GENOME-WIDE EFFECTS OF DNA REPLICATION ON NUCLEOTIDE EXCISION REPAIR OF UV-INDUCED DNA LESIONS.

by
CEM AZGARI

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfilment of
the requirements for the degree of
Master of Science

Sabancı University
August 2020

GENOME-WIDE EFFECTS OF DNA REPLICATION ON NUCLEOTIDE EXCISION REPAIR OF UV-INDUCED DNA LESIONS.

Approved by:

Dr. Ogün Adebali .....
(Thesis Supervisor)

Prof. Dr. Batu Erman .....

Prof. Dr. Halil Kavaklı .....

Date of Approval: August 28, 2020

ABSTRACT

GENOME-WIDE EFFECTS OF DNA REPLICATION ON NUCLEOTIDE EXCISION REPAIR OF UV-INDUCED DNA LESIONS.

CEM AZGARI

MOLECULAR BIOLOGY, GENETICS AND BIOENGINEERING M.S. THESIS, August 2020

Thesis Supervisor: Asst. Prof. Ogün Adebali

Keywords: Nucleotide excision repair, UV damage, (6-4)PP, CPD, XR-seq, Damage-seq, DNA replication, DNA strand asymmetry

Replication can cause unrepaired DNA damages to turn into mutations that might lead to cancer. Nucleotide excision repair is the leading repair mechanism that prevents melanoma cancers by removing UV-induced bulky adducts. However, the role of replication on nucleotide excision repair, in general, is yet to be clarified. Recently developed methods Damage-seq and XR-seq map damage formation and nucleotide excision repair events respectively, in various conditions. Here, we applied Damage-seq and XR-seq methods to UV-irradiated HeLa cells synchronized at two stages of the cell cycle: early S phase, and late S phase. We analyzed the damage and repair events along with replication origins and replication domains of HeLa cells. We found out that in both early and late S phase cells, early replication domains are more efficiently repaired relative to late replication domains. The results also revealed that repair efficiency favors the leading strands around replication origins. Moreover, we observed that the repair efficiency of the strands around replication origins is inversely correlated with the number of melanoma mutations.

ÖZET

UV KAYNAKLI DNA HASARININ KESİP ÇIKARMALI ONARIMI İLE DNA REPLİKASYONUNUN GENOM ÇAPLI ETKİLEŞİMİ.

CEM AZGARİ

MOLEKÜLER BİYOLOJİ, GENETİK VE BİYOMÜHENDİSLİK YÜKSEK LİSANS TEZİ, Ağustos 2020

Tez Danışmanı: Dr. Ogün Adebali

Anahtar Kelimeler: Nükleotid kesip çıkarmalı onarımı, UV hasarı, (6-4)PP, CPD, XR-seq, Damage-seq, DNA replikasyonu, DNA zinciri asimetrisi

Replikasyon, onarılmamış DNA hasarlarının kansere yol açabilecek mutasyonlara dönüşmesine neden olabilir. Nükleotid kesip çıkarmalı onarımı, UV ile indüklenen hacimli DNA katımlarını ortadan kaldırarak melanom kanserlerini önleyen önde gelen onarım mekanizmasıdır. Ancak, replikasyonun nükleotid kesip çıkarmalı onarımındaki rolü henüz açığa kavuşturulmamıştır. Son zamanlarda geliştirilen yöntemler Damage-seq ve XR-seq sırasıyla, hasar oluşumu ve nükleotid kesip çıkarmalı onarımı olaylarını çeşitli koşullar altında haritalandırabilmektedir. Burada, Damage-seq ve XR-seq yöntemlerini hücre döngüsünün erken ve geç S fazlarında senkronize edilip UV ile indüklenen HeLa hücrelerine uyguladık. HeLa hücrelerinin hasar ve onarım olaylarını replikasyon orijini ve replikasyon alanlarıyla birlikte analiz ettik. Hem erken hem de geç S fazlı hücrelerde, erken replikasyon alanlarının geç replikasyon alanlarına göre daha verimli bir şekilde onarıldığını bulduk. Sonuçlar ayrıca onarım verimliliğinin replikasyon orijinleri etrafında DNA'nın öncü ipliklerini desteklediğini ortaya koydu. Dahası, replikasyon orijini etrafındaki ipliklerin onarım etkinliğinin melanom mutasyonlarının sayısı ile ters orantılı olduğunu gözlemledik.

ACKNOWLEDGEMENTS

First of all, I would like to express the deepest appreciation to my thesis advisor Dr. Ogün Adebali. Dr. Ogün Adebali's meticulous comments were an enormous help to me, and without his immense support, patience, and encouragement, this thesis would not have materialized. I will always be grateful for having a chance to study in his lab, where my scientific background, technical knowledge, and skeptical thinking improved considerably. I also want to thank the rest of my thesis jury, Prof. Dr. Batu Erman, and Prof. Dr. Halil Kavaklı for their time and interest in my thesis project.

I would like to thank all the members of ADEBALİLAB; Arda Çetin, Aylin Bircan, Berkay Selçuk, Burak İşlek, and Sezgi Kaya for their help, scientific comments, and especially for their friendship. Whether I need a company while sipping my coffee or I need a piece of insightful advice for my project, they have always been there for me, and I am grateful for that. I am also thankful to my undergraduate students Berk Turhan, Defne Çirci, and Zeynep Kılınç for their enthusiasm and friendship.

I would like to offer my special thanks to all my friends, who never stopped supporting me and relieving my mind when I needed them the most. Without them, my sanity would be at stake. My deepest gratitude goes to my family; my dad, who always trusted me, my mom with her constant care and interest in my studies. If it weren't for the importance she attached to education, I might not be a researcher today. To my grandmother, who is still tracking whether I finished my homework (thesis) or not, and to my brother Nedim Azgari, who without a doubt put the most effort to develop my personality and to boost my interest in learning. I have always admired his enthusiasm to improve and his mindset and taken him as a role model. Lastly, I would like to thank my fiancée Ecem Ornadis, who I considered being my greatest accomplishment. Since I knew her, she stands by my side, never letting me give up or back down. Without her support and trust, I might not even be a Master's student at Sabancı University in the first place.

To my fiancée...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATONS

ERCC1 Excision repair cross-complementation group 1 6

ERD Early replication domain 7, 22, 23, 24, 26, 32, 33, 34, 36

G Guanine 2

GR Globar repair 3, 4, 6, 20

HeLa Henrietta Lacks 12, 14, 15, 16, 20, 22, 24, 32, 33

HMGN1 High mobility group nucleosome-binding domain-containing protein 1 5

ICGC International Cancer Genome Consortium 26

kb Kilobase 13, 19, 23, 26, 27, 29, 31, 35

LRD Late replication domain 7, 22, 23, 24, 26, 32, 33, 34, 35, 36

Mb Megabase 12, 23, 25

OK-seq Okazaki fragment sequencing 25, 26, 27, 30

PBS Phosphate-buffered saline 16

PCNA Proliferating cell nuclear antigen 5, 6

PCR Polymerase chain reaction 10, 11, 17

RAD23B RAD23 Homolog B 4

RFC Replication factor C 5, 6

RFD Replication fork directionality 30, 35, 36

RNA Ribonucleic acid 4, 5

RNAPII RNA polymerase II 4, 5, 10

RPA Replication protein A 6

RPKM Reads per kilobase per million 23, 28, 29

SNS-seq Short nascent strand sequencing 25, 26, 27, 35

ssDNA Small single-stranded DNA 4, 5, 35

T Thymine 2, 17, 20, 26

TCR Transcription-coupled repair 3, 4, 5, 6, 7, 20, 21, 32

TFIIH Transcription initiation factor IIH 5, 6, 7, 11

TTD Trichothiodystrophy 6, 7

UCSC University of California, Santa Cruz 24

USP7 Ubiquitin-specific-processing protease 7 5

UTZ Up transition zone 7, 26

UV Ultraviolet 1, 2, 3, 4, 5, 6, 9, 12, 20, 22, 32, 34, 36

UV-DDB Ultraviolet radiation-DNA damage-binding protein 4

UvrA UvrABC system protein A 4

UvrB UvrABC system protein B 4

UvrC UvrABC system protein C 4

UVSSA UV-stimulated scaffold protein A 5

XAB2 XPA-binding protein 2 5

XP Xeroderma pigmentosum 6

XPA Xeroderma pigmentosum, complementation group A 6

XPB Xeroderma pigmentosum, complementation group B 5, 6, 7

XPC Xeroderma pigmentosum, complementation group C 4, 5, 6

XPD Xeroderma pigmentosum, complementation group D 5, 6, 7

XPE Xeroderma pigmentosum, complementation group E 6

XPF Xeroderma pigmentosum, complementation group F 6

XPG Xeroderma pigmentosum, complementation group G 6, 17

XR-seq Excision repair sequencing xi, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 28, 29, 31, 32

1. INTRODUCTION

1.1 UV-induced damages in humans

Ultraviolet (UV) light is the major cause of skin cancers in humans (Kiefer, 2007). It is a portion of the electromagnetic spectrum which is emitted from the sun together with visible light and heat. Based on its wavelength, UV light divides into three subgroups: UVA (wavelength of 315-400 nm), UVB (wavelength of 280-315 nm), and UVC (wavelength of 100-280 nm). While the less energetic UVA makes up the majority of UV light passing the atmosphere, all UVC and approximately 90% of UVB is either blocked or absorbed by the ozone layer. Even in these conditions, humans are not fully protected from the damaging effects of UV light. So that, UV-irradiation accounts for approximately 30.000 DNA lesions' formation per cell per hour.

The most abundant UV lesions in cellular DNA are pyrimidine dimers (Kielbassa, Roza & Epe, 1997), which are formed by the covalent bonds between the adjacent pyrimidines (Whitmore, Potten, Chadwick, Strickland & Morison, 2001). Different in their chemical structure, two types of pyrimidine dimers are exist; one is called cyclobutene pyrimidine dimers (CPDs), and the other is called pyrimidine (6-4) pyrimidone photoproducts [(6-4)PPs]. While both UVC and UVB can induce these dimer formations, UVA is only capable of inducing CPDs. Nonetheless, UVA induction can convert already formed (6-4)PPs into their Dewar valence isomers. Moreover, UVA can induce oxidative DNA damages through photosensitized reactions (Hu & Adar, 2017). Thanks to the development of time resolved spectroscopy techniques in recent years, dynamics of pyrimidine dimer formation is well known. The formation and biological properties of UV lesions will be briefly discussed in the subsections below.

1.1.1 Cyclobutene pyrimidine dimers (CPDs)

CPDs are the most frequent pyrimidine dimers that are arising from the covalent linkages between the consecutive pyrimidines, and it is characterized by the four-

member ring structure that are double bonded from the pyrimidine 5 and 6 (Whitmore et al., 2001). In vivo, CPDs can be observed in four different configurations: cis-syn, cis-anti, trans-syn, or trans-anti. (Khattak & Wang, 1972) While it is generally observed in cis-syn form when the DNA is double-stranded (Wacker, Dellweg, Träger, Kornhauser, Lodemann, Türck, Selzer, Chandra & Ishimoto, 1964), in denatured DNA and single-stranded regions, trans-syn configuration exists (Taylor & Brockie, 1988). Although it is rare, nonadjacent pyrimidines can also form CPDs in single-stranded regions (Nguyen & Minton, 1988). Moreover, different configurations can affect the ability of repair enzymes to recognize these lesions and correct them, which cause mutability differences between the configurations (Friedberg, Walker, Siede & Wood, 2005).

Apart from the configuration of the lesions, their dipyrimidine doublets (TT, TC, CT, and CC) can contribute to CPD formation at different rates depending on the type of UV exposure or the nucleotide content of the DNA. According to the study of Douki and Cadet, under UVC and UVB exposure, double-stranded mammalian DNA produces TT, TC, CT, and CC CPDs in 100:50:25:10 ratios, respectively (Douki & Cadet, 2001). While TT CPDs accounts for more than half of the total CPDs after the exposure of UVC and UVB, for UVA exposure, this ratio rises to 90% (Mouret, Philippe, Gracia-Chantegrel, Banyasz, Karpati, Markovitsi & Douki, 2010). On the other hand, TT CPDs are the abundant products of UV exposure for mammalian DNA, but the abundance might be greatly influenced by the GC percentage of the DNA. For example, in the bacterial DNA that possess a rich GC percentage, TT CPDs are the minor products of UV exposure (Patrick, 1977).

1.1.2 Pyrimidine (6-4) pyrimidone photoproducts [(6-4)PPs] and their Dewar valence isomers

(6-4)PPs form with occurrence of a pyrimidone ring by the bonding between C6 position of the 5'-end base and C4 position of the 3'-end base. In fact, this structure forms indirectly following the UV exposure, after a cyclic reaction intermediate, which can be either an oxetane if thymine is the 3'-end base, or azetidine if cytosine is the 3'-end base. Because of its indirect formation, (6-4)PPs emerge thousand times slower than CPDs (Schreier, Schrader, Koller, Gilch, Crespo-Hernández, Swaminathan, Carell, Zinth & Kohler, 2007).

Under UVC and UVB exposure, formation of (6-4)PPs is approximately five time less than that of CPDs (Douki & Cadet, 2001). Moreover, TT dipyrimidines, that are the most abundant sites for CPDs, are less frequent for (6-4)PPs. Instead, TC and CCs are the frequent sites for (6-4)PPs, while CT (6-4)PPs are uncommon. Another

unique property of (6-4)PPs is its conversion into Dewar valence isomers with the photoisomerization process (Taylor & Cohrs, 1987). Although UVB irradiation can trigger the process, with the combination of UVB and UVA exposure, the yield increases significantly.

1.2 Nucleotide excision repair in humans

Throughout the evolution, cells manage to evolve highly specialized repair mechanisms to cope with a variety of lesions that threaten the genome integrity and survival. Considering the diversity of these lesions, it would be unexpected to have only a single mechanism that can preserve the integrity of the genome. Hence, there are several repair mechanisms that cells utilize which are eminently conserved between species. Due to the removal of both strands, repair of a double strand break is generally more difficult to repair. There are two mechanisms that can be triggered by double strand breaks: homologous recombination, and non-homologous end-joining. Homologous recombination uses the sister-chromatid as a template to repair double strand breaks in an error-free manner. In addition, if sister-chromatid is not available for use, non-homologous end-joining directs the fusion of broken ends in an error-prone manner. Although being error-prone, non-homologous end-joining is the dominant mechanism for double strand break repair in mammals. Reasons of this dominancy are the distant proximity of chromatids to each other, and the DNA folding that makes the homologous sequence less reachable. In addition, imperfect matches by homologous recombination can lead to tragic outcomes such as creating repeated sequences (Li, Wehrenberg, Waldman & Waldman, 2018).

On the other hand, when a damage occurs on a single strand, the opposite strand can be used as a template. In such cases, DNA excision repair mechanisms remove the lesion site and re-synthesize the gap using the template strand. Base excision repair detects and repairs the oxidation, deamination and alkylation damages (Klungland, Höss, Gunz, Constantinou, Clarkson, Doetsch, Bolton, Wood & Lindahl, 1999). Mismatches that escape proofreading are identified and corrected by mismatch repair (Modrich, 1997). And lastly, bulky adducts caused by UV irradiation, environmental mutagens, and chemotherapeutic agents are removed by nucleotide excision repair (Reardon & Sancar, 2005). Nucleotide excision repair contains two sub pathways that differ from each other at the damage recognition step: global repair (GR) and transcription-coupled repair (TCR). TCR is specialized in recognizing adducts in transcribed regions, while GR can recognize bulky adducts at any site. Subsections below will address the assembly and main properties of nucleotide excision repair in more detail.

1.2.1 Repair of UV-induced damages by nucleotide excision repair

Identified firstly in E. coli by two independent studies published in 1964 (Boyce & Howard-Flanders, 1964; Setlow & Carrier, 1964), nucleotide excision repair can repair variety of bulky adducts from UV-induced pyrimidine dimers to chemotherapeutic agents such as cisplatin (Yimit, Adebali, Sancar & Jiang, 2019). Although repair mechanisms are highly conserved among the species, nucleotide excision repair in humans appeared to be surprisingly different from that of E. coli. While E. coli contains three proteins (UvrA, UvrB, UvrC) for the incision of damaged fragments, human nucleotide excision repair has sixteen proteins for the task. More interestingly, there is not an evolutionarily relevance between these human and E. coli proteins. In addition, the excised fragments are usually around 12 nucleotides long in E. coli. For humans, the length of these fragments are around 30 nucleotides (Sancar, 2016). Human nucleotide excision repair can be generally discussed in three steps: 1) damage recognition, 2) dual incision and excision of damaged fragments, and 3) re-synthesis and ligation.

1.2.1.1 Damage recognition

As mentioned earlier, GR and TCR have distinct damage recognition steps. GR scans the whole genome to detect helix distortions caused by bulky adducts, whereas TCR responds only to a stalled RNA polymerase II (RNAPII) during transcription.

In GR, three proteins (XPC, RAD23B, CETN2) work in coordination to recognize the lesion site (Sugasawa, Ng, Masutani, Iwai, van der Spek, Eker, Hanaoka, Bootsma & Hoeijmakers, 1998). XPC is the first protein to interact with the lesion by binding to the small single-stranded DNA (ssDNA) that is left unpaired due to the pyrimidine dimer formation at the opposite strand. The ability of XPC to bind unpaired ssDNA enables GR to detect a variety of lesions, since the unpaired ssDNA is a common characteristic of bulky adducts. After XPC binding, RAD23B and CETN2 interact with and stabilize XPC. However, helix distortions must be apparent to XPC for an efficient detection. (6-4)PPs are recognized relatively in ease because of having a prominent distortion (Mizukoshi, Kodama, Fujiwara, Furuno, Nakanishi & Iwai, 2001), whereas the distortion of CPDs cause only a 9^{o} unwinding with a 30^{o} bent (Park, Zhang, Ren, Nadji, Sinha, Taylor & Kang, 2002), which can be considered mild. For the detection of CPDs, proteins DDB1 and DDB2 form a complex called ultraviolet radiation-DNA damage-binding protein (UV-DDB). The complex directly interacts with the lesion, and DDB2 kinks the lesion to increase unwinding (Scrima, Koníčková, Czyzewski, Kawasaki, Jeffrey, Groisman, Nakatani,

Iwai, Pavletich & Thomä, 2008), as a result the ssDNA becomes detectable for XPC.

The recognition mechanism of TCR is triggered by the blockage of RNAPII, which transcribes the active gene during transcription elongation. When RNAPII stalls following an encounter with a lesion, it subsequently recruits the nucleotide excision repair proteins (Svejstrup, 2002). Afterwards, RNAPII dynamically interacts with UV-stimulated scaffold protein A (UVSSA), ubiquitin-specific-processing protease 7 (USP7), and Cockayne syndrome protein B CSB. CSB is an ATP-dependent chromatin remodeling factor that contains a helicase motif, surprisingly without a helicase activity (Selby & Sancar, 1997b). Moreover, studies in early 2000s revealed that point mutations in the ATPase domain of CSB protein significantly cripples the cell's ability to escape the inhibited RNA synthesis (Citterio, Rademakers, van der Horst, van Gool, Hoeijmakers & Vermeulen, 1998; Muftuoglu, Selzer, Tuo, Brosh Jr & Bohr, 2002), which suggests that CSB plays a key role for the TCR assembly. Furthermore, recruitment of repair factors that work on incision of the damaged fragment also mediated by CSB (Fousteri, Vermeulen, van Zeeland & Mullenders, 2006). More identified functions of CSB include transcription elongation, chromatin maintenance and remodeling, histone tail binding, and strand annealing (Selby & Sancar, 1997a). Another important Cockayne syndrome protein is CSA, which is also recruited by CSB. CSA mediates the recruitment of PCNA, RFC and DNA polymerase δ. Therefore, it is a key protein for the later events of the repair.

The recruited core nucleotide excision repair factors and some TCR specific factors such as UVSSA, USP7, XPA-binding protein 2 (XAB2) and high mobility group nucleosome-binding domain-containing protein 1 (HMGN1), gather on the lesion site where RNAPII stalls. However, because RNAPII stalls on the lesion, it covers the lesion so that the TCR complex cannot reach it (Tornaletti, Reines & Hanawalt, 1999). To proceed, RNAPII should somehow move from the 35 nucleotides length of strand where it is positioned. There are three proposed mechanisms for that purpose which are degradation, dissociation and backtracking. Because backtracking is already known to be occurring at transcription proofreading and at natural transcription pause sites, it is the most accepted mechanism among these three (Marteijn, Lans, Vermeulen & Hoeijmakers, 2014).

1.2.1.2 Dual incision and excision of damaged fragment

After RNAPII backtracks, transcription initiation factor IIH (TFIIH) initiates to unwind DNA with its helicase subunits. The TFIIH complex is formed of 10 proteins. While XPB and XPD have helicase activity, CDK-activating kinase (CAK) subcomplex is responsible for the initiation of TFIIH complex. The initiation is also

known as DNA damage verification step which is the last reversible step of nucleotide excision repair (Marteijn et al., 2014). With the initiation of TFIIH complex, the lesion becomes ready to be removed. Then XPF-ERCC1 and XPC endonucleases interact with the lesion site to catalyze the lesion from two sides together with the TFIIH complex. Meanwhile, replication protein A (RPA) not only protects the non-damaged single strand, but also interacts with and coordinates most subunits of TFIIH complex. The cleavage of the lesion site that yields 22-30 nucleotides long single stranded gap, is termed dual incision (Marteijn et al., 2014).

1.2.1.3 Re-synthesis and ligation

After the dual incision, the occurred gap must be filled with the ligation process. During replication, the proteins proliferating cell nuclear antigen (PCNA), replication factor C (RFC), DNA polymerase δ, DNA polymerase ϵ and DNA ligase 1 mediates re-synthesis and ligation. However, if the cell is non-replicating, then DNA polymerase κ and XRCC1– DNA ligase 3 fill the gap (Marteijn et al., 2014).

1.2.2 Nucleotide excision repair associated diseases

There are three human diseases that are known to be directly associated with nucleotide excision repair. These diseases are xeroderma pigmentosum (XP), cockayne syndrome (CS) and trichothiodystrophy (TTD) (De Boer & Hoeijmakers, 2000; Lehmann, 2003). XP discovered in 1968 as a hereditary disease that causes a defective nucleotide excision repair (Cleaver, 1968). XP patients are extremely photosensitive, so that they have approximately 5000-fold increased risk of UV-induced skin cancer. Dry parchment skin and pigmentation related anomalies are some of the hallmarks of this disorder (De Boer & Hoeijmakers, 2000). Seven genes that are associated with the disease, known as XP complementation groups (XPA, XPB, XPC, XPD, XPE, XPF, XPG) (Cleaver & Bootsma, 1975), and proteins that are produced by all these genes have a role in GR. Except XPC and XPE, they are also involved in TCR (Van Hoffen, Venema, Meschini, Van Zeeland & Mullenders, 1995).

CS first reported in 1936 as a disease related to deafness and dwarfism (Cockayne, 1936). In the upcoming years, problems at joints, vision, and calcifications in the brain are further reported (Cockayne, 1946; Neill & Dingwall, 1950). Moreover, these patients have aging related issues, and like XP patients, they are photosensitive, though not as severe as XP patients, therefore, their risk of having UV-induced skin cancer is not increased. As a consequence of all these abnormalities, most severe types of CS patients have a lifespan of as short as 7 years. Two genes, CSA and CSB are known to be related to the disease, which are both TCR proteins. Thereby, it

was thought that CS patients are TCR defective. However, since TCR deficiency is not enough to explain all these severe symptoms alone, a deficiency in transcription is also argued (Drapkin, Reardon, Ansari, Huang, Zawel, Ahn, Sancar & Reinberg, 1994).

TTD patients can display a broad range of symptoms from having brittle hair to low fertility and impaired intelligence. If the disorder is caused by one of the XPB, XPD or TTDA genes, which are all code for a component of TFIIH complex, TTD patients can become nucleotide excision repair deficient, hence photosensitive. Even though TFIIH complex can be functional, the levels of TFIIH complex decreases significantly (Giglia-Mari, Miquel, Theil, Mari, Hoogstraten, Ng, Dinant, Hoeijmakers & Vermeulen, 2006).

1.3 Replication and its contribution to mutagenesis

Owing to many potential replication origins, a mammalian cell replicates in approximately 10 hours (Takebayashi, Ogata & Okumura, 2017). During the cell division, only a portion of these replication origins fires, and they fire in an asynchronized manner except the replication origins that are in proximity to each other. By firing simultaneously, these closely packed replication origins coordinates the replication of regions longer than mega bases, termed as "replication domains" (Jackson & Pombo, 1998) (Figure 1.1). Replication domains are divided into 4: early replication domains (ERDs), late replication domains (LRDs) and the zones between these domains are up transition zones (UTZs) and down transition zones (DTZs) (Farkash-Amar, Lipson, Polten, Goren, Helmstetter, Yakhini & Simon, 2008; Hansen, Thomas, Sandstrom, Canfield, Thurman, Weaver, Dorschner, Gartler & Stamatoyannopoulos, 2010; Hiratani, Ryba, Itoh, Yokochi, Schwaiger, Chang, Lyou, Townes, Schübeler & Gilbert, 2008; Koren, Handsaker, Kamitaki, Karlić, Ghosh, Polak, Eggan & McCarroll, 2014; Nakayasu & Berezney, 1989; O'keefe, Henderson & Spector, 1992). Generally, the interior regions of the nucleus are replicated earlier than nuclear periphery regions, thus located at early replication domains (Dimitrova & Berezney, 2002). Multiple studies indicated that these domains are differ from each other in the mutation frequencies. Suggested by genome-wide analysis of mutation rates, early replication domains have reduced levels of mutation comparing to late replication domains (Lawrence, Stojanov, Polak, Kryukov, Cibulskis, Sivachenko, Carter, Stewart, Mermel, Roberts & others, 2013; Stamatoyannopoulos, Adzhubei, Thurman, Kryukov, Mirkin & Sunyaev, 2009). Also, in most cancers, base substitution mutation elevates in late replication domains (Schuster-Böckler & Lehner, 2012).

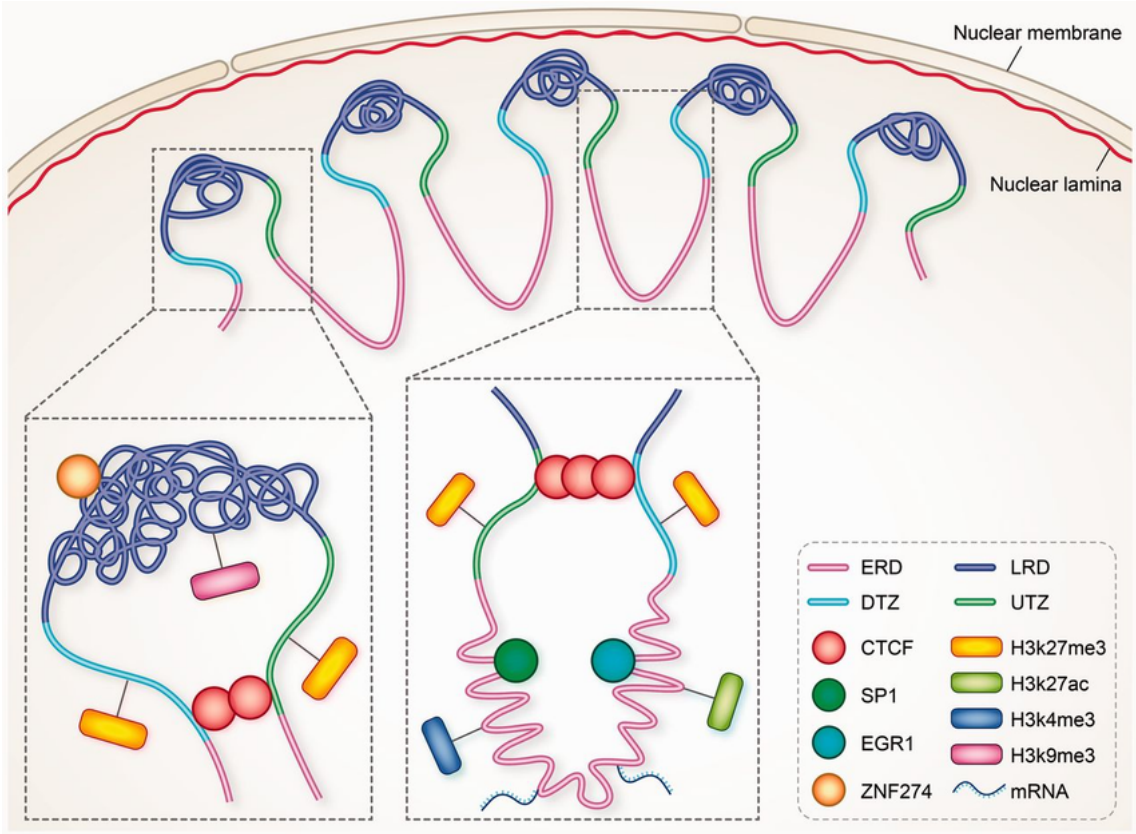

Figure 1.1 Model of replication domains and its chromatin organization (Liu et al., 2016).

Replication is driven by replication fork which is formed when a predefined replication origin fires (Langston, Indiani & O'Donnell, 2009). Replication fork usually proceeds bidirectionally, with the coordinated work of polymerases ϵ and δ. During the movement of the fork, polymerase ϵ continuously synthesizes the leading strand, whereas polymerase δ discontinuously synthesizes the lagging strand. Moreover, bidirectionality creates an asymmetric progress, so that two polymerases work on opposite strands towards different directions. In other words, in the left replicating fork, polymerase ϵ proceeds on the plus strand, while in the right replicating fork, it progresses on the minus strand. Studies suggest that this asymmetric progress of polymerases around the associated replication origins are reflected to the mutation profiles, where lagging strands reported to harbor more mutations than leading strands (Haradhvala, Polak, Stojanov, Covington, Shinbrot, Hess, Rheinbay, Kim, Maruvka, Braunstein & others, 2016; Lujan, Williams, Pursell, Abdulovic-Cui, Clark, McElhinny & Kunkel, 2012; Reijns, Kemp, Ding, de Procé, Jackson & Taylor, 2015; Shinbrot, Henninger, Weinhold, Covington, Göksenin, Schultz, Chao, Doddapaneni, Muzny, Gibbs & others, 2014). Observed asymmetry on mutation profiles is explained by the error-prone bypass mechanism on the lagging strands that makes it vulnerable to mutations (Seplyarskiy, Akkuratov, Akkuratova, Andri-

anova, Nikolaev, Bazykin, Adameyko & Sunyaev, 2019). Other studies argued that the attachment of helicase to leading strands increases the damage response, thus leading to effective repair of the strand (Hedglin & Benkovic, 2017; Yeeles, Poli, Marians & Pasero, 2013). Furthermore, many mutational signatures are reported to have a significant replication strand asymmetry (Tomkova et al., 2018).

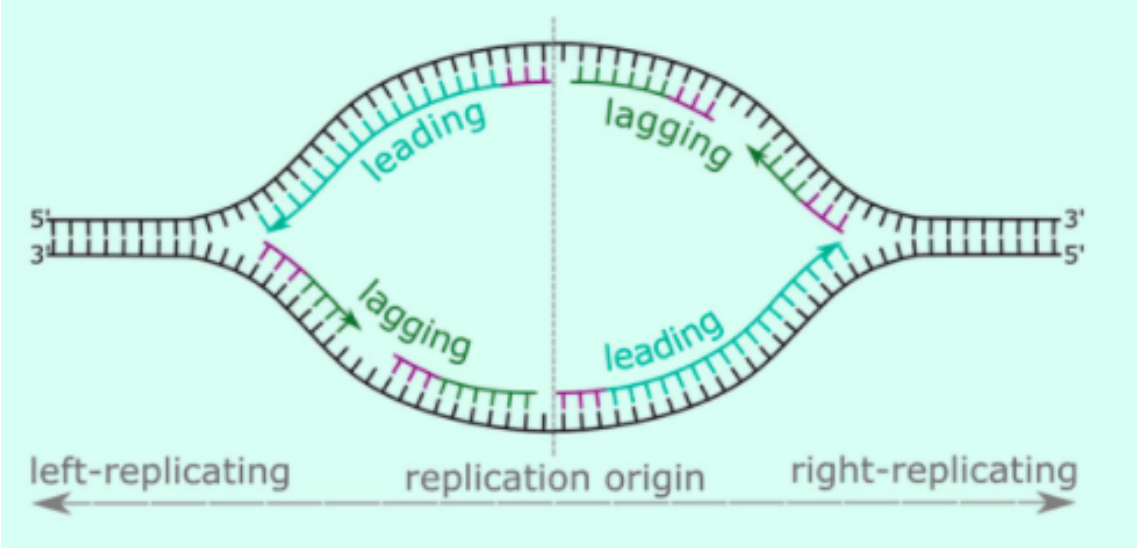

Figure 1.2 A demonstration of asymmetric synthesis of strands around replication origins (Tomkova et al., 2018).

1.4 Mapping damage formation and nucleotide excision repair Events using damage sequencing (Damage-seq) and excision repair sequencing (XR-seq) methods, respectively

Mapping of UV-induced damages and their repair is essential to understand the role of nucleotide excision repair on mutagenesis. Since the birth of the field of DNA repair, which began with the discovery of photolyase in 1958 (Rupert, Goodgal & Herriott, 1958; Sancar, 2016), many methods were introduced to map DNA damage and repair (Li & Sancar, 2020). However, not until the emergence of next-generation sequencing techniques, genome-wide mapping of DNA damage and repair at single-nucleotide resolution could be performed. Today, there are several methods that can perform this task. Among these methods, Damage sequencing (Damage-seq) and eXcision Repair sequencing (XR-seq) can map UV-induced DNA damages and repair of these damages by nucleotide excision repair, respectively, which will be explained in the subsections below.

1.4.1 Damage sequencing (Damage-seq)

Damage-seq mechanism can sensitively detect a variety of DNA lesions such as CPDs, (6-4)PPs, and cisplatin-DNA adducts, mainly using the RNAPII stalling to its advantage (Hu, Lieb, Sancar & Adar, 2016). In fact, the method can be adapted to any DNA damage that stalls RNAPII, where the damage-specific antibody is present (Sancar, 2016). After the induction of the damage, the genomic DNA is sonicated, ligated to first primers, and denaturated. Then, damage sites are immunoprecipitated by damage-specific antibodies and enriched. Following the enrichment, a biotinylated primer is annealed and extended by a polymerase called Q5 DNA polymerase, which extends the primer until it reachs the damage without synthesizing the site of the damage. Next, a second adopter is ligated to the extended primer for amplification by PCR. Lastly, the amplified oligomers can be sequenced and analyzed (Figure 1.3a).

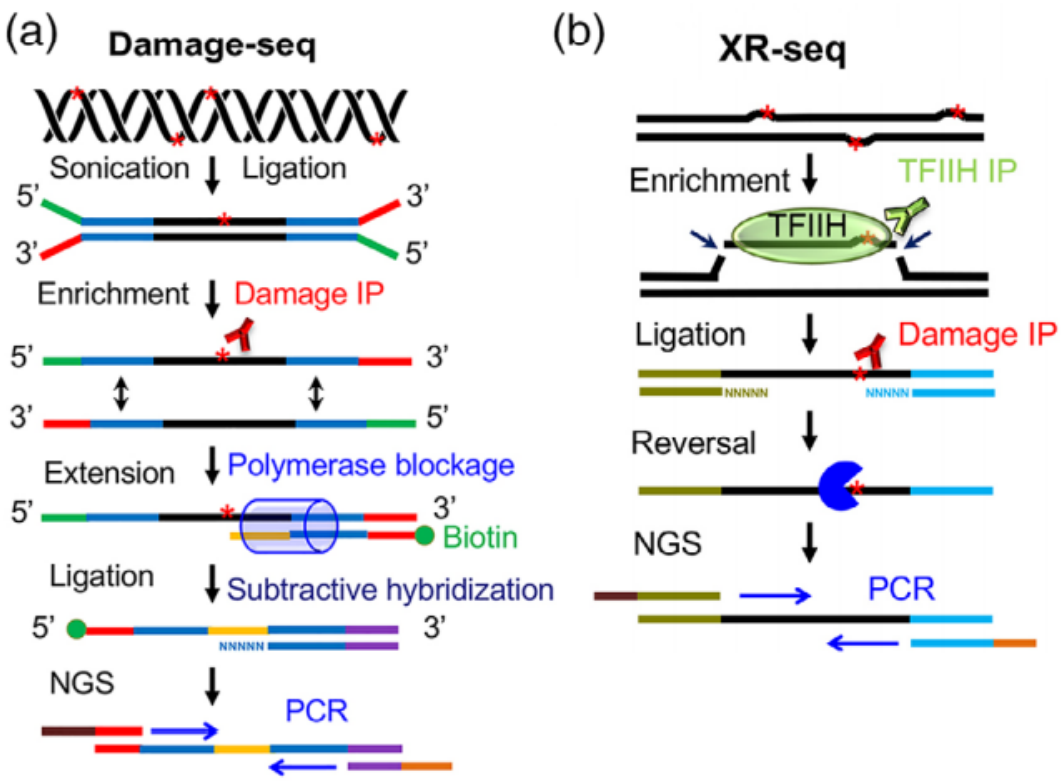

Figure 1.3 Schematic representation of (a) Damage-seq and (b) XR-seq (Li & Sancar, 2020).

1.4.2 Excision repair sequencing (XR-seq)

XR-seq method can measure the repair of DNA damages that is coordinated by nucleotide excision repair, using the 22-30 nucleotides long exiced oligomers that

are produced after the dual incision of lesion site (Hu, Li, Adebali, Yang, Oztas, Selby & Sancar, 2019; Hu et al., 2016). Excised oligomers are immunoprecipitated by TFIIH and ligated by adaptors from both sides. Next, the oligomers are filtered according to the damage of interest by immunoprecipitating with damage-specific antibodies. Then, using photolyases, lesions of the left oligomers are reversed for a proper PCR amplification process and the oligomers are sequenced (Figure 1.3b).

2. THE SCOPE OF THE THESIS

Nucleotide excision repair is the sole mechanism for the removal of bulky adducts. In this study, to assess the influence of replication along with nucleotide excision repair on mutation distribution across replicated sites, we analyzed the Damage-seq and XR-seq data in replicating cells within the context of replication timing. Damage and repair maps are generated for CPDs and (6-4)PPs from UV-irradiated HeLa cells synchronized at two stages of the cell cycle: early S phase, and late S phase. Damage-seq locates and quantifies the regions of UV induced CPD and (6-4)PP damages, while XR-seq captures excised oligomers of the damage site that are removed by the nucleotide excision repair. In combination, two methods provide the genome-wide distribution of UV-induced damages and the differential repair frequency of these damaged sites.

Initially, we examined the quality of the reads that are produced by Damage-seq and XR-seq methods. After quality filtering and performing pre-analysis of damage and repair reads, we located the positions of damage and repair events throughout the human genome. Then, we used these positions together with datasets obtained from public sources to compare the repair rate of nucleotide excision repair in different regions.

In the first part of the study, we mapped the damage and repair events to the replication domains, where closely packed origins of replications fire in a synchronized manner, resulting in simultaneous replication of these Mb-sized regions. Then, we normalized repair events with corresponding damage quantities to eliminate the potential bias caused by the damage formation. By doing so, we managed to observe the differential repair rate between replication domains at different time points on a genome scale. We performed a similar analysis using chromatin states of HeLa cells and examined how chromatin states effect the repair rate of replication domains, while moving from early to late S phase of cell cycle.

Secondly, we aimed to understand whether nucleotide exicision repair contribute to a replicative strand asymmetry. Because nucleotide excision repair is highly associated

with melanoma cancers, replicative strand asymmetry of nucleotide excision repair can correlate with the mutation profiles of this tumor type. We retrieved a somatic melanoma mutation dataset, and quantified the mutations on approximately 20 kb-sized initiation zones where origin of replications are closely positioned. We further separated these initiation zones into their corresponding replication domains before quantifying the mutations. This method enabled us both to compare the mutation count differences of replication domains, and to observe the mutational strand asymmetry on initiation zones. Next, we examined the strand asymmetry of damage and repair events separately on initiation zones. To assess if nucleotide composition of initiation zones contribute to the strand asymmetry, we simulated Damage-seq and XR-seq reads, and compared the signal levels of these reads on initiation zones as well. Lastly, we calculated the repair rate by normalizing repair events with damage quantities to evaluate the asymmetry of relative repair, which we termed repair rate.

3. MATERIALS & METHODS

3.1 Materials

Programming Languages and Tools	Description	Purpose of use	Source
Bash	a shell compatible command language	constructing pipeline, running tools, format conversions	(Ramey, 1998)
Python	a high-level, general purpose programming language	RPKM calculation, aggregating windows of regions	(Rossum, 1995)
R	a language and an environment for graphics and statistics	plotting graphs, correlation analysis	(Ihaka & Gentleman, 1996)
Cutadapt	detects and cuts adaptor sequences	removing adaptors, discarding reads containing adaptors	(Martin, 2011)
Bowtie2	a fast and memory-efficient sequence aligner	aligning reads to the reference genome	(Langmead & Salzberg, 2012)
Samtools	a suit that contains utilities to interact with and manipulate high-throughput sequencing data	sorting, filtering low quality reads	(Li, Handsaker, Wysoker, Fennell, Ruan, Homer, Marth, Abecasis & Durbin, 2009)
Bedtools	a set of utilities to perform genomic analysis	combining paired reads, converting bed files to fasta format, calculating genome coverage, intersecting regions to each other	(Quinlan & Hall, 2010)
BedGraphToBigWig	converts bedGraph files to bigWig	creating bigWig files for visualization	(Kent, Zweig, Barber, Hinrichs & Karolchik, 2010)
Art	a simulation tool that creates a synthetic high-throughput sequencing data	simulating XR-seq and Damage-seq reads	(Huang, Li, Myers & Marth, 2012)

Table 3.1 Programming languages and tools that are used at the study.

