

บทบาทของการฝึกขาดของดีเอ็นเอสายคู่ที่เกิดขึ้นเองในการแก่งของเซลล์ยีสต์ในสภาวะที่ไม่มีการ
สร้างดีเอ็นเอ



นางสาวจิรพรรณ ทองสร้อย

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาชีวเวชศาสตร์ (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2556

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)

เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR) are the thesis authors' files submitted through the University Graduate School.

THE ROLE OF RIND-EDSBs IN CHRONOLOGICAL AGING YEAST CELL

Miss Jirapan Thongsroy



จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Biomedical Sciences

(Interdisciplinary Program)

Graduate School

Chulalongkorn University

Academic Year 2013

Copyright of Chulalongkorn University

Thesis Title	THE ROLE OF RIND-EDSBs IN CHRONOLOGICAL AGING YEAST CELL
By	Miss Jirapan Thongsroy
Field of Study	Biomedical Sciences
Thesis Advisor	ProfessorApiwat Mutirangura, M.D., Ph.D.
Thesis Co-Advisor	Associate ProfessorOranart Matangkasombut, D.D.S., Ph.D

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

.....Dean of the Graduate School
(Associate ProfessorAmon Petsom, Ph.D.)

THESIS COMMITTEE

.....Chairman
(ProfessorPrasit Pavasant, D.D.S., Ph.D.)

.....Thesis Advisor
(ProfessorApiwat Mutirangura, M.D., Ph.D.)

.....Thesis Co-Advisor
(Associate ProfessorOranart Matangkasombut, D.D.S., Ph.D)

.....Examiner
(Associate ProfessorKanya Suphapeetiporn, M.D., Ph.D.)

.....Examiner
(ProfessorNattiya Hirankarn, M.D., Ph. D.)

.....External Examiner
(Siwanon jirawatnotai, Ph.D.)

จิรพรรณ ทองสร้อย : บทบาทของการฉีกขาดของดีเอ็นเอสายคู่ที่เกิดขึ้นเองในการแก่ของเซลล์ยีสต์
ในสถานะที่ไม่มีการสร้างดีเอ็นเอ. (THE ROLE OF RIND-EDSBs IN CHRONOLOGICAL AGING
YEAST CELL) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ. ดร. นพ.อภิวัฒน์ มุทิรางกูร, อ.ที่ปรึกษาวิทยานิพนธ์
ร่วม: รศ. ดร. ทญ.อรนาฎ มาตั้งคสมบัติ, 102 หน้า.

เมื่อไม่นานมานี้เราได้ค้นพบการฉีกขาดของดีเอ็นเอที่เกิดขึ้นเองแบบที่มีลักษณะที่ไม่เคยมีการค้นพบ
มาก่อน เรียกว่า Replication INDependent EDSBs (RIND-EDSBs) ในเซลล์มนุษย์ RIND-EDSBs พบได้ใน
เซลล์ทุกชนิดและทุกระยะของวัฏจักรของเซลล์ เนื่องจาก RIND-EDSBs พบได้ในระยะพัก (G0) RIND-EDSBs
จึงหมายถึง ดีเอ็นเอที่ฉีกขาดที่เกิดขึ้นเองโดยไม่ต้องอาศัยการแบ่งตัวของสายดีเอ็นเอ ในมนุษย์ RIND-EDSBs
จะรักษาสภาพอยู่ในสายดีเอ็นเอที่มีหมู่เมทิลมาก ในโครมาตินที่ขาดหมู่อะเซทิล และ H2AX การซ่อมแซม
RIND-EDSBs มีความแม่นยำกว่าการซ่อมดีเอ็นเอที่ฉีกขาดในสภาวะผิดปกติ RIND-EDSBs ในบริเวณที่มีหมู่
เมทิลสูงในระยะ G0 จะถูกซ่อมแซมแบบต่อปลายโดยตรงโดย ATM ซึ่งเป็นระบบที่มีความแม่นยำสูง ในขณะที่
ดีเอ็นเอที่ฉีกขาดโดยรังสีจะถูกซ่อมแซมโดย DNA-PKc ซึ่งวิธีที่เร็วแต่มีความผิดพลาดได้มากกว่า เราวิเคราะห์
ปริมาณ RIND-EDSBs ในยีสต์ที่เข้าสู่กระบวนการแก่ตามระยะเวลาโดยไม่ได้มีการแบ่งตัว (chronological
aging) และพบว่าเซลล์ที่แก่ไม่เพียงแต่มีอัตราการรอดชีวิตที่ลดลงแต่ยังมีปริมาณ RIND-EDSBs ที่ลดลงอีกด้วย
การศึกษาครั้งนี้แบ่งออกเป็นสามหัวข้อหลัก ประการแรกคือ การสร้างระบบการวัด RIND-EDSBs ในยีสต์สาย
พันธุ์ *Saccharomyces cerevisiae* ประการที่สองคือ การพิสูจน์ว่า RIND-EDSBs มีคุณสมบัติเป็น biological
breaks และมีกระบวนการในระดับโมเลกุลทำหน้าที่ควบคุมการเก็บรักษา RIND-EDSBs ไว้ ประการสุดท้ายคือ
วิเคราะห์หาความสัมพันธ์ระหว่างอัตราการรอดชีวิตและปริมาณ RIND-EDSBs ผลการวิเคราะห์พบว่า ระดับ
RIND-EDSBs เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติในยีสต์ที่ไม่สามารถสร้างโปรตีน Rpd3, เอนไซม์
endonucleases, เอนไซม์ topoisomerase และโมเลกุลที่ทำหน้าที่ควบคุมการซ่อมแซมดีเอ็นเอ (DNA repair
regulators) แต่ในทางตรงกันข้ามเราพบว่า ระดับ RIND-EDSBs ลดลงในยีสต์สายที่ขาดโปรตีนที่ทำหน้าที่
เกี่ยวกับการขจัดตัวของเส้นใยโครมาติน ตัวอย่างเช่นโปรตีนในกลุ่ม the high-mobility group box และโปรตีน
Sir2 นอกจากนี้พบว่า การเกิด RIND-EDSBs มีรูปแบบของลำดับเบสของดีเอ็นเอที่จำเพาะไม่ได้เกิดขึ้นแบบสุ่ม
และพบว่า จะเกิดหลังลำดับเบส “CGK” เพื่อศึกษาบทบาทของ RIND-EDSBs เราจึงทำการชักนำการซ่อมรอย
ขาดของดีเอ็นเอโดยการกระตุ้นการทำงานของเอนไซม์ HO endonuclease เราพบว่าหลังจากทำการกระตุ้น
การทำงานของเอนไซม์เป็นเวลา 3 วัน สามารถตรวจวัดระดับของ RIND-EDSBs และอัตราการรอดชีวิต ได้
ลดลง และยังมีชีวิตต่อสารคาแพนโดยการเพิ่มการซ่อมแซมแบบรวดเร็ว ซึ่งรอยขาดแบบนี้จะมีลำดับเบส
ของดีเอ็นเอที่ไม่ใช่ “CGK” ดังนั้นเราจึงสรุปว่า ระดับของ RIND-EDSBs ถูกเก็บรักษาไว้ในจีโนม มีการควบคุม
แบบไดนามิก ไม่ได้เกิดขึ้นแบบสุ่ม ซึ่งอาจจะเกิดจากอิทธิพลของโครงสร้างจีโนม หรือโครงสร้างของโครมาติน
และประสิทธิภาพของการซ่อมแซมดีเอ็นเอ นอกจากนี้เราพบว่า RIND-EDSBs สามารถแยกได้เป็นสองกลุ่มคือ
RIND-EDSBs ที่ไม่ได้เกิดแบบสุ่ม ซึ่งลำดับเบสจะเป็น “CGK” และอีกกลุ่ม ไม่มีลำดับเบสที่จำเพาะเจาะจง
สุดท้ายนี้พบว่า การลดลงของปริมาณ RIND-EDSBs เกิดขึ้นในขณะที่เซลล์แก่ นำมาซึ่งการเพิ่มขึ้นของการเกิด
พยาธิสภาพ RIND-EDSBs แบบที่ไม่ได้มีลำดับเบสแบบ CGK และนำมาซึ่งการลดลงของอัตราการรอดชีวิต จาก
ผลการทดลองยังชี้ให้เห็นว่า CGK-RIND-EDSBs อาจจะเป็นหนึ่งในการควบคุมแบบเหนือพันธุกรรม ซึ่งมีบทบาท
สำคัญในการป้องกันความไม่เสถียรของจีโนมและกลไกการแก่ของเซลล์ โดยเซลล์ที่มีระดับของ RIND-EDSBs
ลดลงอาจจะมีผลกระทบต่อการใช้ชีวิตของเซลล์

สาขาวิชา ชีวเวชศาสตร์

ปีการศึกษา 2556

ลายมือชื่อนิสิต

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม

5287760220 : MAJOR BIOMEDICAL SCIENCES

KEYWORDS: ENDOGENOUS DNA DOUBLE-STRAND BREAKS / AGING CELLS / GENOMIC INSTABILITY

JIRAPAN THONGSROY: THE ROLE OF RIND-EDSBs IN CHRONOLOGICAL AGING YEAST CELL.
 ADVISOR: PROF.APIWAT MUTIRANGURA, M.D., Ph.D., CO-ADVISOR: ASSOC. PROF.ORANART
 MATANGKASOMBUT, D.D.S., 102 pp.

Recently, we discovered a new class of endogenous DNA double strand breaks (EDSBs). They are found in all cell types and phases. Because these EDSBs are found in G₀, we named them replication independent EDSBs (RIND-EDSBs). In human, RIND-EDSBs are retained in hypermethylated DNA within deacetylated chromatin but without H2AX. RIND-EDSB repair is more precise than pathologic EDSB repair. While methylated RIND-EDSBs are repaired by precise ATM mediated non-homologous end joining (NHEJ) during G₀, radiation induced DSBs are repaired by fast and more error prone DNA-PKcs mediated NHEJ pathway. We evaluated RIND-EDSBs in chronological aging yeast. Not only the viability but also RIND-EDSBs are decreased in aged yeasts. This study was divided into three main parts. First is to set up the method for measure RIND-EDSBs in *Saccharomyces cerevisiae*. Second is to prove that RIND-EDSBs are biological breaks that have the molecular mechanisms regulation in retained RIND-EDSBs. Finally, the relationship between viability and RIND-EDSBs in yeast was evaluated. We found that the RIND-EDSB levels increased significantly in yeast strains lacking proteins involved in some histone deacetylases, endonucleases, topoisomerase, and DNA repair regulators. In contrast, RIND-EDSB levels were downregulated in the mutants that lack chromatin-condensing proteins, such as the high-mobility group box proteins, and Sir2. Moreover, RIND-EDSB occurrences have specificity in terms of sequence pattern, that is, the majority of breaks occurred right after the sequence "CGK". The specificity in "CGK" sequence is not due to chance as "CGK" sequence in yeast genome occurs non-primarily. To evaluate the roles of RIND-EDSBs, we induced global EDSB repair by HO induction. We found that yeasts at 3 days after HO induction had lower RIND-EDSB levels. The lower RIND-EDSB levels were associated with lowering viability, increasing caffeine sensitivity and increasing fast repaired non-CGK RIND-EDSBs. Therefore, we concluded that, the genomic levels of RIND-EDSBs are evolutionally conserved, dynamically regulated, non-randomly presented and may be influenced by genome topology, chromatin structure, and the efficiency of DNA repair systems. Furthermore, we found two classes of RIND-EDSBs. The first occurs non-randomly, preceding the break by CGK sequence. The others do not possess specific sequence. Finally, reduction of RIND-EDSBs during chronological aging led to increased production of pathologic fast repair non-CGK RIND-EDSBs and consequently reduce cell viability. These evidences suggested that the CGK-RIND-EDSBs may be epigenetic marks, playing important role in preventing genomic instability and cellular aging. Cells may normally maintain a certain level of RIND-EDSBs, a reduction of which may affect cell viability.

Field of Study: Biomedical Sciences

Academic Year: 2013

Student's Signature

Advisor's Signature

Co-Advisor's Signature

ACKNOWLEDGEMENTS

I would like to express my appreciation and sincere thanks to my thesis advisor, Professor Dr. Apiwat Mutirangura for his invaluable help, excellent advice and constant encouragement throughout the course of this research. I am most grateful for his teaching and advice, not only the research methodologies but also many methodologies on life.

In addition, I am extremely grateful to Associate Professor Oranart Matangkasombut for helpful suggestions, and insightful comments during my study. I want to thank my other committee and special members, Professor Prasit Pavasant, Associate Professor Kanya Suphapeetiporn, Professor Nattiya Hirankarn, and Dr. Siwanon jirawatnotai their stimulating suggestions.

I also appreciate, Miss Maturada, Miss Araya, Mr. Prakasit, Dr. Monnat, and Dr. Nakarin for their assistance and encouragement, this work would not have been accomplished without them. Finally, I would like to express my deepest gratitude to my family for their love and understanding all the time. This thesis was partially supported by Chulalongkorn University under the office of the Higher Education Commission.

CONTENTS

	Page
THAI ABSTRACT	iv
ENGLISH ABSTRACT	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF TABLES	viii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi
CHAPTER I INTRODUCTION.....	1
CHAPTER II LITERATURE REVIEW	12
CHAPTER III METERIALS AND METHODS.....	26
CHAPTER IV RESULTS	36
CHAPTER V DISCUSSION	63
REFERENCES	70
APPENDIX.....	77
VITA.....	102

LIST OF TABLES

Table	Page
1 Yeast strains used in this study.....	28



LIST OF FIGURES

Figure	Page
1 Sequence logo of RIND-EDSBs in yeast.....	3
2 The correlation between RIND-EDSBs prior to vanillin..... treatments and cell survival	14
3 Double strand break (DSB) repair in yeast.....	15
4 A detailed description of the mechanism of..... non-homologous end joining	17
5 The hallmarks of aging.....	19
6 Epigenetic change in aging.....	19
7 Paradigm of cellular aging cell in yeast.....	20
8 Two Major pathways of chronological life span regulation in yeast.....	22
9 Euchromatin and heterochromatin are characterized by the..... different level of methylation, acetylation, and phosphorylation	23
10 Loss of DNA methylation in aging.....	25
11 Association between age and IRS methylation of Alu.....	25
12 Schematic representation of DNA preparation for Ion Torrent sequencing.....	34
13 Ty1-EDSB-LMPCR assay.....	37
14 EDSBs in various phases of the cell cycle.....	40
15 EDSBs were detected in different DNA preparations.....	41

LIST OF FIGURES

Figure	Page
16 RIND-EDSB levels using HMW DNA preparation and..... intranuclear ligation protocols	43
17 Levels of RIND-EDSBs in yeast strains with mutations..... in DSB repair pathways	45
18 Levels of RIND-EDSBs in yeast strains with deletions of gene..... encoding proteins with the High-Mobility Group B (HMGB) domain	47
19 Levels of RIND-EDSBs in HeLa cells transfected with HMGB1 siRNA.....	48
20 RIND-EDSBs in strains with mutations in genes..... encoding topoisomerases, their partners, and endonucleases	49
21 RIND-EDSBs and chromatin regulators.....	51
22 The levels of EDSBs decreased during chronological aging.....	53
23 Levels of EDSBs in strains with inducible HO endonuclease..... expression	55
24 Correlation between RIND-EDSB levels and %viability of cells.....	56
25 Caffeine treatment in JKM179 yeast strains.....	58
26 Sequence logo of RIND-EDSBs in HO induction experiments.....	61
27 Number of retained break vs number of fast repaired break.....	62

LIST OF ABBREVIATIONS

COBRA	Combined bisulfite restriction analysis
DSB	DNA Double Strand Breaks
EDSBs	Endogenous DNA Double Strand Breaks
HDAC	Histone deacetylase
HMGB1	High Mobility Group Box 1
HMW	Hight-molecular-weight
HR	Homologous recombination
IRSPCR	Interspersed repetitive sequence PCR
LINE-1	Long Intersprsed Nuclear Element Type 1
LMPCR	Ligation-mediated polymerase chain reaction
NHEJ	Non-homologous end-joining
PCR	Polymerase chain reaction
RIND-EDSBs	Replication Independent Endogenous DNA Double Strand Breaks
RT-PCR	Reverse transcriptase polymerase chain reaction
TSA	Trichostatin A

CHAPTER I INTRODUCTION

Background and Rationale

Recently, we discovered replication independent EDSBs (RIND-EDSBs) in human cells without exposure to radiation or DNA breaking agents [1, 2]. Unlike pathogenic DSBs, RIND-EDSBs may possess physiologic function to the cells. In human, RIND-EDSBs are located in regions of methylated genome, retained in facultative heterochromatin and repaired by a more precise ATM-dependent pathway. Moreover, we sequenced yeast RIND-EDSBs and found that RIND-EDSBs were non-randomly distributed in the genome. Interestingly, retained break occurrences have specificity in terms of sequence pattern; the majority of this break occurred right after “CGT” sequence. Because, RIND-EDSBs occurred non-randomly, they may be produced or retained by specific mechanisms. Thus, we propose that RIND-EDSBs are epigenetic marks that may play a role in preventing genome breakage and cellular aging. Cells may normally maintain a certain levels of RIND-EDSBs, a reduction of which may have an effect on cell survival and lifespan of cell.

I. Replication Independent Endogenous DNA Double-strand Breaks and Genome instability

In our previous work, we detected endogenous DNA double strand breaks (EDSBs) that occurred in non-replicating cells without using any inductive agents, named replication independent EDSBs (RIND-EDSBs) [1, 2]. The causes and consequences of RIND-EDSBs were different from DSBs which occur by replication or irradiation. Whereas, irradiation-induced DSBs halt cell cycle, lead to mutations, and induce cell

death [3], RIND-EDSBs did not harm the cell. In conclusion, RIND-EDSBs are not only ubiquitously expressed in human cells. Therefore, we postulated that RIND-EDSBs have an important function and occur non-randomly. Characteristics of RIND-EDSBs were determined and we found that breaks scattered along the chromosomes and the number of breaks positively correlated with chromosome size. Most importantly, breaks occurred most frequently after “CGT” sequence (Dr. Monnat, unpublished data). The frequent occurrence of RIND-EDSBs after “CGT” drew some attention to the “CGT” pattern itself. In addition, RIND-EDSBs were more likely retained in heterochromatin. Since heterochromatin typically occurred at CpG sites, this could be connected to having “CGT” pattern at breaks (Figure 1). The specificity of “CGT” was not a coincidence either, as adenine and thymine was more abundance than cytosine and guanine in the yeast genome. This suggested that the generations of RIND-EDSBs are results of specific mechanisms. The mechanism could resemble how restriction enzymes recognize specific nucleotide sequences and produce a cut, or it might consider physical property of dinucleotide sequences. It was unlikely that mechanism was chromosome location dependent since we observed break locations scattered along chromosome suggesting distinct RIND-EDSBs locations in each cell. Although, the mechanism is still unknown, we hypothesize that there is specific mechanism for RIND-EDSBs occurrences. Since RIND-EDSBs are conserved from yeasts to humans, the RIND-EDSBs producing mechanism should playing important role in preventing genomic instability.

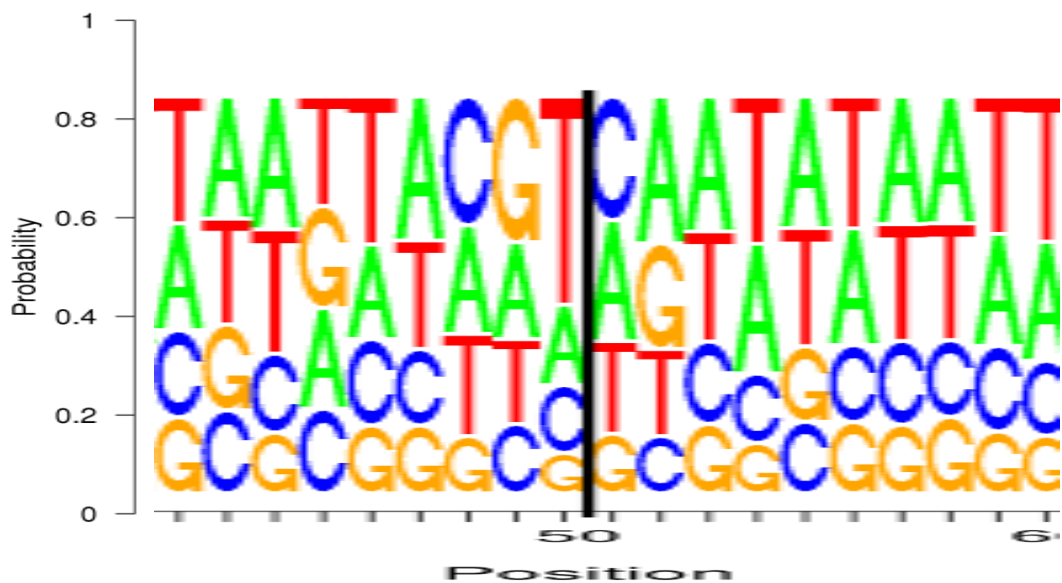


Figure 1. Sequence logo of RIND-EDSBs in yeast. The break is between position 50 and 51 marked by vertical line (Dr. Monnat, unpublished data).

II. Aging and genome instability

Aging or senescence is a complex process that leads to tissue degenerations and relates to several diseases. Molecular mechanisms underlying cellular aging have been proposed. These include loss of mitochondrial function, shortening of telomeres, DNA double strand breaks (DSBs), and epigenetic changes [4]. Aging can be defined as many process of alterative molecular accumulation by genetic and epigenetic events in the organism that lead to loss of some phenotype over time. Epigenetic modifications, including DNA methylation and histone modification, are important for development of normal process and for the maintenance of cellular functions in an adult organism. An important epigenetic alteration global hypomethylation, which has been identified as one of the hallmarks of aging. Because, global DNA hypomethylation and genome instability are common to both

cancer and aging cells [5]. The loss of genomic DNA methylation commonly coexists with histone acetylation and genomic instability, characterized by an increase in mutation rate and chromosome aberration.

III. Concluding Hypothesis

Our recent findings may provide an explanation for the mechanism how global hypomethylation and histone acetylation lead to genomic instability. We found that there exist a number of retained endogenous DNA double strand breaks (EDSBs) in hypermethylated genome and deacetylated heterochromatin [2]. These EDSBs were also found in G₀, therefore we named them Replication-INDependent-EDSBs (RIND-EDSBs). The compact structure of methylated heterochromatin might prevent rapid cellular responses to repair methylated RIND-EDSBs and allow these breaks to be repaired by a more precise DSB repair pathway. In contrast, euchromatin-associated EDSBs could be repaired by a faster but less precise pathway. Consequently, spontaneous mutations in hypomethylated genomic euchromatin may be produced at a faster rate because unmethylated EDSBs are unable to avoid the error-prone DSB repair mechanisms. Therefore, the retention of RIND-EDSBs in heterochromatin may be an epigenetic mark that may play an important role to prevent genomic instability and aging cells.

To further investigate the role of RIND-EDSBs in aging, we chose to use yeast as a model system. Yeast has proved to be a valuable model for aging research and many of the discoveries are proved to be applicable to higher organisms [6]. Thus, the main objective of this project is to investigate if RIND-EDSBs are epigenetic marks

that play a role in preventing genome breakage and cellular aging. So, genomic instability may cause aging via reduction of RIND-EDSBs.

Research Question

1. Are RIND-EDSBs detected in yeast cells?
2. Are RIND-EDSBs retained in heterochromatin in yeast cells?
3. What is the pathway used to repair RIND-EDSBs?
4. What is the mechanism producing or retaining RIND-EDSBs?
5. How are retained RIND-EDSBs associated with chronological aging cells?

