteomische Zusammensetzung aktiver Replikationsgabeln zu untersuchen, während ich ungeplante Transkriptions-Replikationskonflikte in einer humanen Brustkrebszelllinie induzierte. Als Ergebnis fand ich eine Liste von 88 Kandidatenproteinen, von denen 13 Faktoren näher charakterisiert wurden hinsichtlich ihrer potenziellen Rolle bei der Regulation von Transkriptions- Replikationskonflikten. Die Ergebnisse zeigten, dass das Fehlen des CGG-repeat bindenden Proteins CGGBP1 zu Transkriptions-induzierten Replikationsstress führte. CGGBP1 ist essenziell und bindet an CGG-repeats in Promotoren, die dazu neigen, ungewöhnliche DNA-Sekundärstrukturen zu bilden, wie G-Quadruplexe und R-Loops. Immunfluoreszenzbildgebung enthüllte eine Rolle von CGGBP1 bei der Verhinderung der Akkumulation von RNA-Polymerase II auf Chromatin. Interessanterweise führte die Depletion von CGGBP1 zu einer Zunahme von Signal im RNA-Polymerase II-PCNA-Proximity-Ligation-Assay, was auf ein erhöhtes Niveau von Transkriptions-Replikationskonflikten in CGGBP1-defizienten Zellen hinweist. Schließlich führte die Depletion von CGGBP1 zur Bildung von DNA:RNA-Hybriden an einer episomalen Transkriptionseinheit, die CGG-repeats enthält, was darauf hindeutet, dass dieses Protein der Bildung von DNA- Sekundärstrukturen an CGG-repeats entgegenwirkt und so die Akkumulation von RNAPII und Transkriptions-Replikationskonflikten verhindert. Zusammenfassend identifizierte ich eine Liste von Kandidatenfaktoren für Transkriptions- Replikationskonflikte und fand eine Rolle von CGGBP1 bei der Verhinderung von Konflikten.

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Denotation	Abbreviation	Denotation
5' UTR	5' untranslated region	HU	Hydroxy urea
Acetyl-CoA	Acetyl coenzyme A	ICL	Interstrand cross-link
AP	Affinity purification	IP	Immunoprecipitation
BAP	Biotin acceptor peptide	kb	Kilobase
BCCP	Biotin carboxyl carrier protein	lncRNA	Long non-coding RNA
BioAMP	Biotinoyl-5′-adenylate	mRNA	Messenger RNA
CD	Co-directional	MS	Mass spectrometry
CFS	Common fragile site	NLS	Nuclear localization signal/sequence
CMG	Cdc45-mcm-gins	PBS	Phosphate-buffered saline
DMEM	Dulbecco's modified eagle medium	PCA	Protein-fragment complementation assay
DNA	Deoxyribonucleic acid	PDS	Pyridostatin
DOX	Doxycycline	PFA	Paraformaldehyde
DPBS	Dulbecco's phosphate-buffered saline	PLA	Proximity ligation assay
DPC	DNA-protein cross-link	POI	Protein of interest
DRB	5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside	RNA	Ribonucleic acid
DRIP	DNA-RNA immunoprecipitation	RNAi	RNA interference
E2	Estradiol	RNAPI	RNA polymerase I
EdU	5-ethynyl 2′-deoxyuridine	RNAPII	RNA polymerase II
ERF	Early fragile site	RNAPIII	RNA polymerase III
EtOH	Ethanol	rRNA	Ribosomal RNA
EU	5-Ethynyl-uridine	siRNA	Small interfering RNA
FASP	Filter-aided sample preparation	TES	Transcrition end site
FBS	Fetal bovine serum	TRC	Transcriptin-replication conflict
FXS	Fragile X syndrome	tRNA	Transfer RNA
HO	Head-on	TSS	Transcription start site
HRP	Horseradish peroxidase		

1 Introduction

1.1 Transcription-Replication Conflicts

1.1.1 What is a transcription-replication conflict?

In each organism, the genetic information stored in the DNA is read and processed by transcription and duplicated by replication for subsequent cell division. Transcription of ribosomal RNA (rRNA) and transfer RNA (tRNA) is carried out by RNA polymerase I and III, respectively. RNA polymerase II (RNAPII) synthesizes messenger RNAs (mRNA) from the ~19,000 protein-coding genes, but also ~19,000 long non-coding RNAs (lncRNA) are produced by this essential machinery (Frankish et al., 2020). The human body is composed of ~ 200 different cell types that share the identical genomic sequence, but display dramatic differences in their transcriptional output and underlying gene expression landscapes. While rRNA and tRNA is transcribed from a few hundred loci scattered throughout the human chromosomes, large amounts of RNAPII need to progress through thousands of loci simultaneously to ensure mRNA homeostasis and cell functionality.

Similarly, DNA replication initiates at thousands of origins across the genome. Although confined to the S phase of the cell cycle, some sets of genes perform essential functions during the S phase, such as replication factors and core histone genes that work in concert to duplicate the DNA and facilitate the packaging of the newly synthesized DNA into a protective chromatin structure (Kurat et al., 2011; Marzluff et al., 2008). Additionally, ribosomal RNA genes ensure a continuous production of ribosomes (Zaidi et al., 2016), and certain long genes, which are initiated in the G1 phase, continue their transcriptional cycle into the S phase (Helmrich et al., 2011). Thus, a key unanswered question is how the cell coordinates transcription and replication on the same sequence template during S phase to prevent physical transcription-replication conflicts (TRCs).

1.1.2 Which types of TRCs exist?

Transcription and replication machineries can encounter each other with varying frequencies and impact on genome integrity. While our knowledge of these events remains incomplete, it is increasingly clear that the nature and severity of collisions on genome stability are influenced by the orientation of the conflict and functional

state of the transcription block. Depending on which side the replication fork enters a gene, the conflict can either happen in a head-on (HO) or co-directional (CD) orientation (**Figure 1A**).
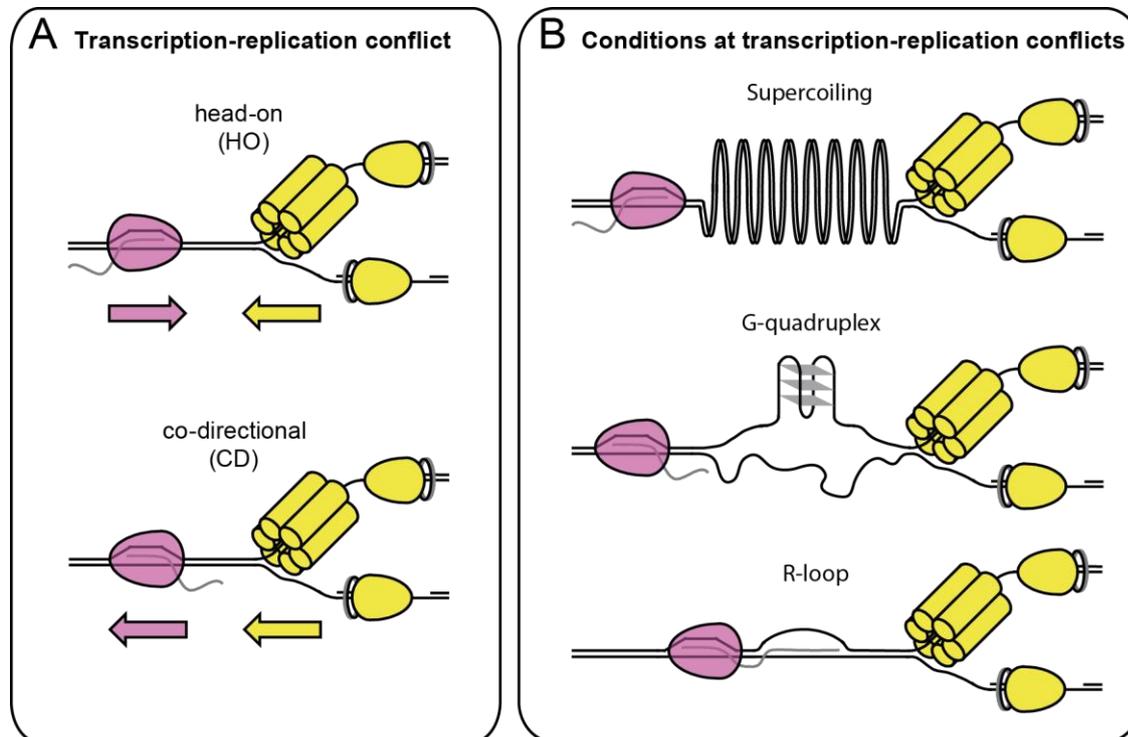
Figure 1: Illustration of transcription-replication conflicts. **A**, Illustration of the two possible orientations head-on and co-directional. **B**, Illustration of different conditions that can occur at transcription-replication conflicts. Replication forks are colored yellow and transcription machineries are depicted in magenta.

It is clear that RNA polymerases are physical obstacles to replication forks regardless of orientation (Azvolinsky et al., 2009; Bedinger et al., 1983; Merrikh et al., 2011). However, a large number of studies indicates that collisions in HO orientation are a more potent block to replication forks and thus can lead to more severe genomic alterations (Hamperl et al., 2017; Lang et al., 2017; Lang & Merrikh, 2021; Prado & Aguilera, 2005; Srivatsan et al., 2010). The most likely explanation to why a HO conflict poses a greater threat for cells is that CD conflicts can be resolved more easily, because the replication fork can continue progression once the RNA polymerase reaches the end of the gene and terminates. In addition to the orientation, other circumstances can also influence the occurrence or severity of conflicts. Both transcription and replication positively supercoil the DNA

ahead of the machineries, which introduces torsional stress that is usually resolved by DNA topoisomerases (**Figure 1B**). Upon depletion of DNA topoisomerases in yeast and human cells, there is an increase in TRCs likely due to the buildup of torsional stress (Bermejo et al., 2007; Tuduri et al., 2009). These elevated TRC levels are associated with DNA damage and slowing of the replication fork in S phase, which can be rescued by inhibition of transcription (Bermejo et al., 2007; Tuduri et al., 2009). In addition, secondary non-B DNA structures forming primarily at repetitive regions can interfere with both transcription and replication, creating genomic instability hotspots (J. Zhao et al., 2010). These structures, such as hairpins or G-quadruplexes, can be formed by the exposed ssDNA that is present during transcription or replication. Studies on mutants of the budding yeast DNA helicase Pif1 and the fission yeast Pfh1 (a homologue of Pif1) have provided evidence to support the idea that non-B DNA structures may cause transcription-mediated replication fork stalling. These studies have shown that Pif1 and Pfh1 can unwind G-quadruplexes *in vitro*. When Pif1 or Pfh1 are missing, replication is slowed or stopped in areas with high G-quadruplex density, as well as in genes that are highly transcribed by RNA Pol II and Pol III (Paeschke et al., 2011; Sabouri et al., 2012).

Another DNA secondary structure associated with transcription-replication conflicts are co-transcriptional R-loops. R-loops are nucleic acid structures composed of RNA-DNA hybrids and a displaced DNA strand. They are created when newly synthesized RNA temporarily re-joins with the template DNA strand located behind the RNA polymerase. These transient, reversible structures can arise at thousands of genes and play important roles in transcription, DNA replication, and DNA repair (Crossley et al., 2019; Ginno et al., 2012, 2013; Lim et al., 2015; Nadel et al., 2015; Sanz et al., 2016). However, they are also a transcriptional barrier to replication that can inhibit fork progression and induce DNA damage (Gan et al., 2011; Hamperl et al., 2017; Lang et al., 2017; Stork et al., 2016). Furthermore, R-loops are likely exacerbating the functional consequences of TRCs, as evidenced by the fact that overexpression of RNase H can decrease DNA damage and replication fork slowing in the presence of hormone or oncogene-induced replication stress (Kotsantis et al., 2016; Stork et al., 2016). In addition, the studies conducted by Lang et al. (2017) on an engineered bacterial genome and by Hamperl et al. (2017)

on a mammalian episomal system showed that R-loops hindered replication and caused significant genome instability when encountered by replication forks in HO orientation. However, when replication and transcription occurred in the same direction (CD), the effects were more manageable. Furthermore, FANCA and FANCD2, two components of the Fanconi anemia DNA repair pathway (Moldovan & D'Andrea, 2009), have a role in inhibiting R-loops and thereby preventing DNA damage caused by TRCs (García-Rubio et al., 2015; Okamoto et al., 2019; Schwab et al., 2015).

1.1.3 Strategies to prevent or minimize TRCs

There are many strategies in place to prevent cellular TRCs. First, replication is confined to S phase and early microscopy studies have shown that active sites of transcription and replication are mostly separated throughout S phase in mouse cells (Wei et al., 1998). However, the analysis of transcription-replication segregation in human cancer cell lines showed conflicting findings (Hassan et al., 1994; Wansink et al., 1994), suggesting that cell-to-cell heterogeneities may exist. Later, nascent RNA sequencing at different time points in S phase supported an inverse correlation between transcription and replication (Meryet-Figuiere et al., 2014). This study showed that late replicating genes were maximally transcribed during early S phase and early replicating genes had their peak transcription during late S phase. Together, this proposed a model where replication and transcription form distinct nuclear foci and when replication moves through a particular gene, transcription is transiently downregulated in that region. However, transcription cannot be shut down in genes necessary for S phase progression, such as replication factors, core histone genes and ribosomal RNA (Kurat et al., 2011; Marzluff et al., 2008; Zaidi et al., 2016). Additionally, transcription of long genes can continue into S phase (Helmrich et al., 2011). Consequently, TRCs are bound to happen at these loci or if the cell mismanages the separation of the two processes during S phase.

Second, as the consequences of HO-TRCs are more detrimental, cells have developed strategies to minimize them. In bacteria, there is a bias for genes to be orientated in the same direction as replication. Importantly, the highly expressed and essential rRNA and tRNA genes show a preferential co-orientation bias with replication (Guy & Roten, 2004; McLean et al., 1998; Rocha & Danchin, 2003a,

2003b). In human cells, many origins of replication are located near transcription start sites (TSS), leading to a CD bias and effectively minimizing HO-TRCs (Chen et al., 2019; Petryk et al., 2016). Additionally, eukaryotic cells have other strategies to prevent HO-TRCs in highly transcribed regions, e.g. a replication fork barrier at the 3'-end of the ribosomal DNA gene clusters (Akamatsu & Kobayashi, 2015; Brewer & Fangman, 1988).

Third, there are a plethora of factors described that are traveling with the replication fork or transcription machinery to prevent or resolve TRCs. Stalled and backtracked RNA polymerases along the genome constitute a major problem for replication forks. Therefore, factors ensuring proper transcription elongation can prevent TRCs. For example, the bacterial GreA and GreB can degrade the nascent RNA of backtracked RNA polymerases, effectively restarting transcription (Opalka et al., 2003). In addition, the transcription factor DksA facilitates elongation by decreasing the occurrence of nucleotide misincorporation, which can cause RNA polymerase pausing and backtracking leading to a reduction in transcriptional roadblocks (Roghanian et al., 2015; Tehranchi et al., 2010). In eukaryotes, the transcription elongation factor TFIIS functions similar to the bacterial factors mentioned, as it promotes transcript cleavage to enable the resumption of arrested RNAP complexes through a comparable mechanism (Cheung & Cramer, 2011). In contrast, the human RECQL5 helicase prevents TRCs by reducing the transcription elongation rate to a more controlled level that leads to less RNA polymerase stalling or backtracking (Saponaro et al., 2014).

1.1.4 Conflict resolution

Once a TRC has occurred, there are three main possibilities to overcome them. RNA polymerase may undergo degradation or eviction from chromatin, enabling the resumption of DNA replication. Alternatively, the replication fork can be briefly cleaved and re-ligated to facilitate RNAP's continuation of transcription and movement beyond the replication fork. Finally, if the conflict cannot be resolved or bypassed, the replisome can be dismantled, and replication can be completed using an upstream or downstream converging replication fork originating from a neighboring dormant origin. Replicative helicases are major factors able to remove DNA-associated protein complexes, including RNA polymerases. For example, *E. coli* Rep, DinG and UrvD as well as Rrm3 in yeast are known to have this function

(Hawkins et al., 2019; Ivessa et al., 2003). The transcriptional machinery also has associating factors that can remove or degrade chromatin-bound RNA polymerases. For example, it has been demonstrated that BRD4, a co-activator of transcription, promotes the movement of replication forks through active genes. Depletion of BRD4 results in elevated R-loop levels, which in turn leads to slower replication forks and DNA damage (F. C. Lam et al., 2020). Furthermore, in *S. pombe*, the eviction of RNA polymerases from centromeric repeats is reliant on the RNAi pathway. When this pathway is disrupted, stalled replication forks accumulate and the homologous recombination pathway is necessary to restart DNA replication (Zaratiegui et al., 2011). Consistently, the absence of the homologous recombination repair factors BRCA2 or RAD51 both lead to elevated TRC levels and transcription-induced DNA damage in human cells (Bhowmick et al., 2022; Groelly et al., 2022). In contrast, the Dicer protein Drc1 is involved in regulating transcription termination and removal of RNAP II independently of the RNAi pathway (Castel et al., 2014). Furthermore, RNAP II accumulates in Drc1Δ cells at sites where replication forks pause, such as the rDNA genes, tRNA genes or actively transcribed genes. This indicates that Drc1 has a secondary, RNAi-independent function in removing RNAP II from chromatin at TRC sites and highlights the wide variety of pathways for RNA polymerase removal. Chromatin context also seems to be important for efficient RNA polymerase removal. For example, the DNA replication checkpoint sensor Mec1 (the yeast homologue of ATR), the chromatin remodeling complex INO80C, and the transcription complex PAF1C work together to prevent conflicts between transcription and replication under hydroxyurea(HU)-induced replication stress (Poli et al., 2016). These factors are essential for the efficient removal of a subset of RNA polymerase II from transcribed genes near early firing origins under HU treatment, reducing the likelihood of transcription-replication fork collisions. Failure to remove RNAP II correlates inversely with recovery from replication stress, suggesting that these factors play a crucial role in preserving genome stability (Poli et al., 2016).

When the replisome encounters a stalled RNAP co-directionally, it can remove the RNAP and utilize the nascent RNA molecule as a starting point to restart replication, which helps to avoid transcription-replication conflicts (TRC skipping) (**Figure 2**, Pomerantz & O'Donnell, 2008). *In vitro*, reconstituted E. coli replisomes

were only temporarily slowed when confronted with a CD RNAP and most RNAPs could be displaced by the replicative helicase (Bayona-Feliu et al., 2021; Brüning & Marians, 2020; Gómez-González & Aguilera, 2019). Even though these controlled *in vitro* reactions offer an understanding of the kinetics of conflicts, the scenario *in vivo* is likely more complex as for example, multiple RNAPs simultaneously transcribe the same gene while interacting with additional proteins not present *in vitro*. Multiple *in vitro* studies set out to address how the *E. coli* replisome manages complex RNAP arrays and replication-R-loop conflicts. The findings were in line with earlier studies, which showed that RNAP complexes in CD orientation caused temporary obstructions, whereas those in the HO orientation resulted in significant fork stalling, especially when faced with multiple RNAPs (García-Rubio et al., 2018; N. Kim et al., 2007; Sankar et al., 2016; Schauer et al., 2020). An alternative to repriming on the nascent RNA involves bypassing the transcriptional block and continuing DNA replication by generating a new Okazaki fragment behind the transcription machinery (Conti & Smogorzewska, 2020; Gómez-González & Aguilera, 2019). This straightforward and efficient approach utilizes the inherent gaps in the process of lagging strand replication to navigate around any impediments in the lagging strand. In contrast, leading strand skipping and repriming necessitates interruption of ongoing DNA synthesis at the replication fork, which ultimately results in the formation of single-stranded DNA gaps that must be processed by post-replication repair mechanisms (Conti & Smogorzewska, 2020; Gómez-González & Aguilera, 2019). The PRIMPOL protein utilizes its dual polymerase and primase activities to carry out this function in human cells (García-Gómez et al., 2013; Mourón et al., 2013). Studies indicate that cells lacking PRIMPOL exhibit heightened sensitivity to replication stress, highlighting the significance of this protein in leading-strand lesion skipping and repriming (Bianchi et al., 2013; Kobayashi et al., 2016). Moreover, PRIMPOL is not traveling along with the replication fork, indicating that it specifically recognizes and is actively recruited to stalled replication forks (Wessel et al., 2019). Additionally, it has been demonstrated that PRIMPOL can reprime downstream of non-B-DNA co-transcriptional structures such as G-quadruplexes (G4s) and R-loops, further supporting its suggested role to promote TRC skipping and repriming (Gómez-González & Aguilera, 2019; Šviković et al., 2019). It should be emphasized that the ability of the replisome to successfully traverse a stalled RNAP complex remains

uncertain. *In vitro* experiments using the T4 replisome have shown that it can pass the RNAP without separating it from the DNA, while in contrast, the *E. coli* replisome displaces both CD and HO RNAPs and can reprime downstream (Pomerantz & O'Donnell, 2008, 2010). In eukaryotic cells, bulky protein-DNA adducts and interstrand cross-links (ICLs) have been considered absolute barriers for the replicative helicase. Intriguingly, replication can resume after encountering an ICL, indicating that the replicative helicase may be capable of adopting an open-ring conformation to move past these obstacles (Huang et al., 2013; Trakselis et al., 2017). Additionally, with the help of accessory helicases like RTEL1, the replicative helicase seems to be able to bypass covalent DNA-protein cross-links (DPCs) on both the leading and lagging strands, thereby enabling it to resume replication downstream of the DPC (Sparks et al., 2019). However, it is still unclear whether such bypass mechanisms can be employed to overcome larger complexes such as RNAP.

If there is a replication block that prevents skipping, repriming and RNAP removal, the replication fork can be stabilized by a process known as fork reversal. Fork reversal is a process where two nascent strands are joined to form a "chicken foot" structure, which stabilizes and protects the stalled replication fork from nuclease attacks (Quinet et al., 2017). This allows for more time to remove the replication block. Recently, it was shown that transcriptionally challenged replication forks can go through a cycle of cleavage and re-ligation (Chappidi et al., 2020). This enables RNAP to move beyond the replication fork and restart DNA replication. This mechanism involves various proteins and is a multistep process. First, reversed forks at head-on TRC sites are remodeled back to the standard three-way fork configuration by RECQL. Next, RAD51 is removed by RECQL5 to inhibit further fork reversal, allowing for MUS81/EME1-mediated cleavage of the fork. The resulting relief of torsional stress facilitates the continuation of transcription beyond the site of conflict for the stalled RNAP complex. Consequently, if transcription is inhibited after the generation of a TRC, replication restart cannot occur as the RNAP complex blocks it. Lastly, RAD52 and DNA ligase IV collaborate to catalyze fork re-ligation and thereby aid in replication restart (Chappidi et al., 2020). It is noteworthy that the mechanism of fork cleavage/re-ligation is currently the only one

discovered that prioritizes the transcribing RNAP over the replisome, enabling it to finish transcription rather than expelling it from the chromatin.
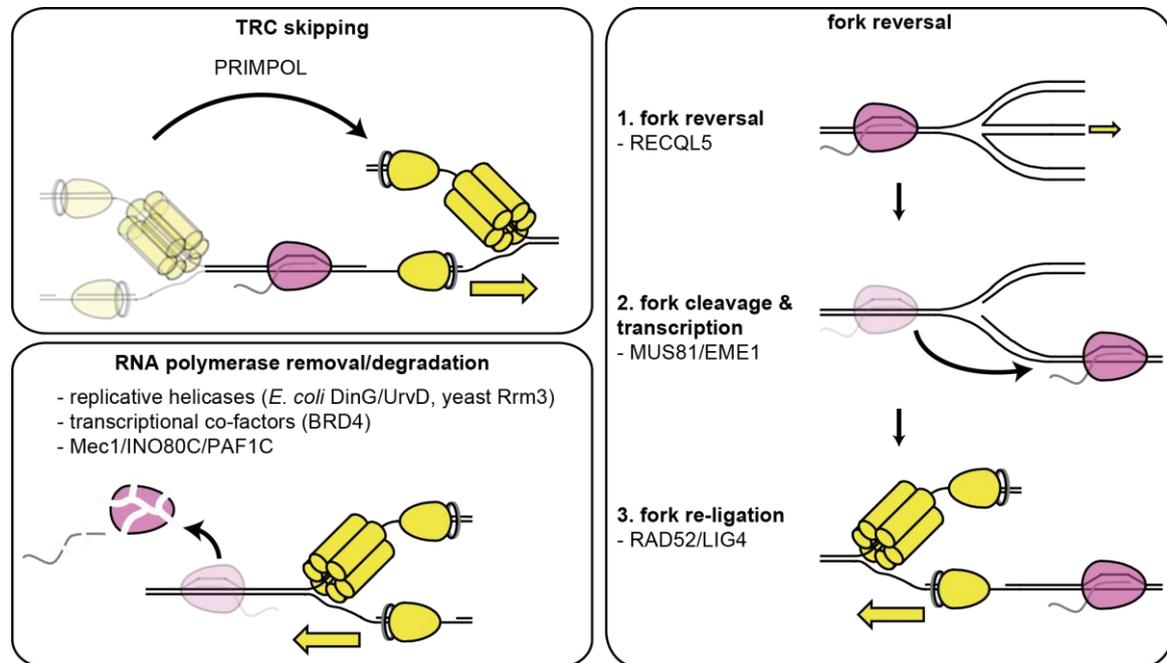
Figure 2: Illustration of transcription-replication conflict resolution pathways. Replication forks are colored yellow and transcription machineries are depicted in magenta.

1.1.5 Why is studying TRCs difficult?

Studying TRCs, especially *in vivo* in human cells, has proven challenging. First, they are transient and likely short-lived events. Second, in higher eukaryotes, the direction of replication at a given gene is difficult to predict. Origins fire stochastically in terms of timing and efficiency, meaning there is a distinct replication landscape in every cell and cell cycle (Czajkowsky et al., 2008; Hyrien, 2015). Furthermore, even if replication forks stall, the excess of origins allows completion of DNA synthesis by the firing of dormant origins from the opposite direction (Ekundayo & Bleichert, 2019), providing a second chance for DNA replication to overcome the obstacle and complete DNA synthesis. Third, RNA polymerases can exist in different functional states during a transcription cycle, such as promoter-proximal pausing, active elongation, backtracking or stalling/arresting at DNA lesion or DNA secondary structures such as R-loops. The

precise outcome of a replication fork collision with such distinct transcriptional complexes is not well understood and an area of active research (Kumar & Remus, 2023). Together, these factors make it very challenging to predict when and where transcription and replication collide in eukaryotic cells and what the consequences are (**Figure 3**).
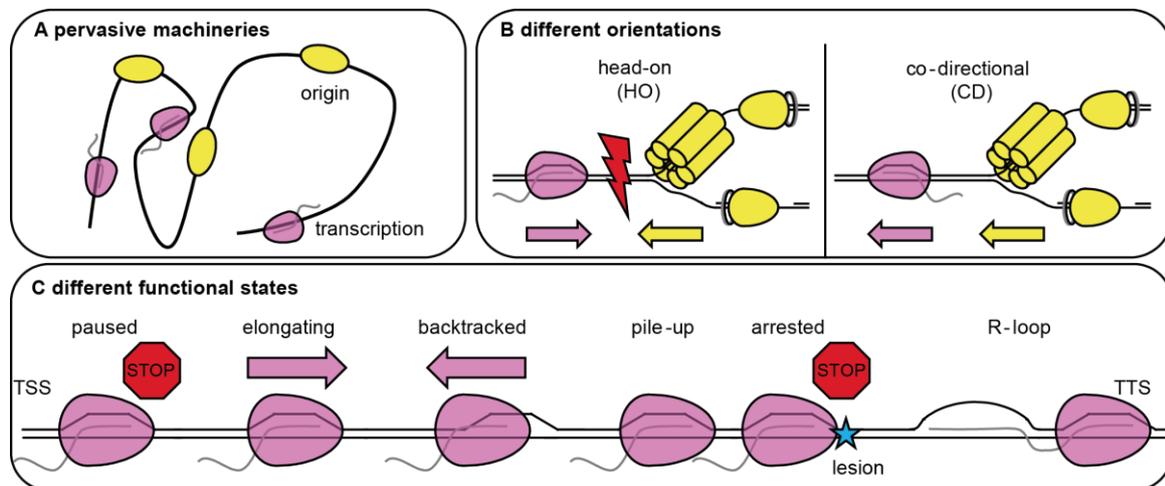
Figure 3: Illustration of different challenges complicating the investigation of TRCs *in vivo*. Replication forks or origins are colored yellow and transcription machineries are depicted in magenta **A**, Pervasiveness of both transcription and replication machineries that can initiate at multiple chromosomal locations in eukaryotic cells. **B**, TRCs can occur in HO or CD orientation **C**, Replication forks can encounter different functional states of RNA polymerases during one transcription cycle. TSS: Transcription Start Site, TTS: Transcription Termination Site.

In order to study the mechanisms and consequences of transcription-replication conflicts, a number of strategies have been used. One approach is to use artificial reporters or plasmid constructs to study defined TRCs. Prokaryotes typically have only one circular chromosome with a single origin of replication, making them ideal for studying different TRC orientations. By inserting inducible reporter or resistance genes into the genome, differences between HO or CD collisions can be specifically detected. For instance, introducing the inducible lacZ and luxABCDE reporter genes on one arm of the *B. subtilis* chromosome revealed that only HO conflicts led to the widespread formation of R-loops (Lang et al., 2017). If the R-loops are not resolved by the RNase HIII enzyme, replication blocks occur, leading

to an increase in genomic instability. These findings suggest that RNA:DNA hybrids at the conflict site partially contribute to the instability resulting from TRCs (Lang et al., 2017). Another approach to study TRCs was to use centromeric plasmids in budding yeast. Using a plasmid containing two direct repeats of a 0.6-kb internal fragment of the *LEU2* gene under an inducible promoter (Prado & Aguilera, 2005), a selectable functional copy of *LEU2* could only be reconstituted upon recombination of the two repeat sequences. This construct was placed next to an early-firing replication origin in either HO or CD orientation. Interestingly, induction of transcription led to recombination exclusively in the HO orientation (Prado & Aguilera, 2005).

The detailed knowledge of replication origins and their activation timing in budding yeast makes yeast an excellent model system for studying engineered chromosomal CD- or HO-TRCs. For instance, a galactose-inducible promoter controlling the LEU2 gene was inserted in HO orientation to the early-firing ARS315 origin (García-Rubio et al., 2018). The resulting chromosomal TRC system showed accumulation of transcription-induced R-loops and phosphorylation of histone H2A, a DNA damage marker, on the reporter gene. These findings indicate that HO transcription can cause DNA breaks in this system. Another study inserted the tetracycline-inducible LYS2 reporter gene in the opposite or same orientation to a highly efficient origin on yeast chromosome III (N. Kim et al., 2007). Consistently, it was shown that transcription of the reporter gene resulted in a large mutational spectrum of the underlying DNA sequence, including complex deletions, insertions as well as -2 frameshift mutations. Crucially, the occurrence of such mutations was higher in the HO orientation, suggesting that orientation matters for the functional impact of TRCs on genome integrity. In a different approach, the two bacterial operator arrays lacO and tetO were introduced downstream of the efficient ARS413 origin of replication (Tsirkas et al., 2022). Additionally, the GAL10 gene was introduced as a transcriptional unit in either HO or CD orientation. Then, fluorescent tags were used to visualize and track the movement of RNA polymerase and the replication machinery in real-time. Intriguingly, when the replication fork was ~20 kb upstream of GAL10, it caused partial transcriptional repression. This protective measure could be the reason that neither HO nor CD transcription caused replication fork slowing in this system. However, disrupting

topoisomerase 1 function or preventing dissociation of RNAPII from DNA led to decreased replisome progression when inducing transcription in HO or CD orientation (Tsirkas et al., 2022).

To study defined TRCs in human cells, episomal constructs containing inducible transcriptional units were used (Hamperl et al., 2017). To ensure unidirectional replication, the Epstein-Barr virus replication origin OriP was combined with expression and binding of the viral EBNA1 protein to the origin, thereby creating a replication fork barrier and ensuring cell-cycle regulated, unidirectional replication during S phase (Gahn & Schildkraut, 1989; Kirchmaier & Sugden, 1995; Dhar et al., 2001). Then, either an R-loop forming or a control gene was introduced in HO or CD orientation. Intriguingly, this approach revealed the occurrence of DNA breaks in both HO and CD orientation only when R-loops were present. However, their processing appeared to be influenced by the orientation, leading to the activation of different DNA damage responses. Specifically, the R-loop HO construct activated ATR kinase, while the R-loop CD construct activated ATM kinase (Hamperl et al., 2017).

Another approach to study TRCs is to investigate expected hotspots at specific genomic loci. These include rRNA genes, tRNA genes, common fragile sites and early replication fragile sites (Lalonde et al., 2021). For example, CD collisions were shown to interfere with replication at highly transcribed rRNA genes in *B. subtilis* and the helicase loader proteins DnaB and DnaD are necessary to restart replication at these loci (Merrikh et al., 2012). Also in yeast, highly transcribed RNAPII genes and tRNA genes slowed replication forks and helicases were essential for fork progression (Azvolinsky et al., 2009; Tran et al., 2017). In mammalian cells, TRCs have been proposed to have a potential role in certain genomic locations that exhibit gaps or breaks on metaphase chromosomes after aphidicolin treatment, known as common fragile sites (CFSs) (Glover et al., 1984; Hecht & Glover, 1984). Properties of CFSs are late replication timing, sparse replication origins, and persistence of abnormal mitotic structures like micronuclei or ultrafine anaphase bridges (Oestergaard & Lisby, 2017). Notably, many CFSs co-localize with exceptionally large genes with long intronic sequences. The resulting long transcription cycle makes them especially vulnerable to encountering a replication fork, resulting in incomplete replication, chromosomal instability, and

deletions (Helmrich et al., 2006; Pentzold et al., 2018; Wilson et al., 2015). However, an examination of nascent transcription, replication origin positioning, and fork directionality at the genome-wide level demonstrated that the fragility of CFSs can be accounted for by the low density of replication origins alone. In this study, inhibition of transcription in S phase was unable to alleviate aphidicolin-induced CFS fragility (Brison et al., 2019), suggesting that the fragility of these loci is not dependent on active transcription. Consequently, the role of TRCs in CFS stability has been controversial and more studies are necessary in the future to unravel the detailed mechanisms how these sites become fragile in the genome.

Another set of loci proposed to be a hotspot of TRCs are early replication fragile sites (ERFs) (Barlow et al., 2013). The properties of these fragile genomic regions are an enrichment of repetitive sequences, high CpG density, and proximity to early replication origins and highly expressed genes. Importantly, fragility is increasing with higher transcriptional activity, indicating TRCs as a contributing cause (Barlow et al., 2013). Interestingly, loss of the transcription elongation factor RECQL5 and the resulting transcription stress such as RNAP backtracking, stalling or arrest, led to a replication-dependent increase of chromosomal DNA breaks that were particularly enriched in the transcribed regions of genes (Saponaro et al., 2014), supporting the view that endogenous TRCs can be an underlying cause for genomic instability *in vivo*.

In this regard, the transcription-replication research field is very limited in terms of reliable methods to study TRCs, especially endogenous TRCs *in vivo*. Currently, proximity ligation assay (PLA) using the antibodies against the replication fork component PCNA and RNAPII (phospho-Ser2) as a mark of transcription is one of the most employed methods to detect differences in TRC levels across conditions (Bayona-Feliu et al., 2021; Hamperl et al., 2017; Landsverk et al., 2020; Papadopoulos et al., 2022; Sanchez et al., 2020). However, the assay is not able to identify the exact location of the TRCs at a molecular level or provide additional mechanistic insights. Considering this limitation, many studies have focused on interrupting the coordination of transcription or replication and measuring the genomic consequences, which indirectly implied a contribution of TRCs. For example, it was demonstrated that exposing cells to HU led to the transcriptional activation of a group of genes that are particularly susceptible to DNA breaks due

to the destabilization of replication forks that come into contact with transcriptional complexes (Hoffman et al., 2015). Similarly, short-term suppression of cyclin-dependent protein kinase 9 (CDK9) results in the occurrence of RNAPII stalling and immediate accumulation of recombination repair factors at potential TRC sites (Shao et al., 2020). Or, as mentioned above, estrogen stimulation of the breast epithelial cancer cell line MCF7 results in dramatic changes of the transcriptional landscape, increasing co-transcriptional R-loop levels and as a potential result unscheduled TRCs (Stork et al., 2016).

1.2 Proximity labeling

The most widely used methods for studying protein-protein interactions in chromatin are affinity purification or immunoprecipitation followed by mass spectrometry (AP-MS/IP-MS). These methods have limitations, such as the loss of transient interactors during the purification process and the inability to study interactions in the native context of living cells.

Proximity labeling followed by mass spectrometry analysis is an alternative method that overcomes these limitations. In this method, a covalent biotin tag is introduced to proteins in the vicinity of a target in living cells. Currently, three major enzymes are used for proximity labeling: biotin ligase, horse radish peroxidase, and engineered ascorbate peroxidase. The biotinylated proteins are then extracted and analyzed by mass spectrometry.

The advantages of proximity labeling over conventional methods include the ability to analyze interactions in the native context of living cells, the detection of transient interactions and low abundance proteins. Furthermore, the covalent binding of biotin allows for stringent washing conditions during the pull-down, effectively reducing background noise in mass spectrometry analysis (Ummethum & Hamperl, 2020). Additionally, proximity labeling enables the study of insoluble cell compartments such as the nuclear matrix, nucleoli, and other nuclear structures, which are difficult to study with conventional methods (Brecht et al., 2020; Kochanova et al., 2020; Ummethum & Hamperl, 2020; Vishnoi et al., 2020).

1.2.1 Biotin Ligase (BioID)

The enzyme BirA biotin ligase found in *Escherichia coli* converts biotin and ATP into biotinoyl-5′-adenylate (bioAMP) (Barker & Campbell, 1981a, 1981b; Eisenberg et al., 1982). One of the BirA-bioAMP complex's important functions is to target the lysine residue in the biotin carboxyl carrier protein (BCCP) subunit of acetyl coenzyme A (acetyl-CoA) carboxylase, which is the only biotinylation site in *E. coli*. To leverage this highly specific reaction, a short synthetic peptide substrate was created (Beckett et al., 1999; Schatz, 1993). This biotin acceptor peptide (BAP) can be fused with proteins of interest (POI) and co-expressed with BirA, which then recognizes and conjugates biotin on BAP's lysine residue (Smith et al., 1998). Subsequently, streptavidin pull-down can efficiently purify the newly biotinylated protein (de Boer et al., 2003). Alternatively, the system was used to explore protein-protein interactions by fusing BirA and BAP with two interacting proteins (Fernández-Suárez et al., 2008). However, prior knowledge of the interacting protein pairs is required.

By mutating *E. coli* BirA* (R118G), bioAMP does not stay bound to the BirA, which allows bioAMP to diffuse from the enzyme and readily react with lysine residues of any protein (Kwon et al., 2000; Kwon & Beckett, 2000). In this way, an unbiased approach was created. Intriguingly, proximity-dependent biotinylation efficiency was observed in *in vitro* experiments, indicating that substrates closer to BirA∗ were more readily biotinylated (Choi-Rhee et al., 2004; Cronan, 2005). To promiscuously biotinylate proteins in mammalian cells, a codon-optimized BirA∗ was created and fused with the protein of interest (Roux et al., 2012). This method, termed BioID, makes it possible to identify the proximal proteome of virtually any protein of interest.

As a first development, Kim et al. used the *Aquifex aeolicus* biotin ligase instead of *E. coli* biotin ligase, decreasing the size of the BioID moiety from 35 to 28 kDa (D. I. Kim et al., 2016). Subsequently, a mutated *E. coli* biotin ligase called TurboID was developed, which reduced the labeling time from 6 hours to 10 minutes (Branon et al., 2018). Meanwhile, a mutated and truncated biotin ligase was derived from *Bacillus subtilis* (BASU), which demonstrated efficient labeling for LC-MS/MS analysis within 30 minutes (Ramanathan et al., 2018). However, the improved activity of BASU was demonstrated only in a specific context where BirA∗

is fused to a small peptide recognizing RNA motifs. Moreover, during the development of TurboID, BASU showed kinetics comparable to BioID and BioID2 (Branon et al., 2018).