Databases	Data Obtained	Source
The European Bioinformatics Institute FTP Server	Genome Reference Consortium Human Build 37 (GRCh37)	(Church, Schneider, Graves, Auger, Cunningham, Bouk, Chen, Agarwala, McLaren, Ritchie & others, 2011)
Gene Expression Omnibus (GEO)	processed Repli-Seq data of HeLa-S3 (accession no: GSE53984), SNS-Seq data of HeLa-S3 (accession no: GSE37757)	(Besnard, Babled, Lapasset, Milhavet, Parrinello, Dantec, Marin & Lemaitre, 2012; Liu et al., 2016)
UCSC Genome Browser	ChromHMM segmentation from HeLa-S3 ChIP-Seq data	(Ernst & Kellis, 2017)
Sequence Read Archive (SRA)	OK-Seq data (accession no: SRP065949)	(Petryk, Kahli, d'Aubenton Carafa, Jaszczyszyn, Shen, Silvain, Thermes, Chen & Hyrien, 2016)
International Cancer Genome Consortium (ICGC)	Simple somatic mutations of Melanoma	(Hayward, Wilmott, Waddell, Johansson, Field, Nones, Patch, Kakavand, Alexandrov, Burke & others, 2017)

Table 3.2 Retrieved datasets and their databases.

cell line	product	method	release	time	replicate
HeLa-S3	CPD	XR-seq	early	120	A
HeLa-S3	CPD	XR-seq	late	120	A
HeLa-S3	CPD	XR-seq	early	120	B
HeLa-S3	CPD	XR-seq	late	120	B
HeLa-S3	CPD	Damage-seq	early	120	A
HeLa-S3	CPD	Damage-seq	late	120	A
HeLa-S3	CPD	Damage-seq	early	120	B
HeLa-S3	CPD	Damage-seq	late	120	B
HeLa-S3	(6-4)PP	XR-seq	async	12	A
HeLa-S3	(6-4)PP	XR-seq	async	12	B
HeLa-S3	(6-4)PP	XR-seq	early	12	A
HeLa-S3	(6-4)PP	XR-seq	early	12	B
HeLa-S3	(6-4)PP	XR-seq	late	12	A
HeLa-S3	(6-4)PP	XR-seq	late	12	B
HeLa-S3	CPD	XR-seq	async	12	A
HeLa-S3	CPD	XR-seq	async	12	B
HeLa-S3	CPD	XR-seq	early	12	A
HeLa-S3	CPD	XR-seq	early	12	B
HeLa-S3	CPD	XR-seq	late	12	A
HeLa-S3	CPD	XR-seq	late	12	B
HeLa-S3	(6-4)PP	Damage-seq	async	12	A
HeLa-S3	(6-4)PP	Damage-seq	async	12	B
HeLa-S3	(6-4)PP	Damage-seq	early	12	A
HeLa-S3	(6-4)PP	Damage-seq	early	12	B
HeLa-S3	(6-4)PP	Damage-seq	late	12	A
HeLa-S3	(6-4)PP	Damage-seq	late	12	B
HeLa-S3	CPD	Damage-seq	async	12	A
HeLa-S3	CPD	Damage-seq	async	12	B
HeLa-S3	CPD	Damage-seq	early	12	A
HeLa-S3	CPD	Damage-seq	early	12	B
HeLa-S3	CPD	Damage-seq	late	12	A
HeLa-S3	CPD	Damage-seq	late	12	B

Table 3.3 Information of samples that are produced for this study.

3.2 Methods

The experiments were performed at the laboratories of Aziz Sancar (University of North Carolina at Chapel Hill) and Jinchuan Hu (Fudan University), whereas analyses of the data were carried by us.

3.2.1 Cell culture and treatments

HeLa-S3 cell lines that were purchased from ATCC were cultured in DMEM medium supplemented with 10% FBS and 1% penicillin/streptomycin at 37°C in a 5% atmosphere CO_2 humidified chamber. By double-thymidine treatment, cells were synchronized at late G1 phase, and released into S phase after the removal of thymidine. Thymidine at 50% confluence was added to the cells to a final concentration of 2 mM for the initial thymidine treatment. After 18 hours, the cells were washed withPBS for their release 18 hours after the initial thymidine treatment, and cultured in fresh medium for 9 hours. Then for 15 hours, cells were treated with 2mM thymidine and released into S phase for designated time before UV irradiation. Cells were irradiated with $20J/m^2$ of UVC, then collected either immediately or after incubation at 37°C for designated time for the following assays.

3.2.2 Flow cytometry analysis

HeLa-S3 cell lines were initially trypsinized, and thenPBS washed. After washing, for 2 hours, cells were fixed in 70% (v/v) ethanol at -20°C, then for 30 minutes, stained in the staining solution at room temperature. Lastly, the progression of the cells throughout the S phase was analyzed by a flow cytometer.

3.2.3 Damage-seq and XR-seq libraries preparation and sequencing

After HeLa-S3 cell lines were harvested in ice-coldPBS at designated time, Damage-seq and XR-seq methods were applied. For Damage-seq, using PureLink Genomic DNA Mini Kit, genomic DNA was taken out and then, cut into fragments by sonication using Q800 Sonicator. After sonication, DNA fragments ($1\mu g$) were subjected to end repair, dA-tailing and ligation using the first adaptor. Then, the fragments were denaturated and immunoprecipitated with either anti-(6-4)PP or anti-CPD antibody. A primer called Bio3U was bound to the fragment and extended with Q5 DNA polymerase until the primer reaches the lesion site. Next, the extended primer fragments were purified and annealed to oligo SH for subtractive hybridization process. After the substractive hybridization, oligo SH was removed using streptavidin

C1 and the fragments were ligated to the second adapter for PCR amplification process. For XR-seq, cells were lysed with a homogenizer and centrifuged to remove chromatin DNA. To extract the nucleotide excision repair products, lysed cells were immunoprecipitated with anti-XPG antibody, which precipitates the excision products. Then, purified fragments were ligated with adaptors from both ends. The fragments were further immunoprecipitation with either anti-(6-4)PP or anti-CPD antibody and lesion sites were repaired by photolyase. After PCR amplification and gel purification, the products were sequenced via Hiseq 2000/2500 platform by the University of North Carolina High-Throughput Sequencing Facility, or Hiseq X platform by the WuXiNextCODE Company.

3.2.4 Damage-seq sequence pre-analysis

The sequenced reads with adapter sequence GACTGGTTCCAATTGAAAGTGCTCTTCCGATCT at 5' end, were discarded via cutadapt with default parameters for both single-end and paired-end reads (Martin, 2011). The remaining reads were aligned to the hg19 human genome using bowtie2 with 4 threads (`-p`) (Langmead & Salzberg, 2012). For paired-end reads, maximum fragment length (`-X`), which means the maximum accepted total length of mated reads and the gap between them, was chosen as 1000. Using samtools, aligned paired-end reads were converted to bam format, sorted using `samtools sort -n` command, and properly mapped reads with a mapping quality greater than 20 were filtered using the command `samtools view -q 20 -bf 0x2` in the respective order (Li et al., 2009). Then, resulting bam files were converted into bed format using `bedtools bamtobed -bedpe -mate1` command (Quinlan & Hall, 2010). The aligned single-end reads were directly converted into bam format after the removal of low quality reads (mapping quality smaller than 20) and further converted into bed format with `bedtools bamtobed` command (Quinlan & Hall, 2010). Because the exact damage sites should be positioned at two nucleotides upstream of the reads (Li et al., 2009), bedtools flank and slop command were used to obtain 10 nucleotide long positions bearing damage sites at the center (5. and 6. positions) (Quinlan & Hall, 2010). The reads that have the same starting and ending positions, were reduced to a single read for deduplication and remaining reads were sorted with the command `sort -u -k1,1 -k2,2n -k3,3n`. Then, reads that did not contain dipyrimidines (TT, TC, CT, CC) at their damage site (5. and 6. positions) were filtered out to eliminate all the reads that do not harbor a UV damage. Lastly, only the reads that were aligned to common chromosomes (chromosome 1-22 + X) were held for further analysis.

3.2.5 XR-seq sequence pre-analysis

TGGAATTCTCGGGTGCCAAGGAACTCCAGTNNNNNNACGATCTCGTATGCCGTCTTCTGCTTG adaptor sequence at the 3' of the reads were trimmed and sequences without the adaptor sequences were discarded using cutadapt with default parameters (Martin, 2011). Bowtie2 was used with 4 threads (`-p`) to align the reads to the hg19 human genome (Langmead & Salzberg, 2012). Then reads with mapping quality smaller than 20 were removed by samtools (Li et al., 2009). Bam files obtained from samtools were converted into bed format by bedtools (Quinlan & Hall, 2010). Multiple reads that were aligned to the same position, were reduced to a single read to prevent duplication effect and remaining reads were sorted with the command `sort -u -k1,1 -k2,2n -k3,3n`. Lastly, only the reads that were aligned to common chromosomes were held for further analysis.

3.2.6 Dna-seq sequence pre-analysis

Paired-end reads were aligned to hg19 human genome via bowtie2 with 4 threads (`-p`) and maximum fragment length (`-X`) chosen as 1000 (Langmead & Salzberg, 2012). Sam files were converted into bed format as it was performed at Damage-seq paired-end reads. Duplicates were removed and reads were sorted with `sort -u -k1,1 -k2,2n -k3,3n` command. Lastly, the reads that did not align to the common chromosomes were discarded.

3.2.7 XR-seq and Damage-seq simulation

Art simulator was used to produce synthetic reads with the parameters `-l 26 -f 2, -l 10 -f 2` for XR-seq and Damage-seq, respectively (Huang et al., 2012). To better represent our filtered real reads, read length (`-l`) parameter was chosen as the most frequent read length after pre-analysis done. The fastq file that Art produced, were filtered according to our reads by calculating a score using nucleotide frequency of the real reads and obtaining most similar 10 million simulated reads. The filtering was done by `filter_syn_fasta.go` script, which is available at the repository: https://github.com/compGenomeLab/lemurRepair. Filtered files were preceded by pre-analysis again for further analysis.

3.2.8 Quantification of melanoma mutations

Melanoma somatic mutations of 183 tumor samples were obtained from the data portal of International Cancer Genome Consortium (ICGC) as compressed tsv files which is publicly available at

https://dcc.icgc.org/releases/release_28/Projects/MELA-AU. Single base substitution mutations were extracted, and only the mutations of common chromosomes were used. To obtain the mutations that could be caused by UV-induced photoproducts, C -> T mutations that have a pyrimidine at the upstream nucleotide was further extracted. Later on, mutations were quantified on 20 kb long initiation zones that were separated into their corresponding replication domains using `bedtools intersect` command with the `-wa -c -F 0.5` options.

3.2.9 Further analysis

In order to separate a region data (replication domains, initiation zones, or replication origins) into chosen number of (201) bins, the start and end positions of all the regions set to a desired range with the unix command: `awk -v a="$intervalLen" -v b="$windowNum" -v c="$name" '{print $1"\t"int(($2+$3)/2-a/2-a*(b-1)/2)"\t"int(($2+$3)/2+a/2+a*(b-1)/2) "\t"$4"\t"."\t"$6}'`. Then, any intersecting regions or regions crossing the borders of its chromosomes were filtered to eliminate the possibility of signal's canceling out effect. After that, `bedtools makewindows` command was used with the `-n 201 -i srcwinnum` options to create a bed file containing the bins. To quantify the XR-seq and Damage-seq profiles on the prepared bed file, `bedtools intersect` command was used to intersect as it was performed for mutation data. Then all bins were aggregated given their bin numbers, and the mean of the total value of each bin were calculated. Lastly RPKM normalization was performed and the plots were produced using ggplot2 in R programming language.

4. RESULTS

4.1 Genome-wide mapping of UV-induced damages and their repair synchronized at two stages of the cell cycle: early S phase, and late S phase

This study presents a set of experiments yielding NGS datasets, followed by bioinformatic analyses of genomic data, where we purified and sequenced fragments of UV-induced damages and their repair in HeLa cells that are synchronized either at early and late S phases. After synchronizing cells using double-thymidine treatment, we further treated cells with $20J/m^2$ UVB exposure. Immediately after the exposure, we adopted Damage-seq to quantify occurred damages by the exposure, before nucleotide excision repair initiates. To quantify repair, we adopted XR-seq and quantified CPD repair at 12 minutes and 2 hours; while (6-4)PP repair were quantified only at 12 minutes (Figure 4.1). We performed each experiment twice to obtain two biological replicates for each sample.

Quality control analyses were performed on early S phased (6-4)PPs at 12 minutes (Figure 4.1B-D) and other samples (Figure 6.2-6.9). The data indicate high qualities and consistent results between replicates. In agreement with the dual incision mechanism of nucleotide excision repair (Huang, Svoboda, Reardon & Sancar, 1992; Li, Hu, Adebali, Adar, Yang, Chiou & Sancar, 2017; Reardon & Sancar, 2005), XR-seq oligomers are in the size range of 20-30 nucleotides, with a median of 26 nucleotides. Moreover, dipyrimidine content of 26 nucleotides long oligomers enriches at position 19-20 (Figure 4.1B), where the DNA lesion occurs (Huang et al., 1992). Also, (6-4)PP samples exhibit high levels of TC dipyrimidine repair (Figure 4.1B, 6.2-6.4), whereas CPD samples exhibit elevated TT dipyrimidine repair (Figure 6.5-6.9), which are the most abundant sites for formation of these photoproducts (Mouret et al., 2010). Because this study focuses on the GR, contribution of TCR can create a bias. Importantly, the repair levels at transcribed and non-transcribed strand are equivalent for samples at 12 minutes (Figure 4.1C, 6.2-6.7), suggesting no contribution of TCR. On the other hand, CPD samples at 2 hours indicate slight

increase towards transcribed strands, which might be a bias caused by TCR (Figure 6.8, 6.9). Correlation plots between the biological replicates indicates reasonable reproducibility, having correlation coefficients 0.86 and above (Figure 4.1D, 6.2-6.9).

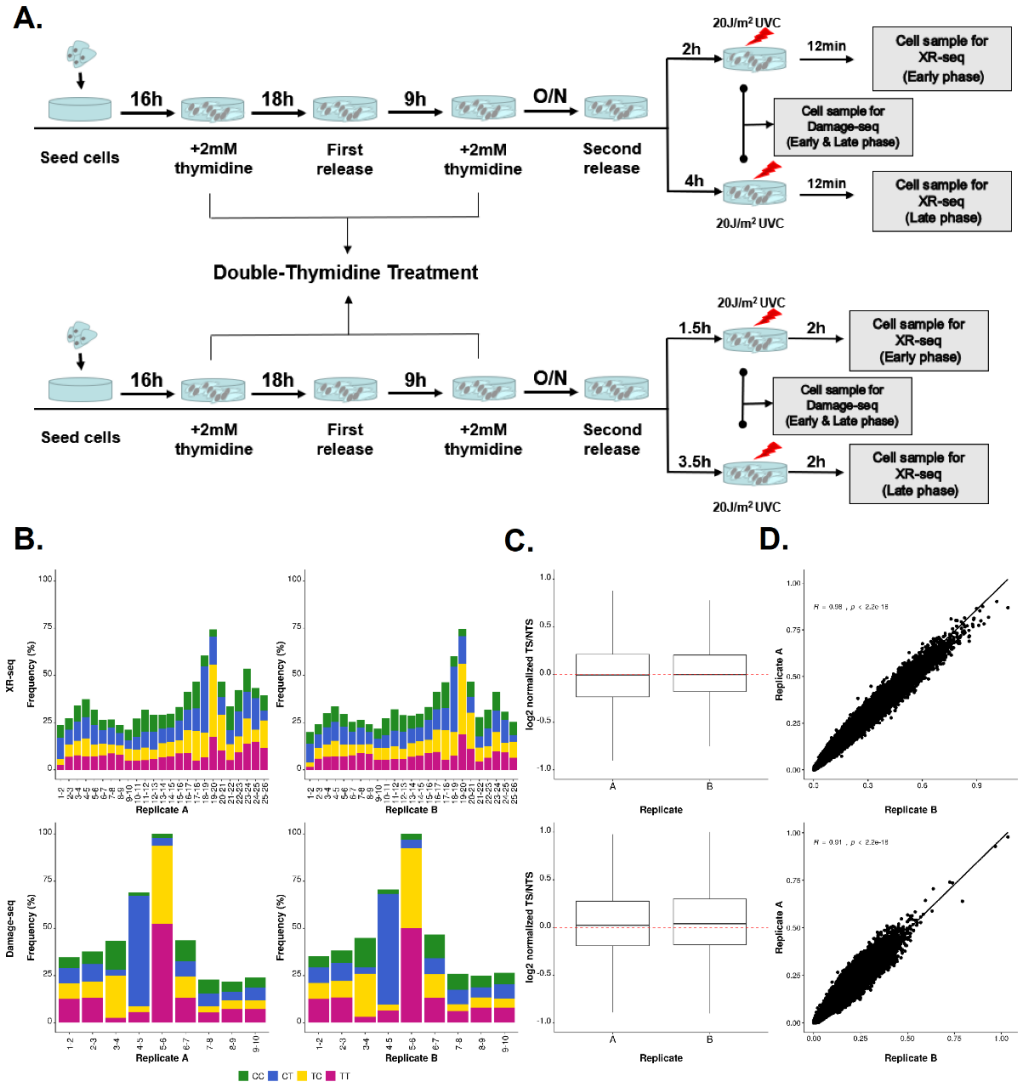

Figure 4.1 A) Experimental setup. B-D) Control figures of (6-4)PP early phased samples at 12 minutes. B) The dinucleotide composition frequency of replicate A and B, respectively. C) log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. D) The correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

4.2 Early replication domains are repaired more efficiently than late replication domains, however, the repair rate of late replication domains elevates while replication proceeds.

To determine how excision repair rates are influenced by replication domains during replication, we compared repair efficiency of early replication domains (ERDs) and late replication domains (LRDs). We obtained replication domains of HeLa cells from a study where a supervised method called Deep Neural Network-Hidden Markov Model was developed to define replication domains from Repli-seq data (Liu et al., 2016). We mapped damage and repair events to corresponding replication domains. To eliminate the effect of a potential bias in damage formation, we normalized repair quantities (XR-seq) by the captured damage events (Damage-seq) in each genomic window (Figure 4.2). This approach enabled us to assess the efficiency of repair per damage at a given region, which we refer to as repair rate. Based on an analysis with a Hi-C dataset, the human genome was classified into A/B compartments, which are associated with open and closed chromatin regions, respectively (Lieberman-Aiden, Van Berkum, Williams, Imakaev, Ragoczy, Telling, Amit, Lajoie, Sabo, Dorschner & others, 2009). Recently, it was also shown that ERDs and LRDs strongly correlate with A/B compartments respectively (Pope, Ryba, Dileep, Yue, Wu, Denas, Vera, Wang, Hansen, Canfield & others, 2014; Ryba, Hiratani, Lu, Itoh, Kulik, Zhang, Schulz, Robins, Dalton & Gilbert, 2010). Because ERDs are correlated to open chromatins, these regions are more reachable for excision repair machinery than LRDs. Expectedly, repair rates are elevated at the center of ERDs and gradually reduced towards flanking sites, while LRDs exhibit an opposite pattern (Figure 4.2A, 6.18-6.23). These results suggest that ERDs and their flanking regions are efficiently repaired, whereas less reachable LRDs are poorly repaired. Moreover, LRDs are known to contain higher mutation frequency than other regions (Lawrence et al., 2013; Stamatoyannopoulos et al., 2009), hence; low repair rate of UV damages located at LRDs might be a key factor of mutagenesis in cancer associated with NER such as melanoma.

On the other hand, the difference between early and late S phases indicates that repair rate is elevated in favor of LRDs when replication timing moves from early to late S phase (Figure 4.2A-B, 6.24-6.29). This time dependent increase in the repair rate of LRDs is likely to be caused by the unfolding of heterochromatin during replication. With the unfolding of the chromatin, more LRD regions will be accessible where the DNA lesions can be efficiently recognized and removed by nucleotide excision repair. Also, we observe a reduction of repair rate in ERDs,

however this reduction might be caused by the relativity of the XR-seq method; increased repair rate in LRDs results in relative decrease in the repair rate in ERDs, even if repair rate does not quantitatively change in ERDs. In addition, (6-4)PP repair at 12 minutes exhibits minor differences between early and late S phases (4.2), potentially because of its fast repair after the damage occurrence (Hu, Adebali, Adar & Sancar, 2017). Conversely, CPD repair rate at 12 minutes and 2 hours demonstrate significant increase for LRDs and decrease for ERDs (4.2B, p-values < 2.2e-16).

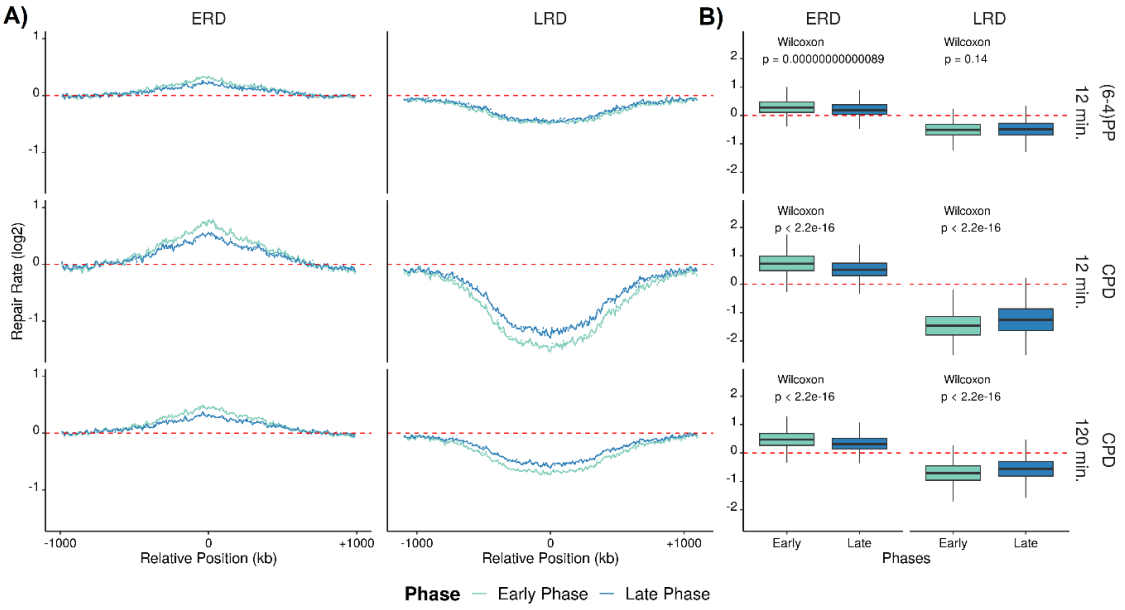

Figure 4.2 The shift of repair efficiency at replication domains during replication timing. A) Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 2 Mb regions with 10 kb intervals, which early replication domains (ERDs, left) and late replication domains (LRDs, right) positioned at the center of the region. B) RPKM values of XR-seq samples are divided by Damage-seq samples (Repair Rate) for both ERDs (left) and LRDs (right) and log2 transformed. Wilcoxon test is used to assess the significance of difference between early and late S phases. The light blue lines are the early phase repair rate values and dark blue lines are the late phase repair rate values. Above the red horizontal dashed line demonstrates that repair is higher than damage, below demonstrates that damage is higher. Analysis is performed on replicate A.

4.3 Variety of chromatin states are associated with differential repair efficiency.

Active chromatin states are repaired effectively; basically because those regions are more accessible to nucleotide excision repair (Adar, Hu, Lieb & Sancar, 2016). We addressed how repair efficiency in ERDs, and LRDs is differentially influenced by

the chromatin states during replication. We retrieved chromatin states of HeLa cells segmented by ChromHMM from UCSC website (Ernst & Kellis, 2017). We intersected the chromatin states with replication domains and mapped damage and repair reads to those regions, for each chromosome. After calculating the repair rates (Figure 4.3A, 6.13A-6.17A), we further assessed early S phase repair relative to late S phase (*early/late repair/damage*) to observe the replication timing differences in efficiency in the function of chromatin states (Figure 4.3B, 6.13B-6.17B). Generally, repair efficiency is higher in the active chromatin states such as promoters and strong enhancers, which is in agreement with the previous studies (Adar et al., 2016; Hu et al., 2016). Those regions sustain high repair rates, even in LRDs during the early S phase, that should be condensed and harder to reach (Figure 4.3A). On the other hand, all the transcription-associated chromatin states together with “FaireW” and “Low” chromatin states are highly affected by the replication timing, generally increasing in ERDs and LRDs in early and late S phases, respectively (Figure 4.3B). “FaireW” represents the regions that are associated to the regulatory activities (Giresi, Kim, McDaniell, Iyer & Lieb, 2007), whereas “Low” stands for low activity regions that neighboring active sites. In ERDs, although both chromatin states have relatively low repair in early and late S phases, they demonstrate a drastic increase when replication proceeds from early to late S phase. However, in LRDs, some transcription-associated chromatin states exhibit a high variance across chromosomes, thus expending the interquartile range of boxplots (Figure 4.3B). Therefore, the effect of replication timing on transcription-associated chromatin states such as "Gen5’", "Gen3’", and "Pol2" is not prominent (Figure 4.3B).

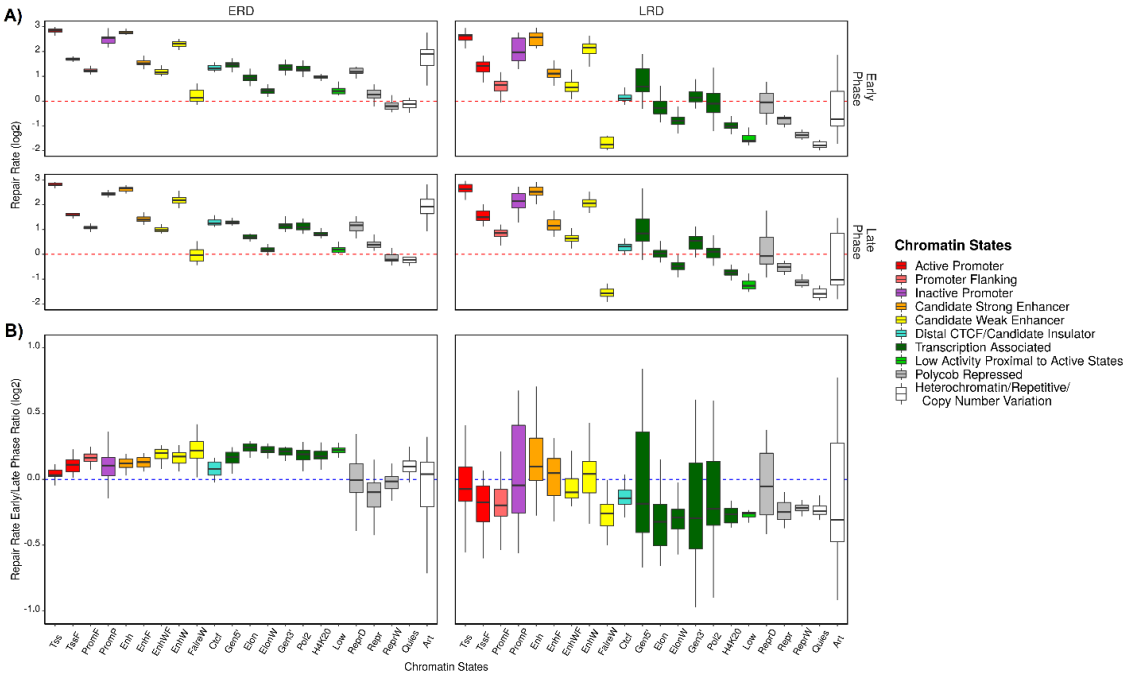

Figure 4.3 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of CPD samples at 12 minutes are calculated, log2 transformed, B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). Analysis is performed on replicate A.

4.4 Origins of replications display distinct melanoma mutation counts and strand asymmetry based on their replication domains.

Replication domains are 1 to 2 Mb-sized DNA chunks that involves many small replication origins. The genome-wide effect of replication timing on nucleotide excision repair can be demonstrated by the differential repair rate in replication domains, while replication proceeds. However, the association of replication origins and nucleotide excision repair cannot be explained using Mb-sized regions. Therefore, we retrieved two independent datasets that are derived from two different methods: okazaki fragment sequencing (OK-seq) and short nascent strand sequencing (SNS-seq). OK-seq quantifies the replication initiation zones that are the sets of closely positioned replication origins using highly purified Okazaki fragments (Petryk et al., 2016), whereas SNS-seq can precisely identifies individual replication origins (Besnard et al., 2012; Langley, Gräf, Smith & Krude, 2016). Using these datasets together with a melanoma mutation dataset that we retrieved from the International

Cancer Genome Consortium (ICGC) data portal (Hayward et al., 2017), we examined the mutation profiles at the sites of replication origins where the replication initiates. Because nucleotide excision repair is highly associated with melanoma cancers, we argued that this relation must be reflected to the mutation counts of melanoma. For both OK-seq and SNS-seq data, we assorted genomic regions based on their corresponding replication domains to detect how mutation profiles of replication origins affected by the domains they are located. Then, we counted the C to T mutations nearby these regions that are centered around individual replication origins (SNS-seq data, Figure 4.4A) or initiation zones (OK-seq data, Figure 4.4B-C) and normalized the C to T mutations with cytosine counts in each bin, to eliminate any nucleotide content bias. Also, we gradually extended the region length of initiation zones from 20 kb to 200 kb (Figure 4.4B-C) to observe the replication effect at a range of scales.

Mutation counts of initiation zones differ depending on the replication domains they are located. In agreement with previous studies (Lawrence et al., 2013; Schuster-Böckler & Lehner, 2012; Stamatoyannopoulos et al., 2009), the mutation counts of initiation zones in LRDs elevate, while ERDs contain the initiation zones with the lowest mutation counts (Figure 4.4B-C). These differences that are related to the replication domains are also persistent for the individual replication origins (Figure 4.4A). Furthermore, initiation zones in up (UTZs) and down transition zones (DTZs), which are the domains that connect ERDs to LRDs, have mutation numbers higher and lower than ERDs and LRDs, respectively. Moreover, the flanking sites of initiation zones in transition zones that are close to ERDs have lower counts, whereas the sites that are close to LRDs have higher (Figure 4.4C, left).

Mutation counts not only exhibit a replication domain-related difference but also reveal a strand asymmetry around the initiation zones (Figure 4.4B-C). The asymmetry suggests that lagging strand (lagging strand template) (minus strand at left direction; plus strand at right direction) have more mutations than leading strand (leading strand template), independent of the replication domains. While the initiation zones in LRDs show an explicit strand asymmetry compared to the initiation zones in ERDs, the initiation zones in ERDs have a wider strand asymmetry than that of LRDs. A possible reason can be the amount of replication origins they harbor; earlier studies suggest that ERDs contain significantly higher number of replication origins (Besnard et al., 2012), and the cumulative effect of these replication origins can create a strand asymmetry that is visible on a wider region. Additionally, replication fork movement at LRDs (1.5–2.3 kb/min) is faster than it is at ERDs (1.1–1.2 kb/min) (Takebayashi, Sugimura, Saito, Sato, Fukushima, Taguchi & Okumura, 2005), which might cause more mutation and increased asymmetry in LRDs.

Conversely, individual replication origins obtained from SNS-seq data do not show an explicit strand asymmetry (Figure 4.4A).

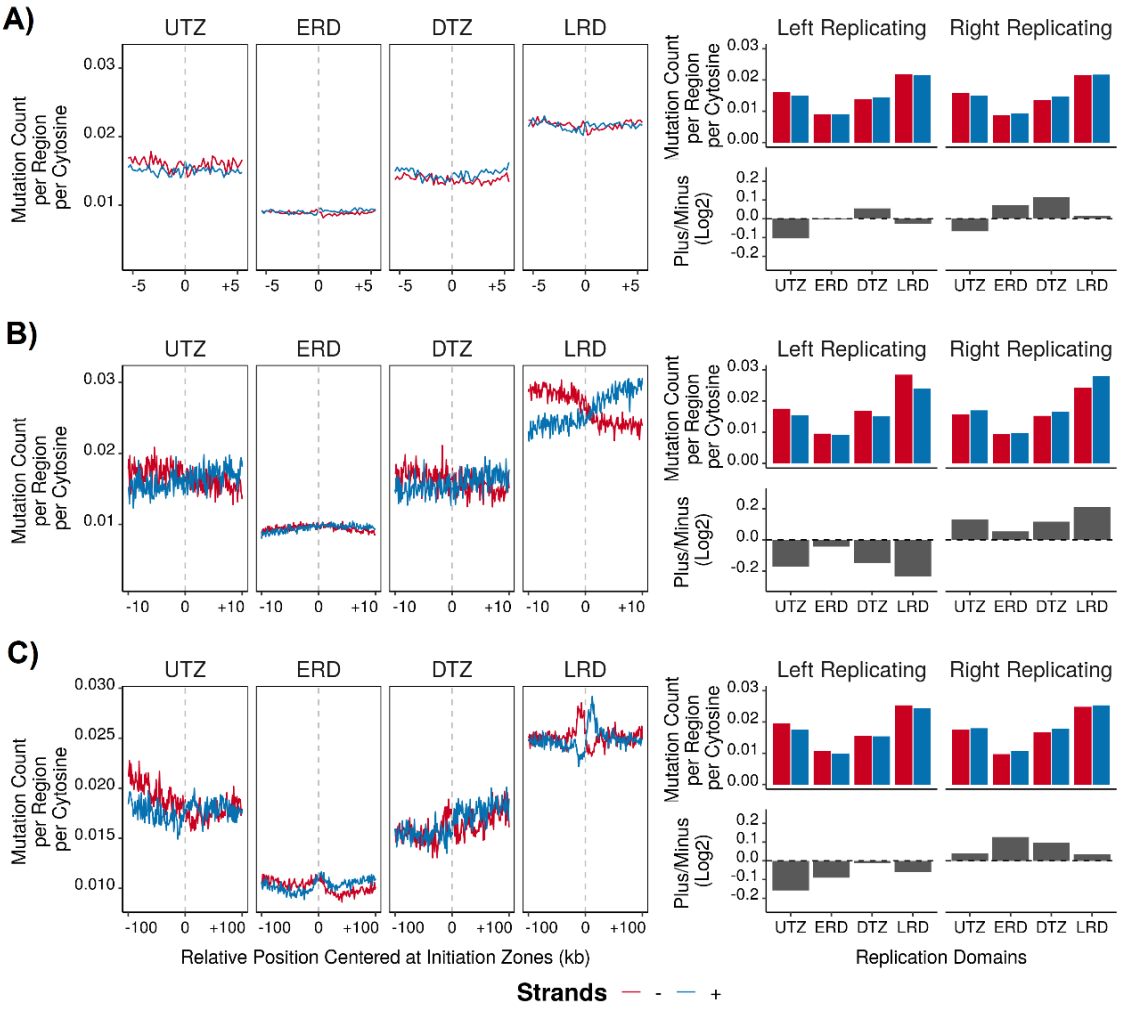

Figure 4.4 Tumor mutation profiles around replication origins and initiation zones for each replication domain. A) C to T mutations are mapped to Replication Origins (SNS-seq) and counted in 10 kb regions with 100 base pair intervals. C to T mutations are mapped to Initiation Zones (OK-seq) B) counted in 20 kb regions with 100 base pair intervals, C) and counted in 200 kb regions with 1000 base pair intervals. Counts are normalized by the number of regions and cytosine counts of each region. Red lines are the plus strands and blue lines are the minus strands. Gray vertical dashed line shows the center of the region. Upper right part demonstrates the strand differences at left (left part of the gray line) and right (right part of the gray line) replicating directions by taking the mean of the intervals, separately for the strands. Below that, strands are divided to each other (Plus/Minus) and log2 transformed to better visualize the asymmetry at each replication domain.

4.5 Asymmetric damage around initiation zones causes asymmetric repair profiles.

To reveal whether there is a strand asymmetry at repair and damage profiles similar to melanoma mutations, we mapped repair and damage events to initiation zones independently. Interestingly, a strand asymmetry around the initiation zones is visible for both repair and damage profiles (Figure 4.5). The asymmetry suggests that lagging template strand harbors more damages and attracts more repair accordingly. Reasoning that nucleotide composition of initiation zones might contribute to the strand asymmetry, we decided to simulate damage and repair signals based on their nucleotide compositions. We simulated signals via Art simulator (Huang et al., 2012), filtered the signals that resembled the observed nucleotide composition from Damage-seq and XR-seq, and mapped the filtered ones to the human genome. The simulated signals indeed display a similar strand asymmetry, indicating the contribution of nucleotide composition of the genomic regions surrounding the initiation zones (Figure 4.5). Nonetheless, simulated signals have lower RPKM values in general (Figure 4.5). Although nucleotide composition of these signals and real ones are similar, the real repair and damage events occurred at other regions. Therefore, it is expected to observe lower RPKM values for simulated signals.

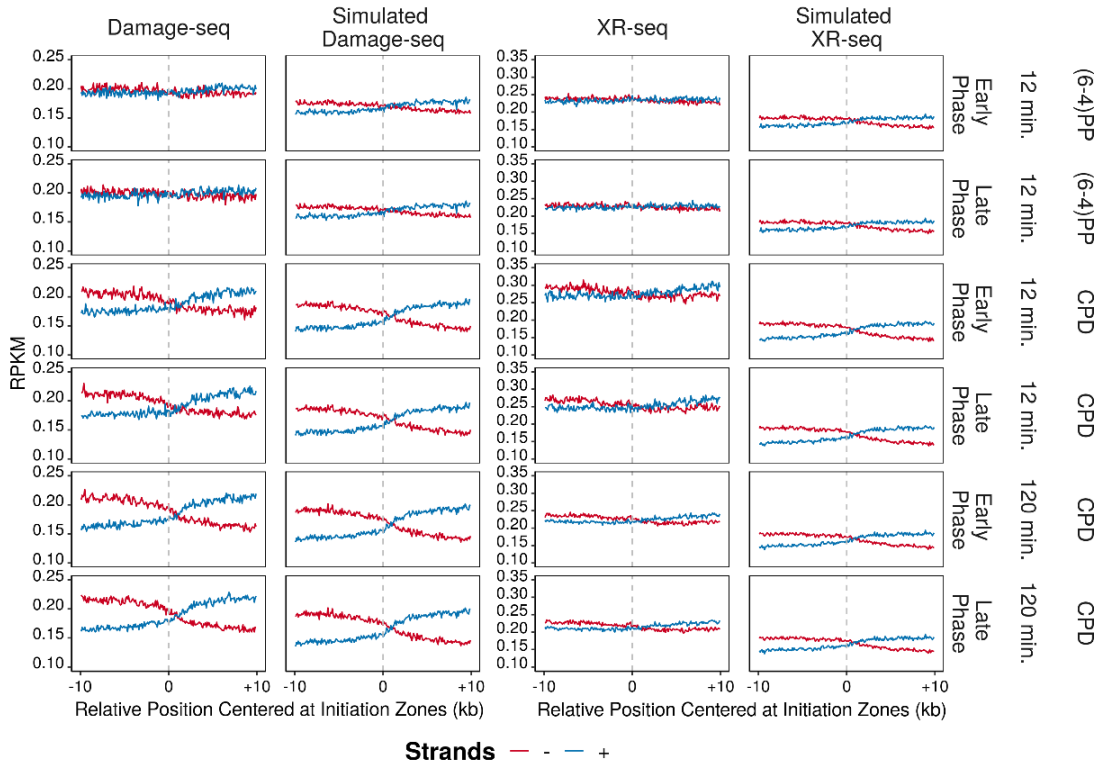

Figure 4.5 Strand asymmetry around initiation zones caused by sequence content. RPKM values of real and simulated Damage-seq samples (left) and XR-seq samples (right) are calculated in 20 kb windows with 100 base pair intervals, which Initiation Zones are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region.

4.6 Strand asymmetry of excision repair rate

After observing strand asymmetry of mutation counts of melanoma, and damage and repair events independently, we examined the repair rates of (6-4)PP and CPD samples around initiation zones for an asymmetry. Interestingly, a strand asymmetry that is inversely correlated with mutation counts of melanoma is prominent among the CPD damages (Figure 4.6). The asymmetry indicates an efficient repair of leading strands, which is in agreement with the mutation counts that displayed low mutation on leading strands. This pattern becomes more explicit at a wider scale (Figure 6.36, 6.37). In addition, CPD repair at 2 hours after damage have elevated asymmetry at a wider scale compared to the CPD samples at 12 minutes (Figure 4.6, 6.36, 6.37, 6.42-6.45), because CPDs are effectively repaired 1 hour after the damage formation. On the contrary, samples with (6-4)PP damages do not exhibit any strand difference, possibly because of the fast repair ability of nucleotide excision repair for these photo-products. Although we do not observe a distinct mutational

strand asymmetry at the individual replication origins, we analyzed the damage and repair events individually, and repair rates around replication origins. Surprisingly, we observed a minor strand asymmetry at individual replication origins (Figure 6.46-6.53, 6.58-6.61). While the samples at 2 hours demonstrates an efficient relative repair at leading templates, samples at 12 minutes display an asymmetry that favors the repair at lagging templates.

Even though replication forks often move bidirectionally, at some regions, forks tend to move continuously in one direction, either by the moving long distances as a single fork, or multiple forks that are fired simultaneously (Takebayashi et al., 2017). To observe the effect of replication fork movement, we decided to use the regions that replication forks move in one direction. We retrieved a data which is produced by OK-seq and contains regions that are dominantly replicated in a single direction, termed high replication fork directionality (RFDs) (Petryk et al., 2016). Repair rates at high RFDs display a decrease at the direction of replication fork on wider regions (Figure 6.62-6.67). This decrease might be caused by the replication fork itself; considering that the regions replication fork had passed remain as open chromatin and becomes reachable to nucleotide excision repair, while the downstream will be relatively condensed. Also, CPDs at 2 hours demonstrate a visible strand asymmetry at both directions in favor of leading template strand (Figure 6.62-6.67).