Hypothesis

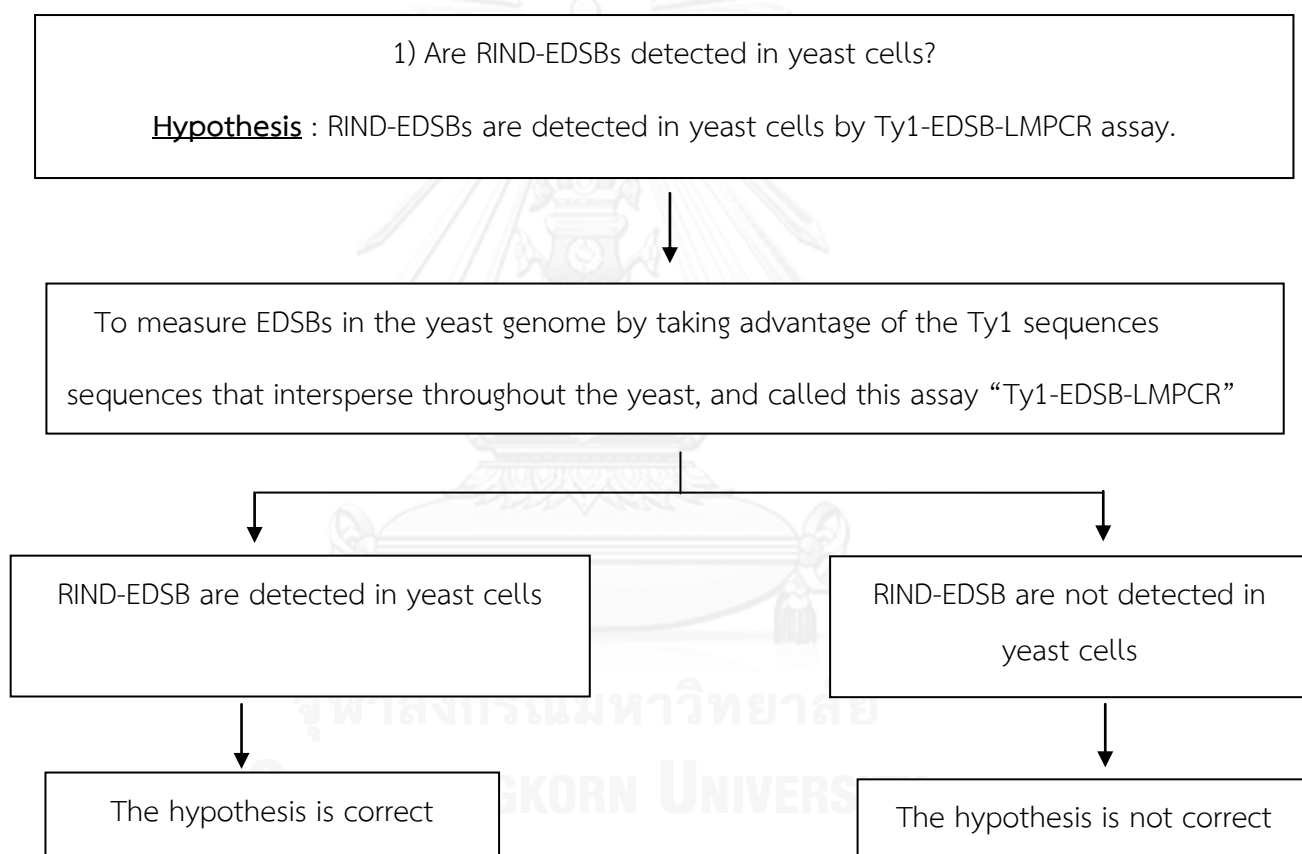
1. RIND-EDSBs are detected in yeast cells by Ty1-EDSB-LMPCR assay.
2. RIND-EDSBs are retained in heterochromatin.
3. The level of RIND-EDSBs would be increased in yeast strains which lack RIND-EDSBs repair pathways.
4. The level of RIND-EDSBs would be decreased in yeast strains lacking genes involved in RIND-EDSBs production or retention.
5. RIND-EDSB levels should change in aging cells, and altered RIND-EDSBs in aging cells should lead to certain phenotype.

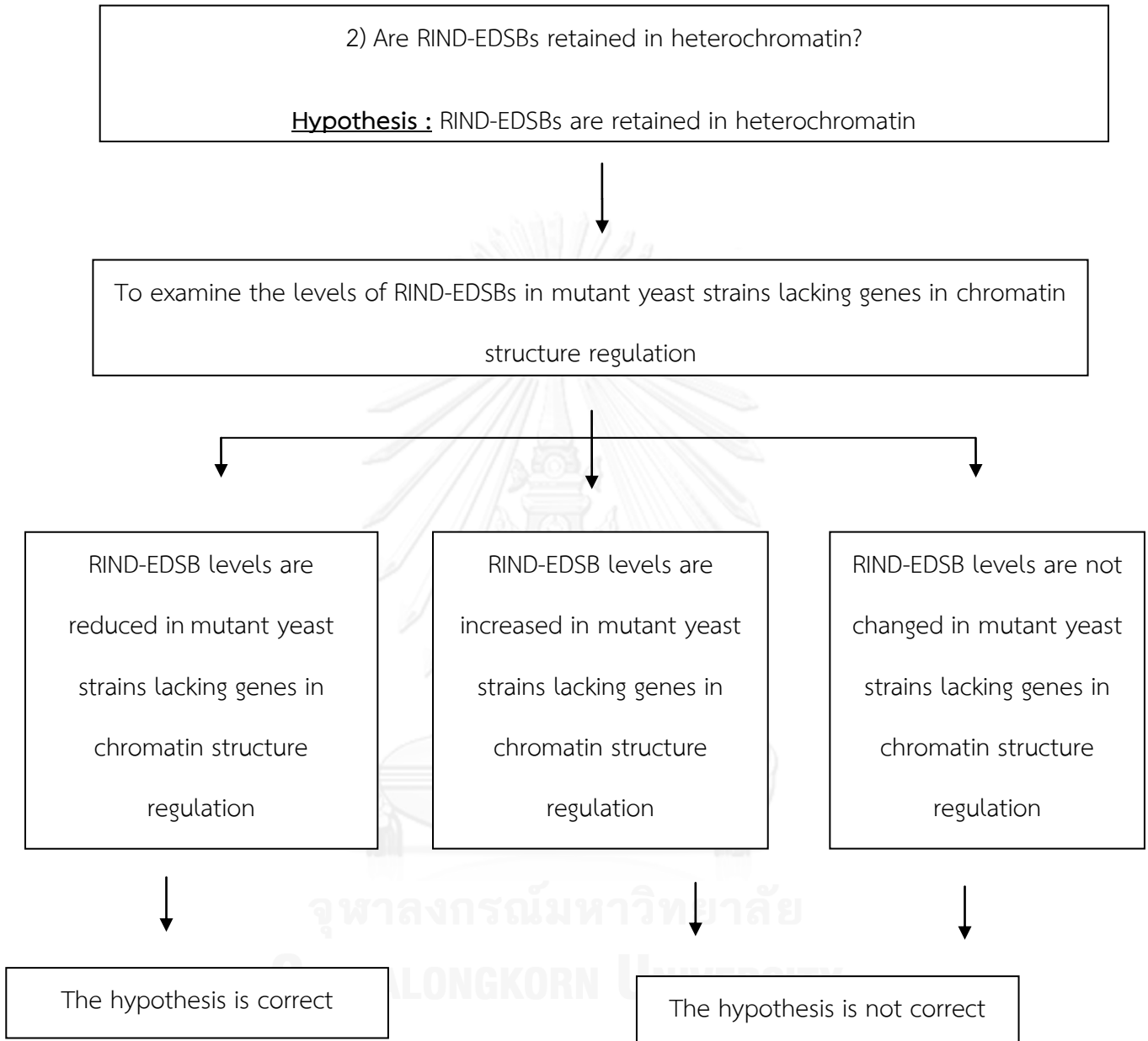
Objectives

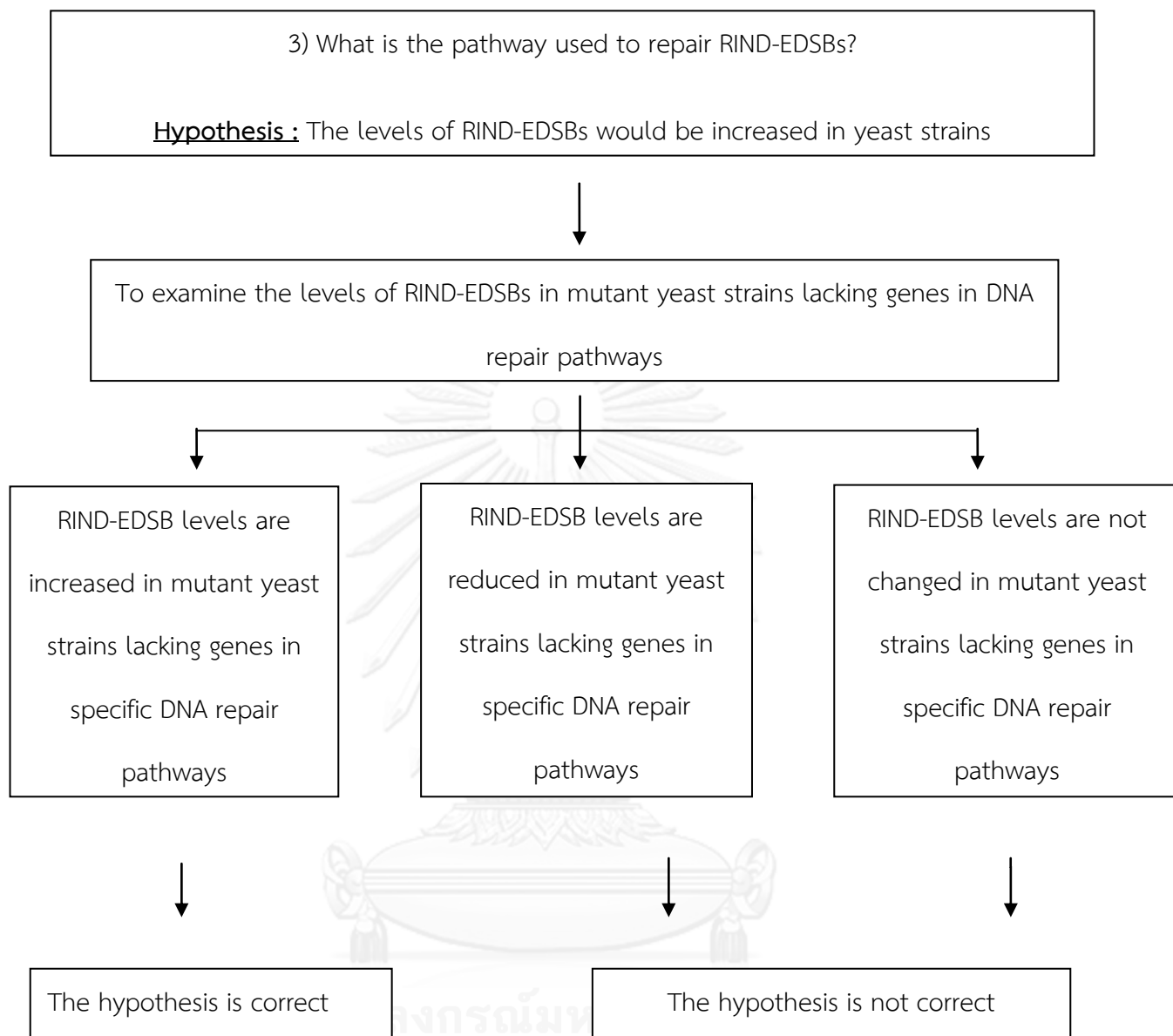
1. To setup the method for EDSBs measurement in the yeast genome by using Ty1 sequences, and called this assay "Ty1-EDSB-LMPCR".
2. To examine if RIND-EDSBs are retained in heterochromatin.

3. To examine the levels of RIND-EDSBs in a collection of mutant yeast strains lacking genes in repair pathway.
4. To explore the molecular mechanisms regulating the level of retained RIND-EDSBs.
5. To establish the relationship between retained RIND-EDSB levels and survival of yeast cells during chronological aging.

Conceptual framework







4) What is the pathway used to produce or retain RIND-EDSBs?

Hypothesis: The level of RIND-EDSBs would be decreased in yeast strains lacking gene in the production or retention of RIND-EDSBs.

To examine the levels of RIND-EDSBs in mutant yeast strains lacking genes in HMGB1, endonucleases, topoisomerase pathways

RIND-EDSB levels are reduced in mutant yeast strains lacking genes in HMGB1, endonucleases, topoisomerase pathway

The hypothesis is correct

RIND-EDSB levels are increased in mutant yeast strains lacking genes in HMGB1, endonucleases, topoisomerase pathway

The hypothesis is not correct

RIND-EDSB levels are not changed in mutant yeast strains lacking genes in HMGB1, endonucleases, topoisomerase pathway

5) How are retained RIND-EDSBs associated with senescence of yeast cell

Hypothesis : RIND-EDSB levels should be changed in aging cells.

To establish the relationship between retained RIND-EDSB levels and survival of yeast cells during aging.

RIND-EDSB levels are reduced in aging cells

The hypothesis is correct

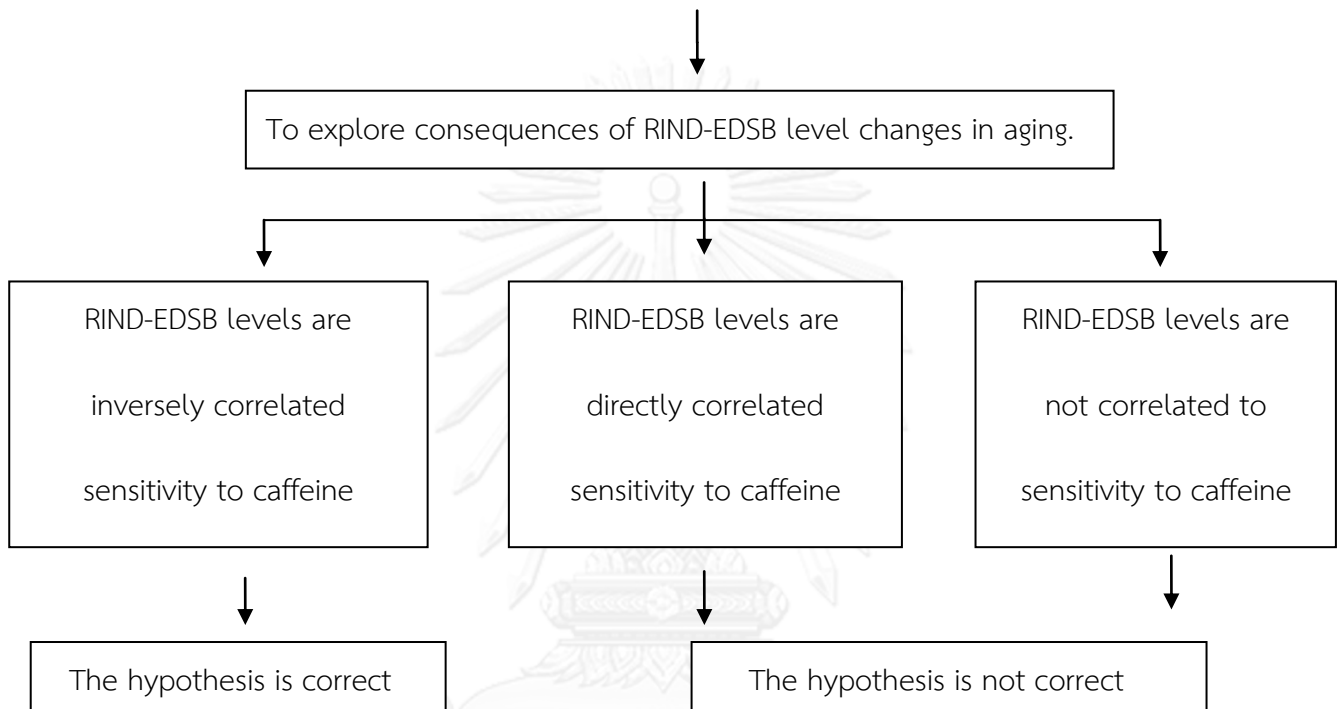
RIND-EDSB levels are increased in aging cells

The hypothesis is not correct

RIND-EDSB levels are not changed in aging cells

5.2. RIND-EDSBs level change should lead to certain phenotype.

Hypothesis : Lower RIND-EDSBs should increase sensitivity to caffeine treatment.



Key Words

Endogenous DNA double-strand breaks, Aging, Genomic instability

Expected Benefit

The results of this study will help us to understand the mechanism of RIND-EDSBs that may be epigenetic marks, playing important role in preventing genomic instability.

CHAPTER II

LITERATURE REVIEW

2.1. Replication independent endogenous DNA double strand breaks (RIND-EDSBs)

DNA damage can occur by environmental factors such as chemicals, radiation or ultraviolet light. Moreover, DNA damage is induced by some mistakes inside the cell during DNA replication or normal cellular metabolism. One principally harmful form of DNA is double strand breaks (DSBs). DSBs can easily lead to gross chromosomal aberrations if not rejoined quickly. Even if it is repaired, the repair process may be error-prone, and may eventually be detrimental to the organism. The process of DSB repair is also influenced by histone modification. In eukaryotic cells, a modification that occurs specifically at the location DSBs is phosphorylation of histone H2AX. This phosphorylated histone is often referred as γ -H2AX, and is one of the earliest DSB repair responses. In yeast, Tel1 and Mec1 phosphorylate H2AX (γ -H2AX) on serine 129 after the production of a double-strand break. In addition, the present of γ -H2AX induced many proteins in homologous recombination (HR) or non-homologous end joining (NHEJ) pathway at the damage site [7, 8].

Endogenous DNA double strand breaks (EDSBs) can occur spontaneously without any exogenous insults [9]. EDSBs are generally believed to result from a variety of events, such as stalled DNA replication through single stranded lesions and mechanical stress [10]. In 2003, Vilenchik and Knudson proposed that EDSBs could arise as often as 50 times per cell cycle in human cells, but most are rapidly repaired in normal cells [9]. Although the majority of spontaneous DSBs are efficiently

repaired, inaccurate repair of EDSBs could be a cause of carcinogenic mutations [9]. Therefore, in normal cells, there should exist mechanisms to avoid error-prone repair of EDSBs that could protect the genome from potentially hazardous mutations or chromosomal rearrangements [1, 2].

Recently, we discovered replication independent EDSBs (RIND-EDSBs) in all cell types without exposure to radiation or DNA breaking agents [2]. Therefore, unlike pathogenic DSBs, RIND-EDSBs may possess physiologic function to the cells. We demonstrated that methylated RIND-EDSBs in human cells were repaired by a more precise ATM-dependent non-homologous end joining (NHEJ), whereas unwanted DSBs were generally repaired by a fast, more-error prone Ku-mediated NHEJ. The compact structure of heterochromatin prevented H2AX phosphorylation, a conventional DSBs repair response, resulting in methylated RIND-EDSBs escaping error-prone NHEJ repair [2].

Currently, we found that the retained EDSBs in heterochromatin may possess biological role particularly to maintain genomic instability. By observing cell survival from vanillin, a DNA-PKcs inhibitor, we found that cell possessed more retained EDSBs survived more than the others (Figure 2).

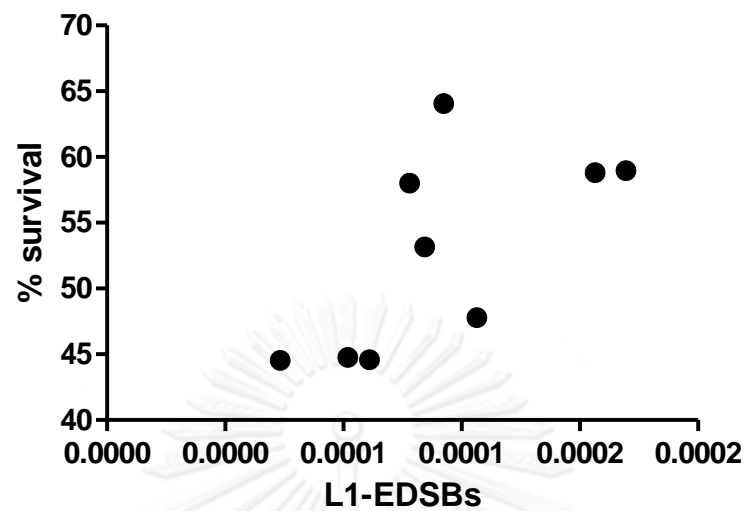


Figure 2. The correlation between RIND-EDSBs in vanillin treatments and cell survival. X-axis shows RIND-EDSB levels and Y-axis shows %survival after vanillin treatment in HeLa cell line.

2.2. DNA repair

DSB repair mechanisms in eukaryotic cells are classified into two major classes as homologous recombination (HR) and non-homologous end joining (NHEJ). Homologous recombination requires a template for the repair of DSB using a sister chromatid in the S/G2 phases of cell cycle. In contrast, non-homologous end joining ligates DNA ends without the need for a homologous template. NHEJ pathway can function in all phases, even in G0 phase (Figure 3).

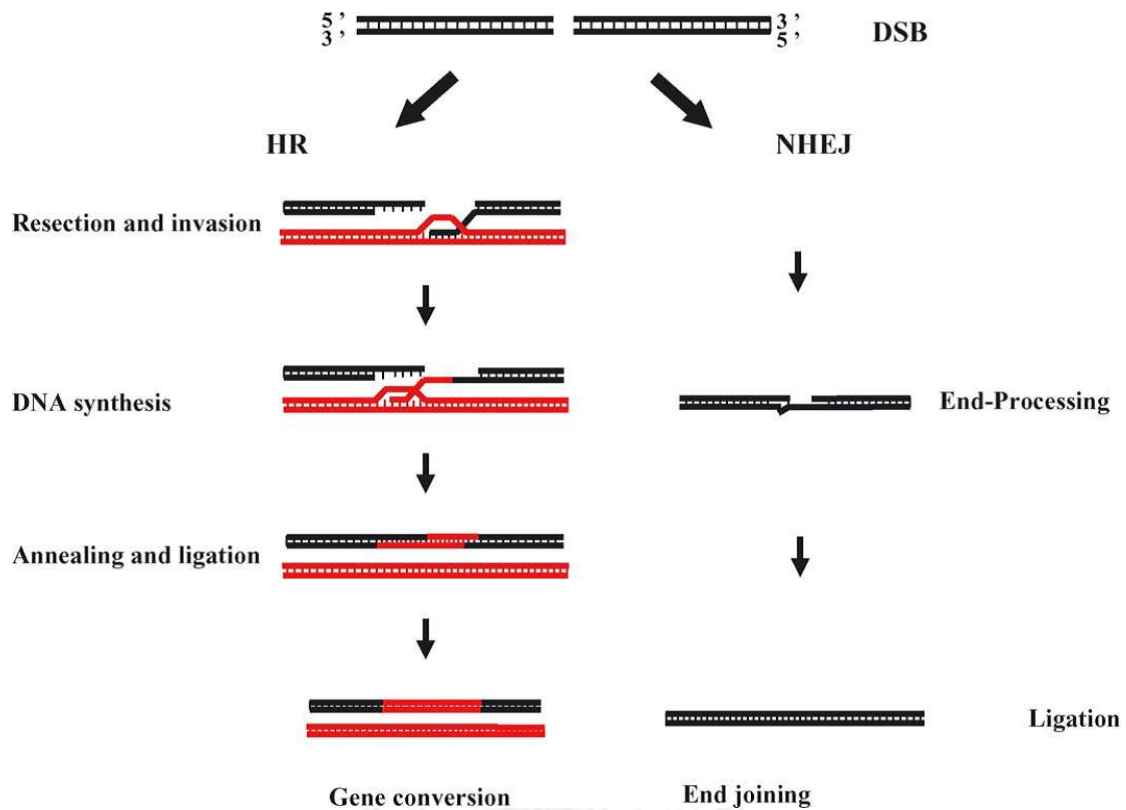


Figure 3. Double strand break (DSB) repair in yeast. DSBs are repaired by two major pathways: homologous recombination (HR) and non-homologous end joining (NHEJ). HR requires DNA synthesis using the intact homologous donor as a template. In NHEJ, non-complementary DNA sequences are processed and are ligated to restore the integrity of the chromosome [10].

2.2.1. Non homologous end joining repair in yeast

Non homologous end joining repair (NHEJ) uses little or no homology to couple DNA ends. In mammals, even though the NHEJ pathway is a major pathway, yeast mostly depends on homologous recombination. However, yeast can repair DSB by NHEJ, and most of the protein components required for NHEJ appear to be evolutionary conserved [11]. The three steps of NHEJ pathway in yeast are :(i) end

binding, (ii) end processing, (iii) end ligation. In the first step, Yku70 and Yku80 form a heterodimer that binds DNA ends [12]. The end binding activity of Yku70/Yku80 complex indicated its early role in the NHEJ process. The heterodimers are very abundant, and bind with high affinity to the ends of DNA. They are efficient for binding to a variety of terminal structures. Therefore, the Yku70/Yku80 heterodimer might bridge the DNA ends [13]. In the absence of Yku70/80 complex, not only the efficiency of end joining, but also the precision is decreased [14]. In addition, the Yku70/80 heterodimer has been shown to be necessary for the recruitment of other proteins and for the activation in vitro of the ligation reaction [15]. The second step, end processing was particularly important for repairing in non-complementary ends. The ends require terminal processing via nucleolysis and polymerization before the final ligation step. Several proteins have been implicated to function in this step of NHEJ composing Mre11, Rad50 and Xrs2 (MRX complex). Mre11 possesses single-stranded endonuclease and ATP-dependent 3'-5' exonuclease activities and Rad50 is a highly conserved coiled-coil member of the structural maintenance of chromosomes family of proteins. The role of *XRS2* appeared to be the recruitment of the MRX complex to DSBs. Thus, the MRX complex could join linear DNA ends and stimulate Yku complex [16]. In yeast strains deleted for the *yku70*, *yku80*, *mre11*, *rad50* or *xrs2* genes, the proportion of imprecise repair increases, indicating that the two complexes might either play roles in end processing, or might protect the ends from degradation [17]. In the third step, Yku70/80 was required for the recruitment of the Dnl4-Lif1 complex and Nej1 that are responsible for the rejoining step (Figure 4) [18].