1.2.2 Engineered Ascorbate Peroxidase (APEX)

Peroxidases exhibit versatility in biochemistry applications through their ability to oxidize different chromogenic substrates in the presence of H_2O_2. One such example is the use of horseradish peroxidase (HRP), which can enhance electron microscopy contrast by polymerizing 3,3′-diaminobenzidine after treatment with OsO4, as demonstrated in studies by Graham and Karnovsky and Li et al. (Graham & Karnovsky, 1966; J. Li et al., 2010). In addition to their ability to oxidize chromogenic substrates in the presence of H_2O_2, peroxidases are also capable of catalyzing the oxidation of phenol derivatives to phenoxyl radicals (Gross & Sizer, 1959). This chemical reaction forms the basis of the widely-used technique tyramide signal amplification for immunostaining (Mayer & Bendayan, 1997). Additionally, phenoxyl radicals can react with electron-rich amino acids, particularly tyrosine (>95%), as well as tryptophan and cysteine (together only ~ 2 %) (Gross & Sizer, 1959; Rhee et al., 2013; Udeshi et al., 2017). However, due to the short lifespan of these radicals (<1 ms), they can only react with amino acid residues that are in close proximity to the peroxidase (Mortensen & Skibsted, 1997).
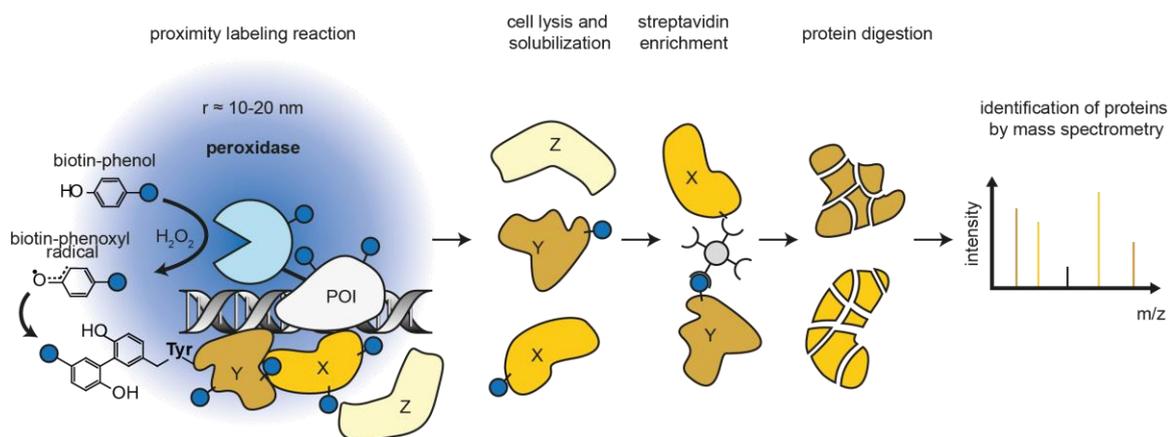
Figure 4: General workflow of a proximity labeling experiment. In this example, a peroxidase is fused to the protein of interest (POI) and expressed in cells. The proximity labeling reaction is initiated by addition of biotin-phenol and H_2O_2. After quenching, cells are lysed and solubilized, biotinylated proteins are enriched with

streptavidin, digested and subsequently analyzed by mass spectrometry. Adapted from Ummethum & Hamperl, 2020.

Initial biotin-based proximity labeling experiments involved the use of HRP and aryl azide-biotin as a substrate (Kotani et al., 2008). The major drawback of HRP is the lack of activity in the mammalian cytosol due to the absence of two crucial disulfide bridges that are unable to form in the reducing environment (Martell et al., 2012). To overcome this limitation, an engineered ascorbate peroxidase (APEX) derived from pea was introduced (Rhee et al., 2013). APEX is functional in all cell compartments and is capable of labeling surrounding proteins via incubation with H_2O_2 and biotin-phenol (**Figure 4**). Biotin-phenoxyl radicals primarily target tyrosine residues located on the surface of proteins. APEX, with a molecular weight of 28 kDa, has a smaller size compared to HRP, which has a molecular weight of 44 kDa. This smaller size reduces the likelihood of the fusion protein interfering with the native structure or function of the protein. However, the initial APEX version had low catalytic activity, prompting the development of a more efficient enzyme via directed protein evolution. This led to the creation of the A134P mutated version, named APEX2 (S. S. Lam et al., 2015).

1.2.3 General considerations for proximity labeling

Proximity labeling is a powerful tool for identifying potential interaction partners of a protein of interest. However, the relatively large size of the enzymes (27-28 kDa) may influence the function and localization of the bait protein (D. I. Kim & Roux, 2016; Roux, 2013; Roux et al., 2012). In addition, the labeling radius is not clear, particularly for biotin ligases. Reactive bioAMP may diffuse away from the biotin ligase, as suggested by Rhee et al. (2013). However, a BioID study on the nuclear pore complex found an effective labeling radius of ~10 nm (D. I. Kim et al., 2014). The labeling radius can be expanded by inserting a flexible linker in the fusion protein (D. I. Kim et al., 2014). Biotin-phenoxyl radicals produced by APEX2 have a very brief lifespan (less than 1ms), resulting in a gradual reduction of biotinylation levels when moving away from the peroxidase (Hung et al., 2016). Another difficulty is the strength of the biotin-streptavidin interaction, which does not allow for efficient elution of biotinylated proteins from the beads, leading to many false

negative detected proteins. Typically, on-bead digestion is utilized to avoid this issue, but false positives may arise due to cleaved peptides of non-biotinylated proteins that interact non-specifically with the beads. Moreover, the biotinylated peptides that still contain part of the streptavidin after cleavage have an unpredictable shift in their molecular weights, resulting in the loss of crucial peptides for subsequent analysis. Therefore, appropriate controls are mandatory to limit discovery of both false positive and false negative interaction partners. Crucially, similar expression levels of bait and control fusion proteins should be achieved. In addition, biotinylation does not necessarily reflect the strength of association between bait and target proteins. In fact, the efficiency of biotinylation depends on the accessibility and number of amino acid residues exposed on the surface, primarily lysine or tyrosine. This implies that proximity labeling studies may be affected by intrinsically disordered regions in proteins, which are susceptible to variations in pH, salt concentration, and post-translational modifications (Minde et al., 2020). This bias may arise due to the high prevalence of lysines in intrinsically disordered regions, which are targeted more often by BioID compared to tyrosines favored by APEX. Recently, there have been advancements in combining proximity labeling and protein-fragment complementation assays (PCA). A PCA involves fusing two proteins of interest (POIs) to two halves of a split reporter protein, such as an enzyme or fluorescent protein. Then, the reporter protein is reconstituted only upon interaction of the POIs. However, the exact dynamics of such a system are not fully understood, for example whether the reconstitution is reversible. So far, Split-BioID and split-APEX2 have been reported for proximity labeling (Han et al., 2019; Munter et al., 2017; Schopp et al., 2017; Xue et al., 2017). This approach is advantageous as biotinylation is dependent on the correct localization of both targeted factors, which significantly reduces the number of false positives (Munter et al., 2017). For transient protein interactions, this approach is specifically interesting as labeling occurs only at the right time and the right site when a protein complex is formed, or a biological process has been initiated. Additionally, splitting the reporter enzyme leads to smaller tags for the POI, potentially reducing the risk of affecting protein functionality.

1.3 The CGG binding protein 1 (CGGBP1) is a DNA-sequence-specific binding factor with multifaceted roles in transcription, replication and DNA repair

1.3.1 Discovery of CGGBP1

The CGG binding protein 1 (CGGBP1) was discovered as a CGG triplet repeat binding protein *in vitro* (Deissler et al., 1996, 1997; Richards et al., 1993; Yellan et al., 2021). The goal of the initial studies was to uncover proteins binding to the CGG repeat tract in the 5' untranslated region (5' UTR) of the FMR1 locus. The expansion of this $(CGG)_{30}$ repeat tract is the molecular basis of the human neurodegenerative disease Fragile X syndrome (FXS). In diseased cells, expansion of the CGG repeat tract has surpassed 200 repeats, leading to hypermethylation and subsequent transcriptional suppression of the FMR1 gene. Consequently, the encoded protein FMRP, a specific RNA-binding protein involved in various cellular functions like regulating dendritic mRNA transport and local protein synthesis at synapses (Feng et al., 1997; Fu et al., 1991; Pieretti et al., 1991) is lost (reviewed in Davis and Broadie, 2017). How and when this expansion occurs is still mostly unknown, but the underlying cause is most likely formation of DNA secondary structures of the CGG repeats (reviewed in Hayward & Usdin, 2021; X.-N. Zhao & Usdin, 2016).

Repetitive sequences are prone to form structures such as hairpins, cruciforms, triplexes (H-DNA) and G4s that can interfere with transcription and replication (reviewed in Mellor et al., 2022). Moreover, disease-length repetitive CGG-tracts are prone to form R-loops (Groh et al., 2014; Loomis et al., 2014). While the $(CGG)_{30}$ repeat in the FMR1 locus has been extensively studied, repeat tracts with n<10 are scattered across the genome and their impact on transcription and replication are largely unknown (Clark et al., 2006; Willems et al., 2014). Intriguingly, even short $(CGG)_7$ oligonucleotides are able to form stable G-quadruplex *in vitro* (Fry & Loeb, 1994; Kettani et al., 1995). Conversely, these short repeat tracts, as opposed to longer tracts, are not linked to high R-loop levels and transcriptional interference (Groh et al., 2014). The question remains if this is due to the inability of DNA secondary structures to form at short repeats *in vivo* or they are suppressed by DNA interacting proteins such as helicases. CGGBP1 is a

protein that binds to CGG repeats and thus could play a role in the prevention of DNA secondary structure formation.

1.3.2 Protein structure

CGGBP1 is a 20 kDa protein that contains a nuclear localization signal and a predicted C2H2-type Zn finger DNA-binding domain (Müller-Hartmann et al., 2000; Singh & Westermark, 2015). Electrophoretic mobility shift assays with CGG-repeat containing oligonucleotides confirmed strong affinity to CGG-repeat sequences *in vitro* (Deissler et al., 1996; Müller-Hartmann et al., 2000), which has also been verified *in vivo* at the CGG repeats in the 5' UTR of the FMR1 locus (Goracci et al., 2016). CGGBP1 is specific to vertebrates and conserved between all amniotes, but absent from amphibians. (Singh & Westermark, 2015). Many aspects of the CGGBP1 structure are still unclear because there is no published crystal structure, NMR data, or cryo-EM data. Two regions of the protein are predicted to be ordered by both Alphafold 2 and I-TASSER (Jumper et al., 2021; Roy et al., 2010; Yang & Zhang, 2015): the proposed C2H2 Zn finger domain (aa 38-95) and a C-terminal region between aa 95-167 predicted to form 3-4 alpha helices (**Figure 5A,C-D**). As determined by mutational analysis, the structured region at the C-terminus as well as aa 69-71 are essential for the ability of CGGBP1 to bind to CGG repeats (Müller-Hartmann et al., 2000). It has been suggested that CGGBP1 oligomerizes and the C-terminal structure could be a dimerization domain (Müller-Hartmann et al., 2000; Singh & Westermark, 2015). A dimer prediction via Alphafold 2 supports the suggestion of a dimerization domain specifically in the C-terminal domain (**Figure 5B**). CGGBP1 was initially thought to be a strict CGG triplet repeat binding protein, however ChIP-seq data suggest that CGGBP1 can bind to other DNA motifs *in vivo* such as the Alu transcription enhancer sequence (Agarwal et al., 2014). There are two more closely studied phosphorylation sites in CGGBP1: Y20 and S164 (**Figure 5D**). Phosphorylation of Y20 ensures proper nuclear localization (Agarwal et al., 2014). In addition, a genome-wide screen for phosphorylation substrates of the DNA damage-sensing kinase ATR revealed the serine at position 164 as a phosphorylation target (Matsuoka et al., 2007).
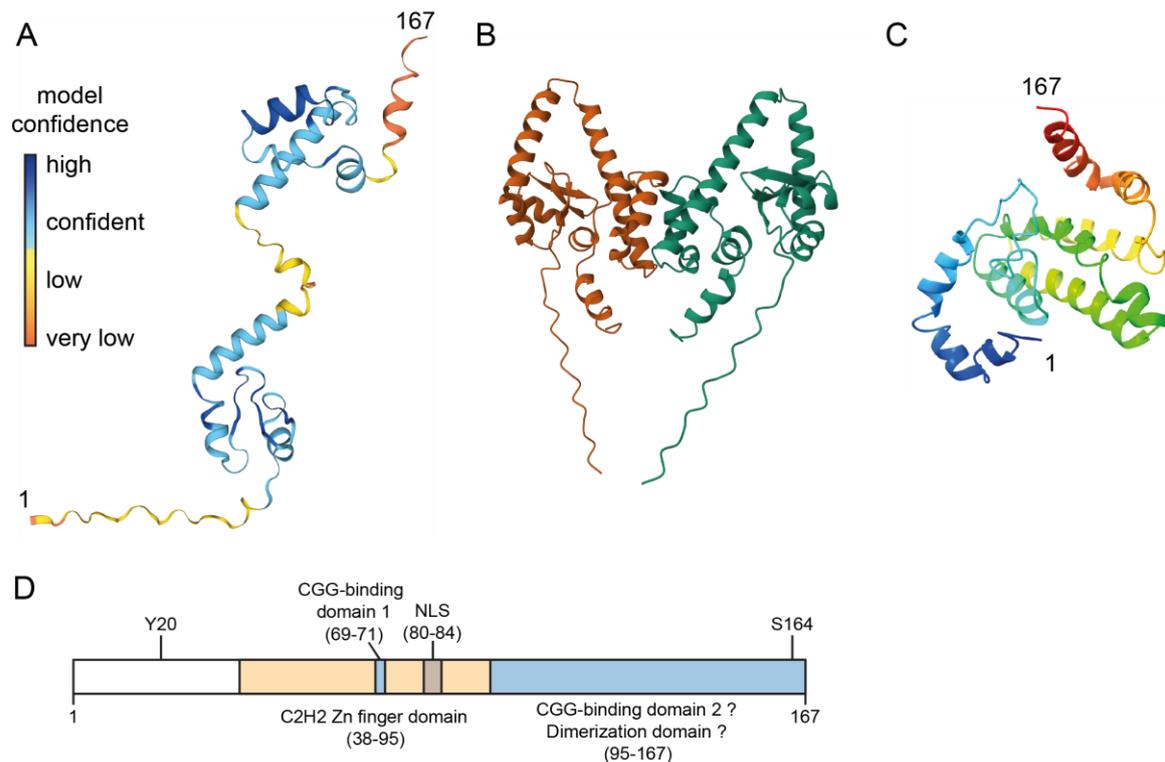
Figure 5: Predicted structure of CGGBP1. **A**, Alphafold 2 predicted structure of CGGBP1 with confidence score indicated by color. **B**, Alphafold 2 predicted structure of a CGGBP1 dimer with each monomer uniquely colored. **C**, I-TASSER structure prediction with rainbow coloring for better 3D visualization. **D**, Schematic of important protein domains and important amino acid residues of CGGBP1. NLS: Nuclear Localization Signal.

1.3.3 Interaction partners

The functional importance of CGGBP1 binding to CGG repeats *in vivo* is not well understood. To gain further insight into the cellular functions of this protein, a first step is to map direct interaction partners of this protein. However, most known interactions of CGGBP1 have only been mapped by high-throughput methods such as yeast two-hybrid or large scale affinity capture-MS (**Figure 6**, overview on BioGRID, Stark et al., 2006) . 38 of these 48 interactors are nuclear, suggesting a preferential role of this protein in the nucleus as expected. However, the described functions of the interaction partners are highly diverse, including DNA damage repair, transcriptional regulation as well as [2Fe–2S] cluster assembly. Importantly, most of these potential interactions have not been verified by complementary

approaches, so that the *bona fide* interaction network of CGGBP1 remains ambiguous.
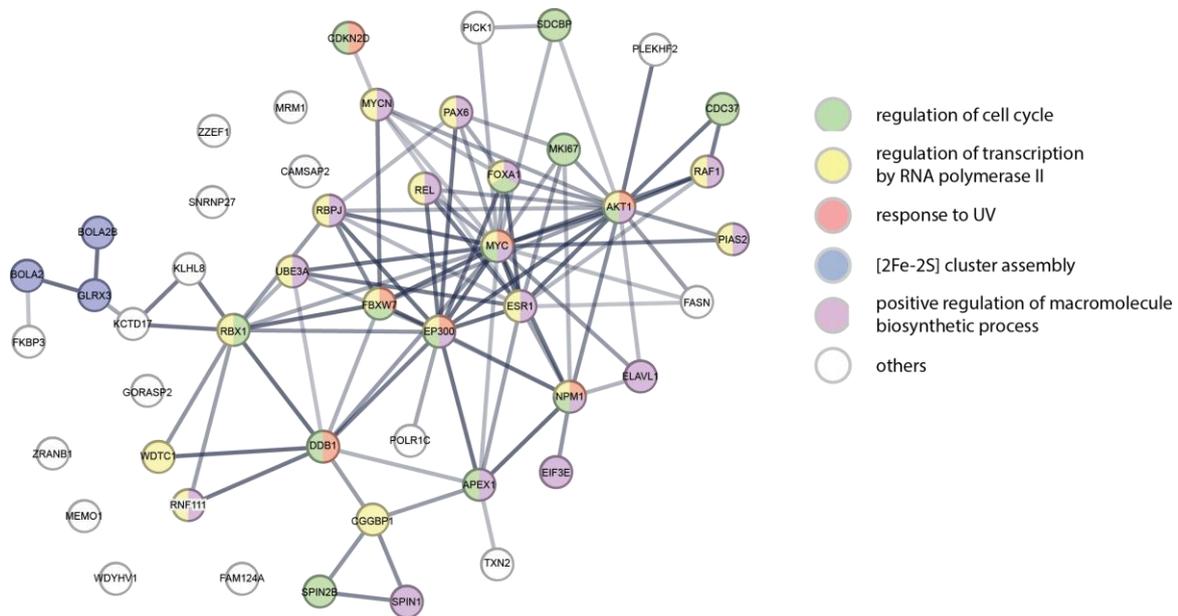
Figure 6: STRING network analysis of CGGBP1 interactors identified by high-throughput methods. The nodes are colored according to gene ontology (GO) biological processes.

However, there are few exceptions of high confidence CGGBP1 interacting factors such as CTCF, for which interaction has been demonstrated by co-immunoprecipitation (Patel et al., 2019). CCCTC-Binding Factor (CTCF) is a major chromatin architecture regulator and its presence at repeats is dependent on CGGBP1. Furthermore, there is evidence that CGGBP1 interacts with the heat shock proteins NFIX and HMGN1 (Singh et al., 2009). This study showed that CGGBP1 is responsible for assembling a transcriptional complex at short CGG repeats in the promoter of the HSF1 gene. In conditions of prolonged heat shock, NFIX utilizes CGGBP1 and HMGN1 to bind to this promoter and in turn affect their DNA binding activity (Singh et al., 2009). In conclusion, CGGBP1 appears to have multi-faceted roles based on the widespread functions of potential interaction partners, but this view is compromised by the lack of experimentally verified interaction partners.

1.3.4 DNA Methylation

A primary mechanism for transcriptional silencing is methylation of CpG dinucleotides (Schübeler, 2015). Interestingly, CGGBP1 binding to CGG repeats was shown to have increased affinity to unmethylated CGG repeats *in vitro*, suggesting that it may be involved in preventing CpG methylation at specific sequences and could therefore be used as a transcriptional regulatory mechanism (Deissler et al., 1996; Goracci et al., 2016). Indeed, CGGBP1-depleted cells show a global increase in CpG methylation levels measured by bisulfite sequencing (Agarwal et al., 2015). However, this increase is observed at regions that contain prior DNA methylation rather than unmethylated regions that acquire *de novo* CpG methylation. This finding does not support the notion that CGGBP1 protects unmethylated regions from CpG methylation *in vivo*. Nevertheless, CGGBP1 was depleted for four days in these experiments, which could be too short for a detectable increase of *de novo* DNA methylation.

1.3.5 Transcription

Initially, CGGBP1 was studied in the context of transcriptional regulation of the FMR1 locus and its implication in FXS. Compellingly, when CGGBP1 is overexpressed, binding at the FMR1 5' UTR is increased and leads to transcriptional repression of the FMR1 gene (Müller-Hartmann et al., 2000). This suggests that an overabundance of CGGBP1-bound repeats could exert a dominant negative influence on transcriptional elongation. In contrast, it appears that depleting CGGBP1 does not impact the transcriptional activity of FMR1 in active alleles of cell lines containing expanded, but not methylated FMR1 CGG alleles (Goracci et al., 2016). CGGBP1 has been shown to repress transcription of GAS1 and CDKN1, as knockdown leads to increased transcription (Singh et al., 2011). Intriguingly, CGGBP1 binds to the promoter regions of these genes and upon depletion, the transcriptionally repressive histone modification H3K9me3 is decreased. This is evidence for CGGBP1 being a direct negative regulator of RNA polymerase II (RNAPII) transcription at the gene where it is bound to. However, there is also evidence for CGGBP1 as a *trans*-regulator of RNAPII transcription. Specifically, CGGBP1 depletion in growth-stimulated cells inhibits RNAPII activity globally and indirectly through an increase in repressive Alu RNA (Agarwal et al., 2014). Additionally, it has been speculated that CGGBP1 influences RNAPI

transcription, as it binds to rRNA genes *in vitro* and *in situ*, but there is no data on rRNA transcription changes upon CGGBP1 depletion (Müller-Hartmann et al., 2000). Together, these differential modes of action on transcriptional regulation by CGGBP1 indicate that there may exist multiple locus-, context-, and cell type-specific mechanisms in place how this factor affects transcriptional output of a given gene.

1.3.6 Cell Cycle arrest and Genomic integrity

Depletion of CGGBP1 leads to a G0/G1 phase arrest in a multitude of cancer cell lines, such as U2OS and HCT116 (Singh et al., 2011). Cells accumulate in G0/G1 phase and the S phase population is decreased. In contrast, normal human fibroblasts (1064Sk) are arrested at the G2/M phase at the expense of a reduced S phase population (Singh & Westermark, 2011). Furthermore, knockdown of CGGBP1 in 1064Sk cells leads to DNA damage measured by increased γH2AX (phospho-S139) levels (Singh et al., 2014). The presence of γH2AX is an indicator of DNA double-strand breaks (Rogakou et al., 1998). As S164 in CGGBP1 has been identified as an ATR target, a role in DNA damage response is likely (**Figure 5D**, Matsuoka et al., 2007). Indeed, overexpression of the phospho-dead mutant CGGBP1-S164A in 1064Sk cells mimics the increased γH2AX levels in CGGBP1-depleted cells (Singh et al., 2014). A fraction of the γH2AX foci co-localizes with telomeres and CGGBP1-S164A overexpression reduces the binding of the telomere protective protein POT1 at telomeric repeats. Thus, an effect on telomere stability could explain the cell cycle defects in 1064Sk cells upon CGGBP1 depletion. However, it is unclear if telomere instability or general DNA damage-induced checkpoint activation is the main cause of the observed cell cycle arrest in the studied cancer cell lines.

2 Aims

2.1 Development of an unbiased approach to identify factors involved in transcription-replication conflicts

As described in the introduction, the transcription-replication conflict research field currently only has very few methods available to study the mechanisms of TRC resolution *in vivo*. Therefore, one goal of this thesis was to develop an unbiased approach to identify unknown factors involved in TRCs. For this, I used a screening approach based on proximity labeling under TRC-inducing conditions in MCF7 breast cancer cells.

2.2 Investigation of potential candidates in the context of transcription-replication conflicts

The second aim of this thesis was to further characterize the candidates found in the screening approach of the first aim. A series of functional assays was performed to answer the question whether and what role the identified protein factors play in transcription, replication, DNA damage, R-loop levels and if any effects can be directly linked to TRCs. As part of this characterization, the CGGBP1 protein was found as a strong candidate factor to have a role in preventing transcription-replication conflicts by counteracting the formation of secondary DNA structures at short CGG repeat tracts in RNAP II promoter sites.

3 Material and Methods

3.1 Material

Reagent	Supplier	Supplier cat. nr.
(+)-Sodium L-ascorbate	Santa Cruz Biotechnology	sc-215877
4′,6-Diamidino-2-phenylindole dihydrochloride (DAPI)	Sigma Aldrich	32670
5,6-Dichlorobenzimidazole riboside (DRB)	Santa Cruz Biotechnology	Sc-200581
5-Bromo-2'-Deoxy-Uridine (BrdU)	Sigma Aldrich	B5002-1G
5-Ethynyl-uridine (5-EU)	Jena Bioscience	CLK-N002-10
6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox)	Sigma Aldrich	238813-5G
Ampicillin	Fisher Bioreagents	BP1760-5
Amplex UltraRed Reagent	ThermoFisher Scientific	A36006
Azide-SS-biotin	BroadPharm	BP-22877
Biotin	Sigma Aldrich	B4051-100MG
Biotinyl Tyramide (biotin phenol)	Cayman Chemical	27997
BSA Fraction V	Sigma Aldrich	10735078001
Dimethyl sulfoxide (DMSO)	SERVA	20385.01
Dithiothreitol (DTT)	ThermoFisher Scientific	R0861
DMEM (phenol red free)	ThermoFisher Scientific	21063-029
DMEM (with phenol red)	ThermoFisher Scientific	41966-029
Doxycycline hydrochloride (DOX)	Sigma Aldrich	D3447-500MG
DPBS	ThermoFisher Scientific	14190-094
Duolink In Situ Wash Buffers	Sigma Aldrich	DUO82049
Dynabeads MyOne Streptavidin C1	ThermoFisher Scientific	65001
ECL reagent WB	ThermoFisher Scientific	34580

Reagent	Supplier	Supplier cat. nr.
EDTA	Roth	8043.2
EGTA	Santa Cruz Biotechnology	sc-3593B
Estrogen E2 (β-Estradiol)	Sigma Aldrich	E2758-250MG
Ethanol	Sigma Aldrich	1.00983.1000
FBS (Charcoal stripped)	Sigma Aldrich	F6765-100ML
FBS (Tet system approved)	Takara Bio USA	631106
Gibson Assembly Master Mix	New England Biolabs	E2611L
Glycerol	Fisher Bioreagents	BP229-1
Glycogen	ThermoFisher Scientific	AM9510
HCl	PanReac AppliChem	182109.1211
HEPES	Sigma Aldrich	H3375-1KG
Hydrogen peroxid solution 30%(w/w) (H_2O_2)	Sigma Aldrich	H1009-100ml
Hydroxyurea (HU)	Biomol	H9120.10
Hygromycin B-solution 20 mL (50 mg/mL) sterile	Carl Roth	CP12.2
Igepal CA-630 (chemically identical to NP-40)	Sigma Aldrich	I3021-100ML
iTaq SYBR Green Supermix	Biorad	1725121
LDS sample buffer 4x	ThermoFisher Scientific	2399549
Lipofectamine RNAiMAX Transfection Reagent	ThermoFisher Scientific	13778150
Magnesium chloride	Sigma Aldrich	M8266-100G
Methanol	Sigma Aldrich	1.06009.2500
MOPS	Merck	475922-100GM
Paraformaldehyde (PFA)	ThermoFisher Scientific	047377.9M
Pen/Strep/Glutamine 100x	ThermoFisher Scientific	10378-016
Phenol:Chloroform:Isoamyl Alcohol	ThermoFisher Scientific	15593-031
Pierce Protein G Magnetic Beads	ThermoFisher Scientific	88847

Reagent	Supplier	Supplier cat. nr.
polyvinylidene difluoride membrane	Merck	IPVH00010
Protease and Phosphatase Inhibitor Cocktail 100x	ThermoFisher Scientific	78446
Protease inhibitor cocktail 100x (Halt, EDTA-free)	ThermoFisher Scientific	78445
Puromycin	Sigma Aldrich	P9620-10ML
Pyridostatin	Cayman Chemical	18013
SDS running buffer 20x	ThermoFisher Scientific	NP0002
SiR-DNA	Tebu-bio	251SC007
Sodium chloride	Merck	106404
Sodium deoxycholate	Sigma Aldrich	D6750-100G
Sodium n-dodecyl sulfate 99% (dry wt.) (SDS)	Alfa Aesar	A11183
Sodium phosphate	Sigma Aldrich	71496-1KG
Sodium azide	Santa Cruz Biotechnology	sc-208393
Sucrose	Sigma Aldrich	S9378-500G
Tergitol (NP-40) solution 70%	Sigma Aldrich	NP40S-100ML
TransIT-LT1 Transfection Reagent	Mirus	MIR2300
TRIS	Merck	1.08382.2500
Triton X-100	Sigma Aldrich	X100-100ML
TRIzol	ThermoFisher Scientific	15596026
Trypsin-EDTA	ThermoFisher Scientific	25200-072
Tween20	Kraft	18014332
β-Mercaptoethanol	Sigma Aldrich	07604-100ml

Enzyme	Supplier	Supplier cat. nr.
DNase I	New England Biolabs	M0303S

EcoRI	New England Biolabs	R3101L
Proteinase K	SERVA Electrophoresis	33756
RNase H	New England Biolabs	M0297S
RNase A	ThermoFisher Scientific	EN0531

siRNA target	Supplier	Supplier cat. nr.
CGGBP1	Dharmacon	L-015703-00-0020
control	Dharmacon	D-001810-10-20
DDB1	Dharmacon	L-012890-00-0005
EMSY	Dharmacon	L-004081-00-0005
HDAC2	Dharmacon	L-003495-02-0005
NUP155	Dharmacon	L-011967-01-0005
RBBP6	Dharmacon	L-006551-00-0005
SCAF8	Dharmacon	L-020498-02-0005
TASOR	Dharmacon	L-020306-02-0005
UHRF1	Dharmacon	L-006977-00-0005
WAC	Dharmacon	L-013325-00-0005
ZMAT2	Dharmacon	L-017090-02-0005
ZNF703	Dharmacon	L-014572-01-0005

Commercial plasmid	Supplier	Supplier cat. nr.
FKBP-V5-AP nes_pLX304	Addgene	120912
pcDNA3-APEX2-NES	Addgene	49386
pCMV(CAT)T7-SB100	Addgene	34879
pSBtet-Hyg	Addgene	60508

Antibody target	Conjugation	Supplier	Supplier cat. nr.	Application	Dilution
Biotin	HRP	ThermoFisher Scientific	43-4323	Western Blot	1:2000
Biotin	Dylight 488	ThermoFisher Scientific	22832	IF	1:1000
BrdU		BD Biosciences	347580	Flow cytometry	1:100
CDC45		Cell Signaling Technology	3673S	Western Blot	1:1000
CGGBP1		SantaCruz Biotechnology	sc-376482	Western Blot, PLA	1:300, 1:100
FANCD2		NovusBio	NB100-182	IF	1:1000
FLAG-tag		ThermoFisher Scientific	PA1-984B	Western Blot	1:500
GAPDH		Merck	CB1001-500UG	Western Blot	1:10000
HA-tag		SantaCruz Biotechnology	sc-7392	Western Blot, IF	1:10000, 1:1000
Histone H3		Abcam	ab1791	Western Blot	1:20000
Mouse IgG	HRP	ThermoFisher Scientific	G21040	Western Blot	1:10000
Mouse IgG	Alexa 488	ThermoFisher Scientific	A32723	IF, flow cytometry	1:1000, 1:500
Mouse IgG	Alexa 594	ThermoFisher Scientific	A11032	IF	1:1000
Mouse IgG	Alexa 647	ThermoFisher Scientific	A32728	IF	1:1000
PCNA		Abcam	ab18197	PLA	1:2000
Phospho RPA32 (Ser33)		Bethyl laboratories inc	A300-246A	PLA	1:2000
Rabbit IgG	HRP	ThermoFisher Scientific	G21234	Western Blot	1:10000
Rabbit IgG	Alexa 488	ThermoFisher Scientific	A11008	IF	1:1000
Rabbit IgG	Alexa 594	ThermoFisher Scientific	A11037	IF	1:1000
Rabbit IgG	Alexa 647	ThermoFisher Scientific	A21245	IF	1:1000
RNA:DNA hybrid (S9.6)		Merck	MABE1095	DRIP	7 µg per sample
RNAPII		Merck	05-623	IF	1:2000

Antibody target	Conjugation	Supplier	Supplier cat. nr.	Application	Dilution
RNAPII phospho Ser5		Abcam	ab5131	PLA	1:2000
RNAPII-phospho Ser2		Biozol Diagnostica	BLD-920204	PLA	1:2000
RNAPII-phospho Ser2		Abcam	ab5095	IF	1:2000
γH2AX		Cell Signaling Technology	9718S	IF	1:300

Kit	Supplier	Supplier cat. nr.
Bioanalyzer Agilent High Sensitivity DNA kit	Agilent Technologies	5067-4626
Click-iT EdU Cell Proliferation Kit for Imaging, Alexa Fluor 594 dye	ThermoFisher Scientific	C10339
Click-iT Nascent RNA capture Kit	ThermoFisher Scientific	C10365
Duolink In Situ PLA® Probe Anti-Mouse MINUS	Sigma Aldrich	duo92004-100RXN
Duolink In Situ PLA® Probe Anti-Rabbit PLUS	Sigma Aldrich	duo92002-100RXN
Duolink in situ detection reagents green (PLA amplification kit)	Sigma Aldrich	DUO92014-100RXN
Duolink in situ detection reagents FarRed (PLA amplification kit)	Sigma Aldrich	DUO92013-100RXN
IDT for Illumina – TruSeq RNA UD Indexes (96 Indexes, 96 samples)	Illumina	20022371
SuperScript III First-Strand Synthesis System	ThermoFisher Scientific	18080051
TruSeq Stranded Total RNA Library Prep Gold (48 Samples)	Illumina	20020598

3.2 Methods

3.2.1 Cell culture

All cell lines used are summarized in **Table 1**. In general, either the breast epithelial adenocarcinoma-derived MCF7 (HTB-22) or osteosarcoma-derived U2OS (Tet-On, Clontech) cell lines were used.

Table 1: Cell lines

Cell line	Insert	Clonality	Expression
MCF7		Monoclonal	None (parental)
MCF7	Plasmid K181 via sleeping beauty transposase	Polyclonal	MCM2-FLAG-APEX2 Puromycin resistance Tetracycline-regulated transactivator DsRed2
MCF7	Plasmid K183 via sleeping beauty transposase	Polyclonal	APEX2-V5-NLS Puromycin resistance Tetracycline-regulated transactivator DsRed2
U2OS	Tet-On	Monoclonal	Tetracycline-regulated transactivator
U2OS	Plasmid pHU43 via stable transfection	Monoclonal	Tetracycline-regulated transactivator Transcription through $(CGG)_{10}$ Hygromycin resistance EBNA1

U2OS Tet-ON cells were cultured in DMEM (GIBCO) supplemented with 10% Tet-ON approved FBS, 4.5 g/L D-Glucose and L-Glutamine, 110 mg/L sodium pyruvate and penicillin/streptomycin in 5% CO_2 at 37°C. MCF7 cells were cultured in phenol-red free DMEM (GIBCO) supplemented with 10% FBS, 4.5 g/L D-Glucose and L-Glutamine, 25 mM HEPES and penicillin/streptomycin in 5% CO_2 at 37°. For S phase release and simultaneous estrogen stimulation, MCF7 cells were first hormone starved by addition of phenol-red free DMEM with 10% charcoal-stripped FBS (Sigma Aldrich) for 24 h, then G1 arrested by adding phenol-red free DMEM

without FBS for 24 h. For S phase release, MCF7 cells were cultured in phenol-red free DMEM with 10 % FBS and after 20-22 h, estrogen was added 30 min before harvesting. Between all medium changes, cells were washed 2 times with 1X PBS.

3.2.2 Plasmid cloning

For most human genes, e.g. MCM2, CDC45 and POLR2E, the cDNA was obtained by reverse transcriptase amplification of total RNA from human cells. Longer genes were synthesized. Standard techniques were used for cloning of plasmids. In short, Gibson Assembly was used to combine PCR fragments, followed by transformation in *E. coli* DH5α and selection with ampicillin. Colonies were cultured in LB medium with ampicillin and plasmids were isolated with silica columns. All plasmids were sequenced to confirm successful cloning.

3.2.3 Plasmid and siRNA transfections

All plasmid and siRNA transfections were carried out using TransIT-LT1 (Mirus) or Lipofectamine RNAiMAX (ThermoFisher) transfection reagents in accordance with the manufacturer's instructions. Plasmid transfections were performed by mixing at a ratio of 2 µg plasmid DNA per 6 µL transfection reagent, followed by incubation for 30 minutes and then adding the mixture dropwise to the cells. For siRNA transfections, a ratio of 50 nM siRNA per 2 µL transfection reagent was used, followed by incubation for 20 minutes before adding the mixture dropwise to the cells. After incubation at 37°C for 12-16 hours, the transfection-medium mix was replaced with fresh medium.

3.2.4 Generation of monoclonal U2OS cell line containing episomal system

Monoclonal U2OS Tet-ON cell lines were obtained by transfecting cells with the pHU43 (Insertion of KpnI and HindIII cut Q-block fragment containing $(CGG)_{10}$ into pSH36 from Hamperl et al., 2017. Cloned following Scior et al., 2011) vector and then subjecting them to selection with 200 µg/ml hygromycin B for a period of 2-3 weeks. After the selection process, surviving cells were collected and sorted into 96-well plates using flow cytometry (BD FACSMelody). The cells were then cultured and expanded for 5-6 weeks while being subjected to selection with 200 µg/ml hygromycin.

3.2.5 Genomic integration of APEX2 constructs with sleeping beauty transposase

MCF7 cells were seeded in 6-wells (~500,000 cells) and cultured normally for one day. Then, using the transfection reagent TransIT-LT1 (Mirus), a mixture of 1 µg transposase containing vector pCMV(CAT)T7-SB100 and 1 µg of a vector carrying a puromycin resistance and the expression cassette of either MCM2-APEX2 or APEX2-NLS was transfected according to the manufacturer's manual. After one day, the transfection medium was replaced by fresh medium. 48h after transfection, selection was started by adding 1 µg/ml puromycin to the culture medium. The next day, the dosage was upped to 2 µg/ml and after 2 more days the surviving cells were frozen and stored in liquid nitrogen.

3.2.6 Immunofluorescence imaging

In general, cells were seeded into 96-well culture plates with square wells and clear polymer coverslip bottom (Ibidi). For fixation, cells were incubated with 4% PFA/PBS for 15 min, washed once with PBS, permeabilized with 0.5% Triton X-100 for 10 min and washed once more. For EdU or EU staining, cells were incubated with 10 µM EdU or 0.5 mM EU 30 min before fixation and a commercial kit was used to attach fluorescent dyes with a click reaction (ThermoFisher). Next, cells were subjected to a blocking step in 3% BSA/PBS at 25°C for 1 hour. Primary antibodies diluted in 3% BSA/PBS were added and left overnight at 4°C, followed by three washes with PBS. Subsequently, secondary antibodies diluted in 3% BSA/PBS and DAPI (5 µg/ml) were added and incubated for 90 minutes at 25°C, followed by three washes with PBS. The cells were imaged with a 40x objective using a spinning disc confocal microscope (Andor Dragonfly), and the resulting images were analyzed using a customized ImageJ script. This script enabled nuclei segmentation on the DAPI channel using StarDist (Schmidt et al., 2018). For creating nucleolar masks in the EU incorporation experiments, the threshhold_yen function from the scikit-image python package was employed (van der Walt et al., 2014; Yen et al., 1995). For imaging of chromatin bound proteins, cells were pre-extracted with CSK100 buffer (100 mM NaCl, 300 mM Sucrose, 3 mM MgCl2, 10 mM MOPS, 0.5% Triton X-100 in PBS) and washed once with PBS before fixation.

3.2.7 Proximity ligation assay (PLA)

Concordant to immunofluorescence imaging, cells were seeded into 96-well culture plates with square wells and clear polymer coverslip bottom (Ibidi). Before fixation, cells were pre-extracted with CSK100 buffer (100 mM NaCl, 300 mM Sucrose, 3 mM MgCl2, 10 mM MOPS, 0.5% Triton X-100 in PBS) for 3-5 mins, depending on confluency, and washed once with PBS. After that, the cells were fixed with 4% PFA/PBS for 15 minutes, washed once with PBS, and permeabilized with 0.5% Triton X-100 for 10 minutes before being washed again with PBS. To perform EdU staining, a commercial kit was used to attach fluorescent dyes with a click reaction (ThermoFisher). The cells were blocked with 5% BSA/PBS at 25°C for 1 hour and then incubated overnight at 4°C with primary antibodies diluted in 5% BSA/PBS. After incubation, the cells were washed twice with PBS, and the PLA was performed according to the manufacturer's protocol (Sigma) with slight modifications. Specifically, a reaction volume of 70 µl was used for the probe incubation, ligation and amplification steps. In brief, the cells were incubated with 1:10 diluted PLA probes in Duolink antibody diluent for 1 hour at 37°C, followed by two 5-minute washes in wash buffer A. To ligate the probes, ligase was added at a ratio of 1:70 to 1x ligation buffer, and the cells were incubated for 30 minutes at 37°C with this mix. Ligation was followed by two 5-minute washes in wash buffer A. For amplification, polymerase was added at a ratio of 1:140 to 1x amplification buffer, and the cells were incubated for 100 minutes at 37°C with this mix. Amplification was followed by two 10-minute washes in wash buffer B. To stain the DNA, DAPI (5 µg/ml) diluted in PBS was added and incubated for 90 minutes before being washed twice with PBS. Finally, the cells were imaged with a 40x objective using a spinning disc confocal microscope (Andor Dragonfly). The number of PLA foci was quantified using ImageJ after nuclei segmentation on the DAPI channel (see section 3.2.6).