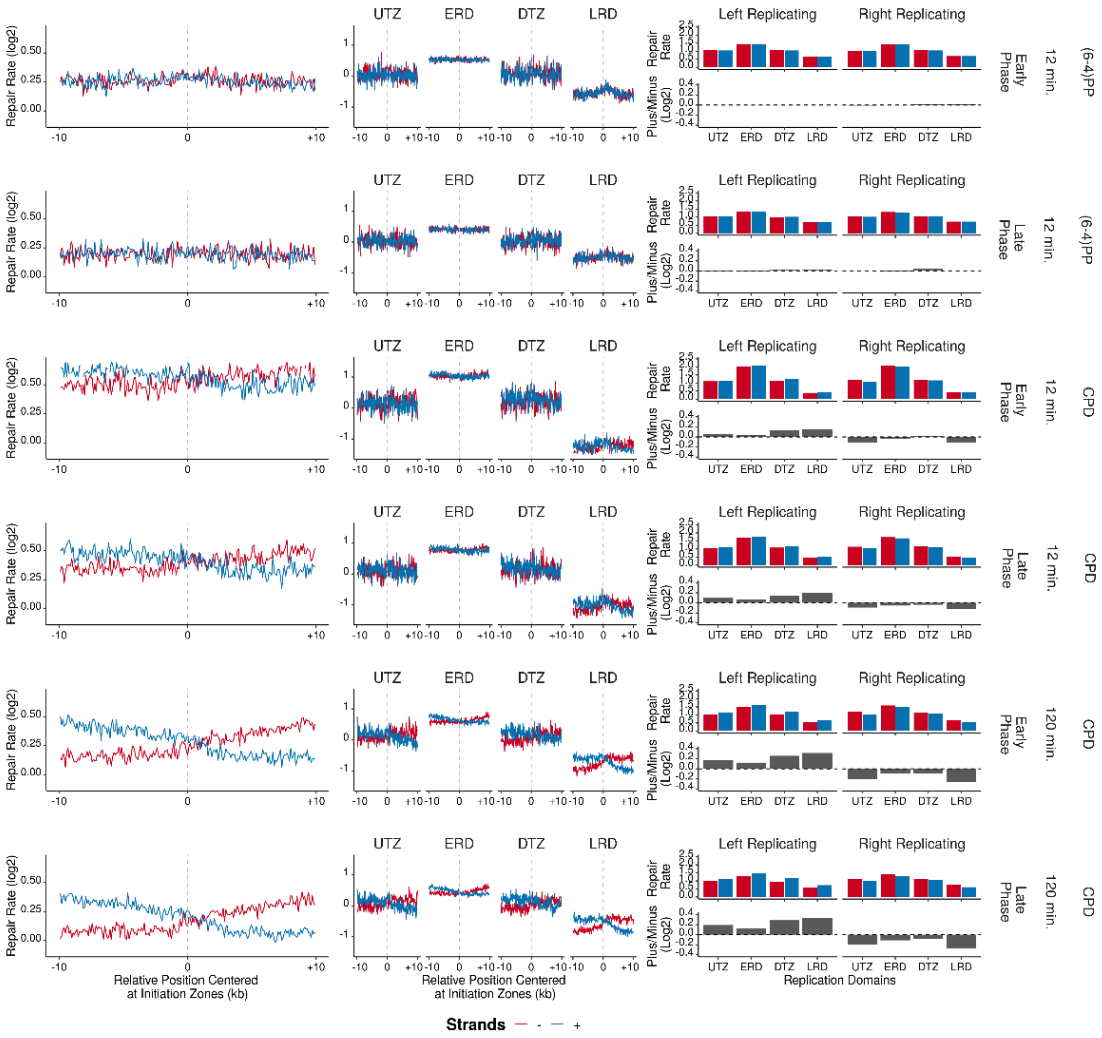

Figure 4.6 Repair rate asymmetry around initiation zones and replication domains. (Left) Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which Initiation Zones are positioned at the center of the region. (Middle) Same analysis performed, however initiation zones separated into their corresponding replication domains. (Right) The strand differences at left (left part of the gray line) and right (right part of the gray line) replicating directions are shown by taking the mean of the intervals, separately for the strands. Below that, strands are divided to each other (Plus/Minus) and log2 transformed to better visualize the asymmetry at each replication domain. Blue lines are the plus strands and red lines are the minus strands. Gray vertical dashed line shows the center of the region.

5. DISCUSSION

DNA replication is a highly conserved and regulated temporal process that is essential to genome inheritance. Yet, stochastic effects of DNA replication might cause mutagenesis contributing to cancer (Tomasetti & Vogelstein, 2015). Therefore, an accurate and properly coordinated DNA replication is needed to prevent errors and to preserve DNA fidelity which is constantly threatened by both endogenous and exogenous sources during DNA replication. Considering 70,000 lesions occur in a single cell per day (Lindahl & Barnes, 2000), DNA lesions must be removed before the next cell division, to avoid their permanent conversion into mutations.

On the other hand, excision repair mechanisms are known to relentlessly coup with DNA damages that are potential sites of mutations. In deficiencies of mismatch repair and nucleotide excision repair, there are specific mutational signatures associated which contribute to different cancer types (Helleday, Eshtad & Nik-Zainal, 2014). Nucleotide excision repair-associated signature 7 exhibits replication timing differences and replication related strand asymmetry (Tomkova et al., 2018). Furthermore, because ERDs are more reachable relative to LRDs, mismatch repair causes a mutagenesis bias between these domains by effectively repairing the mismatches in ERDs (Supek & Lehner, 2015). Similarly, TCR creates a transcriptional strand asymmetry by repairing adducts only at transcribed strands and leaving the opposite strand untouched (Zheng, Wang, Chung, Moslehi, Sanborn, Hur, Collisson, Vemula, Naujokas, Chiotti & others, 2014). Even though signature 7 is linked with DNA replication timing differences and mutational strand asymmetry (Tomkova et al., 2018), the contribution of replication to nucleotide excision repair efficiency, and to the resulted mutation distribution is still unclear. In this study, we performed Damage-seq and XR-seq methods on UV-irradiated HeLa cells that are synchronized at early and late S phases to quantify (6-4)PP and CPD damages and their repair events.

There are some challenges in measuring the efficiency of nucleotide excision repair during replication. Firstly, because genome is copied once during the replication process, for both early and late S phases, only a fraction of the genome was replicated

when the experiments were conducted. Thus, we might observe a weak replication effect on repair rates throughout the genome, even though there is a strong effect in the regions that are replicated. Moreover, bulky DNA adducts can interrupt replication fork, which initiates inter-S phase checkpoint and therefore, delays the replication (Minca & Kowalski, 2011). Delaying of replication prevents us to successfully locate the exact sites of replication forks, which might reveal the local replication effect on nucleotide excision repair. To overcome these challenges, we focused our analyses only on replication-associated regions such as defined replication domains (Liu et al., 2016), initiation zones (Petryk et al., 2016), and replication origins (Besnard et al., 2012) of HeLa cells.

5.1 DNA replication elevates local nucleotide excision repair.

In the first part of the study, we examined the repair rate of nucleotide excision repair at large regions, while replication moves from early S phase to late S phase. We normalized repair events to damage (repair rate) for 2 reasons; firstly, to eliminate the sequence context bias that can lead to more damage formation, thus more repair. Secondly, DNA content of replication domains during replication are not uniform. Because, ERDs are replicated earlier, it is expected to have more DNA content from these domains. After calculating the repair rates in replication domains, we found that ERDs are repaired more efficiently that LRDs in both early and late S phases (Figure 4.2). Because ERDs are usually corresponding to open chromatin sites, they are more reachable for nucleotide excision repair, which in turn, promotes efficient repair. This result suggests that, like mismatch repair (Supek & Lehner, 2015), nucleotide excision repair creates a mutational difference between ERDs and LRDs by efficiently repairing ERDs. Moreover, replication domains are repaired in better efficiency when they are being replicated, which can be an indirect effect of replication progress by opening closely packed chromatin and thereby, promoting the nucleotide excision repair initiation (Figure 4.2). This effect is less prominent for (6-4)PP damages than that of CPDs, because they are repair faster and less effeted by chromatin states. In addition, CPDs at 2 hours exhibits less replication effect than that of CPDs at 12 minutes, possibly caused by the occurred DNA damages that interrupt replication.

Then, we examined the differences in repair rates of chromatin states for both ERDs and LRDs. The chromatin states that are associated with close regions are more varying in ERDs, likewise, the chromatin states that are associated with open regions displays same phenomenon in LRDs, simply because ERDs have less hete-

rochromatin regions, while LRDs have less open regions. Expectedly, the active promoters and enhancers are repaired efficiently for both ERDs and LRDs. Moreover, the major difference between phases occurred on close chromatin states (Figure 4.3, 6.13-6.17). This result indicates that movement of replication elevates the repair rates of close chromatin states by mediating chromatin opening. Lastly, we proposed a simple model to deminstrate the replication effect on repair (Figure 5.1).

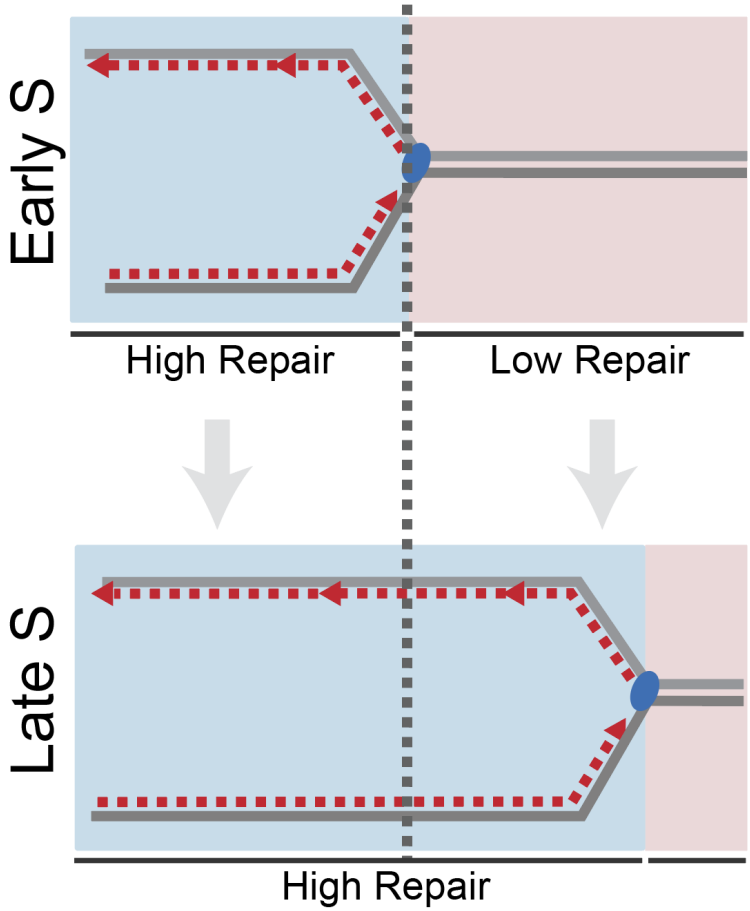

Figure 5.1 Repair preferences of Nucleotide Excision Repair during replication. During the early S phase, open chromatin regions (blue) which mostly correspond to ERDs, are repaired better than the condensed regions (red) because nucleotide excision repair can reach the open chromatin sites more efficiently than it reaches the condensed regions. While replication continues, those condensed regions loosen and become more reachable for excision repair.

5.2 Mutagenesis, UV-induced DNA damage, and repair display replicational strand asymmetry

Secondly, we examined a possible replicational strand asymmetry of mutagenesis, UV-induced DNA damage, and repair, respectively. Bidirectional movement of repli-

cation forks leads to asymmetric progress of polymerases ϵ and δ. Considering that these polymerases act differently when they are introduced with a lesion, the initiation of nucleotide excision repair might not occur at the same efficiency, thus might create a replicational strand asymmetry. On the lagging strands that are synthesized by polymerase δ, replication will not be inhibited when the replication fork encounters a damage, because of the discontinuous synthesis of Okazaki fragments. Instead, small gaps are left opposite to the damage site, which might be repaired post-replicative mechanisms. Conversely on the leading strands, while helicase continues to zipping double-stranded DNA, polymerase ϵ will be stalled. This disagreement then leads to ssDNA gaps that induce ATR pathway, thus increase the chance of repair (Byun, Pacek, Yee, Walter & Cimprich, 2005). In fact, multiple studies reported that the leading strands harbors less mutation than the lagging strands in multiple cancers (Haradhvala et al., 2016; Lujan et al., 2012; Reijns et al., 2015; Shinbrot et al., 2014). Even though this difference usually associated with replication-related processes such as POLE- and APOBEC-related mutagenesis (Haradhvala et al., 2016), we observed a significant mutational strand asymmetry at melanoma cancers, which are related to UV induced damages.

To assess the contribution of nucleotide excision repair on the mutagenesis, we further mapped damage and repair events on these regions separately. Remarkably, we observed a repair bias towards the leading strands (Figure 4.4), which is both in aggreament with mutational strand asymmetry of melanoma mutations and results of a recent study (Seplyarskiy et al., 2019). Furthermore, by revealing that both simulated and observed damage and repair events prefer the leading strands (Figure 4.5), we suggested that the strand asymmetry is somewhat caused by the sequence context. This preference is more prominent on CPDs at 2 hours, whereas (6-4)PPs and CPDs at 12 minutes display a weak or no preference to the leading strands. One possible reason can be that 12 minutes might not be enough to resolve polymerase-blocking at the lesion site. However, the pattern of CPD repair at 2 hours indicates the effect of replication on repair. Similarly, CPDs at 2 hours on high RFD exhibits a 100 to 200 kb long repair rate difference towards the leading strands, while (6-4)PPs and CPDs at 12 minutes display no difference between strands (Figure 6.62-6.67). However, (6-4)PPs and CPDs at 12 minutes around individual replication origins that are defined by SNS-seq data, exhibit a strand asymmetry on repair rate that favor the lagging strands, which is not consistent with other findings (Figure 4.6, 6.36, 6.37, 6.42-6.45). In fact, the difference between the repair and damage events are so small that can only be observed in a 5 kb region. Moreover, CPDs at 2 hours indicate a strand asymmetry on repair rate that favors the leading strands, which is in aggreament with other findings. Notably, LRDs have higher asymmetry of repair

rates in both early and late S phases. Even though it is expected to observe this in late S phases, simply because LRD repair is elevated when it is replicated (Figure 4.6), however, it not expected to observe this phenomenon in early S phase. Considering that replication is a dynamic process that is not fully strict on its order, there can be some initiation of replication forks in LRDs, which can create the asymmetry.

In conclusion, DNA replication can impact the efficiency of nucleotide excision repair on two levels. In genome-wide scale, ERDs are repaired faster than LRDs in both early and late S phases. Also, by opening the heterochromatins, replication promotes the repair of LRDs. In the scale of the replication origins, a replicational strand asymmetry is persistent in multiple replication-associated regions (individual replication origins, initiation zones, high RFD regions) and around initiation zones, there is a significant mutational strand asymmetry as well. By the stalling of polymerase ϵ during replication, the leading strands repaired more efficiently than the lagging strands, especially when the time after the UV exposure is increased. These findings reveal the contribution of nucleotide excision repair together with replication timing to mutational strand asymmetry of cancer genome.

BIBLIOGRAPHY

Adar, S., Hu, J., Lieb, J. D., & Sancar, A. (2016). Genome-wide kinetics of dna excision repair in relation to chromatin state and mutagenesis. *Proceedings of the National Academy of Sciences*, *113*(15), E2124–E2133.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.-M., & Lemaitre, J.-M. (2012). Unraveling cell type–specific and reprogrammable human replication origin signatures associated with g-quadruplex consensus motifs. *Nature structural & molecular biology*, *19*(8), 837.

Boyce, R. P. & Howard-Flanders, P. (1964). Release of ultraviolet light-induced thymine dimers from dna in e. coli k-12. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(2), 293.

Byun, T. S., Pacek, M., Yee, M.-c., Walter, J. C., & Cimprich, K. A. (2005). Functional uncoupling of mcm helicase and dna polymerase activities activates the atr-dependent checkpoint. *Genes & development*, *19*(9), 1040–1052.

Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., Chen, H.-C., Agarwala, R., McLaren, W. M., Ritchie, G. R., et al. (2011). Modernizing reference genome assemblies. *PLoS Biol*, *9*(7), e1001091.

Citterio, E., Rademakers, S., van der Horst, G. T., van Gool, A. J., Hoeijmakers, J. H., & Vermeulen, W. (1998). Biochemical and biological characterization of wild-type and atpase-deficient cockayne syndrome b repair protein. *Journal of Biological Chemistry*, *273*(19), 11844–11851.

Cleaver, J. (1968). Defective repair replication of dna in xeroderma pigmentosum. *nature*, *218*(5142), 652–656.

Cleaver, J. E. & Bootsma, D. (1975). Xeroderma pigmentosum: biochemical and genetic characteristics. *Annual review of genetics*, *9*(1), 19–38.

Cockayne, E. A. (1936). Dwarfism with retinal atrophy and deafness. *Archives of disease in childhood*, *11*(61), 1.

Cockayne, E. A. (1946). Dwarfism with retinal atrophy and deafness. *Archives of disease in childhood*, *21*(105), 52–54.

De Boer, J. & Hoeijmakers, J. H. (2000). Nucleotide excision repair and human syndromes. *Carcinogenesis*, *21*(3), 453–460.

Dimitrova, D. S. & Berezney, R. (2002). The spatio-temporal organization of dna replication sites is identical in primary, immortalized and transformed mammalian cells. *Journal of cell science*, *115*(21), 4037–4051.

Douki, T. & Cadet, J. (2001). Individual determination of the yield of the main uv-induced dimeric pyrimidine photoproducts in dna suggests a high mutagenicity of cc photolesions. *Biochemistry*, *40*(8), 2495–2501.

Drapkin, R., Reardon, J. T., Ansari, A., Huang, J.-C., Zawel, L., Ahn, K., Sancar, A., & Reinberg, D. (1994). Dual role of tfiih in dna excision repair and in transcription by rna polymerase ii. *Nature*, *368*(6473), 769–772.

Ernst, J. & Kellis, M. (2017). Chromatin-state discovery and genome annotation with chromhmm. *Nature protocols*, *12*(12), 2478.

Farkash-Amar, S., Lipson, D., Polten, A., Goren, A., Helmstetter, C., Yakhini, Z., & Simon, I. (2008). Global organization of replication time zones of the mouse

genome. *Genome research*, *18*(10), 1562–1570.

Fousteri, M., Vermeulen, W., van Zeeland, A. A., & Mullenders, L. H. (2006). Cockayne syndrome a and b proteins differentially regulate recruitment of chromatin remodeling and repair factors to stalled rna polymerase ii in vivo. *Molecular cell*, *23*(4), 471–482.

Friedberg, E. C., Walker, G. C., Siede, W., & Wood, R. D. (2005). *DNA repair and mutagenesis*. American Society for Microbiology Press.

Giglia-Mari, G., Miquel, C., Theil, A. F., Mari, P.-O., Hoogstraten, D., Ng, J. M., Dinant, C., Hoeijmakers, J. H., & Vermeulen, W. (2006). Dynamic interaction of ttda with tfiih is stabilized by nucleotide excision repair in living cells. *PLoS Biol*, *4*(6), e156.

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R., & Lieb, J. D. (2007). Faire (formaldehyde-assisted isolation of regulatory elements) isolates active regulatory elements from human chromatin. *Genome research*, *17*(6), 877–885.

Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., Dorschner, M. O., Gartler, S. M., & Stamatoyannopoulos, J. A. (2010). Sequencing newly replicated dna reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences*, *107*(1), 139–144.

Haradhvala, N. J., Polak, P., Stojanov, P., Covington, K. R., Shinbrot, E., Hess, J. M., Rheinbay, E., Kim, J., Maruvka, Y. E., Braunstein, L. Z., et al. (2016). Mutational strand asymmetries in cancer genomes reveal mechanisms of dna damage and repair. *Cell*, *164*(3), 538–549.

Hayward, N. K., Wilmott, J. S., Waddell, N., Johansson, P. A., Field, M. A., Nones, K., Patch, A.-M., Kakavand, H., Alexandrov, L. B., Burke, H., et al. (2017). Whole-genome landscapes of major melanoma subtypes. *Nature*, *545*(7653), 175–180.

Hedglin, M. & Benkovic, S. J. (2017). Eukaryotic translesion dna synthesis on the leading and lagging strands: Unique detours around the same obstacle. *Chemical reviews*, *117*(12), 7857–7877.

Helleday, T., Eshtad, S., & Nik-Zainal, S. (2014). Mechanisms underlying mutational signatures in human cancers. *Nature Reviews Genetics*, *15*(9), 585–598.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T. M., Schübeler, D., & Gilbert, D. M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol*, *6*(10), e245.

Hu, J. & Adar, S. (2017). The cartography of uv-induced dna damage formation and dna repair. *Photochemistry and photobiology*, *93*(1), 199–206.

Hu, J., Adebali, O., Adar, S., & Sancar, A. (2017). Dynamic maps of uv damage formation and repair for the human genome. *Proceedings of the National Academy of Sciences*, *114*(26), 6758–6763.

Hu, J., Li, W., Adebali, O., Yang, Y., Oztas, O., Selby, C. P., & Sancar, A. (2019). Genome-wide mapping of nucleotide excision repair with xr-seq. *Nature protocols*, *14*(1), 248–282.

Hu, J., Lieb, J. D., Sancar, A., & Adar, S. (2016). Cisplatin dna damage and repair maps of the human genome at single-nucleotide resolution. *Proceedings of the National Academy of Sciences*, *113*(41), 11507–11512.

Huang, J.-C., Svoboda, D. L., Reardon, J. T., & Sancar, A. (1992). Human nu-

cleotide excision nuclease removes thymine dimers from dna by incising the 22nd phosphodiester bond 5'and the 6th phosphodiester bond 3'to the photodimer. *Proceedings of the National Academy of Sciences*, *89*(8), 3664–3668.

Huang, W., Li, L., Myers, J. R., & Marth, G. T. (2012). Art: a next-generation sequencing read simulator. *Bioinformatics*, *28*(4), 593–594.

Ihaka, R. & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, *5*(3), 299–314.

Jackson, D. A. & Pombo, A. (1998). Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of s phase in human cells. *The Journal of cell biology*, *140*(6), 1285–1295.

Kent, W. J., Zweig, A. S., Barber, G., Hinrichs, A. S., & Karolchik, D. (2010). Bigwig and bigbed: enabling browsing of large distributed datasets. *Bioinformatics*, *26*(17), 2204–2207.

Khattak, M. & Wang, S. Y. (1972). The photochemical mechanism of pyrimidine cyclobutyl dimerization. *Tetrahedron*, *28*(4), 945–957.

Kiefer, J. (2007). Effects of ultraviolet radiation on dna. In *Chromosomal Alterations* (pp. 39–53). Springer.

Kielbassa, C., Roza, L., & Epe, B. (1997). Wavelength dependence of oxidative dna damage induced by uv and visible light. *Carcinogenesis*, *18*(4), 811–816.

Klungland, A., Höss, M., Gunz, D., Constantinou, A., Clarkson, S. G., Doetsch, P. W., Bolton, P. H., Wood, R. D., & Lindahl, T. (1999). Base excision repair of oxidative dna damage activated by xpg protein. *Molecular cell*, *3*(1), 33–42.

Koren, A., Handsaker, R. E., Kamitaki, N., Karlić, R., Ghosh, S., Polak, P., Eggan, K., & McCarroll, S. A. (2014). Genetic variation in human dna replication timing. *Cell*, *159*(5), 1015–1026.

Langley, A. R., Gräf, S., Smith, J. C., & Krude, T. (2016). Genome-wide identification and characterisation of human dna replication origins by initiation site sequencing (ini-seq). *Nucleic acids research*, *44*(21), 10230–10247.

Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nature methods*, *9*(4), 357.

Langston, L. D., Indiani, C., & O'Donnell, M. (2009). Whither the replisome: emerging perspectives on the dynamic nature of the dna replication machinery. *Cell Cycle*, *8*(17), 2686–2691.

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., Carter, S. L., Stewart, C., Mermel, C. H., Roberts, S. A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, *499*(7457), 214–218.

Lehmann, A. R. (2003). Dna repair-deficient diseases, xeroderma pigmentosum, cockayne syndrome and trichothiodystrophy. *Biochimie*, *85*(11), 1101–1111.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and samtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, S., Wehrenberg, B., Waldman, B. C., & Waldman, A. S. (2018). Mismatch tolerance during homologous recombination in mammalian cells. *DNA repair*, *70*, 25–36.

Li, W., Hu, J., Adebali, O., Adar, S., Yang, Y., Chiou, Y.-Y., & Sancar, A. (2017). Human genome-wide repair map of dna damage caused by the cigarette smoke

carcinogen benzo [a] pyrene. *Proceedings of the National Academy of Sciences*, *114*(26), 6752–6757.

Li, W. & Sancar, A. (2020). Methodologies for detecting environmentally induced dna damage and repair. *Environmental and Molecular Mutagenesis*, *61*.

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, *326*(5950), 289–293.

Lindahl, T. & Barnes, D. (2000). Repair of endogenous dna damage. In *Cold Spring Harbor symposia on quantitative biology*, volume 65, (pp. 127–134). Cold Spring Harbor Laboratory Press.

Liu, F., Ren, C., Li, H., Zhou, P., Bo, X., & Shu, W. (2016). De novo identification of replication-timing domains in the human genome by deep learning. *Bioinformatics*, *32*(5), 641–649.

Lujan, S. A., Williams, J. S., Pursell, Z. F., Abdulovic-Cui, A. A., Clark, A. B., McElhinny, S. A. N., & Kunkel, T. A. (2012). Mismatch repair balances leading and lagging strand dna replication fidelity. *PLoS Genet*, *8*(10), e1003016.

Marteijn, J. A., Lans, H., Vermeulen, W., & Hoeijmakers, J. H. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. *Nature reviews Molecular cell biology*, *15*(7), 465–481.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10–12.

Minca, E. C. & Kowalski, D. (2011). Replication fork stalling by bulky dna damage: localization at active origins and checkpoint modulation. *Nucleic acids research*, *39*(7), 2610–2623.

Mizukoshi, T., Kodama, T. S., Fujiwara, Y., Furuno, T., Nakanishi, M., & Iwai, S. (2001). Structural study of dna duplexes containing the (6–4) photoproduct by fluorescence resonance energy transfer. *Nucleic acids research*, *29*(24), 4948–4954.

Modrich, P. (1997). Strand-specific mismatch repair in mammalian cells. *Journal of Biological Chemistry*, *272*(40), 24727–24730.

Mouret, S., Philippe, C., Gracia-Chantegrel, J., Banyasz, A., Karpati, S., Markovitsi, D., & Douki, T. (2010). Uva-induced cyclobutane pyrimidine dimers in dna: a direct photochemical mechanism? *Organic & biomolecular chemistry*, *8*(7), 1706–1711.

Muftuoglu, M., Selzer, R., Tuo, J., Brosh Jr, R. M., & Bohr, V. A. (2002). Phenotypic consequences of mutations in the conserved motifs of the putative helicase domain of the human cockayne syndrome group b gene. *Gene*, *283*(1-2), 27–40.

Nakayasu, H. & Berezney, R. (1989). Mapping replicational sites in the eucaryotic cell nucleus. *Journal of Cell Biology*, *108*(1), 1–11.

Neill, C. A. & Dingwall, M. M. (1950). A syndrome resembling progeria: A review of two cases. *Archives of disease in childhood*, *25*(123), 213.

Nguyen, H. T. & Minton, K. W. (1988). Ultraviolet-induced dimerization of nonadjacent pyrimidines: A potential mechanism for the targeted- 1 frameshift mutation. *Journal of molecular biology*, *200*(4), 681–693.

O'keefe, R. T., Henderson, S. C., & Spector, D. L. (1992). Dynamic organization of dna replication in mammalian cell nuclei: spatially and temporally defined

replication of chromosome-specific alpha-satellite dna sequences. *The Journal of cell biology*, *116*(5), 1095–1110.

Park, H., Zhang, K., Ren, Y., Nadji, S., Sinha, N., Taylor, J.-S., & Kang, C. (2002). Crystal structure of a dna decamer containing a cis-syn thymine dimer. *Proceedings of the National Academy of Sciences*, *99*(25), 15965–15970.

Patrick, M. H. (1977). Studies on thymine-derived uv photoproducts in dna—i. formation and biological role of pyrimidine adducts in dna. *Photochemistry and photobiology*, *25*(4), 357–372.

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), 1–13.

Pope, B. D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D. L., Wang, Y., Hansen, R. S., Canfield, T. K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. *Nature*, *515*(7527), 402–405.

Quinlan, A. R. & Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Ramey, C. (1998). Bash reference manual. *Network Theory Limited*, *15*.

Reardon, J. T. & Sancar, A. (2005). Nucleotide excision repair. *Progress in nucleic acid research and molecular biology*, *79*, 183–235.

Reijns, M. A., Kemp, H., Ding, J., de Procé, S. M., Jackson, A. P., & Taylor, M. S. (2015). Lagging-strand replication shapes the mutational landscape of the genome. *Nature*, *518*(7540), 502–506.

Rossum, G. (1995). Python reference manual. Technical report, NLD.

Rupert, C. S., Goodgal, S. H., & Herriott, R. M. (1958). Photoreactivation in vitro of ultraviolet inactivated hemophilus influenzae transforming factor. *The Journal of general physiology*, *41*(3), 451.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T. C., Robins, A. J., Dalton, S., & Gilbert, D. M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome research*, *20*(6), 761–770.

Sancar, A. (2016). Mechanisms of dna repair by photolyase and excision nuclease (nobel lecture). *Angewandte Chemie International Edition*, *55*(30), 8502–8527.

Schreier, W. J., Schrader, T. E., Koller, F. O., Gilch, P., Crespo-Hernández, C. E., Swaminathan, V. N., Carell, T., Zinth, W., & Kohler, B. (2007). Thymine dimerization in dna is an ultrafast photoreaction. *Science*, *315*(5812), 625–629.

Schuster-Böckler, B. & Lehner, B. (2012). Chromatin organization is a major influence on regional mutation rates in human cancer cells. *nature*, *488*(7412), 504–507.

Scrima, A., Koníčková, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., Nakatani, Y., Iwai, S., Pavletich, N. P., & Thomä, N. H. (2008). Structural basis of uv dna-damage recognition by the ddb1–ddb2 complex. *Cell*, *135*(7), 1213–1223.

Selby, C. P. & Sancar, A. (1997a). Cockayne syndrome group b protein enhances elongation by rna polymerase ii. *Proceedings of the National Academy of Sciences*, *94*(21), 11205–11209.

Selby, C. P. & Sancar, A. (1997b). Human transcription-repair coupling factor csb/ercc6 is a dna-stimulated atpase but is not a helicase and does not disrupt

the ternary transcription complex of stalled rna polymerase ii. *Journal of Biological Chemistry*, *272*(3), 1885–1890.

Seplyarskiy, V. B., Akkuratov, E. E., Akkuratova, N., Andrianova, M. A., Nikolaev, S. I., Bazykin, G. A., Adameyko, I., & Sunyaev, S. R. (2019). Error-prone bypass of dna lesions during lagging-strand replication is a common source of germline and cancer mutations. *Nature genetics*, *51*(1), 36–41.

Setlow, R. & Carrier, W. (1964). The disappearance of thymine dimers from dna: an error-correcting mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, *51*(2), 226.

Shinbrot, E., Henninger, E. E., Weinhold, N., Covington, K. R., Göksenin, A. Y., Schultz, N., Chao, H., Doddapaneni, H., Muzny, D. M., Gibbs, R. A., et al. (2014). Exonuclease mutations in dna polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. *Genome research*, *24*(11), 1740–1750.

Stamatoyannopoulos, J. A., Adzhubei, I., Thurman, R. E., Kryukov, G. V., Mirkin, S. M., & Sunyaev, S. R. (2009). Human mutation rate associated with dna replication timing. *Nature genetics*, *41*(4), 393–395.

Sugasawa, K., Ng, J. M., Masutani, C., Iwai, S., van der Spek, P. J., Eker, A. P., Hanaoka, F., Bootsma, D., & Hoeijmakers, J. H. (1998). Xeroderma pigmentosum group c protein complex is the initiator of global genome nucleotide excision repair. *Molecular cell*, *2*(2), 223–232.

Supek, F. & Lehner, B. (2015). Differential dna mismatch repair underlies mutation rate variation across the human genome. *Nature*, *521*(7550), 81–84.

Svejstrup, J. Q. (2002). Mechanisms of transcription-coupled dna repair. *Nature Reviews Molecular Cell Biology*, *3*(1), 21–29.

Takebayashi, S.-i., Ogata, M., & Okumura, K. (2017). Anatomy of mammalian replication domains. *Genes*, *8*(4), 110.

Takebayashi, S.-i., Sugimura, K., Saito, T., Sato, C., Fukushima, Y., Taguchi, H., & Okumura, K. (2005). Regulation of replication at the r/g chromosomal band boundary and pericentromeric heterochromatin of mammalian cells. *Experimental cell research*, *304*(1), 162–174.

Taylor, J.-S. & Brockie, I. R. (1988). Synthesis of a trans-syn thymine dimer building block. solid phase synthesis of cgtat [t, s] tatgc. *Nucleic acids research*, *16*(11), 5123–5136.

Taylor, J. S. & Cohrs, M. P. (1987). Dna, light, and dewar pyrimidinones: the structure and biological significance to tpt3. *Journal of the American Chemical Society*, *109*(9), 2834–2835.

Tomasetti, C. & Vogelstein, B. (2015). Variation in cancer risk among tissues can be explained by the number of stem cell divisions. *Science*, *347*(6217), 78–81.

Tomkova, M., Tomek, J., Kriaucionis, S., & Schuster-Böckler, B. (2018). Mutational signature distribution varies with dna replication timing and strand asymmetry. *Genome biology*, *19*(1), 1–12.

Tornaletti, S., Reines, D., & Hanawalt, P. C. (1999). Structural characterization of rna polymerase ii complexes arrested by a cyclobutane pyrimidine dimer in the transcribed strand of template dna. *Journal of Biological Chemistry*, *274*(34), 24124–24130.

Van Hoffen, A., Venema, J., Meschini, R., Van Zeeland, A., & Mullenders, L. (1995). Transcription-coupled repair removes both cyclobutane pyrimidine dimers and

6-4 photoproducts with equal efficiency and in a sequential way from transcribed dna in xeroderma pigmentosum group c fibroblasts. *The EMBO journal*, *14*(2), 360–367.

Wacker, A., Dellweg, H., Träger, L., Kornhauser, A., Lodemann, E., Türck, G., Selzer, R., Chandra, P., & Ishimoto, M. (1964). Organic photochemistry of nucleic acids. *Photochemistry and Photobiology*, *3*(4), 369–394.

Whitmore, S., Potten, C., Chadwick, C., Strickland, P. T., & Morison, W. (2001). Effect of photoreactivating light on uv radiation-induced alterations in human skin. *Photodermatology, photoimmunology & photomedicine*, *17*(5), 213–217.

Yeeles, J. T., Poli, J., Marians, K. J., & Pasero, P. (2013). Rescuing stalled or damaged replication forks. *Cold Spring Harbor perspectives in biology*, *5*(5), a012815.

Yimit, A., Adebali, O., Sancar, A., & Jiang, Y. (2019). Differential damage and repair of dna-adducts induced by anti-cancer drug cisplatin across mouse organs. *Nature communications*, *10*(1), 1–11.

Zheng, C. L., Wang, N. J., Chung, J., Moslehi, H., Sanborn, J. Z., Hur, J. S., Collisson, E. A., Vemula, S. S., Naujokas, A., Chiotti, K. E., et al. (2014). Transcription restores dna repair to heterochromatin, determining regional mutation rates in cancer genomes. *Cell reports*, *9*(4), 1228–1234.

6. APPENDIX

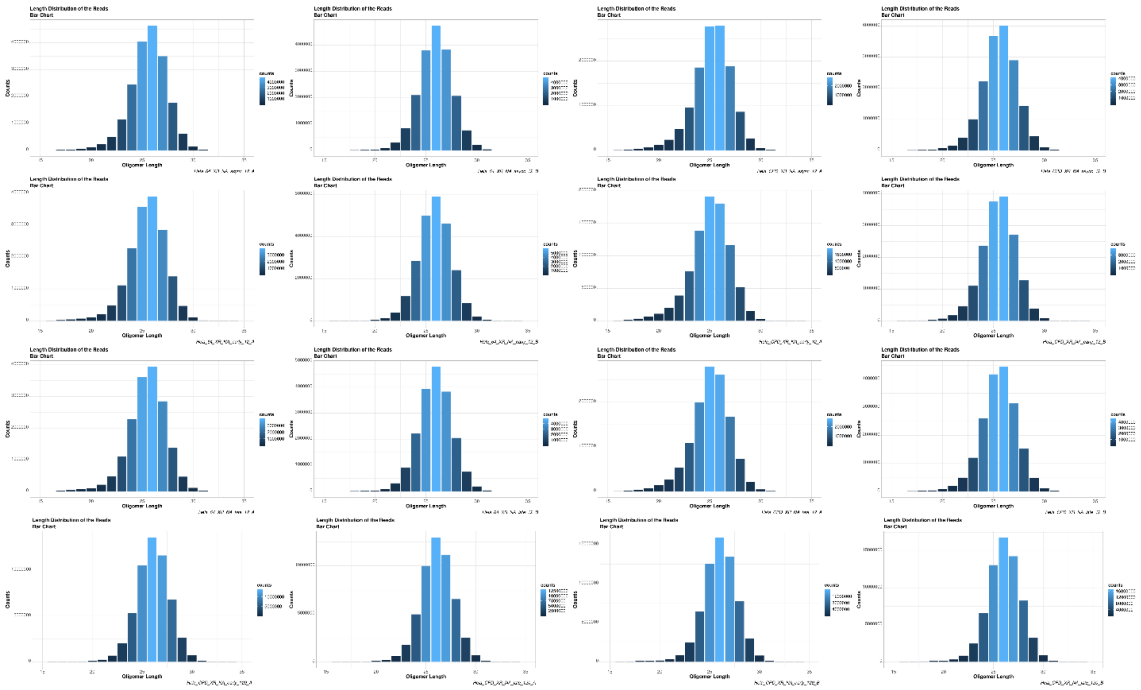

Figure 6.1 Length distribution of excised oligomers of XR-seq samples after adaptor trimming and duplicate removal. Majority of the oligomers are 26 nucleotides long.