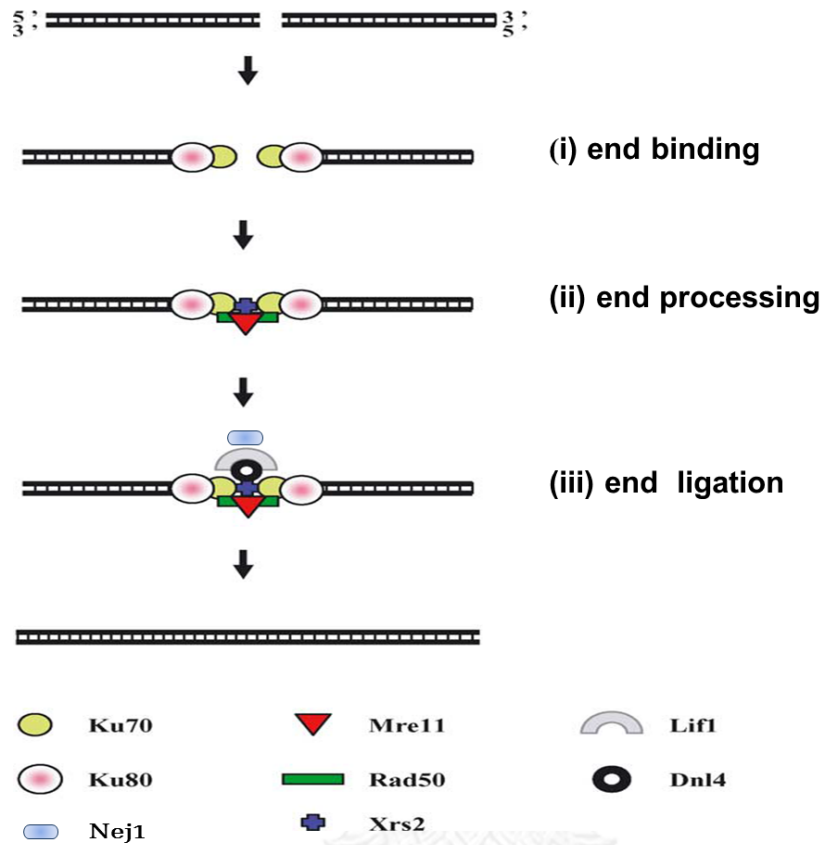


Figure 4. A detailed description of the mechanism of non-homologous end joining. Since the order of recruitment was still unsolved, steps (i) and (ii) could be reversed. Following DSBs creation, the yku70/80 heterodimer is recruited to the broken ends (i). The Mre11/Rad50/Xrs2 (MRX) complex is then recruited (ii). The Lif1/Dnl4 complex is then recruited (iii). The MRX and Yku complexes promoted activity of Lif1/Dnl4 and Nej1, resulting in ligation of the broken ends [10].

2.3. Cellular aging

Aging can be defined as the time-dependent functional decline that affects most living organisms. Characteristic morphology can be detected in aging cells such as increased nuclear and cell size, increased number of vacuoles in the endoplasmic reticulum and cytoplasm, and large lysosomal bodies. Research in cellular aging has been stimulated based on the fact that normal human diploid cells can change a limited of cell division [19]. Degeneration are the risk factor for major human pathologies, including cancers, cardiovascular disorders, diabetes, and neurological diseases.

Common phenomena that are contributed to aging in different organism has been identified and categorized as hallmarks of aging. These hallmarks are genomic instability, telomere shortening, epigenetic change, deregulated nutrient sensing, loss of proteostasis, mitochondrial dysfunction, altered intercellular communication, and cellular senescence (Figure 5) [20]. Defining hallmarks of aging may be considered to contribute a framework for future studies on molecular mechanisms of aging and improvement of human life span. One interesting of nine hallmarks of aging is epigenetic changes which are alterations in the methylation of DNA or methylation and acetylation of histone. Thus, these changes can induce instability of cell that contributed to the aging process (Figure 6) [20].



Figure 5. The hallmarks of aging. The nine hallmarks of aging include genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication [20].

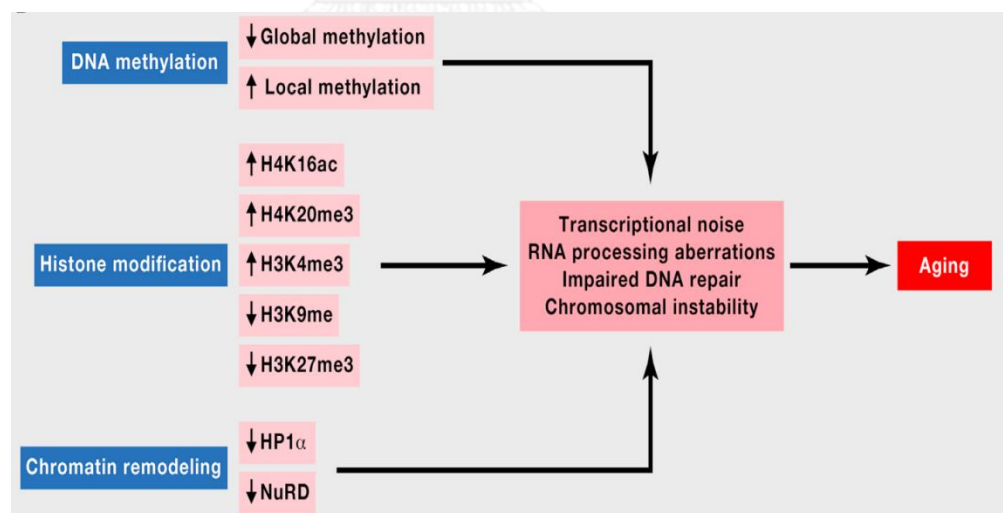


Figure 6. Epigenetic changes in aging. The methylation of DNA or acetylation and methylation of histones alter to associate with many proteins that affect to epigenetic changes and contribute to the aging process [20].

2.3.1. Aging in yeasts

The budding yeast *Saccharomyces cerevisiae* can be further considered a preeminent model for aging studies. Furthermore, the aging process of yeast cells is conserved with multicellular eukaryotes. Thus, yeast is used widely as a model organism in aging study. Cellular aging in yeast can be classified into two classes as replicative life span (RLS) and chronological life span (CLS). The replicative life span is measured by the number of times that daughter cells can divide from the mother cell until cell division stops, whereas chronological life span is measured by the length of time non-dividing yeast cells can survive [21] (Figure 7).

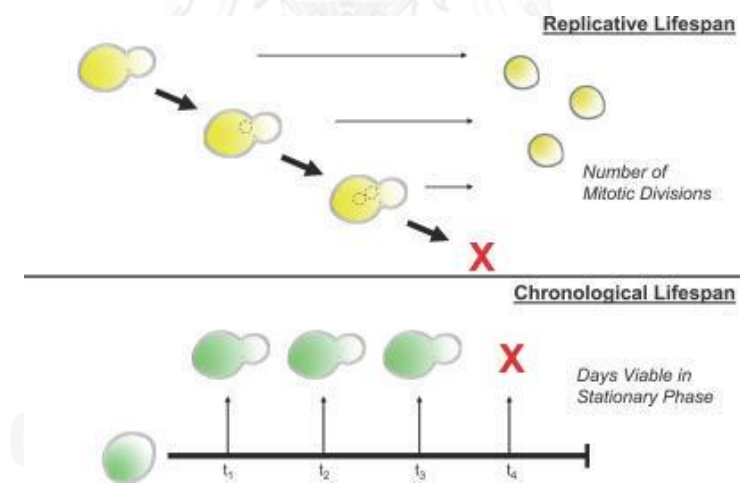


Figure 7. Paradigms of cellular aging in yeast. (A) RLS in yeast is measured life span by the number of times that the daughter cells can generate from mother cell before senescence. (B) CLS is measured by the time cells in a stationary culture can survive before cell death [22].

2.3.2. Chronological life span

Chronological life span (CLS) is currently used in several laboratories worldwide [23]. *Saccharomyces cerevisiae* is a simple model to study chronological aging. CLS is typically measured as survival of non-dividing cells. Generally, cells stop dividing within 2-7 days after starting culture, and viability of cells is monitored by colony forming units (CFUs) [24]. Yeast enters stationary phase when nutrient is limited, several proteins were discovered playing a role in life span regulation. Various studies in genetic utilization, nutrient-sensing, and mitochondrial respiration have indicated that many pathways are worked in the mechanism of chronological aging. The yeast model has been successful to identify genes which implicated to control life span. Two major regulated aging pathways in yeast cell are Tor/Sch9 pathway and Ras/Adenylate cyclase (AC)/PKA pathway [23, 25]. Fabrizio *et al.* found that life span extension noticed in inhibiting the two major pathways. Tor is a serine-threonine protein kinase, which was functioned to upstream of Sch9 to control nutrient-sensing in growing cells. *RAS2* encodes a highly conserved G-protein that activates PKA via adenylate cyclase [23]. PKA and Tor/Sch9 pathways concentrate at transcription factors that mediate the expression of stress-responsive genes (Figure 8). In addition, many results showed that characteristics of chronological aging in yeast are DNA damage and stress pro-senescence pathways.

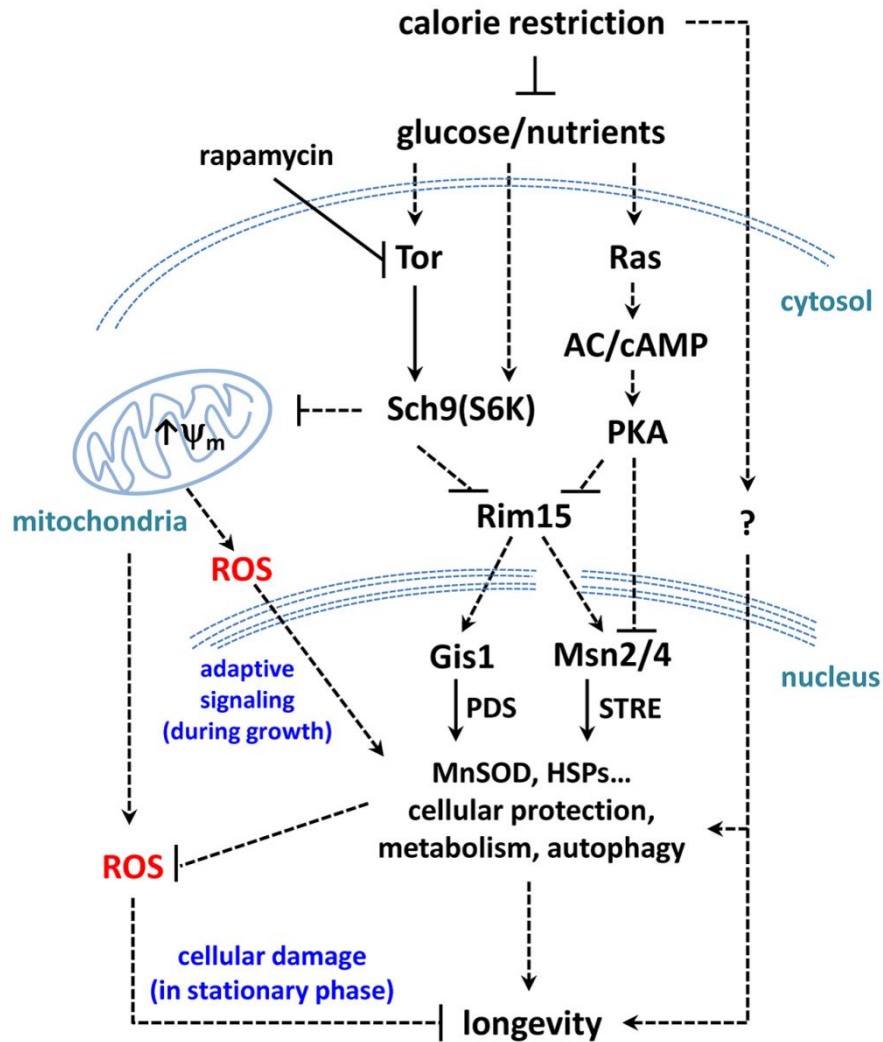


Figure 8. Two Major pathways of chronological life span regulation in yeast. Sch9, Tor, and Ras converge on the protein kinase Rim15 to control nutrient-sensing pathways. The down-stream pathways are activated by the Rim15-controlled Msn2/4 and Gis1 stress-responsive transcription factors [21].

2.3.3. Aging and genomic hypomethylation

DNA methylation means the addition of methyl group, mainly to the cytosine position in dinucleotide CpG site [26]. Normally, DNA methylation is crucial for normal development and maintained genomic stability. Between 60% and 90% of CpG dinucleotides are methylated in the mammalian genomes [27]. Moreover, DNA methylation is intrinsically controlled with processing in chromatin remodeling. Chromatin is divided in two groups which are active or euchromatin and inactive or heterochromatin. Methylated DNA is associated with inactive chromatin (heterochromatin). In contrast, active chromatin is characterized by the loose structure and lack of DNA methylation [28] (Figure 9).

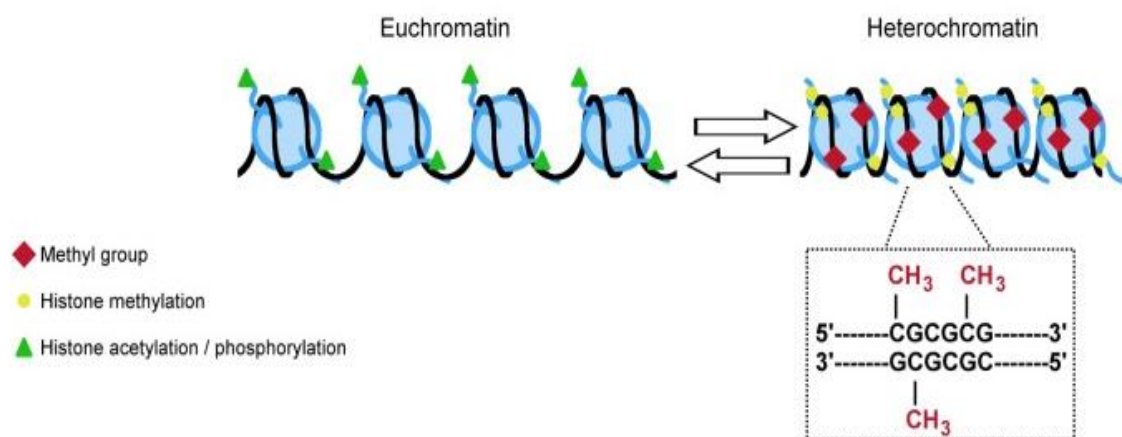


Figure 9. Euchromatin and heterochromatin are characterized by the different levels of DNA and histone methylation, acetylation, and phosphorylation. DNA hypermethylation typically links to heterochromatin while hypomethylation regularly links to euchromatin [29].

DNA methylation can be altered by loss of DNA methylation throughout the genome. The precise status of DNA methylation is balanced in mature cells but this balance is strongly shifted to DNA hypomethylation in aging cells [30]. DNA demethylation could also have an influence on aging cells [31]. The first evidence of the age-dependent loss of genome methylation found that during ontogenesis, the content of 5-methylcytosine (5meC) in DNA isolated from the diverse organs of humpback salmon was decreased [32, 33]. This finding was supported by later studies that age-dependent DNA hypomethylation was found in many mammalian tissues, embryonic tissues, and newborn animals were gradually the amount of 5meC in DNA isolated from aging cells are decreased [34, 35]. Vanyushin *et al.* found that human fibroblasts treated with demethylating agent 5-aza-2'-deoxycytidine (5-aza-dC) to induce DNA demethylation have a considerably shortened life span [36, 37]. Loss of DNA methylation during the aging process was primarily associated with demethylated DNA that lead to genomic instability (Figure 10). Moreover, the methylation status of numerous single-copy genes decreased with age in the brain, liver, spleen, mammary glands, and other tissues.

Genomic hypomethylation is a main characteristic of aging cells. In 2010, Jintaridh *et al.* found that cells quickly lost DNA methylation when the organism enter middle age, and found that DNA methylation can be increased in the elderly (Figure 11) [38]. So if we can prevent the loss of DNA methylation, we may be also to prevent aging.

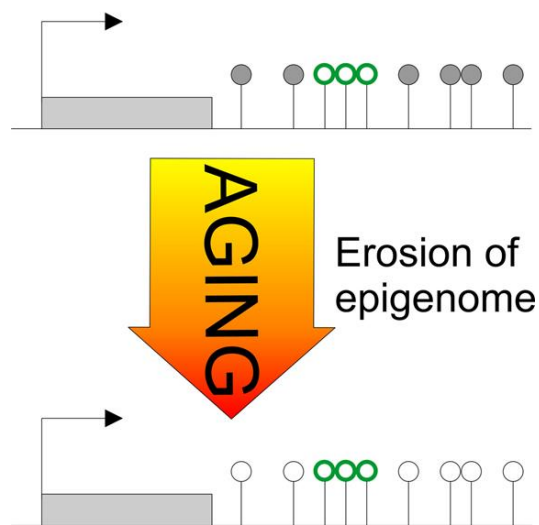


Figure 10. Loss of DNA methylation in aging. Early aging induces hypomethylation centers on CpG sites. Age-associated hypomethylation pattern shows reduced (white circle) the methylation levels from young cell (gray circle) [39].

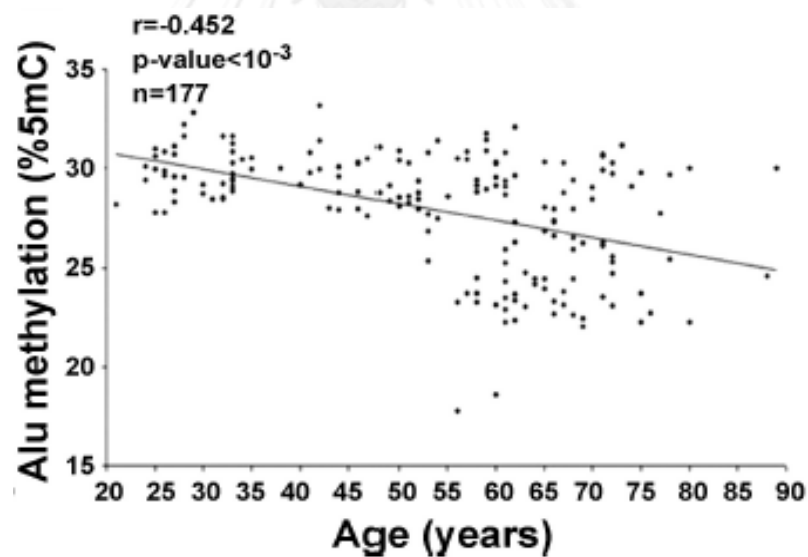


Figure 11. Association between age and IRS methylation of Alu. That cells quickly lost DNA methylation when entering middle age, significance of correlation coefficients (r) between age and IRS methylation was set at $P < 0.05$ [38].

CHAPTER III

MATERIALS AND METHODS

Yeast strains, media and growth conditions

Yeast strains used in this study are listed in Table 1. Asynchronous yeast cultures were grown in YPD (Sigma, USA) to log phase (OD₆₀₀ 0.4-0.6). For the cell cycle experiments, yeast cells were arrested at G₀, G₁, S, and M phase by culturing in YP medium containing 2% raffinose (Sigma, USA) for 48 hours, YPD in the presence of 5 μ M α -factor (Sigma, USA), of 0.2 M hydroxyurea (Sigma, USA), and of 15 μ g/ml nocodazole (Sigma, USA) for 180 minutes at 30°C, respectively. Cell cycle phases were confirmed by phase-contrast microscopy. For the G₀ phase, most cells were small and without buds. Cells arrested in the G₁ phase had enlarged schmoo morphology. In the early stages of the S phase, cells had large buds and short mitotic spindles. Finally, cells had large buds but no mitotic spindles in the M phase [40]. Apoptosis was induced by the addition of 175 mM acetic acid (pH 3.0) to the YPD for 200 minutes at 30°C [41]. Nine independent preparations of G₀ cells from each mutant strain were used in all subsequent experiments to determine the RIND-EDSB levels. In most strains, over 80% of the cells were unbudded. (Proportions of unbudded, small budded and large budded cells of all strains are shown in supplementary Table 1). To inhibit the activity of histone deacetylases (HDACs), triplicates of stationary cultures were treated with 10 μ M of trichostatin A (TSA; Sigma, USA) for 4 hours [2, 42]

For chronological aging experiments, yeast cells were grown in YP + 2% glucose at 30°C 250 rpm overnight and switched to YP + 2% raffinose for 2 days before being washed and resuspended in sterile ddH₂O. For caffeine treatment to inhibit the activity of ATM/ATR-dependent DNA repair pathways, yeast cells were treated with 10 mM of caffeine (Sigma, USA) 48 hours [44]. Cell viability was monitored by plating an aliquot of the culture on YPD plate every 10 days until cell viability reached approximately 10%.

For HO induction and caffeine treatment, JKM179 strain was used in this method. JKM179 was a yeast strain that carries galactose-inducible HO endonuclease and a specific irreparable HO specific site at the MAT locus with a deletion of the homologous sequence at HML and HMR loci [43]. Therefore the galactose-induced DSB cannot be repaired by the process of homologous recombination. The expression of HO endonuclease was induced by the addition of galactose to the media for 3 hours. After induction, cells were washed with sterile deionized water and placed in media containing 2% glucose to repress HO expression. Approximately 1×10^8 cells/ml were resuspended in sterile deionized water and were kept at 30°C in a rotary shaker at 250 rpm. To inhibit the activity of ATM/ATR-dependent DNA repair pathways, JKM179 cells were treated with 10 mM of caffeine (Sigma, USA) 48 hours [44]. After 3 days of HO induction, cell viability was measured by plating an aliquot of the culture on YPD plate and counting the number of visible colony forming units (CFUs) after 3 days of incubation at 30°C.

Table 1. Yeast strains used in this study.

Yeast strains	Genotype	Source
BY4741	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0</i>	G.R. Fink
<i>ybr136wΔ (mec1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mec1Δ sml1Δ</i>	M.C. Keogh [1]
<i>ybl088cΔ (tel1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 tel1Δ::KanMX</i>	Open biosystems
<i>ymr224cΔ (mre11Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 mre11Δ::KanMX</i>	Open biosystems
<i>ymr284wΔ (yku70Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yku70Δ::KanMX</i>	Open biosystems
<i>ymr106cΔ (yku80Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 yku80Δ::KanMX</i>	Open biosystems
<i>ylr265cΔ (nej1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nej1Δ::KanMX</i>	Open biosystems
<i>yer095wΔ (rad51Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad51Δ::KanMX</i>	Open biosystems
<i>ypr052cΔ (nhp6aΔ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nhp6aΔ::KanMX</i>	Open biosystems
<i>ybr089c-aΔ (nhp6bΔ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nhp6bΔ::KanMX</i>	Open biosystems
<i>ydl002cΔ (nhp10Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 nhp10Δ::KanMX</i>	Open biosystems
<i>ypr065wΔ (rox1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rox1Δ::KanMX</i>	Open biosystems
<i>ykl032cΔ (ixr1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ixr1Δ::KanMX</i>	Open biosystems
<i>ydr174wΔ (hmo1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 hmo1Δ::KanMX</i>	Open biosystems
<i>ymr072wΔ (abf2Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 abf2Δ::KanMX</i>	Open biosystems

Table 1. Yeast strains used in this study.