3.2.8 Chromatin fractionation and Western Blotting

Chromatin fractionation was performed as described (Méndez and Stillman, 2000) with slight modifications. In detail, 1×10^6 cells were resuspended in buffer A (10 mM HEPES pH7.9, 10 mM KCL, 1.5 mM MgCl$_2$, 340 mM sucrose, 10% Glycerol, 1 mM DTT) and lysed by incubation with 0.1% Triton X-100 for 5 min on ice. The nuclei were obtained by low-speed centrifugation at 1300 g for 4 minutes at 4°C

and washed once with buffer A before being lysed in buffer B (3 mM EDTA, 0.2 mM EGTA, 1 mM DTT, 1x protease inhibitor cocktail) by incubation for 30min on ice. The insoluble chromatin was obtained by centrifugation at 1700 g for 4 minutes at 4°C and washed once with buffer B before resuspending the final chromatin pellet in 2x LDS sample buffer. To obtain whole cell extracts, cells were lysed with RIPA buffer (50 mM Tris HCl, pH 7.4, 150 mM NaCl, 0.5% deoxycholate, 0.1% sodium dodecyl sulfate, 1% NP-40, and 1x protease inhibitor cocktail). Both whole cell extracts and chromatin fractions were sonicated for 10 minutes using a Bioruptor UCD-200, alternating 2.5 seconds on and 2.5 seconds off, before being separated by electrophoresis and transferred onto polyvinylidene difluoride membranes. The membranes were blocked overnight at 4°C in 5% skimmed milk dissolved in 0.1% Tween/PBS. Primary antibodies were added and incubated overnight at 4°C in 3% BSA/PBS, followed by three washes with 0.1% Tween/PBS. HRP-linked secondary antibodies were added and incubated for 1 hour at 25°C before three washes and signal detection using ECL reagent.

3.2.9 Proximity Labeling for Immunofluorescence Imaging

For each condition, 10.000 MCF7 cells were seeded in one well of a 96-well plate (μ-plate 96 well, Ibidi). APEX2 constructs were expressed by adding doxycycline 48 h before fixation. The cells were incubated with 500 μM biotin phenol for 30 min before proximity labeling. To initiate the proximity labeling reaction, 1 mM H_2O_2 was added for 1 min. After quick aspiration, the reaction was stopped by washing 2 times with quencher solution (10mM sodium azide, 10 mM sodium ascorbate, 5 mM Trolox in DPBS). Cells were fixed in 4% paraformaldehyde for 15 min at room temperature, washed once in DPBS and permeabilized by incubation in 0.25% Triton X-100 in DPBS. After blocking with 3% BSA/DPBS for 2 h at room temperature, primary antibodies diluted in 3% BSA/DPBS were added overnight at 4°C. After 3 washes with DBPS, secondary antibodies diluted in 3% BSA/DPBS including 1:1000 SiR-DNA were added for 1 h. After 2 final washes with DBPS, cells were imaged at 40x on a spinning disc confocal microscope (Andor Dragonfly). Primary antibodies: V5 (1:1000), FLAG (1:1000). Secondary antibodies: Neutravidin-DyLight 488 (1:1000), anti-rabbit-Alexa 594 (1:1000), anti-mouse Alexa-594 (1:1000).

3.2.10 Proximity labeling and immunoprecipitation for mass spectrometry

For each condition and biological replicate, 3,000,000 cells were seeded in a 10 cm dish. The polyclonal MCF7 cell lines either contained tetracycline controlled expression cassettes for MCM2-FLAG-APEX2 or APEX2-V5-NLS. The cells were arrested in G1 and released back into S phase as described above. At the time of serum starvation, 100 or 200 ng/µl doxycycline were added until the end of the experiment for the APEX2-NLS and MCM2-APEX2 cell lines, respectively. To decrease transcription-associated replication stress, 100 µM DRB was added 2 h before harvesting. The cells were incubated with 500 µM biotin phenol for 1 h and with 100 nM estrogen for 30min before proximity labeling and harvest. To initiate the proximity labeling reaction, 1 mM H_2O_2 was added for 1 min. After quick aspiration, the reaction was stopped by washing 3 times with quencher solution (10mM sodium azide, 10 mM sodium ascorbate, 5 mM Trolox in DPBS). The cells were trypsinized, resuspended in DMEM medium and centrifuged at 600xg for 2 min. Pellet was resuspended in 1 ml quencher, transferred to protein low-bind tubes and kept on ice for the remaining steps. Chromatin fractionation was performed as described above. The final chromatin pellet was resuspended in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton X-100, 0.5% sodium doxycholate, 0.1% SDS in H_2O). Chromatin was solubilized by sonication with 5% duty factor, 200 cycles/burst, 140W peak incident power for 10 min in milliTUBEs (Covaris). Non-solubilized material was cleared by high-speed centrifugation at 15000xg for 15 min. To extract biotinylated proteins, 100 µl magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1, ThermoFisher) were pre-washed twice in RIPA buffer, added to the solubilized chromatin and rotated overnight at 4°C. Beads were washed once with SDS buffer (50 mM Tris-HCl pH 7.4, 2% SDS in H_2O), twice with washing buffer (50 mM Tris-HCl, 150 mM NaCl, 0.4% SDS, 1% IGEPAL-CA630, 1 mM EGTA, 1.5 mM $MgCl_2$ in H_2O) and once in washing buffer without detergents (50 mM Tris-HCl, 150 mM NaCl, 1 mM EGTA, 1.5 mM $MgCl_2$ in H_2O). Biotinylated proteins were eluted with 2 M NH_4OH containing 25 mM biotin at 95°C for 10 min.

3.2.11 Mass spectrometry sample preparation and measurement

This part was not done by me, but the experimental procedure is still provided for the sake of completeness. A total of 10 µg proteins was proteolysed using a filter-

aided sample preparation (FASP) protocol (Grosche et al., 2016; Wiśniewski et al., 2009). Protein samples were diluted to a final volume of 400 µL with UA buffer (8 M urea in 0.1 M Tris/HCl pH 8.5) and reduced for 30 min at room temperature using 1 µL 1 M dithiothreitol (DTT). Next, the samples were alkylated by adding 10 µL 300 mM iodacetamide (IAA) and incubated in the dark for 30 min at room temperature. Through adding 2 µL 1 M DTT, the unreacted IAA was quenched. The 30 kDa molecular weight cut-off centrifuge filters (Sartorius) were equilibrated by adding 200 µL UA buffer and subsequent centrifugation for 15 min at 15000 g. The samples were transferred to the filter, followed by centrifugation for 15 min at 15000 g. Next, three washing steps with 200 µL UA buffer and three with 100 µL 50 mM ABC (ammonium bicarbonate) succeeded. The filter was then placed into a new PCR-grade filtrate tube (Sartorius). Proteins on the filter were pre-digested by adding 40 µL ABC and 0.5 µL Lys-C and incubating for 2 h at room temperature. To fully digest the proteins, 10 µL ABC and 1 µg trypsin were added and after sealing the tubes with parafilm, the filter was incubated overnight at 37 °C. The peptides were centrifuged for 15 min at 15000 g. 20 µL ABC containing 5 % acetonitrile was added and another centrifugation step followed. The eluate was acidified by addition of 1.3 µL of trifluoroacetic acid to reach a pH of 2. The samples were stored at -20 °C until mass spectrometry analysis. LC-MS/MS mass spectrometry was performed on a Q Exactive HF mass spectrometer (Thermo Scientific). A rapid separation liquid chromatography (RSLC) system (Ultimate 3000, Dionex part of Thermo Scientific) was equipped with a nano trap column (300 µm inner diameter x 5 mm, packed with Acclaim PepMap100 C18, 5 µm particle size, 100 Å pore size, LC packings) and an analytical column (Acquity UPLC M-Class HSS T3 Column, C18, 75 µm inner diameter x 250 mm, 1.8 µm particle size, 100 Å pore size, Waters). 6 µL of sample was automatically loaded onto the nano trap column.

3.2.12 Cell cycle analysis

The cells were treated with 25 µM of 5-Bromo-2′-deoxyuridine (BrdU) for 30 min and then washed with PBS before harvesting by trypsinization. The cells were fixed with 70% ice-cold ethanol and permeabilized with 0.25% Triton X-100/PBS for 15 min on ice. To denature DNA to single strands, the cells were exposed to 2N HCl for 15 min at 25°C and then neutralized with 100 mM sodium borate pH 8.5. Next,

the cells were blocked in 1% BSA/PBS with 0.1% Tween-20 for 15 min at 25°C and incubated in primary BrdU antibody (1:100) for 3h in the dark at 25°C with gentle agitation. After washing with PBS, the cells were incubated with AlexaFluoro-488 secondary antibody for 1h in the dark at 25°C with gentle agitation, followed by three PBS washes. After transferring to flow cytometry tubes, the DNA was stained by adding propidium iodide (PI; 0.1 mg/mL; Sigma) and RNA was removed by RNase A (10 mg/mL) treatment for 45min. Finally, the cells were analyzed using a FACSMelody device (BD Bioscience) and cell cycle profiles were determined using FlowJo software.

3.2.13 Plasmid copy number analysis

Genomic DNA was obtained from 2-4 × 10^5 cells through trypsinization, washing in 1x PBS, and resuspension in TE buffer (10mM Tris-HCl at pH7.4, 2mM EDTA). IRN buffer (50mM Tris-HCl at pH 8, 20 mM EDTA, and 0.5 M NaCl), 0.5% SDS, and 10 µg Proteinase K were added, followed by incubation at 37°C for 1 hour. The DNA was then extracted with phenol/chloroform and treated with 20 µg RNase A at 37°C for 1 hour. Following chloroform extraction, the DNA was precipitated with EtOH/sodium acetate, washed with 70% EtOH, and resuspended in TE. EcoRI (NEB) restriction enzyme was used to digest the DNA overnight at 37°C. The plasmid copy number was determined by quantitative PCR on a Roche LightCycler 480 Instrument II using SYBR-Green Supermix (Biorad) and primer pairs listed in **Table S2**, which targeted either the oriP region of the plasmid or a region of the genomic beta-actin gene. The relative plasmid copy number was calculated as the ratio of the amount of 2x oriP to the amount of the genomic β-Actin.

3.2.14 Nascent RNA sequencing with EU

For this experiment the conditions were matched with the proximity labeling approach. For each condition and biological replicate, 250,000 cells were seeded in one well of a 6-well plate. The polyclonal MCF7 cell line contained a tetracycline-controlled expression cassette for MCM2-FLAG-APEX2. The cells were arrested in G1 and released back into S phase as described above. At the time of serum starvation, 200 ng/µl doxycycline was added until the end of the experiment. The cells were incubated with 500 µM biotin phenol for 1 h and with 0.5 mM 5-ethynyluridine (EU) 30 min before harvest. Additionally, either no estrogen or 100

nM estrogen was added 30 min before harvest. For harvesting, the medium was removed and 6-wells were washed once with PBS. After removing the PBS, 6-well plates were wrapped in aluminum foil and stored at -80°C until further processing. RNA was extracted by directly adding 500 µl TRIzol reagent (ThermoFisher) to 6-wells, resuspending until cells were completely lysed and detached and transferring the resulting mixture to DNA low-binding tubes. RNA was then isolated according to the manufacturer's protocol. Then, RNA was treated with RNase-free DNase I at 37°C for 30 min, followed by inactivation with 5 mM EDTA and incubation for 10 min at 75°C. Subsequently, the RNA concentration was measured with the Qubit RNA HS kit (ThermoFisher) on a Qubit fluorometer (ThermoFisher). For each sample, 1.6 µg RNA was used for subsequent steps. To covalently bind biotin to EU, the Click-iT nascent RNA capture kit (ThermoFisher) was used according to the manufacturer's manual with minor changes. In particular, 0.5 mM azide-SS-biotin (BroadPharm) was used in the reaction and not the biotin azide provided by the kit. Next, for each sample 50 µl of magnetic beads (Dynabeads MyOne Streptavidin C1, ThermoFisher) were pre-washed three times in 1x binding and washing buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.5% Tween-20), two times in solution A (0.1 M NaOH, 0.05 M NaCl) and two times in solution B (0.1 M NaCl) before resuspending in 100 µl 2x binding and washing buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2 M NaCL, 1% Tween-20). After incubating the RNA for 5 min at 70°C and briefly cooling on ice it was added to the beads and incubated for 30 min at 25°C while slowly rotating. Next, the beads were washed three times in 1x binding and washing buffer and once with RNase-free water. To Elute EU-RNA, 100 µl of 2% β-mercaptoethanol in RNase-free water was added and the beads were incubated for 60 min at 25°C while rotating. For total RNA sequencing, the TruSeq Stranded Total RNA Library Prep Gold kit (Illumina) was used following the manufacturer's manual. Final cDNA libraries were checked for quality and quantified with a Bioanalyzer (Agilent). The final pool of libraries was loaded in the flow cell at 5 nM concentration. Sequencing was carried out on a NovaSeq 6000 system (Illumina) with a 150-bp paired-end protocol, according to Illumina's instructions. Reads were trimmed with Trim Galore (https://github.com/FelixKrueger/TrimGalore). Next, Rsubread was used to align the reads to the hg38 build of the human genome (Liao et al., 2019). Subsequently,

FeatureCounts was used to count the reads per gene (Liao et al., 2014). Finally, voom was used for differential gene expression analysis (Law et al., 2014).

3.2.15 Reverse transcription-qPCR

TRIzol reagent (ThermoFisher) was used to isolate total RNA from harvested cells according to the manufacturer's protocol. Subsequently, the RNA concentration was measured with the Qubit RNA HS kit (ThermoFisher) on a Qubit fluorometer (ThermoFisher). Then, 1.5 µg RNA of each sample was treated with RNase-free DNase I at 37°C for 30 minutes, followed by inactivation with 5 mM EDTA and incubation for 10min at 75°C. Each sample was split into two to provide -RT controls. Reverse transcription was carried out using the SuperScript III first strand synthesis kit (ThermoFisher) following the manufacturer's instructions with random hexamers. To measure mRNA expression levels, equal amounts of cDNA were mixed with iTaq SYBR Green Supermix (Biorad) and subjected to analysis on a Roche LightCycler 480 Instrument II. The change in comparative threshold cycles was used to determine mRNA expression levels with primer pairs detailed in **Table S2**.

3.2.16 DNA:RNA immunoprecipitation (DRIP)

The methodology outlined in Ginno et al., 2012 was used to carry out the DRIP assay with slight modifications. In brief, DNA was extracted using the phenol/chloroform method, precipitated using of ethanol/sodium acetate, washed with 70% ethanol, and finally resuspended in in TE buffer (10mM Tris-HCl at pH7.4, 2mM EDTA). Sonication of DNA was conducted using a 10% duty factor, 200 cycles/burst, and 140W peak incident power for 4 min (Covaris). For RNase H-treated samples, 4 µg of DNA was subjected to RNase H treatment at 37°C overnight. The DNA was purified using phenol/chloroform extraction and EtOH/sodium acetate precipitation. Next, 4 µg of DNA was combined with 7 µg of S9.6 antibody in 1X binding buffer (10 mM $NaPO_4$ pH 7, 140 mM NaCl, 0.05% Triton X-100) and allowed to bind overnight at 4°C while rotating. Protein G magnetic beads were then added and incubated for 2 h rotating at 4°C. The beads were washed thrice with the binding buffer and elution was carried out using elution buffer (50 mM Tris pH 8, 10 mM EDTA, 0.5% SDS, Proteinase K) for 45 minutes at 55°C. Finally, DNA was retrieved using phenol/chloroform extraction and

EtOH/sodium acetate precipitation. The quantitative PCR analysis of immunoprecipitated DNA fragments was performed using SYBR-Green Supermix (Biorad) on a Roche LightCycler 480 Instrument II.

3.2.17 Mass spectrometry data analysis

To identify peptides and proteins, Proteome Discoverer 2.5 software (Thermo Fisher Scientific; version 2.5.0.400) was employed, utilizing a database search (Sequest HT search engine) against the Swissprot human database (The UniProt Consortium, 2023), with full tryptic specificity and allowance for up to two missed tryptic cleavage sites. Precursor mass tolerance was set at 10 ppm and fragment mass tolerance at 0.02 Da. The static modification of carbamidomethylation of Cys was included, while dynamic modifications comprised deamidation of Asn, Gln, and Arg, oxidation of Pro and Met, and a combination of Met loss with acetylation on the protein N-terminus. The Percolator was employed to validate peptide spectrum matches and peptides, accepting only the top-scoring hit for each spectrum and meeting the FDR <5% and posterior error probability <0.01 cut-off values. The list of proteins at the end complied to the strict parsimony principle. To quantify the proteins, abundance values were used for unique peptides and normalized based on the total peptide amount to account for sample loading errors. The protein abundances were then calculated by adding up the abundance values for admissible peptides. The final protein ratio was determined by taking the median abundance values from three replicate analyses. To determine the statistical significance of the ratio change, the T-test approach described in Navarro et al., 2014 was used. This approach is based on the assumption that there are expression changes in only a few proteins compared to the total number of proteins being quantified. The quantification variability of the non-changing "background" proteins was used to determine which proteins had a statistically significant change in expression.

3.2.18 ChIP-seq data analysis

For CGGBP1 ChIP-seq analysis, we used publicly available ENCODE data of K562 cells with the following identifiers: ENCSR763FNU and ENCSR334KTB (Luo et al., 2020; The ENCODE Project Consortium, 2012). For most analyses, tools from the web platform galaxy were used (Afgan et al., 2018, https://usegalaxy.eu).

To annotate CGGBP1 peaks, bed files were analyzed by ChIPseeker. For metagene plots of CGGBP1 ChIP-seq signal, bam files of the two replicates and the control were converted to bigWig with bamCoverage, heatmap values were calculated by deeptools computeMatrix and the plot created by deeptools plotHeatmap. For genome browser snapshots, these bigWig files were loaded into IGV (Robinson et al., 2011). For motif enrichment analysis, I used MEME-ChIP (Bailey et al., 2015) with 1st order Markov backround model, minimum motif width of 3, maximum motif width of 30 and any number of motif repetitions allowed. To analyze G4 hunter score, DRIP signal, GC% and GC skew of a set of ~2600 transcription start sites (TSS) that are proximal (<3kb away from promoter) to at least one CGGBP1 peak was selected. As controls, two sets of ~2600 randomly selected TSS were created. To obtain G4 quadruplex formation prediction scores, the G4 Hunter python script was used on FASTA files of 2kb windows around the TSS with –w 25 –s 0.0 parameters (https://github.com/AnimaTardeb/G4-hunter.git). This calculates a mean G4 Hunter score at the middle of each 25 bp window sliding by 1bp across the 2kb. DRIP signal was obtained from a published data set (Castillo-Guzman et al., 2020, GSM4478670). The K562 DRIP-seq signal bigWig file was converted from build hg19 to hg38 with the CrossMap tool (H. Zhao et al., 2014). Heatmap values and the plot for the TSS sets were created by computeMatrix and plotHeatmap, respectively. For GC% and GC skew, a 10 kb window around the TSS was split into 200 bp windows using a sliding window of 1bp with bedtools MakeWindowsBed. Nucleotide content was calculated with bedtools NucBed and GC skew was then calculated by measuring (G-C)/(G+C).

3.2.19 Software

For cloning, the design of constructs and confirmation of sequencing results was carried out with SnapGene software (www.snapgene.com). Planning of Gibson Assemblies was done with the online tool of NEB (https://nebuilder.neb.com/). GraphPad Prism (www.graphpad.com) and Adobe Illustrator (https://adobe.com/products/illustrator) were used for figures. For mass spectrometry data, network analysis and GO term enrichment was done with STRING (Szklarczyk et al., 2015).

3.2.20 Quantification and statistical analyses

Statistical parameters including the number of biological replicates, standard deviation and statistical significance are reported in the figures and the figure legends. Statistical significance is determined by the value of $p < 0.05$ by One-Way ANOVA test or two-tailed t-test, where appropriate. Where appropriate, I confirmed that sample sizes were large enough that any deviations from normality did not affect the statistical test results.

4 Results

4.1 Design and cloning of split-APEX2 constructs for proximity labeling at transcription-replication conflict sites

The first objective of my doctoral thesis was to develop a new tool to study transcription-replication conflicts and uncover in an unbiased way candidate proteins that are associated and physically bound to transcription-replication conflict sites in the genome. An ideal approach to achieve this is based on split-APEX2 with the simple idea to fuse one of the inactive split-APEX2 fragments to the transcription machinery and the other to the replication fork. As a result, APEX2 only reconstitutes if these two complexes are in close proximity, which would be the case at transcription-replication conflict sites. Thus, the reconstituted, active APEX2 would specifically label proteins in proximity to TRCs. After streptavidin enrichment, these TRC-specific proteins could then be identified by LC-MS/MS. However, at the beginning of this thesis project, split-APEX2 had only been used successfully when studying interaction of two members of the same protein complex (Han et al., 2019; Xue et al., 2017). In my approach, reconstitution of split-APEX2 would only work if two different complexes come close enough to each other. This uncertainty as well as using the proximity labeling enzyme APEX2 raised multiple questions: Are the two complexes close enough to each other during a TRC for split-APEX2? Does the fusion of split-APEX2 to the members of the two complexes influence their localization or function? Are flexible peptide linkers within the fusion proteins necessary to successfully reconstitute split-APEX2 and is it active upon reconstitution? If split-APEX2 reconstitutes, does it dissociate again or does it lock the complexes together? How can I avoid or navigate the notoriously high background generated by proximity labeling?

To answer most of these questions, I cloned multiple expression vectors carrying split-APEX2 fused to either members of the replication fork or transcription machinery (**Table S1**). These included MCM2, MCM3, MCM5, CDC45, GINS2, GINS3 of the CDC45-MCM-GINS (CMG) complex of the replicative helicase and POLR2B, POLR2E, CDC73 of the transcription machinery. Split-APEX2 fragments were fused to either the C- or N-terminus of the respective factors based on previously published studies with other fusion proteins or by finding surface

exposed ends in the structure. These vectors were then used to perform double transfection experiments and test the activity of APEX2. Additionally, I introduced glycine-serine linkers (Xue et al., 2017) or GR6, a long spacer helix from rat plectin (Griesenbeck et al., 2003; Steinböck & Wiche, 1999) to provide more flexibility and bridge a larger distance for successful split-APEX2 reconstitution. As a positive control, full length APEX2 was fused to the respective replication or transcription factors. To test APEX2 activity, MCF7 cells were transfected and after fusion protein expression incubated with either biotin phenol or the Amplex UltraRed reagent. Addition of H_2O_2 initiated APEX2 proximity labeling, which led to biotinylation or conversion of Amplex UltraRed to the fluorescent dye resorufin. Split-APEX2 activity was then assessed by immunofluorescence (IF) imaging in comparison to the full length APEX2 positive control. In addition, each fusion protein also expressed either a FLAG or V5 epitope tag that were additionally stained in separate IF channels to monitor transfection efficiency and the presence of both fusion proteins within the same cell.

Initial experiments suffered from low transfection efficiency, which resulted in a very low number of cells that were successfully transfected with both constructs (data not shown). Consequently, I stably transfected cells by selecting a polyclonal pool of cells that constitutively expressed different split-APEX2 fusion proteins and then transiently transfected the counterpart construct for APEX2 activity testing (**Figure 7A**). Independent of which factor was used or the linker type, no combination led to reconstituted, active APEX2 (**Figure 7B-C**). Surprisingly, expression of the two unfused split-APEX2 fragments AP and EX did not reconstitute to active APEX2 (**Figure 7C**). Based on this large series of negative results, and the multiple uncertainties with this approach raised above, it was clear that the necessary troubleshooting to obtain a positive combination of two constructs would not be feasible in time. However, inspired by the strong activity of full-length MCM2-APEX2 as the positive control in these experiments, I developed an alternative strategy to induce TRCs and subsequently proximity label transcriptionally challenged replication forks with full length MCM2-APEX2.
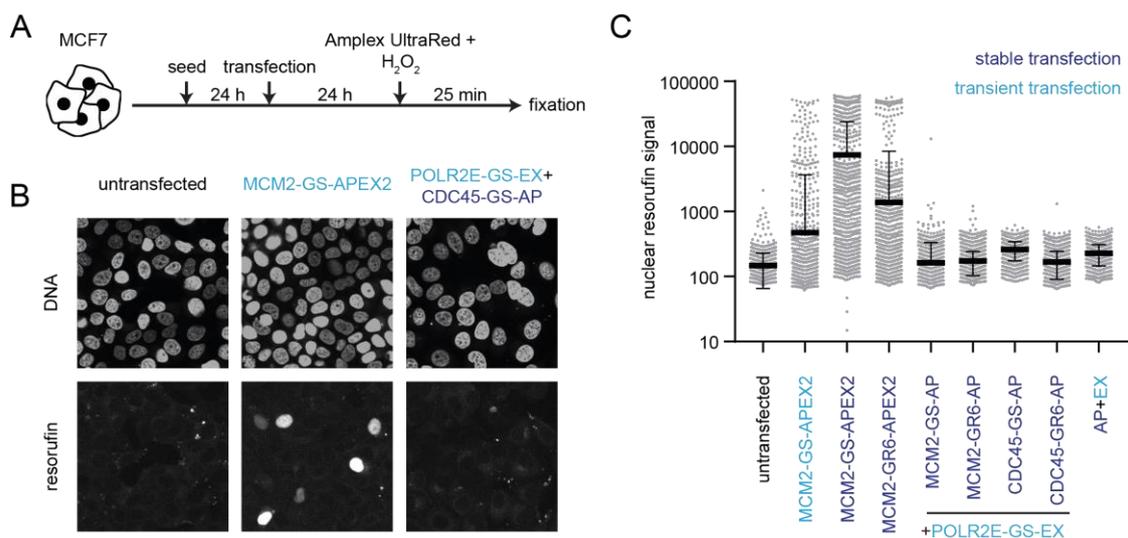
Figure 7: Split-APEX2 proximity labeling approach. **A**, Experimental layout for proximity labeling activity testing via transient transfection of either WT MCF7 or stably transfected cell lines with split-APEX2 constructs. **B**, Example IF images of split-APEX2 activity testing in MCF7 cells. **C**, Quantification of nuclear resorufin signal of IF imaging after proximity labeling with transient and stable transfections of TRC split-APEX2 combinations or MCM2-APEX2 as a positive control. Data is represented as mean ± standard deviation. GS Glycine Serine Linker, GR6 rat plectin spacer helix

4.2 Creating cell lines containing MCM2-APEX2 expression cassettes

To gain insight into the dynamic protein interactions at a transcriptionally challenged DNA replication fork, I generated MCF7 cell lines that allow inducible expression of APEX2 as a fusion protein with the MCM2 subunit of the MCM2-7 replicative helicase complex (MCM2-APEX2). Additionally, I engineered MCF7 cells to express APEX2 fused to a nuclear localization signal (APEX2-NLS) as a spatial nuclear proximity labeling control. The constructs driving the expression of the fusion proteins under a Tet-ON promoter were genomically integrated using the sleeping beauty transposase system (Kowarz et al., 2015, **Figure 8A**). MCM2 as well as other members of the MCM complex have been successfully deployed for proximity labeling with BioID in previous studies (Dubois et al., 2016). APEX2 as the proximity labeling enzyme was chosen due to its fast kinetics superior to

BioID. The polyclonal MCM2-APEX2 and APEX2-NLS cell lines contained ~35-70 and ~20-40 copies of the expression cassettes, respectively, that were randomly integrated in the genome (**Figure 8B**).
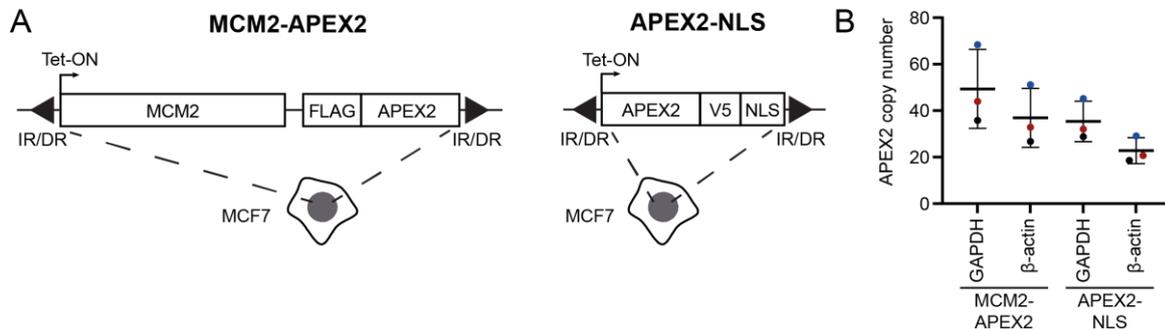
Figure 8: Genomic integration of MCM2-APEX2 and APEX2-NLS expression cassettes. **A**, Expression cassettes for MCM2-APEX2 and APEX2-NLS sleeping beauty transposon strategy. Tet-ON: DOX inducible promoter, IR/DR region: inverted repeats/direct repeats for sleeping beauty transposase. **B**, qPCR analysis of genomic insertion copy numbers calculated with the relative ratio 2xAPEX2/GAPDH or 2xAPEX2/β-actin to account for the diploid reference gene copy number. Different colors represent distinct primer pairs for APEX2 detection. Data is represented as mean of the primer pairs ± standard deviation.

Expression of FLAG and V5 tagged fusion proteins was not detectable below 50 ng/ml DOX and increased with rising DOX concentrations (**Figure 9A-B**). As expected, biotinylation was also dependent on DOX concentration (**Figure 9A, C-D**). Regardless of DOX concentrations, biotinylation levels were generally higher for APEX2-NLS than for MCM2-APEX2. Immunofluorescence staining confirmed the nuclear localization of biotinylation for both fusion proteins (**Figure 9A**). Nuclear fractionation after DOX induction revealed the presence of MCM2-APEX2 in the chromatin fraction (Chr.), suggesting chromatin incorporation of MCM2-APEX2 and therefore expression of a functional protein.
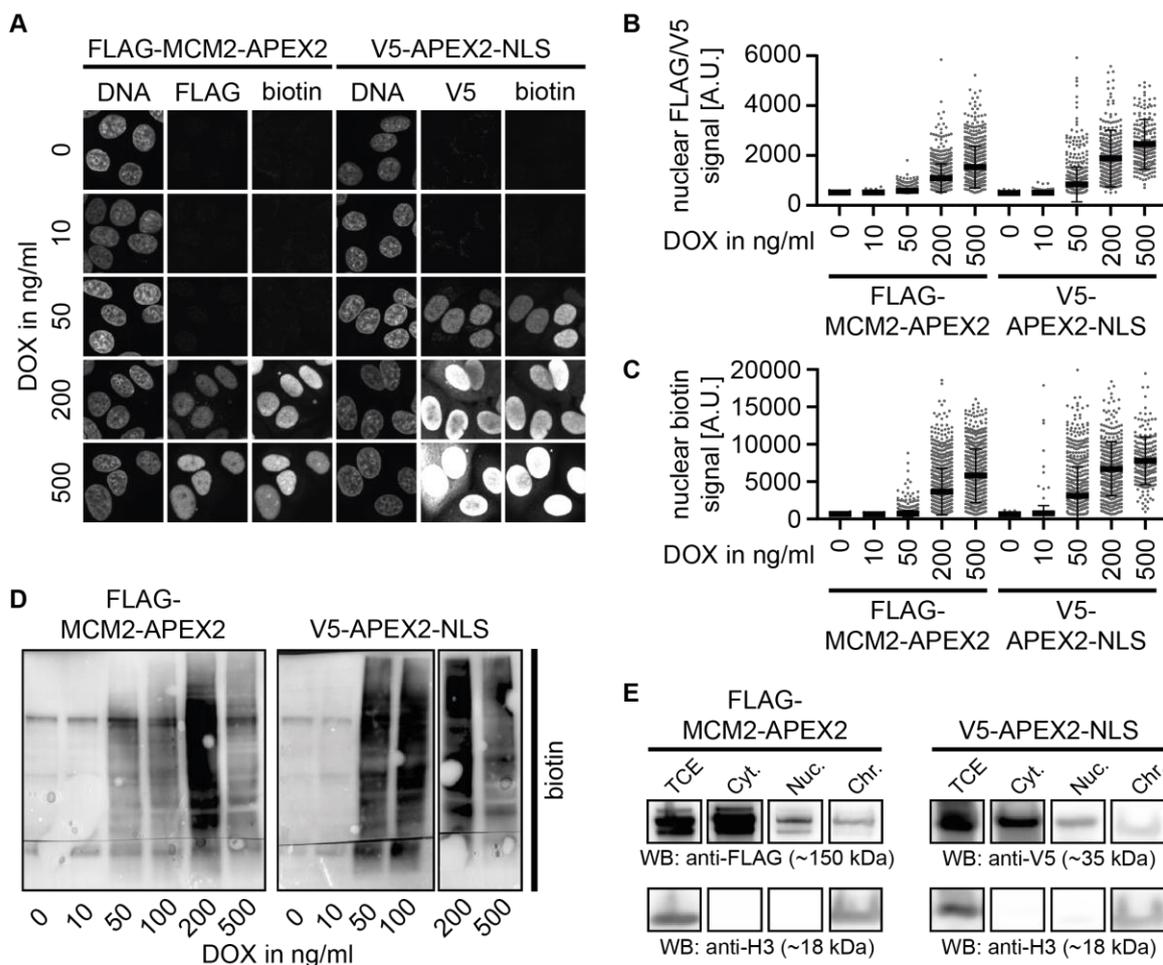
Figure 9: Expression and proximity labeling activity of MCM2-APEX2 and APEX2-NLS upon DOX treatment. **A**, Immunofluorescence images of proximity labeling in the MCM2-APEX2 and APEX2-NLS expressing MCF7 cell lines with varying DOX concentrations. DOX was added 48h before fixation. Cells were incubated with 0.5 mM biotin phenol for 60 min before proximity labeling and fixation. **B**, Quantification of nuclear biotin signal in immunofluorescence images. Total number of analysed nuclei is >200 for each condition. **C**, Quantification of nuclear FLAG/V5 signal in immunofluorescence images. Total number of analysed nuclei is >200 for each condition. **D**, Western Blot of proximity labeling with varying DOX concentrations. DOX was added 48h before sample collection. Cells were incubated with 0.5 mM biotin phenol for 60 min before proximity labeling and sample collection. **E**, Western Blot of nuclear fractionation in both cell lines after treatment with 200 ng/µL DOX for 48 h. TCE: total cell extract, Cyt: cytoplasmic, Nuc: nucleoplasmic, Chr: chromatin fraction.

In contrast, APEX2-NLS was mostly detected in the nucleoplasmic fraction (Nuc.) as expected. (**Figure 9E**). Of note, MCM2-APEX2 was also prominently present in the cytoplasmic and nucleoplasmic fractions (**Figure 9E**), which is likely due to the DOX-induced overexpression of the protein causing mislocalization. However, biotinylation during proximity labeling exclusively took place in the nucleus of cells (**Figure 9A**), indicating that the active fraction of the protein creating the biotinylation signal was correctly localized in the nuclear compartment. In summary, these results suggested that MCM2-APEX2 expression could be controlled in MCF7 cells and the resulting fusion protein was active.

To test the impact of DOX concentration on the final proximity labeling reaction and detection of biotinylated proteins, I performed a preliminary mass spectrometry analysis. For MCM2-APEX2, I tested 50 and 200 ng/ml DOX, for APEX2-NLS 25 and 200 ng/ml DOX (**Figure 10**). For both constructs, 200 ng/ml DOX resulted in a large number of biotinylated proteins visible in the Western Blot, while 25 or 50 ng/ml DOX led to barely visible bands (**Figure 10A**). Other members of the MCM2-7 complex served as an indicator for successful enrichment of proteins in proximity of MCM2. The MCM2-APEX2 200 ng/ml DOX sample did not enrich for many MCM2-7 complex members when compared to the APEX2-NLS 200 ng/ml DOX sample (**Figure 10B**). However, in comparison to low APEX2-NLS (25 ng/ml DOX) or the negative control (-DOX), all members of the MCM2-7 complex were enriched (**Figure 10C-D**). This indicated that high DOX concentrations for the APEX2-NLS cell line led to unspecific biotinylation of chromatin associating factors such as the MCM2-7 complex. In addition to testing DOX concentrations, I assessed if chromatin isolation could increase specificity by enriching for chromatin associating proteins. Indeed, the replication fork factor PCNA and actively elongating RNA polymerase II were enriched upon chromatin isolation (**Figure 10E**). As a result of these optimization experiments, chromatin enrichment and lower APEX2-NLS DOX concentration were used in later proximity labeling experiments.
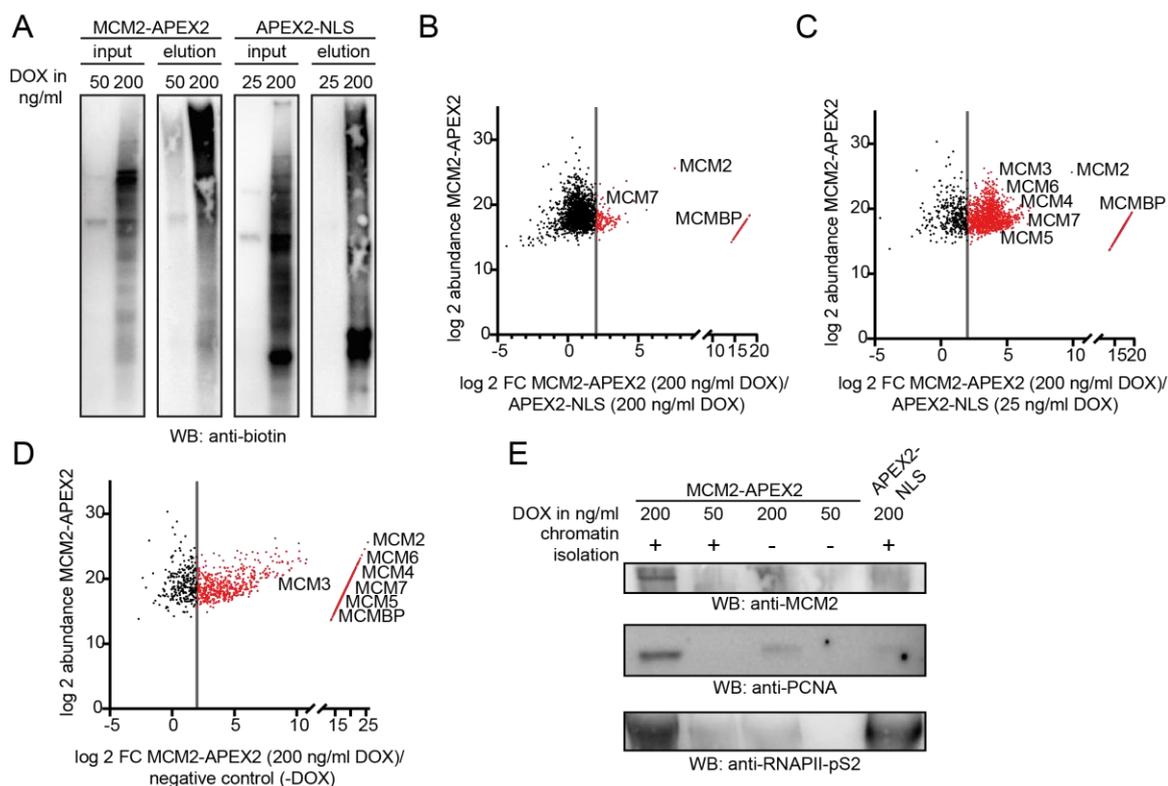
Figure 10: Optimization experiments for proximity labeling in the MCM2-APEX2 and APEX2-NLS expressing MCF7 cell lines. DOX was added 48h before harvest. Cells were incubated with 0.5 mM biotin phenol for 60 min before proximity labeling and fixation. **A**, Western Blots of biotinylated proteins in input and elution for both cell lines with indicated DOX concentrations. **B-D**, Plots showing the indicated protein abundance fold-changes measured by mass spectrometry. MCM2-7 complex members are labeled if they are > 2-fold enriched over the APEX2-NLS control. **E**, Western Blot of replicative and transcriptional factors in the elution with or without preceding chromatin isolation and with indicated DOX concentrations.