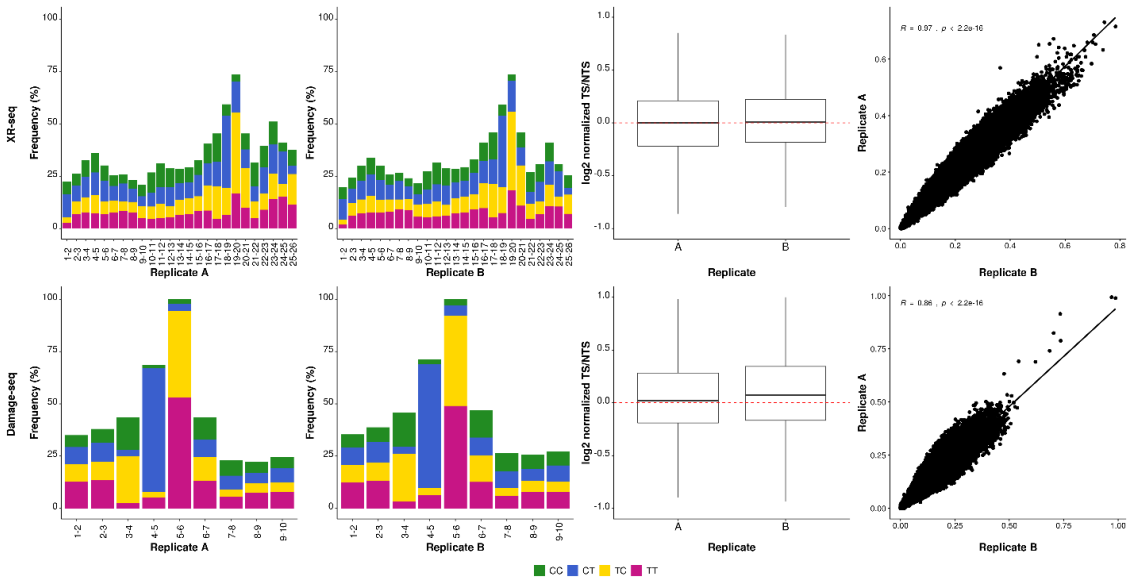

Figure 6.2 Control figures of (6-4)PP asynchronized samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

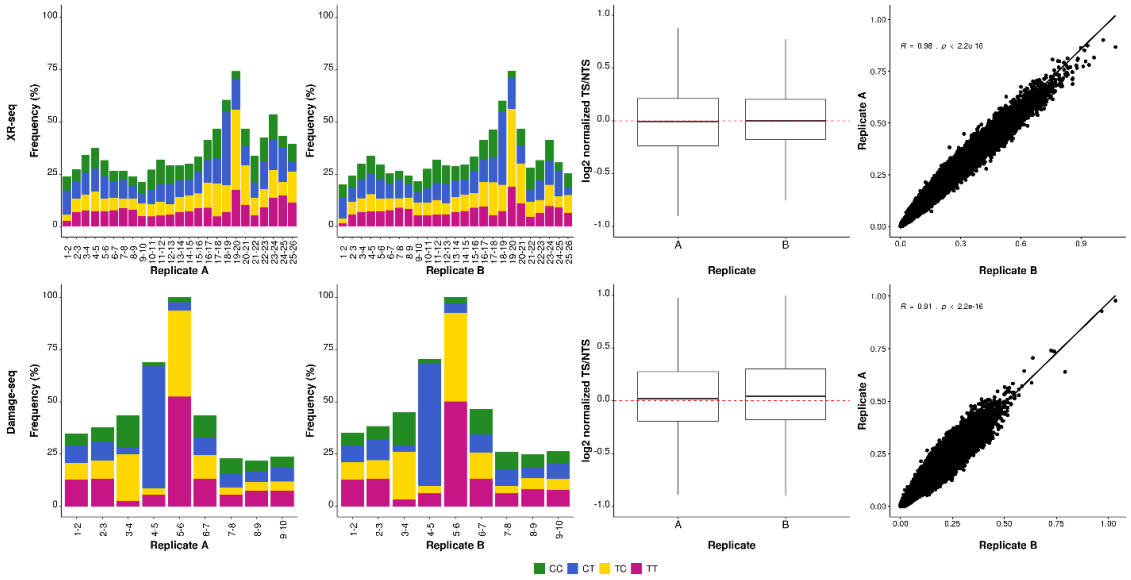

Figure 6.3 Control figures of (6-4)PP early phased samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

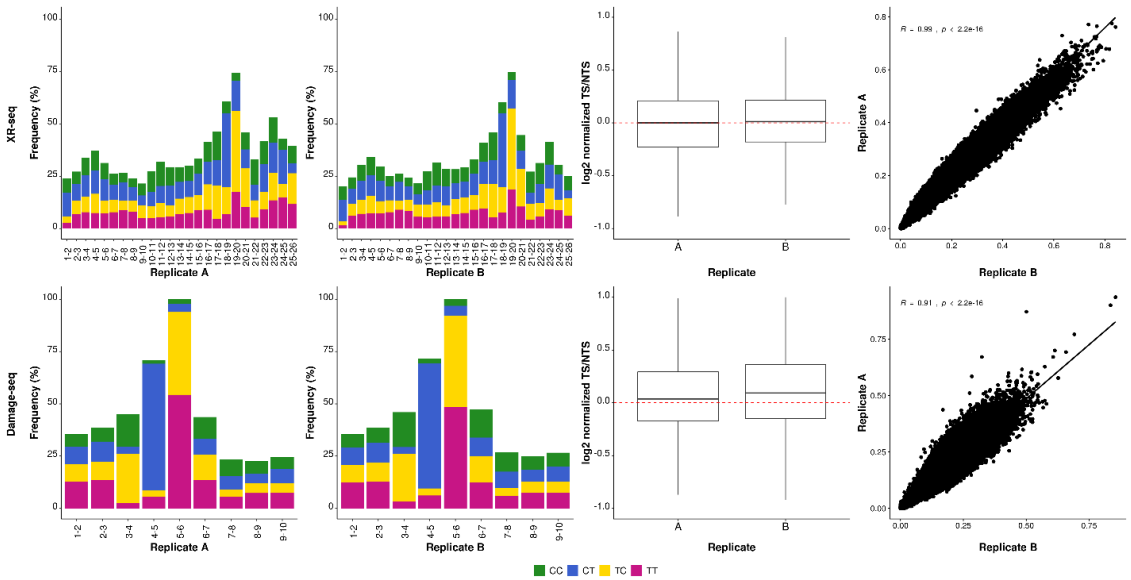

Figure 6.4 Control figures of (6-4)PP late phased samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

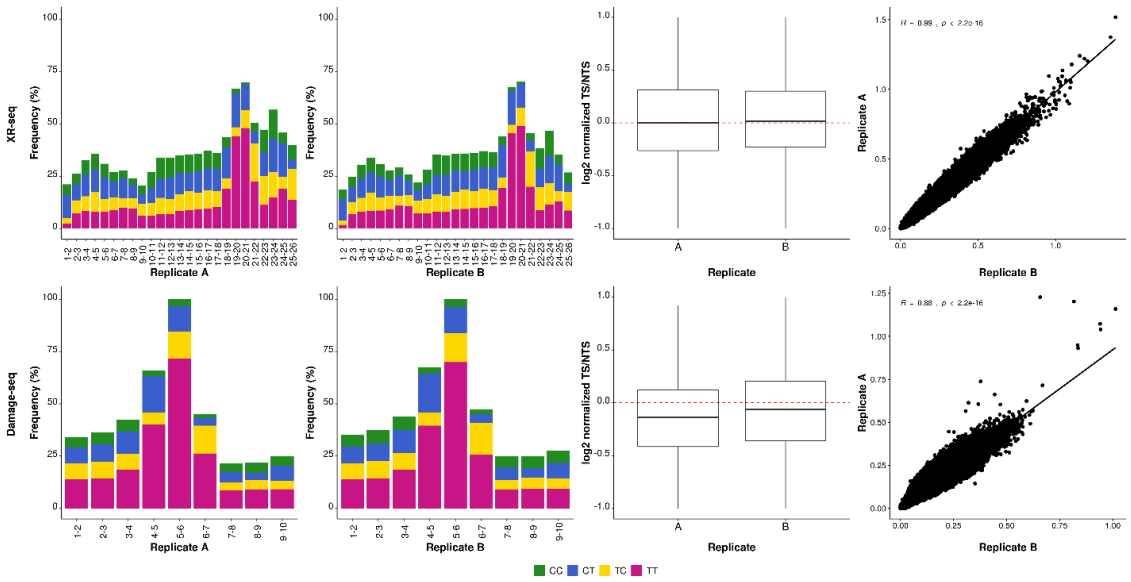

Figure 6.5 Control figures of CPD asynchronized samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

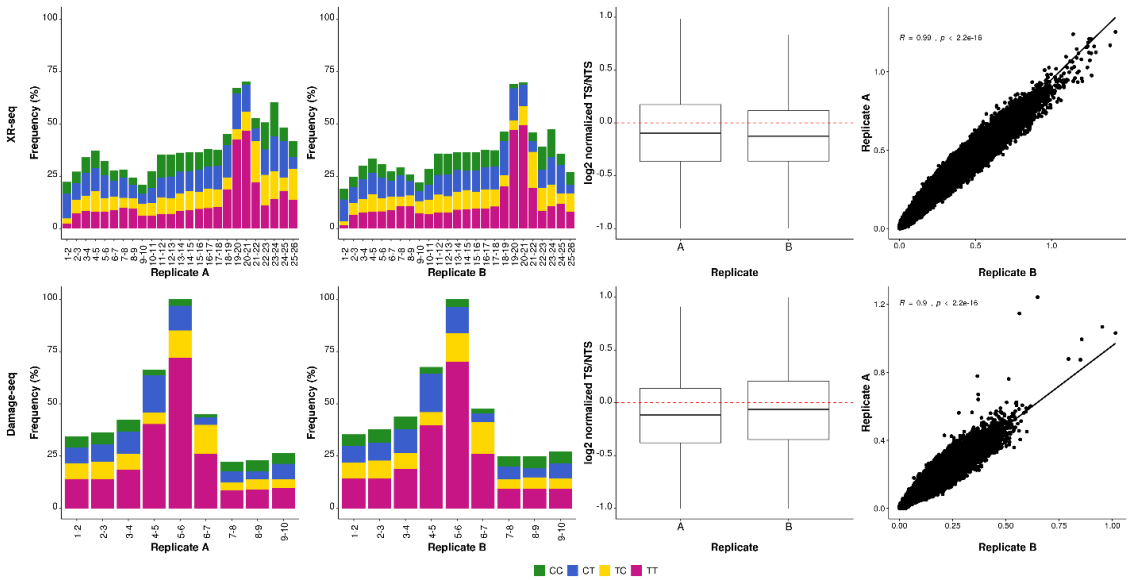

Figure 6.6 Control figures of CPD early phased samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

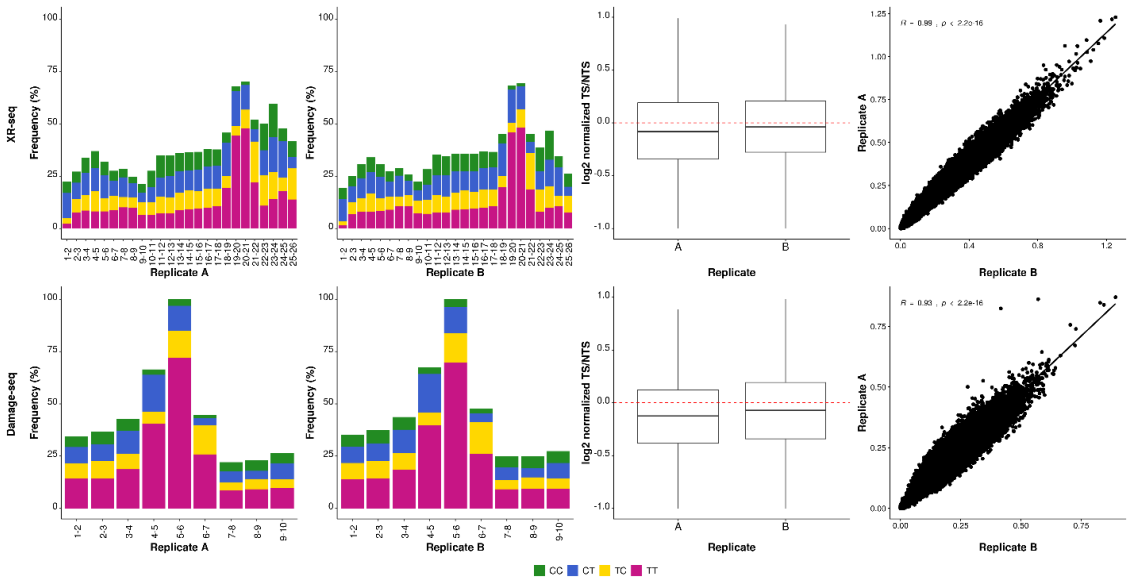

Figure 6.7 Control figures of CPD late phased samples at 12 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

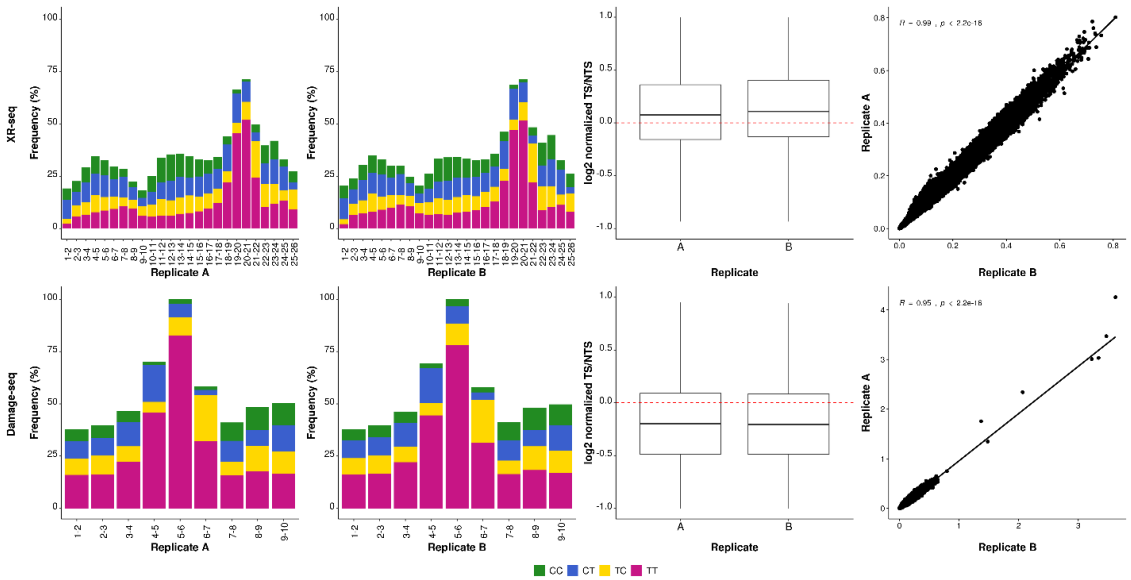

Figure 6.8 Control figures of CPD early phased samples at 120 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

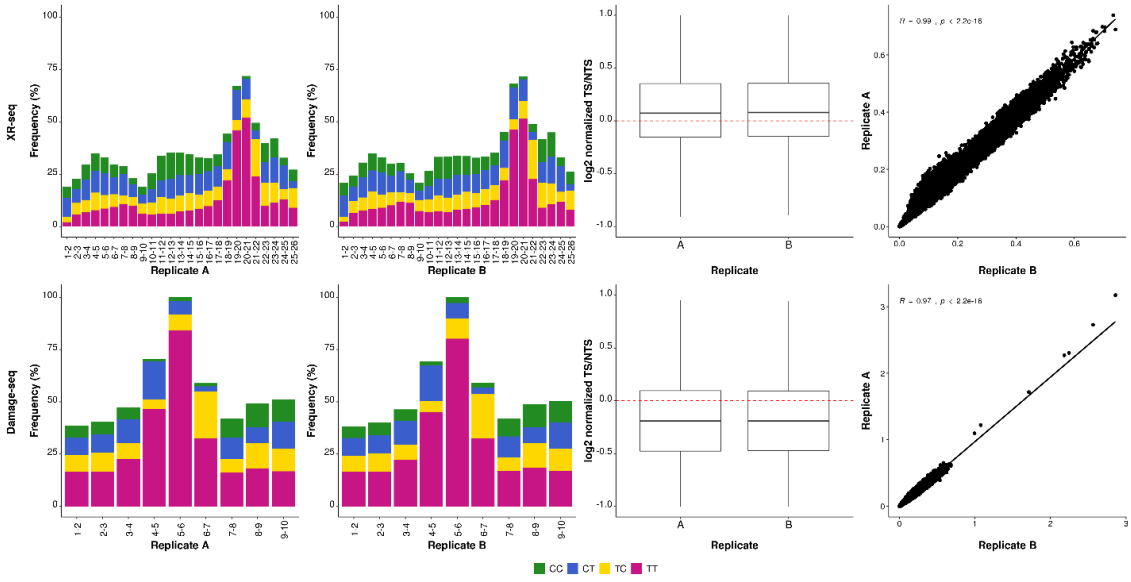

Figure 6.9 Control figures of CPD late phased samples at 120 minutes. Column 1 is the correlation plot of the biological replicates (A & B). Column 1 and 2 displays the dinucleotide composition frequency of replicate A and B, respectively. Column 3 is the log2 transformed TS/NTS ratios of replicate A and B. Row 1 is the results of XR-seq samples, and row 2 is the results of Damage-seq samples. Column 4 is the correlation plot of the biological replicates (A & B). Correlation coefficient is calculated by Spearman's rank correlation test.

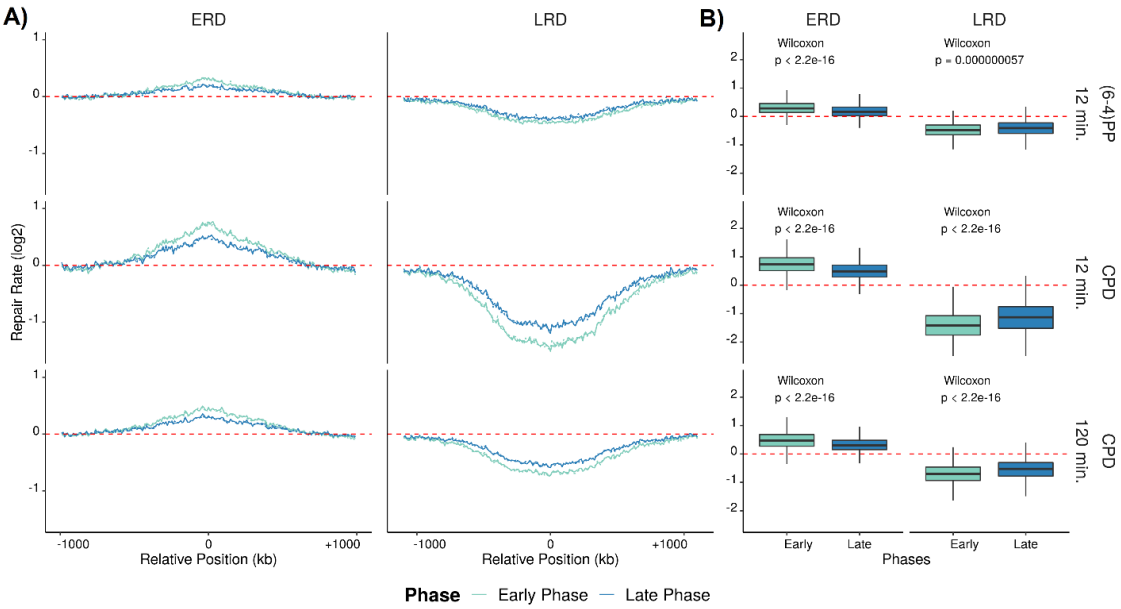

Figure 6.10 The shift of repair efficiency at replication domains during replication. A) Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 2 Mb regions with 10 kb intervals, which early replication domains (ERDs, left) and late replication domains (LRDs, right) positioned at the center of the region. B) RPKM values of XR-seq samples are divided by Damage-seq samples (Repair Rate) for both ERDs (left) and LRDs (right) and log2 transformed. Wilcoxon test is used to assess the significance of difference between early and late S phases. The light blue lines are the early phase repair rate values and dark blue lines are the late phase repair rate values. Above the red horizontal dashed line demonstrates that repair is higher than damage, below demonstrates that damage is higher. (Same analysis as Figure 2, it is performed with replicate B)

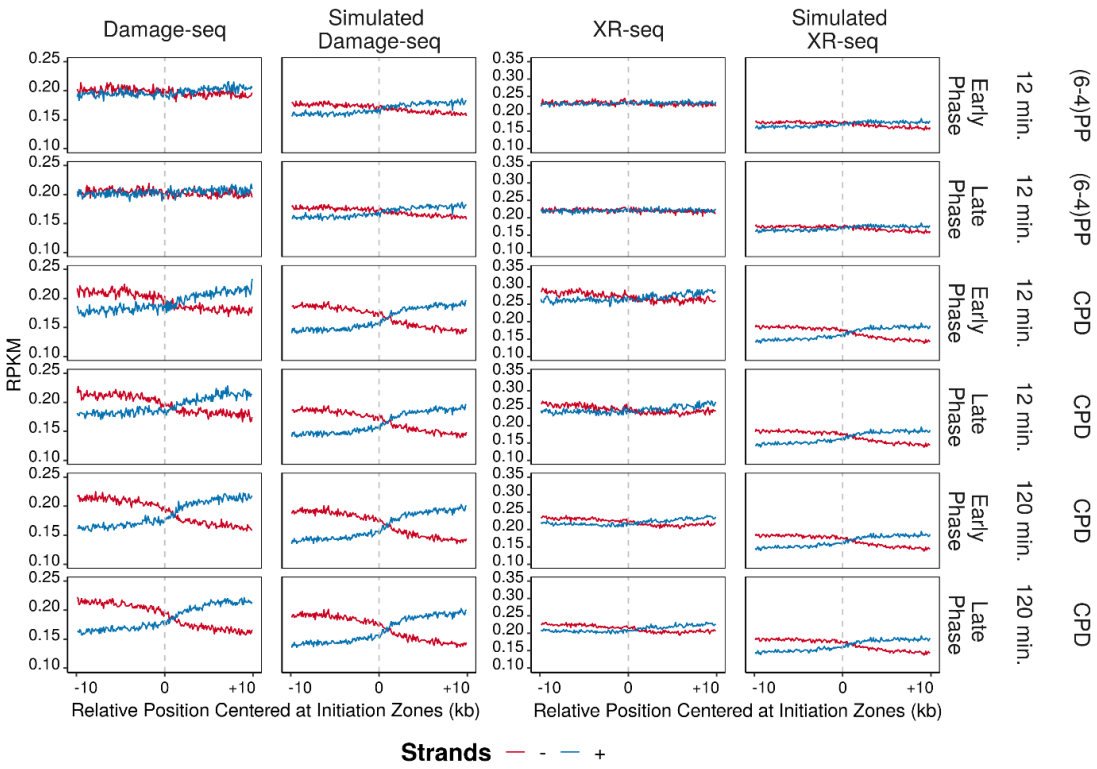

Figure 6.11 Strand asymmetry around initiation zones caused by nucleotide bias. RPKM values of real and simulated Damage-seq samples (left) and XR-seq samples (right) are calculated in 20 kb windows with 100 base pair intervals, which Initiation Zones are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region. (Same analysis as Figure 4, it is performed with replicate B)

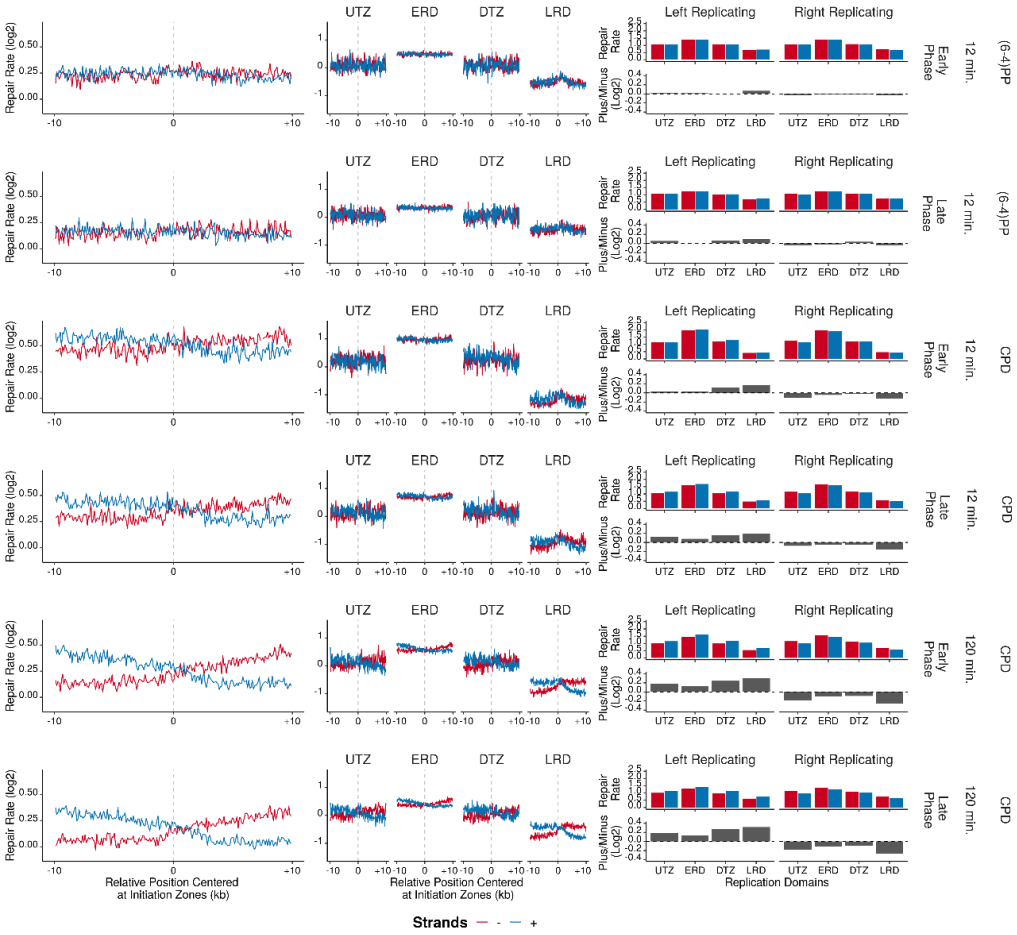

Figure 6.12 Repair rate asymmetry around initiation zones and replication domains. (Left) Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which Initiation Zones are positioned at the center of the region. (Middle) Same analysis performed, however initiation zones separated into their corresponding replication domains. (Right) The strand differences at left (left part of the gray line) and right (right part of the gray line) replicating directions are shown by taking the mean of the intervals, separately for the strands. Below that, strands are divided to each other (Plus/Minus) and log2 transformed to better visualize the asymmetry at each replication domain. Blue lines are the plus strands and red lines are the minus strands. Gray vertical dashed line shows the center of the region. (Same analysis as Figure 6, it is performed with replicate B)

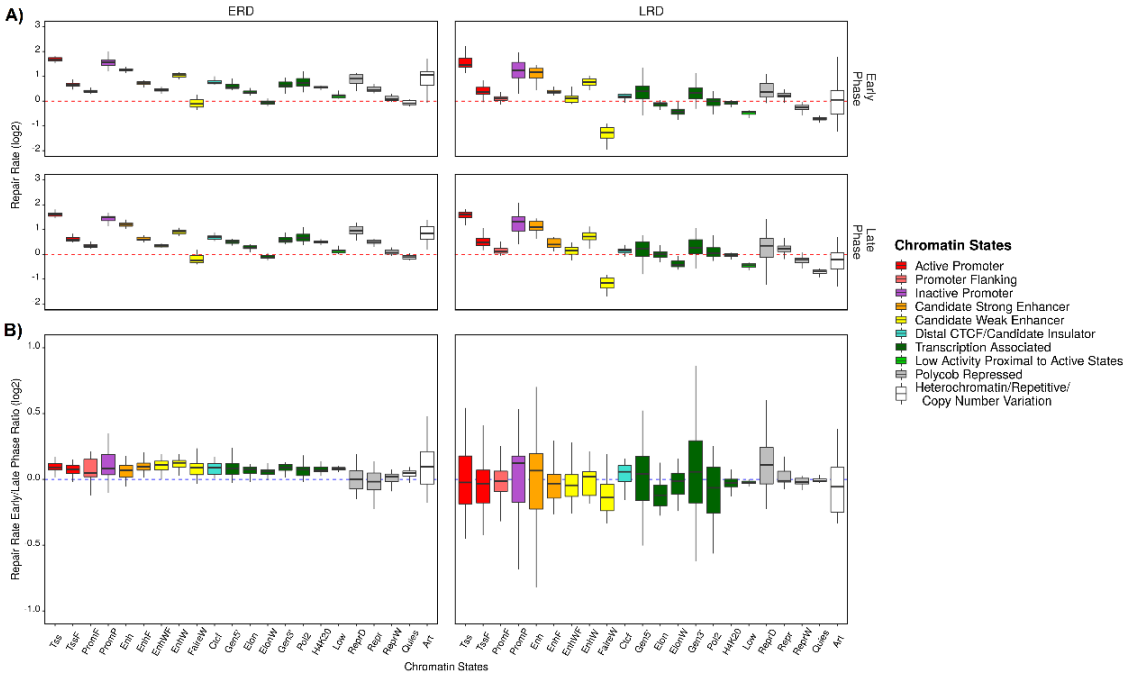

Figure 6.13 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of (6-4)PP samples at 12 minutes are calculated, log2 transformed B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). Analysis is performed on replicate A.

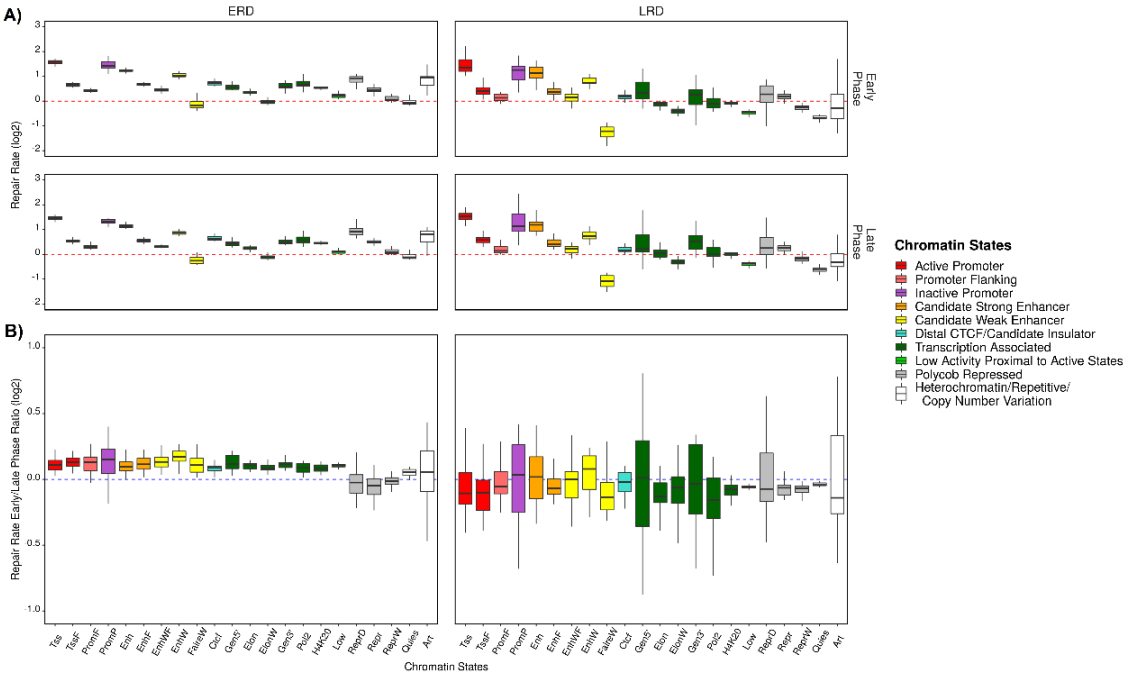

Figure 6.14 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of (6-4)PP samples at 12 minutes are calculated, log2 transformed, B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). Analysis is performed on replicate B.

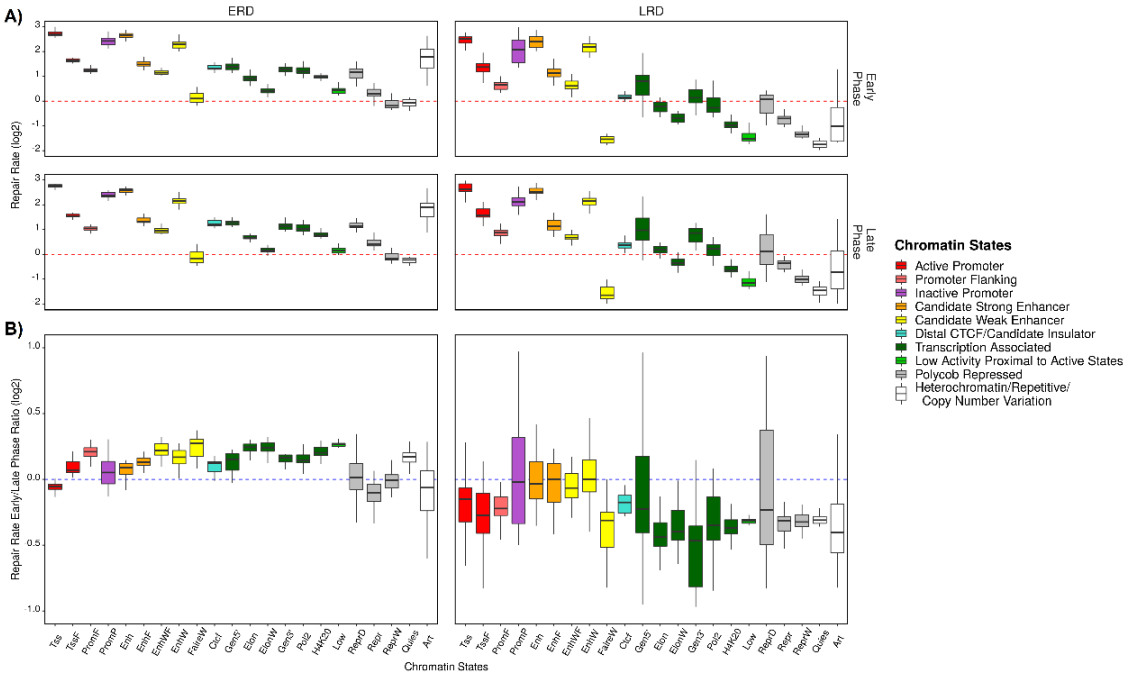

Figure 6.15 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of CPD samples at 12 minutes are calculated, log2 transformed, B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). (Same analysis as Figure 3, it is performed with replicate B)

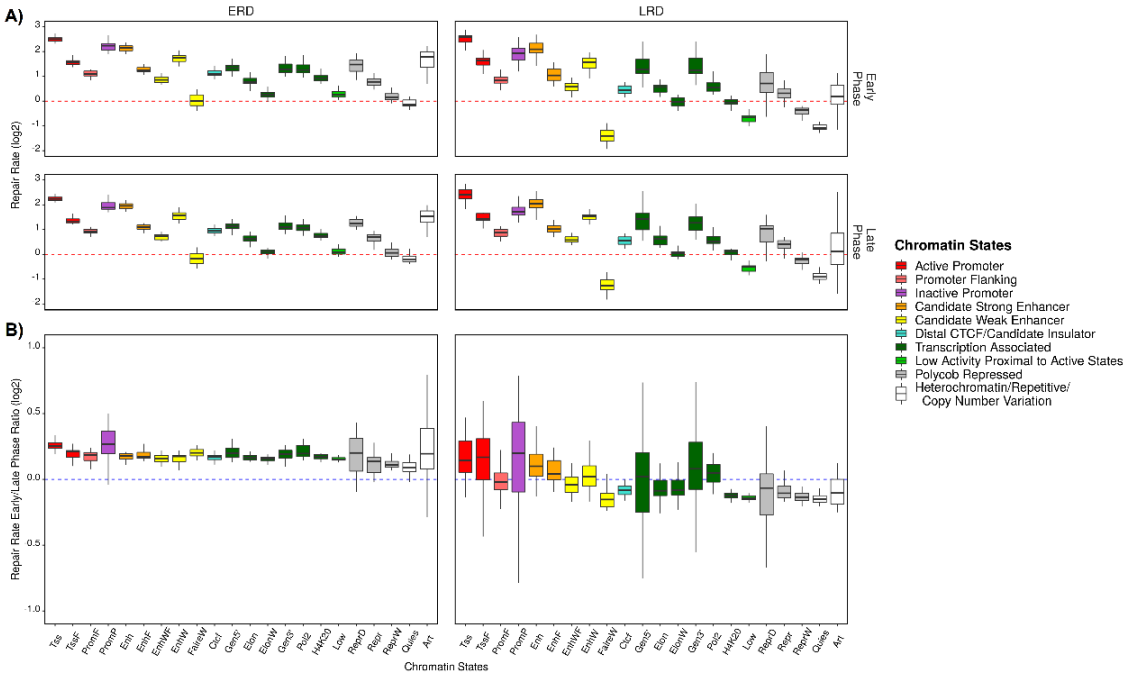

Figure 6.16 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of CPD samples at 120 minutes are calculated, log2 transformed, B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). Analysis is performed on replicate A.

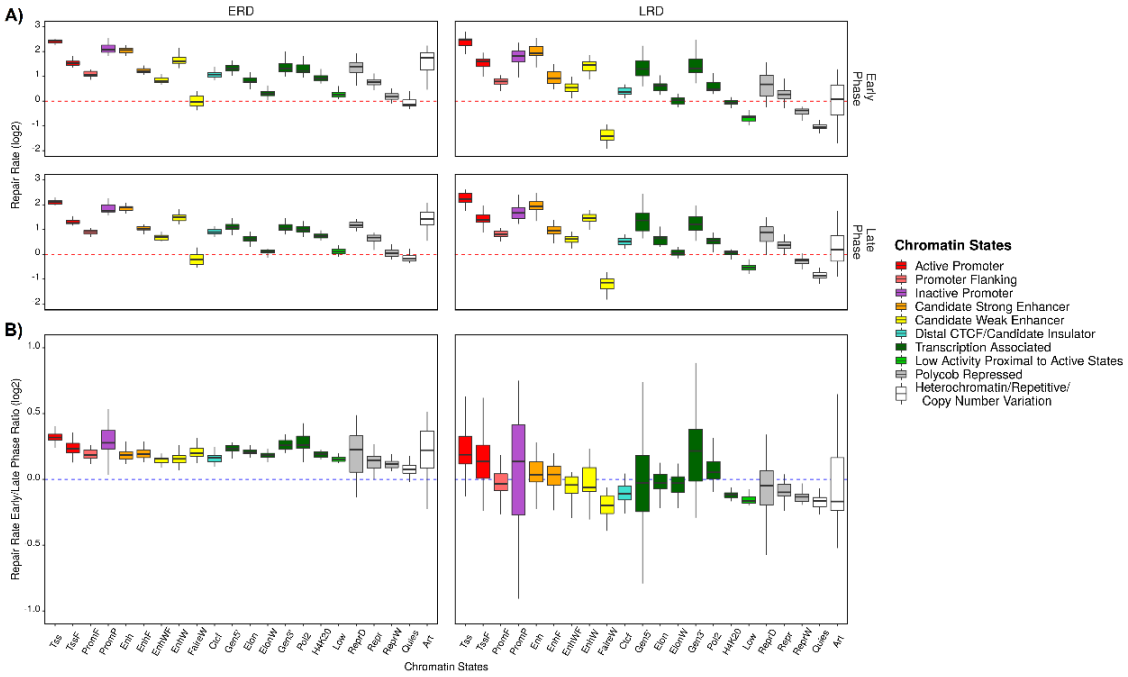

Figure 6.17 The effect of Chromatin States to repair efficiency of replication domains. A) Repair rates (XR-seq/Damage-seq) of CPD samples at 120 minutes are calculated, log2 transformed B) and for every region, the repair rates at early S phase divided by repair rates at late S phase to spot the chromatin states that are repaired dominant at a phase. Above the red horizontal dashed line demonstrates that repair is higher than damage (A), and the blue horizontal dashed line demonstrates that the chromatin state has higher repair efficiency at early S phase than it has at late S phase (B). Analysis is performed on replicate B.