Yeast strains	Genotype	Source
<i>ycr077cΔ (pat1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 pat1Δ::KanMX</i>	Open biosystems
<i>yol006cΔ (top1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 top1Δ::KanMX</i>	Open biosystems
<i>ylr234wΔ (top3Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 top3Δ::KanMX</i>	Open biosystems
<i>ygl175cΔ (sae2Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sae2Δ::KanMX</i>	Open biosystems
<i>ykl113cΔ (rad27Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rad27Δ::KanMX</i>	Open biosystems
<i>ykl114cΔ (apn1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 apn1Δ::KanMX</i>	Open biosystems
<i>yhl022cΔ (spo11Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spo11Δ::KanMX</i>	Open biosystems
<i>ykr101wΔ (sir1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sir1Δ::KanMX</i>	Open biosystems
<i>ydl042cΔ (sir2Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sir2Δ::KanMX</i>	Open biosystems
<i>ylr442cΔ (sir3Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sir3Δ::KanMX</i>	Open biosystems
<i>ydr227wΔ (sir4Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 sir4Δ::KanMX</i>	Open biosystems
<i>ynl330cΔ (rpd3Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rpd3Δ::KanMX</i>	Open biosystems
<i>ynl021wΔ (hda1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 hda1Δ::KanMX</i>	Open biosystems
<i>ydr334wΔ (swr1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 swr1Δ::KanMX</i>	Open biosystems
<i>yol012cΔ (htz1Δ)</i>	<i>MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 htz1Δ::KanMX</i>	Open biosystems

High-Molecular weight (HMW) DNA preparation for yeast and Ty1-EDSB-LMPCR

To prepare HMW DNA, yeast cells were treated with 1 mg/ml lyticase (70 U/mg) (Sigma, USA) for 2 hours and embedded in 1% low melting point agarose at a concentration of 2×10^8 cells per plug. Embedded cells were digested in 400 μ l of digestion buffer (1 mg/ml proteinase K, 50 mM Tris, pH 8.0, 20 mM EDTA, 1% sodium lauryl sarcosine) at 37°C overnight. The plugs were rinsed 6 times in TE buffer for 40 minutes. EDSBs with cohesive ends were polished by incubating with T4 DNA polymerase (New England Biolabs, Beverly, MA, USA) and dNTPs for 1 hour. The enzyme was inactivated by adding EDTA at a final concentration of 20 mM, for 5 minutes, and rinsed 6 times in TE buffer for 40 minutes. The modified LMPCR linkers were prepared from oligonucleotides: 5'-AGGTAACGAGTCAGAC CACCGATCGCTCGGAA GCTTACCTCGTGGACGT-3' and 5'-ACGTCCACGAG-3' (Sigma, Singapore) [2]. The linkers (50 pmol) were ligated to the polished EDSB ends in the HMW DNA preparations using T4 DNA ligase (New England Biolabs) at 25°C overnight. Linker-ligated DNA was then extracted from the agarose plugs using a QIAquick gel extraction kit (Qiagen, Basel, Switzerland) [2]. The quantity of EDSBs was measured by real-time PCR using an ABI PRISM® 7500 system (Applied Biosystems, Carlsbad, CA, USA) with Ty1 primer 5'-AATGGAATCCCAACAATTATCTCAA-3' (Biodesign, Thailand), the linker primer and the Taqman probe homologous to the 3' linker sequence (6-fam) ACGTCCACGAGG TAAGCTTCCGAGCGA (tamra, phosphate) [Sigma, Singapore] [2]. DNA amplification was performed with 0.2 μ M of each primer, 0.3 μ M Taqman probe, 0.025 U of HotStarTaq, 1x TaqMan® Universal PCR Master Mix (Applied Biosystems) and 10 ng of ligated DNA. Initial denaturation was at 95°C for 15 minutes, followed by denaturation at 95°C for 5 seconds, annealing at 58°C for 5 seconds, and extension for 2 minutes at

69°C for up to 60 cycles, with quantification after each extension step [2]. To normalize potential differences in the amount of Ty1 per genome, genomic DNA of each mutant was used as its own control DNA. Control DNA was digested with AluI and ligated to the LMPCR linkers. The numbers of EDSBs were compared with the AluI-digested ligated control DNA and reported in arbitrary units of Ty1-EDSB-LMPCR templates per genome. The Ty1-EDSB-LMPCR units were estimated from the number of AluI sites in yeast genome, and converted to the number of EDSBs.

Yeast nuclei isolation and intranuclear linker ligation

To isolate the nuclei, yeast cells were treated with 1 mg/ml lyticase (70 U/mg) (Sigma, USA) for 2 hours and digested in SPC digestion buffer (1M sorbitol, 20 mM 1,4-piperazinediethanesulfonic acid (Pipes), pH 6.3, 0.1 mM CaCl₂), and the nuclei were collected in SPC buffer with 9% ficoll, as previously described [3]. LMPCR linkers were ligated in situ with the nuclei preparation and Ty1-EDSB-LMPCR were performed as described for HMW DNA [2].

HMGB1si cells and RT-PCR of HMGB1

The commercially available oligonucleotides HSS142453, HSS142454, and HSS142455 from the Stealth RNAi system (Invitrogen) were used for the specific knockdown of HMGB1 gene in HeLa cells. Transfection was carried out with the Lipofectamine2000 transfection reagent (Invitrogen). A negative control siRNA (Invitrogen) was transfected in parallel. After 72 hours, an aliquot of transfected cells was collected to determine the level of HMGB1 mRNA. RNA extraction was performed and 5 µg of RNA was reverse transcribed with RevertAid™ First Strand cDNA Synthesis Kit (Fermentas). The total cDNA of each sample was analyzed in

triplicate by a quantitative - comparative CT (DDCT) study in an ABI PRISM® 7500 instrument (Applied Biosystems, Carlsbad, CA, USA) with the HMGB1 forward primer 5'ATATGGCAAAGCGGACAAG-3' and the HMGB1 reverse primer 5'GCAACATCACCAATG GACAG-3' [Sigma, Singapore][4]. The relative expression of HMGB1 was normalized to GAPDH expression.

HMW DNA preparation for HMGB1si cells and LINE-1-EDSB-LMPCR

HMW DNA was prepared as previously described [2]. Approximately 5x10⁵ cells were embedded in 1% low melting point agarose, lysed and digested in 400 µl of digestion buffer (1 mg/ml proteinase K, 50 mM Tris, pH 8.0, 20 mM EDTA, 1% sodium lauryl sarcosine) at 37°C overnight. EDSB end polishing and linker ligation were carried out as described for yeast HMW DNA.

For human cells, Long Interspersed Element1 (LINE-1 or L1) sequences were used instead of Ty1 for LMPCR. The number of L1-EDSBs was measured by real-time PCR as previously described [2]. Control HeLa DNA was digested with EcoRV/AluI and ligated to the LMPCR linkers. The amounts of EDSBs were compared with the EcoRV/AluI-digested ligated control DNA and the arbitrary units of L1-EDSB-LMPCR templates per genome were converted to EDSB numbers.

COmbined Bisulfite Restriction Analysis (COBRA)-LINE-1 and COBRA-LINE-1-EDSB

Nested COBRA-LINE-1 LMPCR was used to measure the methylation of LINE-1 sequences proximal to EDSBs [5]. In the first round, we performed a PCR with 5 µl (approximately 250 ng) of bisulfite-treated DNA, 0.3 µM of AMETLINKP primer (5'-GTTTGGAAAGTTTATTTTGTGGAT-3') and 0.3 µM of LINEMSPCR 270 & 280 reverse primer

(5' RTAAAACCCTCCRAACCAAATA TAAA3'). The PCR conditions were 95°C for 15 minutes, followed by 30 cycles at 95°C for 1 minute, 48°C for 1 minute, and 72°C for 2 minutes and 1 final cycle at 72°C for 7 minutes. In the second round, we performed PCR with 2 µl of PCR amplicons from the first PCR step and 0.3 µM of the L1 primers, the FCOBRALINE-I forward primer (5' CGTAAGGGGTTAGGGAG TTTTT 3') and the LINEMSPCR 270 & 280 reverse primers. The PCR conditions were 95°C for 15 minutes, followed by 40 cycles at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute and 1 final cycle at 72°C for 7 minutes. The restriction analysis was performed by digesting 8 µl of the PCR products with 2 U of TaqI and 2 U of TspI at 65°C overnight; the digested products were then electrophoresed in an 8% non-denaturing polyacrylamide gel and stained with SYBR green. DNA methylation levels were determined as previously described [6]. Briefly, the intensities of 4 out of 5 COBRA LINE-1 fragments (the 160, 98, 80 and 62 bp fragments but not the 18 bp fragments) in the polyacrylamide gel were quantified using a phosphoimager and ImageQuant Software (Molecular Dynamics, GE Healthcare, Slough, UK). The intensity of each band was divided by its double-stranded DNA length to normalize the intensity to fragment sizes as follows: 160 bp/160 (A), 98 bp/94 (B), 80 bp/78 (C) and 62 bp/62 (D). The LINE-1 methylation level was calculated by the following formula: $(C + A) / (C + A + A + B + D) \times 100$. DNA from HeLa cells was used as a control to normalize the inter-assay methylation variation for all of the experiments.

Statistical analyses

Statistical analyses were performed using Student's t-test for data with normal distribution or a Mann-Whitney test for data that do not have normal distribution, as specified.

Library preparation and sequencing

After ligation with two linkers (Figure12), HMW DNA was subjected to 60 cycles of PCR with two primers, First-L-F (5'-AGGTAACGAGTCAGACCACCGA-3') and Second-L-R (5'-GGTACCGGTAGGGCCTACGGGT-3' (Sigma, Singapore) with an annealing temperature of 50°C. PCR product was purified with QIAquick PCR Purification Kit (Qiagen, Switzerland) and sequenced by Ion Torrent sequencer (Ion Torrent™ Personal Genome Machine® (PGM), Life Technologies, USA)

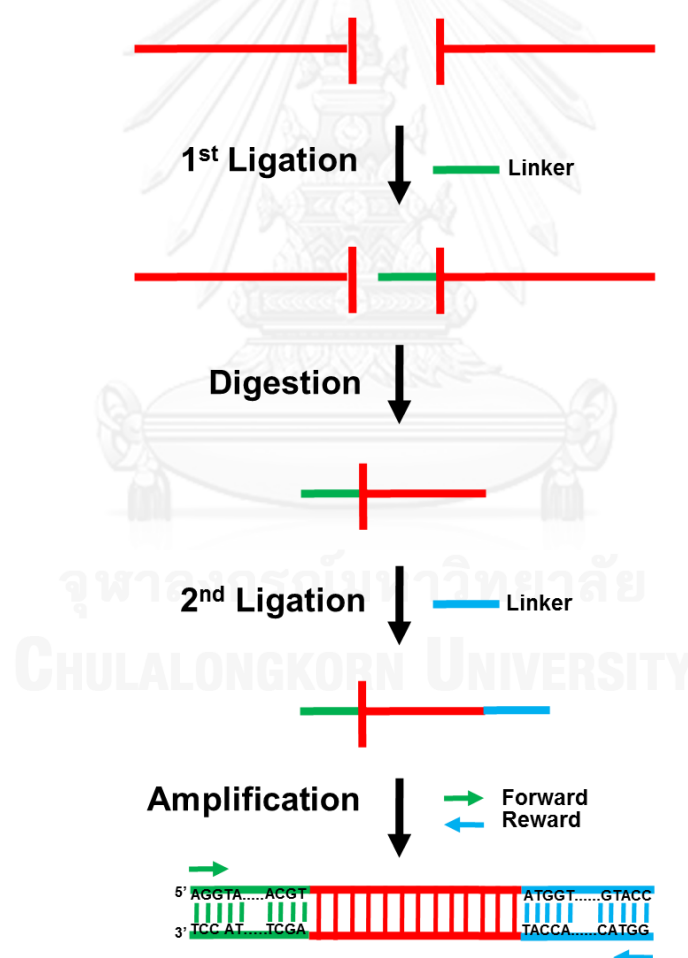


Figure 12. Schematic representation of DNA preparation for Ion Torrent sequencing showed EDSBs in HMW DNA ligated with two linkers oligonucleotides.

Data processing and mapping

We performed the same steps as in our another study. In brief, we trimmed the adaptor from both ends by aligning the adaptor sequences to reads using “pairwise Alignment” function from BioStrings package. Taking only reads with head adaptor sequence without the last ten bases missing, we blasted them against *S. cerevisiae* strain S288C genome (BLAST 2.2.27). We further mapped the reads that were previously unaligned by BLAST against our 179 whole genome sequence data using BLAST. We considered the aligned results as uniquely mapped if the preceding four bases before the breakpoint of the mapped whole genome reads were all the same. We combined these two uniquely aligned results for further analysis.

Data Analysis of Sequence logo

We extracted the first 50 bases of each uniquely mapped read and connected them to its preceding 4 bases. These 54 bases were then used to plot a variation of sequence logo where all columns have the same height. An R “seqLogo” package was used for plotting. For statistical analysis, we calculated two sets of Fisher’s exact test p-value. For the set, we constructed a 2x2 contingency as follows: rows represented the number of breaks in the two samples (i.e., JKM179 vs. JKM179-CA and JKM179-CA vs. JKM179-HO-CA-6) and columns represented the number of CGK breaks and non-CGK breaks. Fisher’s exact test p-value were then computed from these contingency tables.

CHAPTER IV

RESULTS

Establishment of Ty1-EDSB-LMPCR assay

We previously established a method to detect EDSBs that occur rather rarely in the human genome. In this assay, genomic DNA was extracted from human cells by the high molecular weight DNA (HMW) preparation protocol [2]. Existing DNA breaks preserved in the genome were first ligated to linker oligonucleotides. They were then detected by PCR using a pair of primers complementary to the linker sequence and the LINE-1 repetitive sequences in human genome. Because this method can detect low numbers of EDSBs occurring in proximity to the LINE-1 sequences, it is called “L1-EDSB-LMPCR” [2]. Here, we modified this method to measure EDSBs in the yeast genome by taking advantage of the Ty1 sequences (as opposed to the LINE-1 sequence in human cells), and called this assay “Ty1-EDSB-LMPCR”. Ty1 sequences are abundant repetitive sequences that intersperse throughout the yeast genome. The presence of EDSBs was quantitatively analyzed by real-time PCR using primers complementary to both the linker and the Ty1 sequences and a Taqman probe complementary to the linker oligonucleotides (Figure 13A). Thus, this assay favourably detected EDSBs located near Ty1 (Ty1-EDSBs). To be efficiently amplified by real-time PCR, in human cells, EDSBs should locate approximately within 300 bp, from LINE1 sequences [9]. Analyses of varying amounts of control AluI digested DNA showed that our assay quantitatively detected DSB ends and did not detect any Ty1-EDSB-LMPCR product without linker ligation (Figure 13B).

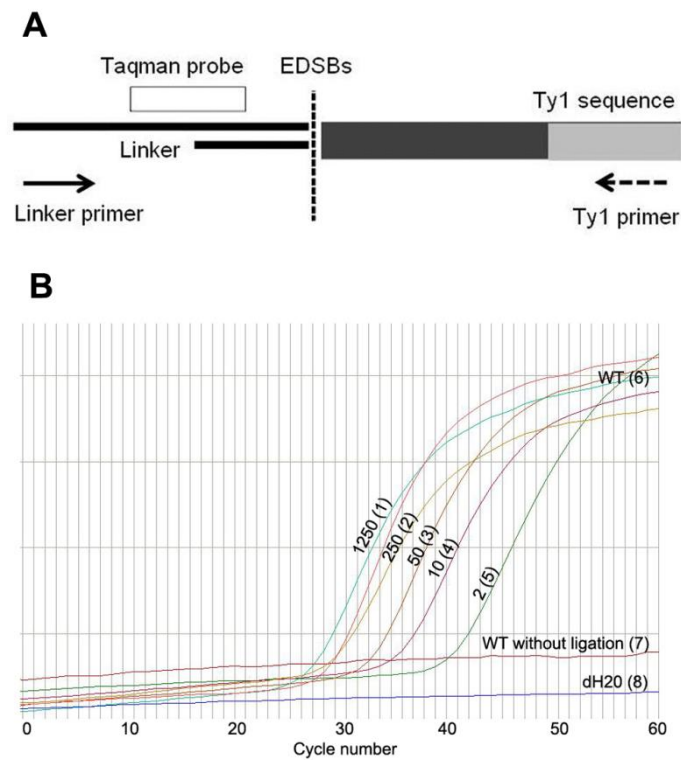


Figure 13. Ty1-EDSB-LMPCR assay. (A) A schematic representation of the Ty1-EDSB-LMPCR assay shows EDSBs in HMW DNA ligated with linker oligonucleotides. EDSBs located close to the Ty1 sequence are measured by real-time PCR using the Ty1 primer (dash arrow), linker primer (solid arrow), and the Taqman probe complementary to the linker sequence (white bar). (B) A representative real-time PCR result showed the level of Ty1-EDSB-LMPCR products in the HMW-DNA of a wild-type yeast strain (WT, dashed line) relative to control DSBs generated by Alu1 digested restriction enzymes (equivalent to 2, 10, 50, 250 and 1250 cells/ μ l; dark green, purple, brown, yellow, and light green lines, respectively). Note that no Ty1-EDSB-LMPCR product could be detected in HMW without linker ligation (magenta line) and in water control (dH₂O, blue line).

Detection of EDSBs in yeast

We used the Ty1-EDSB-LMPCR assay to estimate the levels of EDSBs during the G₀, G₁, S and M phases of the cell cycle (Figure 14A). Here, we estimated the total amount of EDSBs under an assumption that the AluI restriction endonuclease generates DSBs every 256 bp on average, and that EDSBs are distributed equally throughout the genome. The results were similar to the findings in human cells; the level of EDSBs was highest in S phases, but still detectable at a lower level in G₀ phase (Figure 14A). Next, we tried to verify if the fragmented DNA from apoptotic cells could interfere with the detection of the genuine EDSBs. Therefore, we tested whether Ty1-EDSB-LMPCR could detect fragmented DNA generated by apoptosis. Yeast cells undergo apoptotic cell death, with characteristic DNA fragmentation, upon treatment with acetic acid [10]. We found that Ty1-EDSB-LMPCR did not detect fragmented apoptotic DNA, prepared by the HMW DNA extraction protocol (Figure 14B). This is likely due to the nature of the DSB ends of apoptotic DNA fragments which was reported to be staggered, not efficiently blunted by Klenow treatment, and thus are not ligated to the linkers [10]. Ty1-EDSB-LMPCR specifically detected only EDSBs from the genomic DNA, which was added to the sample to normalize the total amount of DNA, and did not detect any signal from the sample containing 100% fragmented DNA from apoptotic cells (Figure 14B, at 100% Apoptotic DNA). This indicated that apoptotic DNA did not interfere with the Ty1-EDSB-LMPCR measurement of EDSBs. Therefore, the Ty1-EDSB-LMPCR is an accurate and sensitive method to study EDSBs and RIND-EDSBs in the yeast genome.

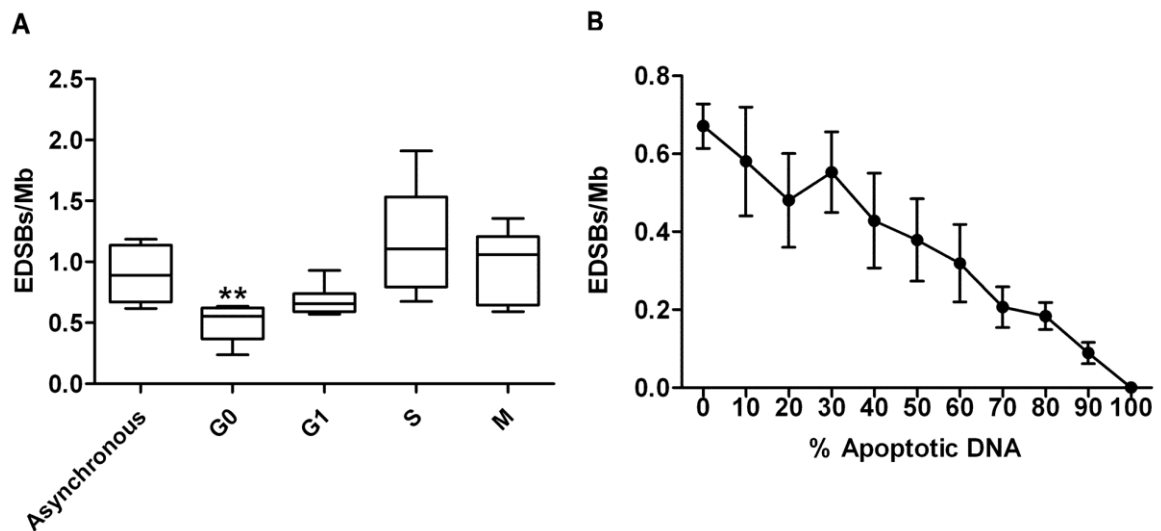


Figure 14. EDSBs in various phases of the cell cycle. (A) EDSBs were measured in asynchronous culture and in yeast cells arrested in G0, G1, S, and M phases. The levels of EDSBs from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. There was a significant decrease in EDSBs in G0 cells compared to asynchronous culture, (** $P < 0.001$, Mann-Whitney test). (B) HMW DNA was isolated from apoptotic yeast cells, mixed with control DNA at varying percentages, and analyzed by Ty1-EDSB-LMPCR. The graph represents the mean levels of EDSBs with error bars representing standard deviations. The Ty1-EDSB-LMPCR assay could not detect DNA fragments from apoptotic cells (at 100% apoptotic DNA). Furthermore, apoptotic DNA fragments did not interfere with quantitative measurement of EDSBs.

We also determined if the signals we observed by Ty1-EDSB-LMPCR were due to our DNA preparation protocol. Similarly to the experiments performed in human cells [2], we compared the levels of EDSBs of genomic DNA prepared with different protocols, including an in-gel (HMW-DNA), a liquid DNA, and a combined preparation protocols (Figure 15A). Only a minimal difference was derived when we subtracted the DSBs levels of DNA prepared with the liquid DNA protocol (cell→liquid) from that of DNA prepared with the combined in-gel followed by liquid DNA protocol (cell→gel→liquid) (Figure 15B). This suggests that adding an in-gel preparation step did not increase the number of DSBs significantly. Thus, our method for in-gel preparation of genomic DNA produced an insignificant number of breaks, as measured by Ty1-EDSB-LMPCR.

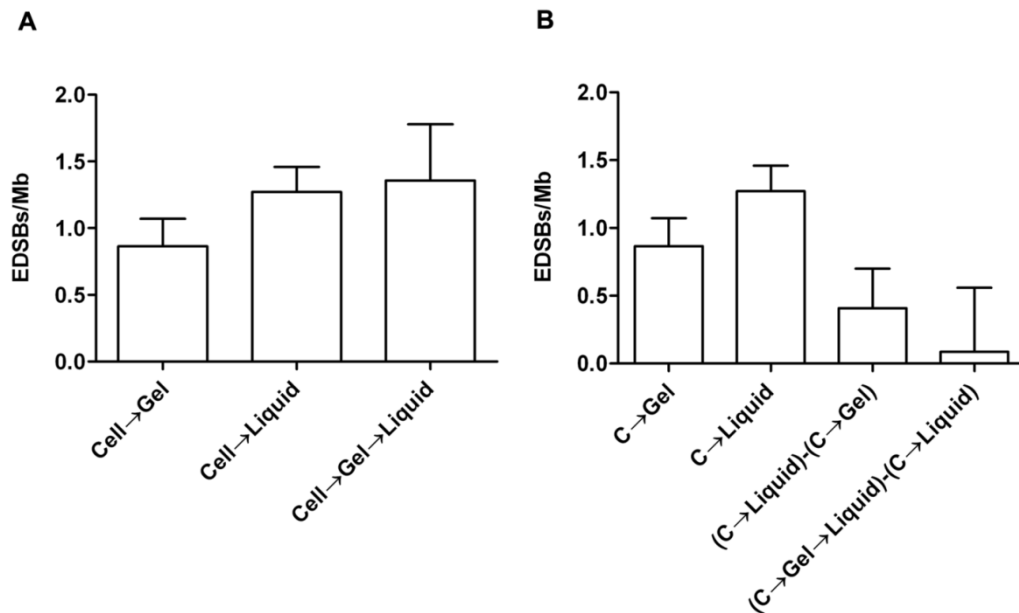


Figure 15. EDSBs were detected in different DNA preparations including HMW DNA (cell→gel), liquid DNA (cell→liquid), and liquid DNA extracted from in-gel HMW DNA (cell→gel→liquid). (A) The levels of EDSBs from different DNA preparation methods. (B) Subtracted DSBs levels between liquid DNA and other methods. When comparing cell→gel→liquid with cell→liquid, adding in gel preparation step did not increase the number of DSBs significantly. The average levels of EDSBs from 9 independent experiments are shown as histograms with error bars representing SEM.

To further confirm that Ty-EDSBs are not artifacts from HMW DNA preparation, we isolated the yeast nuclei and performed linker ligation in situ. Using this intranuclear ligation method, the DNA was protected in the nuclear membrane. We compared the levels of Ty-EDSBs of various mutant yeast strains that harbor different levels of Ty-EDSBs (see below) using the HMW DNA preparation and the intranuclear ligation protocols (Figure 16A and B, respectively). The results demonstrated that the assay could detect DNA breaks within the nuclei. Importantly, we observed the same pattern of differences in the levels of EDSBs among different yeast mutant strains using the two techniques, suggesting that the HMW-DNA-based assay could reflect the situations in the nuclei. Although the levels of EDSB detected with intranuclear ligation were lower than those detected with the HMW DNA preparations, this could be expected from the lower efficiency of the ligation reaction in the complex nuclear architecture. It is also unlikely that the mutations in various genes would lead to the same effect in these two different protocols if they affect artificially induced DNA breaks. Therefore, we believe that the assay provide a sensitive means to measure the low level of randomly occurring EDSBs that reflect the levels in vivo.

In our previous study in human cells, when we treated the cells with a histone deacetylase inhibitor trichostatin A (TSA), histones became hyperacetylated and the retained RIND-EDSBs were repaired [9]. However, the levels of RIND-EDSBs increased when cells were treated simultaneously with TSA and inhibitors of DNA-PKcs and ATM [9]. Intriguingly, similar results were observed in yeast cells suggesting that RIND-EDSBs were also similarly regulated by chromatin structure in yeast (Figure 16C). The levels of RIND-EDSBs were decreased in TSA treated cells and increased in TSA-

treated *mec1* Δ cells lacking the ATR checkpoint kinase homolog. In this experiment, we demonstrated again that both HMW DNA preparation and intranuclear ligation protocols yielded the same results (Figure 16C and 16D).