4.3 Induction of Transcription-Replication Conflicts in MCF7 cells

High levels of transcription can induce genomic instability owing to an increased incidence of R-loops and/or transcription-replication conflicts. Previously, it was shown that estrogen-responsive breast epithelial MCF7 cancer cells accumulate DNA damage upon prolonged treatment with high concentrations of estrogen (Stork et al., 2016). This damage was specific to S phase cells and dependent on

the massive E2-induced changes in R-loop formation in S phase, implying a role for perturbed transcription-replication coordination and/or R-loop mediated transcription-replication conflicts in this system. To induce a transcriptional burst in S phase MCF7 cells, I first arrested cells in G0/G1 by a combination of serum and hormone starvation (**Figure 11A**). Cells were then allowed to re-enter the cell cycle in complete medium. To monitor S phase entry, the incorporation of the thymidine analogue bromodeoxyuridine (BrdU) was analyzed by flow cytometry. After 20h release, ~30-45% of cells had progressed synchronously into early S phase, at which point I mock-treated or stimulated the cells with 100 nM E2 for 30min (**Figure 11B-C**).
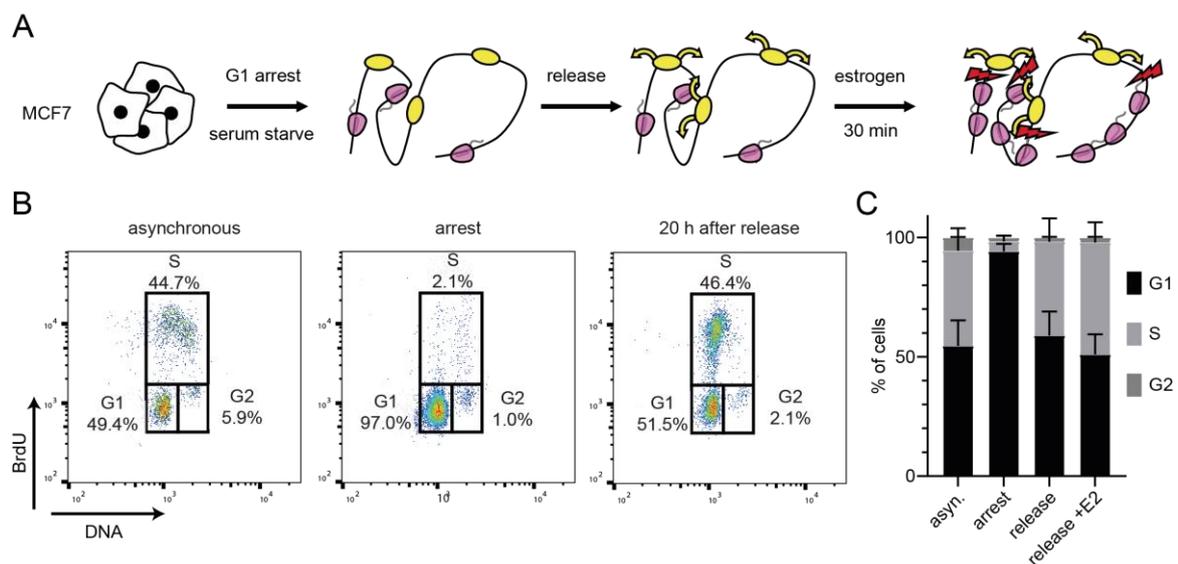
Figure 11: Inducing unscheduled transcription-replication interference in MCF7 cells. **A**, Schematic of the G1 arrest and S phase release with simultaneous estrogen treatment in MCF7 cells. Origins and replication forks are depicted in yellow and transcription complexes are magenta. Red lightning bolt illustrates transcription-replication conflicts. **B**, BrdU-propidium iodide (PI) cell cytometry profiles showing MCF7 cells growing asynchronously, arrested in G1 and 20 hours after adding back serum containing medium. Cells were categorized into G1, S or G2 phase by the amount of incorporated BrdU and DNA content. **C**, Bar graph showing the percentage of cells in G1-, S-, or G2 phase in the cell cytometry profiles of **B**. Error bars indicate the standard deviation. N=3.

Importantly, cells were treated with E2 for 16 h directly after hormone starvation in the Stork et al. study (2016). However, a study on global nascent RNA production changes in MCF7 cells upon short estrogen treatments (10, 40, 160 min) revealed the most dramatic up- and downregulation of thousands of genes already at the 10 and 40 min mark (Hah et al., 2011). Consequently, I treated the synchronized cells that were just entering early S phase for 30 min with E2 to induce unscheduled conflicts between the highly upregulated genes and replication forks. To confirm if estrogen-induced a transcriptional response in this setup, I employed nascent RNA sequencing with EU-seq (Jao & Salic, 2008; Yokoyama et al., 2016). To get insight into global transcriptional changes, I pulse-labeled MCF7 cells with the uridine analogue 5-ethynyluridine (EU). This compound is incorporated into nascent RNA and can be fluorescently labeled with click-chemistry for subsequent analysis of nascent RNA levels by RNA sequencing (Jao & Salic, 2008). Indeed, EU-seq revealed that transcription of the known estrogen upregulated genes was increased upon the short E2 stimulation (**Figure 12**), whereas cells released without E2 showed comparably lower transcription levels similar to an asynchronous cell population.

Next, I wanted to analyze whether this system leads to an increase in unscheduled conflicts between transcription and replication. For this, I determined the frequency of interactions between the transcription and replication machineries with a proximity ligation assay (PLA). I used antibodies against RNAPII phospho Ser2 and PCNA to target active transcription and replication, respectively **(Figure 13A)**.
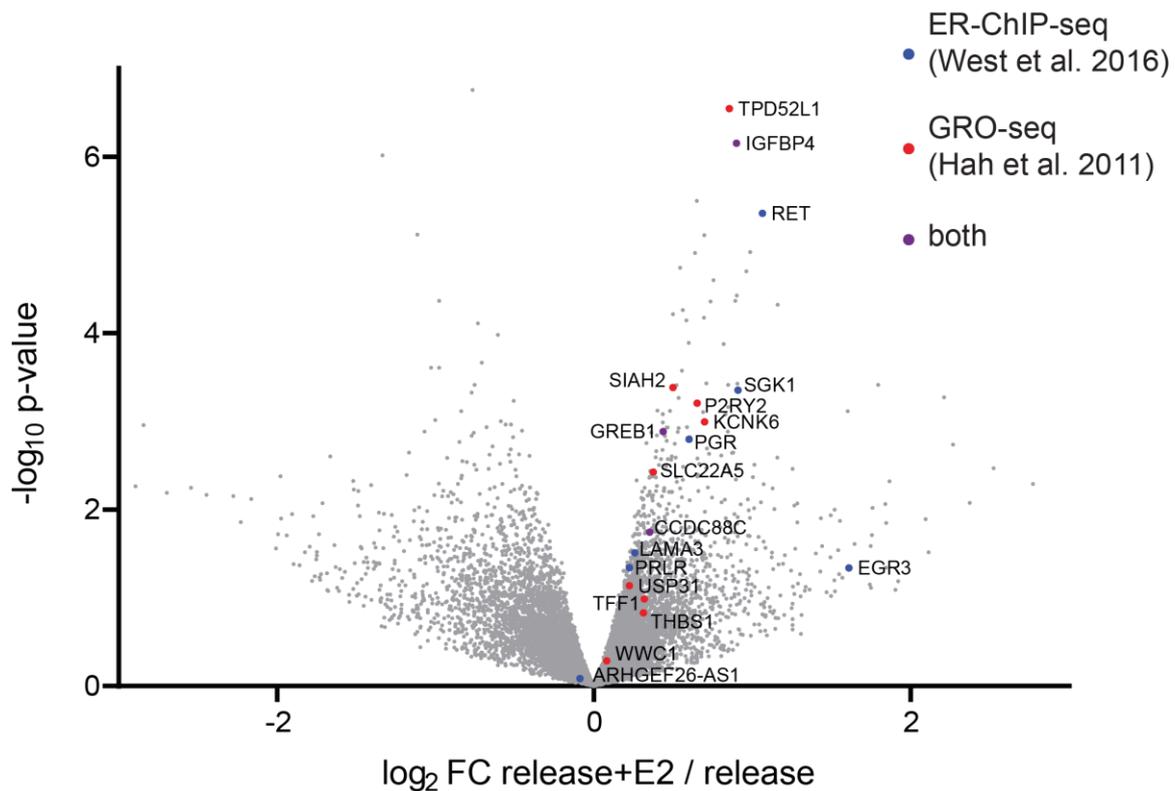
Figure 12: Nascent RNA-sequencing confirms global transcriptional changes upon short estrogen treatment in MCF7 cells. Volcano plot showing the up- and downregulation of nascent transcripts upon 30 min of estrogen treatment in MCF7 cells during S phase release. Transcripts known to be induced in the indicated studies have been colored accordingly.

As expected, PLA foci were mostly specific to S phase cells marked by EdU labeling (**Figure 13B-C**). Moreover, based on the distinct EdU incorporation pattern, early S phase cells generally score higher in their number of PLA foci than mid or late S phase cells (**Figure 13D**), consistent with the preferential early replication timing of transcriptionally active chromatin compartments (Petryk et al., 2016). As early S phase cells are enriched in these release conditions, this likely explains the significant increase in S phase specific colocalization of RNAPIIPS2 – PCNA compared to the asynchronous condition (**Figure 13C**, compare condition 2 with condition 4). Addition of the transcription elongation inhibitor 5,6-Dichlorobenzimidazole 1-β-D-ribofuranoside (DRB) lowered the PLA frequency after release regardless of whether cells were stimulated with estrogen or not (**Figure 13B-C**). However, estrogen stimulation did not further increase the number

of PLA foci per nucleus (**Figure 13C**, compare condition 4 with condition 8), suggesting that the frequency of TRC formation was not markedly increased.
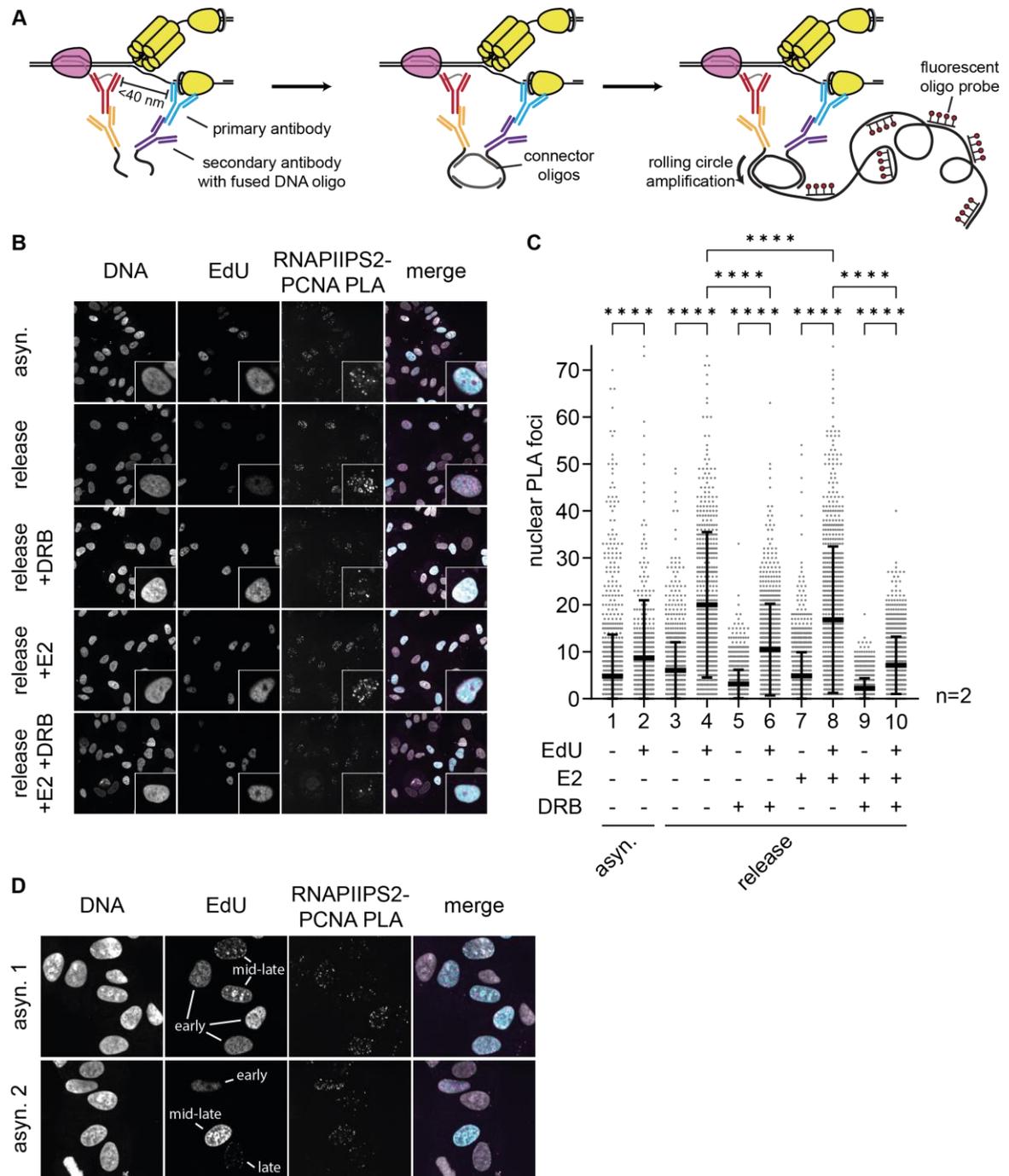
Figure 13: Imaging of TRC-PLA foci upon short estrogen treatment in MCF7 cells. **A**, Cartoon of the *in situ* proximity ligation assay. Only if the targets of two primary antibodies are in proximity, fluorescent oligo probes detect DNA generated by rolling circle amplification. **B**, Representative immunofluorescence images of TRC-PLA foci, DNA staining and EdU incorporation with or without 30 min of estrogen treatment in MCF7 cells during S phase release. For transcriptional inhibition, cells

were treated with 100 µM DRB 2 hours before fixation. In the bottom right of each image, one enlarged example EdU+ cell is shown. **C**, Quantification of TRC-PLA foci in all cells. The cell population for each condition was split into EdU- and EdU+ cells. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. Data is pooled from two biological replicates. **D**, Example immunofluorescence images of TRC-PLA foci and S-phase progression determined by distinct EdU patterns.

4.4 Transcription-Replication Conflict Proximity Labeling with MCM2-APEX2

Next, I used the MCM2-APEX2 and APEX2-NLS cell lines to screen for factors that are present at the MCM complex during unscheduled E2-induced transcription perturbation. For this, I used the arrest and release system described in section 4.2 with the MCM2-APEX2 and APEX2-NLS cell lines followed by proximity labeling and mass spectrometry (**Figure 14A**).
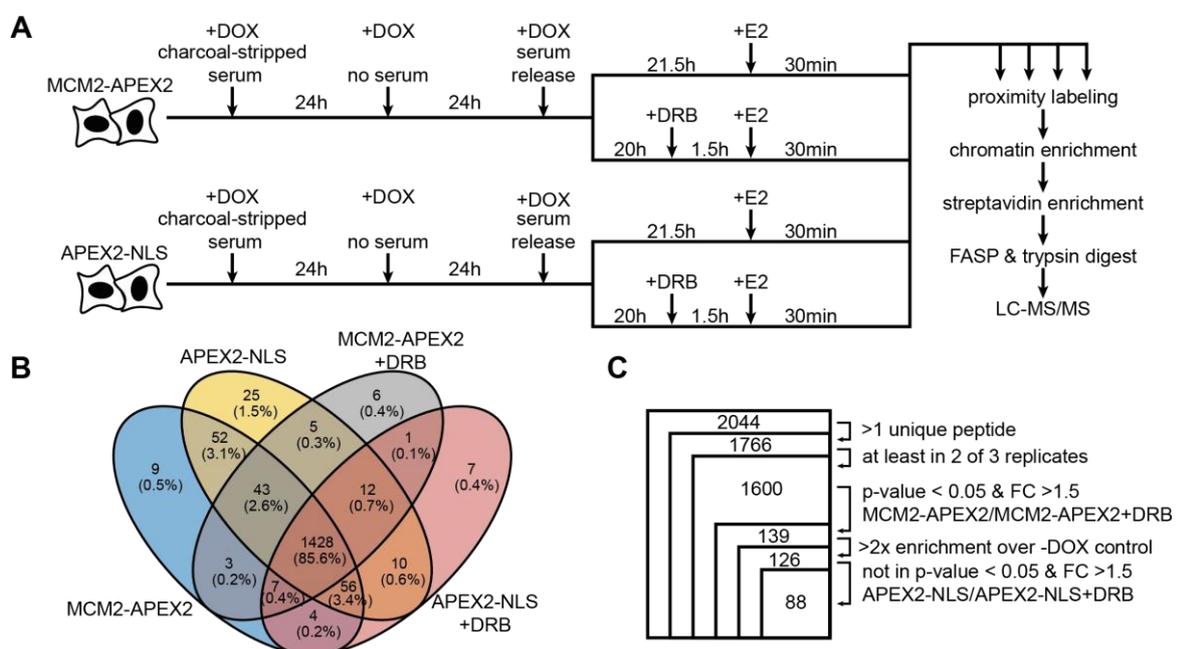
Figure 14: Transcription-replication conflict proximity labeling. **A**, Experimental setup with MCM2-APEX2 (200 ng/ml DOX) and APEX2-NLS (100 ng/ml DOX) expressing MCF7 cell lines. Cells were incubated with 0.5 mM biotin phenol for 60 min before 1 min of proximity labeling and harvest. **B**, Venn diagram of overlapping

proteins in the four conditions. **C**, Filtering steps to obtain the final list of 88 candidates.

Transcription inhibition with DRB served as a negative control, as it significantly reduces RNAPIIPS2-PCNA PLA foci levels (**Figure 13**). As an additional negative control, the same set of samples was analyzed in the absence of DOX and therefore no biotinylation was expected in the cells (see **Figure 9**A, 0 DOX). In total, 2044 proteins were identified across the two cell lines and four conditions (**Figure 14A-C**) by LC-MS after proximity labeling, chromatin enrichment and streptavidin pulldown of biotinylated proteins. There was a large overlap of identified proteins across all samples (**Figure 14B**), suggesting that many of the identified proteins are derived from unspecific interactions with the beads and/or affinity reagents. Members of the replication fork were only poorly enriched in the MCM2-APEX2 proximity labeling when compared to APEX2-NLS (**Figure 15**), indicating that the overexpression condition of MCM2-APEX2 could also effectively label the general nuclear proteome.
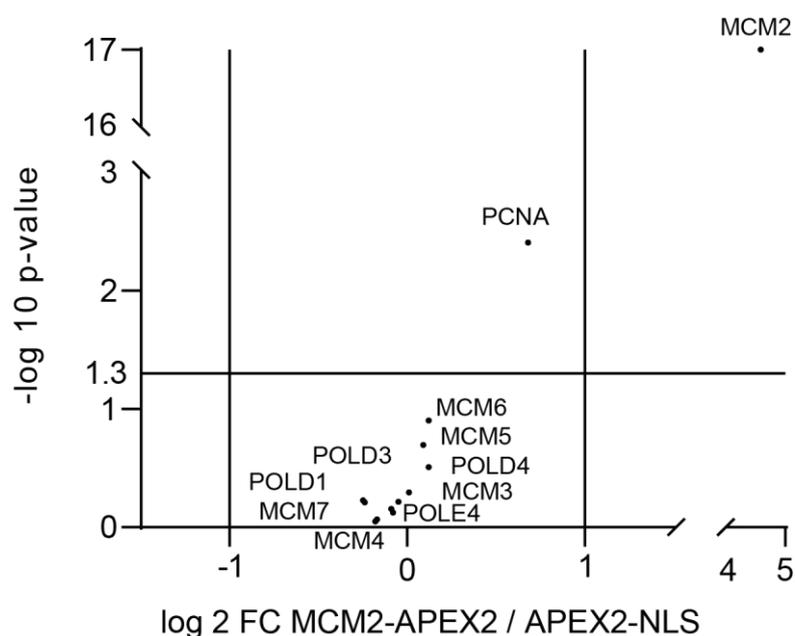
Figure 15: Low enrichment of replication factors in MCM2-APEX2 vs. APEX2-NLS samples. Volcano plot showing the abundance ratio of selected replication factors

in MCM2-APEX2 (200 ng/ml DOX) vs. AEPX2-NLS (100 ng/ml DOX). Cut-off lines are x=±1 and y= 1.3 (p<0.05).

This is also consistent with the results of the chromatin fractionation showing a large amount of MCM2-APEX2 in the nucleoplasmic fraction (**Figure 9E**). Therefore, to find proteins specifically enriched upon induction of unscheduled E2 transcription perturbation, I used label-free quantification and employed a series of filtering steps that excludes proteins either not identified with high confidence (>1 unique peptide, present in at least 2 of 3 biological replicates), or the level of enrichment and p-value statistics when compared to -DOX or DRB conditions (**Figure 14C**). This led to a final list of 88 candidates that were enriched in the MCM2-APEX2 cell line under conditions of estrogen-induced transcription perturbation (**Figure 16** and **Figure 17**).
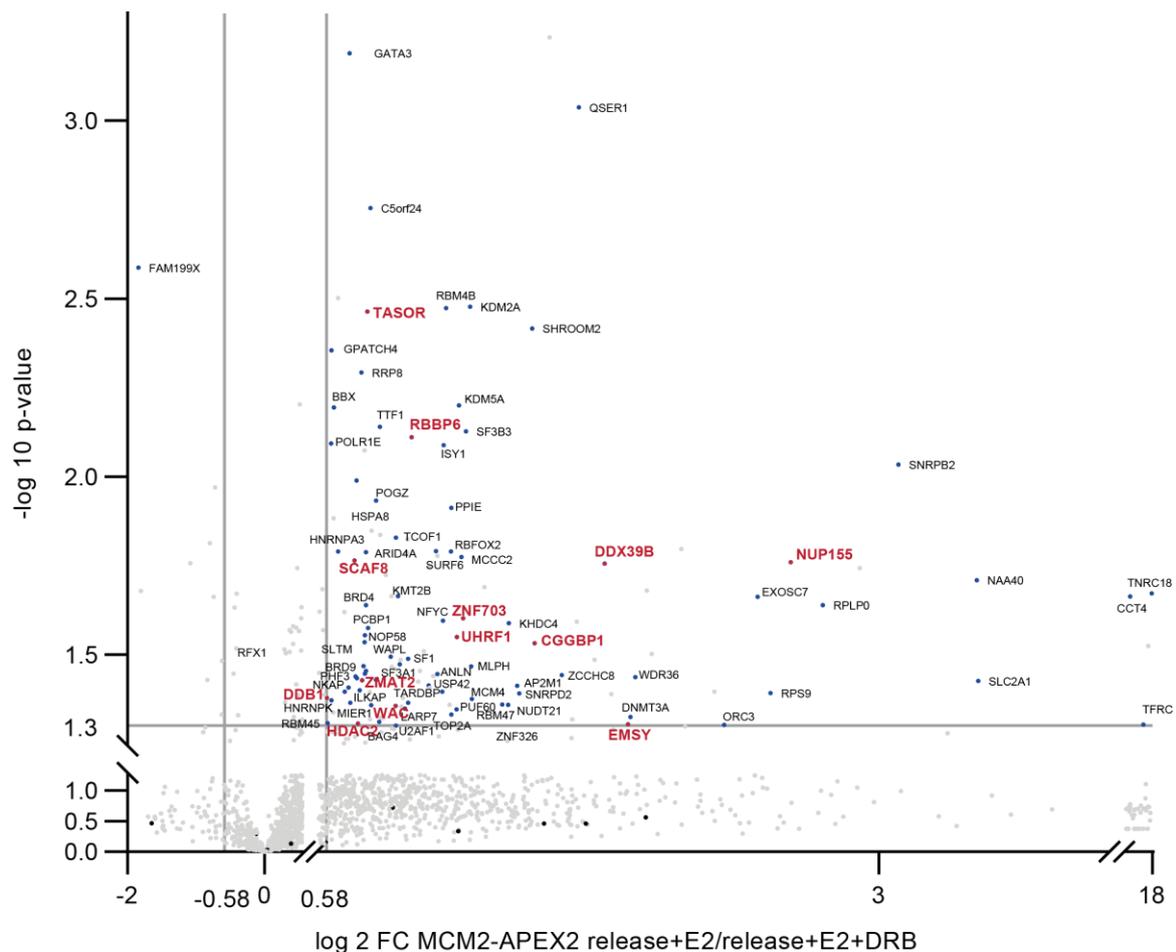
Figure 16: Volcano plot showing enrichment of TRC associating candidates. Abundance ratio of MCM2-APEX2 release + E2 vs. release + E2 + DRB. Cut-off

lines are x=±0.58 (FC>1.5) and y= 1.3 (p<0.05). Red labeled factors represent candidate factors that were selected for further characterization.

Interestingly, this list contained factors that have already been implicated in transcription-replication conflicts, e.g. BRD4 (F. C. Lam et al., 2020) and DDX39B (Pérez-Calero et al., 2020). Next, I performed a network analysis and Gene Ontology (GO) enrichment analysis on the candidates and found that the factors clustered in two main functional annotations, namely splicing and mRNA processing as well as DNA binding and chromatin organization factors (**Figure 17**).

Figure 17: STRING network of the 88 candidate TRC factors. The thickness of the connections illustrates the confidence level. Red labeled factors represent candidate factors that were selected for further characterization.

To characterize some of these candidates in detail, I followed up on 13 factors that were of particular interest based on their described function and previous literature with connections to TRCs (**Table 2**).

Table 2: List of selected candidates potentially playing a role in TRCs

Gene	Name	Connection to TRCs	Reference
CGGBP1	CGG triplet repeat-binding protein 1	Involvement in transcription and DNA repair	(Singh & Westermark, 2015)
DDB1	DNA damage-binding protein 1	CUL4-DDB1 complex prevents re-replication	(Zhong et al., 2003)
DDX39B	Spliceosome RNA helicase DDX39B	R-loop processing helicase	(Pérez-Calero et al., 2020)
EMSY	BRCA2-interacting transcriptional repressor	H3K4me3 demethylase, H3K4me3 directly implicated in TRCs	(Varier et al., 2016; Chong et al., 2020)
HDAC2	Histone deacetylase 2	HDAC2 inhibitor: Reduced replication fork speed, Increased replication stress	(Bhaskara et al., 2013)
NUP155	Nuclear pore complex protein	Involved in topological stress relieve upon replication stress	(Bermejo et al., 2011)
RBBP6	E3 ubiquitin-protein ligase	RBBP6 knockdown: Reduced replication fork speed, increased damage at common fragile sites	(Miotto et al., 2014)
SCAF8	SR-related and CTD-associated factor 8	interacts with RNAPII and RECQL5	(Gregersen et al., 2019)
TASOR	Protein TASOR	Associates with RNA processing	(Douse et al., 2020)
UHRF1	E3 ubiquitin-protein ligase	Associates with the replication fork, Plays role in DNA repair	(Sharif et al., 2007; Motnenko et al., 2018)
WAC	WW domain-containing adapter protein with coiled-coil	Interacts with RNAPII at TSSs, Cell cycle checkpoint activator	(Zhang & Yu, 2011)
ZMAT2	Zinc finger matrin-type protein 2	Involved in pre-mRNA splicing	(Tanis et al., 2018)
ZNF703	Zinc finger protein 703	Transcriptional corepressor, HDAC dependent	(Nakamura et al., 2008)

Next, I performed experiments in MCF7 cells to assess and validate the consequences of depleting these factors on transcription-replication conflicts and associated replication stress.

4.5 Validation and characterization of selected candidates

RT-qPCR confirmed the depletion of the candidate transcripts by ~70-100% when comparing the control siRNA treatment with the specific candidate siRNA treatment (**Figure 18**).
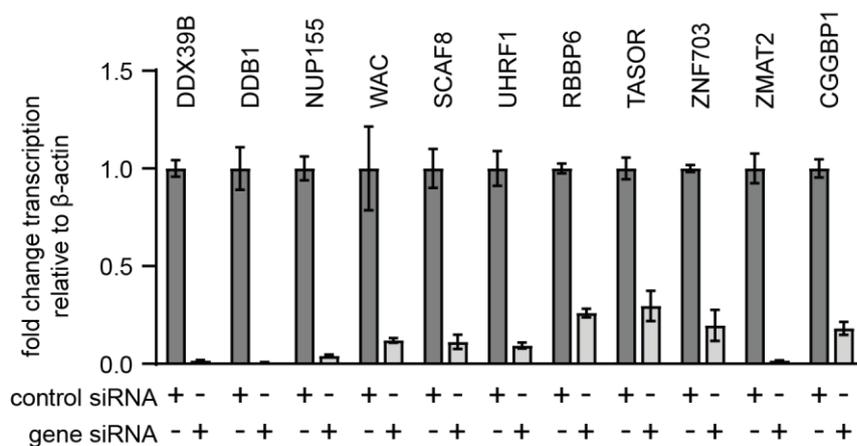
Figure 18: Confirmation of siRNA knockdown of candidate genes. Transcript level analysis via RT-qPCR of candidate siRNAs compared to control siRNA relative to β-actin after 72 h of siRNA treatment. Data is represented as mean of three technical replicates ± standard deviation.

First, I first measured if siRNA-mediated depletion of the factors leads to changes in γH2AX foci count per EdU+ nucleus. Indeed, depletion of several factors led to significantly increased γH2AX foci formation compared to the control siRNA (**Figure 19A-B**). In particular, siRNA-mediated knockdown of RBBP6 and TASOR led to the largest increase of foci of the DNA damage marker γH2AX (**Figure 19A-B**). Next, measured if depletion of the factors leads to changes in RNAPIIPS2-PCNA PLA foci count per EdU+ nucleus. Interestingly, siRNA knockdown of DDB1, WAC and TASOR significantly increased the number of PLA foci (**Figure 19C-D**). In contrast, depletion of HDAC2, CGGBP1 and EMSY led to a significant decrease in TRC-PLA foci (**Figure 19C-D**). In summary, this indicates a potential involvement of DDB1, WAC, TASOR, HDAC2, CGGBP1 and EMSY in the regulation of cellular TRC levels.
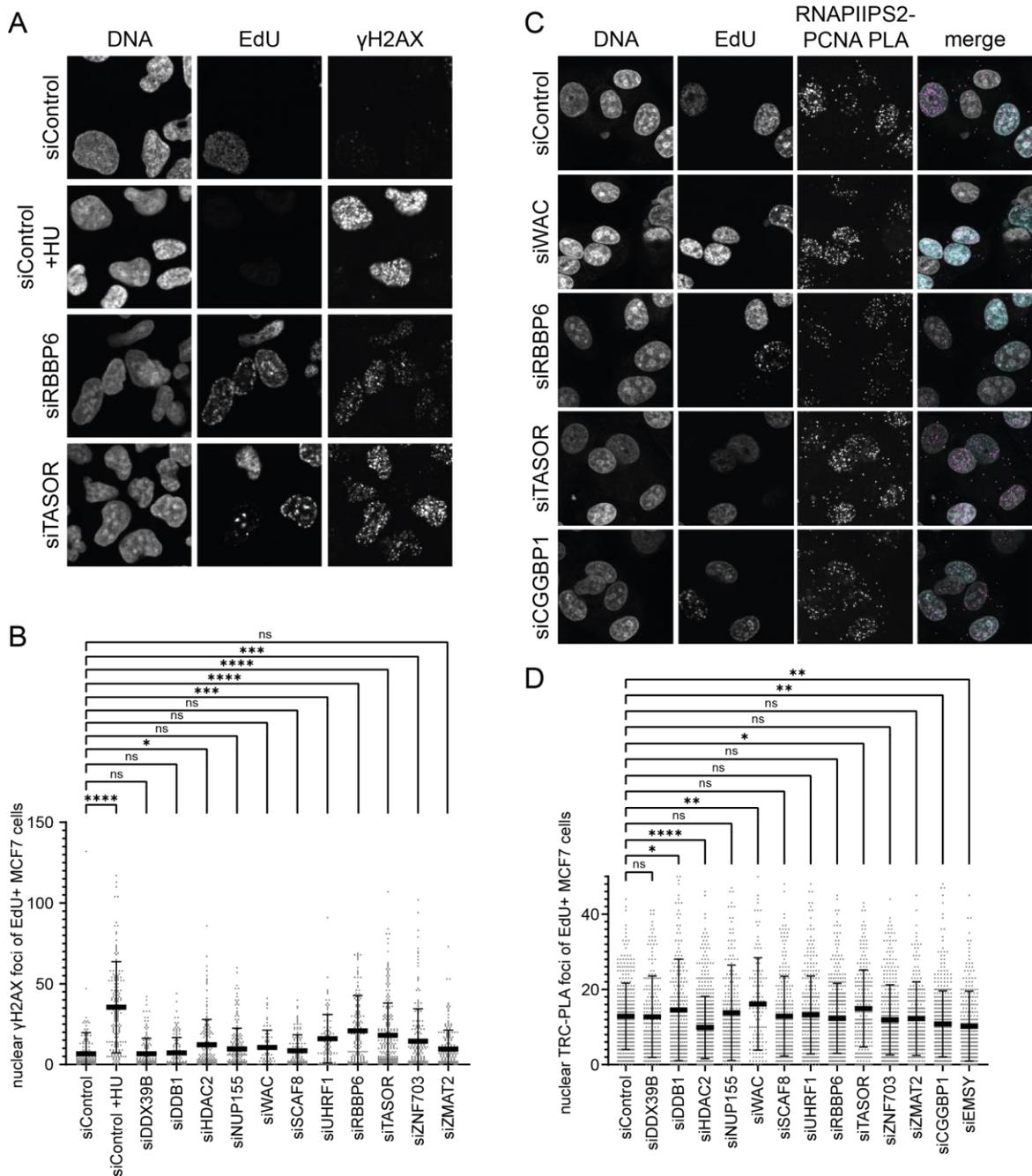
Figure 19: Imaging of γH2AX and TRC-PLA foci formation upon depletion of candidate proteins. **A**, Example immunofluorescence images of γH2AX foci, DNA staining and EdU incorporation in MCF7 cells with indicated 72 h siRNA knockdown. **B**, Quantification of γH2AX foci in EdU+ MCF7 cells. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well with >100 nuclei was analysed.**C**, Example immunofluorescence images of TRC-PLA foci, DNA staining and EdU incorporation in MCF7 cells with indicated 72 h siRNA knockdown. **D**, Quantification of TRC-PLA foci in EdU+ MCF7 cells. Data is represented as mean

± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well with >150 nuclei was analysed.

As siRNA-mediated knockdown of the four proteins WAC, RBBP6, TASOR and CGGBP1 led to significant changes in γH2AX or TRC-PLA foci formation, I decided to focus on these candidates. Next, I measured FANCD2 foci formation upon knockdown of these factors. FANCD2 is a general DNA damage marker that accumulates at stalled replication forks (Nepal et al., 2017). Interestingly, depletion of WAC, RBBP6, TASOR and CGGBP1 increased FANCD2 foci formation in an S phase specific manner compared to non-targeting siRNA control knockdown (**Figure 20A-B**).
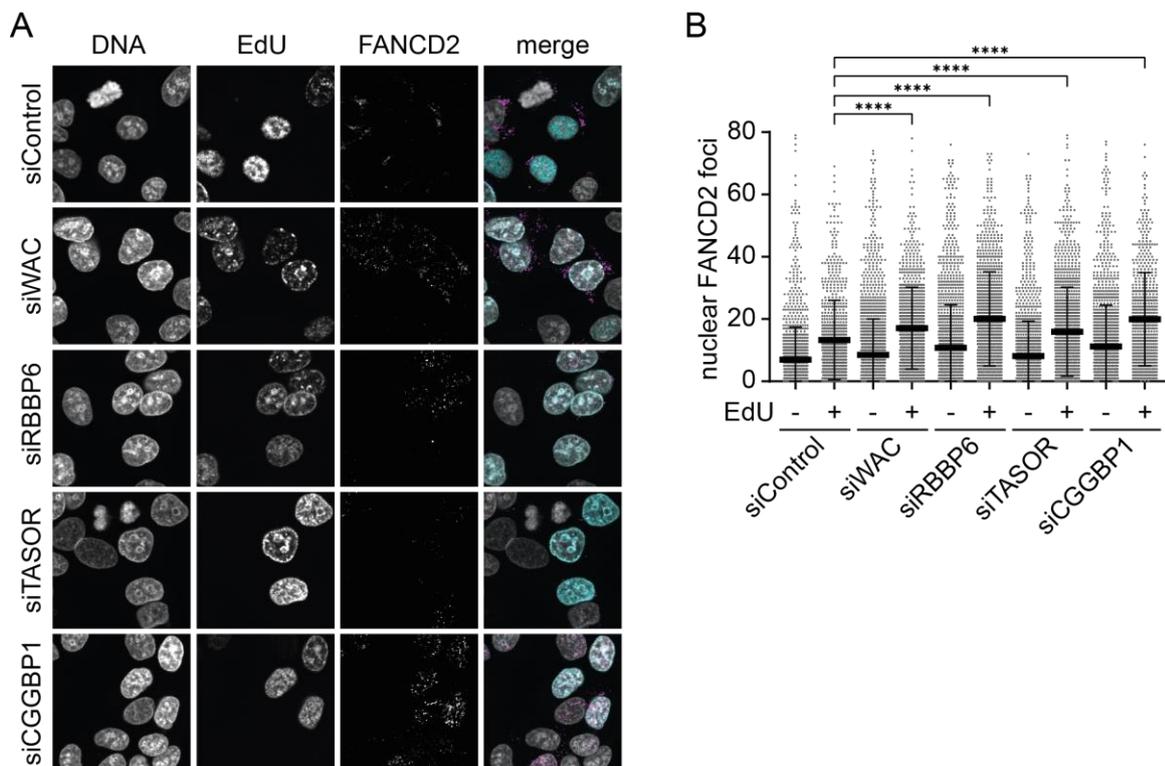
Figure 20: Imaging of TRC-PLA and FANCD2 foci upon depletion of candidate proteins. **A**, Example immunofluorescence images of FANCD2 foci, DNA staining and EdU incorporation in MCF7 cells with indicated 72 h siRNA knockdown. **B**, Quantification of FANCD2 foci in MCF7 cells. The cell population for each condition was split into EdU- and EdU+ cells. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, three wells with each >500 nuclei were analysed. **D**, Example

immunofluorescence images of TRC-PLA foci, DNA staining and EdU incorporation in MCF7 cells with indicated 72 h siRNA knockdown. **E**, Example immunofluorescence images of FANCD2 foci, DNA staining and EdU incorporation in MCF7 cells with indicated 72 h siRNA knockdown.

Together, this data suggests that WAC, TASOR and CGGBP1 could play a specific role in TRC resolution, whereas RBBP6 more likely takes over a TRC-independent but important role for maintaining genome stability during S phase. In summary, this proximity labeling based strategy provided a candidate list of potential proteins that are specifically enriched at a transcriptionally challenged replication fork and will become a useful resource for future projects. CGGBP1 was, among the shortlisted candidates, one of the most interesting factors that was further characterized as part of this PhD thesis project.

4.6 CGGBP1

4.6.1 Global profiling of CGGBP1 binding sites in the human genome

As introduced, CGGBP1 was first characterized as a protein binding specifically to CGG triplet repeats. To determine binding sites across the genome, I used a publicly available ENCODE CGGBP1 Chromatin Immunoprecipitation-Sequencing (ChIP-Seq) data set in human K562 cells (The ENCODE Project Consortium, 2012). By comparing two biological replicates to the untagged control, 3459 high-confidence CGGBP1 binding sites could be identified. Interestingly, peak annotation revealed ~73% of binding peaks within 1 kb of promoters and only ~10% distal intergenic peaks (**Figure 21A-B**). This was confirmed by a gene metaplot of the ChIP-Seq signal, which showed an elevated signal at the transcription start sites (TSSs) of most genes. Only a few genes had either no signal at all or signal throughout the gene body, as shown by examples in **Figure 21A**. As expected, the top three binding motifs, uncovered by motif enrichment analysis of the peaks, were CGG repeats, but interestingly only a short motif of 3-4 repeats was sufficient for CGGBP1 binding (**Figure 21D**). It is necessary to point out that CGGBP1 has been shown to bind to telomeres, which could not be mapped by the ChIP-Seq approach here that excluded repetitive regions of the genome. Similarly, other repetitive regions e.g. rDNA could also not be analyzed.