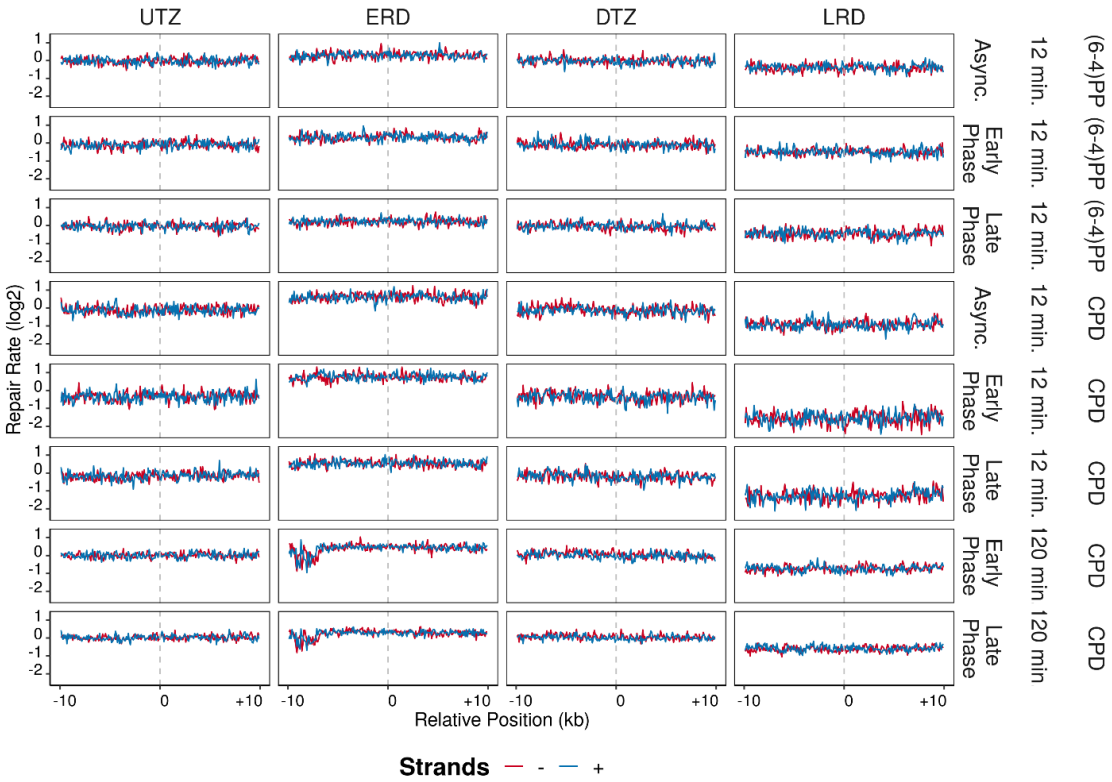

Figure 6.18 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

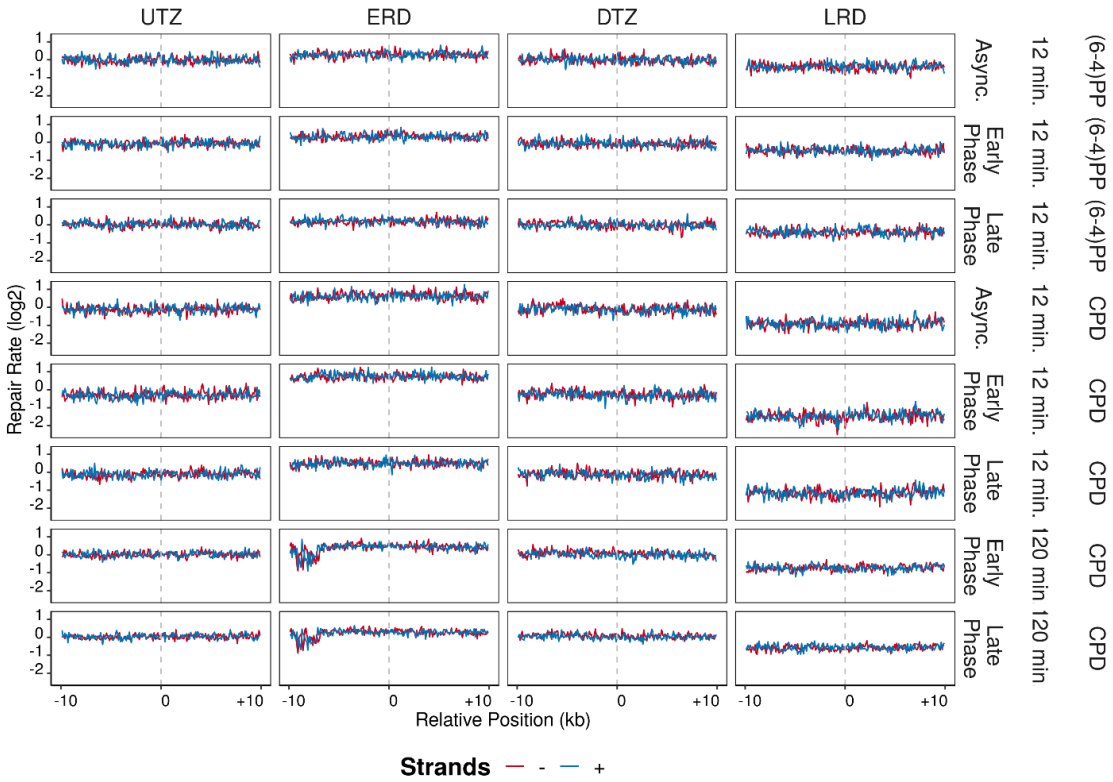

Figure 6.19 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

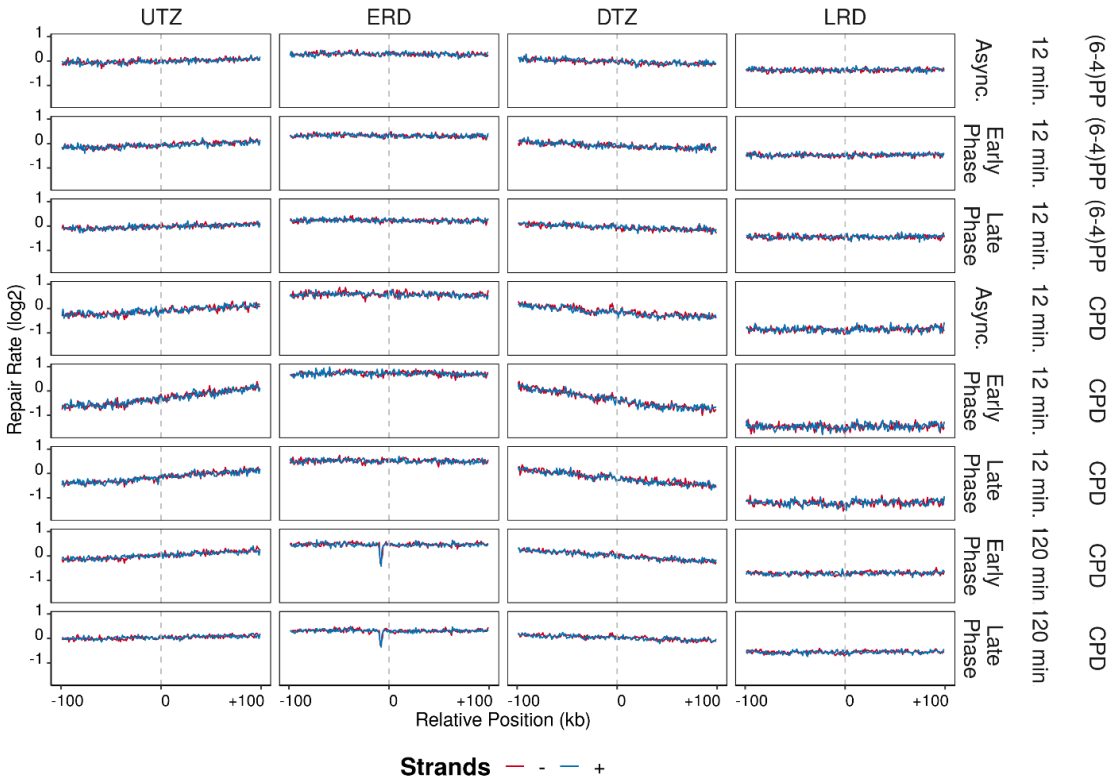

Figure 6.20 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

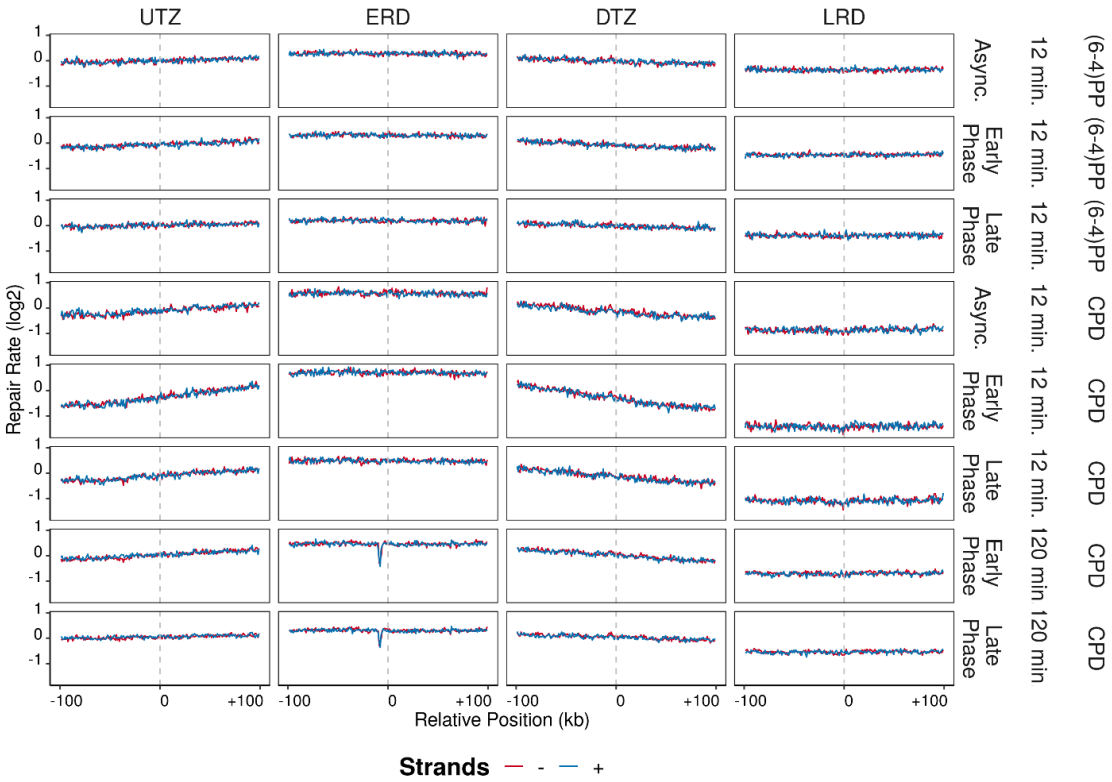

Figure 6.21 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

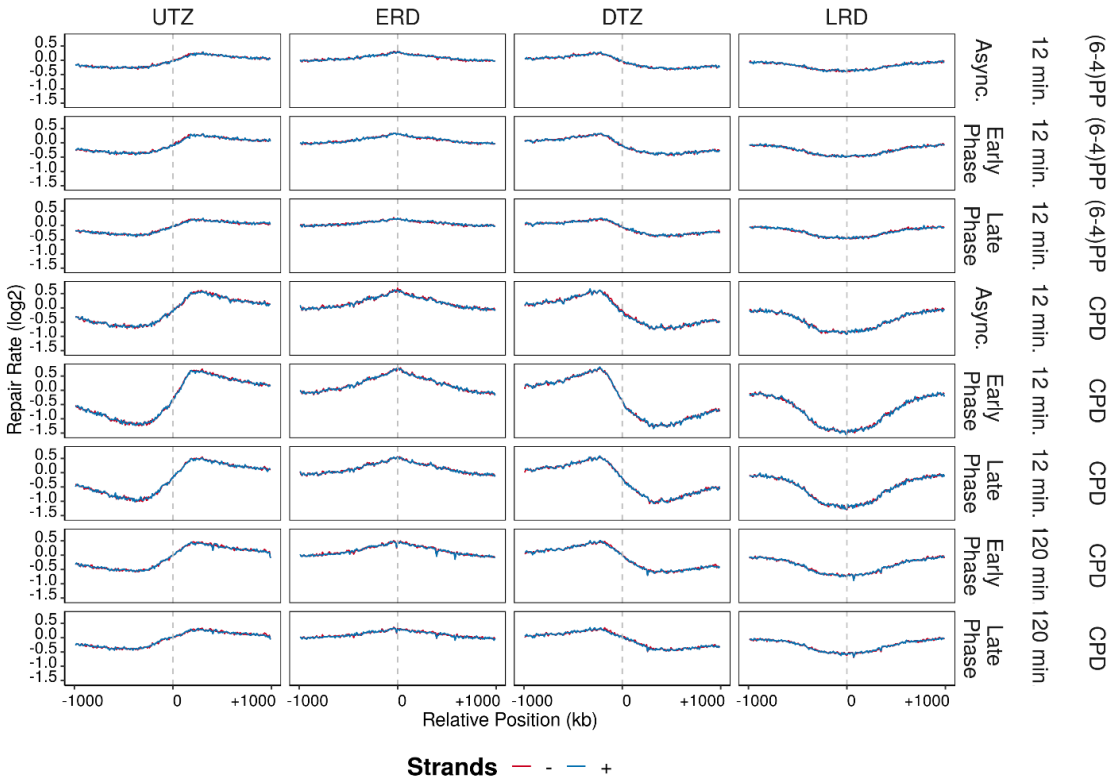

Figure 6.22 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

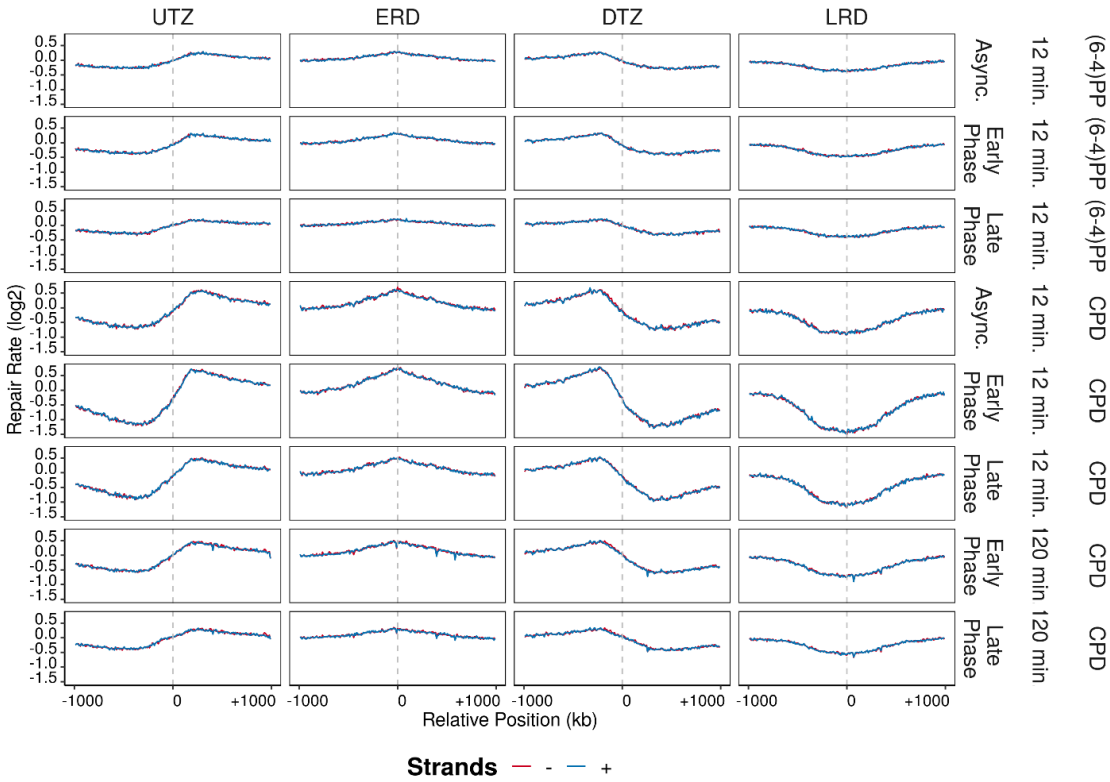

Figure 6.23 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

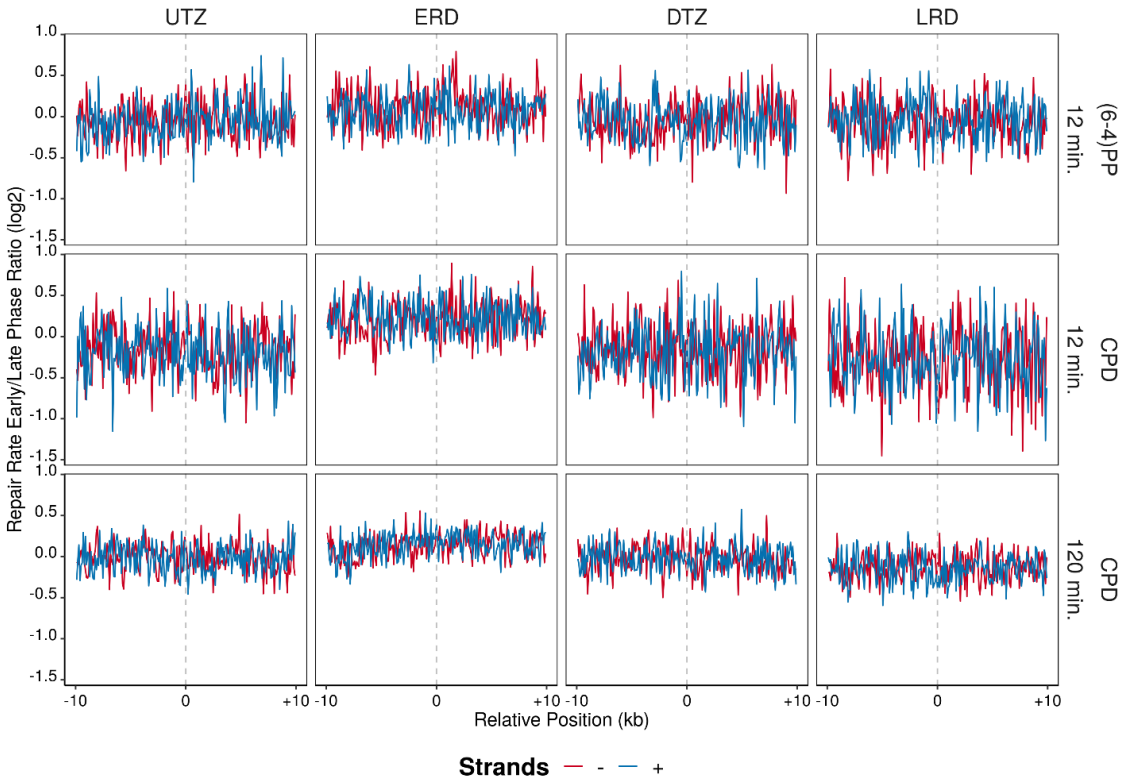

Figure 6.24 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

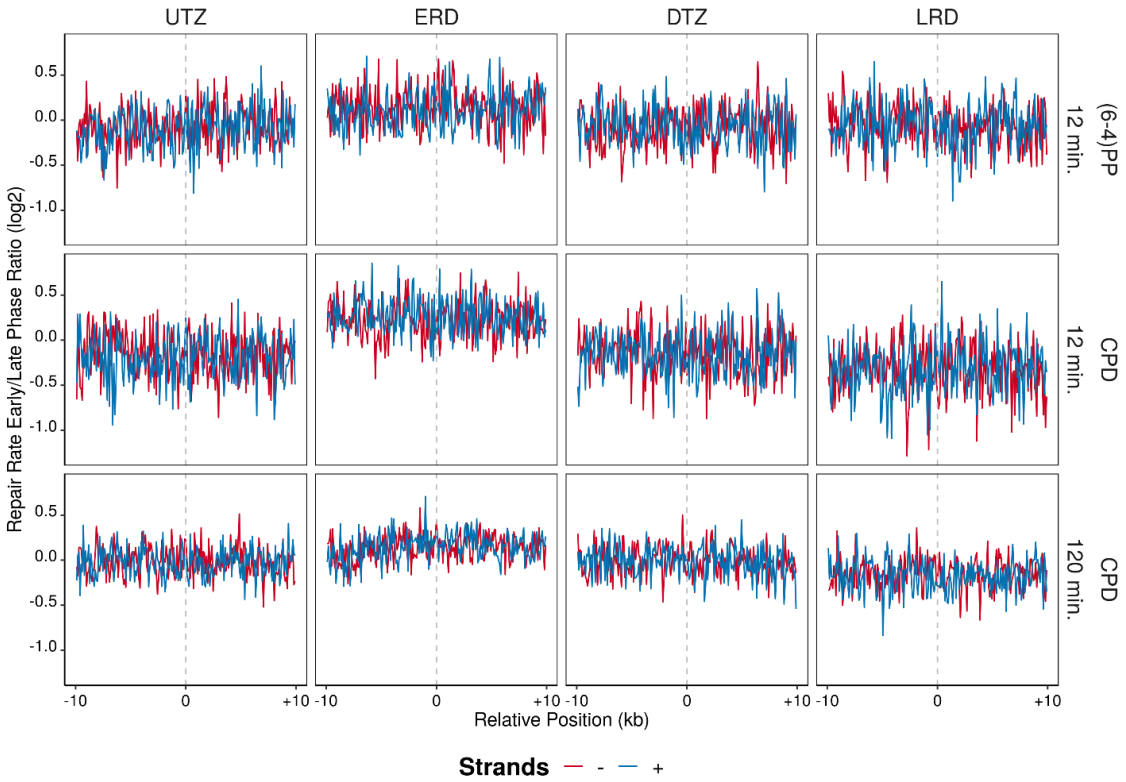

Figure 6.25 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

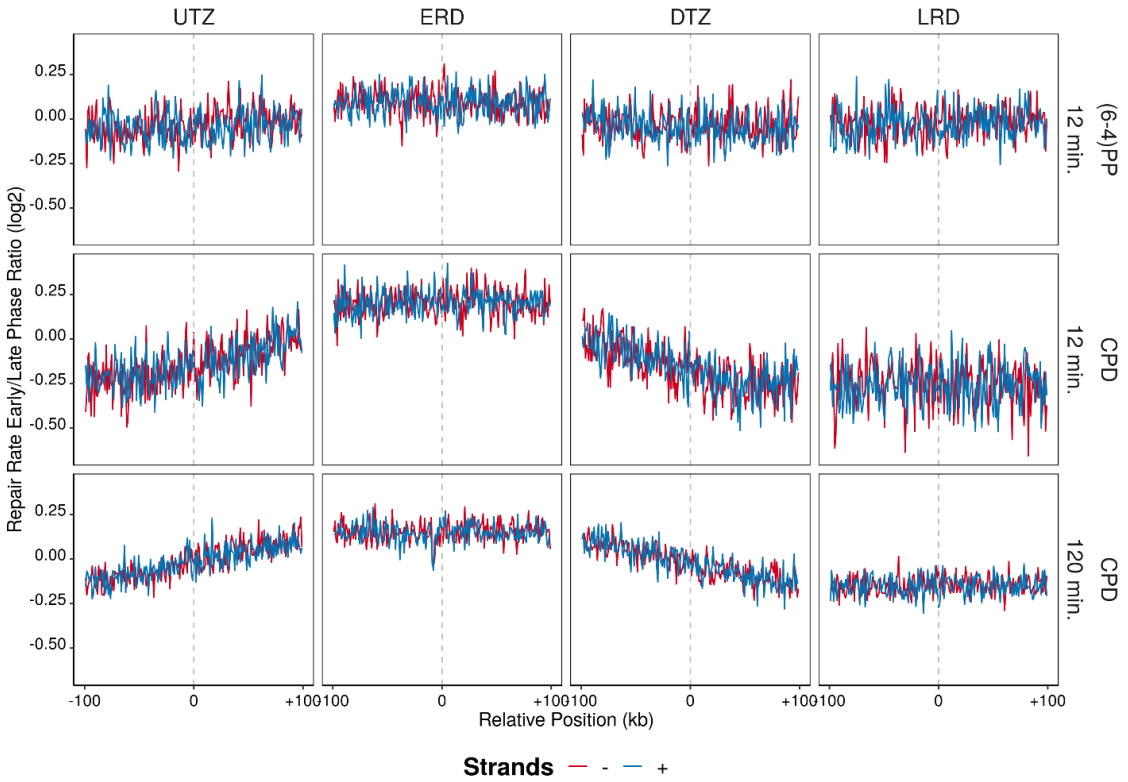

Figure 6.26 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

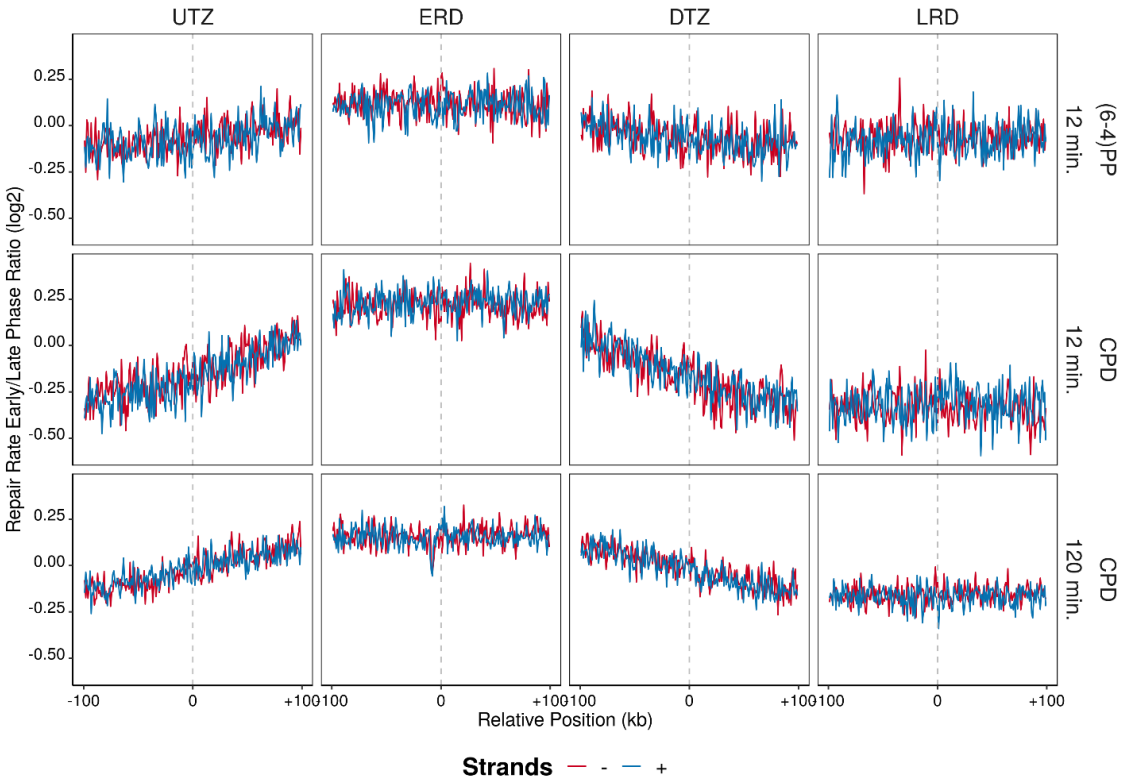

Figure 6.27 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

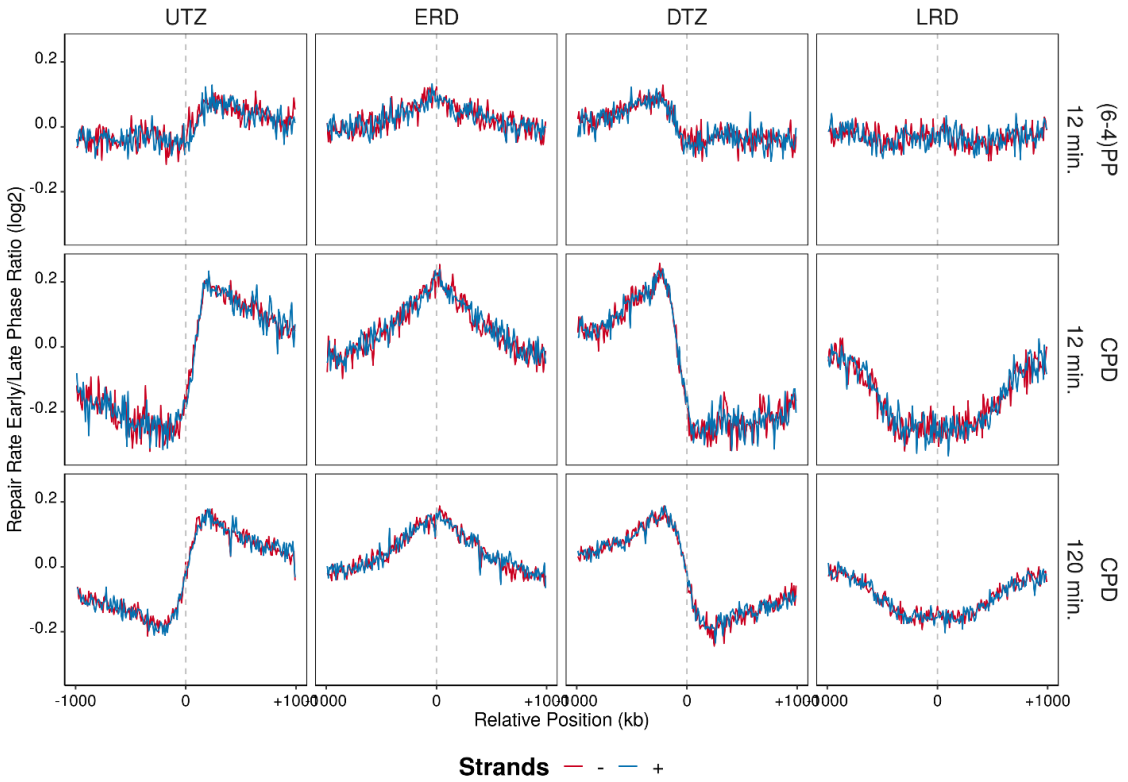

Figure 6.28 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

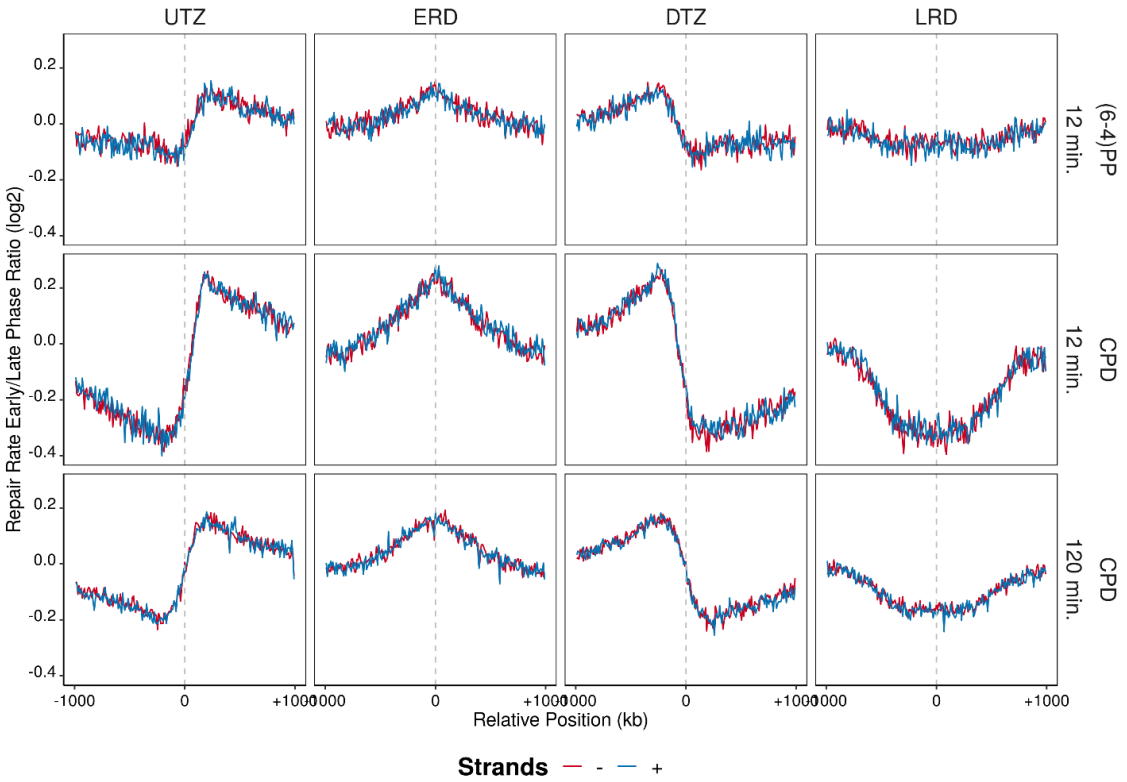

Figure 6.29 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

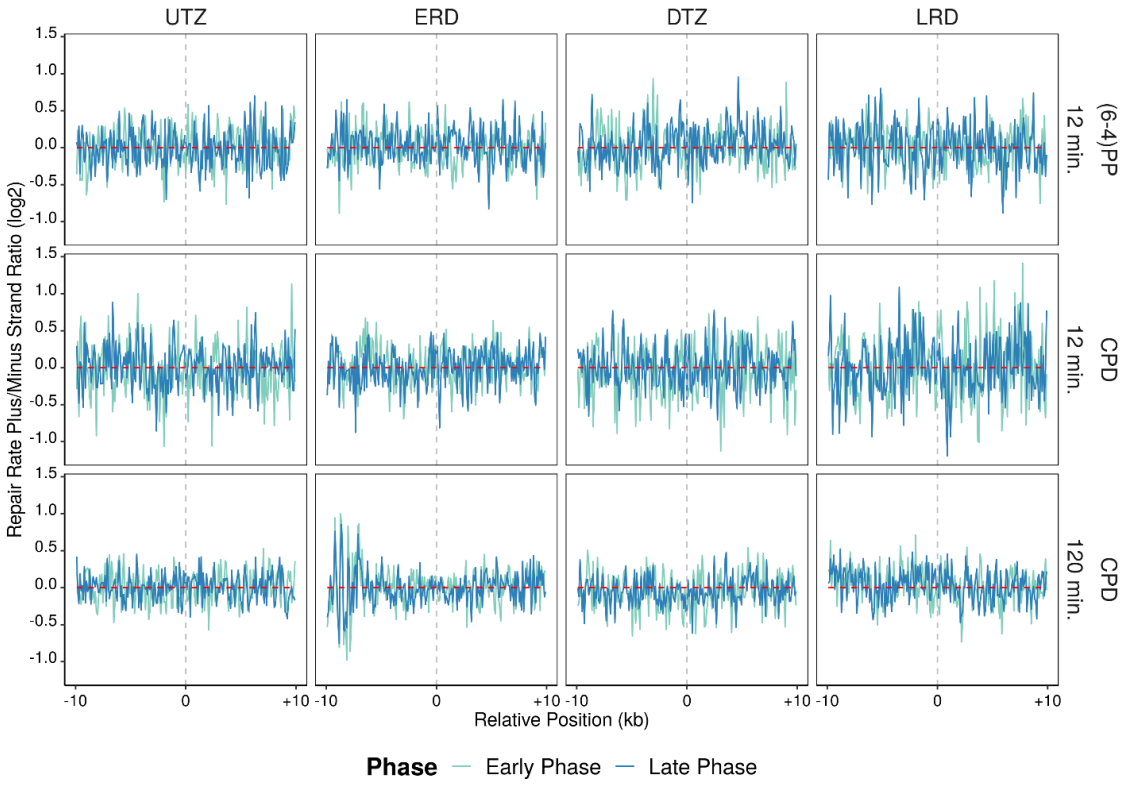

Figure 6.30 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

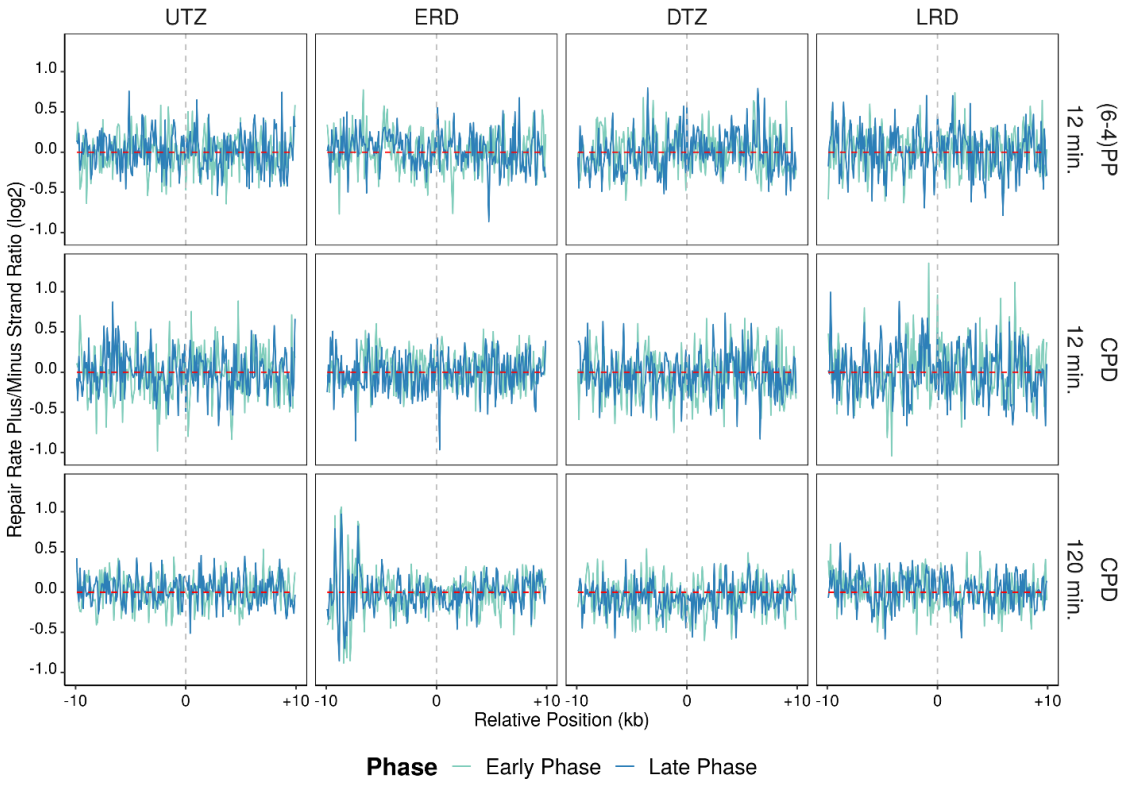

Figure 6.31 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

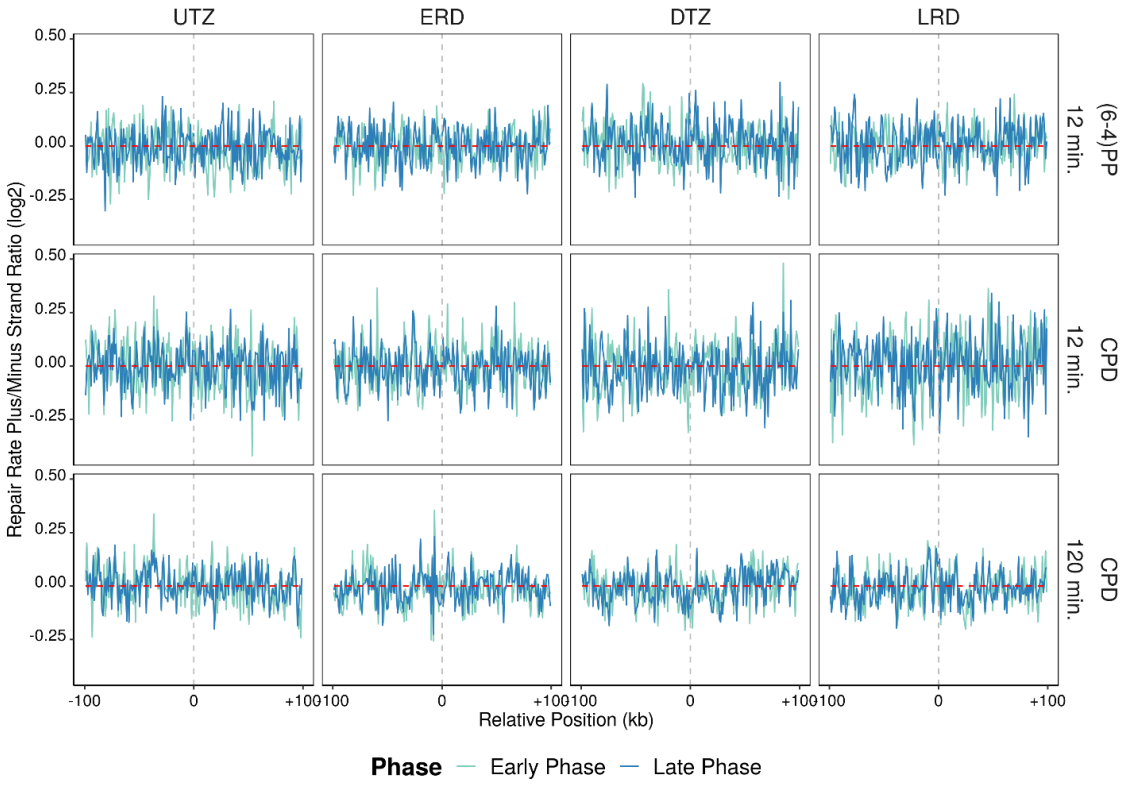

Figure 6.32 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

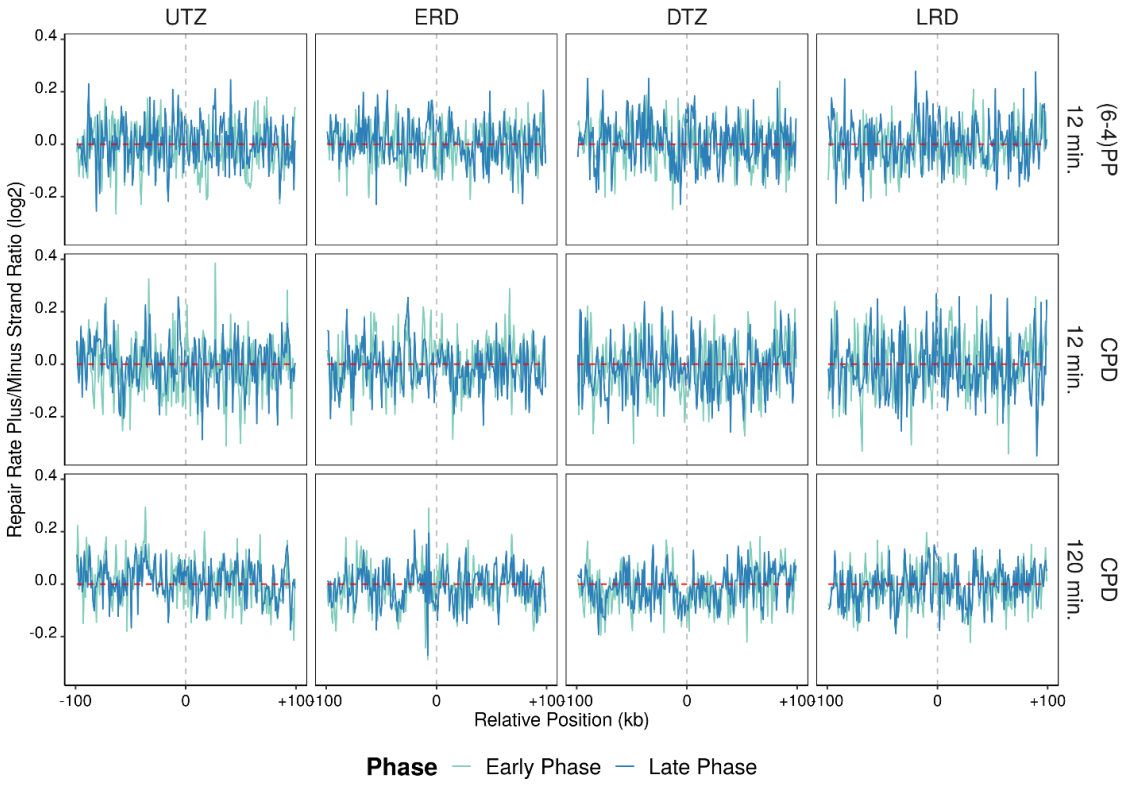

Figure 6.33 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

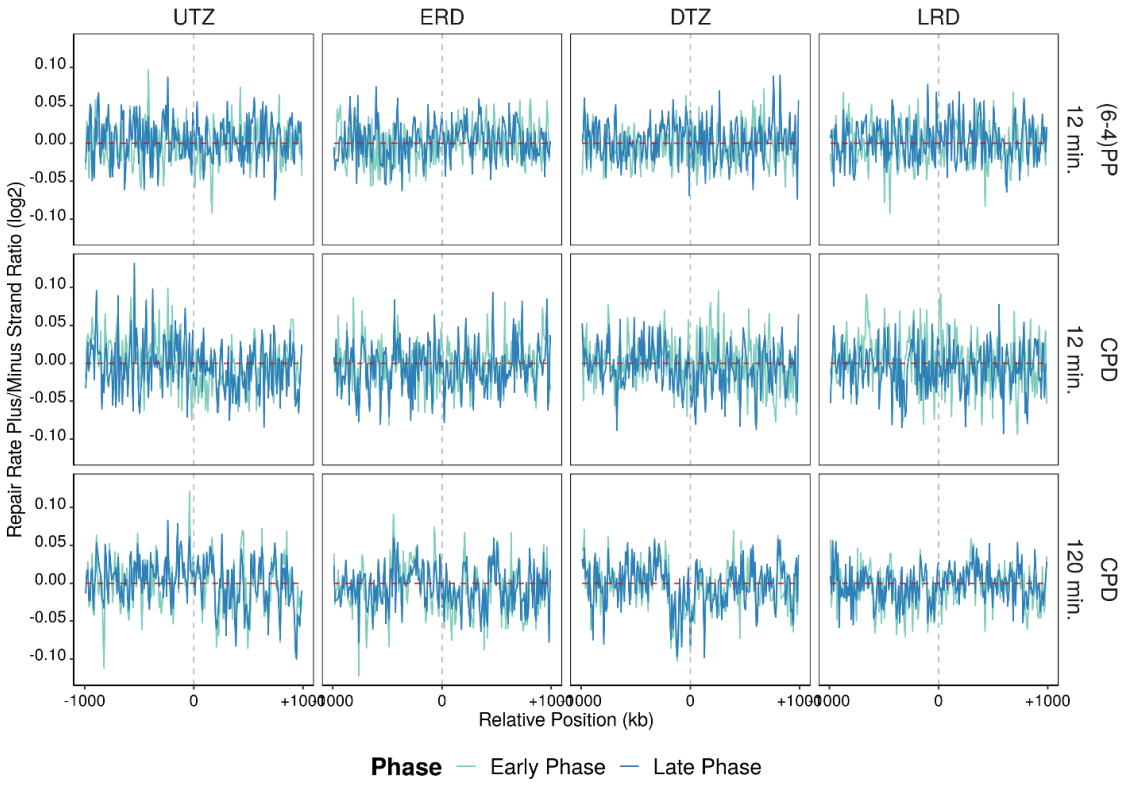

Figure 6.34 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

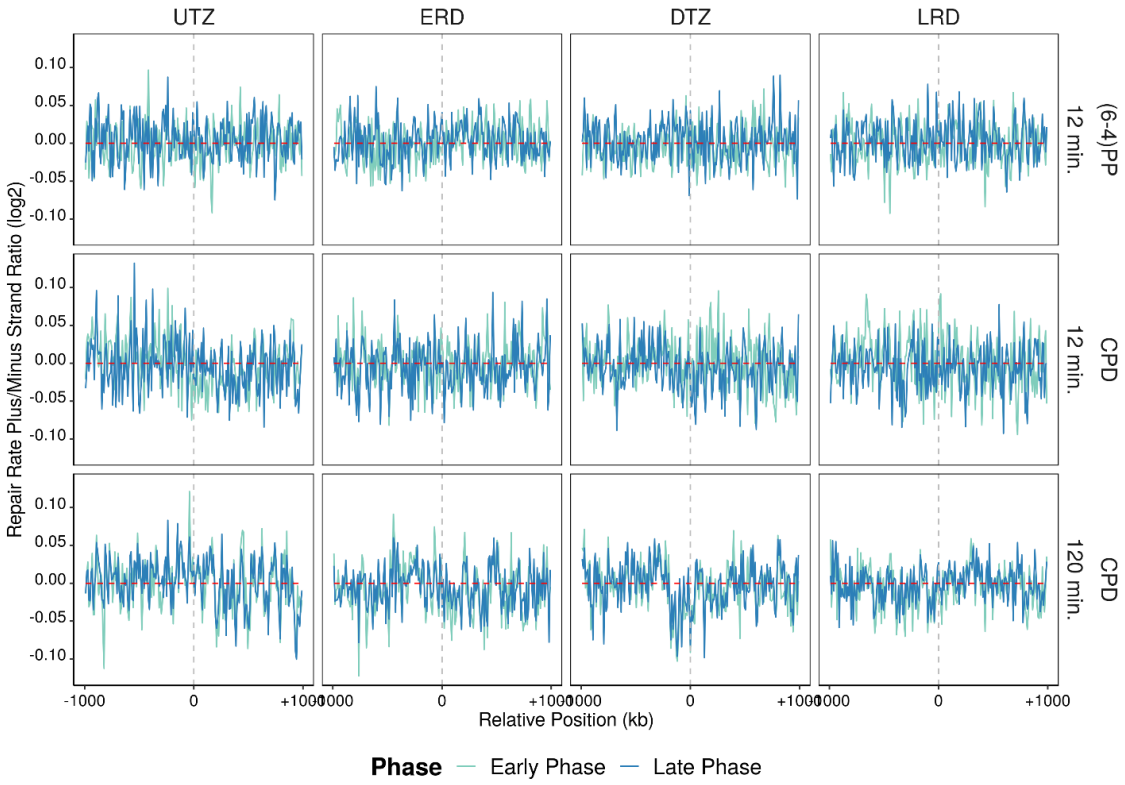

Figure 6.35 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 2 Mb windows with 10 kb intervals, which replication domains are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