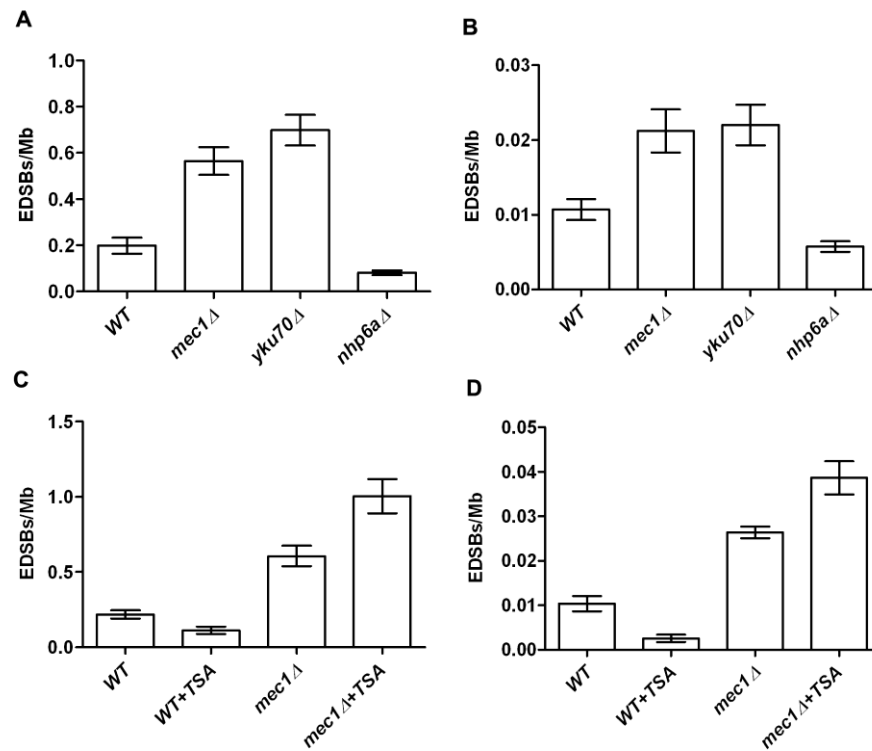


Figure 16. RIND-EDSB levels using HMW DNA preparation and intranuclear ligation protocols. (A, B) The levels of RIND-EDSBs were measured in G0 cells of WT, *mec1* Δ , *yku70* Δ , *nhp6a* Δ strains using HMW DNA (A) and intranuclear ligation (B) protocols. (C, D) The levels of RIND-EDSBs in controls and TSA-treated WT and *mec1* Δ strains using HMW DNA (C) and intranuclear ligation (D) protocols. Bar graphs represent average values and error bars represent standard deviation of triplicate experiments.

Repair of RIND-EDSBs

The sensitivity of our PCR-based DSB detection method allowed us to quantitatively analyze low levels of EDSBs in non-replicative (G0) cells. To explore the roles of DNA repair pathways in RIND-EDSB repair, we examined the levels of RIND-EDSBs in several yeast strains with deletions of genes encoding components of the DNA damage response (Figure 17). The levels of RIND-EDSBs in G0 yeast cells were significantly increased in the *mec1* Δ , *tel1* Δ , and *mre11* Δ strains, which lack key DNA damage sensor genes. We then examined the levels of RIND-EDSBs in strains lacking genes important for NHEJ repair (*yku70* Δ , *yku80* Δ , and *nej1* Δ). The levels of RIND-EDSBs were significantly increased in strains *yku70* Δ and *yku80* Δ . However, when *nej1* was deleted, there was no change in the RIND-EDSB level. During G0 phase of haploid yeast cells, we did not expect that the RIND-EDSBs would be repaired by homologous-recombination (HR). However, we observed a large increase in RIND-EDSBs in a yeast strain lacking Rad51, a key protein in the HR pathway.

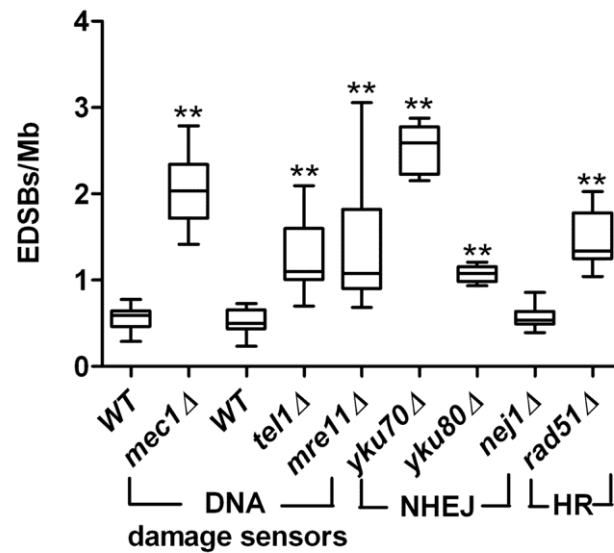


Figure 17. Levels of RIND-EDSBs in yeast strains with mutations in DSB repair pathways. The levels of RIND-EDSBs were significantly increased in G0 cells of *mec1*Δ, *tel1*Δ, *mre11*Δ, *yku70*Δ, *yku80*Δ, and *rad51*Δ but not in *nej1*Δ strains. The levels of EDSBs from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. **P < 0.001 (Mann-Whitney test).

Regulations of RIND-EDSBs by chromosomal stress, and endonucleases

Whether there is any cellular process that influences RIND-EDSB levels is still not known. Without DNA replication, the RIND-EDSBs could be a result of chromosomal stress or an endonuclease. We focused on 3 groups of genes whose functions promote DNA breaks, including topoisomerases, endonucleases and High-Mobility Group B (HMGB). We hypothesized that RIND-EDSBs would be lower in yeast strains lacking any genes involved in the production or the retention of RIND-EDSBs.

First, we examined the role of genes encoding proteins with HMGB domains. We measured the levels of RIND-EDSBs in yeast strains with deletions of each of the seven genes in the HMGB family, i.e., *NHP6A*, *NHP6B*, *NHP10*, *ROX1*, *IXR1*, *HMO1*, and *ABF2* [11]. The levels of RIND-EDSBs were reduced in all of the mutant strains tested, but significantly decreased in the *nhp6aΔ*, *rox1Δ*, *ixr1Δ*, and *hmo1Δ* strains (Figure 18). This finding led us to investigate the role of *HMGB1*, the human homolog of budding yeast *HMO1*, in human cells. Indeed, siRNA depletion of *HMGB1* in human cervical carcinoma HeLa cells also resulted in a decrease in the level of RIND-EDSBs (Figure 19).

Our previous study showed that RIND-EDSBs are preferentially retained in methylated DNA [3]. We therefore examined the methylation of genomic LINE1 (Figure 19C) and of LINE1 located close to EDSBs in HeLa cells (Figure 19D). Intriguingly, the depletion of *HMGB1* significantly decreased the methylation level of L1-EDSBs (Figure 19D). This result indicates that *HMGB1* is involved in the production or retention of hypermethylated RIND-EDSBs in HeLa cells.

Next, we examined a potential role of topoisomerases, their partners, and endonucleases in the generation of EDSBs. We found that, while the levels of RIND-EDSBs did not change in the *top1Δ*, *apn1Δ*, and *spo11Δ* strains, they significantly increased in the *pat1Δ*, *top3Δ*, *sae2Δ*, and *rad27Δ* strains (Figure 20).

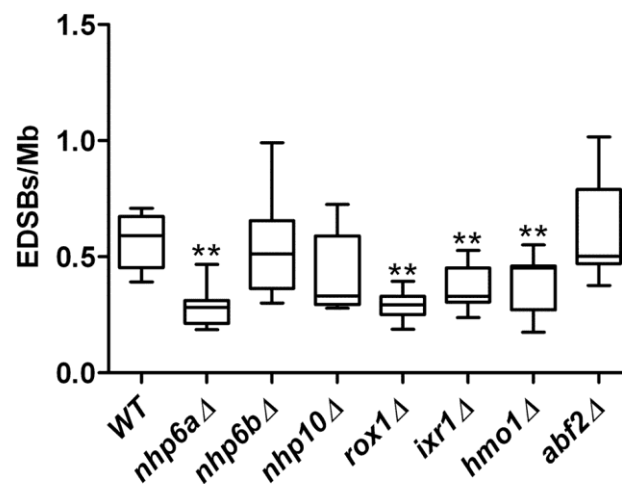


Figure 18. Levels of RIND-EDSBs in yeast strains with deletions of genes encoding proteins with the High-Mobility Group B (HMGB) domain. The levels of EDSBs were significantly decreased in G0 cells of yeast strains lacking *NHP6A*, *IXR1*, *ROX1*, and *HMO1*, suggesting that they play an important role in the production or retention of RIND-EDSBs. Nevertheless, the levels of RIND-EDSBs in *nhp6bΔ*, *nhp10Δ*, and *abf2Δ* strains were unchanged. The levels of EDSBs from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. ** $P < 0.001$ (Mann-Whitney test).

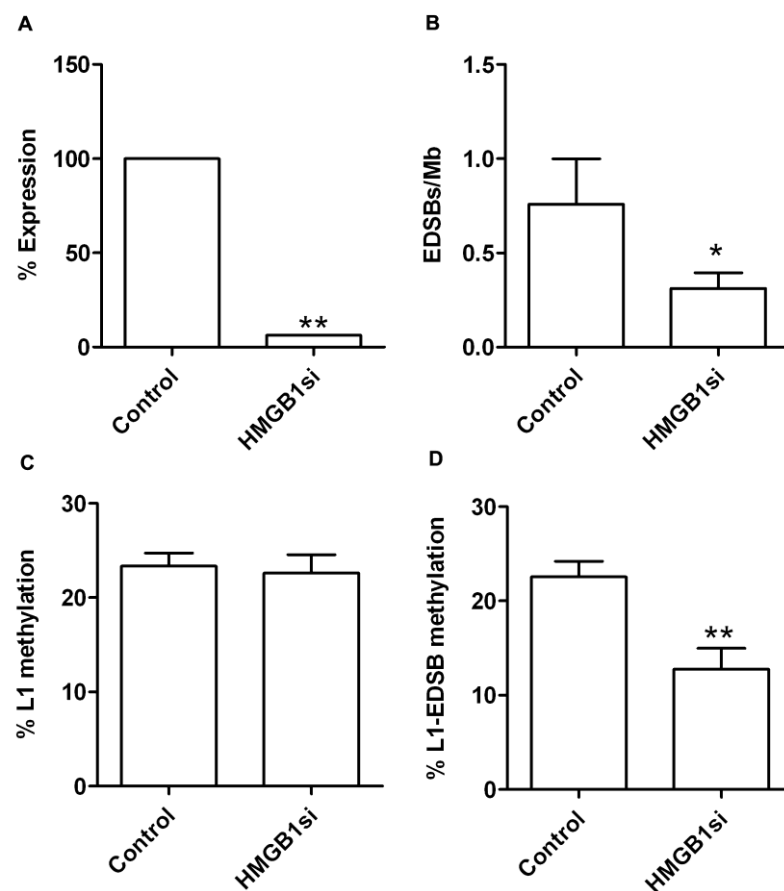


Figure 19. Levels of RIND-EDSBs in HeLa cells transfected with HMGB1 siRNA. (A) The level of expression of HMGB1 mRNA was significantly downregulated in HMGB1 siRNA HeLa cells compared to control siRNA. ** $P < 0.001$ (Paired t-test). (B) Downregulation of HMGB1 by HMGB1 siRNA in HeLa cells resulted in a decreased level of RIND-EDSBs when compared to the control. (C) Levels of L1 methylation as measured by COBRA-L1 assay were not changed in HMGB1 siRNA transfected cells. (D) L1-EDSB methylation levels were significantly lower in HMGB1 siRNA cells than in the control. The mean values from 9 independent experiments are shown as histograms with error bars representing SEM. * $P < 0.05$, ** $P < 0.001$ (t-test).

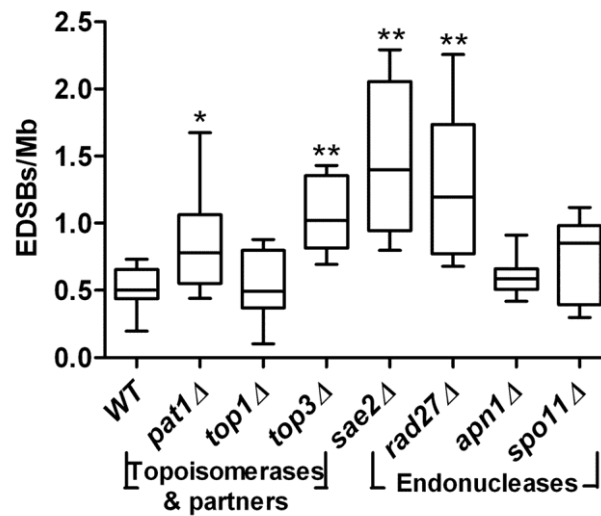


Figure 20. RIND-EDSBs in strains with mutations in genes encoding topoisomerases, their partners, and endonucleases. Deletions of genes encoding topoisomerases, their partners, and endonucleases did not reduce the levels of RIND-EDSBs in G0 cells. On the contrary, the levels of RIND-EDSBs were increased in G0 cells of *top3*Δ, *rad27*Δ, and *sae2*Δ strains. The values from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. * $P < 0.05$, ** $P < 0.001$ (Mann-Whitney test).

Heterochromatin and RIND-EDSBs levels

Our previous study demonstrated a relationship between RIND-EDSBs and chromatin acetylation [9]. To determine if these hold true in yeast, we examined the levels of RIND-EDSBs in yeast strains lacking genes important for heterochromatin formation, including the two histone deacetylases *SIR2* and *HDA1*. We observed a significantly lower level of RIND-EDSBs in *sir2Δ*, as predicted, but not in the *hda1Δ* (Figure 21A). We also examined the level of RIND-EDSBs in a mutant strain lacking *RPD3*, a distinct group of histone deacetylase. Unlike *sir2Δ*, we found a significant increase of RIND-EDSB levels in the *rpd3Δ* strain (Figure 21A).

The fact that we detected a lower level of RIND-EDSBs in the *sir2Δ* strain, together with the well-described role of Sir2 in heterochromatin formation, led us to propose that the levels of RIND-EDSBs could be indirectly controlled by Sir2 via its role in heterochromatin formation. Of note, we did not see any significant decrease in the levels of RIND-EDSBs in yeast strains lacking Sir1, Sir3, and Sir4, proteins that are related to Sir2 [12], suggesting that Sir2 may hold a specific role in the regulation of RIND-EDSBs (Figure 21B). On the other hand, when genes that suppress heterochromatin spreading, such as *HTZ1* and *SWR1* [13], were deleted, the levels of RIND-EDSBs increased significantly (*htz1Δ* and *swr1Δ*, Figure 21C).

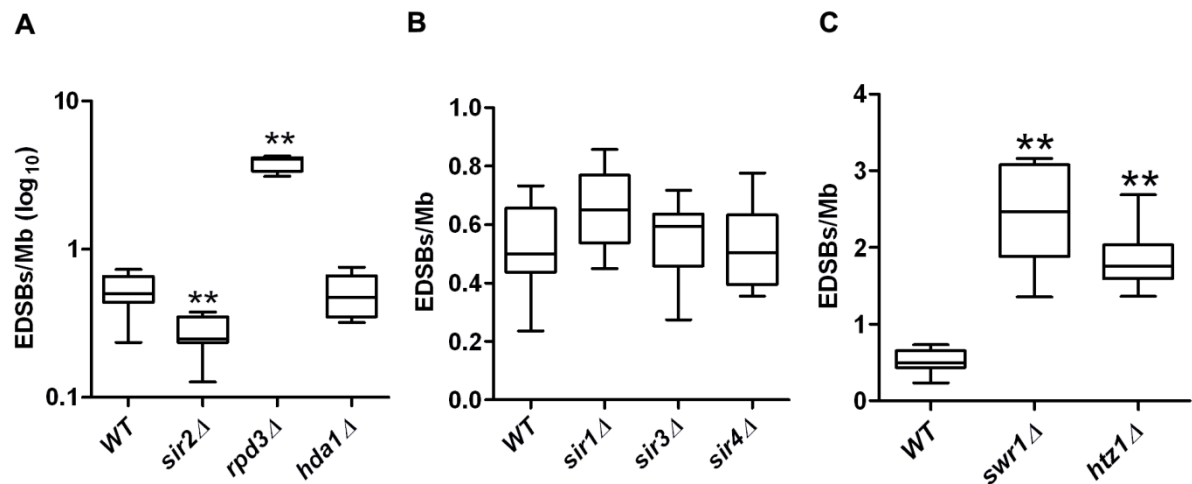


Figure 21. RIND-EDSBs and chromatin regulators. (A) The levels of RIND-EDSBs were measured in G0 cells of yeast strains lacking histone deacetylase genes, *SIR2*, *RPD3*, and *HDA1*. The level was decreased in the *sir2Δ* strain, while it was increased in *rpd3Δ* strain. (B) No significant change in the level of RIND-EDSBs was observed in yeast strains lacking the silent information regulator genes, *SIR1*, *SIR3*, or *SIR4*. (C) In contrast, deletions of *HTZ1* and *SWR1*, genes required for the prevention of heterochromatin spreading, led to significantly increased levels of RIND EDSBs. The values from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. **P < 0.001 (Mann-Whitney test).

RIND-EDSBs and chronological aging yeast cells

First, we measured the level of EDSBs in chronological aging cell and found that EDSB levels decreased in aging cells (Figure 22A, B). The levels of EDSBs were significantly decreased during 20-50 days when yeasts were kept in water. When cells treated with caffeine, a ATM/ATR inhibitor, during aging, we detected lower viability of cells than control. The decrease in viability was significantly different at day 50 (Figure 22C).



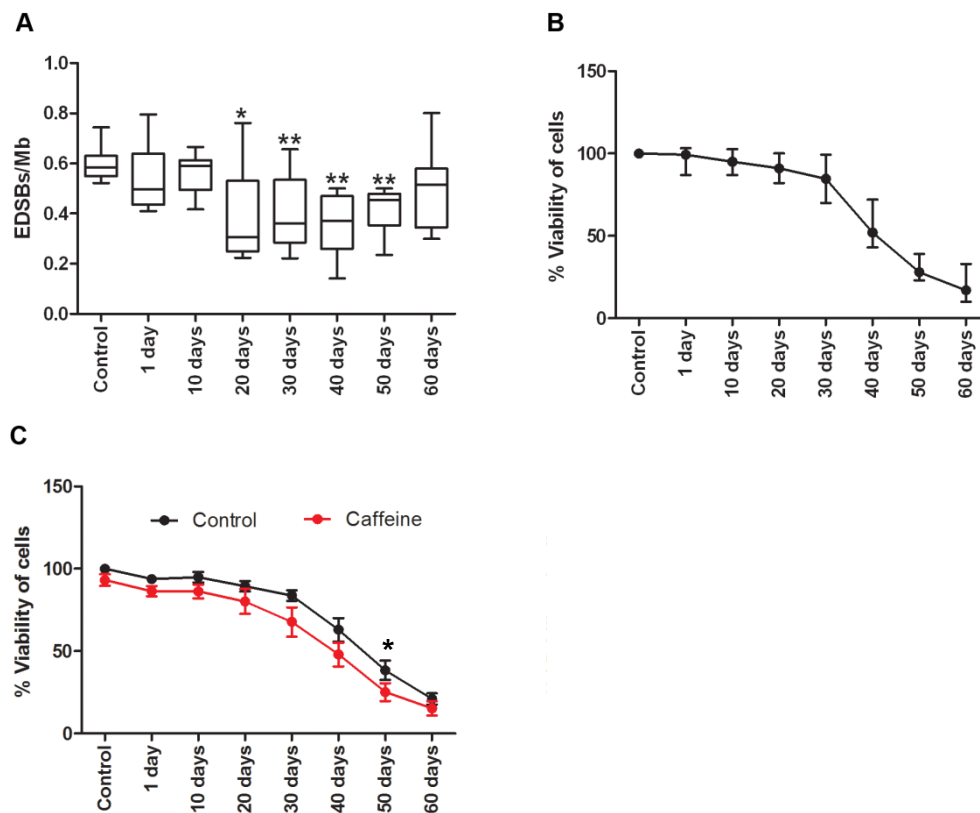


Figure 22. The levels of EDSBs decreased during chronological aging. (A) Ty1-EDSB levels and (B) %viability of wild type cells during chronological aging. (C) %viability of caffeine treatment when compare with control during chronological aging. The levels of EDSBs from 9 repeats are shown in as boxplots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. ** $P < 0.001$ (Mann-Whitney test).

To further support this result, we set up the following set of experiments in the JKM179 yeast strain. Previously, we found that the expression of HO endonuclease in JKM179 strain led to a reduction in RIND-EDSBs (Figure 23). HO induction generates a single DSB and the cells immediately try to repair the break. The reduction in EDSBs level is likely due to DSB repair activity as no reduction was observed from HO induction in *mec1*Δ strain (YAA25) (Figure 23). This suggests that the reduction in EDSB level could result from the Mec1-dependent repair of existing EDSBs when cellular DNA damage response and DNA repair activity was activated by HO-induced DSB formation.

Using this model, we could analyze the effect of low level of RIND-EDSBs on cell viability without other immediate effects of the treatment. We therefore chose to use HO induction as a means to reduce RIND-EDSBs. First, we evaluated the correlation between RIND-EDSB levels and cell viability. With 30 repeats at 3 days after HO induction, HO induced cells had lower RIND-EDSBs and lower viability (Figure 24A, B). Interestingly, we observed a strong direct correlation between the levels of change in RIND-EDSBs and change in %viability of HO induced cells (Pearson $r = 0.6$, p value < 0.0001) (Figure 24C).

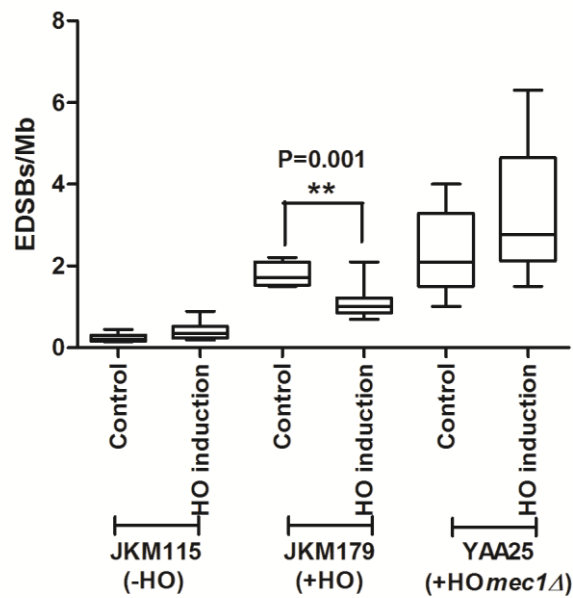


Figure 23. Levels of EDSBs in strains with inducible HO endonuclease expression. Cells were grown to log phase and HO expression was induced by the addition of galactose to the media for 3 hours and then turned off by switching to media containing glucose. JKM115 is a control strain without HO expression. JKM179 carries the HO gene under the control of galactose inducible promoter and the deletion of the homologous sequence at HML and HMR loci. YAA25 is isogenic to JKM179 with *sml1Δ* and *mec1Δ*. The levels of Ty1-EDSBs were significantly decreased in JKM179 yeast strain when HO was induced, while no significant change in the level of Ty1-EDSBs was observed in YAA25 or the control strains after HO induction. * denotes statistical significance at $p < 0.05$.