To summarize, most detected binding events at non-repetitive regions of the genome are located at CGG repeats inside of genes, with a strong preference for promoter regions.
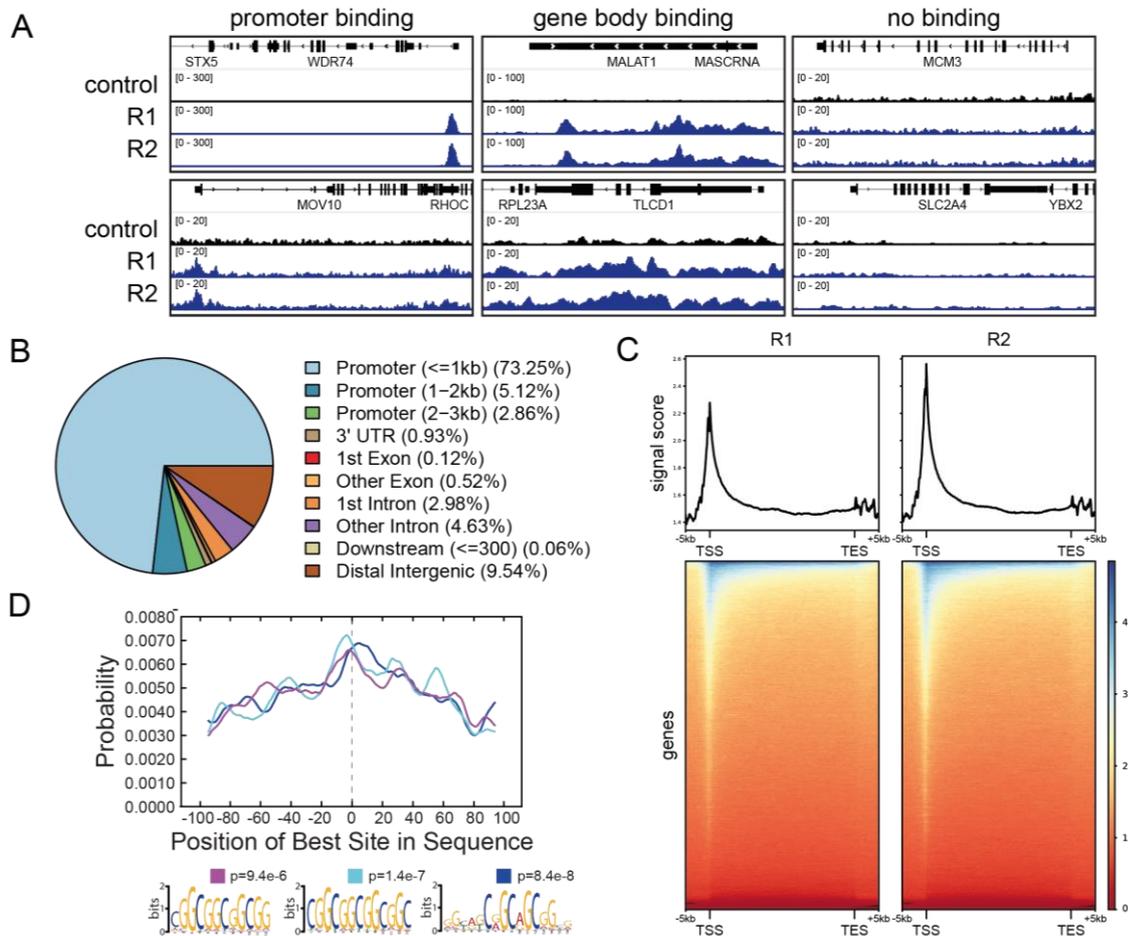
Figure 21: Analysis of CGGBP1 ChIP-Seq data from K562 cells. **A**, Example tracks of CGGBP1 ChIP-seq signal in K562 cells at different genes of two replicates (R1 and R2) compared to the control. **B**, Pie chart showing the distribution of CGGBP1 ChIP-seq peak annotation relative to the position in the genome. **C**, Metagene plots and underlying heatmaps of CGGBP1 ChIP-seq signal of the two replicates R1 and R2 across all annotated transcripts. **D**, Motif probability graph and DNA sequence logos of the most significant hits found in the CGGBP1 ChIP-seq peak sequences.

Next, I investigated whether the TSSs with a CGGBP1 ChIP-Seq peak had special DNA sequence properties regarding the GC percentage, GC skew, G-quadruplex prediction or R-loop formation. As expected, the GC percentage of the 2688

CGGBP1 peak TSSs was increased compared to two independent sets of TSSs that were randomly chosen from the genome (data not shown). To account for the fact that GC skew, G4 prediction and R-loop formation are strongly influenced by GC percentage of the underlying DNA sequence, I selected two independent sets of control TSS that are not bound by CGGBP1 but show a comparable GC content to the CGGBP1 peak containing TSS (**Figure 22A**). The GC-skew is defined as an asymmetric excess of C over G on a given strand of the DNA. As expected, the GC skew is very similar for the CGGBP1 peak TSSs compared to the two random control sets of TSSs with matched GC percentage (**Figure 22B**). However, G4 formation is predicted to be significantly higher in a narrow window of 400 bp around the TSS, but no difference to the random control TSS is observed directly upstream and downstream of this TSS window (**Figure 22C**). Most strikingly, publicly available DRIP-Seq data in K562 cells showed a strong enrichment of R-loop levels around CGGBP1 peak TSSs compared to the two control TSSs (**Figure 22D**). Together these results indicate that DNA secondary structures, specifically G-quadruplexes and R-loops, are more likely to form at TSSs with CGGBP1 binding.
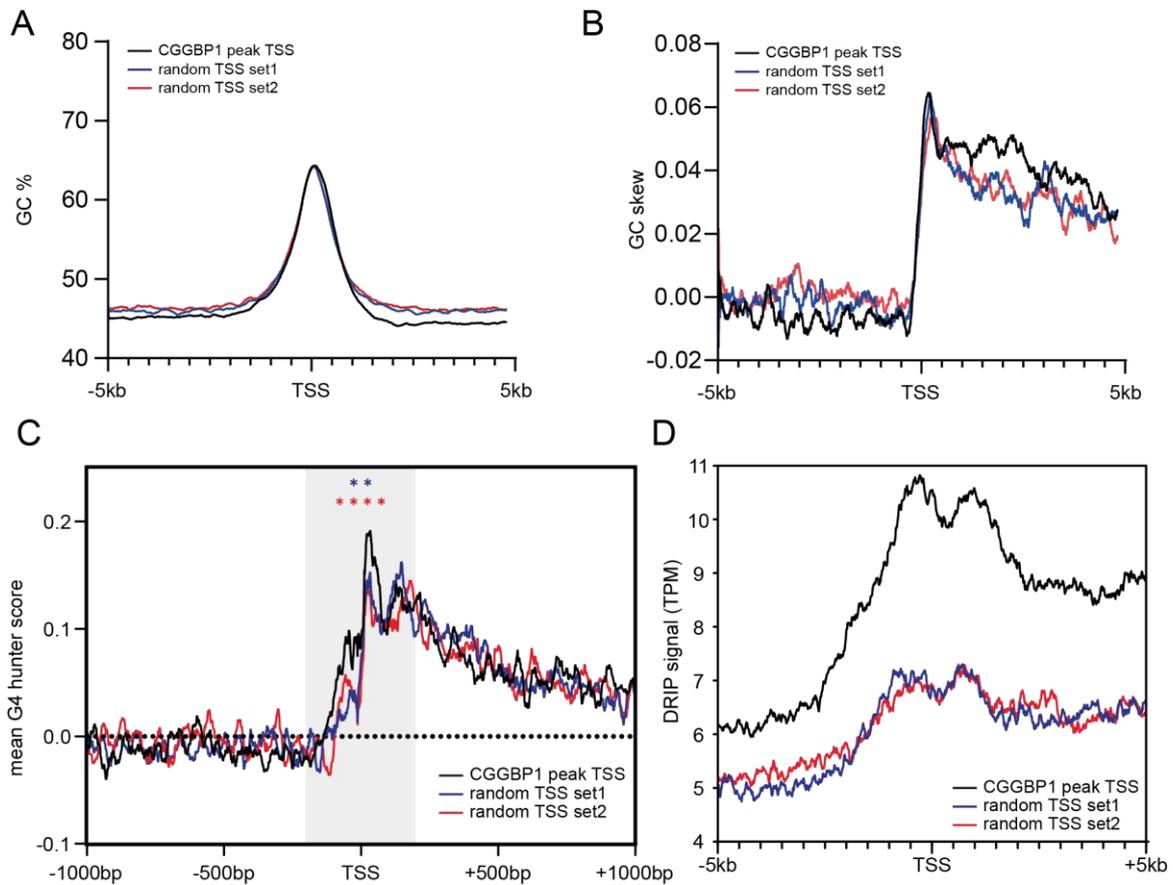
Figure 22: CGGBP1-binding TSSs show higher G4-forming potential and R-loop formation. **A**, Summary plot of GC% of the DNA sequence at either TSSs with a CGGBP1 binding peak in proximity or two random control sets of TSSs with a matched GC% to the CGGBP1 peak containing TSS. **B**, Summary plot of GC skew of the DNA sequence at either TSSs with a CGGBP1 binding peak in proximity or the two random sets of TSSs. GC skew was calculated with the formula (G-C)/(G+C). **C**, Summary plot of predicted G4 formation scores at either TSSs with a CGGBP1 binding peak in proximity or the two random control sets of TSSs. Statistical significance was calculated using two tailed student's t-test on the medians of 10 bp windows at the TSS +/- 100 bp indicated by the grey box. Significance was tested against both controls in individual t-tests indicated by the color of the stars. **D**, Summary plot of DRIP-seq signal at either TSSs with a CGGBP1 binding peak or the two random control sets of TSSs.

4.6.2 Altering cellular CGGBP1 levels leads to global transcriptional changes

CGGBP1 is primarily binding to thousands of RNAPII promoter sites, but how depletion or overexpression of CGGBP1 influences the transcriptional landscape of cells has not been comprehensively addressed. To answer this question, I knocked down CGGBP1 using a pool of four distinct siRNAs targeting CGGBP1. Western Blot of total cell extracts revealed efficient depletion of CGGBP1 at all timepoints after siRNA transfection (**Figure 23A**). To get insight into global transcriptional activity, I pulse-labeled the cells with EU, which can be fluorescently labeled by click-chemistry for subsequent microscopy analysis (Jao & Salic, 2008). Strikingly, CGGBP1-depleted cells show significantly increased transcriptional activity compared to control cells (**Figure 23B-C**). As expected, treatment of cells with the transcription elongation inhibitor DRB reduced nuclear EU levels in CGGBP1-depleted and control cells. Despite the low levels of EU incorporation upon DRB treatment, CGGBP1-depleted cells maintained a small but significant increase in transcriptional activity compared to control cells, indicating that this effect may not only be dependent on RNAPII transcription, but the other two nuclear RNA Polymerases I and III may also contribute to the elevated EU levels in CGGBP1-depleted cells (**Figure 23B-C**).
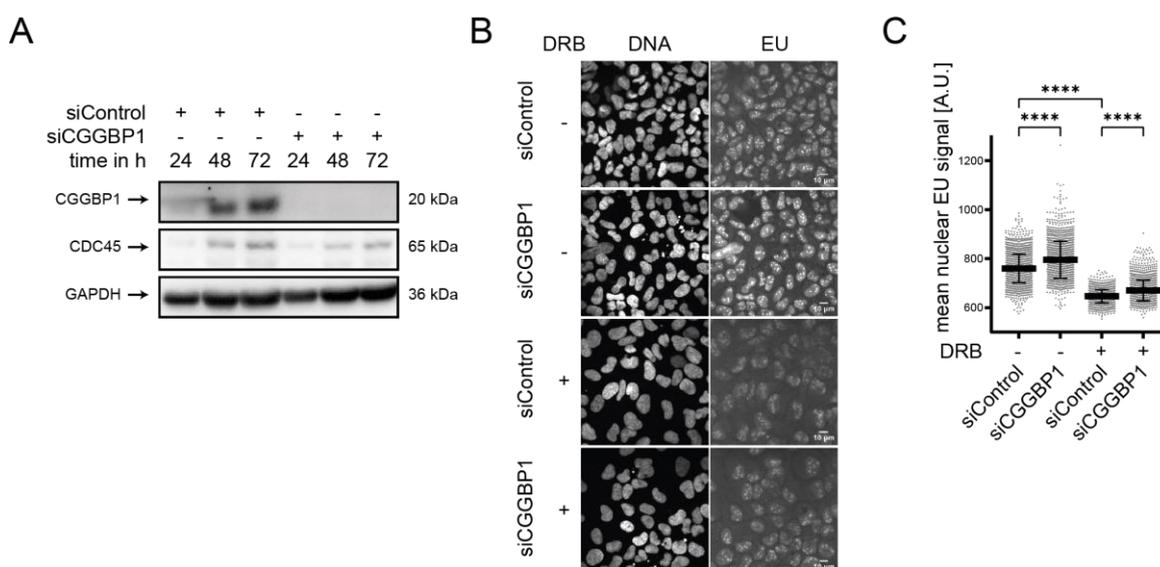
Figure 23: Global transcriptional activity is increased upon siRNA-mediated CGGBP1 knockdown in U2OS cells. **A**, Western Blot of U2OS total cell extracts after 24, 48 and 72 h of CGGBP1 knockdown compared to control siRNA. CDC45

and GAPDH are loading controls. N=1. **B**, Example immunofluorescence (IF) images of EU incorporation in U2OS cells upon 72 h CGGBP1 knockdown compared to control siRNA. For transcriptional inhibition, 100 µM DRB was added 2 h before fixation. **C**, Quantification of mean nuclear EU signal in all fields similar to B. For transcriptional inhibition, 100 µM DRB was added 2 h before fixation. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, two wells with each >1000 nuclei were analysed.

EU labeling measures the total nascent transcriptional activity of cells and cannot discriminate between RNAPII transcribed genes with or without CGGBB1 binding. To answer this question for some candidate genes, I measured the transcriptional output of three genes that contain CGGBP1 binding sites (MOV10, MALAT1 and TLCD1, **Figure 24A**) and two genes without CGGBP1 binding sites (UHRF1 and ZNF703) by RT-qPCR. Interestingly, transcriptional output of the CGGBP1 binding genes was reduced to ~40-50% in CGGBP1-depleted cells compared to control cells. This effect was less pronounced or not visible for the two non-CGGBP1 binding genes. This indicated – at least for the few genes that were tested in this assay – that CGGBP1 is a potential positive regulator of transcription at genes with CGGBP1 binding at its own promoter.
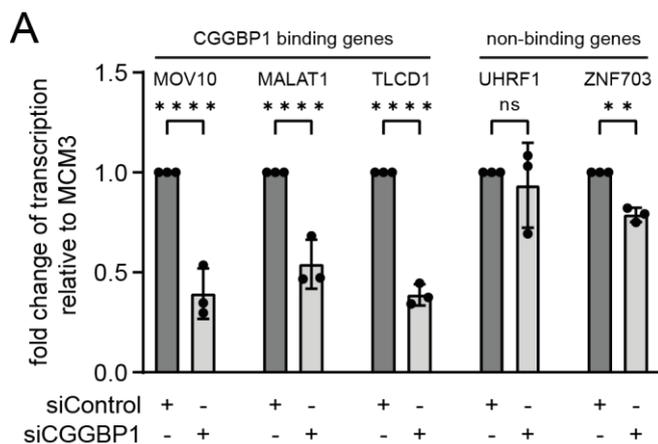
Figure 24: Transcriptional output of example genes with or without CGGBP1 binding in U2OS cells. **A**, Gene expression of selected example genes measured by RT-qPCR of cDNA from U2OS Tet-ON pHU43 clone 2 (see **Figure 33**) treated with siControl or siCGGBP1 for 72 h and no DOX. Shown is the fold change relative

to MCM3 and normalized to siControl. Statistical significance was calculated using one-way ANOVA. N=3.

Next, I wanted to investigate whether overexpression of CGGBP1 would have the opposite effect on the transcriptional landscape in cells compared to the knockdown experiments. To study this, I transiently transfected cells with an overexpression vector containing an HA-tagged CGGBP1 open reading frame. Western Blot of total cell extracts confirmed strong HA-CGGBP1 overexpression 48h and 72h after transfection compared to an empty vector control (**Figure 25A**). To assess global transcriptional changes, I measured EU incorporation in HA (-) and HA (+) cells with and without DRB treatment. Overexpression of HA-CGGBP1 decreased global transcriptional activity in comparison to the empty control vector, which also held true for DRB treated cells (**Figure 25B-C**). Similar to the knockdown experiments, this result indicated an additional involvement of the other RNA polymerases I and III in this decrease. Importantly, a direct comparison of transfected and therefore HA (+) versus non-transfected HA (-) cells within the same experimental condition revealed a significant reduction in EU signal in the HA (+) cells, supporting the notion that higher CGGBP1 levels indeed reduce nascent RNA levels (**Figure 25D**).
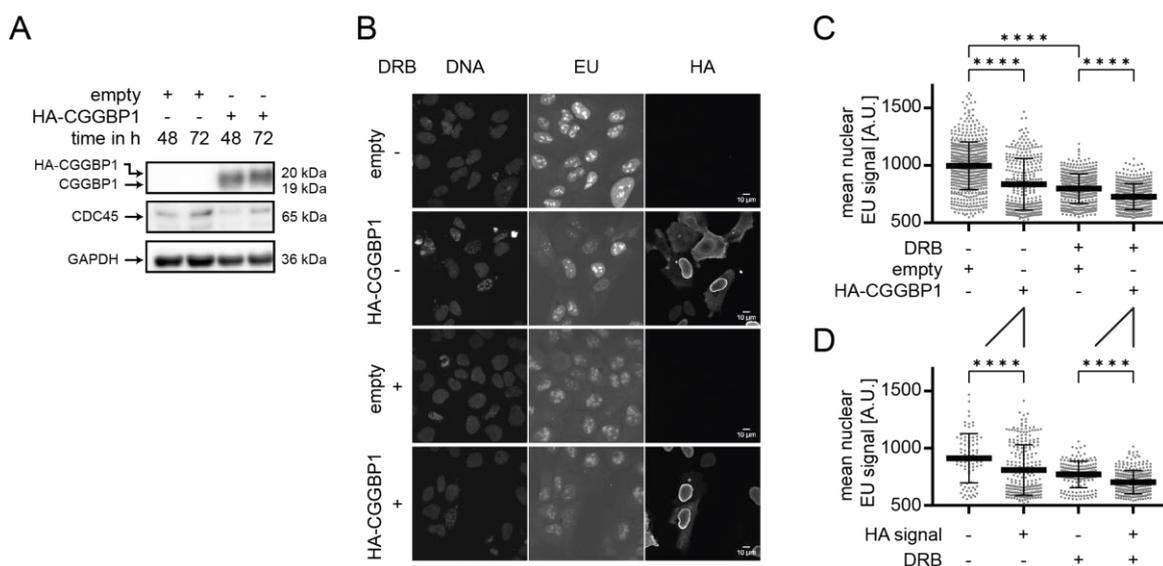
Figure 25: Global transcriptional activity is decreased upon overexpression of HA-CGGBP1 in U2OS cells. **A**, Western Blot of HA-CGGBP1 overexpression in U2OS cells. Total cell extracts were taken 48 or 72 h after transfection of the

overexpression plasmid. CDC45 and GAPDH are loading controls. WT CGGBP1 is not visible here due to the excess amount of HA-CGGBP1. **B**, Example IF images of EU incorporation and HA signal in U2OS cells upon HA-CGGBP1 overexpression compared to an empty vector control. For transcriptional inhibition, 100 µM DRB was added 2 h before fixation. **C**, Quantification of mean nuclear EU signal in all fields similar to B). Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, results are from one well containing >1000 cells. **D**, Quantification of mean nuclear EU signal in all fields similar to B). Only the overexpression condition is shown with cells split into HA+ or HA-. Statistical significance was calculated using one-way ANOVA. For each condition, one well each containing >1000 nuclei was analysed.

The CGGBP1 depletion and overexpression experiments showed a clear adverse effect on transcriptional activity. Thus, I next asked how this transcriptional response to different CGGBP1 levels affects cell cycle progression. To address this, I performed BrdU or EdU labeling and subsequent flow cytometry or microscopy analysis, respectively. When CGGBP1 is depleted for 48h or 72h, cells accumulate in G1 phase and the S phase population is reduced (**Figure 26A-C**). Interestingly, when overexpressing CGGBP1, a similar reduction of S phase cells is observed (**Figure 26D**).
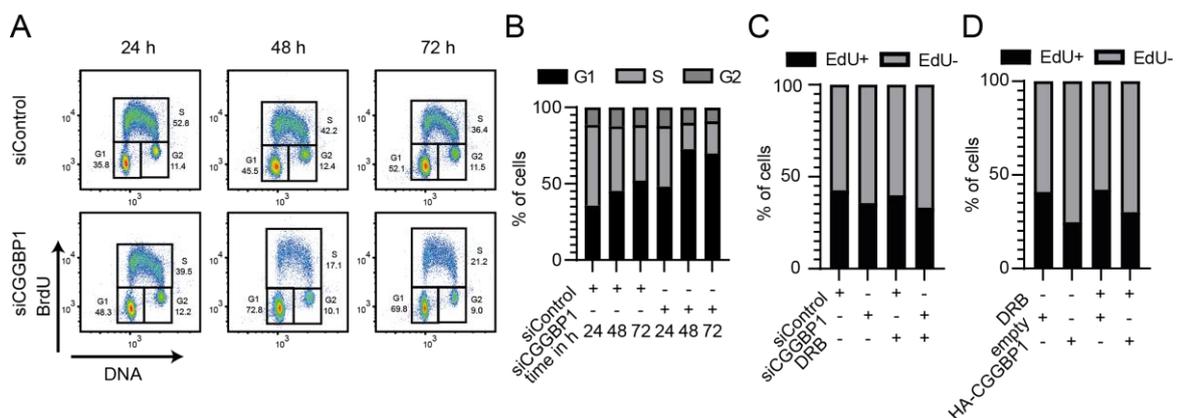
Figure 26: Depletion and overexpression of CGGBP1 affects cell cycle progression **A**, BrdU-PI flow cytometry profiles of U2OS cells upon treatment with siControl or siCGGBP1 for 24, 48 and 72 h. The percentage of cells in G1, S or G2 are quantified next to the gates. For each condition >30000 cells were analysed.

N=1. **B**, Quantification of cell cycle distribution in BrdU cell cycle analysis by flow cytometry upon treatment of U2OS cells with siControl or siCGGBP1 for 24, 48 and 72 h. N=1. **C**, Bar graph showing the percentage of EdU+ and EdU- cells in all fields similar to **Figure 27A**. For each condition, two wells with each >900 nuclei were analysed. **D**, Bar graph showing the percentage of EdU+ and EdU- cells in all fields similar to **Figure 28A**. For each condition, three wells with each >900 nuclei were analysed.

In summary, balanced cellular CGGBP1 levels seem to be crucial for cell cycle progression and proper transcriptional activity, as knockdown and overexpression of CGGBP1 positively and negatively influence transcription, respectively.

4.6.3 Altering cellular CGGBP1 levels impacts the level of chromatin-bound RNAPII complexes

CGGBP1 depletion led to dramatic changes in the RNA output of RNAPII transcribed genes with CGGBP1 binding sites (**Figure 24A**), but the mechanistic basis for this effect is unclear. One plausible explanation could be an effect on RNAPII elongation, which could affect the amount of RNAPII complexes on chromatin. To address this, I performed immunofluorescence imaging on pre-extracted cells targeting in each cell simultaneously chromatin-bound total RNAPII and the elongating form of RNAPII phospho-Ser2 (RNAPIIPS2) with specific antibodies. In addition, EdU pulse labeling was used to determine the population of S phase cells. CGGBP1-depleted cells exhibited a small but significant increase in both total chromatin-bound RNAPII and RNAPIIPS2 signals (**Figure 27A-C**). As expected, the RNAPII-specific transcription elongation inhibitor DRB dramatically reduced total RNAPII and RNAPIIPS2 signals, but again CGGBP1-depleted cells showed significantly more RNAPII chromatin occupancy compared to siControl cells (**Figure 27A-C**). To determine if RNAPIIPS2 is accumulating on chromatin, I calculated the ratio of the RNAPIIPS2 versus total RNAPII signal per cell. Interestingly, this ratio was increased in EdU (+) cells compared to EdU (-) cells in both control and CGGBP1 knockdown conditions (**Figure 27D**). Importantly, RNAPIIPS2 accumulated significantly more in CGGBP1-depleted than in the

control S phase cells (**Figure 27D**). Of note, the ratio is high in CGGBP1-depleted S phase cells, because of unchanged RNAPIIPS2 levels, but significantly less total RNAPII compared to the control S phase cells (**Figure 27E-F**). Together, these results indicate that CGGBP1 alleviates the accumulation of chromatin-bound elongating RNAPII during DNA replication. Importantly, no significant differences can be observed between control and CGGBP1-depleted cells in the RNAPIIPS2/RNAPII ratio upon DRB treatment, suggesting that this effect is transcription-dependent (**Figure 27D**).
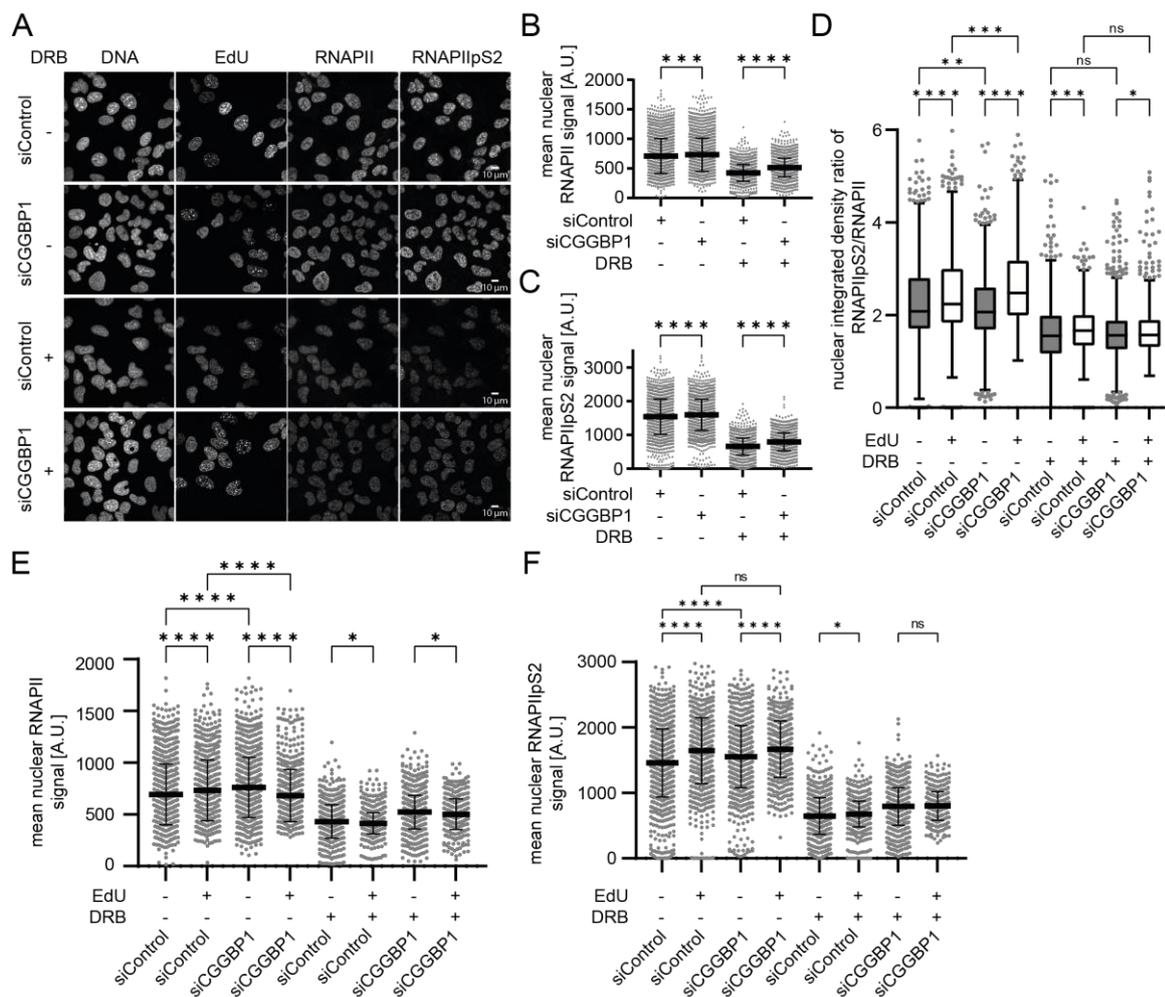
Figure 27: Chromatin-bound elongating RNAPII accumulates in S phase upon CGGBP1 depletion in U2OS cells. **A**, Example IF images of total RNAPII, RNAPIIPS2 and EdU incorporation upon treatment of U2OS cells with siControl or siCGGBP1 for 72 h. For transcriptional inhibition, 100 µM DRB were added 2 h before fixation. **B**, Quantification of mean nuclear total RNAPII signal in all fields similar to A. **C**, Quantification of mean nuclear RNAPIIPS2 signal in all fields similar to A. **D**, Box plot of ratios between RNAPIIPS2 and total RNAPII mean integrated

nuclear signal in the same cell calculated for EdU+ and EdU- cells in each condition of A. **E**, Quantification of mean nuclear total RNAPII signal in all fields similar to A. Cells of each condition are split into EdU- and EdU+. **F**, Quantification of mean nuclear RNAPIIPS2 signal signal in all fields similar to A. Cells of each condition are split into EdU- and EdU+. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, two wells with each >900 nuclei were analysed.

To assess if CGGBP1 overexpression has an impact on chromatin-bound total RNAPII and RNAPIIPS2 levels, I used a similar immunofluorescence imaging approach. However, to be able to track which cells were overexpressing HA-CGGBP1, it was technically not possible to measure both total RNAPII and RNAPIIPS2 signals in the same cell, resulting in two independent experiments with separate stainings (**Figure 28A-G**). Strikingly, CGGBP1 overexpression dramatically decreased chromatin-bound RNAPIIPS2 levels, while only a mild decrease was observed for total RNAPII (**Figure 28A-C**). Importantly, when splitting the HA-CGGBP1 transfected cell population in HA (+) and HA (-) cells, this decrease is specific to the HA (+) cells (**Figure 28D-E**), suggesting that this effect is specific to CGGBP1 overexpressing cells. Interestingly, RNAPIIPS2 levels are specifically reduced in non-S phase cells, suggesting that excess CGGBP1 does not impair the elongation of RNAPII into the gene body during replication (**Figure 28G**). In summary, these results provide further evidence that CGGBP1 is binding to promoter regions of RNAPII genes and a higher abundance of CGGBP1 at these promoter sites can physically block transcription elongation.
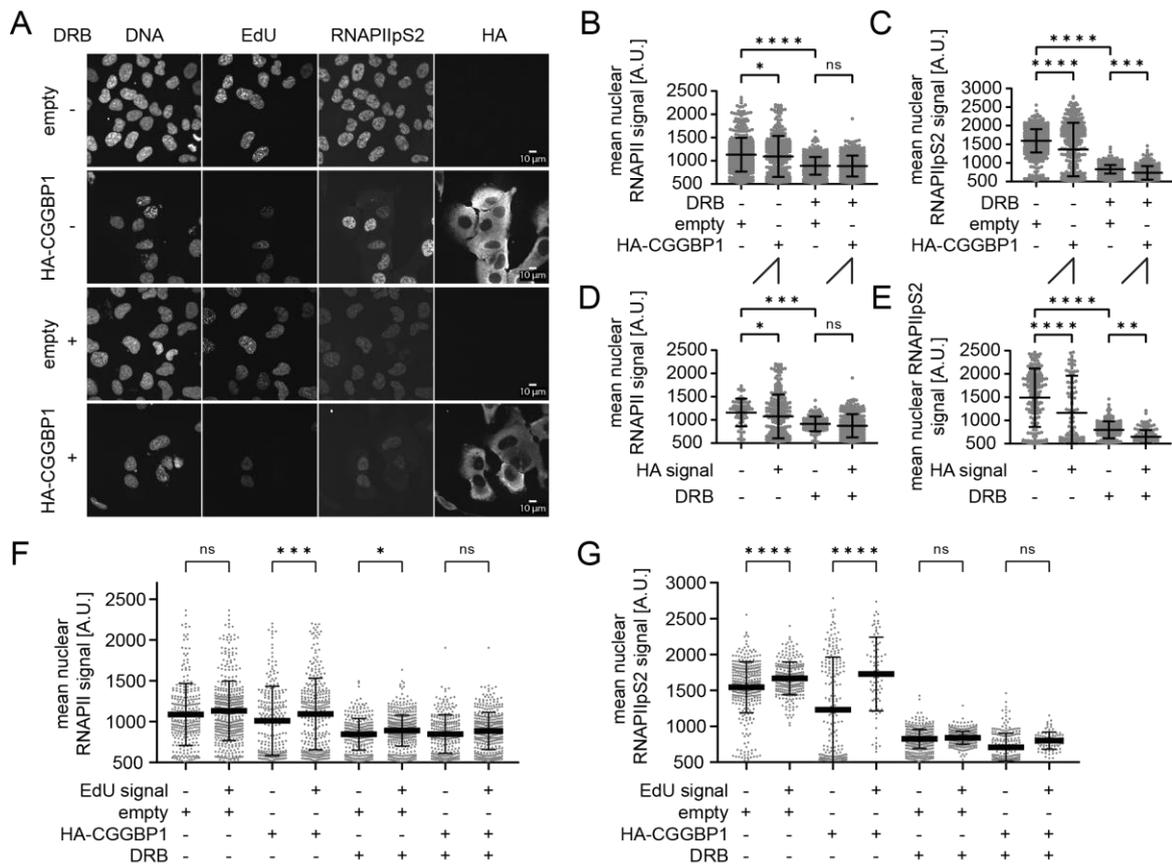
Figure 28: Overexpression of HA-CGGBP1 decreases chromatin-bound elongating RNAPII outside of S phase in U2OS cells. **A**, Example IF images of RNAPIIPS2 and EdU incorporation in U2OS cells upon HA-CGGBP1 overexpression compared to an empty overexpression plasmid control. For transcriptional inhibition, 100 μM DRB was added 2 h before fixation. **B**, Quantification of mean nuclear total RNAPII signal in in all conditions of A. **C**, Quantification of mean nuclear RNAPIIPS2 signal in all fields similar to A. **D**, Quantification of mean nuclear total RNAPII signal in all conditions of A. Only the overexpression condition is shown with cells split into HA+ or HA-. **E**, Quantification of mean nuclear RNAPIIPS2 signal in all fields similar to A. Only the overexpression condition is shown with cells split into HA+ or HA-. **F**, Quantification of mean nuclear total RNAPII signal in all fields similar to A. Cells of each condition are split into EdU- and EdU+. **G**, Quantification of mean nuclear RNAPIIPS2 signal in all fields similar to A. Cells of each condition are split into EdU- and EdU. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well with >1000 nuclei was analysed.

4.6.4 CGGBP1 levels are important to mitigate transcription-replication interference

As chromatin-bound RNAPIIPS2 accumulated in CGGBP1-depleted S phase cells (**Figure 27D**), I next investigated whether this may cause increased interference with replication. Strikingly, CGGBP1 knockdown led to a considerable increase in TRC-PLA foci in EdU (+) cells (**Figure 29A-B**). Furthermore, I could observe reduced EdU incorporation rates in CGGBP1-depleted S phase cells compared to the control, indicating a slowing of the replication fork (**Figure 29C**).
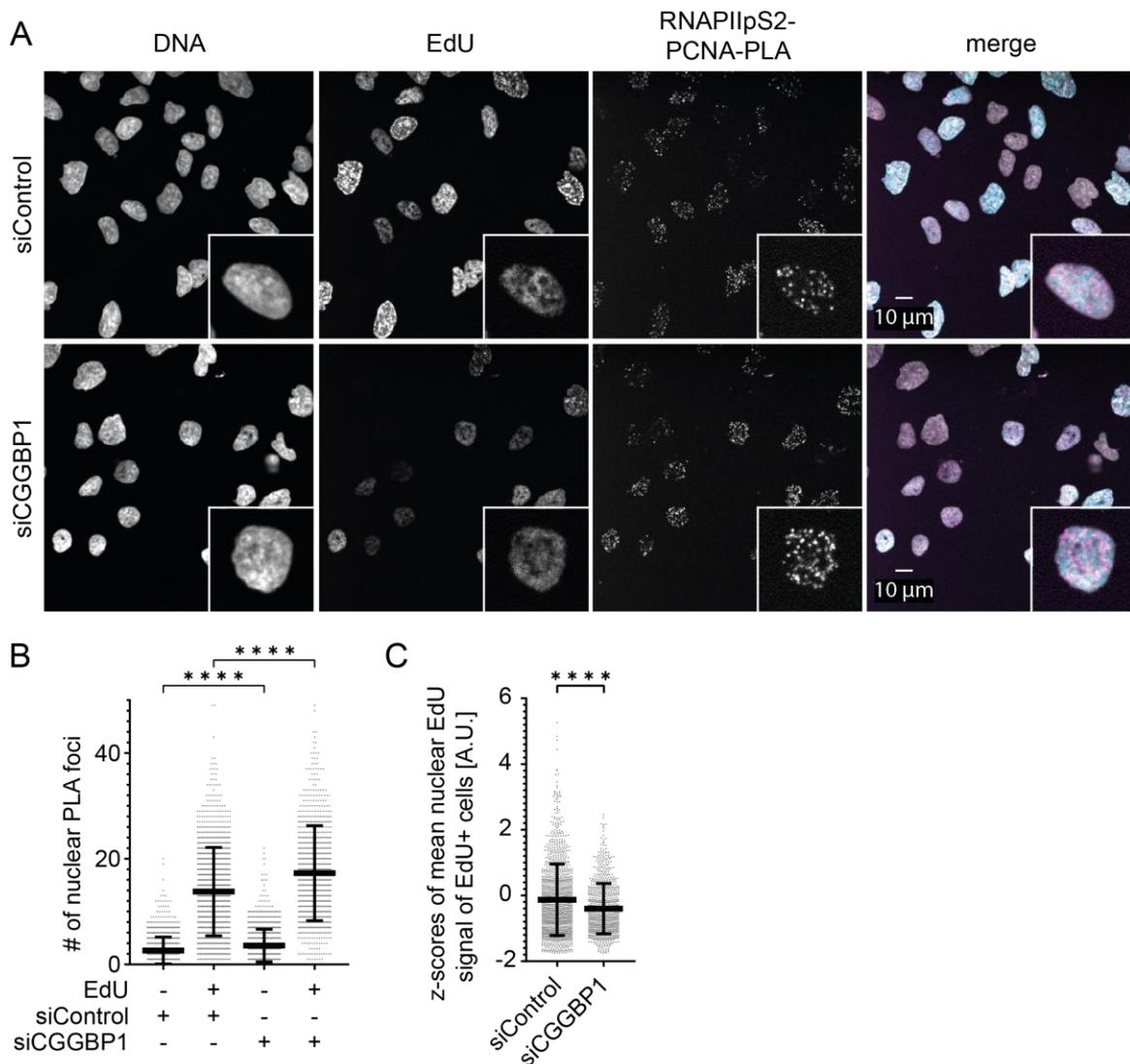
Figure 29: CGGBP1 depletion increases TRC-PLA foci frequency in U2OS. **A**, Example IF images of EdU incorporation and RNAPIIPS2 - PCNA proximity ligation assay foci in U2OS cells upon 48 h CGGBP1 knockdown compared to control siRNA. N=2. **B**, Quantification of nuclear PLA foci in EdU- and EdU+ cells in all fields similar to A. Data is represented as mean ± standard deviation. Statistical

significance was calculated using one-way ANOVA. **C**, Quantification of mean nuclear EdU signal of EdU+ cells in all fields similar to A. Data is represented as mean of the z-scores ± standard deviation. Statistical significance was calculated using two-tailed student's t-test. For each condition, three wells with each >100 EdU+ nuclei were analysed. Results of two biological repeats were pooled.