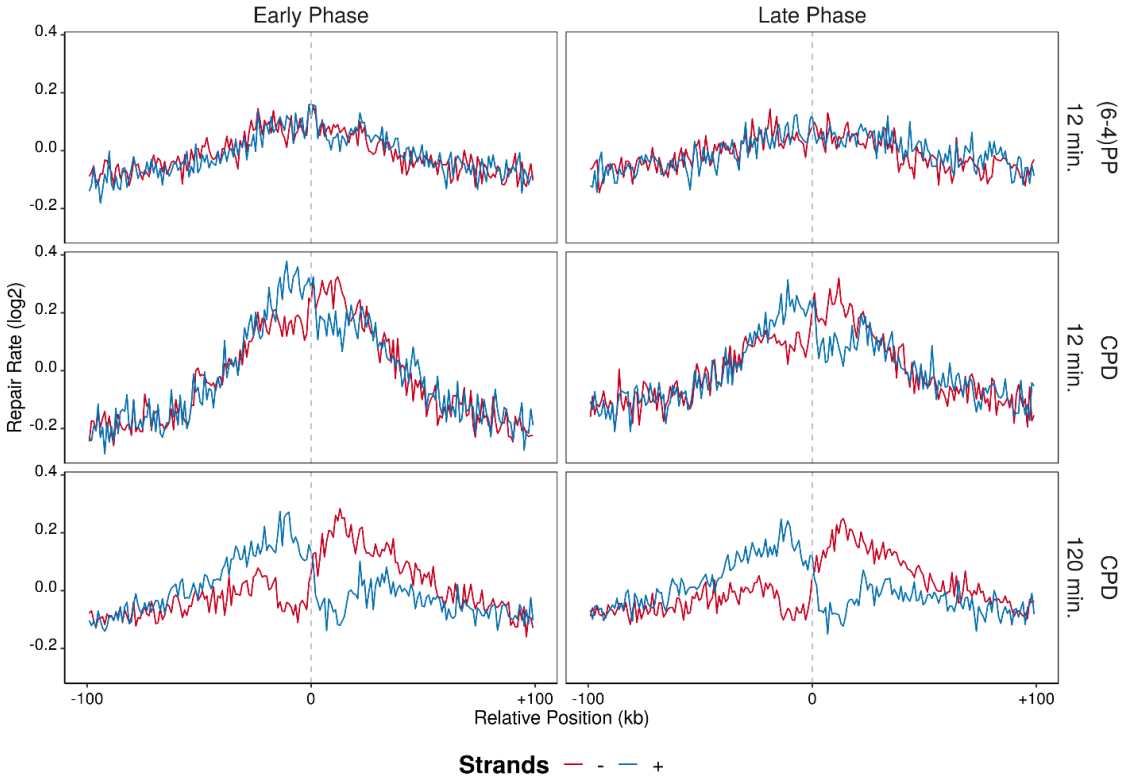

Figure 6.36 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

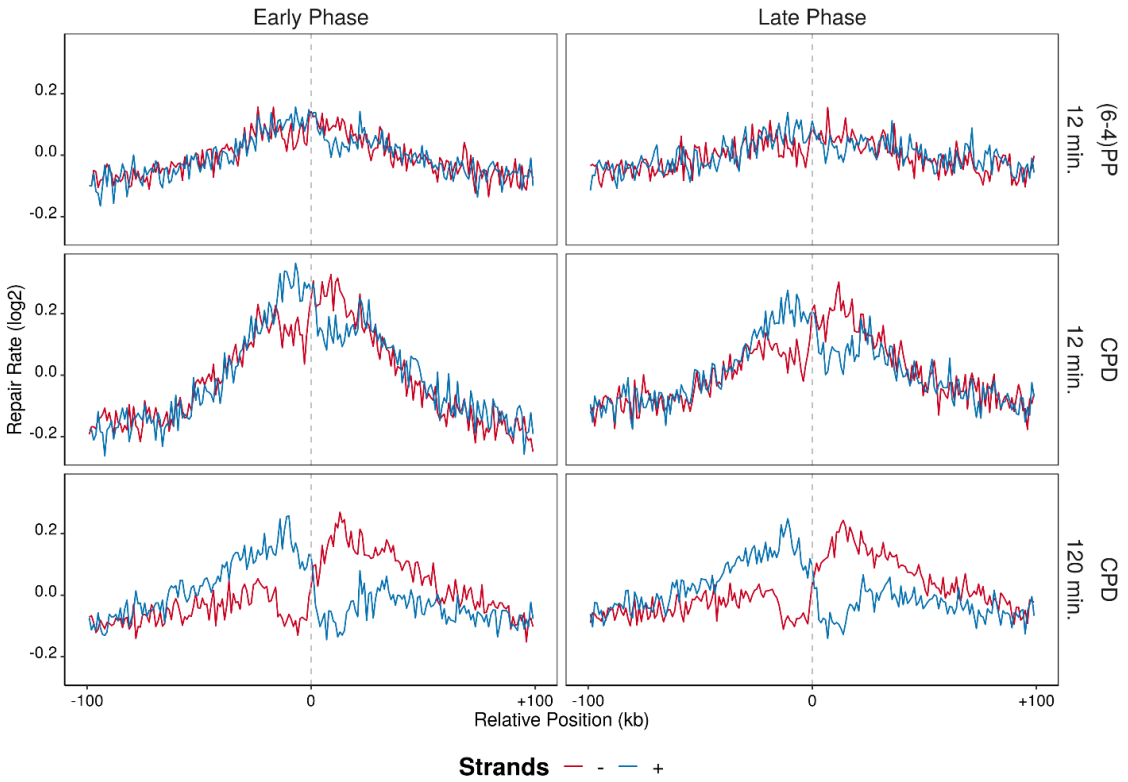

Figure 6.37 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

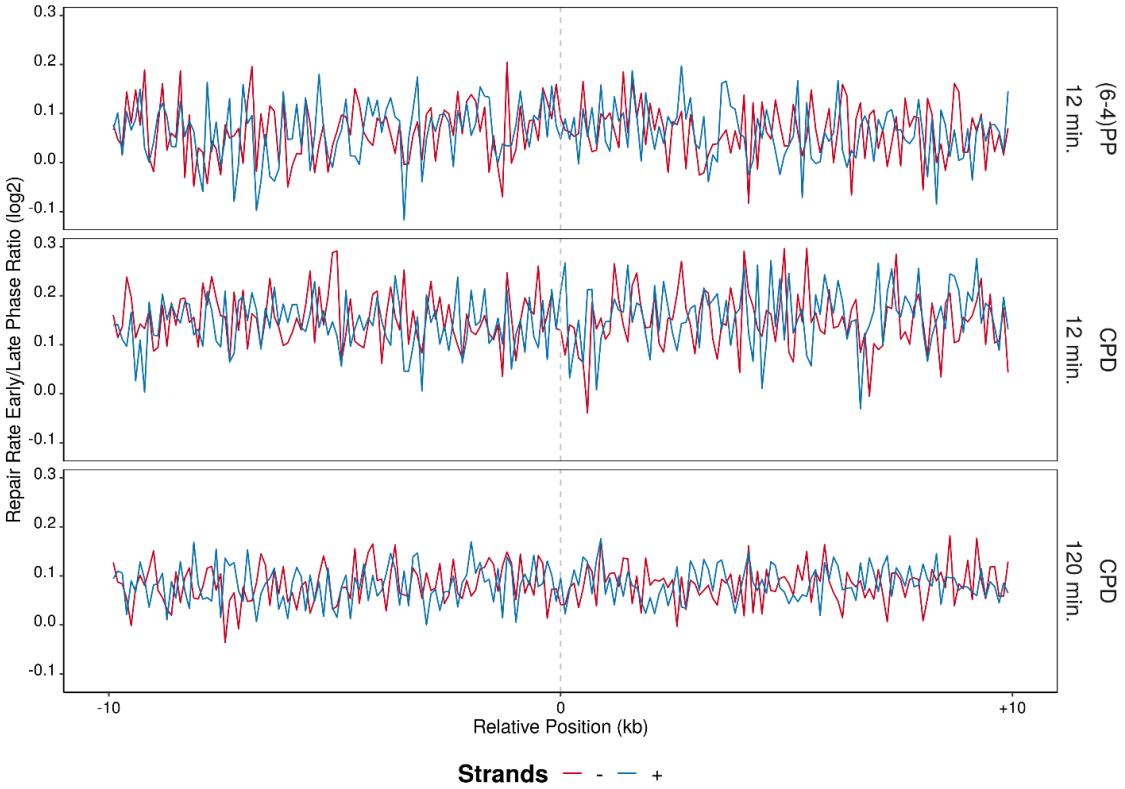

Figure 6.38 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

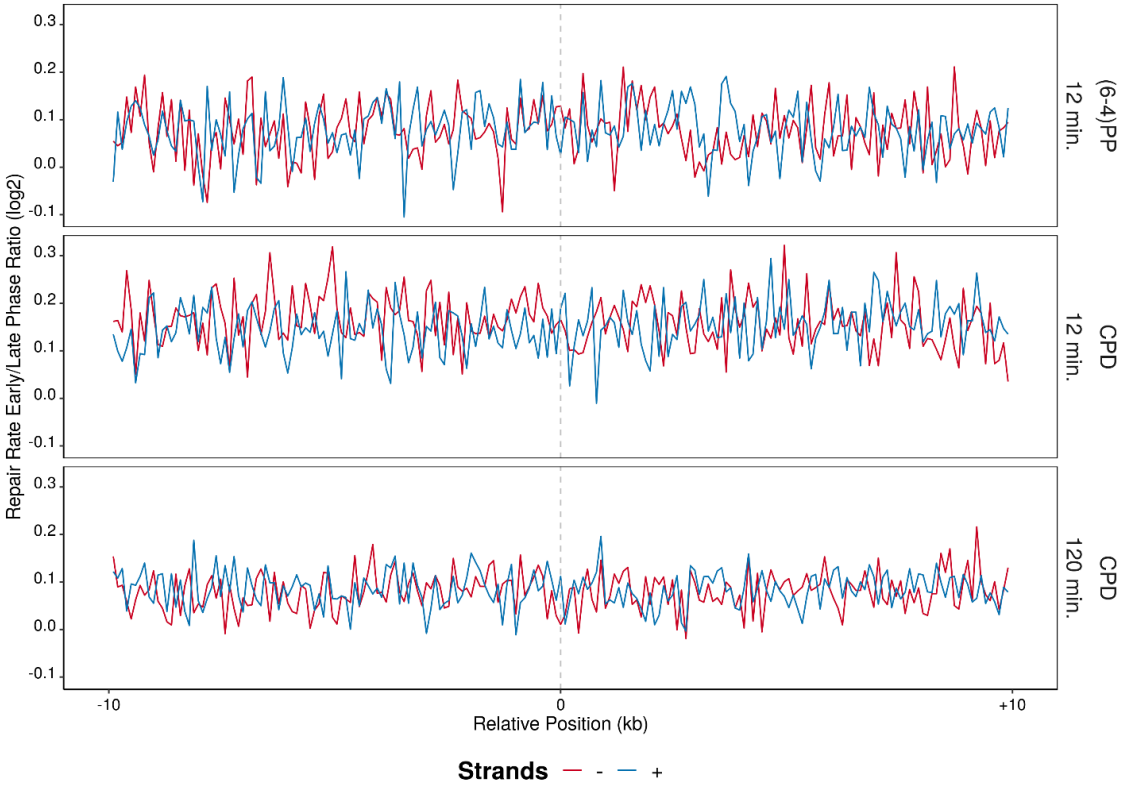

Figure 6.39 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

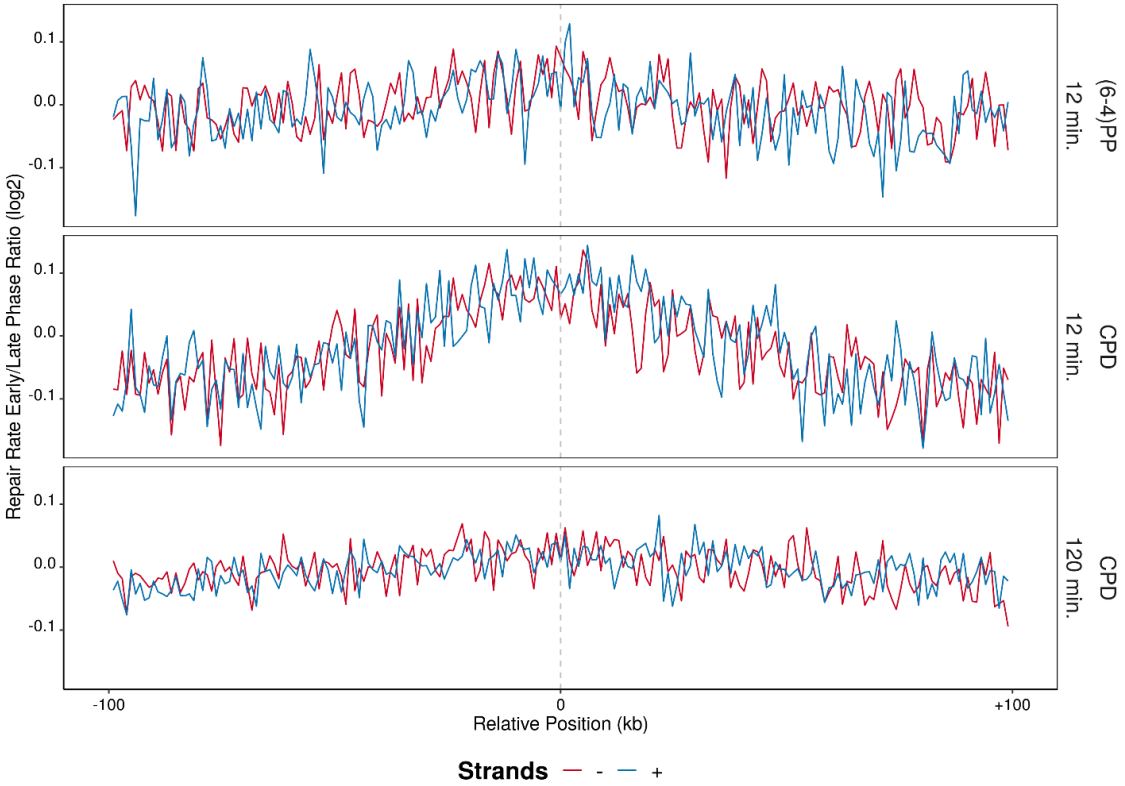

Figure 6.40 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

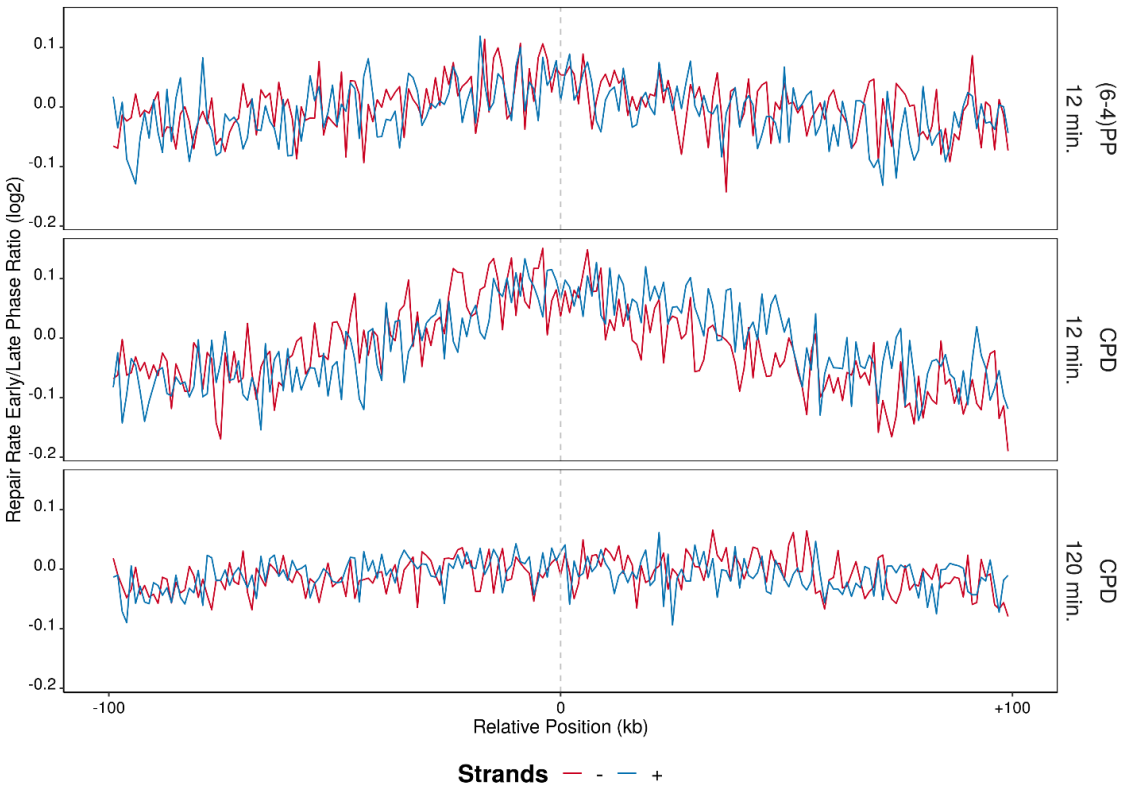

Figure 6.41 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

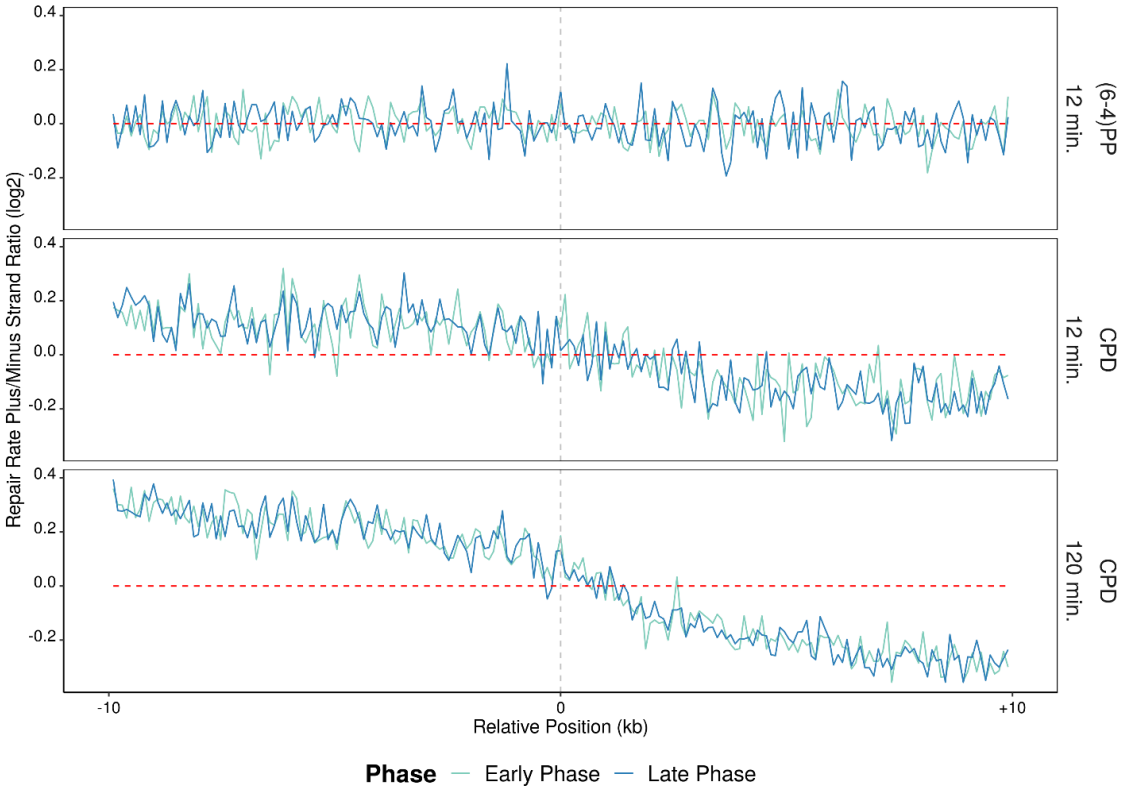

Figure 6.42 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which initiation zones are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

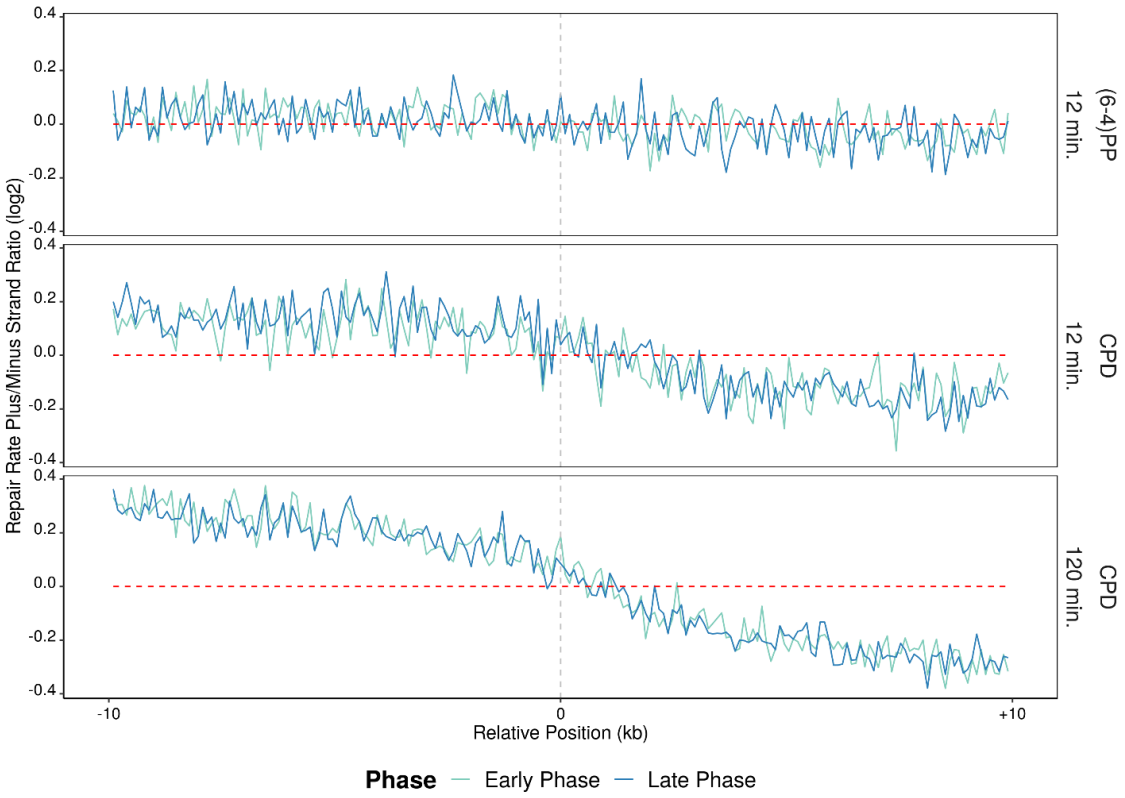

Figure 6.43 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which initiation zones are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

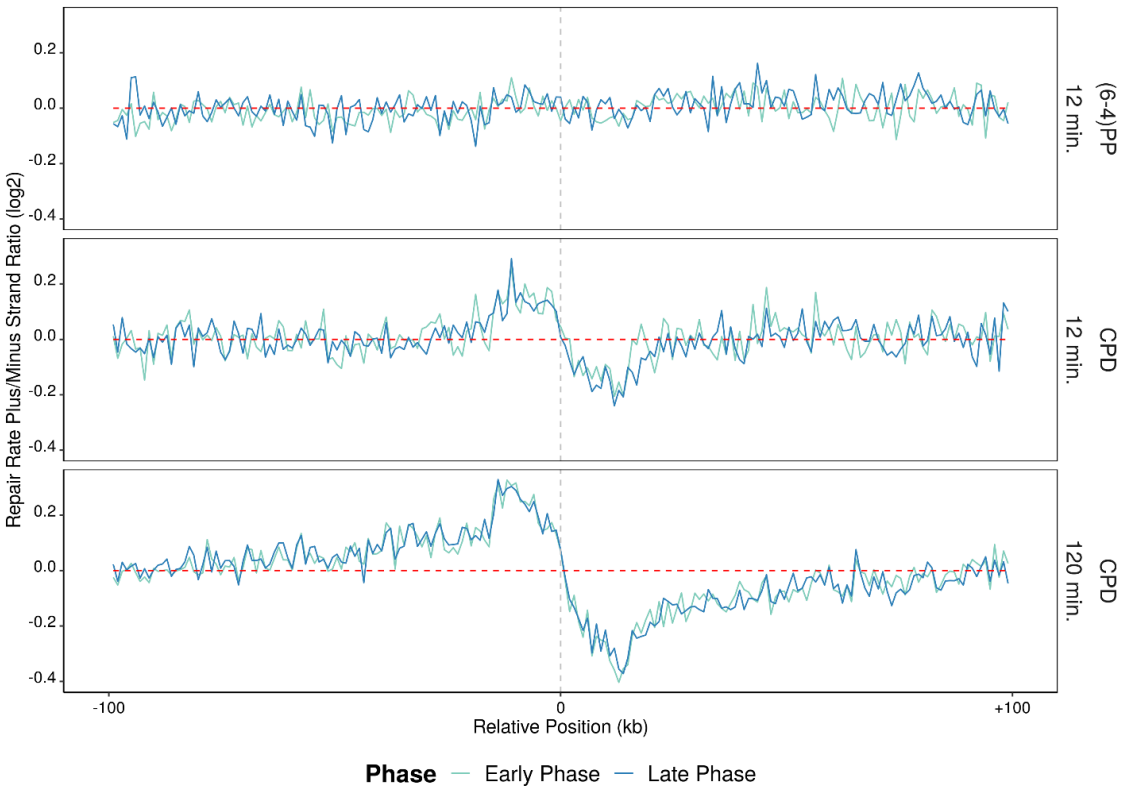

Figure 6.44 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

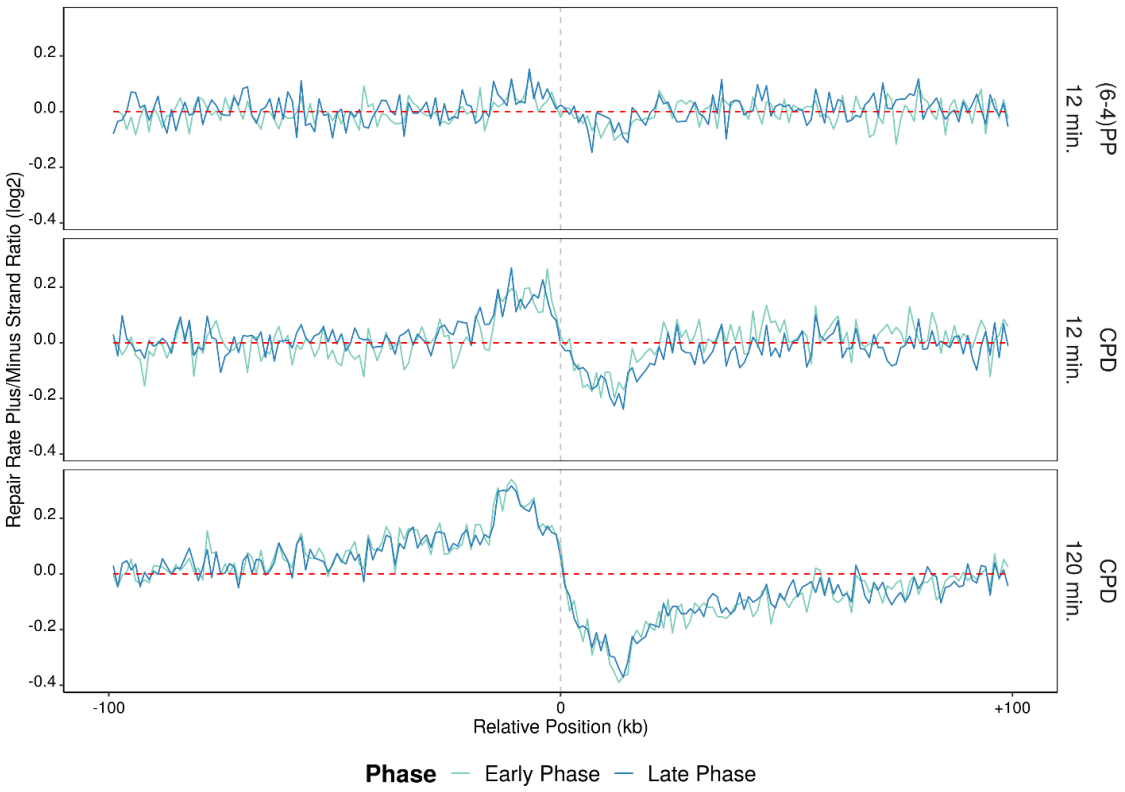

Figure 6.45 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals, which initiation zones are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

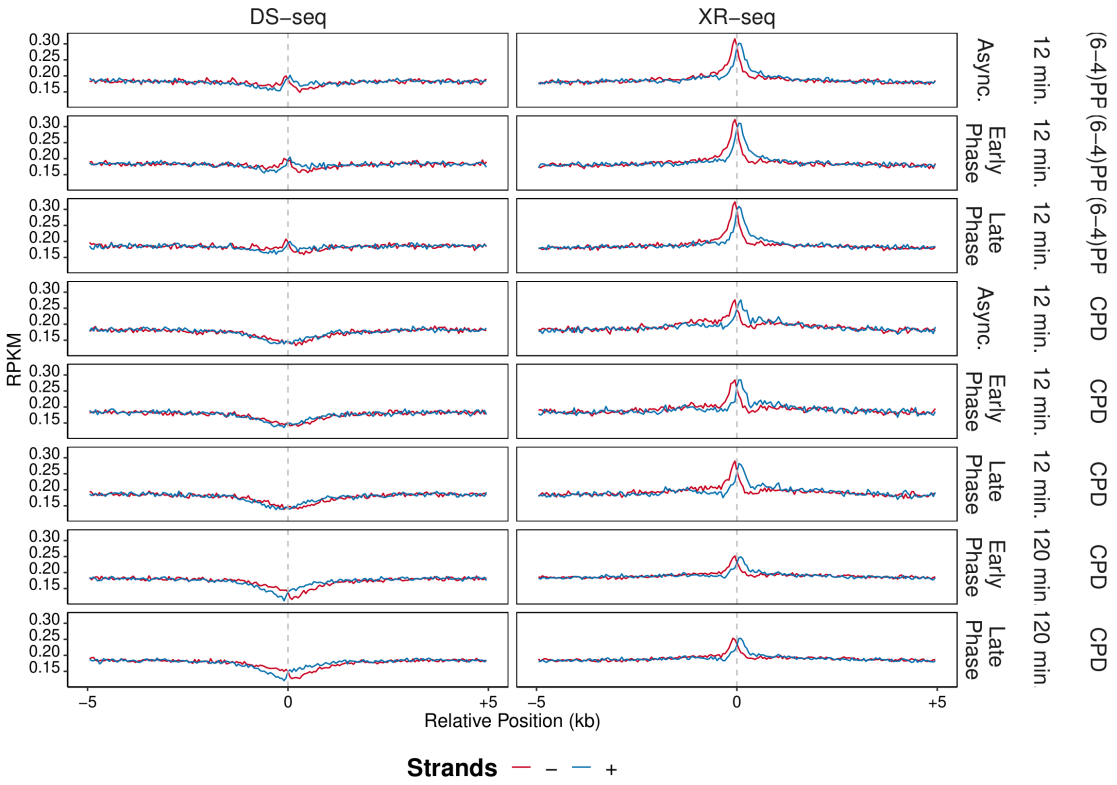

Figure 6.46 RPKM values of Damage-seq samples (left) and XR-seq samples (right) are calculated in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region. Analysis is performed on replicate A.

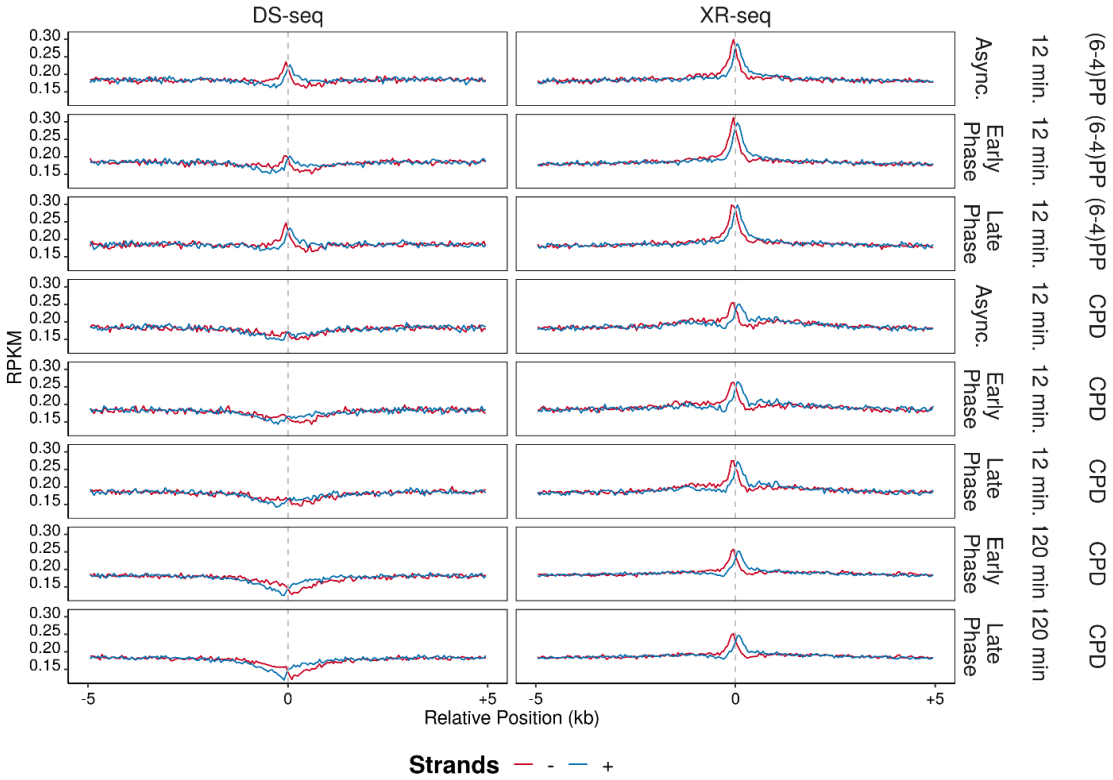

Figure 6.47 RPKM values of Damage-seq samples (left) and XR-seq samples (right) are calculated in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region. Analysis is performed on replicate B.

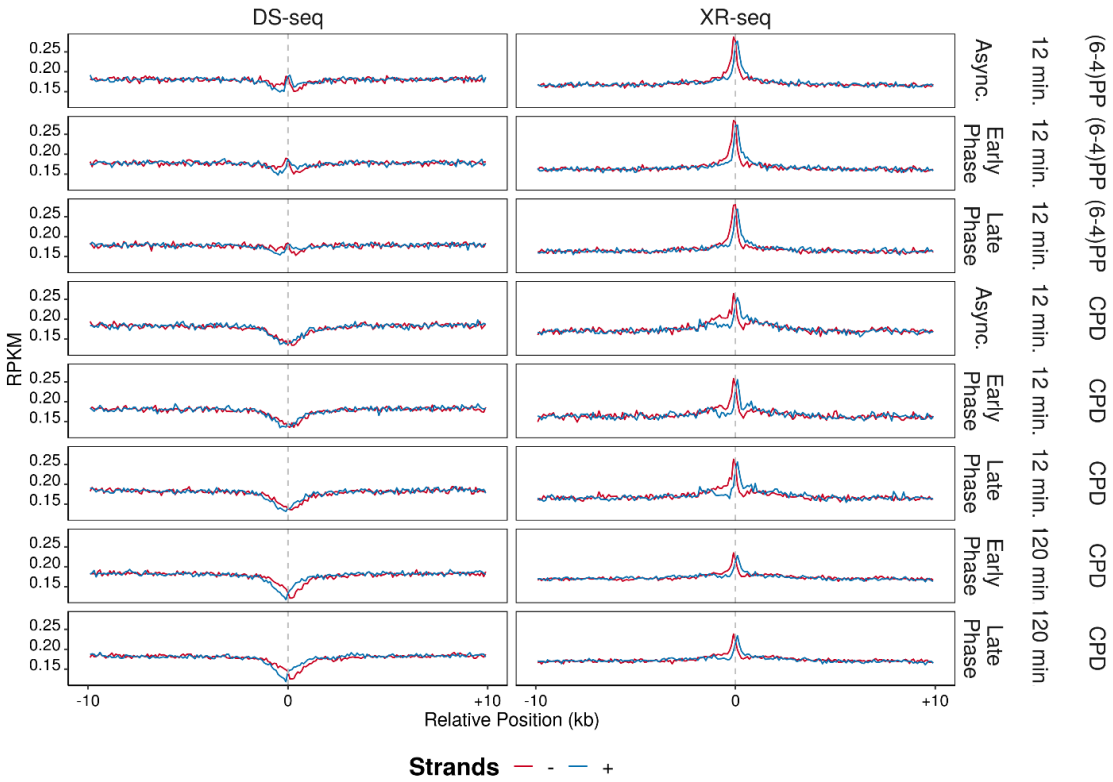

Figure 6.48 RPKM values of Damage-seq samples (left) and XR-seq samples (right) are calculated in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region. Analysis is performed on replicate A.

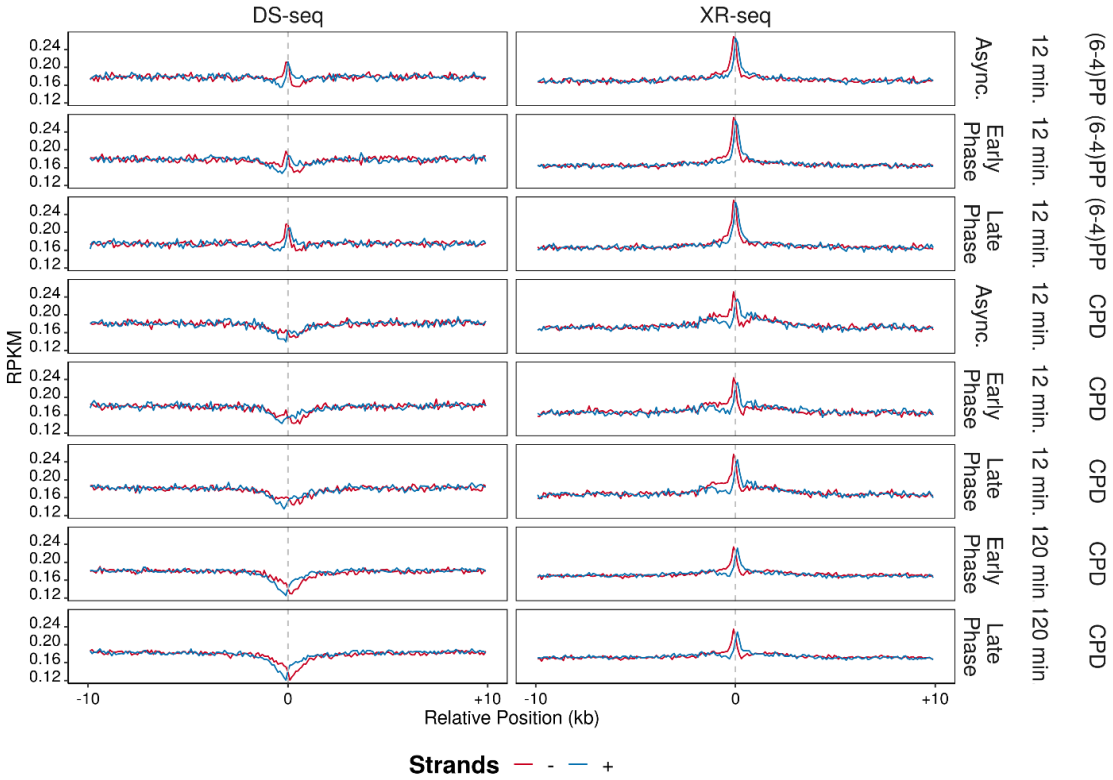

Figure 6.49 RPKM values of Damage-seq samples (left) and XR-seq samples (right) are calculated in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Blue lines represent the plus strands and red lines represent the minus strands. Gray vertical dashed line shows the center of the region. Analysis is performed on replicate B.