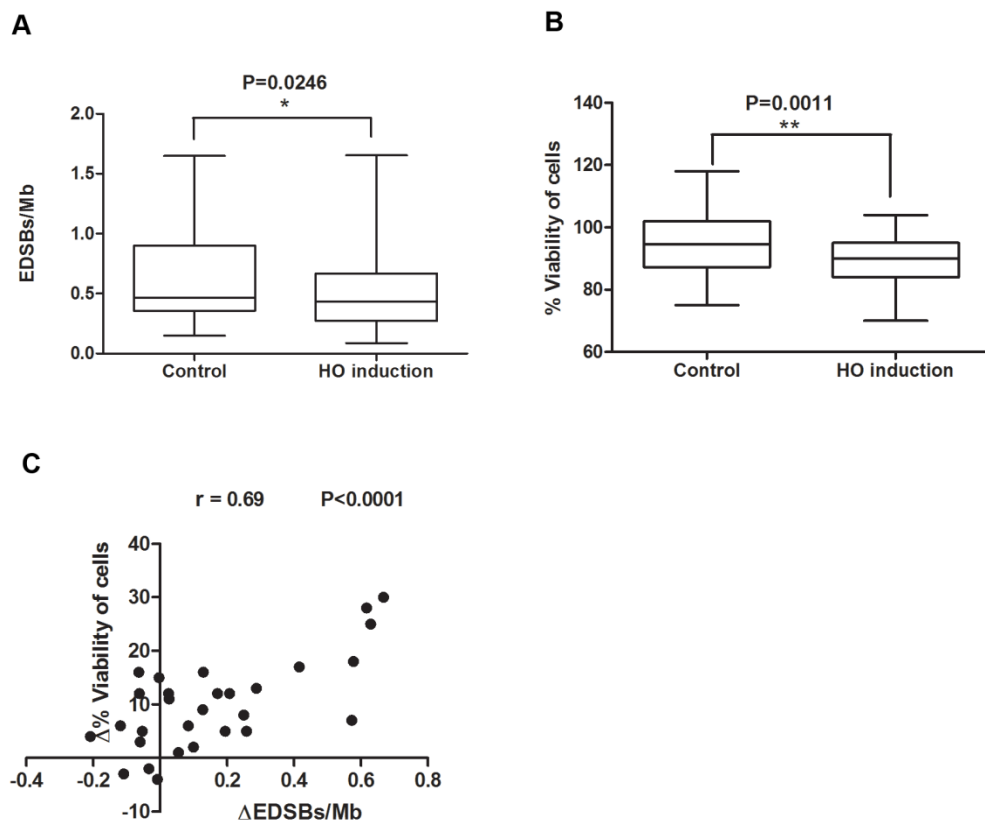


Figure 24. Correlation between RIND-EDSB levels and %viability of cells. (A) RIND-EDSB levels and (B) %viability of JKM179 cells at 3 days after HO induction. * $p < 0.05$, ** $p < 0.01$ (Mann Whitney). (C) The changes in RIND-EDSB levels (Δ EDSBs) before and after HO induction were plotted on the X axis and change in %viability of cells on the Y axis. ($n = 30$, Pearson $r = 0.6$, p value < 0.0001).

Upon caffeine treatment, the levels of RIND-EDSBs were significantly increased in JKM179 strain expressing HO endonuclease (Figure 25A). Similarly, the lower viability was observed in HO induced cells treated with 10mM caffeine (Figure 25B). As caffeine is an inhibitor of Mec1/Tel1 kinase, this result further support an earlier hypothesis that HO induction led to a reduction in RIND-EDSBs level that is Mec1/Tel1 dependent. Furthermore, the lower level of viability upon caffeine treatment in HO expressing cells and aging cells suggests that low level of RIND-EDSBs could lead to an increase in fast repair EDSBs, which could be error-prone and subsequently lead to an increased mutations and reduced viability of cells.

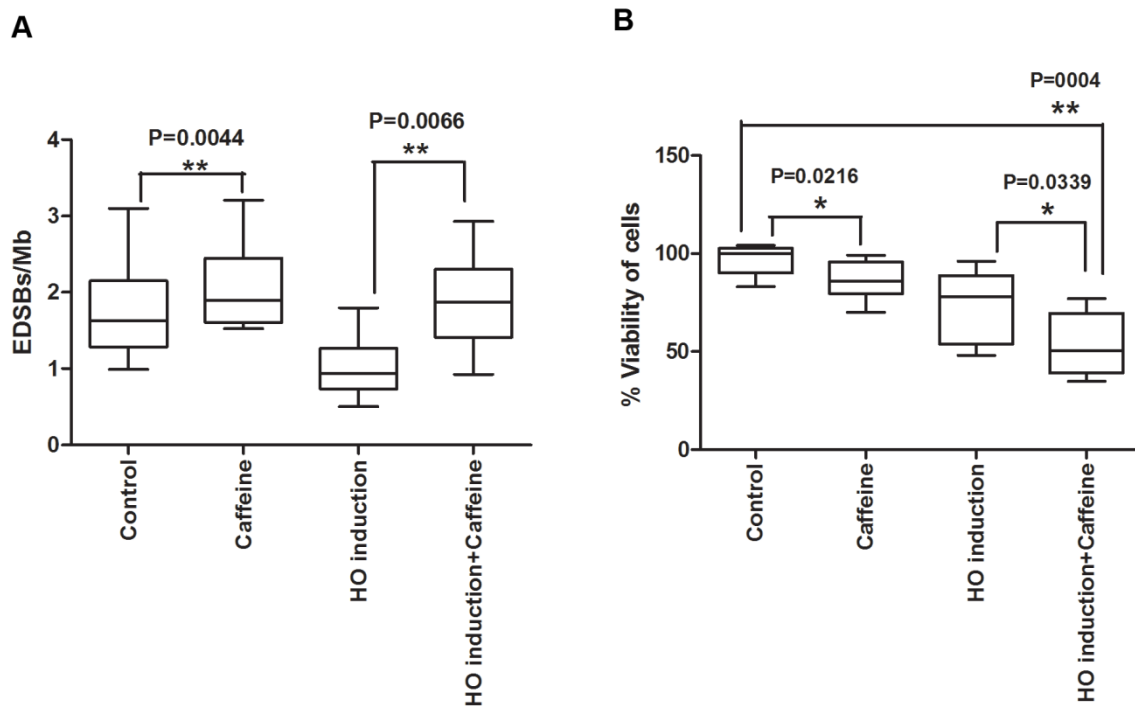


Figure 25. Caffeine treatment in JKM179 yeast strains. JKM179 carries the HO gene under the control of a galactose-inducible promoter and the deletion of the homologous sequence at *HML* and *HMR* loci. At day 3 after induction, JKM179 cells were treated with 10 mM caffeine for 48 hours. (A) The levels of EDSBs from 9 independent experiments are shown as box plots, with the boxes representing the interquartile ranges (25th to 75th percentile) and the median lines representing the 50th percentile. The whiskers represent the minimum and the maximum values. ** $P < 0.01$ (Mann-Whitney test). (B) % of cell survival in JKM179 strain at with inducible HO endonuclease expression and caffeine treatment. * $P < 0.05$, ** $P < 0.01$

Sequence pattern neighbouring break points

To investigate if the RIND-EDSBs may inversely correlate with H2AX-associated pathogenic breaks and sporadic deletion as well as insertion from non-homologous end joining errors using Ion Torrent DNA sequencing. We conducted our experiment on four samples, namely, JKM179 (179), JKM179-caffeine treatment (179-CA), JKM179 HO endonuclease (179-HO), and JKM179 HO endonuclease -caffeine treatment (179-CA-HO-6). The 179 sample was not induced nor treated with any chemicals and, therefore, was regarded as control. The 179-CA sample was treated with caffeine which suppresses the repair mechanism. The 179-HO sample was induced to express HO endonuclease, resulting in the reduced number of breaks. 179-HO-CA-6 was 179-HO sample treated with caffeine two days later.

There were initially 34,348, 30,722, 48,182 and 29,061 reads in 179, 179-CA, 179-HO and 179-HO-CA-6, respectively. The number of reads, which contained head adaptor without the last ten bases missing and had sequence length greater than ten bases, was 4,101, 3,810, 4,611 and 1,519 in the same order as above. Finally, 478, 745, 424 and 400 reads were uniquely mapped to S288C genome or whole genome reads.

In another study, we observed a consensus sequence “CGT|X” with “|” denoted the breakpoints (Dr.Monnat, unpublished data). Breaks still occurred after “CGT” in all samples except the 179-HO-CA-6 (Figure 26). Therefore, we used Fisher’s exact test to determine if “CGT” or other 4-mer sequences had relationship with breaks in the 179 sample (Table 1 and Supplementary Table 1). “CGTG” and “CGGG” were the two significant patterns. We regarded these two patterns as retained breaks and the

rest of 4-mer sequences as fast repair breaks. The number of retained breaks in the 179 sample was higher than the 179-CA however, the 179-CA had greater number of fast repair breaks (Figure 27). We, thus, performed Fisher's exact test to examine the relationship between samples and types of break (i.e., retained vs fast repaired breaks). Blocking the repair mechanism by caffeine treatment (179-CA) showed an increase in the number of fast repair breaks. In 179-HO, the number of both types of breaks was reduced when compared with the 179. When we treated the 179-HO sample with caffeine (179-HO-CA-6), the number of fast repair breaks significant increased (Figure 27)

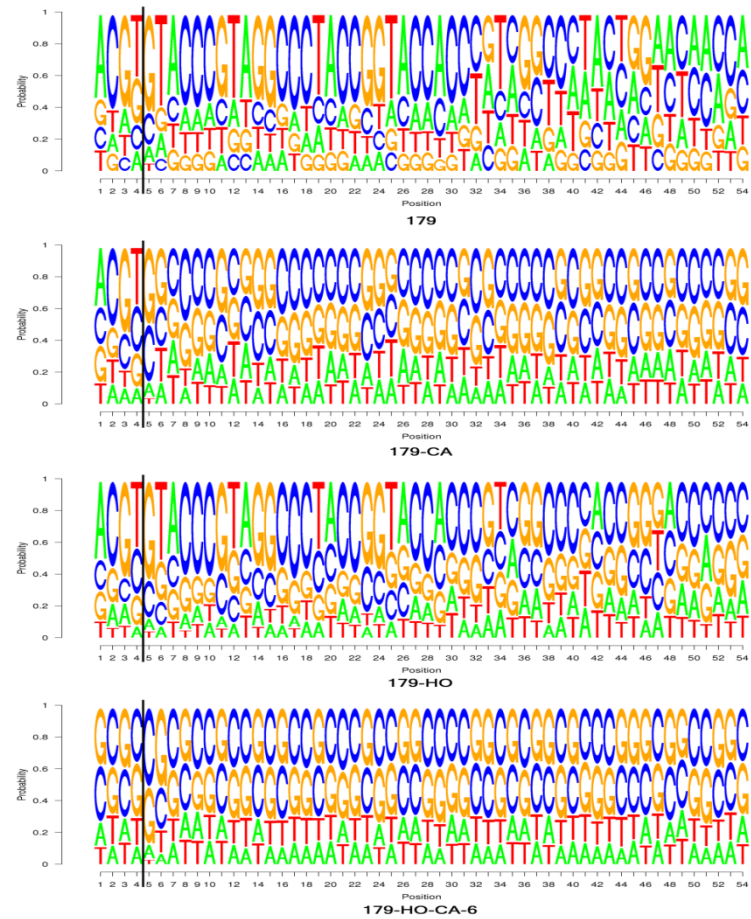


Figure 26. Sequence logo of DSBs where all columns have the same height. The break is between position 4 and 5 as marked by a vertical line.

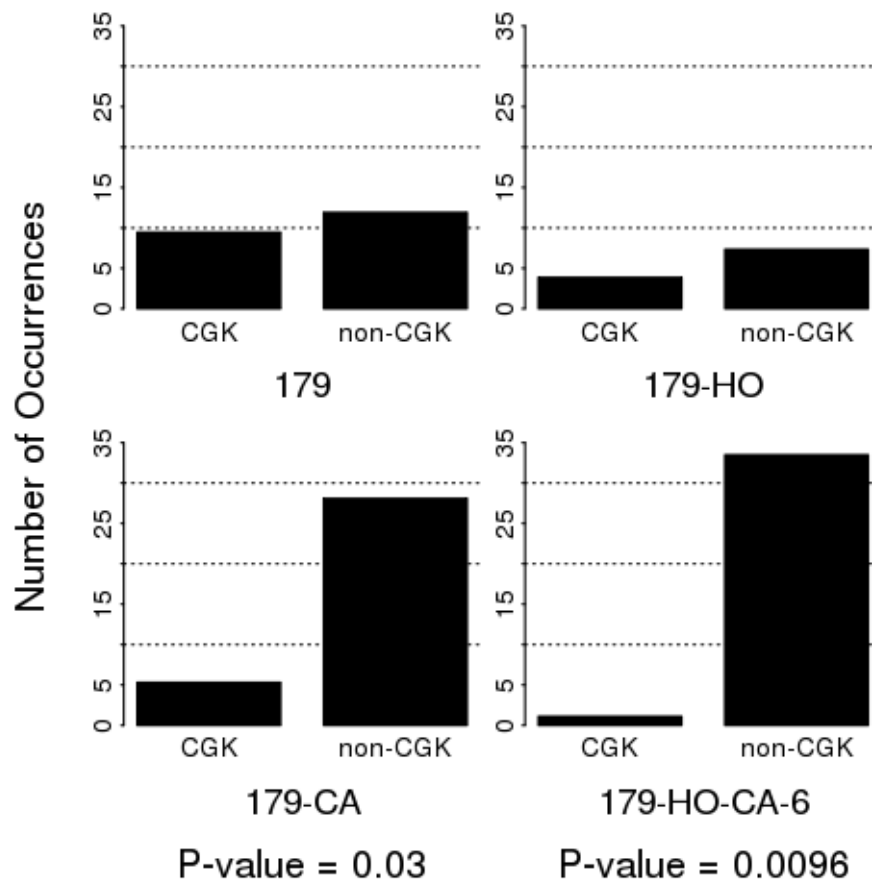


Figure 27. Number of retained break (CGK) vs number of fast repaired break (non-CGK) in the four samples

CHAPTER V

DISCUSSION

In this study, we established an assay for the measurement of RIND-EDSBs in budding yeast, and investigated the roles of various genes involved in the regulation of chromatin and DNA repair on the levels of RIND-EDSBs. From our results, we propose that RIND-EDSBs could be another epigenetic mark, because it satisfies two criteria of an epigenetic mark. First, there must be a non-random underlying mechanism. For example, DNA methylation is specific to the cytosine in CpG dinucleotides because of the DNA methyltransferase function [14, 15]. Second, epigenetic mark must have a specific function in controlling gene expression or affecting genomic integrity such as aging.

RIND-EDSBs are evolutionarily conserved

Similar to what previously observed in human cells [2], we found that EDSBs were present in the yeast genome during all phases of the cell cycle (Figure 14A). The detection of Ty1-EDSBs during the G₀ phase indicates that non-dividing yeast cells harbored RIND-EDSBs. Additionally, the detection of high levels of S-phase EDSBs was consistent with the hypothesis that DNA replication converts single-strand lesions to replication-dependent EDSBs [10]. The similarities between EDSB patterns observed in the mammalian and yeast genomes imply that the existence of EDSBs, and perhaps RIND-EDSBs, are evolutionarily conserved and may be required for cell homeostasis.

We have shown here and previously that, in both human and yeast cells, the levels of RIND-EDSB are actively regulated. Though evolutionarily distant, both species retain mechanisms, more specifically genes, which regulate the baseline

levels of the RIND-EDSBs. HMGB domain proteins are abundant proteins that bind DNA in a sequence-independent manner [16]. They are intricately involved in the regulation of chromatin structure and affect many DNA metabolic processes. Hmgb proteins bind to certain DNA lesions and either inhibit or facilitate their removal [11]. Interestingly, the levels of RIND-EDSBs were reduced in cells lacking Hmgb proteins (Figure 18 and 19). Therefore, Hmgb proteins may have a positive role in retaining the levels of RIND-EDSBs in the yeast genome. The *HMGB* genes are known to possess diverse functions in the maintenance of chromatin structures. The variation in the degrees of RIND-EDSB reduction in different deletion strains may be a result of functional redundancies within the HMGB family. Nhp6A and Nhp6B are homologous proteins that are functionally redundant. In budding yeast, *NHP6A* is expressed much more robustly than *NHP6B*. Therefore, the *nhp6a* deletion resulted in a more pronounced phenotype [11].

We observed here that yeast strains lacking specific genes involving in the generation of DNA breaks, topoisomerases and endonucleases, increased RIND-EDSB levels (Figure 20). These results appear contradictory to our initial hypothesis and suggest that these genes are not involved in the production or retention of RIND-EDSBs. The increase in RIND-EDSB levels when the genes encoding certain topoisomerases or endonucleases were deleted may provide a hint for a potentially essential role of RIND-EDSBs. Both topoisomerases and endonucleases generate DNA breaks to mediate their biological roles, which is to release genome's physical stress. Thus, an increase of RIND-EDSBs upon the deletions of these genes may represent a crucial compensatory mechanism for the loss of function of certain topoisomerases or endonucleases. The conserved existence of RIND-EDSBs over long evolutionary

time implies that they may provide an advantage for the organisms such that this feature survives through natural selection.

Levels of RIND-EDSB and repair of the breaks are regulated

In human cells, compact heterochromatin-associated RIND-EDSBs are repaired by an ATM-dependent pathway. However, Ku-mediated NHEJ can repair euchromatin-associated EDSBs [9]. Here we found that the levels of RIND-EDSBs were significantly increased in yeast strains with deletions of genes encoding components of the DNA damage response (Figure 17). This result suggested that the levels of RIND-EDSBs are constantly monitored and controlled by these DNA damage sensors. Thus, deletion of the DNA damage sensors abolished that control mechanism.

Two major pathways that repair double stranded breaks are non-homologous recombination (NHEJ) and homologous-mediated recombination (HR) [17]. In this study, we focused on a set of genes that operate in the NHEJ pathway. During the non-replicating phase, the haploid budding yeasts contain only one copy of the genome. Thus, it is generally presumed that the conventional HR-mediated DNA repair, which requires another copy of the genome as a template for repair, does not operate during this stage. The levels of RIND-EDSBs were significantly increased in *yku70Δ* and *yku80Δ* strains. Therefore, the NHEJ pathway regulated, at least partly, the levels of RIND-EDSBs in non-replicating yeast. These data are in accordance with our previous finding that RIND-EDSBs could be repaired by the NHEJ pathways in human cell [9]. Interestingly, *nej1* deletion did not change the RIND-EDSB levels. There may be other factors that compensate for the loss of *NEJ1* in yeast. Although it is generally believed that in the non-replicative stage, budding yeasts are not able to repair DNA breaks by the HR-mediated pathway, we observed a significant increase

in RIND-EDSBs in a yeast strain lacking Rad51, a key protein in the HR pathway. This result suggests that there might be an alternative Rad51-mediated pathway to repair RIND-EDSBs in non-replicative yeast.

RIND-EDSBs and heterochromatin

We previously demonstrated, in human cells, that areas containing RIND-EDSBs are hypermethylated and are within the facultative heterochromatin [9]. We also found that the RIND-EDSBs were devoid of γ H2AX, and that trichostatin A (TSA) treatment increased histone acetylation, produced spontaneous DNA breakages, triggered H2AX phosphorylation and allowed RIND-EDSB repair [9]. Therefore, we hypothesized that RIND-EDSB levels and break repair are at least partly regulated by specific pathways and are influenced by the genome topology and chromatin structures.

Studies in yeast indicated that the level of RIND-EDSBs is connected to the level of heterochromatin and may be controlled indirectly by the proteins that regulate the spreading of heterochromatin (Figure 21). Our findings were also consistent with our prior report, which suggested that RIND-EDSBs are likely retained in heterochromatin. Low levels of RIND-EDSBs were found in yeast strain lacking the histone deacetylase Sir2. Moreover, *htz1 Δ* and *swr1 Δ* strains, lacking genes that suppress heterochromatin spreading [13], possessed high levels of RIND-EDSBs. Nevertheless, a significant increase in RIND-EDSB levels in the *rpd3 Δ* strain was observed. *Rpd3* is a histone deacetylase that has a controversial role in heterochromatin formation. Traditionally, *Rpd3* has been associated with telomere stability [18]. However, recent evidence suggests that it may antagonize Sir2-

dependent heterochromatin spreading [19]. This result supports the anti-Sir2 role of Rpd3.

RIND-EDSBs associated with chronological aging

The main objective of our project is to investigate replication independent endogenous DNA double-strand breaks (RIND-EDSBs) in preventing genome breakage and cellular aging. We hypothesize that maintained RIND-EDSBs have important role to protect genome integrity in cells. Cells may maintain a certain level of RIND-EDSBs, a reduction of which may have adverse effects on cell survival and lifespan and could contribute to genomic instability during cellular aging.

There are two main models of aging in yeast: replicative aging which measures the number of times a mother-cell can divide before senescence, and chronological aging which reflects the time non-dividing cells can live in a quiescent state [20]. Since chronological aging model represents the situation often found in post-mitotic differentiated cells in higher organisms, we chose this model for our study.

In this study, the result showed that we found a reduction of RIND-EDSBs during aging in yeast (Figure 22). In addition, RIND-EDSB function possibly associates with genomic instability and aging. As the retention of RIND-EDSBs may prevent error-prone repair and genomic instability, lower level of RIND-EDSBs could provide to the aging process. Thus, we propose the role of RIND-EDSBs in maintaining the integrity of the genome and prevent aging. To test this hypothesis, we set up the experiments in the JKM179 yeast strain which cause a galactose induction HO endonuclease gene. HO endonuclease generates a DSB during mating type switching. This DSB is repaired by the process of homologous recombination, using the silenced copy at HML α or

HMRa on chromosome III as template to repair. In JKM179 strain was deleted the homologous sequences HML α and HMRa were detected, therefore a galactose-induced DSB at MAT could not be repaired by homologous recombination [21].

We found that the expression of HO in JKM179 strain led to reduced RIND-EDSBs similar to what we observed during aging (Figure 23). Hence, we chose to use HO induction as a means to investigate the effect of reduced RIND-EDSBs instead of aging. Furthermore, we discovered the association between RIND-EDSB levels and cell viability. Our result also showed that HO induction led to a lower level of RIND-EDSBs and lower cell viability. Thus, the retention of RIND-EDSBs might arrest error-prone repair and prevent genomic instability. Moreover, caffeine treatment in aging and HO induction experiments suggest that a lower level of RIND-EDSBs could lead to an increase in the fast repair of EDSBs, which could be error-prone and subsequently lead to an increased mutations and reduced viability of cells. We hypothesized that RIND-EDSBs are not pathologic DNA lesions and may have important biological function. In addition, we studied characteristics of yeast RIND-EDSBs using Ion torrent DNA sequencing. We conducted our experiment on four samples (179, 179-CA, 179-HO and 179-CA-HO-6). The locations and sequences around RIND-EDSBs, we found a consensus sequence "CGT" in all samples except the 179-HO-CA-6 (Figure 27). However, the proportions of CGT sequence copy were different in each sample. For instance, the number of retained breaks (CGT sequence) in the 179 was higher than the 179-CA because the 179-CA had greater number of fast repair breaks (Figure 7). Blocking the repair mechanism by caffeine treatment (179-CA) showed an increased number of fast repair breaks. In 179-HO, the number of both types of breaks reduced when compared with the 179. When we treated the 179-HO sample with caffeine (179-HO-CA-6), the number of fast repaired breaks was significantly increase (Figure 27). These results show that RIND-EDSBs occurred non-randomly distributed in genome. Interestingly, retained breaks occur with specificity in terms of sequence pattern. So, RIND-EDSBs are produced and retained by specific mechanisms. Although, the mechanism is not clear, our results showed that RIND-EDSBs have a non-random underlying mechanism. Moreover, RIND-

EDSBs have a specific function or affecting genomic integrity in aging. From our results, we propose that RIND-EDSBs may be epigenetic mark which was specific mechanism to preserve genomic integrity and play a role in cellular aging.



REFERENCES

1. Pornthanakasem, W., et al., LINE-1 methylation status of endogenous DNA double-strand breaks. *Nucleic Acids Res*, 2008. 36(11): p. 3667-75.
2. Kongruttanachok, N., et al., Replication independent DNA double-strand break retention may prevent genomic instability. *Mol Cancer*, 2010. 9: p. 70.
3. Hoeijmakers, J.H., Genome maintenance mechanisms for preventing cancer. *Nature*, 2001. 411(6835): p. 366-74.
4. Blagosklonny, M.V., et al., Impact papers on aging in 2009. *Aging (Albany NY)*, 2010. 2(3): p. 111-21.
5. Fraga, M.F., R. Agrelo, and M. Esteller, Cross-talk between aging and cancer: the epigenetic language. *Ann N Y Acad Sci*, 2007. 1100: p. 60-74.
6. Kaeberlein, M., Lessons on longevity from budding yeast. *Nature*, 2010. 464(7288): p. 513-9.
7. Revet, I., et al., Functional relevance of the histone gammaH2Ax in the response to DNA damaging agents. *Proc Natl Acad Sci U S A*, 2011. 108(21): p. 8663-7.
8. Lobrich, M., et al., gammaH2AX foci analysis for monitoring DNA double-strand break repair: strengths, limitations and optimization. *Cell Cycle*, 2010. 9(4): p. 662-9.