CGGBP1 preferentially binds to promoter sites (**Figure 21**) and knockdown leads to increased TRC-PLA foci and replication stress (**Figure 29**). To test whether CGGBP1 is really in proximity of transcription and replication, I used PLA with CGGBP1 as one of the targets. First, I tested whether CGGBP1 co-localizes with RNAPII phospho-Ser5 (RNAPII-pS5), a marker for promoter bound RNAPII. Indeed, there were significantly more CGGBP1-RNAPII-pS5 PLA foci compared to both single antibody controls regardless of S phase status, suggesting an interaction of CGGBP1 with RNAPII at promoters (**Figure 30A, D**). Second, I observed significantly more CGGBP1-PCNA PLA foci, indicating proximity of CGGBP1 to replication forks (**Figure 30B, D**). Lastly, I also tested whether CGGBP1 co-localizes with RPA32-pS33, a marker of stalled replication forks in conditions of increased replication stress (Vassin et al., 2009). Intriguingly, there were more CGGBP1-RPA32-pS33 PLA foci compared to the single antibody controls, indicating a proximity to stalled replication forks (**Figure 30C, D**).
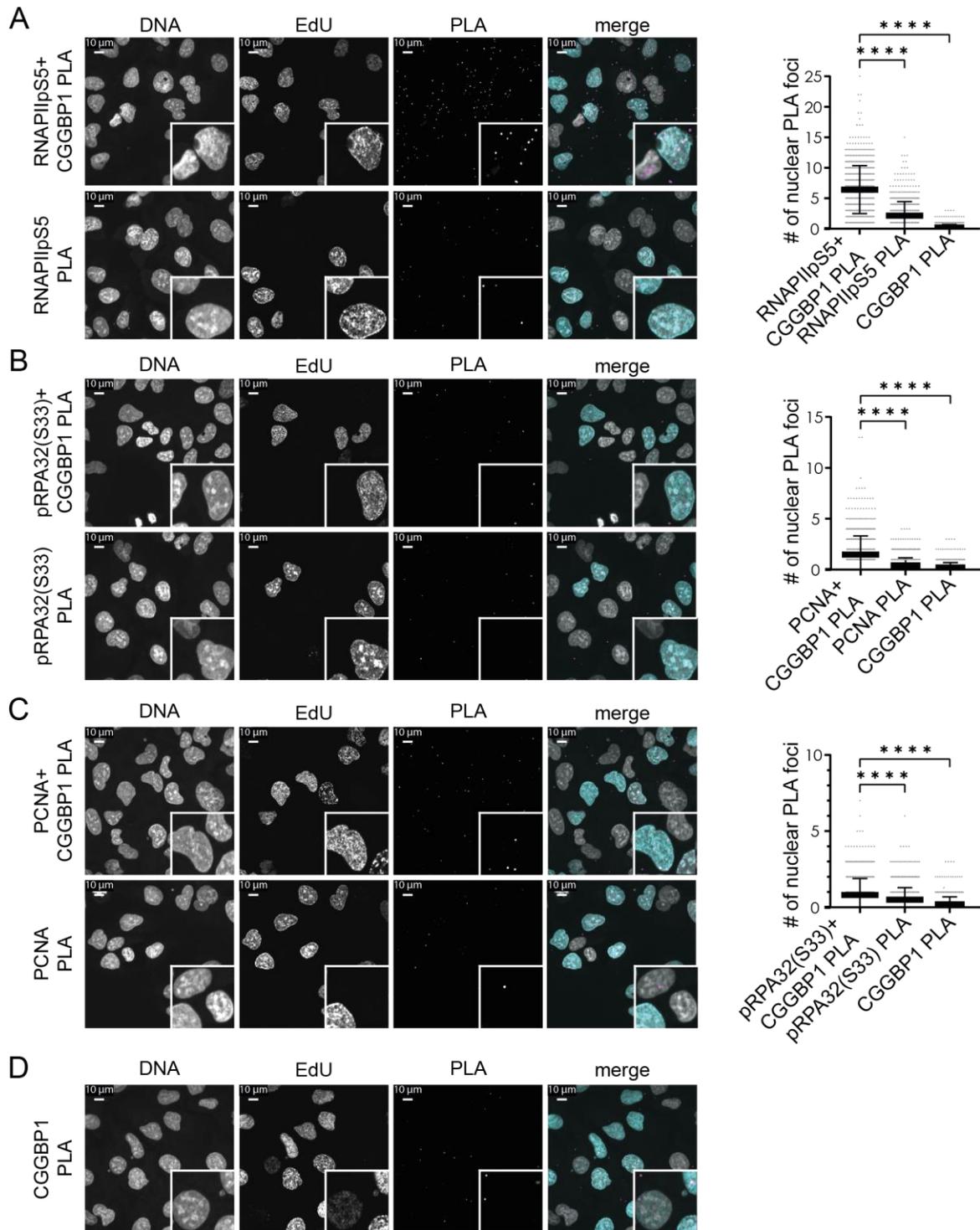
Figure 30: CGGBP1 is in proximity to replication, transcription and replication stress. **A**, Example IF images of EdU incorporation and proximity ligation assay foci of the RNAPII-pS5 – CGGBP1 antibody combination. The RNAPII-pS5 single antibody control is shown in the bottom panels. Quantification of PLA foci is shown on the right. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well

with >500 nuclei was analysed. **B**, Example IF images of EdU incorporation and proximity ligation assay foci of the PCNA – CGGBP1 antibody combination. The PCNA single antibody control is shown in the bottom panels. Quantification of PLA foci is shown on the right. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well with >500 nuclei was analysed. **C**, Example IF images of EdU incorporation and proximity ligation assay foci of the pRPA32(S33) – CGGBP1 antibody combination. The pRPA32(S33) single antibody control is shown in the bottom panels. Quantification of PLA foci is shown on the right. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, one well with >500 nuclei was analysed. **D**, Example IF images of EdU incorporation and proximity ligation assay foci of the CGGBP1 single antibody control. For comparison, the same quantification of this control is shown in **A**, **B**, and **C**.

4.6.5 Pyridostatin-induced DNA damage in S phase cells is dependent on cellular CGGBP1 levels

As depletion of CGGBP1 seemed to induce replication stress and interfered with RNAPII transcription, I hypothesized that CGGBP1 binding at promoter site CGG repeats could prevent the formation of DNA secondary structures such as G4 and R-loops. These structures are potent roadblocks for the replication fork and could explain increased replication stress and transcription-replication interference. To study this, I treated cells with pyridostatin (PDS), a small molecule that selectively binds to and stabilizes G-quadruplex structures independent of sequence variability (Müller et al., 2010; Rodriguez et al., 2008, 2012). Importantly, PDS is able to displace specific G4-binding proteins, as the binding affinity between G4s and PDS is very high. This also includes transcription factors, meaning PDS can interfere with gene expression (L. Li et al., 2020, 2021). This raised several questions for CGGBP1: Can PDS treatment affect the relationship between CGGBP1 and transcription? Does depletion of CGGBP1 lead to more PDS stabilized G4s and thus more replication stress and DNA damage? To test this, I performed immunofluorescence imaging of EU incorporation and γH2AX in

CGGBP1-depleted and control cells with or without PDS treatment. Interestingly, EU signal was dramatically increased in control cells upon addition of PDS (**Figure 31A-B**), which could indicate that PDS binds to endogenous G4s involved in recruiting transcription factors (Spiegel et al., 2021). Furthermore, addition of PDS to control cells led to an increase of γH2AX foci (**Figure 31A, C**), confirming the finding that PDS-stabilized G4s are more severe blocks for replication forks, leading to double-strand breaks (Broxson et al., 2011; De & Michor, 2011; Lemmens et al., 2015; Sarkies et al., 2010, 2012). Of note, CGGBP1 depletion without PDS treatment did not lead to an increase in γH2AX foci formation in U2OS cells, which is in contrast to reports in other cell lines (Singh et al., 2014). Surprisingly, depletion of CGGBP1 combined with PDS treatment mostly rescued these effects, with decreased EU signal and less γH2AX foci (**Figure 31A-C**). This EU signal reduction was mostly specific to non-nucleolar regions, indicating that the presence of CGGBP1 upon PDS treatment increases EU incorporation mainly in RNAPII transcribed regions (**Figure 31A, D-E**). In summary, these results suggested that if CGGBP1 is present, G4s stabilized by PDS are a more powerful obstacle for replication and transcription, resulting in more DNA damage.
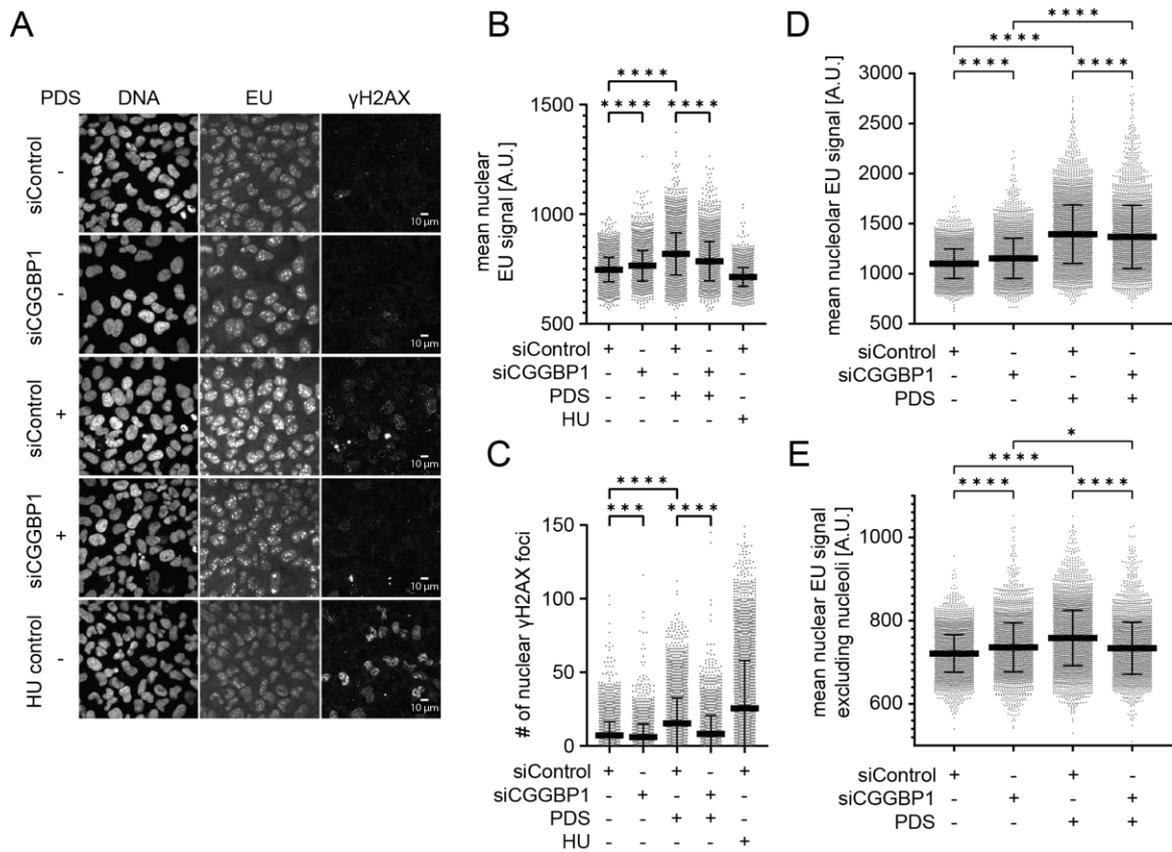
Figure 31: CGGBP1 depletion alleviates pyridostatin-induced DNA damage in U2OS cells. **A**, Example IF images of γH2AX signal and EU incorporation in U2OS cells upon 72 h CGGBP1 knockdown compared to control siRNA. Cells were treated with 20 μM pyridostatin (PDS) or with 500 μM hydroxyurea for 4h or 2h before fixation, respectively. **B**, Quantification of mean nuclear EU signal in all fields similar to A. The data of siControl and siCGGBP1 without treatment are already shown in Figure 23C, but plotted here again for direct comparison. **C**, Quantification of nuclear γH2AX foci in all fields similar to A. **D**, Quantification of mean nucleolar EU signal in all fields similar to A. **E**, Quantification of mean nuclear EU signal excluding nucleoli in all fields similar to A. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, two wells with each >1000 nuclei were analysed.

In order to test whether these effects are specific to DNA replication in S phase, I repeated the experiment with EdU incorporation instead of EU. In control cells, PDS treatment increased γH2AX foci mainly in EdU (+) cells, indicating that most PDS-induced DNA damage was present during S phase (**Figure 32A-B)**. In

support, EdU incorporation was lower in PDS treated control cells (**Figure 32C**). In the absence of CGGBP1 and upon PDS treatment, γH2AX foci formation was reduced especially in EdU (+) cells and EdU incorporation was restored to some extent. Together, these results indicate that PDS-stabilized G4s in the presence of CGGBP1 are severe impediments for replication forks, ultimately leading to excessive DNA damage.
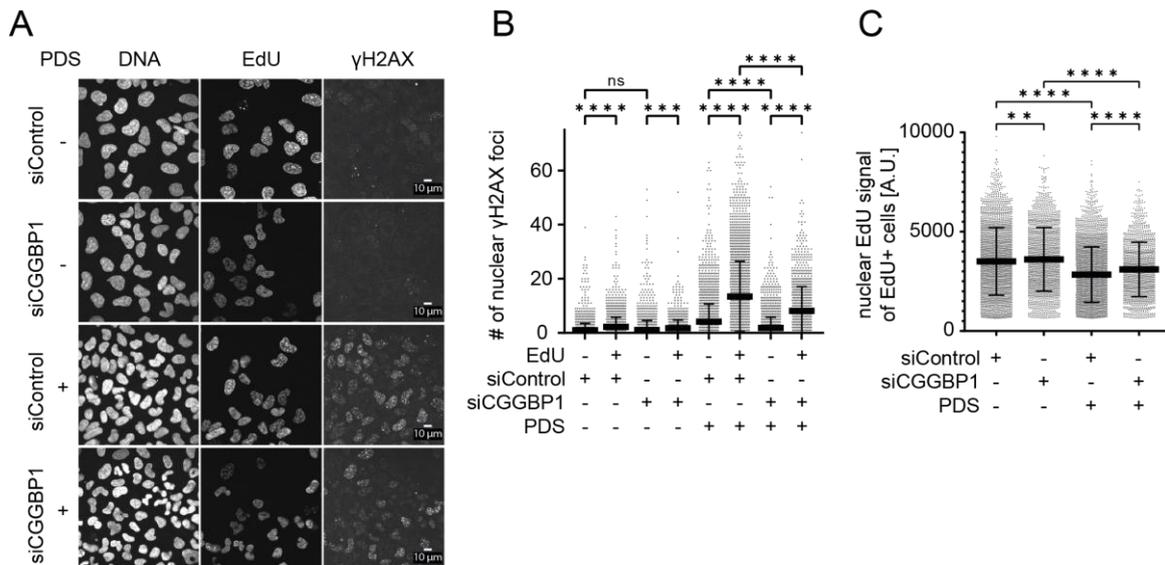
Figure 32: CGGBP1 depletion alleviates pyridostatin-induced DNA damage especially during S phase in U2OS cells. **A**, Example IF images of γH2AX signal and EdU incorporation in U2OS cells upon 72 h CGGBP1 knockdown compared to control siRNA. Cells were treated with 20 μM PDS for 4h before fixation. **B**, Quantification of nuclear γH2AX foci in all fields similar to A. Cells of the same condition are split into EdU- and EdU+. **C**, Quantification of mean nuclear EdU signal of EdU+ cells in all fields similar to A. Data is represented as mean ± standard deviation. Statistical significance was calculated using one-way ANOVA. For each condition, three wells with each >900 nuclei were analysed.

4.6.6 R-loops are enriched at a CGG-repeat containing episomal transcription unit after CGGBP1 depletion

To study how CGGBP1 binding influences the formation of co-transcriptional R-loops, I used a previously established episomal system for studying TRCs (Hamperl et al., 2017). To mimic endogenous CGGBP1-binding, I cloned 10 CGG

repeats ($(CGG)_{10}$) 30 bp downstream of the doxycycline (DOX) inducible Tet-ON promoter (**Figure 33A**). In addition, this plasmid contains an unidirectional Epstein-Barr virus replication origin oriP orientated head-on to transcription, because robust R-loop formation was mainly observed in this orientation and not in co-directional TRCs (Hamperl et al., 2017). I transfected this construct into U2OS cells and generated stable monoclonal cell lines with each clone containing ~5-70 plasmid copies per cell (**Figure 33B**). After expansion, clone 2 maintained ~30 copies during the passaging of untreated cells (**Figure 33C**). To test the involvement of CGGBP1 in TRCs, I depleted CGGBP1 by siRNA treatment and simultaneously induced transcription of the episomal $(CGG)_{10}$ transcription unit with DOX (**Figure 33D**). Surprisingly, plasmid copy numbers did not significantly change upon DOX induction or CGGBP1 depletion (**Figure 33E**), indicating no strong increase in replication stress, which was not the case in the previous study using a strong R-loop forming transcription unit (Hamperl et al., 2017). As expected, in control cells the transcriptional output measured by RT-qPCR was DOX dose-dependent (**Figure 33F**). However, upon CGGBP1 depletion, transcription was greatly reduced at a high DOX concentration compared to control cells (**Figure 33F**), indicating that CGGBP1 is necessary for faithful RNAPII transcription through the $(CGG)_{10}$ repeat.

In order to test whether this reduced transcriptional activity can be explained by the formation of R-loops on the $(CGG)_{10}$ repeat in the absence of CGGBP1, I performed DNA-RNA immunoprecipitation (DRIP) and qPCR at the $(CGG)_{10}$ repeat with the S9.6 antibody, which is specific to RNA-DNA hybrids (Boguslawski et al., 1986). To confirm the DRIP signal, RNase H was added as a control, because it specifically degrades RNA:DNA hybrids. In control cells, RNA-DNA hybrid formation was not significantly increased upon transcriptional induction via DOX (**Figure 33G**). However, compared to the control cells, the DRIP signal was significantly higher in CGGBP1-depleted cells upon DOX induction (**Figure 33G**). Furthermore, this RNA:DNA hybrid increase could be observed despite a lower transcriptional output under this condition (**Figure 33F**). Together, these results indicate a protective role for CGGBP1 against the formation of RNA:DNA hybrids at short CGG repeats during transcription and consequently ensuring proper mRNA production.
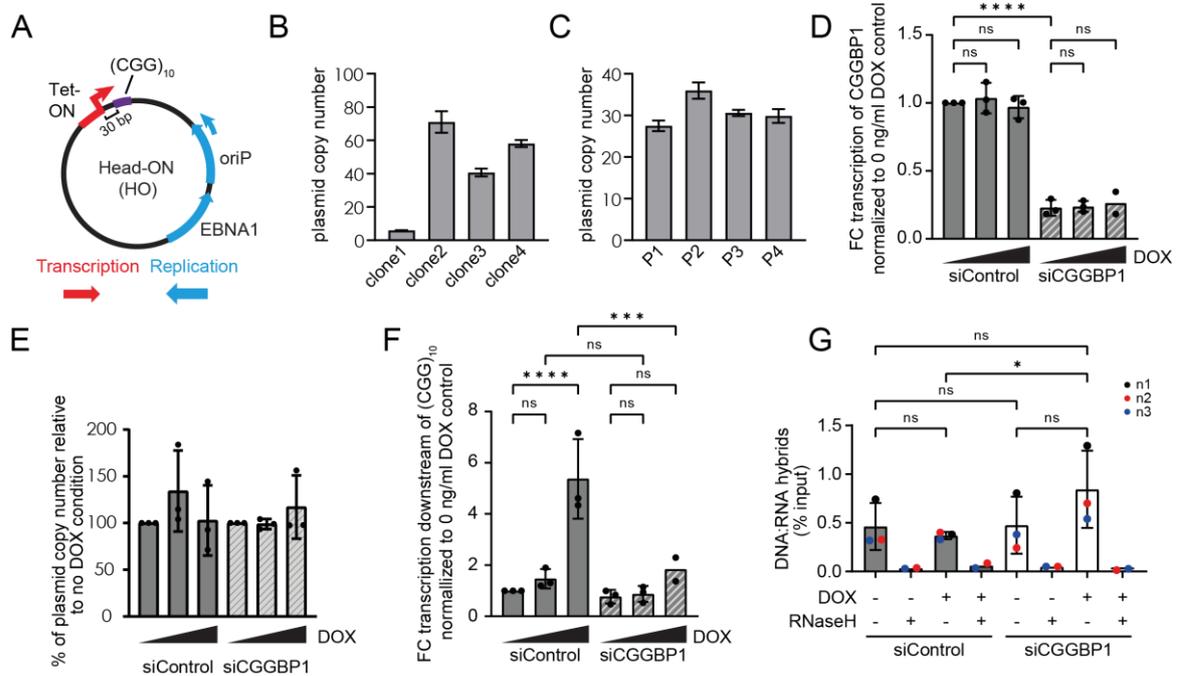
Figure 33: DNA:RNA hybrids are enriched at a CGG repeat containing episomal transcription unit upon CGGBP1 depletion in U2OS cells. **A**, Plasmid map of the episomal transcription-replication conflict system with 10 CGG repeats inserted at the Tet-ON promoter. EBNA1 forces unidirectional replication at the oriP. Doxycycline treatment induces transcription leading to head-on conflicts with replication. **B**, Initial plasmid copy numbers of different U2OS Tet-ON monoclonal cell lines carrying the episomal system measured by qPCR. Plasmid copy numbers were calculated with the relative ratio 2xOriP/β-actin to account for the diploid reference gene copy number. Data is represented as mean ± standard deviation of three technical qPCR replicates. **C**, Plasmid copy numbers of U2OS Tet-ON pHU43 clone 2 during passaging measured by qPCR. The time between each passage was 3-4 days. DOX and siRNA treatment experiments were done between passages 1 and 4. Data is represented as mean ± standard deviation of three technical qPCR replicates. **D**, Gene expression of CGGBP1 measured by RT-qPCR of cDNA from U2OS Tet-ON pHU43 clone 2 carrying the episomal system and treated with siControl or siCGGBP1 for 72 h. Cells were treated with 0, 100 or 1000 ng/ml DOX for 72 h. Shown is the fold change relative to MCM3 and normalized to the 0 ng/ml DOX control. Data is represented as mean ± standard deviation of three (except for the siCGGBP1 1000 ng/ml condition) biological replicates. **E**, Plasmid copy number changes during DOX treatment of U2OS Tet-ON pHU43 clone 2. Data is represented as mean ± standard deviation

of three biological replicates. **F**, Gene expression downstream of $(CGG)_{10}$ measured by RT-qPCR of cDNA from U2OS Tet-ON pHU43 clone 2 carrying the episomal system and treated with siControl or siCGGBP1 for 72 h. Cells were treated with 0, 100 or 1000 ng/ml DOX for 72 h. Shown is the fold change relative to MCM3 and normalized to the 0 ng/ml DOX siControl condition. Data is represented as mean ± standard deviation of three biological replicates (except for two biological replicates for the siCGGBP1 1000 ng/ml condition). **G**, DNA:RNA hybrid immunoprecipitation (DRIP) with the S9.6 antibody downstream of the $(CGG)_{10}$ repeat measured by qPCR of U2OS Tet-ON pHU43 clone 2 treated with siControl or siCGGBP1 and 0 or 1000 ng/ml DOX for 72h. Data is represented as mean ± standard deviation of three biological replicates (except for two biological replicates for RNaseH controls). Statistical significance was calculated using one-way ANOVA.

5 Discussion

5.1 Development of an unbiased approach to identify factors involved in transcription-replication conflicts

5.1.1 Optimization of MCM2-APEX2 epxression and proximity labeling in MCF7 cells

Integration of MCM2-APEX2 and APEX2-NLS expression cassettes via sleeping beauty transposase into MCF7 cells was successful and allowed DOX-dependent regulation of expression levels (**Figure 9A-B**). However, it is important to note that the endogenous MCM2 alleles were not replaced or deleted in the cell line, resulting in co-expression and therefore potential competition between the wildtype and APEX2-tagged versions of the protein. This will likely result in a mixed situation in cells where some replication origins will incorporate the wildtype MCM2 allele and others the MCM2-APEX2 fusion protein into the MCM2-7 double hexamer during replication origin licensing (Costa & Diffley, 2022). Therefore, only a fraction of replication forks will contain the APEX2 moiety allowing for the proximity labeling reaction on active replication forks, whereas excess MCM2-APEX2 fusion protein might diffuse freely in the nucleoplasm similar to the APEX2-NLS fusion construct and therefore increase the background of the proximity labeling reaction. Indeed, nuclear fractionation showed incorporation of MCM2-APEX2 in the chromatin fraction, but also in all other fractions including the cytoplasmic fraction (**Figure 9E**). However, immunofluorescence imaging of biotinylation by proximity labeling confirmed nuclear labeling (**Figure 9A**). Initial mass spectrometry results of MCM2-APEX2 proximity labeling showed a clear enrichment of other members of the MCM2-7 complex compared to the APEX2-NLS control, especially when APEX2-NLS expression was controlled at a low level (**Figure 10B-C**). These results all indicated that at least a fraction of MCM2-APEX2 is functionally integrated into the MCM2-7 complex and therefore into active replication forks during S phase. Nevertheless, future optimizations of this system should include endogenous APEX2 tagging of the MCM2 endogenous allele to overcome this limitation and reduce the background of the proximity labeling approach.

5.1.2 Induction of unscheduled Transcription-Replication Conflicts in MCF7 cells

To provoke increased collisions between transcription and replication, I used estrogen stimulation of MCF7 cells that showed clear transcription- and replication-dependent DNA damage in S phase cells (Stork et al., 2016). Thus, the idea was to induce an acute estrogen-mediated transcriptional burst while the MCF7 cells simultaneously enter S phase, thereby likely causing unscheduled transcription-replication interference. A major difference to the previous study (Stork et al., 2016) was that cells were only treated for 30 min with estrogen compared to the published treatment duration of 16 hours. Interestingly, this short pulse is able to influence transcription drastically at thousands of genes across the MCF7 genome (Hah et al., 2011). Consistently, EU-Seq data did show major changes in the transcriptional landscape (**Figure 12**). Importantly, many known estrogen-responsive genes such as GREB1 and IGFBP4 were upregulated under these conditions (**Figure 12** and Hah et al., 2011; West et al., 2016), suggesting successful estrogen stimulation of the MCF7 cells. However, the number of TRC-PLA foci did not further increase upon estrogen stimulation (**Figure 13**), indicating no increase in the relative frequency of TRCs between estrogen-stimulated and DMSO-treated control cells. A reasonable explanation for this result is the fact that the transcriptomic analysis revealed a similar number of up- and downregulated genes (**Figure 12**), suggesting that the overall transcriptional output of the cell has not increased, but rather the precise locations of active transcription has changed, thereby not changing the frequency but rather the genomic locations of TRCs in an unscheduled and acute manner. On the contrary, treatment of cells with the transcription inhibitor DRB strongly reduced overall TRC frequency in estrogen-treated cells (**Figure 12**), which provided the best control condition for the proximity labeling experiments to determine the proteomic composition of replication forks encountering transcription complexes.

5.1.3 Transcription-Replication Conflict Proximity Labeling with MCM2-APEX2

In the proximity labeling experiment with the MCM2-APEX2 and APEX2-NLS expressing MCF7 cell lines, thousands of proteins were identified in all four conditions (**Figure 14B**). This is rather a high number of identified proteins for

proximity labeling experiments (Ummethum & Hamperl, 2020). One explanation could be the excess number of MCM2-7 complexes on chromatin, which consequently leads to higher biotinylation levels. However, APEX2-NLS control samples similarly contributed to this high number of identified proteins, which could be a result of the overexpression system and resulting pan-nuclear biotinylation. This could also explain the poor enrichment of replication factors when comparing the MCM2-APEX2 vs. APEX2-NLS samples (**Figure 15**). Alternatively, as mentioned above, a high fraction of non chromatinized, nucleoplasmic MCM2-APEX2 could also contribute to this as it would lead to biotin labeling of proteins not associating with replication forks similar to the samples with free diffusing APEX2-NLS. Lastly, the high overlap in identified proteins between the four conditions can also arise from unspecific interactions with the streptavidin beads. Despite these technical limitations, the direct comparison of estrogen-treated versus DRB-treated cells revealed 88 candidate factors that were specifically enriched at transcriptionally challenged replication forks (**Figure 14C**, **Figure 16** and **Figure 17**). Importantly, several of these factors have been previously linked to TRCs such as the topoisomerase TOP2A, the R-loop helicase DDX39B as well as the chromatin reader BRD4 (Yeo et al., 2016; Pérez-Calero et al., 2020; F. C. Lam et al., 2020), therefore validating the approach. As part of this thesis, I have followed up on 13 candidate factors in this list by siRNA-mediated knockdown experiments (**Table 2**). Initial depletion experiments revealed a role of multiple candidates in protecting against DNA damage measured by γH2AX foci formation (**Figure 19A-B**). In addition, significant changes in TRC-PLA foci formation pointed to an involvement in TRC resolution for several candidates (**Figure 19C-D**). A follow-up experiment revealed a role of WAC, RBBP6, TASOR and CGGBP1 in protecting against S phase specific DNA damage as measured by FANCD2 levels as a marker of replication stress (**Figure 20A-B**). As CGGBP1 showed most consistently an effect among these assays, a potential role of this protein in transcription-replication conflicts was further analyzed by functional assays in the second part of this thesis.

5.2 CGGBP1

5.2.1 Global profiling of CGGBP1 binding sites in the human genome

As CGGBP1 has initially been discovered as a protein binding to CGG repeats at the FMR1 locus, a major question has been, where else CGGBP1 binds in the genome. To answer this, I analyzed publicly available ChIP-Seq data from the ENCODE project (Luo et al., 2020; The ENCODE Project Consortium, 2012). This analysis revealed binding of CGGBP1 to the promoter regions (<1 kb away from TSS) of thousands of RNAPII genes with short CGG repeat tracts as an enriched binding motif (**Figure 21**). However, there are also binding sites with no such binding motif, indicating that CGGBP1 does not exclusively bind to CGG repeats. Interestingly, an independent CGGBP1 ChIP-Seq experiment did not show such a strong binding preference to gene promoters and did not show apparent CGG repeat binding preference (Agarwal et al., 2014). It is unclear where this discrepancy from the two datasets comes from, but one plausible explanation is the difference in antibodies used, as high-quality ChIP-grade antibodies are crucial for reliable ChIP-Seq data. The ENCODE project uses standardized protein tags with respective established antibodies for the ChIP-Seq studies (Luo et al., 2020; The ENCODE Project Consortium, 2012), whereas, a mixture of polyclonal CGGBP1 antibodies was used in the second study, which might introduce a large number of false positives. In addition, these two ChIP-Seq data sets were obtained from different cell lines, further complicating side-by-side comparison of the two datasets. Interestingly, a recent study confirmed the CGG repeat binding specificity of human CGGBP1 by a universal protein-binding microarray of all possible DNA 10-mers (Yellan et al., 2021). Therefore, it is most likely that CGGBP1 primarily binds to CGG motifs, but the binding of CGGBP1 to chromatin might be influenced by other chromatin characteristics *in vivo*, including a potential interaction with nucleosomes or other chromatin-binding interaction partners. Previous research has shown that CGGBP1 is a member of a heterotrimeric complex with NFIX and the high-mobility group protein HMGN1 that regulates HSF1 transcription bidirectionally (Singh et al., 2009). Consequently, some binding partners of CGGBP1 could facilitate further associations with specific chromatin features, potentially clarifying the discrepancy between these findings. Nevertheless, the

concentration of CGGBP1 at gene promoters was evident, regardless of its exact targeting mechanism.

5.2.2 Altering cellular CGGBP1 levels leads to global transcriptional changes and impacts the level of chromatin-bound RNAPII complexes

Prior research has shown that higher levels of DNA secondary structures, such as G4s and co-transcriptional RNA:DNA hybrids can hinder both RNA and DNA polymerases, leading to difficulties in transcription elongation and replication fork stalling (Belotserkovskii & Hanawalt, 2015; Gan et al., 2011; Groh et al., 2014; Tous & Aguilera, 2007). In agreement with this, my findings indicate that cells lacking CGGBP1 have an accumulation of active RNAPII on chromatin that is dependent on both transcription and S phase (**Figure 27** & **Figure 28**). This accumulation results in replication fork impediment and an increase in transcription-replication interference, as assessed by a TRC-PLA assay (**Figure 29** & **Figure 30**). I noticed a rise in nascent transcription, as determined by EU incorporation in cells lacking CGGBP1 (**Figure 23B-C**), which is challenging to reconcile with my hypothesis of impaired RNAPII transcription elongation during S phase. One potential explanation is that CGGBP1 also regulates genes transcribed by other RNA polymerases. Notably, ribosomal RNA (rRNA) genes transcribed by RNAPI have a high concentration of CGG repeats, and CGGBP1 binds preferentially to rRNA gene clusters located on the small arms of acrocentric chromosomes (Müller-Hartmann et al., 2000). This is in line with the observation that the EU signal increases in nucleoli following CGGBP1 depletion (**Figure 31D-E**). Furthermore, CGGBP1 was discovered to bind to RNAPIII promoters, effectively inhibiting transcription, particularly at repetitive Alu-SINE elements (Agarwal et al., 2014). As a result, the general increase in EU incorporation detected in cells lacking CGGBP1 may, in part, be attributed to alterations in the transcription dynamics of these other two RNA polymerases. On the other hand, an increase in CGGBP1 expression resulted in a general decline in EU incorporation, as well as a reduction in chromatin-bound RNAPII (**Figure 28**). This suggests that elevated CGGBP1 levels have an inhibitory impact, most likely on all nuclear RNA polymerases. Taken together, these findings imply that maintaining an appropriate balance of CGGBP1 within cells is critical for achieving efficient nuclear transcription rates. However, the precise mechanism by which CGGBP1

governs the transcription elongation rates of different RNA polymerases is likely intricate and may also be due to secondary effects resulting from cell cycle arrest or DNA damage checkpoint activation in CGGBP1-depleted cells.

5.2.3 CGGBP1 mitigates TRCs and prevents G4/R-loop formation

One main objective of my thesis was to assess the role of CGGBP1 in transcription-replication conflicts. Indeed, upon depletion of CGGBP1, the number of TRC-PLA foci increased in U2OS cells, indicating a protective function of CGGBP1 (**Figure 29**). As mentioned, TRCs can be aggravated by non-B DNA secondary structures. Consistently, CGGBP1 binds to genomic regions of higher G4 forming propensity and higher R-loop levels, indicating that CGGBP1 plays a role in preventing DNA secondary structure formation. Interestingly, the stabilization of G4 structures by PDS led to replication impediment and DNA damage in S phase cells, which was partly alleviated by CGGBP1 depletion (**Figure 31** & **Figure 32**). This remarkable finding establishes a clear link between CGGBP1 levels and DNA damage caused by G4 structures. Intriguingly, CGGBP1 features a C-terminal SQ motif that has been identified as a target of the ATR DNA damage response kinase (Matsuoka et al., 2007; Singh & Westermark, 2015, p. 1; Traven & Heierhorst, 2005). Consequently, CGGBP1 may play a functional role in the DNA checkpoint signaling of G4-induced DNA damage, although further research will be necessary to comprehend the mechanistic aspects of this interaction. Intriguingly, a recent study determining the genomic patterns of transcription-replication conflict sites in mouse primary B cells by a sequential pull-down approach of active transcription and replication fork associating DNA found striking similarities to CGGBP1 binding sites (St Germain et al., 2022). In particular, they found an increase in GC%, G4 quadruplex formation propensity as well as the sequence motif GGCGGCGG, further supporting an involvement of short CGG motifs at sites of transcription-replication conflicts.

To assess whether CGGBP1 also protects against the formation of R-loops, I utilized an episomal plasmid system with a short transcription-inducible CGG repeat and showed that CGGBP1 depletion blocks transcription and leads to the formation of RNA-DNA hybrid structures on the repeat sequence (**Figure 33**). Based on my results, I propose a model in which specific cellular factors such as CGGBP1 maintain accurate transcription and replication programs and prevent the

formation of genome-destabilizing secondary structures on short trinucleotide repeats (**Figure 34**). Finally, in addition to the increased DRIP-signal at CGGBP1 binding sites and the results of the episomal system, there is further evidence for a link between R-loops and CGGBP1. In particular, CGGBP1 is spatially associating with R-loops as identified via proximity labeling with catalytically inactive RNase H (Yan et al., 2022).
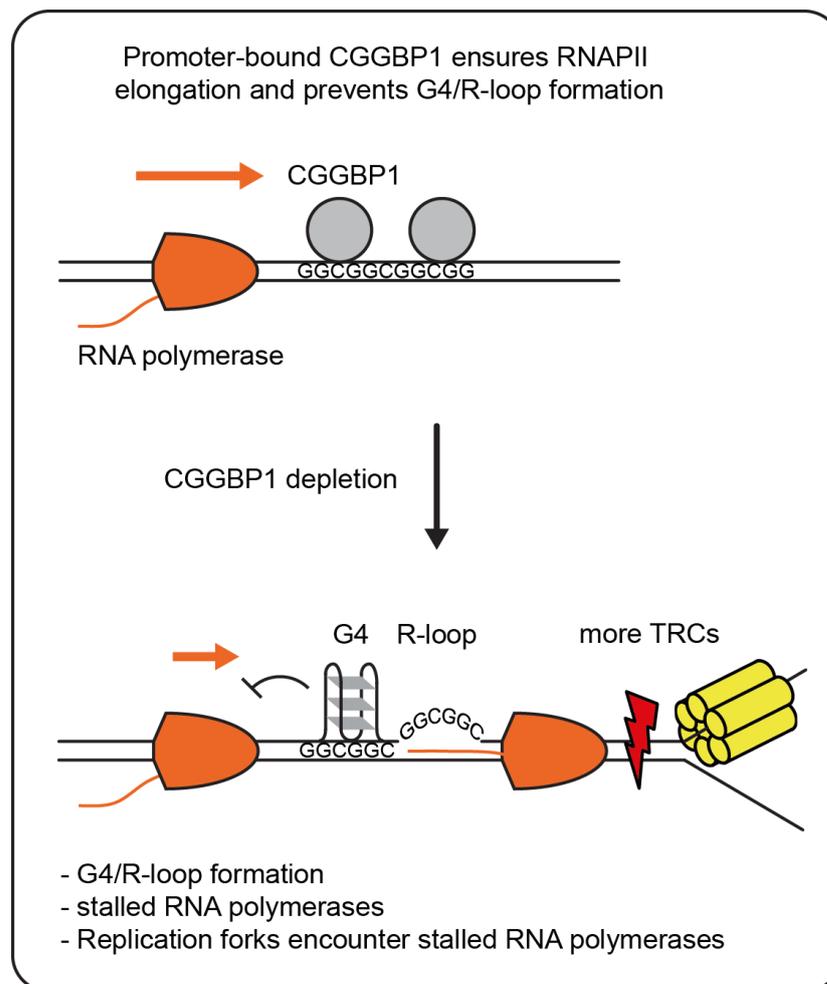
Figure 34: Model of how CGGBP1 binding at short CGG repeats opposes secondary structure formation and promotes transcription elongation. The presence of CGGBP1 at promoter sites prevents the formation of DNA secondary structures capable of blocking RNA polymerases and replication forks. Depletion of CGGBP1 leads to the formation of DNA secondary structures resulting in the accumulation of stalled RNA polymerases. In S-phase, replication forks encounter the stalled RNA polymerases, leading to unscheduled transcription-replication conflicts and replication impairment.

Finally, I want to address the important limitation of this work that most conclusions on CGGBP1 were drawn from experiments conducted in a single cell line (U2OS). This could limit the generality of these conclusions. As an example, CGGBP1 depletion led to an increase in TRC-PLA in U2OS cells (**Figure 29**), but a decrease in MCF7 cells (**Figure 19C-D**). The cause of this discrepancy needs more future investigations, since CGGBP1 is known to bind preferentially to unmethylated DNA (Deissler et al., 1996; Varley et al., 2013), one potential explanation could stem from distinct DNA methylation patterns in the two cell lines. The rationale to focus on U2OS cells as the primary cell line in this research project was mostly for technical limitations: First, setting up the episomal plasmid system containing a short transcription-inducible CGG repeat in MCF7 cells was not achieved and second, high-throughput immunofluorescence imaging proved to be very difficult with the MCF7 cell line. However, both experimental approaches provided important conclusions in the U2OS cell line and more work is clearly needed to extend my findings into other cancer or non-cancerous cell lines.