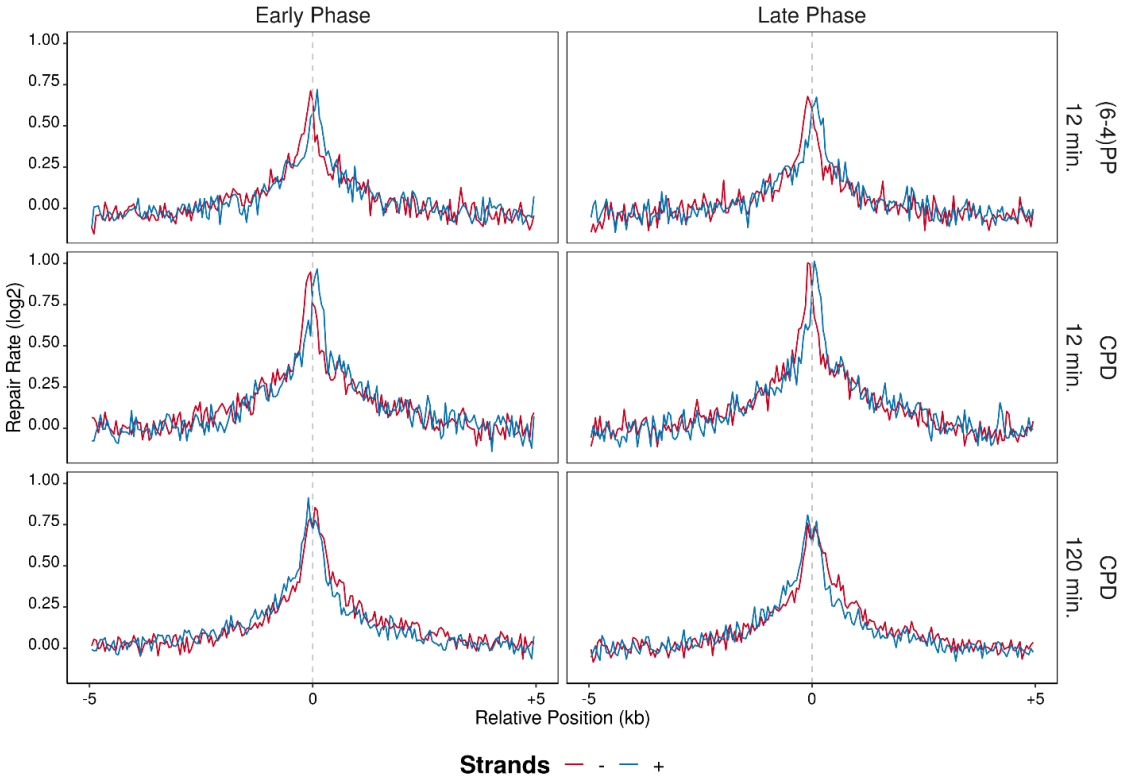

Figure 6.50 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

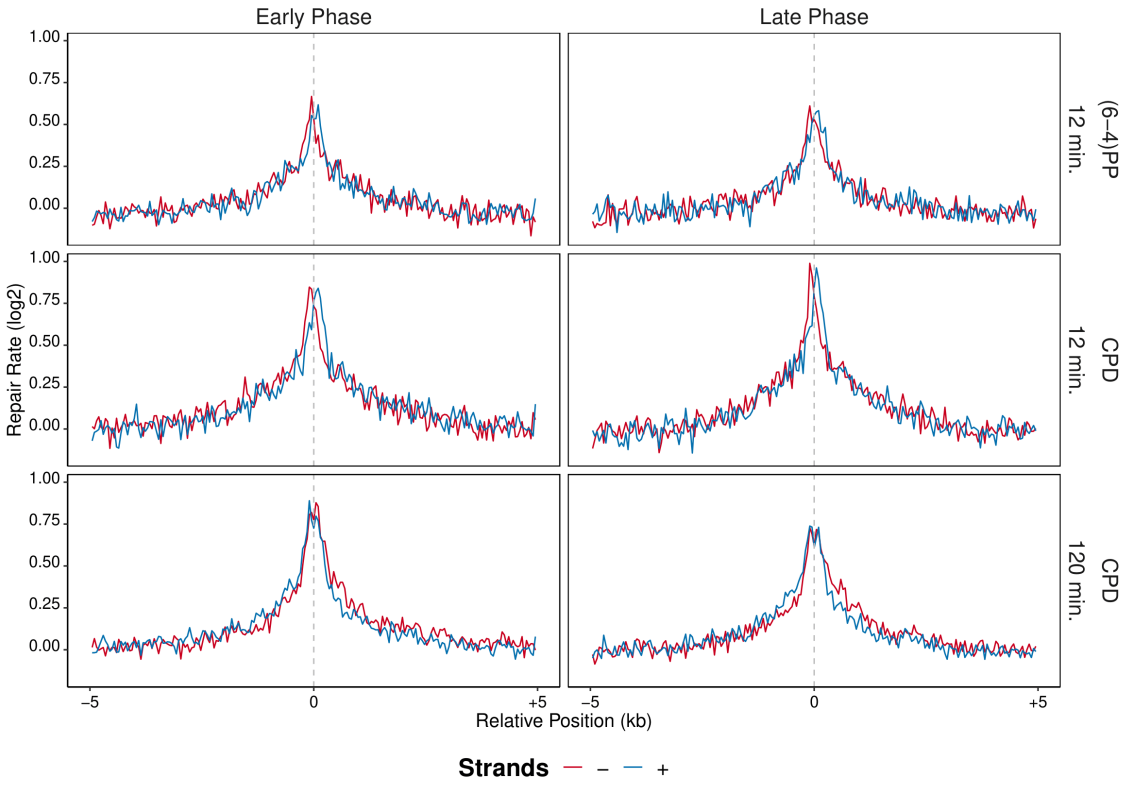

Figure 6.51 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

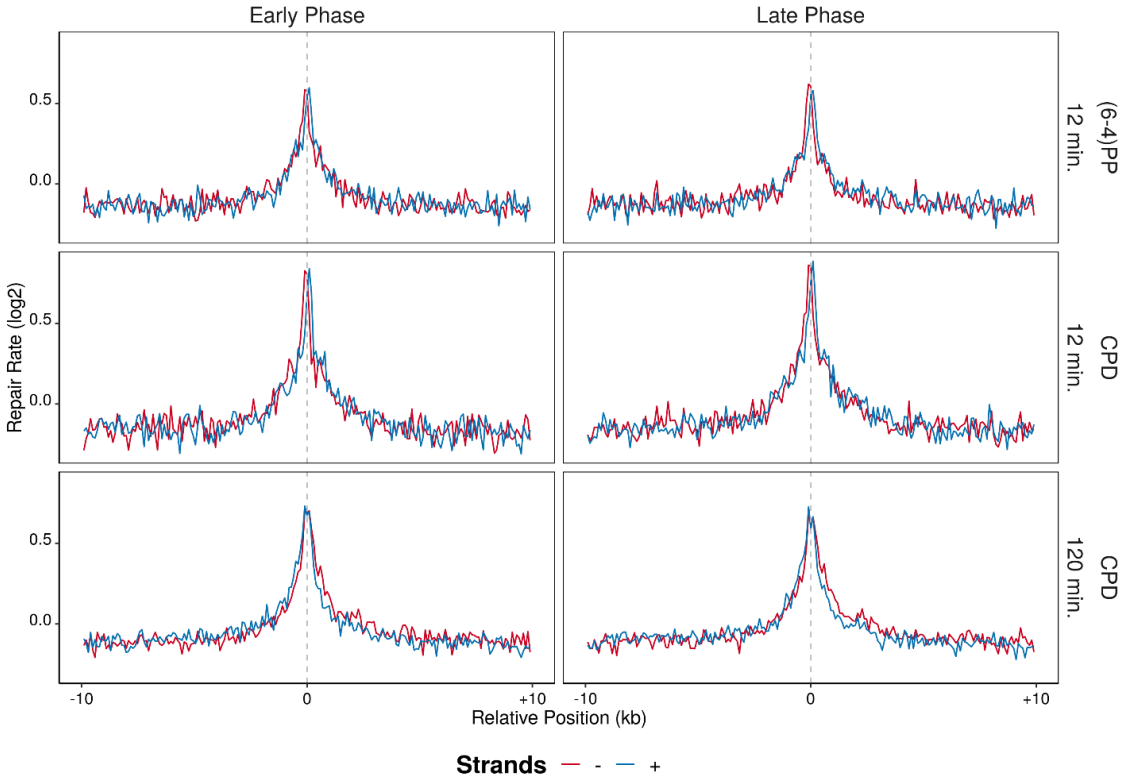

Figure 6.52 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

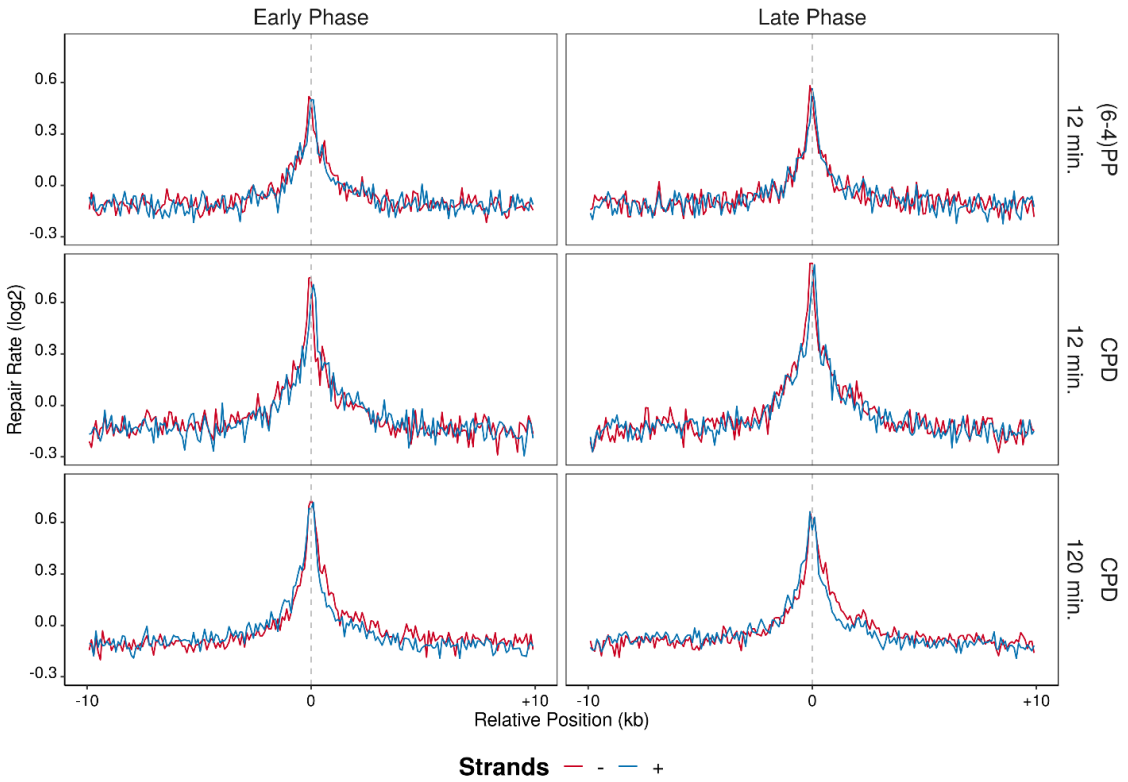

Figure 6.53 Repair rates (XR-seq/Damage-seq) are calculated and log2 transformed in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

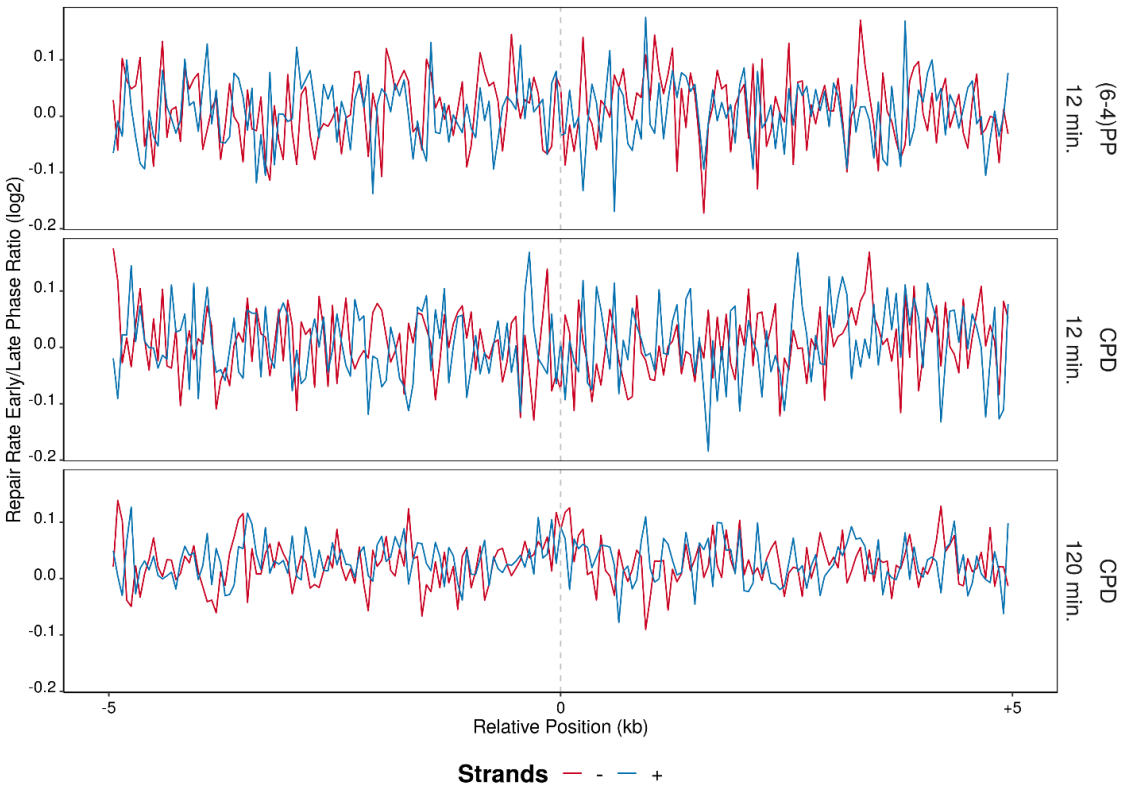

Figure 6.54 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

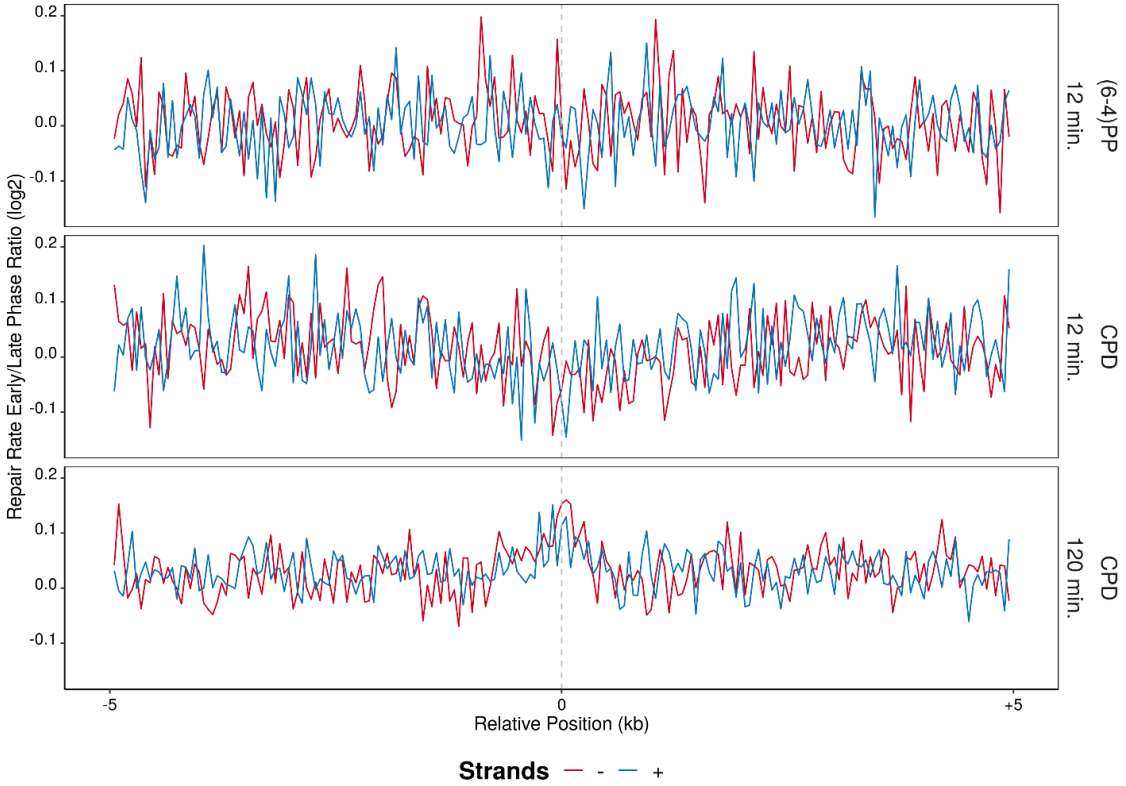

Figure 6.55 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

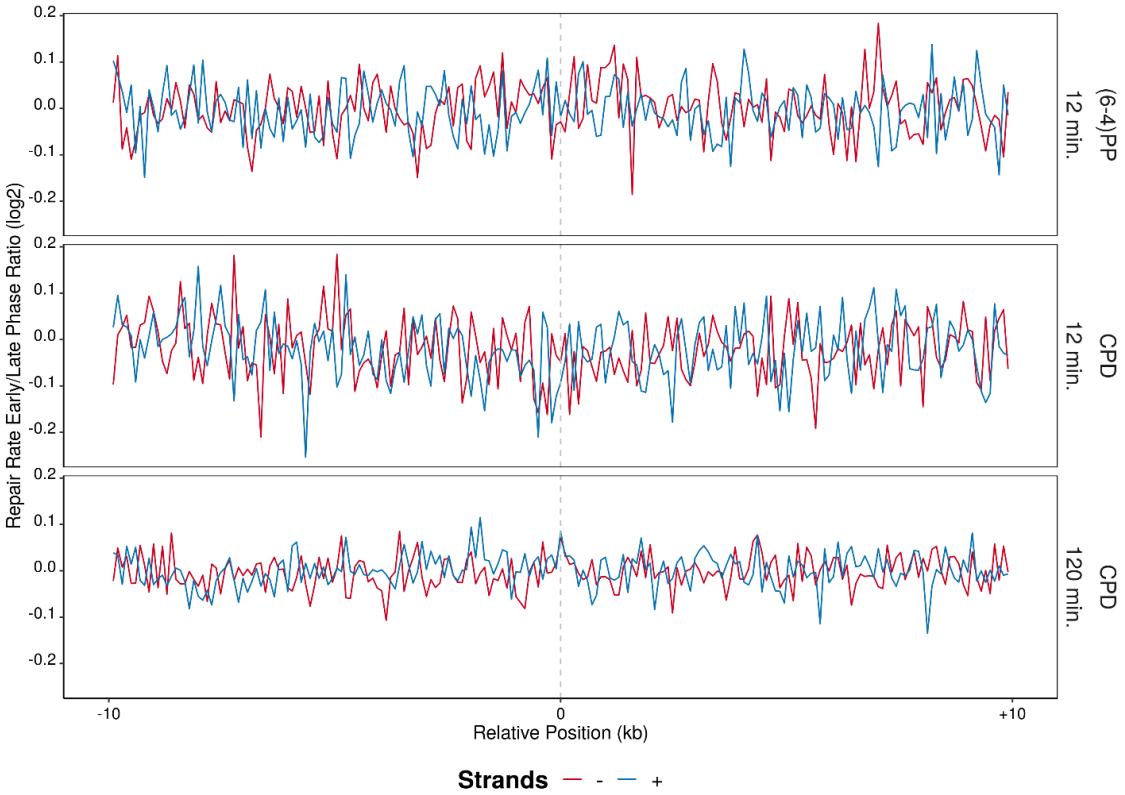

Figure 6.56 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

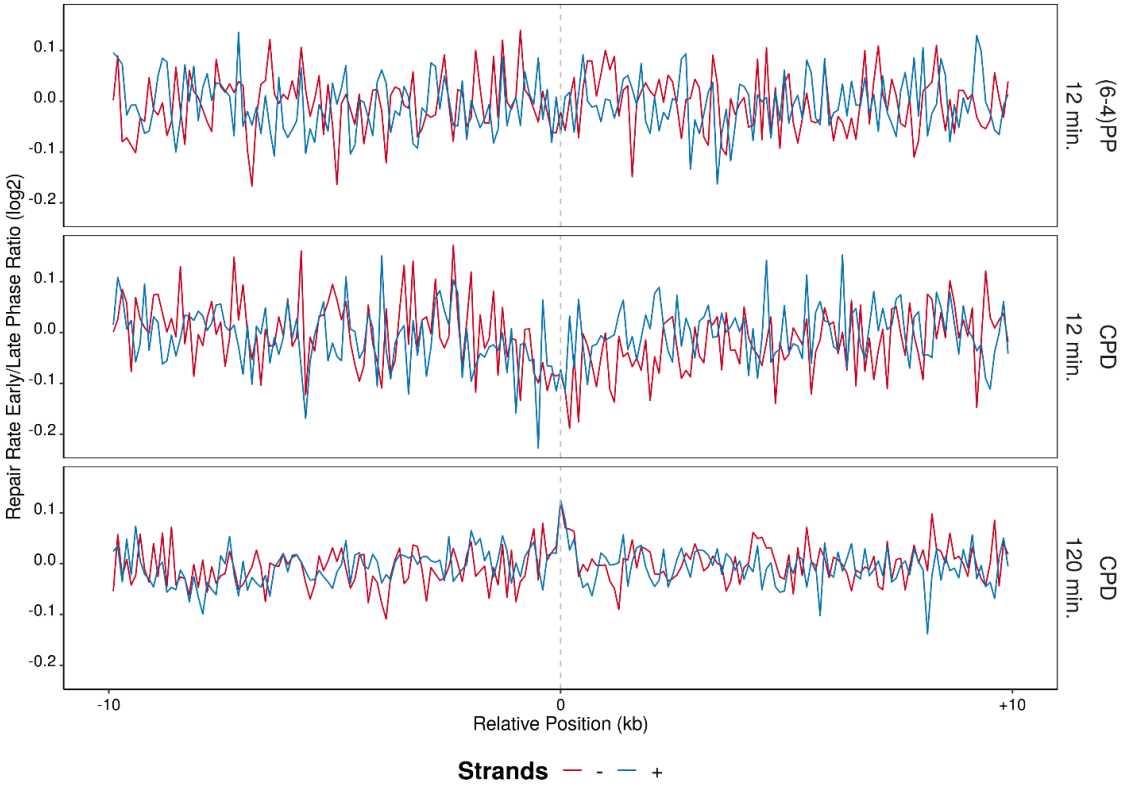

Figure 6.57 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

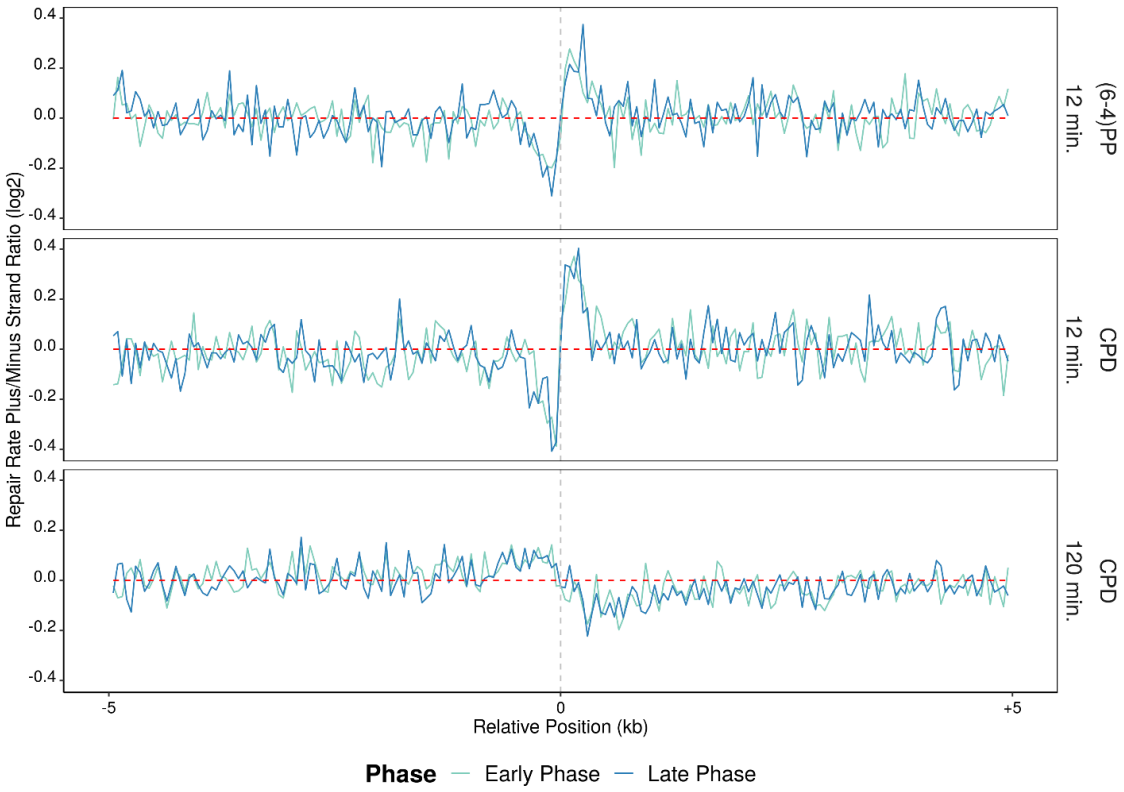

Figure 6.58 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

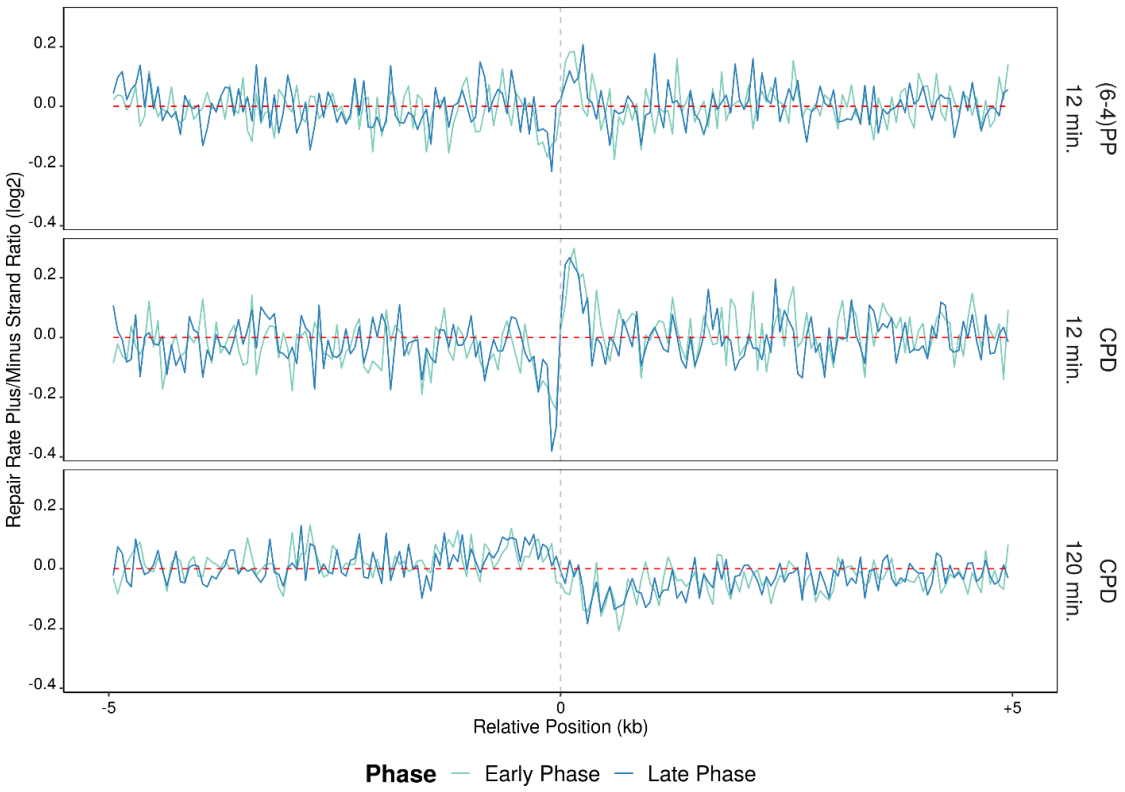

Figure 6.59 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 10 kb windows with 50 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

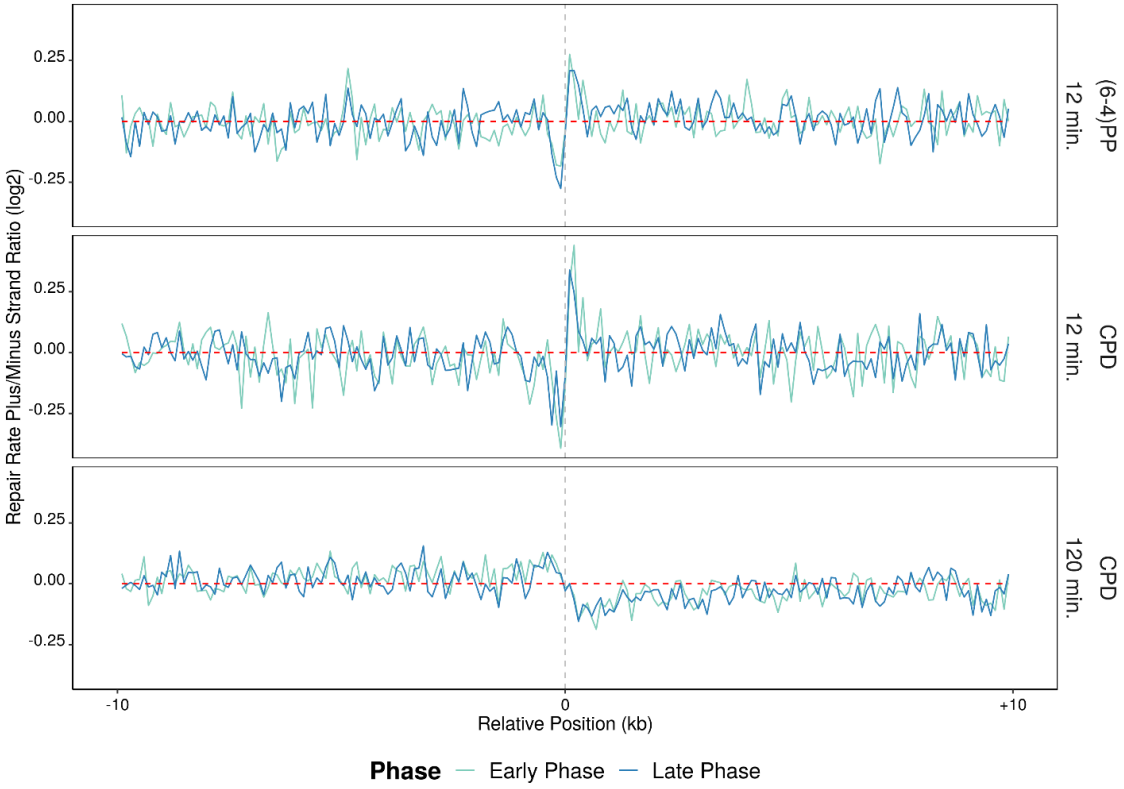

Figure 6.60 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate A.

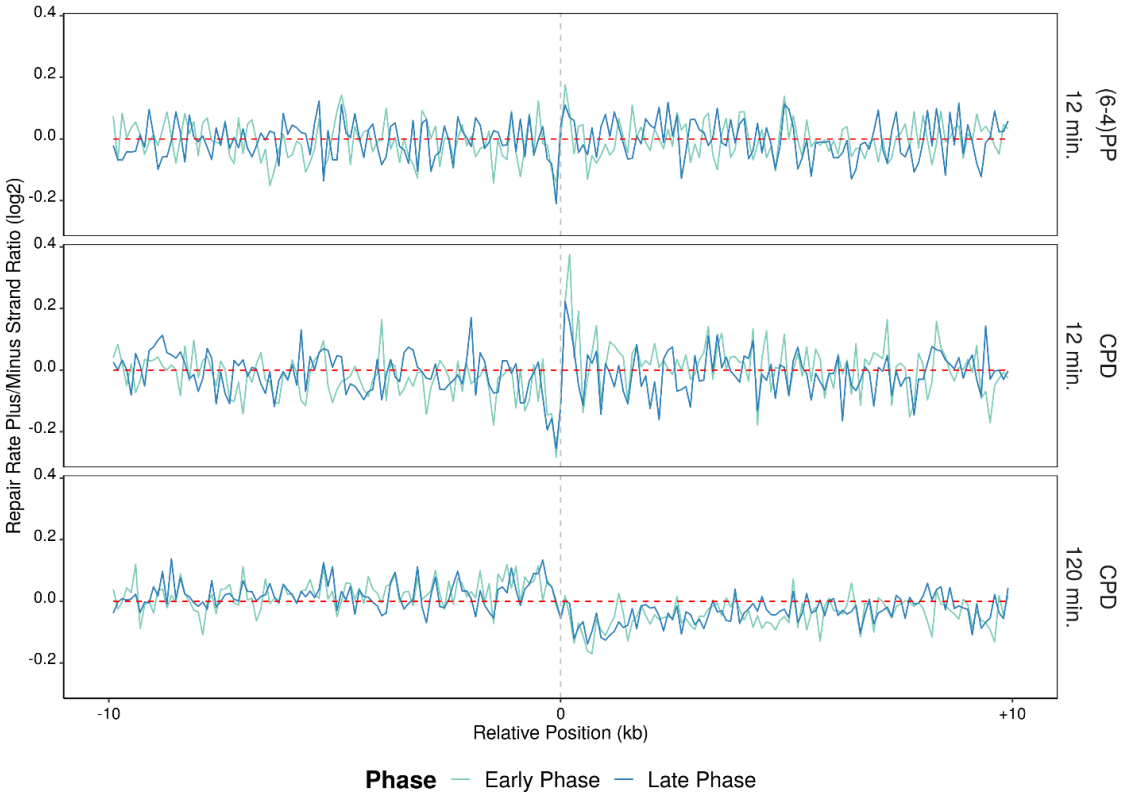

Figure 6.61 After log2 transformed repair rates (XR-seq/Damage-seq) are calculated, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals, which replication origins (SNS-seq) are positioned at the center of the region. Light blue lines are the early phase samples and dark blue lines are the late phase samples. Analysis is performed on replicate B.