9. Vilenchik, M.M. and A.G. Knudson, Endogenous DNA double-strand breaks: production, fidelity of repair, and induction of cancer. *Proc Natl Acad Sci U S A*, 2003. 100(22): p. 12871-6.
10. Aylon, Y. and M. Kupiec, DSB repair: the yeast paradigm. *DNA Repair (Amst)*, 2004. 3(8-9): p. 797-815.
11. Kegel, A., J.O. Sjostrand, and S.U. Astrom, Nej1p, a cell type-specific regulator of nonhomologous end joining in yeast. *Curr Biol*, 2001. 11(20): p. 1611-7.
12. Milne, G.T., et al., Mutations in two Ku homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 1996. 16(8): p. 4189-98.
13. Pang, D., et al., Ku proteins join DNA fragments as shown by atomic force microscopy. *Cancer Res*, 1997. 57(8): p. 1412-5.
14. Boulton, S.J. and S.P. Jackson, Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing. *EMBO J*, 1998. 17(6): p. 1819-28.
15. Chen, L., et al., Promotion of Dnl4-catalyzed DNA end-joining by the Rad50/Mre11/Xrs2 and Hdf1/Hdf2 complexes. *Mol Cell*, 2001. 8(5): p. 1105-15.
16. Anderson, D.E., et al., Structure of the Rad50 x Mre11 DNA repair complex from *Saccharomyces cerevisiae* by electron microscopy. *J Biol Chem*, 2001. 276(40): p. 37027-33.
17. Moore, J.K. and J.E. Haber, Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol*, 1996. 16(5): p. 2164-73.

18. Herrmann, G., T. Lindahl, and P. Schar, *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *EMBO J*, 1998. 17(14): p. 4188-98.
19. Rogina, B. and S.L. Helfand, Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proc Natl Acad Sci U S A*, 2004. 101(45): p. 15998-6003.
20. Lopez-Otin, C., et al., The hallmarks of aging. *Cell*, 2013. 153(6): p. 1194-217.
21. Longo, V.D., et al., Replicative and chronological aging in *Saccharomyces cerevisiae*. *Cell Metab*, 2012. 16(1): p. 18-31.
22. Kaeberlein, M., C.R. Burtner, and B.K. Kennedy, Recent developments in yeast aging. *PLoS Genet*, 2007. 3(5): p. e84.
23. Fabrizio, P. and V.D. Longo, The chronological life span of *Saccharomyces cerevisiae*. *Aging Cell*, 2003. 2(2): p. 73-81.
24. Fabrizio, P. and V.D. Longo, The chronological life span of *Saccharomyces cerevisiae*. *Methods Mol Biol*, 2007. 371: p. 89-95.
25. Fabrizio, P., et al., Regulation of longevity and stress resistance by Sch9 in yeast. *Science*, 2001. 292(5515): p. 288-90.
26. Iqbal, K., et al., Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A*, 2011. 108(9): p. 3642-7.
27. Tucker, K.L., Methylated cytosine and the brain: a new base for neuroscience. *Neuron*, 2001. 30(3): p. 649-52.

28. Wilson, A.S., B.E. Power, and P.L. Molloy, DNA hypomethylation and human diseases. *Biochim Biophys Acta*, 2007. 1775(1): p. 138-62.
29. Raabe, F.J. and D. Spengler, Epigenetic Risk Factors in PTSD and Depression. *Front Psychiatry*, 2013. 4: p. 80.
30. Pogribny, I.P. and B.F. Vanyushin, Age-related genomic hypomethylation, in *Epigenetics of Aging*. 2010, Springer. p. 11-27.
31. Richardson, B., Impact of aging on DNA methylation. *Ageing Res Rev*, 2003. 2(3): p. 245-61.
32. Berdishev, G.D., et al., Nucleotide composition of DNA and RNA from somatic tissues of humpback salmon and its changes during spawning *Biochemistry (Mosc.)*, 1967. 32: p. 988-993.
33. Vanyushin, B.F., et al., Investigation of some characteristics of the primary and secondary structure of DNA from the liver of spawning humpback salmon. *Biochemistry (Mosc.)* 1969. 34: p. 191-198.
34. Vanyushin, B.F., et al., The 5-methylcytosine in DNA of rats. Tissue and age specificity and the changes induced by hydrocortisone and other agents. *Gerontologia*, 1973. 19(3): p. 138-52.
35. Romanov, G.A. and B.F. Vanyushin, Methylation of reiterated sequences in mammalian DNAs. Effects of the tissue type, age, malignancy and hormonal induction. *Biochim Biophys Acta*, 1981. 653(2): p. 204-18.
36. Holliday, R., Strong effects of 5-azacytidine on the in vitro lifespan of human diploid fibroblasts. *Exp Cell Res*, 1986. 166(2): p. 543-52.

37. Fairweather, D.S., M. Fox, and G.P. Margison, The in vitro lifespan of MRC-5 cells is shortened by 5-azacytidine-induced demethylation. *Exp Cell Res*, 1987. 168(1): p. 153-9.
38. Jintaridth, P. and A. Mutirangura, Distinctive patterns of age-dependent hypomethylation in interspersed repetitive sequences. *Physiol Genomics*, 2010. 41(2): p. 194-200.
39. Murgatroyd, C., et al., The Janus face of DNA methylation in aging. *Aging (Albany NY)*, 2010. 2(2): p. 107-10.
40. Burke, D., Dawson, D. Stearns, T, Cold Spring Harbor, in *Methods in yeast genetics*. 2000, Cold Spring Harbor Laboratory Press: NY. p. 145-147.
41. Ribeiro, G.F., M. Corte-Real, and B. Johansson, Characterization of DNA damage in yeast apoptosis induced by hydrogen peroxide, acetic acid, and hyperosmotic shock. *Mol Biol Cell*, 2006. 17(10): p. 4584-91.
42. Bernstein, B.E., J.K. Tong, and S.L. Schreiber, Genomewide studies of histone deacetylase function in yeast. *Proc Natl Acad Sci U S A*, 2000. 97(25): p. 13708-13.
43. Moser, B.A., et al., Mechanism of caffeine-induced checkpoint override in fission yeast. *Mol Cell Biol*, 2000. 20(12): p. 4288-94.
44. Kim, J.A., et al., Heterochromatin is refractory to gamma-H2AX modification in yeast and mammals. *J Cell Biol*, 2007. 178(2): p. 209-18.
45. Bandyopadhyay, S., et al., Rewiring of genetic networks in response to DNA damage. *Science*, 2010. 330(6009): p. 1385-9.

46. Nelson, R.G. and W.L. Fangman, Nucleosome organization of the yeast 2-micrometer DNA plasmid: a eukaryotic minichromosome. *Proc Natl Acad Sci U S A*, 1979. 76(12): p. 6515-9.
47. Yao, X., et al., Overexpression of high-mobility group box 1 correlates with tumor progression and poor prognosis in human colorectal carcinoma. *J Cancer Res Clin Oncol*, 2010. 136(5): p. 677-84.
48. Chalitchagorn, K., et al., Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene*, 2004. 23(54): p. 8841-6.
49. Kitkumthorn, N., et al., LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients. *Clin Chim Acta*, 2012. 413(9-10): p. 869-74.
50. Pages, H., et al., Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.26.3.
51. Altschul, S., et al., Basic local alignment search tool. *J Mol Biol*, 1990. 215(3): p. 403-410.
52. Stillman, D.J., Nhp6: a small but powerful effector of chromatin structure in *Saccharomyces cerevisiae*. *Biochim Biophys Acta*. 1799(1-2): p. 175-80.
53. Gasser, S.M. and M.M. Cockell, The molecular biology of the SIR proteins. *Gene*, 2001. 279(1): p. 1-16.
54. Meneghini, M.D., M. Wu, and H.D. Madhani, Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell*, 2003. 112(5): p. 725-36.

55. Turek-Plewa, J. and P. Jagodzinski, The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cellular and Molecular Biology Letters* 2005. 10(4): p. 631.
56. Sharif, J., et al., The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, 2007. 7171: p. 909-912.
57. Reeves, R., Nuclear functions of the HMG proteins. *Biochim Biophys Acta*, 2010. 1799(1-2): p. 3-14.
58. Sun, Z. and M. Hampsey, A general requirement for the Sin3-Rpd3 histone deacetylase complex in regulating silencing in *Saccharomyces cerevisiae*. 1999. 152: p. 921-932.
59. Zhou, J., et al., Histone deacetylase Rpd3 antagonizes Sir2-dependent silent chromatin propagation. *Nucleic Acids Res*, 2009. 37(11): p. 3699-713



APPENDIX

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

SUBMITTED ARTICLE

Characteristics of Replication-Independent endogenous double-strand breaks in
Saccharomyces cerevisiae

Monnat Pongpanich^{1,4}, Maturada Patchsung², Jirapan Thongsroy², Apiwat
Mutirangura^{3,4§}

¹Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

²Inter-Department Program of BioMedical Sciences, Faculty of Graduate School, Chulalongkorn University, Bangkok, Thailand

³Department of Anatomy, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand

⁴Center for Excellence in Molecular Genetics of Cancer and Human Diseases, Chulalongkorn University, Bangkok, Thailand

[§]Corresponding author

Email addresses:

MPO: monnat.p@chula.ac.th MPA: phukansin@hotmail.com

JT: ju_jirapan@hotmail.com AM: mapiwat@chula.ac.th

Keywords: Replication-Independent Endogenous DNA Double-Strand Breaks, RIND-EDSBs, *Saccharomyces cerevisiae*, next-generation sequencing

Abstract

Background

Replication-INDependent endogenous double-strand breaks (RIND-EDSBs) occur in non-dividing cells, both human and yeast, without exposure to any inductive agents. In human cells, they are hypermethylated, preferentially retained in heterochromatin and unbound by γ -H2AX. In single gene deletion strain yeasts, the RIND-EDSB levels are altered. The number of RIND-EDSBs are increased in strains with deletions of histone deacetylase, endonucleases, topoisomerase, or DNA repair regulators but decreased in strains with deletions of the high-mobility group box proteins or Sir2. In summary, RIND-EDSBs are different from pathologic double-strand breaks in terms of causes and consequences. Here, we determined the nucleotide sequences surrounding RIND-EDSBs and investigated the features of the sequences and break locations.

Results

Using next-generation sequencing of cells in a resting state, we were able to find break positions along *Saccharomyces cerevisiae* chromosomes. We found that break locations did not cluster at specific regions such as the centromere or telomere but instead are scattered along chromosomes. The number of breaks correlated with the size of the chromosome. Most importantly, break occurrences have specificity in terms of sequence pattern. Specifically, the majority of breaks occurred immediately after the sequence “CGT”. The specificity of the “CGT” sequence cannot be attributed to chance because in the yeast genome, the “CGT” sequence occurs non-primarily.

Conclusions

RIND-EDSBs occurred non-randomly; that is, they are produced and retained by specific mechanisms. With particular mechanisms regulating their generation and potentially possessing specific functions, RIND-EDSBs could be another epigenetic mark.

Background

In our previous work, we detected endogenous DNA double strand breaks (EDSBs) that occurred in non-replicating cells without using any inductive agents and named them replication-independent EDSBs (RIND-EDSBs) [1-3]. The causes and consequences of RIND-EDSBs are different from replication-induced EDSBs [4] and irradiation-induced DSBs [5]. Whereas irradiation-induced DSBs halt the cell cycle [6], induce cell death [7] and serve as mediators of mutation [8], RIND-EDSBs do not. RIND-EDSBs are detectable in all human cell types [1] and yeast [3]. Down-regulation or deletions of high mobility group genes in both human and yeast lower RIND-EDSB levels [3]. High mobility group genes are multifunctional genes that regulate multiple DNA-dependent processes such as transcription, replication, recombination, and DNA repair [9, 10]. In conclusion, not only are RIND-EDSBs ubiquitously expressed in eukaryotic cells but also some genes serve to maintain optimal levels of RIND-EDSBs. Therefore, RIND-EDSBs are evolutionally conserved in eukaryotic cells and may possess some essential functions [1-3].

We reported several pieces of evidence that indicate that RIND-EDSBs may help maintain genomic integrity. First, RIND-EDSBs are linked to methylated CpG nucleotides [1]. Genomic hypomethylation is associated with genomic instability [11-13]. Consequently, fragile genomes contain not only low levels of DNA methylation

but also low numbers of RIND-EDSBs. Second, RIND-EDSBs are preferentially retained in heterochromatin and unbound by γ -H2AX [2]. Finally, when RIND-EDSBs were reduced by trichostatin A, the number of γ -H2AX foci increased [2]. γ -H2AX foci are signals of pathologic DSBs [14, 15]. This observation suggests that RIND-EDSBs may prevent unwanted DNA breakage. Interestingly, generation of RIND-EDSBs can be discovered in cells lacking topoisomerase [3]. The function of topoisomerase is to control the topological structure of DNA [16, 17]. Therefore, it is possible that RIND-EDSBs and topoisomerase may have redundant roles.

RIND-EDSBs do not lead to mutation. We demonstrated that methylated RIND-EDSBs in human cells were repaired by the more precise, ATM-dependent type of non-homologous end joining (NHEJ) [2], whereas unwanted DSBs were generally repaired by a faster, more-error prone, Ku-mediated type of NHEJ. The compact structure prevented H2AX phosphorylation, a conventional DSB repair response, which resulted in methylated RIND-EDSBs escaping error-prone NHEJ repair [2].

We hypothesize here that RIND-EDSBs were not pathologic DNA lesions and may have essential biological function. Consequently, RIND-EDSBs should not occur randomly. In this work, we studied characteristics of yeast RIND-EDSBs using high throughput sequencing, which allowed us to analyze the locations and sequences around RIND-EDSBs.

Results

Sequencing produced 105,775 reads. We trimmed adaptors by aligning adaptor sequences against reads. Alignment results varied as reported in Additional file 1: Figure S3. Therefore, we categorized alignments into three groups as follows: (1) reads without a head adaptor sequence, (2) reads with a head adaptor sequence but

missing ten or more of the last bases and (3) reads with a head adaptor sequence and not missing the last ten bases. The number of reads in each category was 28,905, 40,077, and 36,793, respectively. We retained only reads in the third category in order to obtain the most accurate break positions possible. We next discarded 26,824 reads because their length was less than 10 bases, leaving 9,969 reads. Of these, 1,893 reads mapped to the genome, and 1,494 reads aligned uniquely. We excluded uniquely mapped reads that were not aligned from the first base of the read. Our analysis was based on the remaining 1,257 reads.

Distribution and number of RIND-EDSBs

We investigated whether break positions had any relationship with position on the chromosome e.g., centromeric or telomeric. The alignment positions from BLAST corresponded to break positions; thus, we could observe EDSB distribution across all 16 chromosomes. Figure 1 shows 721 EDSB positions determined from the 1,257 filtered reads. The reads originated from a population of cells rather than a single cell. The height of each vertical bar corresponds to the number of reads aligned to that position. As shown in the figure, the break positions did not clump near the centromere or telomere but are scattered along the chromosomes. However, they are distributed unevenly along chromosomes, and breakpoints mostly occurred in clusters. The approach used to identify clusters is explained in the Methods section.

Some chromosomes had significant clusters (orange or red clusters in Figure 1); thus, we simulated the data (See Methods for further details) to verify the significance of the clusters. Clusters also occurred in simulated data (see Additional file 1 for p-values of the simulated clusters), and for chromosomes 1-16, totals of

260, 300, 264, 349, 281, 260, 344, 293, 256, 295, 311, 336, 332, 282, 338 and 286, respectively, out of 1,000 simulated data had significant clusters.

To see if the number of breaks correlated with chromosome size, we plotted the number of breaks against the length of the chromosomes (Figure 2). The number of breaks increased with chromosome size. The Pearson correlation coefficient was 0.93, and the p-value was 2.16E-07.

Sequence patterns neighboring break points

To explore whether specific sequence patterns existed around the breaks, we plotted sequence logos (Figure 3 and Additional file 2: Figure S1) of those 1,257 uniquely mapped reads (See Methods for further details). In Figure 3, we observed a consensus sequence “CGT|C” that spanned over breakpoints, with “|” denoting the break. Base frequency pattern was strikingly different at “CGT|C” columns. In columns other than “CGT|C”, the frequency of C is close to that of G, and that of A is close to that of T; the frequencies of C and G are mostly lower than the frequencies of A and T. This is consistent with the proportions of A, C, G and T in yeast genome, which are 30.9%, 19.2%, 19.1% and 30.8%, respectively. Additional file 2: Figure S1 shows sequence logos where the column height is proportional to information content. Although the information content in each column is low, position 50, which is right before breakpoint, has the highest information content, indicating a lower tolerance for substitutions compared to other positions.

We next investigated the following two questions: (1) Did the sequence “CGT|C” occur at break positions randomly? (2) Did C, G, T and C occur consecutively in the same location? We performed Fisher’s exact test (see Methods) to answer these questions. Forty-two of 1,257 reads had “CGT|C” at the breakpoint and “CGT|C”

occurred 23,057 times whereas other 4-mer sequences occurred 11,777,830 times in the genome. The p-value was $8.89E-37$, which was highly significant. This suggested that “CGT|C” occurred non-randomly and in the same location.

We further validated the non-random occurrence of “CGT|C” sequence using simulations as described in Methods. Unlike the observed data, all 100 sequence logos from simulated data (Additional file 3) showed no specific pattern in any column. In every column of all logos, C and G, whose frequencies are approximately the same, have lower frequency than A and T, which have similar frequencies. This frequency pattern is the same as all of the columns in the observed data except “CGT|C” columns. This further supported our conclusion that the association of “CGT|C” with breaks was nonrandom.

To examine whether breaks were associated with other 4-mer sequences, we calculated Fisher’s exact test and odds ratio for the other possible 4-mer sequences (Additional file 4: Table S1). The significant 4-mer sequences are shown in Table 1. In addition to “CGT|C”, the p-values of “CGT|A”, “CGT|G” and “CGT|T” were extremely significant. In addition, the sequences “AGT|A”, “AGT|C”, “AGT|G”, “CAT|C”, “CTC|C”, “GGT|A”, “TAT|C”, and “TGT|A” were significant with non-zero odds ratios. “AAA|T” and “TGA|A” were also significant, but those sequences did not occur in any reads. We plotted the number of stitched reads with each possible 4-mer sequence at breaks (Figure 4A) and the number of each 4-mer occurrences in the yeast genome (Figure 4B). Comparing the two figures clearly illustrated that the significant 4-mer sequences, which were common in reads, were not predominant in the genome. These results led us to conclude that RIND-EDSBs preferentially occurred at specific sequence patterns.

In conclusion, the most frequent sequence before breaks was “CGT”. The 4-mers starting with “CGT” had exceptionally high odds ratios (red triangles in Figure 5). Moreover, there were other 4-mer sequences that occurred frequently with odds from 3-6 (red stars in Figure 5). This group of 4-mers mostly contained GT before breaks. Conversely, some 4-mer sequences had zero odds (pink dots in Figure 5). These 4-mer sequences were never observed in reads. Two of them, “AAA|T” and TGA|A”, had significant p-values. When these 4-mers were considered together, the p-value was extremely significant ($< 2.2E-16$; last row in Table 1). This finding demonstrated sequence specificity in break occurrences, i.e., the specificity of having no breaks at these 4-mer sequences. The number of 4-mer sequences with no breaks was not trivial (Figure 6). None of these 4-mer sequences had T in the third position.

Discussion

We postulated that RIND-EDSBs have an important function and therefore must occur non-randomly. In this work, we examined characteristics of RIND-EDSBs to determine if RIND-EDSBs systematically occur. We found that breaks were scattered along chromosomes and that the number of breaks positively correlated with chromosome size. Most importantly, breaks occurred most frequently at the sequence “CGT”.

RIND-EDSBs were evolutionarily conserved, as they were present in both human and yeast genomes. In addition, the frequency of RIND-EDSBs decreased in yeast cells depleted of *Hmo1*. The same results occurred in human cells lacking *HMGB1*, which is the human homolog of *Hmo1* [3].

The frequent occurrence of RIND-EDSBs after “CGT” drew some attention to the “CGT” pattern itself. The identification of this pattern did not conflict with our previous findings that, in human cells, RIND-EDSBs were hypermethylated, and methylation pre-existed at break sites [1]. In addition, RIND-EDSBs were more likely to be retained in heterochromatin [2]. Heterochromatin was hypermethylated and DNA methylation typically occurred at CpG sites [18, 19], which could be related to having a “CGT” pattern at breaks. The specificity of “CGT” was not a coincidence either, as adenine and thymine are more abundant than cytosine and guanine in the yeast genome. This suggested that RIND-EDSBs are generated by a specific mechanism. The mechanism could resemble that used by restriction enzymes to recognize and cut a specific nucleotide sequence, or it might depend on a physical property of dinucleotide sequences. It was unlikely that the mechanism was dependent on chromosome location because we observed break locations scattered along chromosomes, suggesting distinct RIND-EDSBs locations in each cell. Although the mechanism is still unknown, we speculate that there is a specific mechanism for RIND-EDSB occurrences. Because RIND-EDSBs were conserved from yeast to human, the mechanism by which RIND-EDSBs are produced should be conserved as well.

RIND-EDSBs could be another epigenetic mark, which would require two properties. First, there must be a non-random underlying mechanism. For example, DNA methylation is specific to the cytosine in CpG dinucleotides because of the DNA methyltransferase function [20, 21]. Second, epigenetic mark must have a specific function in controlling gene expression or affecting genomic integrity. For example, methylation in promoters suppressed gene expression [22, 23], intragenic LINE-1 hypomethylation reduced gene expression [24], and hypomethylation can lead to

genomic instability [11]. From the conservation of RIND-EDSBs in every cell from yeast to human, we deduced that a specific mechanism must exist to produce them. Moreover, in human cells, γ -H2AX, which was a marker of abnormal DNA breakage, increased when RIND-EDSBs were reduced [2]. When HMGB1 was depleted, the level of RIND-EDSBs decreased, suggesting a role for HMGB1 in retaining RIND-EDSBs [3]. A yeast strain lacking genes involved in generating DSBs, e.g., topoisomerases, had an increased level of RIND-EDSBs [3]. This fact implied a compensatory mechanism for the loss of function of certain topoisomerases and indicated a potential role for RIND-EDSBs. Therefore, we inferred that RIND-EDSBs were an epigenetic mark.

Conclusions

In summary, we found that RIND-EDSBs were location independent but preferentially occurred immediately after the sequence “CGT”. This finding suggested that RIND-EDSB production was regulated by a non-random mechanism. The specificity of the “CGT” sequence was consistent with the methylation of CpG sites observed in human cells. Having a specific mechanism and being retained for a specific function signified that RIND-EDSBs are epigenetic marks.

Methods

Yeast strains, media and growth conditions

The wild type yeast *Saccharomyces cerevisiae*, BY4741 (*MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0*), was used in this study. Yeast was grown at 30°C in 10 ml of liquid YPD media (Sigma, USA) for 1 day to stationary phase ($\sim 8 \times 10^6$ cells/mL) and diluted to 1×10^4 cell/mL in YP medium containing 2% raffinose (Sigma, USA) for 48 hours (for

synchronization). Cell cycle phases were confirmed by phase-contrast microscopy. Whole cells were pelleted by centrifugation at 5,000 rpm for 5 minutes.

High-Molecular weight (HMW) DNA preparation

Yeast cell pellets were treated with 1 mg/ml lyticase (70 U/mg) (Sigma, USA) for 2 hours and embedded in 1% low melting point agarose (MO BIO, USA) at a concentration of 2×10^8 cells per plug mold. Plugs with cells were digested in 400 μ l of digestion buffer (1 mg/ml proteinase K, 50 mM Tris, pH 8.0, 20 mM EDTA, 1% sodium lauryl sarcosine) at 37°C overnight. The plugs were washed 6 times in TE buffer for 40 minutes, and the cohesive end EDSBs were polished with T4 DNA polymerase (New England Biolabs, USA). The enzyme was inactivated by incubation in 20 mM EDTA for 5 minutes, which was followed by 6 washes in TE buffer for 40 minutes. The first linkers were prepared from the two oligonucleotides 5'-AGGTAACGAGTCAGACCACCGATCGCTCGGAAGC TTACCTCGT GGACGT-3' and 5'-ACGTCCACGAG-3' (Sigma, Singapore) [1] for ligation to HMW DNA using T4 DNA ligase (New England Biolabs, USA) at 25°C for 2 nights. DNA was extracted from the agarose plugs using a QIAquick gel extraction kit (Qiagen, Switzerland) [1]. DNA was digested in a 100 μ l reaction volume with 2 U of RsaI in 1X NEB buffer (New England Biolabs, USA) 4 at 37°C overnight. Digested DNA was purified by using phenol:chloroform and ethanol precipitation and ligated again with the second linkers (50 μ M), which were prepared from the two oligonucleotides 5'ATGGTACCACCCGTAGGCCCTAC CGGT ACC-3' and 5'-GGTACCGGTAGGGCCTACGGGTGGTACCAT -3' (Sigma, Singapore). Ligated DNA was purified with QIAquick PCR Purification Kit (Qiagen, Switzerland).

We have shown previously that the detected RIND-EDSBs were not due to our DNA preparation protocol [3].