In summary, the primary aim of this thesis was to develop an unbiased approach to identify factors involved in transcription-replication conflicts. I successfully developed a proximity labeling system to investigate the proteomic composition of active replication forks while inducing unscheduled transcription-replication conflicts. As a result, I found a list of candidate proteins that were enriched at transcriptionally challenged replication forks. The secondary aim was to investigate these potential candidates in the context of transcription-replication conflicts. For this, I focused on the CGG-triplet repeat binding protein CGGBP1. I discovered that CGGBP1 primarily binds to CGG repeats at promoters and prevents RNA polymerase II accumulation on chromatin. Intriguingly, CGGBP1 depletion led to an increase in RNA polymerase II – PCNA proximity ligation assay foci in U2OS cells, indicating elevated levels of transcription-replication conflicts in CGGBP1-deficient cells. Furthermore, CGGBP1 depletion led to DNA:RNA hybrid formation, suggesting that this factor counteracts DNA secondary structure formation at CGG repeats, thereby preventing RNAPII accumulation and resulting transcription-replication conflicts. Until now, most factors known to play a role in transcription-replication conflicts are either traveling with the replication fork or transcription

machinery. However, here I uncover that CGGBP1 is preventing DNA secondary structure formation and TRCs by binding to fixed locations on the chromatin. This introduces a new mechanism of transcription-replication conflict prevention that can be explored with similar DNA binding proteins in future studies.

References

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379

Agarwal, P., Collier, P., Fritz, M. H.-Y., Benes, V., Wiklund, H. J., Westermark, B., & Singh, U. (2015). CGGBP1 mitigates cytosine methylation at repetitive DNA sequences. *BMC Genomics*, *16*(1), 390. https://doi.org/10.1186/s12864-015-1593-2

Agarwal, P., Enroth, S., Teichmann, M., Jernberg Wiklund, H., Smit, A., Westermark, B., & Singh, U. (2014). Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. *Cell Cycle*, *15*(12), 1558–1571. https://doi.org/10.4161/15384101.2014.967094

Akamatsu, Y., & Kobayashi, T. (2015). The Human RNA Polymerase I Transcription Terminator Complex Acts as a Replication Fork Barrier That Coordinates the Progress of Replication with rRNA Transcription Activity. *Molecular and Cellular Biology*, *35*(10), 1871–1881. https://doi.org/10.1128/MCB.01521-14

Azvolinsky, A., Giresi, P. G., Lieb, J. D., & Zakian, V. A. (2009). Highly transcribed RNA polymerase II genes are impediments to replication fork progression in Saccharomyces cerevisiae. *Molecular Cell*, *34*(6), 722–734. https://doi.org/10.1016/j.molcel.2009.05.022

Barker, D. F., & Campbell, A. M. (1981a). Genetic and biochemical characterization of the birA gene and its product: Evidence for a direct role of biotin holoenzyme synthetase in repression of the biotin operon in Escherichia coli. *Journal of Molecular Biology*, *146*(4), 469–492. https://doi.org/10.1016/0022-2836(81)90043-7

Barker, D. F., & Campbell, A. M. (1981b). The birA gene of Escherichia coli encodes a biotin holoenzyme synthetase. *Journal of Molecular Biology*, *146*(4), 451–467. https://doi.org/10.1016/0022-2836(81)90042-5

Barlow, J. H., Faryabi, R. B., Callén, E., Wong, N., Malhowski, A., Chen, H. T., Gutierrez-Cruz, G., Sun, H.-W., McKinnon, P., Wright, G., Casellas, R., Robbiani, D. F., Staudt, L., Fernandez-Capetillo, O., & Nussenzweig, A. (2013). Identification of early replicating fragile sites that contribute to genome instability. *Cell*, *152*(3), 620–632. https://doi.org/10.1016/j.cell.2013.01.006

Bayona-Feliu, A., Barroso, S., Muñoz, S., & Aguilera, A. (2021). The SWI/SNF chromatin remodeling complex helps resolve R-loop-mediated

transcription–replication conflicts. *Nature Genetics*, *53*(7), Article 7. https://doi.org/10.1038/s41588-021-00867-2

Beckett, D., Kovaleva, E., & Schatz, P. J. (1999). A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Science: A Publication of the Protein Society*, *8*(4), 921–929. https://doi.org/10.1110/ps.8.4.921

Bedinger, P., Hochstrasser, M., Jongeneel, C. V., & Alberts, B. M. (1983). Properties of the T4 bacteriophage DNA replication apparatus: The T4 dda DNA helicase is required to pass a bound RNA polymerase molecule. *Cell*, *34*(1), 115–123. https://doi.org/10.1016/0092-8674(83)90141-1

Belotserkovskii, B. P., & Hanawalt, P. C. (2015). PNA binding to the non-template DNA strand interferes with transcription, suggesting a blockage mechanism mediated by R-loop formation. *Molecular Carcinogenesis*, *54*(11), 1508–1512. https://doi.org/10.1002/mc.22209

Bermejo, R., Capra, T., Jossen, R., Colosio, A., Frattini, C., Carotenuto, W., Cocito, A., Doksani, Y., Klein, H., Gómez-González, B., Aguilera, A., Katou, Y., Shirahige, K., & Foiani, M. (2011). The Replication Checkpoint Protects Fork Stability by Releasing Transcribed Genes from Nuclear Pores. *Cell*, *146*(2), 233–246. https://doi.org/10.1016/j.cell.2011.06.033

Bermejo, R., Doksani, Y., Capra, T., Katou, Y.-M., Tanaka, H., Shirahige, K., & Foiani, M. (2007). Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation. *Genes & Development*, *21*(15), 1921–1936. https://doi.org/10.1101/gad.432107

Bhaskara, S., Jacques, V., Rusche, J. R., Olson, E. N., Cairns, B. R., & Chandrasekharan, M. B. (2013). Histone deacetylases 1 and 2 maintain S-phase chromatin and DNA replication fork progression. *Epigenetics & Chromatin*, *6*(1), 27. https://doi.org/10.1186/1756-8935-6-27

Bhowmick, R., Lerdrup, M., Gadi, S. A., Rossetti, G. G., Singh, M. I., Liu, Y., Halazonetis, T. D., & Hickson, I. D. (2022). RAD51 protects human cells from transcription-replication conflicts. *Molecular Cell*, *82*(18), 3366-3381.e9. https://doi.org/10.1016/j.molcel.2022.07.010

Bianchi, J., Rudd, S. G., Jozwiakowski, S. K., Bailey, L. J., Soura, V., Taylor, E., Stevanovic, I., Green, A. J., Stracker, T. H., Lindsay, H. D., & Doherty, A. J. (2013). PrimPol bypasses UV photoproducts during eukaryotic chromosomal DNA replication. *Molecular Cell*, *52*(4), 566–573. https://doi.org/10.1016/j.molcel.2013.10.035

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

Branon, T. C., Bosch, J. A., Sanchez, A. D., Udeshi, N. D., Svinkina, T., Carr, S. A., Feldman, J. L., Perrimon, N., & Ting, A. Y. (2018). Efficient proximity

labeling in living cells and organisms with TurboID. *Nature Biotechnology*, *36*(9), 880–887. https://doi.org/10.1038/nbt.4201

Brecht, R. M., Liu, C. C., Beilinson, H. A., Khitun, A., Slavoff, S. A., & Schatz, D. G. (2020). Nucleolar localization of RAG1 modulates V(D)J recombination activity. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1920021117

Brewer, B. J., & Fangman, W. L. (1988). A replication fork barrier at the 3' end of yeast ribosomal RNA genes. *Cell*, *55*(4), 637–643. https://doi.org/10.1016/0092-8674(88)90222-x

Brison, O., El-Hilali, S., Azar, D., Koundrioukoff, S., Schmidt, M., Nähse, V., Jaszczyszyn, Y., Lachages, A.-M., Dutrillaux, B., Thermes, C., Debatisse, M., & Chen, C.-L. (2019). Transcription-mediated organization of the replication initiation program across large genes sets common fragile sites genome-wide. *Nature Communications*, *10*(1), 5693. https://doi.org/10.1038/s41467-019-13674-5

Broxson, C., Beckett, J., & Tornaletti, S. (2011). Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene. *Biochemistry*, *50*(19), 4162–4172. https://doi.org/10.1021/bi2002136

Brüning, J.-G., & Marians, K. J. (2020). Replisome bypass of transcription complexes and R-loops. *Nucleic Acids Research*, *48*(18), 10353–10367. https://doi.org/10.1093/nar/gkaa741

Castel, S. E., Ren, J., Bhattacharjee, S., Chang, A.-Y., Sánchez, M., Valbuena, A., Antequera, F., & Martienssen, R. A. (2014). Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell*, *159*(3), 572–583. https://doi.org/10.1016/j.cell.2014.09.031

Castillo-Guzman, D., Hartono, S. R., Sanz, L. A., & Chédin, F. (2020). *SF3B1-targeted Splicing Inhibition Triggers Global Alterations in Transcriptional Dynamics and R-Loop Metabolism* (p. 2020.06.08.130583). bioRxiv. https://doi.org/10.1101/2020.06.08.130583

Chappidi, N., Nascakova, Z., Boleslavska, B., Zellweger, R., Isik, E., Andrs, M., Menon, S., Dobrovolna, J., Balbo Pogliano, C., Matos, J., Porro, A., Lopes, M., & Janscak, P. (2020). Fork Cleavage-Religation Cycle and Active Transcription Mediate Replication Restart after Fork Stalling at Co-transcriptional R-Loops. *Molecular Cell*, *77*(3), 528-541.e8. https://doi.org/10.1016/j.molcel.2019.10.026

Chen, Y.-H., Keegan, S., Kahli, M., Tonzi, P., Fenyö, D., Huang, T. T., & Smith, D. J. (2019). Transcription shapes DNA replication initiation and termination in human cells. *Nature Structural & Molecular Biology*, *26*(1), 67. https://doi.org/10.1038/s41594-018-0171-0

Cheung, A. C. M., & Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, *471*(7337), 249–253. https://doi.org/10.1038/nature09785

Choi-Rhee, E., Schulman, H., & Cronan, J. E. (2004). Promiscuous protein biotinylation by Escherichia coli biotin protein ligase. *Protein Science: A*

Publication of the Protein Society, *13*(11), 3043–3050. https://doi.org/10.1110/ps.04911804

Chong, S. Y., Cutler, S., Lin, J.-J., Tsai, C.-H., Tsai, H.-K., Biggins, S., Tsukiyama, T., Lo, Y.-C., & Kao, C.-F. (2020). H3K4 methylation at active genes mitigates transcription-replication conflicts during replication stress. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-14595-4

Clark, R. M., Bhaskar, S. S., Miyahara, M., Dalgliesh, G. L., & Bidichandani, S. I. (2006). Expansion of GAA trinucleotide repeats in mammals. *Genomics*, *87*(1), 57–67. https://doi.org/10.1016/j.ygeno.2005.09.006

Conti, B. A., & Smogorzewska, A. (2020). Mechanisms of direct replication restart at stressed replisomes. *DNA Repair*, *95*, 102947. https://doi.org/10.1016/j.dnarep.2020.102947

Costa, A., & Diffley, J. F. X. (2022). The Initiation of Eukaryotic DNA Replication. *Annual Review of Biochemistry*, *91*, 107–131. https://doi.org/10.1146/annurev-biochem-072321-110228

Cronan, J. E. (2005). Targeted and proximity-dependent promiscuous protein biotinylation by a mutant Escherichia coli biotin protein ligase. *The Journal of Nutritional Biochemistry*, *16*(7), 416–418. https://doi.org/10.1016/j.jnutbio.2005.03.017

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

Czajkowsky, D. M., Liu, J., Hamlin, J. L., & Shao, Z. (2008). DNA Combing Reveals Intrinsic Temporal Disorder in the Replication of Yeast Chromosome VI. *Journal of Molecular Biology*, *375*(1), 12–19. https://doi.org/10.1016/j.jmb.2007.10.046

Davis, J. K., & Broadie, K. (2017). Multifarious Functions of the Fragile X Mental Retardation Protein. *Trends in Genetics: TIG*, *33*(10), 703–714. https://doi.org/10.1016/j.tig.2017.07.008

de Boer, E., Rodriguez, P., Bonte, E., Krijgsveld, J., Katsantoni, E., Heck, A., Grosveld, F., & Strouboulis, J. (2003). Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(13), 7480–7485. https://doi.org/10.1073/pnas.1332608100

De, S., & Michor, F. (2011). DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nature Structural & Molecular Biology*, *18*(8), 950–955. https://doi.org/10.1038/nsmb.2089

Deissler, H., Behn-Krappa, A., & Doerfler, W. (1996). Purification of nuclear proteins from human HeLa cells that bind specifically to the unstable tandem repeat (CGG)n in the human FMR1 gene. *Journal of Biological Chemistry*, *271*(8), 4327–4334. https://doi.org/10.1074/jbc.271.8.4327

References

Deissler, H., Wilm, M., Genç, B., Schmitz, B., Ternes, T., Naumann, F., Mann, M., & Doerfler, W. (1997). Rapid protein sequencing by tandem mass spectrometry and cDNA cloning of p20-CGGBP. A novel protein that binds to the unstable triplet repeat 5'-d(CGG)n-3' in the human FMR1 gene. *The Journal of Biological Chemistry*, *272*(27), 16761–16768. https://doi.org/10.1074/jbc.272.27.16761

Dhar, S. K., Yoshida, K., Machida, Y., Khaira, P., Chaudhuri, B., Wohlschlegel, J. A., Leffak, M., Yates, J., & Dutta, A. (2001). Replication from oriP of Epstein-Barr Virus Requires Human ORC and Is Inhibited by Geminin. *Cell*, *106*(3), 287–296. https://doi.org/10.1016/S0092-8674(01)00458-5

Douse, C. H., Tchasovnikarova, I. A., Timms, R. T., Protasio, A. V., Seczynska, M., Prigozhin, D. M., Albecka, A., Wagstaff, J., Williamson, J. C., Freund, S. M. V., Lehner, P. J., & Modis, Y. (2020). TASOR is a pseudo-PARP that directs HUSH complex assembly and epigenetic transposon control. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-18761-6

Dubois, M.-L., Bastin, C., Lévesque, D., & Boisvert, F.-M. (2016). Comprehensive Characterization of Minichromosome Maintenance Complex (MCM) Protein Interactions Using Affinity and Proximity Purifications Coupled to Mass Spectrometry. *Journal of Proteome Research*, *15*(9), 2924–2934. https://doi.org/10.1021/acs.jproteome.5b01081

Eisenberg, M. A., Prakash, O., & Hsiung, S. C. (1982). Purification and properties of the biotin repressor. A bifunctional protein. *Journal of Biological Chemistry*, *257*(24), 15167–15173. https://doi.org/10.1016/S0021-9258(18)33408-2

Ekundayo, B., & Bleichert, F. (2019). Origins of DNA replication. *PLoS Genetics*, *15*(9), e1008320. https://doi.org/10.1371/journal.pgen.1008320

Feng, Y., Gutekunst, C.-A., Eberhart, D. E., Yi, H., Warren, S. T., & Hersch, S. M. (1997). Fragile X Mental Retardation Protein: Nucleocytoplasmic Shuttling and Association with Somatodendritic Ribosomes. *The Journal of Neuroscience*, *17*(5), 1539–1547. https://doi.org/10.1523/JNEUROSCI.17-05-01539.1997

Fernández-Suárez, M., Chen, T. S., & Ting, A. Y. (2008). Protein-protein interaction detection in vitro and in cells by proximity biotinylation. *Journal of the American Chemical Society*, *130*(29), 9251–9253. https://doi.org/10.1021/ja801445p

Frankish, A., Diekhans, M., Jungreis, I., Lagarde, J., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Boix, C., Carbonell Sala, S., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I. T., García Girón, C., … Flicek, P. (2020). GENCODE 2021. *Nucleic Acids Research*, *49*(D1), D916–D923. https://doi.org/10.1093/nar/gkaa1087

Fry, M., & Loeb, L. A. (1994). The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proceedings of the National*

Academy of Sciences, *91*(11), 4950–4954.
https://doi.org/10.1073/pnas.91.11.4950

Fu, Y.-H., Kuhl, D. P. A., Pizzuti, A., Pieretti, M., Sutcliffe, J. S., Richards, S.,
Verkert, A. J. M. H., Holden, J. J. A., Fenwick, R. G., Warren, S. T.,
Oostra, B. A., Nelson, D. L., & Caskey, C. T. (1991). Variation of the CGG
repeat at the fragile X site results in genetic instability: Resolution of the
Sherman paradox. *Cell*, *67*(6), 1047–1058. https://doi.org/10.1016/0092-
8674(91)90283-5

Gahn, T. A., & Schildkraut, C. L. (1989). The Epstein-Barr virus origin of plasmid
replication, oriP, contains both the initiation and termination sites of DNA
replication. *Cell*, *58*(3), 527–535. https://doi.org/10.1016/0092-
8674(89)90433-9

Gan, W., Guan, Z., Liu, J., Gui, T., Shen, K., Manley, J. L., & Li, X. (2011). R-
loop-mediated genomic instability is caused by impairment of replication
fork progression. *Genes & Development*, *25*(19), 2041–2056.
https://doi.org/10.1101/gad.17010011

García-Gómez, S., Reyes, A., Martínez-Jiménez, M. I., Chocrón, E. S., Mourón,
S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I. J., & Blanco, L.
(2013). PrimPol, an archaic primase/polymerase operating in human cells.
Molecular Cell, *52*(4), 541–553.
https://doi.org/10.1016/j.molcel.2013.09.025

García-Rubio, M., Aguilera, P., Lafuente-Barquero, J., Ruiz, J. F., Simon, M.-N.,
Geli, V., Rondón, A. G., & Aguilera, A. (2018). Yra1-bound RNA–DNA
hybrids cause orientation-independent transcription–replication collisions
and telomere instability. *Genes & Development*, *32*(13–14), 965–977.
https://doi.org/10.1101/gad.311274.117

García-Rubio, M., 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., Lim, Y. W., Lott, P. L., Korf, I., & Chédin, F. (2013). GC skew at the
5′ and 3′ ends of human genes links R-loop formation to epigenetic
regulation and transcription termination. *Genome Research*, *23*(10), 1590–
1600. https://doi.org/10.1101/gr.158436.113

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., Berger, C., Coyle, J., & Echo, B. (1984). DNA polymerase alpha
inhibition by aphidicolin induces gaps and breaks at common fragile sites
in human chromosomes. *Human Genetics*, *67*(2), 136–142.
https://doi.org/10.1007/BF00272988

Gómez-González, B., & Aguilera, A. (2019). Transcription-mediated replication hindrance: A major driver of genome instability. *Genes & Development*, *33*(15–16), 1008–1026. https://doi.org/10.1101/gad.324517.119

Goracci, M., Lanni, S., Mancano, G., Palumbo, F., Chiurazzi, P., Neri, G., & Tabolacci, E. (2016). Defining the role of the CGGBP1 protein in FMR1 gene expression. *European Journal of Human Genetics*, *24*(5), 697–703. https://doi.org/10.1038/ejhg.2015.182

Graham, R. C., & Karnovsky, M. J. (1966). The early stages of absorption of injected horseradish peroxidase in the proximal tubules of mouse kidney: Ultrastructural cytochemistry by a new technique. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, *14*(4), 291–302. https://doi.org/10.1177/14.4.291

Gregersen, L. H., Mitter, R., Ugalde, A. P., Nojima, T., Proudfoot, N. J., Agami, R., Stewart, A., & Svejstrup, J. Q. (2019). SCAF4 and SCAF8, mRNA Anti-Terminator Proteins. *Cell*, *177*(7), 1797-1813.e18. https://doi.org/10.1016/j.cell.2019.04.038

Griesenbeck, J., Boeger, H., Strattan, J. S., & Kornberg, R. D. (2003). Affinity Purification of Specific Chromatin Segments from Chromosomal Loci in Yeast. *Molecular and Cellular Biology*, *23*(24), 9275–9282. https://doi.org/10.1128/MCB.23.24.9275-9282.2003

Groelly, F. J., Dagg, R. A., Petropoulos, M., Rossetti, G. G., Prasad, B., Panagopoulos, A., Paulsen, T., Karamichali, A., Jones, S. E., Ochs, F., Dionellis, V. S., Puig Lombardi, E., Miossec, M. J., Lockstone, H., Legube, G., Blackford, A. N., Altmeyer, M., Halazonetis, T. D., & Tarsounas, M. (2022). Mitotic DNA synthesis is caused by transcription-replication conflicts in BRCA2-deficient cells. *Molecular Cell*, *82*(18), 3382-3397.e7. https://doi.org/10.1016/j.molcel.2022.07.011

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

Grosche, A., Hauser, A., Lepper, M. F., Mayo, R., von Toerne, C., Merl-Pham, J., & Hauck, S. M. (2016). The Proteome of Native Adult Müller Glial Cells From Murine Retina. *Molecular & Cellular Proteomics: MCP*, *15*(2), 462–480. https://doi.org/10.1074/mcp.M115.052183

Gross, A. J., & Sizer, I. W. (1959). The Oxidation of Tyramine, Tyrosine, and Related Compounds by Peroxidase. *Journal of Biological Chemistry*, *234*(6), 1611–1614. https://doi.org/10.1016/S0021-9258(18)70059-8

Guy, L., & Roten, C.-A. H. (2004). Genometric analyses of the organization of circular chromosomes: A universal pressure determines the direction of ribosomal RNA genes transcription relative to chromosome replication. *Gene*, *340*(1), 45–52. https://doi.org/10.1016/j.gene.2004.06.056

Hah, N., Danko, C. G., Core, L., Waterfall, J. J., Siepel, A., Lis, J. T., & Kraus, W. L. (2011). A Rapid, Extensive, and Transient Transcriptional Response to

Estrogen Signaling in Breast Cancer Cells. *Cell*, *145*(4), 622–634. https://doi.org/10.1016/j.cell.2011.03.042

Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T., & Cimprich, K. A. (2017). Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell*, *170*(4), 774-786.e19. https://doi.org/10.1016/j.cell.2017.07.043

Han, Y., Branon, T. C., Martell, J. D., Boassa, D., Shechner, D., Ellisman, M. H., & Ting, A. (2019). Directed Evolution of Split APEX2 Peroxidase. *ACS Chemical Biology*, *14*(4), 619–635. https://doi.org/10.1021/acschembio.8b00919

Hassan, A. B., Errington, R. J., White, N. S., Jackson, D. A., & Cook, P. R. (1994). Replication and transcription sites are colocalized in human cells. *Journal of Cell Science*, *107*(2), 425–434. https://doi.org/10.1242/jcs.107.2.425

Hawkins, M., Dimude, J. U., Howard, J. A. L., Smith, A. J., Dillingham, M. S., Savery, N. J., Rudolph, C. J., & McGlynn, P. (2019). Direct removal of RNA polymerase barriers to replication by accessory replicative helicases. *Nucleic Acids Research*, *47*(10), 5100–5113. https://doi.org/10.1093/nar/gkz170

Hayward, B. E., & Usdin, K. (2021). Mechanisms of Genome Instability in the Fragile X-Related Disorders. *Genes*, *12*(10), 1633. https://doi.org/10.3390/genes12101633

Hecht, F., & Glover, T. W. (1984). Cancer chromosome breakpoints and common fragile sites induced by aphidicolin. *Cancer Genetics and Cytogenetics*, *13*(2), 185–188. https://doi.org/10.1016/0165-4608(84)90060-8

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

Helmrich, A., Stout-Weider, K., Hermann, K., Schrock, E., & Heiden, T. (2006). Common fragile sites are conserved features of human and mouse chromosomes and relate to large active genes. *Genome Research*, *16*(10), 1222–1230. https://doi.org/10.1101/gr.5335506

Hoffman, E. A., McCulley, A., Haarer, B., Arnak, R., & Feng, W. (2015). Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Research*, *25*(3), 402–412. https://doi.org/10.1101/gr.180497.114

Huang, J., Liu, S., Bellani, M. A., Thazhathveetil, A. K., Ling, C., de Winter, J. P., Wang, Y., Wang, W., & Seidman, M. M. (2013). The DNA translocase FANCM/MHF promotes replication traverse of DNA interstrand crosslinks. *Molecular Cell*, *52*(3), 434–446. https://doi.org/10.1016/j.molcel.2013.09.021

Hung, V., Udeshi, N. D., Lam, S. S., Loh, K. H., Cox, K. J., Pedram, K., Carr, S. A., & Ting, A. Y. (2016). Spatially resolved proteomic mapping in living

cells with the engineered peroxidase APEX2. *Nature Protocols*, *11*(3), 456–475. https://doi.org/10.1038/nprot.2016.018

Hyrien, O. (2015). Peaks cloaked in the mist: The landscape of mammalian replication origins. *The Journal of Cell Biology*, *208*(2), 147–160. https://doi.org/10.1083/jcb.201407004

Ivessa, A. S., Lenzmeier, B. A., Bessler, J. B., Goudsouzian, L. K., Schnakenberg, S. L., & Zakian, V. A. (2003). The Saccharomyces cerevisiae helicase Rrm3p facilitates replication past nonhistone protein-DNA complexes. *Molecular Cell*, *12*(6), 1525–1536. https://doi.org/10.1016/s1097-2765(03)00456-8

Jao, C. Y., & Salic, A. (2008). Exploring RNA transcription and turnover in vivo by using click chemistry. *Proceedings of the National Academy of Sciences*, *105*(41), 15779–15784. https://doi.org/10.1073/pnas.0808480105

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), Article 7873. https://doi.org/10.1038/s41586-021-03819-2

Kettani, A., Kumar, A. R., & Patel, D. J. (1995). Solution Structure of a DNA Quadruplex Containing the Fragile X Syndrome Triplet Repeat. *Journal of Molecular Biology*, *254*(4), 638–656. https://doi.org/10.1006/jmbi.1995.0644

Kim, D. I., Jensen, S. C., Noble, K. A., KC, B., Roux, K. H., Motamedchaboki, K., & Roux, K. J. (2016). An improved smaller biotin ligase for BioID proximity labeling. *Molecular Biology of the Cell*, *27*(8), 1188–1196. https://doi.org/10.1091/mbc.E15-12-0844

Kim, D. I., Kc, B., Zhu, W., Motamedchaboki, K., Doye, V., & Roux, K. J. (2014). Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proceedings of the National Academy of Sciences*, *111*(24), E2453–E2461. https://doi.org/10.1073/pnas.1406459111

Kim, D. I., & Roux, K. J. (2016). Filling the void: Proximity-based labeling of proteins in living cells. *Trends in Cell Biology*, *26*(11), 804–817. https://doi.org/10.1016/j.tcb.2016.09.004

Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J., & Jinks-Robertson, S. (2007). Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair*, *6*(9), 1285–1296. https://doi.org/10.1016/j.dnarep.2007.02.023

Kirchmaier, A. L., & Sugden, B. (1995). Plasmid maintenance of derivatives of oriP of Epstein-Barr virus. *Journal of Virology*, *69*(2), 1280–1283. https://doi.org/10.1128/jvi.69.2.1280-1283.1995

Kobayashi, K., Guilliam, T. A., Tsuda, M., Yamamoto, J., Bailey, L. J., Iwai, S., Takeda, S., Doherty, A. J., & Hirota, K. (2016). Repriming by PrimPol is

critical for DNA replication restart downstream of lesions and chain-terminating nucleosides. *Cell Cycle (Georgetown, Tex.)*, *15*(15), 1997–2008. https://doi.org/10.1080/15384101.2016.1191711

Kochanova, N. Y., Schauer, T., Mathias, G. P., Lukacs, A., Schmidt, A., Flatley, A., Schepers, A., Thomae, A. W., & Imhof, A. (2020). A multi-layered structure of the interphase chromocenter revealed by proximity-based biotinylation. *Nucleic Acids Research*, *48*(8), 4161–4178. https://doi.org/10.1093/nar/gkaa145

Kotani, N., Gu, J., Isaji, T., Udaka, K., Taniguchi, N., & Honke, K. (2008). Biochemical visualization of cell surface molecular clustering in living cells. *Proceedings of the National Academy of Sciences*, *105*(21), 7405–7409. https://doi.org/10.1073/pnas.0710346105

Kotsantis, P., Silva, L. M., Irmscher, S., Jones, R. M., Folkes, L., Gromak, N., & Petermann, E. (2016). Increased global transcription activity as a mechanism of replication stress in cancer. *Nature Communications*, *7*, 13087. https://doi.org/10.1038/ncomms13087

Kowarz, E., Löscher, D., & Marschalek, R. (2015). Optimized Sleeping Beauty transposons rapidly generate stable transgenic cell lines. *Biotechnology Journal*, *10*(4), 647–653. https://doi.org/10.1002/biot.201400821

Kumar, C., & Remus, D. (2023). Looping out of control: R-loops in transcription-replication conflict. *Chromosoma*. https://doi.org/10.1007/s00412-023-00804-8

Kurat, C. F., Lambert, J.-P., van Dyk, D., Tsui, K., van Bakel, H., Kaluarachchi, S., Friesen, H., Kainth, P., Nislow, C., Figeys, D., Fillingham, J., & Andrews, B. J. (2011). Restriction of histone gene transcription to S phase by phosphorylation of a chromatin boundary protein. *Genes & Development*, *25*(23), 2489–2501. https://doi.org/10.1101/gad.173427.111

Kwon, K., & Beckett, D. (2000). Function of a conserved sequence motif in biotin holoenzyme synthetases. *Protein Science*, *9*(8), 1530–1539. https://doi.org/10.1110/ps.9.8.1530

Kwon, K., Streaker, E. D., Ruparelia, S., & Beckett, D. (2000). Multiple Disordered Loops Function in Corepressor-induced Dimerization of the Biotin Repressor. *Journal of Molecular Biology*, *304*(5), 821–833. https://doi.org/10.1006/jmbi.2000.4249

Lalonde, M., Trauner, M., Werner, M., & Hamperl, S. (2021). Consequences and Resolution of Transcription–Replication Conflicts. *Life*, *11*(7), 637. https://doi.org/10.3390/life11070637

Lam, F. C., Kong, Y. W., Huang, Q., Vu Han, T.-L., Maffa, A. D., Kasper, E. M., & Yaffe, M. B. (2020). BRD4 prevents the accumulation of R-loops and protects against transcription–replication collision events and DNA damage. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-17503-y

Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., & Ting, A. Y. (2015). Directed evolution of APEX2 for electron

microscopy and proximity labeling. *Nature Methods*, *12*(1), 51–54. https://doi.org/10.1038/nmeth.3179

Landsverk, H. B., Sandquist, L. E., Bay, L. T. E., Steurer, B., Campsteijn, C., Landsverk, O. J. B., Marteijn, J. A., Petermann, E., Trinkle-Mulcahy, L., & Syljuåsen, R. G. (2020). WDR82/PNUTS-PP1 Prevents Transcription-Replication Conflicts by Promoting RNA Polymerase II Degradation on Chromatin. *Cell Reports*, *33*(9), 108469. https://doi.org/10.1016/j.celrep.2020.108469

Lang, K. S., Hall, A. N., Merrikh, C. N., Ragheb, M., Tabakh, H., Pollock, A. J., Woodward, J. J., Dreifus, J. E., & Merrikh, H. (2017). Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell*, *170*(4), 787-799.e18. https://doi.org/10.1016/j.cell.2017.07.044

Lang, K. S., & Merrikh, H. (2021). Topological stress is responsible for the detrimental outcomes of head-on replication-transcription conflicts. *Cell Reports*, *34*(9), 108797. https://doi.org/10.1016/j.celrep.2021.108797

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Lemmens, B., van Schendel, R., & Tijsterman, M. (2015). Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nature Communications*, *6*, 8909. https://doi.org/10.1038/ncomms9909

Li, J., Wang, Y., Chiu, S.-L., & Cline, H. T. (2010). Membrane Targeted Horseradish Peroxidase as a Marker for Correlative Fluorescence and Electron Microscopy Studies. *Frontiers in Neural Circuits*, *4*. https://doi.org/10.3389/neuro.04.006.2010

Li, L., Williams, P., Gao, Z., & Wang, Y. (2020). VEZF1–guanine quadruplex DNA interaction regulates alternative polyadenylation and detyrosinase activity of VASH1. *Nucleic Acids Research*, *48*(21), 11994–12003. https://doi.org/10.1093/nar/gkaa1092

Li, L., Williams, P., Ren, W., Wang, M. Y., Gao, Z., Miao, W., Huang, M., Song, J., & Wang, Y. (2021). YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nature Chemical Biology*, *17*(2), 161–168. https://doi.org/10.1038/s41589-020-00695-1

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Liao, Y., Smyth, G. K., & Shi, W. (2019). The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Research*, *47*(8), e47. https://doi.org/10.1093/nar/gkz114

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*, e08007. 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

Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., Baymuradov, U. K., Graham, K., Litton, C., Miyasato, S. R., Strattan, J. S., Jolanki, O., Lee, J.-W., Tanaka, F. Y., Adenekan, P., … Cherry, J. M. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Research*, *48*(D1), D882–D889. https://doi.org/10.1093/nar/gkz1062

Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H., & Ting, A. Y. (2012). Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nature Biotechnology*, *30*(11), 1143–1148. https://doi.org/10.1038/nbt.2375

Marzluff, W. F., Wagner, E. J., & Duronio, R. J. (2008). Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nature Reviews. Genetics*, *9*(11), 843–854. https://doi.org/10.1038/nrg2438

Matsuoka, S., Ballif, B. A., Smogorzewska, A., McDonald, E. R., Hurov, K. E., Luo, J., Bakalarski, C. E., Zhao, Z., Solimini, N., Lerenthal, Y., Shiloh, Y., Gygi, S. P., & Elledge, S. J. (2007). ATM and ATR Substrate Analysis Reveals Extensive Protein Networks Responsive to DNA Damage. *Science*, *316*(5828), 1160–1166. https://doi.org/10.1126/science.1140321

Mayer, G., & Bendayan, M. (1997). Biotinyl–Tyramide: A Novel Approach for Electron Microscopic Immunocytochemistry. *Journal of Histochemistry & Cytochemistry*, *45*(11), 1449–1454. https://doi.org/10.1177/002215549704501101

McLean, M. J., Wolfe, K. H., & Devine, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *Journal of Molecular Evolution*, *47*(6), 691–696. https://doi.org/10.1007/pl00006428

Mellor, C., Perez, C., & Sale, J. E. (2022). Creation and resolution of non-B-DNA structural impediments during replication. *Critical Reviews in Biochemistry and Molecular Biology*, *57*(4), 412–442. https://doi.org/10.1080/10409238.2022.2121803

Merrikh, H., Machón, C., Grainger, W. H., Grossman, A. D., & Soultanas, P. (2011). Co-directional replication–transcription conflicts lead to replication restart. *Nature*, *470*(7335), 554–557. https://doi.org/10.1038/nature09758

Merrikh, H., Zhang, Y., Grossman, A. D., & Wang, J. D. (2012). Replication-transcription conflicts in bacteria. *Nature Reviews. Microbiology*, *10*(7), 449–458. https://doi.org/10.1038/nrmicro2800

References

Meryet-Figuiere, M., Alaei-Mahabadi, B., Ali, M. M., Mitra, S., Subhash, S., Pandey, G. K., Larsson, E., & Kanduri, C. (2014). Temporal separation of replication and transcription during S-phase progression. *Cell Cycle*, *13*(20), 3241–3248. https://doi.org/10.4161/15384101.2014.953876

Minde, D.-P., Ramakrishna, M., & Lilley, K. S. (2020). Biotin proximity tagging favours unfolded proteins and enables the study of intrinsically disordered regions. *Communications Biology*, *3*(1), 1–13. https://doi.org/10.1038/s42003-020-0758-y

Miotto, B., Chibi, M., Xie, P., Koundrioukoff, S., Moolman-Smook, H., Pugh, D., Debatisse, M., He, F., Zhang, L., & Defossez, P.-A. (2014). The RBBP6/ZBTB38/MCM10 Axis Regulates DNA Replication and Common Fragile Site Stability. *Cell Reports*, *7*(2), 575–587. https://doi.org/10.1016/j.celrep.2014.03.030

Moldovan, G.-L., & D'Andrea, A. D. (2009). How the Fanconi Anemia pathway guards the genome. *Annual Review of Genetics*, *43*, 223–249. https://doi.org/10.1146/annurev-genet-102108-134222

Mortensen, A., & Skibsted, L. H. (1997). Importance of Carotenoid Structure in Radical-Scavenging Reactions. *Journal of Agricultural and Food Chemistry*, *45*(8), 2970–2977. https://doi.org/10.1021/jf970010s

Motnenko, A., Liang, C.-C., Yang, D., Lopez-Martinez, D., Yoshikawa, Y., Zhan, B., Ward, K. E., Tian, J., Haas, W., Spingardi, P., Kessler, B. M., Kriaucionis, S., Gygi, S. P., & Cohn, M. A. (2018). Identification of UHRF2 as a novel DNA interstrand crosslink sensor protein. *PLoS Genetics*, *14*(10), e1007643. https://doi.org/10.1371/journal.pgen.1007643

Mourón, S., Rodriguez-Acebes, S., Martínez-Jiménez, M. I., García-Gómez, S., Chocrón, S., Blanco, L., & Méndez, J. (2013). Repriming of DNA synthesis at stalled replication forks by human PrimPol. *Nature Structural & Molecular Biology*, *20*(12), 1383–1389. https://doi.org/10.1038/nsmb.2719

Müller, S., Kumari, S., Rodriguez, R., & Balasubramanian, S. (2010). Small-molecule-mediated G-quadruplex isolation from human cells. *Nature Chemistry*, *2*(12), 1095–1098. https://doi.org/10.1038/nchem.842

Müller-Hartmann, H., Deissler, H., Naumann, F., Schmitz, B., Schröer, J., & Doerfler, W. (2000). The human 20-kDa 5'-(CGG)(n)-3'-binding protein is targeted to the nucleus and affects the activity of the FMR1 promoter. *Journal of Biological Chemistry*, *275*(9), 6447–6452. https://doi.org/10.1074/jbc.275.9.6447

Munter, S. D., Görnemann, J., Derua, R., Lesage, B., Qian, J., Heroes, E., Waelkens, E., Eynde, A. V., Beullens, M., & Bollen, M. (2017). Split-BioID: A proximity biotinylation assay for dimerization-dependent protein interactions. *FEBS Letters*, *591*(2), 415–424. https://doi.org/10.1002/1873-3468.12548

Nadel, J., Athanasiadou, R., Lemetre, C., Wijetunga, N. A., Ó Broin, P., Sato, H., Zhang, Z., Jeddeloh, J., Montagna, C., Golden, A., Seoighe, C., & Greally, J. M. (2015). RNA:DNA hybrids in the human genome have distinctive

nucleotide characteristics, chromatin composition, and transcriptional relationships. *Epigenetics & Chromatin*, 8, 46. https://doi.org/10.1186/s13072-015-0040-6

Nakamura, M., Choe, S.-K., Runko, A. P., Gardner, P. D., & Sagerström, C. G. (2008). Nlz1/Znf703 acts as a repressor of transcription. *BMC Developmental Biology*, 8(1), 108. https://doi.org/10.1186/1471-213X-8-108

Navarro, P., Trevisan-Herraz, M., Bonzon-Kulichenko, E., Núñez, E., Martínez-Acedo, P., Pérez-Hernández, D., Jorge, I., Mesa, R., Calvo, E., Carrascal, M., Hernáez, M. L., García, F., Bárcena, J. A., Ashman, K., Abian, J., Gil, C., Redondo, J. M., & Vázquez, J. (2014). General Statistical Framework for Quantitative Proteomics by Stable Isotope Labeling. *Journal of Proteome Research*, 13(3), 1234–1247. https://doi.org/10.1021/pr4006958

Nepal, M., Che, R., Ma, C., Zhang, J., & Fei, P. (2017). FANCD2 and DNA Damage. *International Journal of Molecular Sciences*, 18(8), 1804. https://doi.org/10.3390/ijms18081804

Oestergaard, V. H., & Lisby, M. (2017). Transcription-replication conflicts at chromosomal fragile sites-consequences in M phase and beyond. *Chromosoma*, 126(2), 213–222. https://doi.org/10.1007/s00412-016-0617-2

Okamoto, Y., Abe, M., Itaya, A., Tomida, J., Ishiai, M., Takaori-Kondo, A., Taoka, M., Isobe, T., & Takata, M. (2019). FANCD2 protects genome stability by recruiting RNA processing enzymes to resolve R-loops during mild replication stress. *The FEBS Journal*, 286(1), 139–150. https://doi.org/10.1111/febs.14700

Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W., & Darst, S. A. (2003). Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase. *Cell*, 114(3), 335–345. https://doi.org/10.1016/S0092-8674(03)00600-7

Paeschke, K., Capra, J. A., & Zakian, V. A. (2011). DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell*, 145(5), 678–691. https://doi.org/10.1016/j.cell.2011.04.015

Papadopoulos, D., Solvie, D., Baluapuri, A., Endres, T., Ha, S. A., Herold, S., Kalb, J., Giansanti, C., Schülein-Völk, C., Ade, C. P., Schneider, C., Gaballa, A., Vos, S., Fischer, U., Dobbelstein, M., Wolf, E., & Eilers, M. (2022). MYCN recruits the nuclear exosome complex to RNA polymerase II to prevent transcription-replication conflicts. *Molecular Cell*, 82(1), 159-176.e12. https://doi.org/10.1016/j.molcel.2021.11.002

Patel, D., Patel, M., Datta, S., & Singh, U. (2019). CGGBP1 regulates CTCF occupancy at repeats. *Epigenetics & Chromatin*, 12(1), 57. https://doi.org/10.1186/s13072-019-0305-6

Pentzold, C., Shah, S. A., Hansen, N. R., Le Tallec, B., Seguin-Orlando, A., Debatisse, M., Lisby, M., & Oestergaard, V. H. (2018). FANCD2 binding identifies conserved fragile sites at large transcribed genes in avian cells.

Nucleic Acids Research, *46*(3), 1280–1294.
https://doi.org/10.1093/nar/gkx1260

Pérez-Calero, C., Bayona-Feliu, A., Xue, X., Barroso, S. I., Muñoz, S., González-Basallote, V. M., Sung, P., & Aguilera, A. (2020). UAP56/DDX39B is a major cotranscriptional RNA–DNA helicase that unwinds harmful R loops genome-wide. *Genes & Development*, *34*(13–14), 898–912. https://doi.org/10.1101/gad.336024.119

Petryk, N., Kahli, M., d'Aubenton-Carafa, Y., Jaszczyszyn, Y., Shen, Y., Silvain, M., Thermes, C., Chen, C.-L., & Hyrien, O. (2016). Replication landscape of the human genome. *Nature Communications*, *7*(1), 10208. https://doi.org/10.1038/ncomms10208

Pieretti, M., Zhang, F. P., Fu, Y. H., Warren, S. T., Oostra, B. A., Caskey, C. T., & Nelson, D. L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell*, *66*(4), 817–822. https://doi.org/10.1016/0092-8674(91)90125-i

Poli, J., Gerhold, C.-B., Tosi, A., Hustedt, N., Seeber, A., Sack, R., Herzog, F., Pasero, P., Shimada, K., Hopfner, K.-P., & Gasser, S. M. (2016). Mec1, INO80, and the PAF1 complex cooperate to limit transcription replication conflicts through RNAPII removal during replication stress. *Genes & Development*, *30*(3), 337–354. https://doi.org/10.1101/gad.273813.115

Pomerantz, R. T., & O'Donnell, M. (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature*, *456*(7223), 762–766. https://doi.org/10.1038/nature07527

Pomerantz, R. T., & O'Donnell, M. (2010). What happens when replication and transcription complexes collide? *Cell Cycle (Georgetown, Tex.)*, *9*(13), 2537–2543. https://doi.org/10.4161/cc.9.13.12122

Prado, F., & Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *The EMBO Journal*, *24*(6), 1267–1276. https://doi.org/10.1038/sj.emboj.7600602

Quinet, A., Lemaçon, D., & Vindigni, A. (2017). Replication Fork Reversal: Players and Guardians. *Molecular Cell*, *68*(5), 830–833. https://doi.org/10.1016/j.molcel.2017.11.022

Ramanathan, M., Majzoub, K., Rao, D. S., Neela, P. H., Zarnegar, B. J., Mondal, S., Roth, J. G., Gai, H., Kovalski, J. R., Siprashvili, Z., Palmer, T. D., Carette, J. E., & Khavari, P. A. (2018). RNA–protein interaction detection in living cells. *Nature Methods*, *15*(3), 207–212. https://doi.org/10.1038/nmeth.4601

Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., & Ting, A. Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science*, *339*(6125), 1328–1331. https://doi.org/10.1126/science.1230593

Richards, R. I., Holman, K., Yu, S., & Sutherland, G. R. (1993). Fragile X syndrome unstable element, p(CCG)n, and other simple tandem repeat sequences are binding sites for specific nuclear proteins. *Human*

Molecular Genetics, *2*(9), 1429–1435.
https://doi.org/10.1093/hmg/2.9.1429

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*(1), 24–26. https://doi.org/10.1038/nbt.1754

Rocha, E. P. C., & Danchin, A. (2003a). Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nature Genetics*, *34*(4), 377–378. https://doi.org/10.1038/ng1209

Rocha, E. P. C., & Danchin, A. (2003b). Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Research*, *31*(22), 6570–6577. https://doi.org/10.1093/nar/gkg859

Rodriguez, R., Miller, K. M., Forment, J. V., Bradshaw, C. R., Nikan, M., Britton, S., Oelschlaegel, T., Xhemalce, B., Balasubramanian, S., & Jackson, S. P. (2012). Small-molecule–induced DNA damage identifies alternative DNA structures in human genes. *Nature Chemical Biology*, *8*(3), 301–310. https://doi.org/10.1038/nchembio.780

Rodriguez, R., Müller, S., Yeoman, J. A., Trentesaux, C., Riou, J.-F., & Balasubramanian, S. (2008). A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *Journal of the American Chemical Society*, *130*(47), 15758–15759. https://doi.org/10.1021/ja805615w

Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S., & Bonner, W. M. (1998). DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139 *. *Journal of Biological Chemistry*, *273*(10), 5858–5868. https://doi.org/10.1074/jbc.273.10.5858

Roghanian, M., Zenkin, N., & Yuzenkova, Y. (2015). Bacterial global regulators DksA/ppGpp increase fidelity of transcription. *Nucleic Acids Research*, *43*(3), 1529–1536. https://doi.org/10.1093/nar/gkv003

Roux, K. J. (2013). Marked by association: Techniques for proximity-dependent labeling of proteins in eukaryotic cells. *Cellular and Molecular Life Sciences*, *70*(19), 3657–3664. https://doi.org/10.1007/s00018-013-1287-3

Roux, K. J., Kim, D. I., Raida, M., & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *The Journal of Cell Biology*, *196*(6), 801–810. https://doi.org/10.1083/jcb.201112098

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, *5*(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Sabouri, N., McDonald, K. R., Webb, C. J., Cristea, I. M., & Zakian, V. A. (2012). DNA replication through hard-to-replicate sites, including both highly transcribed RNA Pol II and Pol III genes, requires the S. pombe Pfh1 helicase. *Genes & Development*, *26*(6), 581–593. https://doi.org/10.1101/gad.184697.111

Sanchez, A., Vivo, A. de, Tonzi, P., Kim, J., Huang, T. T., & Kee, Y. (2020). Transcription-replication conflicts as a source of common fragile site instability caused by BMI1-RNF2 deficiency. *PLOS Genetics*, *16*(3), e1008524. https://doi.org/10.1371/journal.pgen.1008524

Sankar, T. S., Wastuwidyaningtyas, B. D., Dong, Y., Lewis, S. A., & Wang, J. D. (2016). The nature of mutations induced by replication–transcription collisions. *Nature*, *535*(7610), 178–181. https://doi.org/10.1038/nature18316

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

Saponaro, M., Kantidakis, T., Mitter, R., Kelly, G. P., Heron, M., Williams, H., Söding, J., Stewart, A., & Svejstrup, J. Q. (2014). RECQL5 Controls Transcript Elongation and Suppresses Genome Instability Associated with Transcription Stress. *Cell*, *157*(5), 1037–1049. https://doi.org/10.1016/j.cell.2014.03.048

Sarkies, P., Murat, P., Phillips, L. G., Patel, K. J., Balasubramanian, S., & Sale, J. E. (2012). FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Research*, *40*(4), 1485–1498. https://doi.org/10.1093/nar/gkr868

Sarkies, P., Reams, C., Simpson, L. J., & Sale, J. E. (2010). Epigenetic Instability due to Defective Replication of Structured DNA. *Molecular Cell*, *40*(5), 703–713. https://doi.org/10.1016/j.molcel.2010.11.009

Schatz, P. J. (1993). Use of Peptide Libraries to Map the Substrate Specificity of a Peptide-Modifying Enzyme: A 13 Residue Consensus Peptide Specifies Biotinylation in Escherichia coli. *Bio/Technology*, *11*(10), 1138–1143. https://doi.org/10.1038/nbt1093-1138

Schauer, G. D., Spenkelink, L. M., Lewis, J. S., Yurieva, O., Mueller, S. H., van Oijen, A. M., & O'Donnell, M. E. (2020). Replisome bypass of a protein-based R-loop block by Pif1. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(48), 30354–30361. https://doi.org/10.1073/pnas.2020189117

Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell Detection with Star-Convex Polygons. In A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, & G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018* (pp. 265–273). Springer International Publishing. https://doi.org/10.1007/978-3-030-00934-2_30

Schopp, I. M., Amaya Ramirez, C. C., Debeljak, J., Kreibich, E., Skribbe, M., Wild, K., & Béthune, J. (2017). Split-BioID a conditional proteomics approach to monitor the composition of spatiotemporally defined protein complexes. *Nature Communications*, *8*(1), 15690. https://doi.org/10.1038/ncomms15690

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature*, *517*(7534), 321–326. https://doi.org/10.1038/nature14192

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

Scior, A., Preissler, S., Koch, M., & Deuerling, E. (2011). Directed PCR-free engineering of highly repetitive DNA sequences. *BMC Biotechnology*, *11*(1), 87. https://doi.org/10.1186/1472-6750-11-87

Shao, X., Joergensen, A. M., Howlett, N. G., Lisby, M., & Oestergaard, V. H. (2020). A distinct role for recombination repair factors in an early cellular response to transcription–replication conflicts. *Nucleic Acids Research*, *48*(10), 5467–5484. https://doi.org/10.1093/nar/gkaa268

Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., & Koseki, H. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, *450*(7171), 908–912. https://doi.org/10.1038/nature06397

Singh, U., Bongcam-Rudloff, E., & Westermark, B. (2009). A DNA Sequence Directed Mutual Transcription Regulation of HSF1 and NFIX Involves Novel Heat Sensitive Protein Interactions. *PLoS ONE*, *4*(4), e5050. https://doi.org/10.1371/journal.pone.0005050

Singh, U., Maturi, V., Jones, R. E., Paulsson, Y., Baird, D. M., & Westermark, B. (2014). CGGBP1 phosphorylation constitutes a telomere-protection signal. *Cell Cycle*, *13*(1), 96–105. https://doi.org/10.4161/cc.26813

Singh, U., Roswall, P., Uhrbom, L., & Westermark, B. (2011). CGGBP1 regulates cell cycle in cancer cells. *BMC Molecular Biology*, *12*(1), 28. https://doi.org/10.1186/1471-2199-12-28

Singh, U., & Westermark, B. (2011). CGGBP1 is a nuclear and midbody protein regulating abscission. *Experimental Cell Research*, *317*(2), 143–150. https://doi.org/10.1016/j.yexcr.2010.08.019

Singh, U., & Westermark, B. (2015). CGGBP1—An indispensable protein with ubiquitous cytoprotective functions. *Upsala Journal of Medical Sciences*, *120*(4), Article 4. https://doi.org/10.3109/03009734.2015.1086451

Smith, P. A., Tripp, B. C., DiBlasio-Smith, E. A., Lu, Z., LaVallie, E. R., & McCoy, J. M. (1998). A plasmid expression system for quantitative in vivo biotinylation of thioredoxin fusion proteins in Escherichia coli. *Nucleic Acids Research*, *26*(6), 1414–1420. https://doi.org/10.1093/nar/26.6.1414

Sparks, J. L., Chistol, G., Gao, A. O., Räschle, M., Larsen, N. B., Mann, M., Duxin, J. P., & Walter, J. C. (2019). The CMG Helicase Bypasses DNA-Protein Cross-Links to Facilitate Their Repair. *Cell*, *176*(1–2), 167-181.e21. https://doi.org/10.1016/j.cell.2018.10.053

References

Spiegel, J., Cuesta, S. M., Adhikari, S., Hänsel-Hertsch, R., Tannahill, D., & Balasubramanian, S. (2021). G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biology*, *22*(1), 117. https://doi.org/10.1186/s13059-021-02324-z

Srivatsan, A., Tehranchi, A., MacAlpine, D. M., & Wang, J. D. (2010). Co-Orientation of Replication and Transcription Preserves Genome Integrity. *PLOS Genetics*, *6*(1), e1000810. https://doi.org/10.1371/journal.pgen.1000810

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., & Tyers, M. (2006). BioGRID: A general repository for interaction datasets. *Nucleic Acids Research*, *34*(suppl_1), D535–D539. https://doi.org/10.1093/nar/gkj109

Steinböck, F. A., & Wiche, G. (1999). Plectin: A cytolinker by design. *Biological Chemistry*, *380*(2), 151–158. https://doi.org/10.1515/BC.1999.023

St Germain, C. P., Zhao, H., Sinha, V., Sanz, L. A., Chédin, F., & Barlow, J. H. (2022). Genomic patterns of transcription–replication interactions in mouse primary B cells. *Nucleic Acids Research*, *50*(4), 2051–2073. https://doi.org/10.1093/nar/gkac035

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*, e17548. https://doi.org/10.7554/eLife.17548

Šviković, S., Crisp, A., Tan-Wong, S. M., Guilliam, T. A., Doherty, A. J., Proudfoot, N. J., Guilbaud, G., & Sale, J. E. (2019). R-loop formation during S phase is restricted by PrimPol-mediated repriming. *The EMBO Journal*, *38*(3), e99793. https://doi.org/10.15252/embj.201899793

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K. P., Kuhn, M., Bork, P., Jensen, L. J., & von Mering, C. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, *43*(D1), D447–D452. https://doi.org/10.1093/nar/gku1003

Tanis, S. E. J., Jansen, P. W. T. C., Zhou, H., Heeringen, S. J. van, Vermeulen, M., Kretz, M., & Mulder, K. W. (2018). Splicing and Chromatin Factors Jointly Regulate Epidermal Differentiation. *Cell Reports*, *25*(5), 1292-1303.e5. https://doi.org/10.1016/j.celrep.2018.10.017

Tehranchi, A. K., Blankschien, M. D., Zhang, Y., Halliday, J. A., Srivatsan, A., Peng, J., Herman, C., & Wang, J. D. (2010). The transcription factor DksA prevents conflicts between DNA replication and transcription machinery. *Cell*, *141*(4), 595–605. https://doi.org/10.1016/j.cell.2010.03.036

The ENCODE Project Consortium. (2012). An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature, 489*(7414), 57–74. https://doi.org/10.1038/nature11247

References

The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Tous, C., & Aguilera, A. (2007). Impairment of transcription elongation by R-loops in vitro. *Biochemical and Biophysical Research Communications*, *360*(2), 428–432. https://doi.org/10.1016/j.bbrc.2007.06.098

Trakselis, M. A., Seidman, M. M., & Brosh, R. M. (2017). Mechanistic insights into how CMG helicase facilitates replication past DNA roadblocks. *DNA Repair*, *55*, 76–82. https://doi.org/10.1016/j.dnarep.2017.05.005

Tran, P. L. T., Pohl, T. J., Chen, C.-F., Chan, A., Pott, S., & Zakian, V. A. (2017). PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nature Communications*, *8*, 15025. https://doi.org/10.1038/ncomms15025

Traven, A., & Heierhorst, J. (2005). SQ/TQ cluster domains: Concentrated ATM/ATR kinase phosphorylation site regions in DNA-damage-response proteins. *BioEssays*, *27*(4), 397–407. https://doi.org/10.1002/bies.20204

Tsirkas, I., Dovrat, D., Thangaraj, M., Brouwer, I., Cohen, A., Paleiov, Z., Meijler, M. M., Lenstra, T., & Aharoni, A. (2022). Transcription-replication coordination revealed in single live cells. *Nucleic Acids Research*, *50*(4), 2143–2156. https://doi.org/10.1093/nar/gkac069

Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., De Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Coquelle, A., & Pasero, P. (2009). Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nature Cell Biology*, *11*(11), 1315–1324. https://doi.org/10.1038/ncb1984

Udeshi, N. D., Pedram, K., Svinkina, T., Fereshetian, S., Myers, S. A., Aygun, O., Krug, K., Clauser, K., Ryan, D., Ast, T., Mootha, V. K., Ting, A. Y., & Carr, S. A. (2017). Antibodies to biotin enable large-scale detection of biotinylation sites on proteins. *Nature Methods*, *14*(12), 1167–1170. https://doi.org/10.1038/nmeth.4465

Ummethum, H., & Hamperl, S. (2020). Proximity Labeling Techniques to Study Chromatin. *Frontiers in Genetics*, *11*, 450. https://doi.org/10.3389/fgene.2020.00450

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: Image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Varier, R. A., de Santa Pau, E. C., van der Groep, P., Lindeboom, R. G. H., Matarese, F., Mensinga, A., Smits, A. H., Edupuganti, R. R., Baltissen, M. P., Jansen, P. W. T. C., ter Hoeve, N., van Weely, D. R., Poser, I., van Diest, P. J., Stunnenberg, H. G., & Vermeulen, M. (2016). Recruitment of the Mammalian Histone-modifying EMSY Complex to Target Genes Is Regulated by ZNF131*. *Journal of Biological Chemistry*, *291*(14), 7313–7324. https://doi.org/10.1074/jbc.M115.701227

References

Varley, K. E., Gertz, J., Bowling, K. M., Parker, S. L., Reddy, T. E., Pauli-Behn, F., Cross, M. K., Williams, B. A., Stamatoyannopoulos, J. A., Crawford, G. E., Absher, D. M., Wold, B. J., & Myers, R. M. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Research*, *23*(3), 555–567. https://doi.org/10.1101/gr.147942.112

Vassin, V. M., Anantha, R. W., Sokolova, E., Kanner, S., & Borowiec, J. A. (2009). Human RPA phosphorylation by ATR stimulates DNA synthesis and prevents ssDNA accumulation during DNA-replication stress. *Journal of Cell Science*, *122*(22), 4070–4080. https://doi.org/10.1242/jcs.053702

Vishnoi, N., Dhanasekeran, K., Chalfant, M., Surovstev, I., Khokha, M. K., & Lusk, C. P. (2020). Differential turnover of Nup188 controls its levels at centrosomes and role in centriole duplication. *Journal of Cell Biology*, *219*(3). https://doi.org/10.1083/jcb.201906031

Wansink, D. G., Manders, E. E., van der Kraan, I., Aten, J. A., van Driel, R., & de Jong, L. (1994). RNA polymerase II transcription is concentrated outside replication domains throughout S-phase. *Journal of Cell Science*, *107*(6), 1449–1456. https://doi.org/10.1242/jcs.107.6.1449

Wei, X., Samarabandu, J., Devdhar, R. S., Siegel, A. J., Acharya, R., & Berezney, R. (1998). Segregation of transcription and replication sites into higher order domains. *Science*, *281*(5382), 1502–1506. https://doi.org/10.1126/science.281.5382.1502

Wessel, S. R., Mohni, K. N., Luzwick, J. W., Dungrawala, H., & Cortez, D. (2019). Functional Analysis of the Replication Fork Proteome Identifies BET Proteins as PCNA Regulators. *Cell Reports*, *28*(13), 3497-3509.e4. https://doi.org/10.1016/j.celrep.2019.08.051

West, D. C., Pan, D., Tonsing-Carter, E. Y., Hernandez, K. M., Pierce, C. F., Styke, S. C., Bowie, K. R., Garcia, T. I., Kocherginsky, M., & Conzen, S. D. (2016). GR and ER Coactivation Alters the Expression of Differentiation Genes and Associates with Improved ER+ Breast Cancer Outcome. *Molecular Cancer Research*, *14*(8), 707–719. https://doi.org/10.1158/1541-7786.MCR-15-0433

Willems, T., Gymrek, M., Highnam, G., Mittelman, D., & Erlich, Y. (2014). The landscape of human STR variation. *Genome Research*, *24*(11), 1894–1904. https://doi.org/10.1101/gr.177774.114

Wilson, T. E., Arlt, M. F., Park, S. H., Rajendran, S., Paulsen, M., Ljungman, M., & Glover, T. W. (2015). Large transcription units unify copy number variants and common fragile sites arising under replication stress. *Genome Research*, *25*(2), 189–200. https://doi.org/10.1101/gr.177121.114

Wiśniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nature Methods*, *6*(5), 359–362. https://doi.org/10.1038/nmeth.1322

Xue, M., Hou, J., Wang, L., Cheng, D., Lu, J., Zheng, L., & Xu, T. (2017). Optimizing the fragment complementation of APEX2 for detection of

specific protein-protein interactions in live cells. *Scientific Reports*, *7*(1), 12039. https://doi.org/10.1038/s41598-017-12365-9

Yan, Q., Wulfridge, P., Doherty, J., Fernandez-Luna, J. L., Real, P. J., Tang, H.-Y., & Sarma, K. (2022). Proximity labeling identifies a repertoire of site-specific R-loop modulators. *Nature Communications*, *13*(1), 53. https://doi.org/10.1038/s41467-021-27722-6

Yang, J., & Zhang, Y. (2015). I-TASSER server: New development for protein structure and function predictions. *Nucleic Acids Research*, *43*(W1), W174-181. https://doi.org/10.1093/nar/gkv342

Yellan, I., Yang, A. W. H., & Hughes, T. R. (2021). Diverse Eukaryotic CGG-Binding Proteins Produced by Independent Domestications of hAT Transposons. *Molecular Biology and Evolution*, *38*(5), 2070–2075. https://doi.org/10.1093/molbev/msab007

Yen, J. C., Chang, F. J., & Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, *4*(3), 370–378. https://doi.org/10.1109/83.366472

Yeo, C. Q. X., Alexander, I., Lin, Z., Lim, S., Aning, O. A., Kumar, R., Sangthongpitag, K., Pendharkar, V., Ho, V. H. B., & Cheok, C. F. (2016). P53 Maintains Genomic Stability by Preventing Interference between Transcription and Replication. *Cell Reports*, *15*(1), 132–146. https://doi.org/10.1016/j.celrep.2016.03.011

Yokoyama, Y., Zhu, H., Lee, J. H., Kossenkov, A. V., Wu, S. Y., Wickramasinghe, J. M., Yin, X., Palozola, K. C., Gardini, A., Showe, L. C., Zaret, K. S., Liu, Q., Speicher, D., Conejo-Garcia, J. R., Bradner, J. E., Zhang, Z., Sood, A. K., Ordog, T., Bitler, B. G., & Zhang, R. (2016). BET Inhibitors Suppress ALDH Activity by Targeting ALDH1A1 Super-Enhancer in Ovarian Cancer. *Cancer Research*, *76*(21), 6320–6330. https://doi.org/10.1158/0008-5472.CAN-16-0854

Zaidi, S. K., Boyd, J. R., Grandy, R. A., Medina, R., Lian, J. B., Stein, G. S., & Stein, J. L. (2016). Expression of Ribosomal RNA and Protein Genes in Human Embryonic Stem Cells Is Associated With the Activating H3K4me3 Histone Mark. *Journal of Cellular Physiology*, *231*(9), 2007–2013. https://doi.org/10.1002/jcp.25309

Zaratiegui, M., Castel, S., Irvine, D. V., Kloc, A., Ren, J., Li, F., Castro, E. de, Marín, L., Chang, A.-Y., Goto, D., Cande, W. Z., Antequera, F., Arcangioli, B., & Martienssen, R. A. (2011). RNAi promotes heterochromatic silencing through replication-coupled release of RNA polII. *Nature*, *479*(7371), 135. https://doi.org/10.1038/nature10501

Zhang, F., & Yu, X. (2011). WAC, a functional partner of RNF20/40, regulates histone H2B ubiquitination and gene transcription. *Molecular Cell*, *41*(4), 384–397. https://doi.org/10.1016/j.molcel.2011.01.024

Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P., & Wang, L. (2014). CrossMap: A versatile tool for coordinate conversion between genome

assemblies. *Bioinformatics (Oxford, England)*, *30*(7), 1006–1007. https://doi.org/10.1093/bioinformatics/btt730

Zhao, J., Bacolla, A., Wang, G., & Vasquez, K. M. (2010). Non-B DNA structure-induced genetic instability and evolution. *Cellular and Molecular Life Sciences: CMLS*, *67*(1), 43–62. https://doi.org/10.1007/s00018-009-0131-2

Zhao, X.-N., & Usdin, K. (2016). Ups and Downs: Mechanisms of Repeat Instability in the Fragile X-Related Disorders. *Genes*, *7*(9). https://doi.org/10.3390/genes7090070

Zhong, W., Feng, H., Santiago, F. E., & Kipreos, E. T. (2003). CUL-4 ubiquitin ligase maintains genome stability by restraining DNA-replication licensing. *Nature*, *423*(6942), 885–889. https://doi.org/10.1038/nature01747

6 Appendices

6.1 Acknowledgements

I started this project in September 2018 and it has been quite a rollercoaster with some very challenging times, but it has also created a huge number of fond memories. I would like to express my heartfelt gratitude to everyone that helped me reach this tremendous goal.

First and foremost, I would like to deeply thank you, Stephan, for inviting me to be a part of your newly-founded lab. Thank you for your absolutely terrific guidance and unwavering support. I am very grateful for all the things you taught me in the lab, but most importantly for elevating my scientific thinking to the next level. Your extremely positive attitude towards "negative" results kept me sane and helped me enormously every step of the way. I admire the way you led the lab and I aspire to follow in your footsteps should I at some point have my own lab.

Also, a big thank you to Maria-Elena. Your feedback on my project during the Thesis Advisory Committee meetings, but also after my presentations were absolutely crucial for me to be able to finish this thesis.

I would like to thank all the members of the Hamperl lab for providing me with an absolutely fantastic work environment. A special thanks to Elisabeth for the world class lab organization, teaching me numerous techniques in the lab and helping out whenever I had problems. Your positive attitude and great humor brightened many days. A heartfelt thanks to Anna for being the best bench neighbor I could wish for. Your positive attitude helped me tremendously to stay motivated throughout the years. Of course, I would also like to thank the four M's; Matthias, Manuel, Maxime and Marcel. Matthias, thank you for making many days much more fun through your great humor. It was really great to have you in the lab from the beginning to the end of my project. Manuel, thank you for your huge help with all automation questions, especially for the initial image analysis script. It was also very fun to talk about beer with you. Maxime, thank you for answering all my questions and for all the great discussions. I am also very thankful for Maxime's Master Macro for Everything and Anything. Your humor was a great addition to the lab. Marcel, thank you for the great scientific discussions and countless weekend bouldering sessions. Also, a thank you to the newest members of the lab, Clare

and Atiqa, for keeping up the great atmosphere in the lab. A big thank you to Alexandros for helping me a lot in the lab during a very critical time. I am very grateful for the administrative help I received from Marie-Sophie and Juana, as this relieved me of a lot of stress.

I am also grateful for everyone else at the IES for being very nice and helping me whenever they could. I am thankful for the great discussions during seminars and unforgettable discussions at evening get-togethers. It was truly amazing to work at such an amazing international institute. I am very thankful for the great atmosphere and countless events such as the international lunch, cinema club or the sports day. I would also like to thank everyone who joined regularly for lunch. I will never forget some of the conversations we had about truly unique topics.

A special thanks to Andreas for the maintenance of the microscopes and his magnificent help whenever I had questions. I am also very grateful Anne for the swift processing of the mass spectrometry samples and help with any questions I had. I would also like to thank Gabriele for helping me with the RNA sequencing data analysis.

I extend my heartfelt thanks to all the other members of my Thesis Advisory Committee, Prof. Dr. Heinrich Leonhardt and Prof. Dr. Karlene Cimprich for Their constructive criticism and for always pointing me in the right direction throughout the years.

I am also very thankful for my great Großhaderner flat mates throughout the years. Thank you Basti for always animating me to go out and do things in Munich. Meike, thank you for motivating us to cook regularly and introducing the concept of tidiness. Kevin, for inspiring me to join hikes and also live a healthier life in general. I am also truly grateful to all of my Forchheimer friends for their love and support throughout the years. Especially during the pandemic, the weekly online meetings kept me sane.

A special thanks to my brother Janek, Martina and my niece Johanna for their love and showing me that there is also a less stressful part to life. Finally, I would like to thank my parents Annette and Norbert. Your love and support are the main reason I was able to achieve this goal. Thank you for always being there when I needed you.

6.2 Supplementary data

Table S1: Summary of plasmids created for proximity labeling experiments.

ID	Plasmid description	Linker	Tag	Gene	Protein size in kDa	Details
K133	pcDNA3.1_APEX2a-FLAG	-	FLAG	-	22.6	control
K143	pcDNA3.1_CDC45_GR6_FLAG_APEX2a	GR6	FLAG	CDC45	123.3	replication
K144	pcDNA3.1_MCM2_GR6_FLAG_APEX2a	GR6	FLAG	MCM2	159.7	replication
K145	pcDNA3.1_MCM2_GR6_FLAG_APEX2	GR6	FLAG	MCM2	165	replication
K148	pcDNA3.1_CDC45_GlySer_FLAG_APEX2a	GS	FLAG	CDC45	89.5	replication
K149	pcDNA3.1_MCM2_GlySer_FLAG_APEX2a	GS	FLAG	MCM2	125.8	replication
K150	pcDNA3.1_MCM2_GlySer_FLAG_APEX2	GS	FLAG	MCM2	131.1	replication
K158	pcDNA_APEX2b-V5-GlySer-POLR2B_PuroR	GS	V5	POLR2B	142.1	transcription
K159	pcDNA_APEX2b-V5-GlySer-POLR2E_PuroR	GS	V5	POLR2E	32.8	transcription
K160	pcDNA_MCM2-GlySer-V5-APEX2b_PuroR	GS	V5	MCM2	110.1	replication
K161	pcDNA_APEX2b-V5-GR6-POLR2B_PuroR	GR6	V5	POLR2B	176	transcription
K162	pcDNA_APEX2b-V5-GR6-POLR2E_PuroR	GR6	V5	POLR2E	66.6	transcription
K163	pcDNA_MCM2-GR6-V5-APEX2b_PuroR	GR6	V5	MCM2	144	replication
K164	pcDNA_V5-APEX2b_PuroR	-	V5	-	6.9	control
K175	pcDNA_Psf2-GR6-V5-APEX2b_PuroR	GR6	V5	Psf2	63.5	replication
K176	pcDNA_Psf2-GS-V5-APEX2b_PuroR	GS	V5	Psf2	29.6	replication
K177	pcDNA_Psf3-GR6-V5-APEX2b_PuroR	GR6	V5	Psf3	66.6	replication
K178	pcDNA_Psf3-GS-V5-APEX2b_PuroR	GS	V5	Psf3	32.7	replication
K179	pcDNA_Psf3-noEx2-GR6-V5-APEX2b_PuroR	GR6	V5	Psf3	57.7	replication
K180	pcDNA_Psf3-noEx2-GS-V5-APEX2b_PuroR	GS	V5	Psf3	23.8	replication
K220	pcDNA3.1-MCM3-GR6-APEX2	GR6	FLAG	MCM3	154.4	replication
K221	pcDNA3.1-MCM3-GS-APEX2	GS	FLAG	MCM3	120.5	replication
K222	pcDNA3.1-MCM5-GR6-APEX2	GR6	FLAG	MCM5	145.6	replication
K223	pcDNA3.1-MCM5-GS-APEX2	GS	FLAG	MCM5	111.7	replication
K216	pcDNA_APEX2-V5-GlySer-POLR2B-NLS_PuroR	GS	V5	POLR2B	164.5	transcription
K217	pcDNA_APEX2-V5-GlySer-POLR2E-NLS_PuroR	GS	V5	POLR2E	55.2	transcription
K218	pcDNA_APEX2-V5-GlySer-NLS_PuroR	GS	V5	-	30.8	control
K181	pSBtet-Pur_MCM2-FLAG-APEX2	GS	FLAG	MCM2	133.5	replication
K182	pSBtet-Pur_APEX2-V5-POLR2E-NLS	GS	V5	POLR2E	57.7	transcription
K183	pSBtet-Pur_APEX2-V5-NLS	-	V5	-	33.2	control
K225	pSBtet-Pur_CDC45-FLAG-APEX2	GS	FLAG	CDC45	95.4	replication
K226	pSBtet-Pur_MCM3-FLAG-APEX2	GS	FLAG	MCM3	122.9	replication
K234	pSBtet-Pur_CDC45-FRB-FLAG-AP	GS	FLAG	CDC45	103	replication
K235	pSBtet-Pur_MCM2-FRB-FLAG-AP	GS	FLAG	MCM2	139	replication
K236	pSBtet-Hyg_CDC73-FKBP-V5-EX	GS	V5	CDC73	83	transcription
K237	pSBtet-Hyg_EX-V5-FKBP-POLR2E-NLS	GS	V5	POLR2E	48	transcription

Table S2: List of oligonucleotides

Oligo	Sequence	Description
GAPDH_fw	GCTCCCTCTTTCTTTGCAGC	Primer used for qPCR to determine genomic GAPDH DNA amount as reference
GAPDH_fw	ACCATGAGTCCTTCCACGAT	Primer used for qPCR to determine genomic GAPDH DNA amount as reference
APEX2_1_fw	GCGTTCTGGATTTGAGGGTC	Primer used for qPCR to determine APEX2 copy number
APEX2_1_rev	CGTCCGCTGCATATTTGTCA	Primer used for qPCR to determine APEX2 copy number
APEX2_2_fw	CCGTTGAGAAGGCGAAGAAG	Primer used for qPCR to determine APEX2 copy number
APEX2_2_rev	GCGATGTCAAGACCGTTGTT	Primer used for qPCR to determine APEX2 copy number
APEX2_3_fw	TCAGCTACCTTCTGACAAGGC	Primer used for qPCR to determine APEX2 copy number
APEX2_3_rev	CGGAAAGCTTTTGGTGAGCC	Primer used for qPCR to determine APEX2 copy number
CGGBP1_cDNA_fw	TATCCTTACGATGTACCAGACTATGCTGAGCGATTTGTAGTAACAGC	Primer for amplification of CGGBP1 cDNA, see Singh et al., 2014
CGGBP1_cDNA_rev	TATAGCGGCCGCTCAACAATCTTGTGAGTTGAG	Primer for amplification of CGGBP1 cDNA, see Singh et al., 2014
CGGBP1_PCR2_fw	TGTAGGTACCGCCACCATGGGATATCCTTACGATGTACCAGACTATGCT	Primer for amplification of CGGBP1 cDNA with HA tag at 5' end, see Singh et al., 2014
CGGBP1_PCR2_rev	TATAGCGGCCGCTCAACAATCTTGTGAGTTGAG	Primer for amplification of CGGBP1 cDNA with stop codon at 3' end, see Singh et al., 2014
MOV10_ex1_fw	CGGCTGCGGACCATTTATAA	Primer used for RT-qPCR to determine transcription levels of MOV10
MOV10_ex2_rev	CTGCGGTTGGTGGGTTTTAA	Primer used for RT-qPCR to determine transcription levels of MOV10
MALAT1_ex1_fw	AAAAGCAGACCCAGAGCAGT	Primer used for RT-qPCR to determine transcription levels of MALAT1
MALAT1_ex1_rev	CCTGAAAGTGCTCACAAGGC	Primer used for RT-qPCR to determine transcription levels of MALAT1
TLCD1_ex1_fw	CTCTGTCGCCTGCCCCTA	Primer used for RT-qPCR to determine transcription levels of TLCD1
TLCD1_ex1_fw	CGCCGTCTCAATCTCCACTA	Primer used for RT-qPCR to determine transcription levels of TLCD1
UHRF1_ex1_fw	ACTCGCTGTCCAGGCTGA	Primer used for RT-qPCR to determine transcription levels of UHRF1
UHRF1_ex2_rev	ATGGTGTCATTCAGGCGGA	Primer used for RT-qPCR to determine transcription levels of UHRF1
ZNF703_ex1_fw	GTGTCCCTCTTGCCACCG	Primer used for RT-qPCR to determine transcription levels of ZNF703

Oligo	Sequence	Description
ZNF703_ex2_rev	AAGGGGCTCTTCTTGGCG	Primer used for RT-qPCR to determine transcription levels of ZNF703
MCM3_ex2_fw	ACCGGCTGATTGTCAATGTG	Primer used for RT-qPCR to determine transcription levels of MCM3
MCM3_ex3_rev	TGCCTTCCAGTCCTACGTAG	Primer used for RT-qPCR to determine transcription levels of MCM3
CGGBP1_ex1_fw	TGCCATTAGTGACCACCTC	Primer used for RT-qPCR to determine transcription levels of CGGBP1
CGGBP1_ex1_rev	TCTCAAGTGGGATGTTGGC	Primer used for RT-qPCR to determine transcription levels of CGGBP1
Tet-ON_fw	GGCTAGCAAGCTTGATGTG	Primer used for RT-qPCR to determine transcription levels and DRIP enrichment downstream of the $(CGG)_{10}$ repeat on plasmid pHU43
Tet-ON_rev	GCAATAGCATCACAAATTTCACA	Primer used for RT-qPCR to determine transcription levels and DRIP enrichment downstream of the $(CGG)_{10}$ repeat on plasmid pHU43
OriP_fw	TTTTCGCTGCTTGTCCTTTT	Primer used for qPCR to determine plasmid copy number
OriP_rev	TTTTCGCTGCTTGTCCTTTT	Primer used for qPCR to determine plasmid copy number
β-actin_in1_fw	CGGGGTCTTTGTCTGAGC	Primer used for qPCR to determine genomic β-actin DNA amount as reference
β-actin_in1_rev	CAGTTAGCGCCCAAAGGAC	Primer used for qPCR to determine genomic β-actin DNA amount as reference

Curriculum vitae

Henning Ummethum
email: henning.ummethum@gmail.com

EDUCATION	
09.2018 – expected 2023	PhD Biology, Helmholtz Zentrum München/Ludwig Maximilian University of Munich, Munich, Germany
10.2016 – 07.2018	MSc. Biotechnology/Bioengineering, Hochschule Weihenstephan-Triesdorf, Freising, Germany
10.2011 – 09.2016	BSc. Biotechnology, Hochschule Weihenstephan-Triesdorf, Freising, Germany

WORK EXPERIENCE

Doctoral thesis

Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, Hamperl Group

- Developed an unbiased screening based on proximity labeling with APEX2 to identify factors of transcription-replication conflicts (TRCs)
- Performed follow-up experiments on one candidate with the main focus on TRCs, DNA damage, DNA secondary structures via ChIP-seq data, immunofluorescence imaging, proximity ligation assay and DNA:RNA hybrid immunoprecipitation (DRIP)

Master thesis

Research Unit Protein Science, Helmholtz Zentrum München

- Studied the effect of hyperglycemia on the proteome and migration behaviour of keratinocytes via quantitative mass spectrometry
- Main focus on mass spectrometry data analysis

Bachelor thesis

- Cloning and activity testing of enzymes that are part of the plant hyperforin metabolism pathway in *S. cerevisiae*

List of publications

Ummethum, H., Lalonde, M., Werner, M., Trauner, M., Chanou, A., Weiβ, M., Lee, C. S. K., Kruse, E., Ettinger, A., & Hamperl, S. (2023). The CGG triplet repeat binding protein 1 counteracts DNA secondary structure-induced transcription-replication conflicts (p. 2023.03.09.531843). bioRxiv. https://doi.org/10.1101/2023.03.09.531843

Ummethum, H., & Hamperl, S. (2020). Proximity Labeling Techniques to Study Chromatin. *Frontiers in Genetics*, 11, 450. https://doi.org/10.3389/fgene.2020.00450

Lalonde, M., **Ummethum, H.**, Trauner, M., Ettinger, A., & Hamperl, S. (2023). An automated image analysis pipeline to quantify the coordination and overlap of transcription and replication activity in mammalian genomes. *Methods in Cell Biology*. https://doi.org/10.1016/bs.mcb.2023.05.012

Weiβ, M., Chanou, A., Schauer, T., Tvardovskiy, A., Meiser, S., König, A.-C., Schmidt, T., Kruse, E., **Ummethum, H.**, Trauner, M., Werner, M., Lalonde, M., Hauck, S. M., Scialdone, A., & Hamperl, S. (2023). Single-copy locus proteomics of early- and late-firing DNA replication origins identifies a role of Ask1/DASH complex in replication timing control. *Cell Reports*, 42(2), 112045. https://doi.org/10.1016/j.celrep.2023.112045