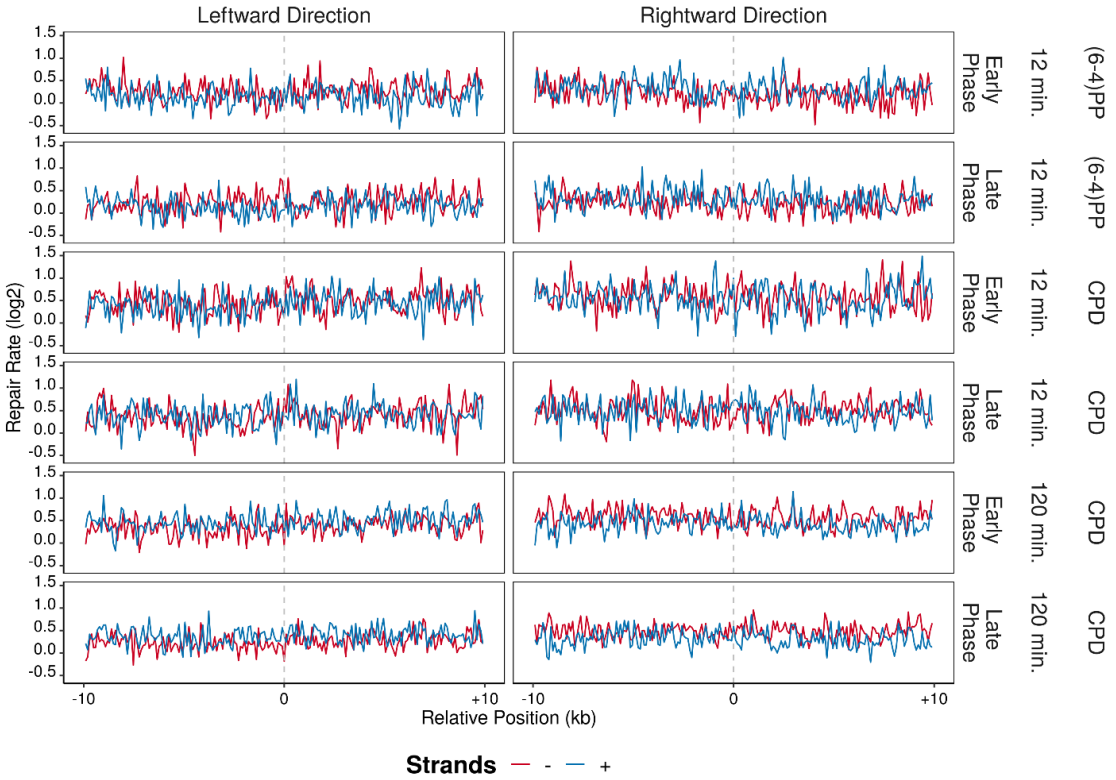

Figure 6.62 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

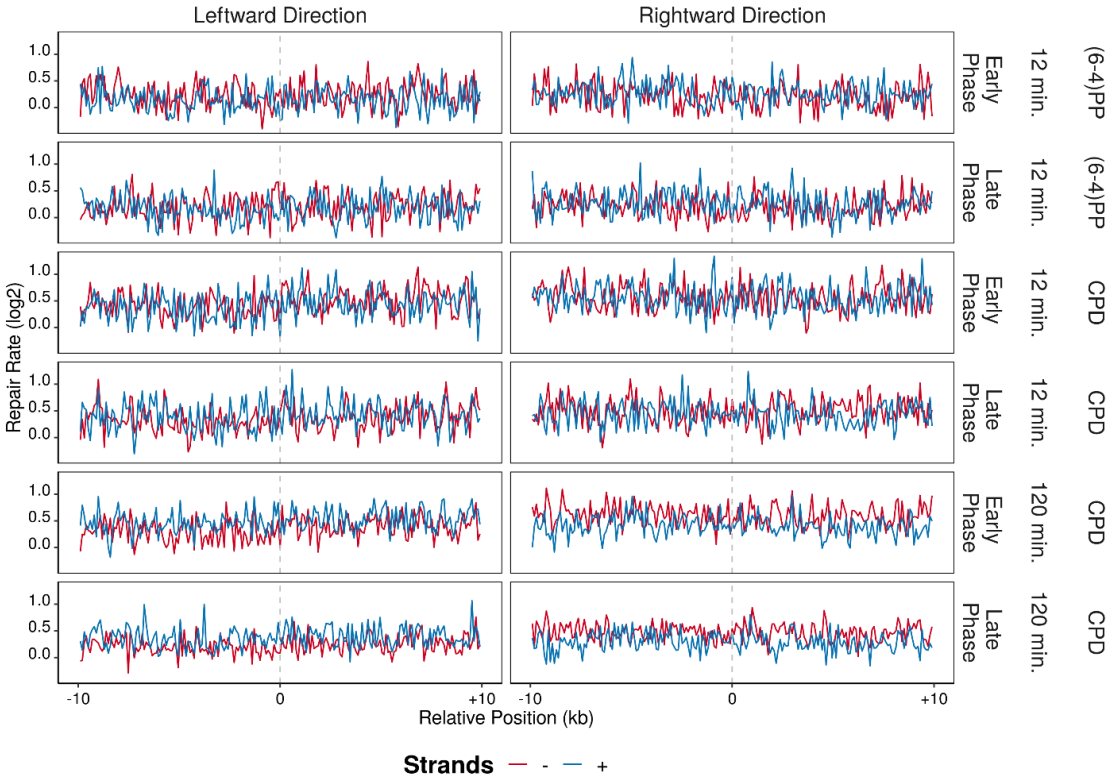

Figure 6.63 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

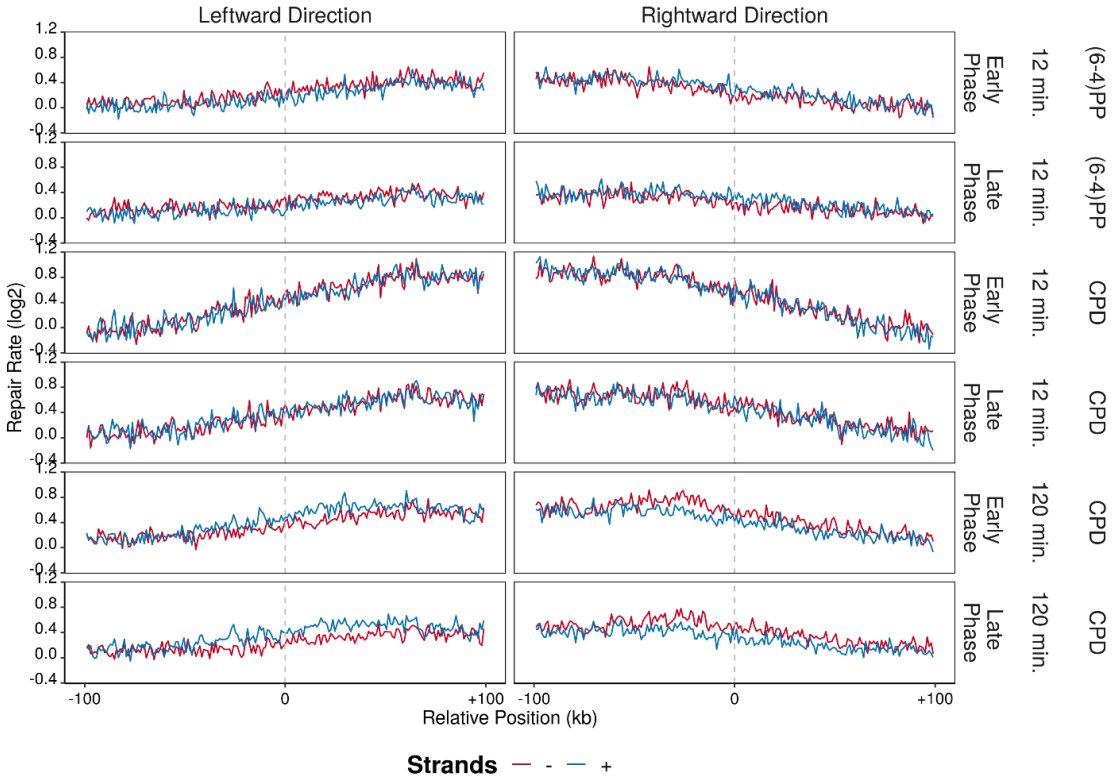

Figure 6.64 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

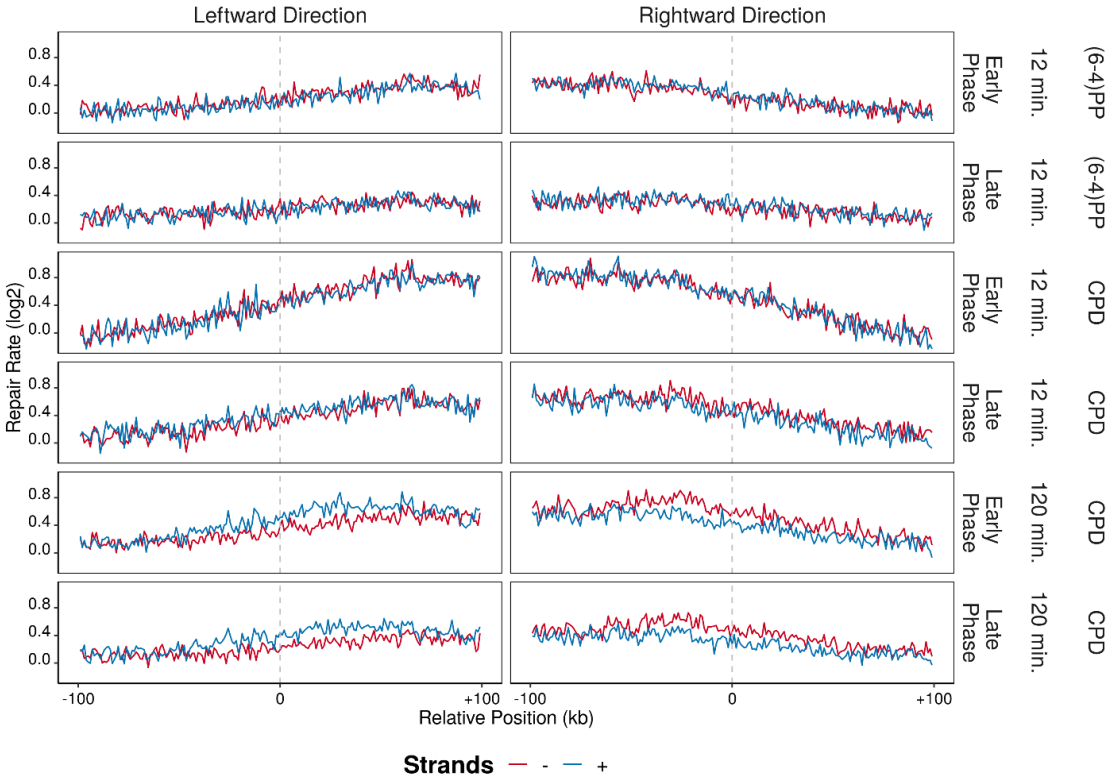

Figure 6.65 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

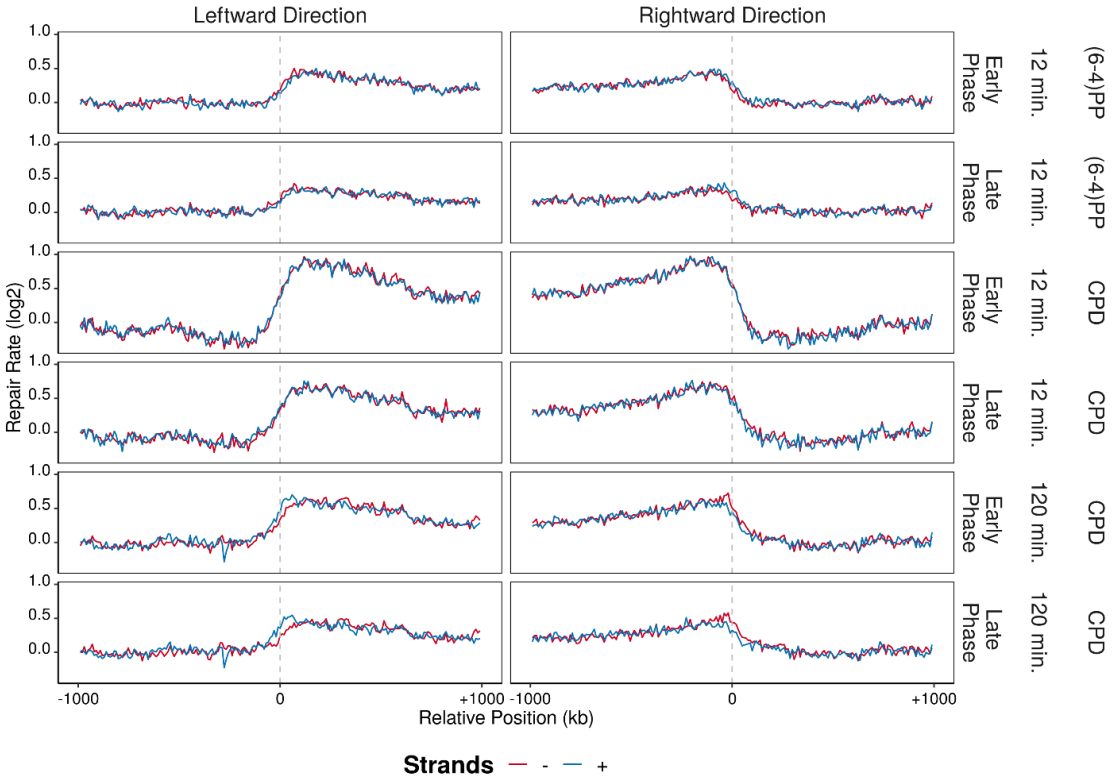

Figure 6.66 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

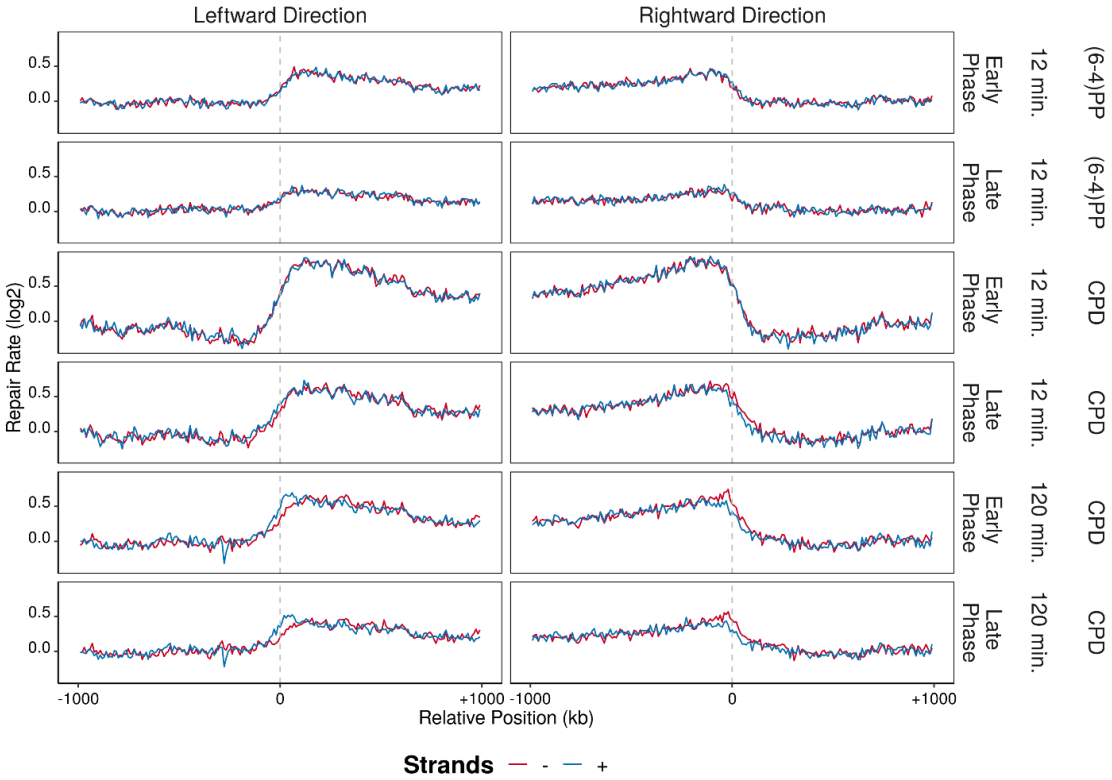

Figure 6.67 Repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions and log2 transformed in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

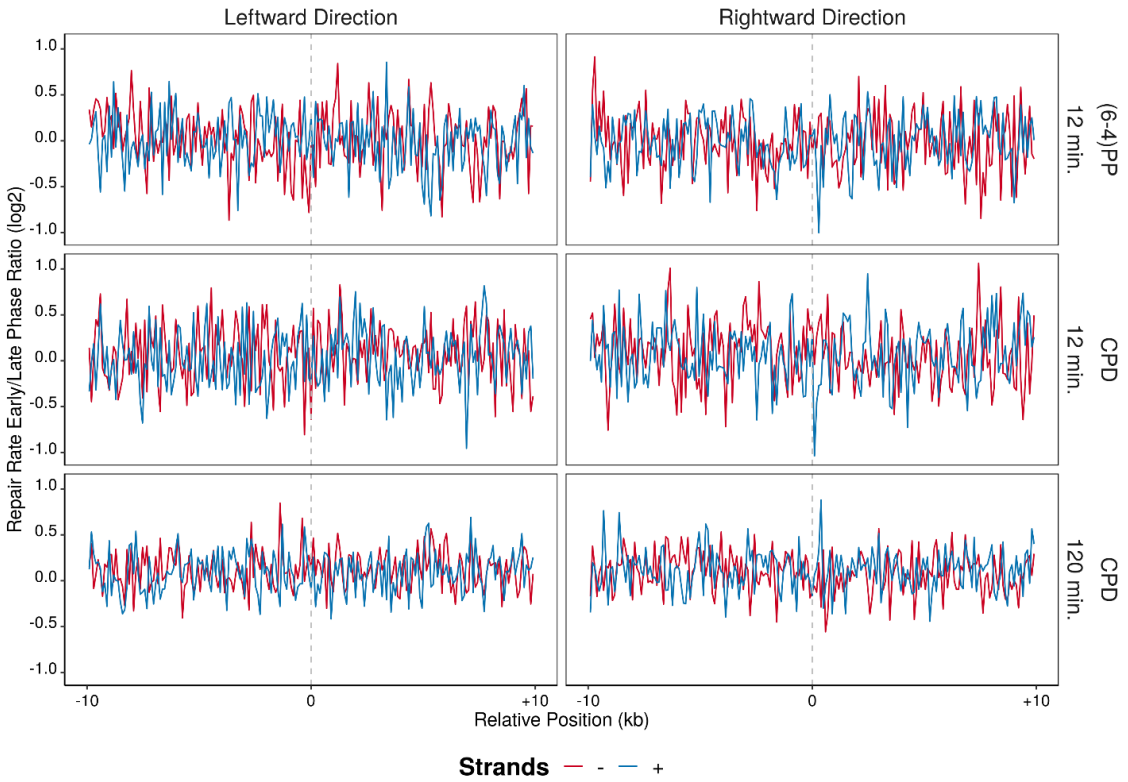

Figure 6.68 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

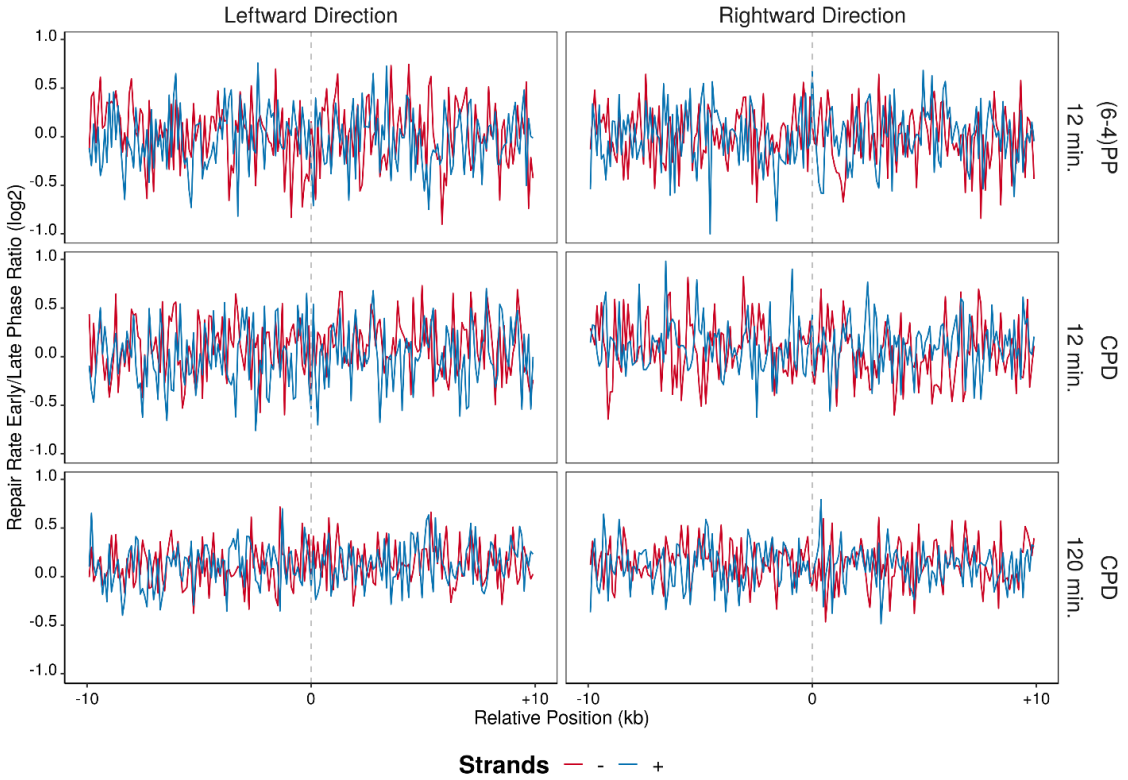

Figure 6.69 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

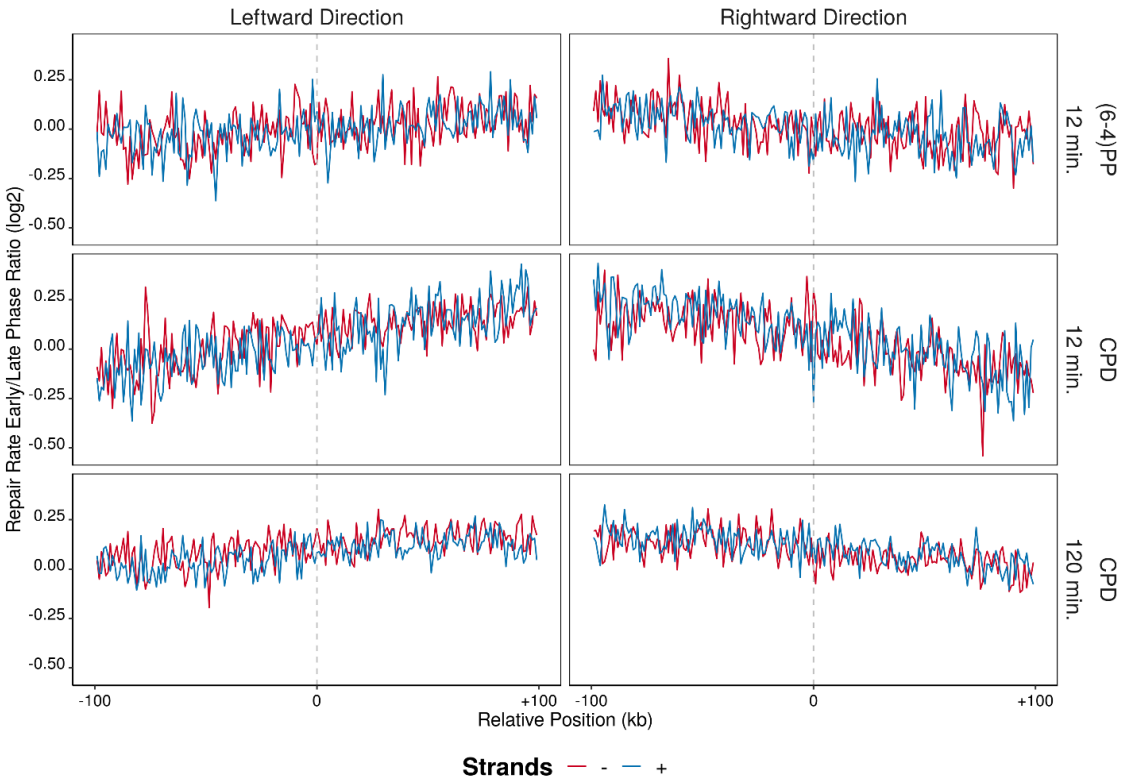

Figure 6.70 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

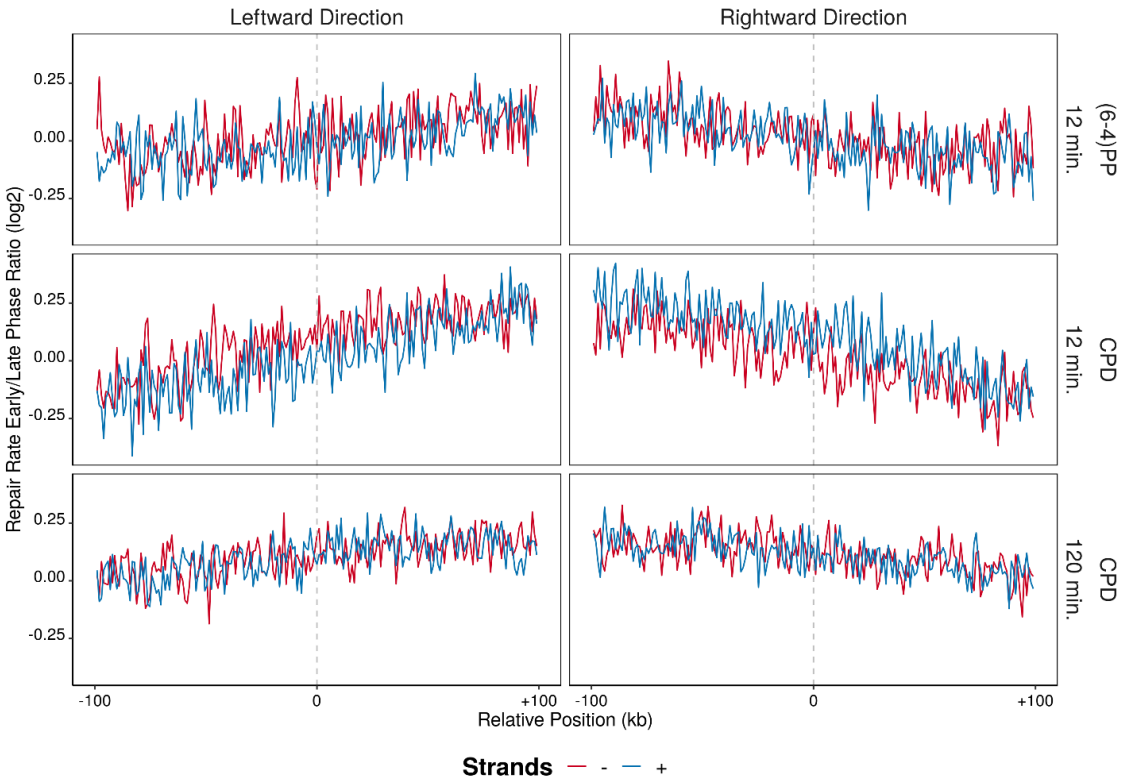

Figure 6.71 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

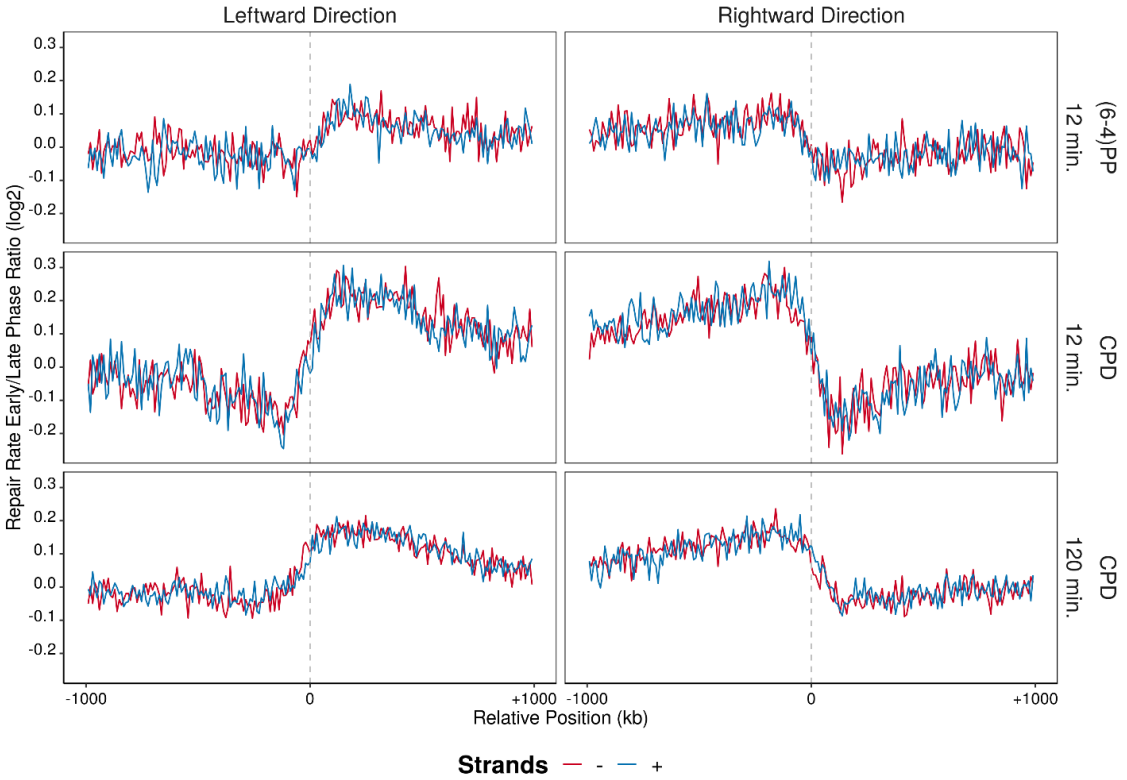

Figure 6.72 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

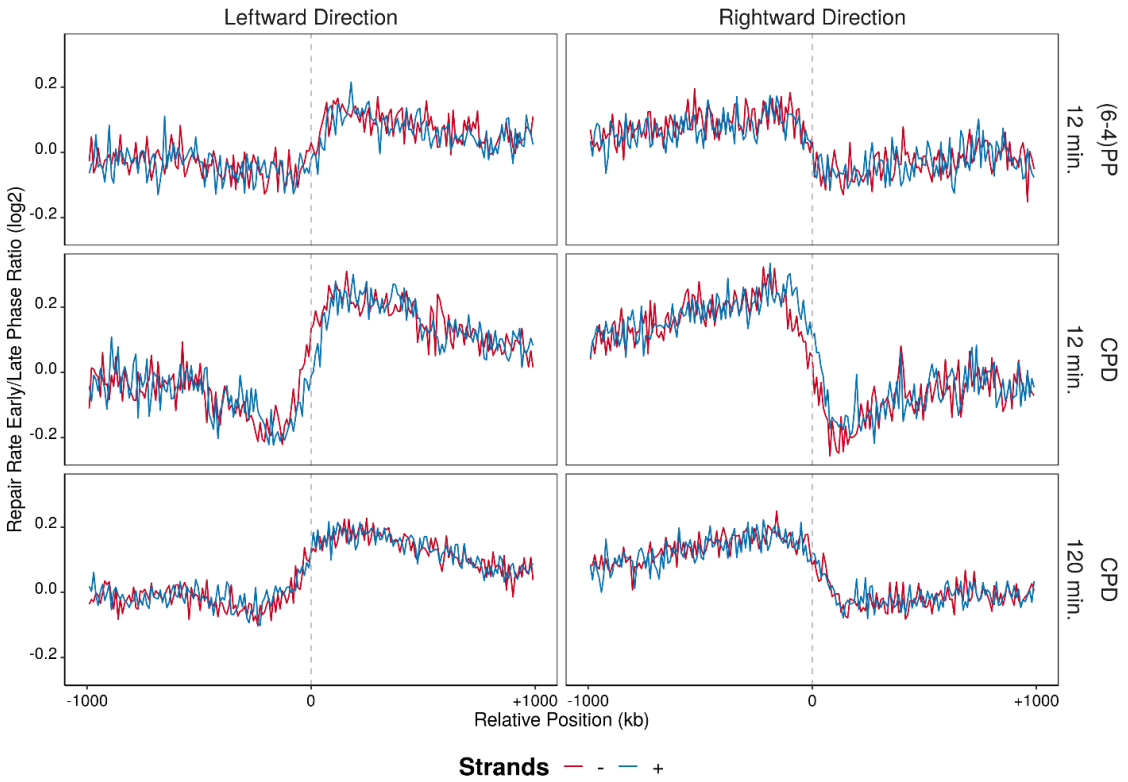

Figure 6.73 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, samples that are at early S phase of the cell cycle are further divided by the ones that are at the late S phase (early/late) in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

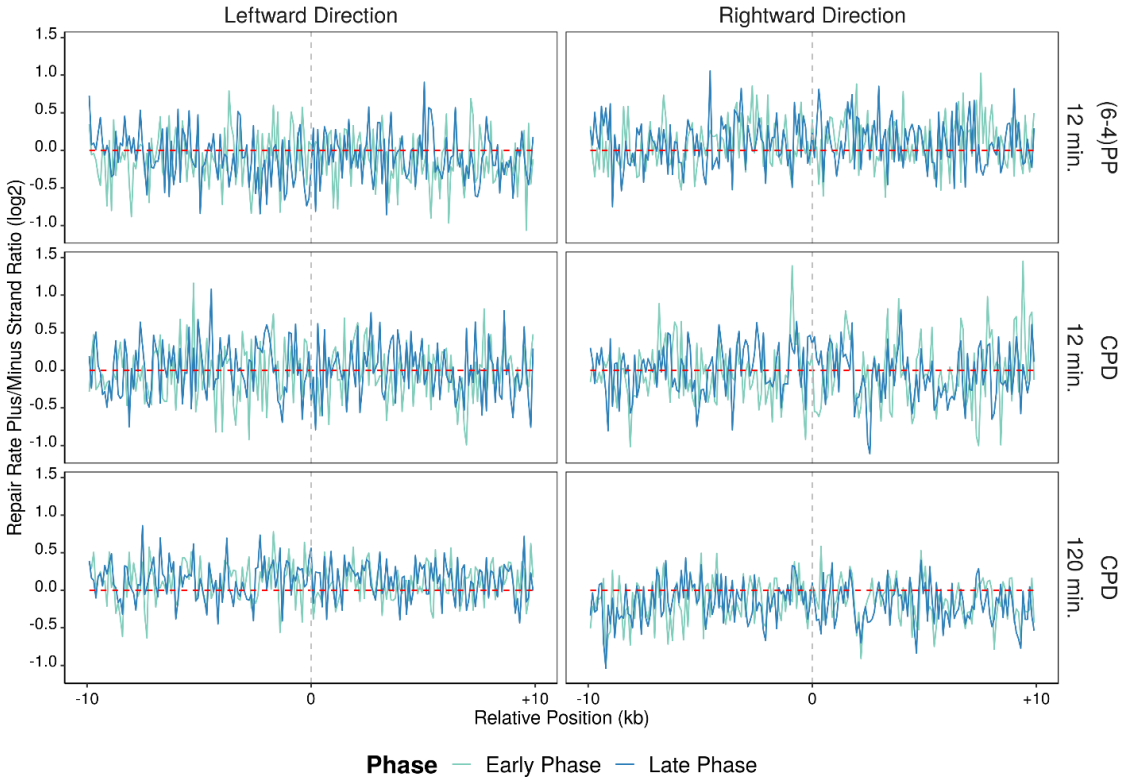

Figure 6.74 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

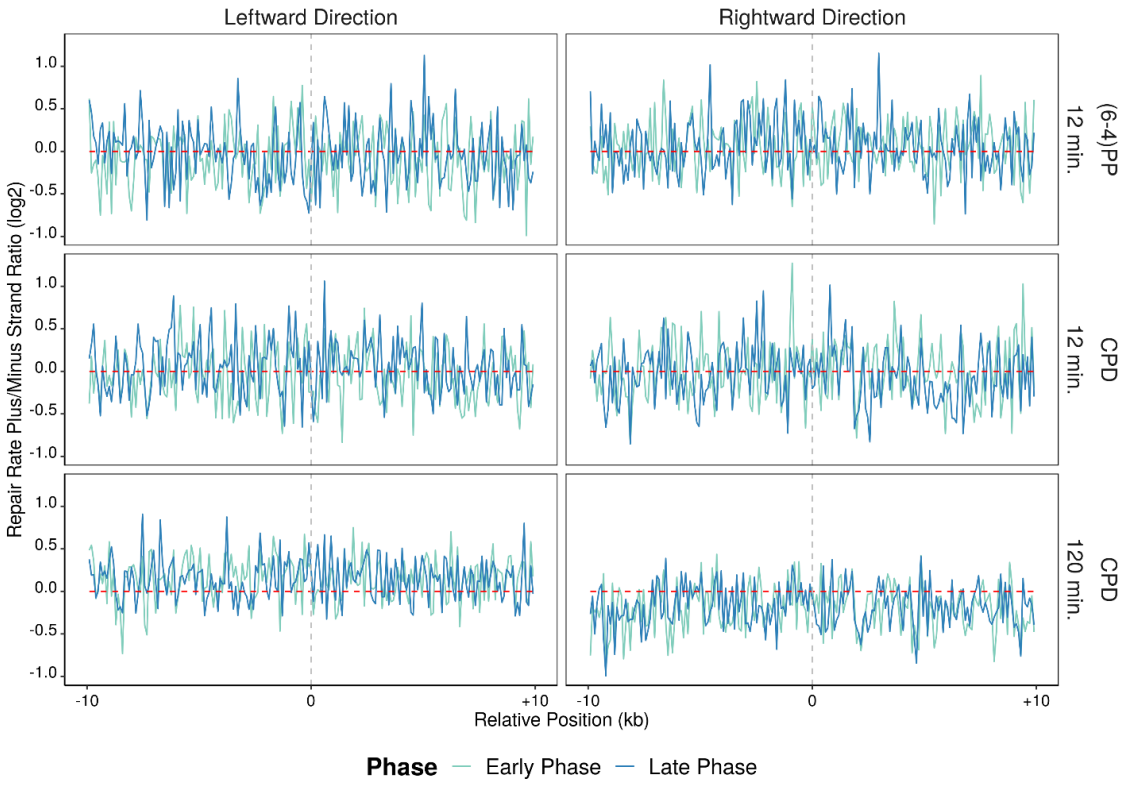

Figure 6.75 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 20 kb windows with 100 base pair intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

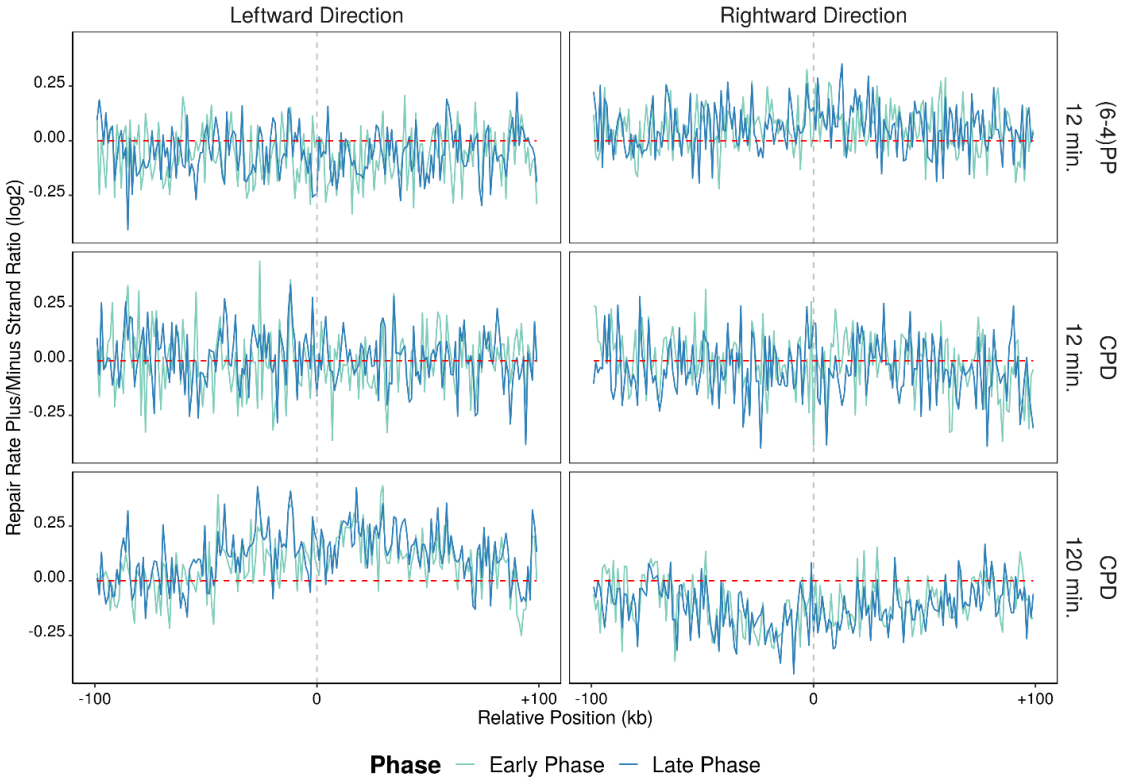

Figure 6.76 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

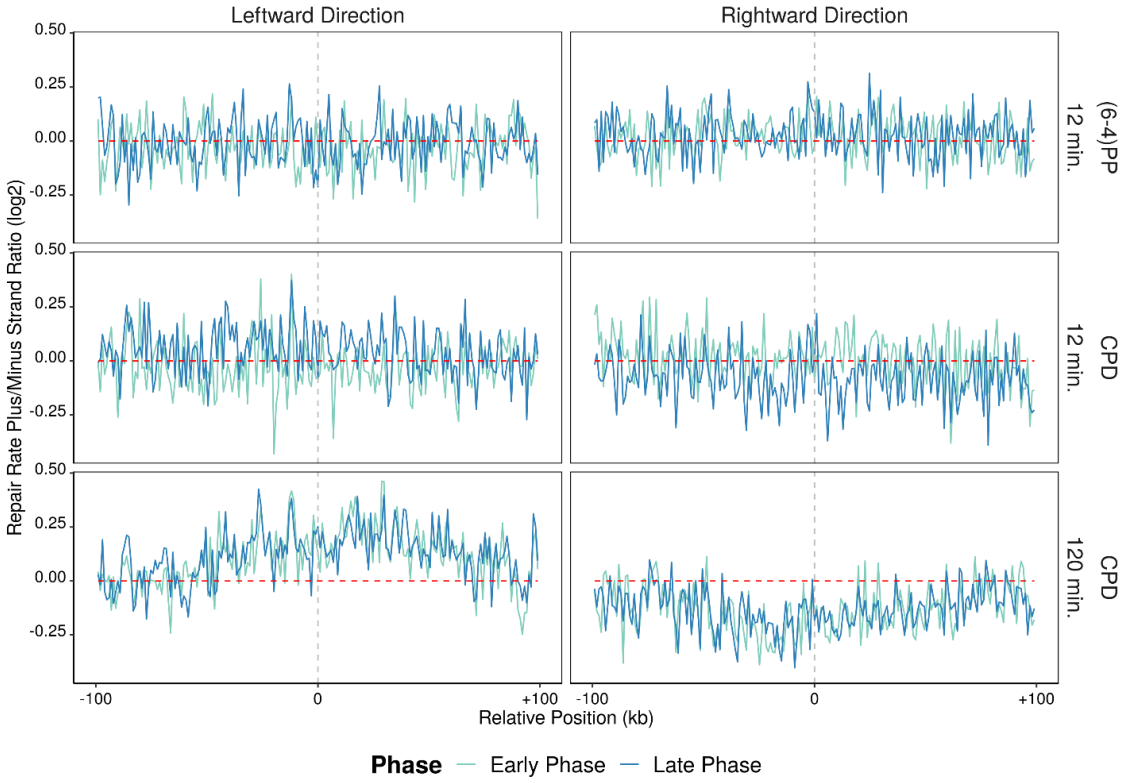

Figure 6.77 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 200 kb windows with 1 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.

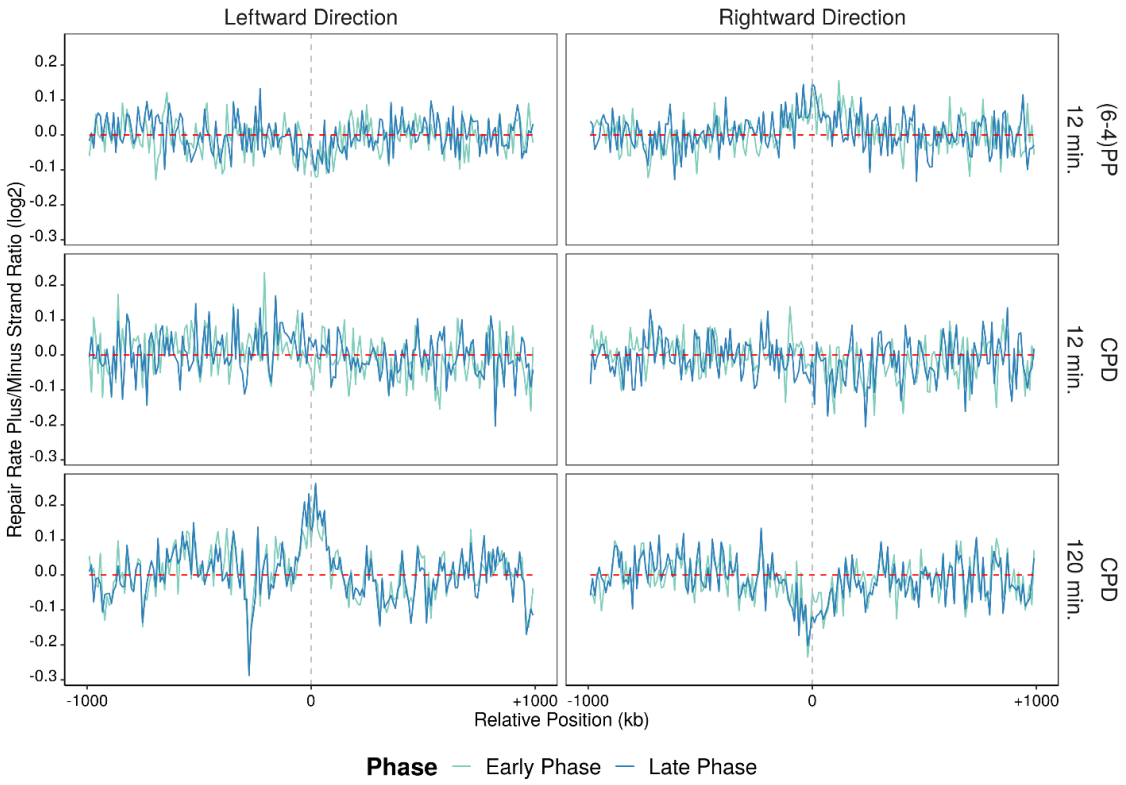

Figure 6.78 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate A.

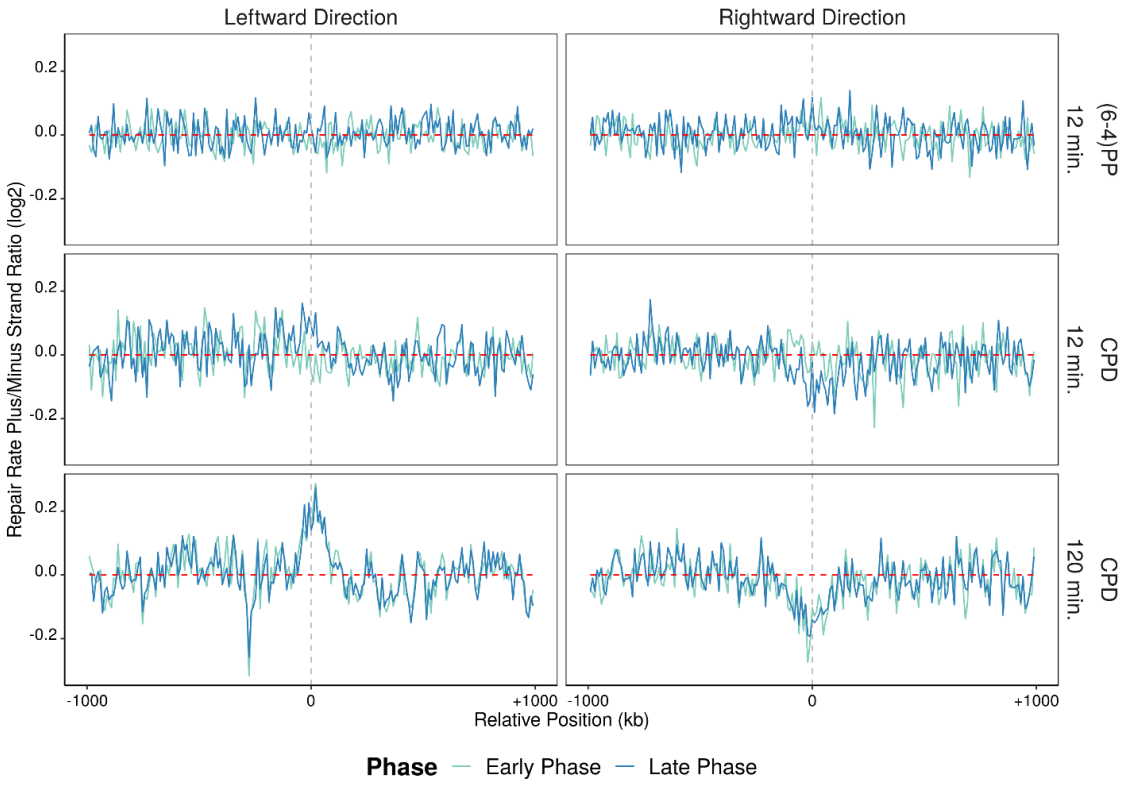

Figure 6.79 After log2 transformed repair rates (XR-seq/Damage-seq) of high RFD regions are calculated at both leftward and rightward directions, plus strands of the samples are further divided to minus strands (plus/minus) in 2 Mb windows with 10 kb intervals. The blue lines are the plus strands and red lines are the minus strands. Analysis is performed on replicate B.