Library preparation and sequencing

After ligation with two linkers, HMW DNA was subjected to 60 cycles of PCR with two primers, First-L-F (5'-AGGTAACGAGTCA GACCACCGA-3') and Second-L-R (5'-GGTACCGGTAG GGCCTACGGGT-3' (Sigma, Singapore) with an annealing temperature of 50°C. PCR product was purified with QIAquick PCR Purification Kit (Qiagen, Switzerland) and sequenced by Ion Torrent sequencer (Ion Torrent™ Personal Genome Machine® (PGM), Life Technologies, USA)

Data processing and mapping

Each sequenced read contained different adaptors at both ends of the read; therefore, we had to trim adaptors from both ends of the reads. We refer to adaptors that are ligated to the beginning and end of reads as head adaptors and tail adaptors, respectively. We trimmed adaptors by aligning head and tail adaptor sequences against read sequences using the “pairwiseAlignment” function (rewarded 1 for a match and penalized 2, 5 and 2 for a mismatch, a gap opening and a gap extension, respectively) from the Biostrings package [25] in R. For each read, we did not know if it came from the plus or minus strand; therefore, we aligned the head and tail adaptor to reads from the fasta file and the reverse complement sequences. If a read was from the plus strand, the alignment against the original sequence would be more successful, while minus strand reads would align more successfully to the reverse complement. The alignment was considered successful if it met the following criteria: (1) the aligned length was greater than 10 bases but less than adaptor length + 5. If the aligned length was much longer than adaptor length, it implied the presence of excessive gaps, and thus, those alignments were not trustworthy. We added 5 to the adaptor length to allow for some gaps in the

alignment, and 5 was arbitrarily chosen. (2) The last aligned base position of the head adaptor was within head adaptor length + 5. Similar logic was applied to the tail adaptor. Adaptors should be aligned at the ends of the reads; therefore, if the aligned position was too far from the end, we did not trust the alignment. (3) The last ten bases of the head adaptor sequence must be matched. As break positions were between the last base of head adaptor sequence and its next base, we could be more certain that break positions were accurate if bases in the tail region of the adaptor were aligned. Ten was arbitrarily chosen. We put no restriction of this type on the tail adaptor.

The trimmed reads whose lengths were longer than 10 bases were aligned using Nucleotide-Nucleotide BLAST 2.2.27+ [26] with default parameters against the *S. cerevisiae* strain S288C genome which was downloaded from Saccharomyces Genome Database (SGD). *S. cerevisiae* strain BY4741 had an incomplete reference genome and was missing chromosomes 3, 5 and 12. Nevertheless, the alignment results against BY4741 in the chromosomes that were present were the same as those using S288C.

Data Analysis

Identifying clusters of breaks

We followed Hackenberg et al. [27] in searching for clusters of break positions. In their work, they presented an algorithm that searched for clusters of CpGs and assigned a p-value associated with each cluster. Here, we regarded break positions as CpGs and used in-house R code to identify clusters.

Sequence logo

We extracted the first 50 bases of each uniquely mapped read and added the preceding 50 bases of sequence, which were obtained from the reference genome. We referred to these as stitched reads. The preceding 50 bases were taken from the 5' direction of the same strand (plus or minus) to which the read was mapped. Therefore, the break position is between the 50th and 51st base. Because each read had variable length, some reads were shorter than 50 bases, and thus the stitched read sequence length was not 100 for every read. Consequently, the number of reads that contributed to plotting each column of sequence logo was different (shown in Additional file 4: Table S2). The height of each sequence logo column is proportional to its information content; however, we also plotted a variation of the sequence logo where all columns have the same height. The package “seqLogo” was used for plotting [28].

Simulations

For each chromosome, we sampled position numbers on that chromosome with replacement. The number of items to sample was the observed number of breaks on the chromosome. Strands were assigned to each sampled number by sampling from 0 and 1 with a probability equals to proportion of reads mapped to plus strand in each chromosome for 0. The sampled numbers were regarded as read alignment positions.

For cluster simulation, we simulated 1,000 datasets and applied the previously described cluster identifying approach to the simulated data. For each chromosome, we counted the number of datasets that contained significant cluster(s). Bonferroni correction was used to account for multiple testing.

For sequence logo simulation, we extracted from the reference genome the sequence starting 50 bases to the left of the sampled position and extending 49 bases to the right of the sampled position and then plotted sequence logos based on these extracted sequences.

Statistical analysis

We constructed 256 2x2 contingency tables. For each table, rows are 4-mer occurrences in stitched reads vs. occurrences in the genome, and columns are one of the 4-mer sequences vs. the rest of the 4-mer sequences. Let “WXYZ” denote a certain 4-mer sequence. The first column of the first row, denoted by A, corresponded to the number of stitched reads that had sequence “WXYZ” at the breaks. The second column of the first row, denoted by B, corresponded to the number of stitched reads that had a sequence other than “WXYZ” at the breaks. In the second row, the first column, denoted by C, contained the number of positions in the genome that had sequence “WXYZ”. The second column of the second row, denoted by D, contained the number of positions in the genome that had a sequence other than “WXYZ”. We then computed Fisher’s exact test p-value from these contingency tables. We used the Bonferroni correction to set the significance cut-off at 0.05/256.

Abbreviations

EDSBs: Endogenous double-strand breaks; RIND-EDSBs: replication-independent EDSBs; NHEJ: Non-homologous end joining.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MPO performed all bioinformatics analysis and wrote the manuscript.

MPA prepared samples for sequencing and helped draft the manuscript.

JT prepared samples for sequencing.

AM conceived and designed the study, analyzed the data, and edited the manuscript.

All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Discovery-Based Development Grant and the Research Chair Grant 2011 from the National Science and Technology Development Agency (NSTDA), Thailand; by the Four Seasons Hotel Bangkok's 4th Cancer Care charity fun run in coordination with the Thai Red Cross Society; and by a grant from Ratchadaphiseksomphot Endowment Fund (grant number GDNS 57-002-23-001),

References

1. Pornthanakasem W, Kongruttanachok N, Phuangphairoj C, Suyarnsestakorn C, Sanghangthum T, Oonsiri S, Ponyeam W, Thanasupawat T, Matangkasombut O, Mutirangura A: LINE-1 methylation status of endogenous DNA double-strand breaks. *Nucleic Acids Res* 2008, 36(11):3667-3675
2. Kongruttanachok N, Phuangphairoj C, Thongnak A, Ponyeam W, Rattanatanyong P, Pornthanakasem W, Mutirangura A: Research Replication independent DNA double-strand break retention may prevent genomic instability. *Molecular cancer* 2010, 9(1)

3. Thongsroy J, Matangkasombut O, Thongnak A, Rattanatanyong P, Jirawatnotai S, Mutirangura A: Replication-Independent Endogenous DNA Double-Strand Breaks in *Saccharomyces cerevisiae* Model. *PLoS one* 2013, 8(8):e72706
4. Helleday T, Lo J, van Gent DC, Engelward BP: DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA repair* 2007, 6(7):923-935
5. Vamvakas S, Vock EH, Lutz WK: On the role of DNA double-strand breaks in toxicity and carcinogenesis. *CRC Crit Rev Toxicol* 1997, 27(2):155-174
6. Hoeijmakers J: Genome maintenance mechanisms for preventing cancer. *Nature* 2001, 411(6835):366-374
7. Negritto M: Repairing Double-Strand DNA Breaks. *Nature Education* 2010, 3(9):26
8. Khanna KK, Jackson SP: DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001, 27(3):247-254
9. Bustin M: Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol* 1999, 19(8):5237-5246
10. Reeves R, Adair JE: Role of high mobility group (HMG) chromatin proteins in DNA repair. *DNA repair* 2005, 4(8):926-938
11. Eden A, Gaudet F, Waghmare A, Jaenisch R: Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* 2003, 300(5618):455-455
12. Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, Jaenisch R: Induction of tumors in mice by genomic hypomethylation. *Science* 2003, 300(5618):489-492

13. Rodriguez J, Frigola J, Vendrell E, Risques R, Fraga MF, Morales C, Moreno V, Esteller M, Capellà G, Ribas M: Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res* 2006, 66(17):8462-9468
14. Pilch DR, Sedelnikova OA, Redon C, Celeste A, Nussenzweig A, Bonner WM: Characteristics of γ -H2AX foci at DNA double-strand breaks sites. *Biochem Cell Biol* 2003, 81(3):123-129
15. Kinner A, Wu W, Staudt C, Iliakis G: γ -H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic Acids Res* 2008, 36(17):5678-5694
16. Roca J: Topoisomerase II: a fitted mechanism for the chromatin landscape. *Nucleic Acids Res* 2009, 37(3):721-730
17. Timsit Y: Local sensing of global DNA topology: from crossover geometry to type II topoisomerase processivity. *Nucleic Acids Res* 2011,39(20):8665-8676
18. Attwood J, Yung R, Richardson B: DNA methylation and the regulation of gene transcription. *Cell Mol Life Sci* 2002, 59(2):241-257
19. Klose RJ, Bird AP: Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006, 31(2):89-97
20. Turek-Plewa J, Jagodzinski P: The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cellular and Molecular Biology Letters* 2005, 10(4):631

21. Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K: The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 2007, 450(7171):908-912
22. Jones PA, Takai D: The role of DNA methylation in mammalian epigenetics. *Science* 2001, 293(5532):1068-1070
23. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y: Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010, 466(7303):253
24. Apornthawan C, Pin-on P, Chaiyaratana N, Pongpanich M, Boonyaratanakornkit V, Mutirangura A: Upstream mononucleotide A-repeats play a cis-regulatory role in mammals through the DICER1 and Ago proteins. *Nucleic Acids Res* 2013, 41(19):8872-8885
25. Pages H, Aboyou P, Gentleman R, DebRoy S: Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.26.3
26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215(3):403-410
27. Hackenberg M, Previti C, Luque-Escamilla PL, Carpena P, Martínez-Aroza J, Oliver JL: CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics* 2006, 7(1):446
28. Bembom O: seqLogo: Sequence logos for DNA sequence alignments. Rpackage version 1.24.0

Figures and Legends

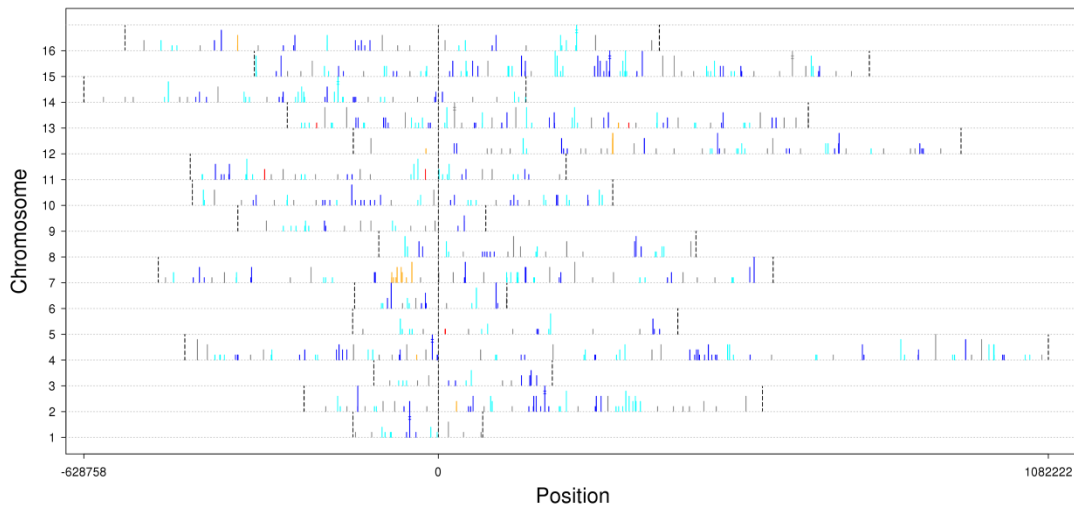


Figure 1 - Distribution of RIND-EDSBs in each chromosome.

The x-axis corresponds to chromosome position where position zero is the centromere (marked by a long vertical dotted gray line). The beginning and the end of each chromosome are marked by short vertical dotted gray lines. The y-axis represents the chromosome where chromosome 1 is at the bottom and chromosome 16 is at the top. Cyan, blue, orange, red and gray lines mark break positions. Blue and cyan lines indicate that the represented breaks are in a non-significant cluster. Cyan and blue are used alternately between each cluster. Orange and red lines indicate significant clusters. Gray lines indicate singleton breaks. The height of each line represents the number of reads aligned to that position. Lines with two horizontal dashes across indicate more than five reads aligned to that position.

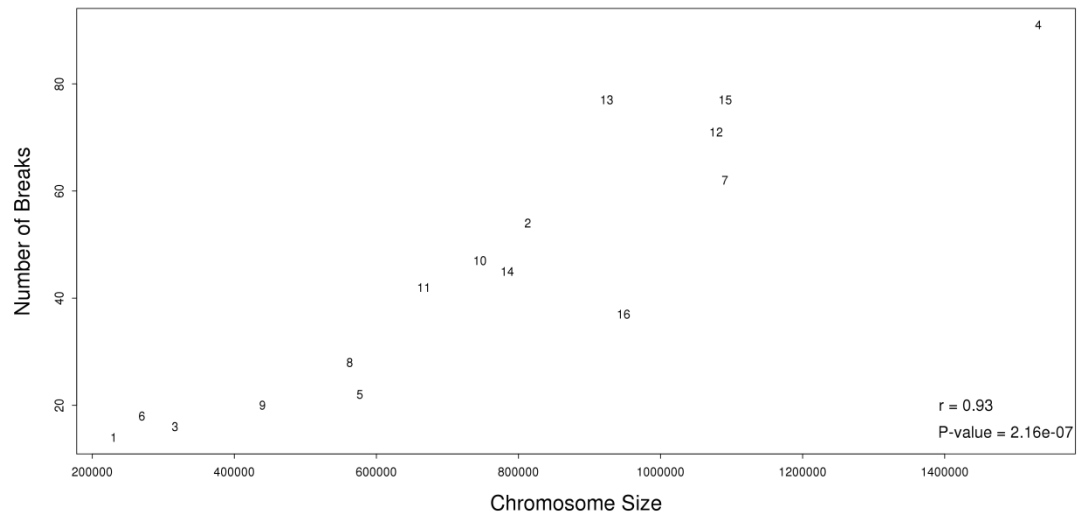


Figure 2 - Number of breaks and chromosome size.

The x-axis represents chromosome size and the y-axis represents the number of breaks on the chromosome. The numbers in the graph correspond to chromosome number.

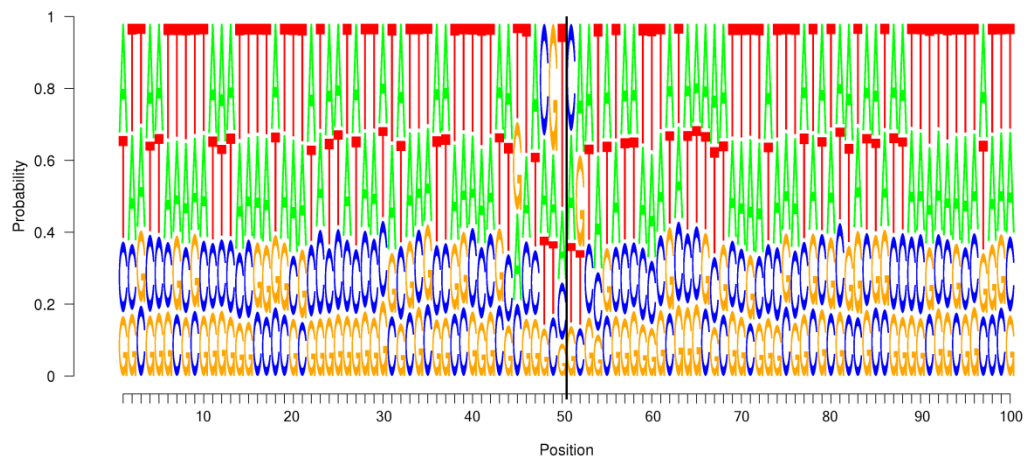


Figure 3 - Sequence logo.

Sequence logo where all columns have the same height. The break is located between position 50 and 51 and is marked by a vertical line.

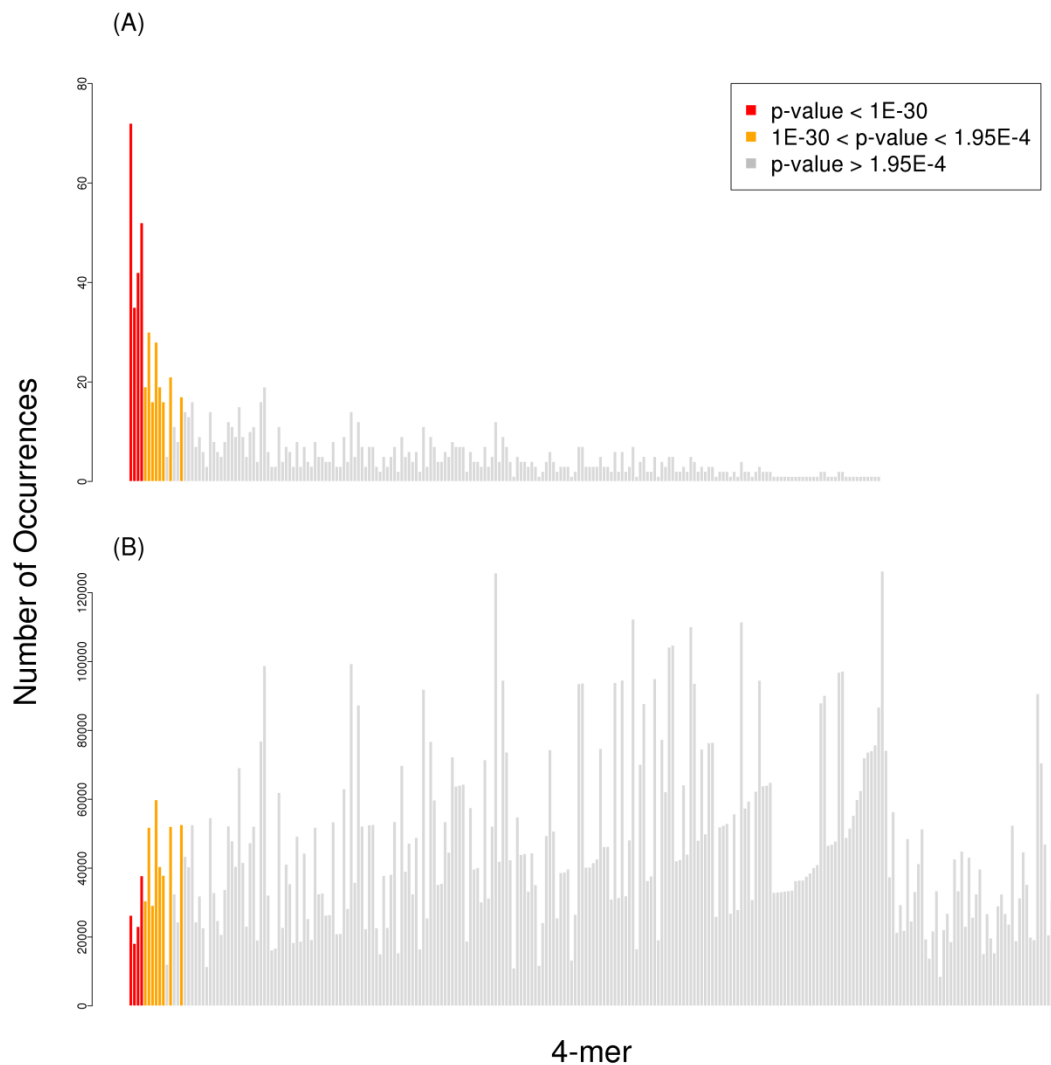


Figure 4 - Number of 4-mer sequences.

(A) Number of reads with each 4-mer sequence spanned over breaks. (B) Number of locations in the genome with each 4-mer sequence. The x-axis represents all 256 possible 4-mer sequences, and the 4-mers are sorted by odds ratio (Additional file 4: Table S1). The y-axis represents the number of occurrences of each 4-mer sequence.

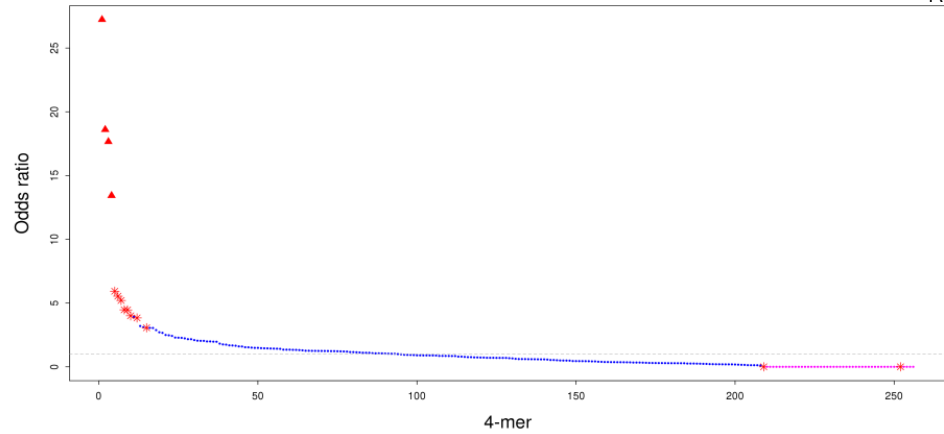


Figure 5 - Sorted odds ratio of each 4-mer.

The x-axis represents all 256 possible 4-mer sequences sorted by odds ratio (Additional file 4: Table S1). The y-axis represents the odds ratio. Red triangles denote significant 4-mer sequences with an odds ratio > 10. Red stars denote significant 4-mer sequences with an odds ratio < 10. Pink dots denote 4-mer sequences with odds ratios of zero.

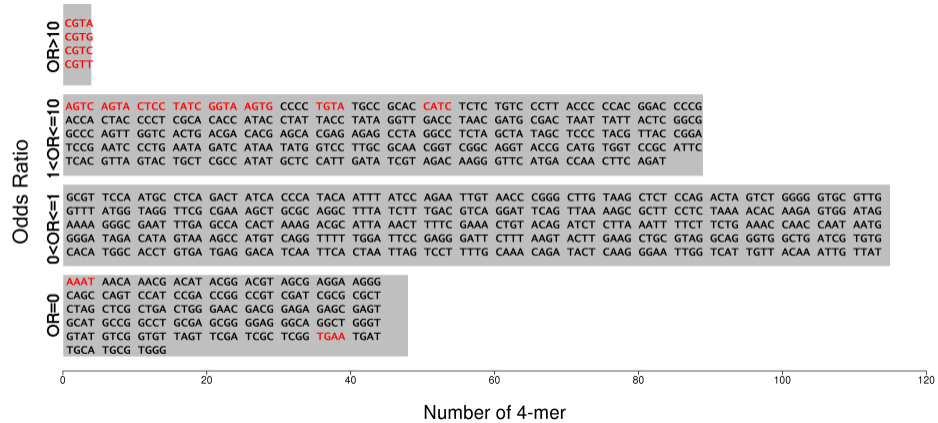


Figure 6 - The number of 4-mer sequences in each range of odds ratio.

The y-axis represents four ranges of odds ratios: OR = 0, $0 < OR \leq 1$, $1 < OR \leq 10$ and $OR > 10$, from the bottom to the top. The x-axis represents the number of 4-mers per range. The sequences of each 4-mer in the range are shown; those in red are significant.

Table

Table 1 - Fisher's exact test P-value.

4-mer	A*	B*	C*	D*	p-value	odds ratio	95% CI
CGTA	72	1185	26285	11774602	2.60E-74	27.23	21.13 - 34.56
CGTG	35	1222	18145	11782742	1.03E-31	18.61	12.89 - 26.01
CGTC	42	1215	23057	11777830	8.89E-37	17.66	12.66 - 24.02
CGTT	52	1205	37776	11763111	2.74E-39	13.43	9.98 - 17.75
AGTC	19	1238	30484	11770403	1.82E-09	5.93	3.55 - 9.29
AGTA	30	1227	51841	11749046	2.67E-13	5.54	3.72 - 7.95
CTCC	16	1241	29164	11771723	1.85E-07	5.20	2.96 - 8.49
TATC	28	1229	59853	11741034	2.01E-10	4.47	2.96 - 6.49
GGTA	19	1238	40440	11760447	1.46E-07	4.46	2.68 - 7.00
AGTG	16	1241	37903	11762984	5.21E-06	4.00	2.28 - 6.52
TGTA	21	1236	52077	11748810	3.92E-07	3.83	2.36 - 5.89
CATC	17	1240	52624	11748263	7.55E-05	3.06	1.78 - 4.92
AAAT	0	1257	126257	11674630	2.91E-06	0.00	0.00 - 0.27
TGAA	0	1257	90661	11710226	0.000128	0.00	0.00 - 0.38
XXXX ¹	0	1257	1685373	10115514	< 2.2e-16	0.00	0.00 - 0.02

* see description in Methods section

¹ combined 4-mer sequences that were never observed in reads

VITA

Name Miss Jirapan Thongsroy

DATE OF BIRTH 7th April 1982

PLACE OF BIRTH Nakorn Sri Tummarat, Thailand

INSTITUTIONS ATTENDED Chulalongkorn University, 2000-2003
Bachelor of science (Botany)

Kasetsart University, 2004-2007
Master of science (Genetic)

RESEARCH GRANT Ph.D. Grant by Chulalongkorn University
under the office of the Higher Education
Commission

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY