ROLES OF HISTONE BIOTINYLATION IN GENE REGULATION OF TRANSPOSABLE ELEMENTS

by

Yap Ching Chew

A DISSERTATION

Presented to the Faculty of

The Graduate College at University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Interdepartmental Area of Nutrition

Under the Supervision of Professor Janos Zempleni

Lincoln, Nebraska

July, 2008

ROLES OF HISTONE BIOTINYLATION IN GENE REGULATION OF TRANSPOSABLE ELEMENTS

Yap Ching Chew, Ph.D.

University of Nebraska, 2008

Adviser: Janos Zempleni

Chromatin modifications play crucial roles in gene regulation and DNA repair. Importantly, "epigenetic marks" such as histone modifications and DNA methylation have been implicated in transcriptional repression of retrotransposons. Repression of retrotransposons is critical for preventing transpositions and aberrant activities of host genes. Covalent binding of the vitamin biotin to histones, mediated by holocarboxylase synthetase, is one of the very few chromatin modifications that directly depend on nutrient supply in the diet. In this dissertation I demonstrate that chromosome stability depends on biotin. I show that binding of biotin to K9 in histone H2A and K12 in histone H4 is the mechanism that causes repression of retrotransposons. I also provide evidence that silencing of retrotransposons by histone biotinylation is mediated by increased transcription of anti-sense RNA. This observation suggests that RNAi is involved in the silencing of retrotransposons by histone biotinylation. In another project of my dissertation, I developed an avidin-based assay to quantify activities of histone debiotinylases. This assay is a useful tool to achieve one of the long-term goals of our

laboratory, i.e., the identification of histone debiotinylases in human cells. Identification of histone debiotinylases is a critical step to understand the regulation of histone biotinylation. Finally, I used mass spectrometry to identify novel biotinylation sites in human histones, and to identify histone marks that co-occur with biotinylation in a single histone molecule. This dissertation provides novel insights into the effects of nutrient status on cancer risk by a diet-dependent epigenetic mechanism of gene regulation, and it offers novel tools for the study of biological functions of histone biotinylation.

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Dr. Janos Zempleni for giving me the opportunity to become a Ph.D. student in his lab.

I wish to express my appreciation to Dr. Timothy Carr, Dr. Andrea Cupp, Dr. Ashraf Raza, Dr. Gautam Sarath and Dr. Marilynn Schnepf for serving on my doctoral committee. My special thanks to Dr. Sarath for all his input and guidance.

I am very thankful to my parents, family, and Marco for their understanding and support.

Gabi, Rocio, and Samudra, thank you for you time and effort in helping with my research.

I wish to thank my other lab members for their continued assistance and support.

I am very grateful to the staffs in Department of Nutrition and Health Sciences especially Donna Hahn, Jeannie Pittam, Lori Rausch, and Connie Wieser for assisting me in all possible ways.

Yap Ching Chew

July 2008

TABLE OF CONTENTS

LIST OF FIGURES …. 7

LIST OF TABLES …. 10

INTRODUCTION 11

REFERENCES …….. 26

CHAPTER I. A diet-dependent epigenetic mechanism that represses transposable elements …… 32

BACKGROUND ...……... 33

MATERIALS AND METHODS …….. 36

RESULTS ……...…….. 41

DISCUSSION ……... 53

REFERENCES ……. 56

CHAPTER II. An avidin-based assay for histone debiotinylase activity in human cell nuclei …… 62

BACKGROUND ...……... 63

MATERIALS AND METHODS …….. 65

RESULTS ……...…….. 68

DISCUSSION ……... 75

REFERENCES ……. 79

CHAPTER III. Identification of novel histone biotinylation sites by mass spectrometry …… 84

BACKGROUND ...……... 85

MATERIALS AND METHODS …….. 86

RESULTS ……...…….. 87

DISCUSSION ……... 88

REFERENCES ……. 90

CONCLUSIONS AND OUTLOOK …….. 92

REFERENCES ……. 96

APPENDIX …….. 98

Repression of LINE-1 elements by histone biotinylation is mediated by increased production of anti-sense RNA in human cells ……..... 98

BACKGROUND ...……... 98

MATERIALS AND METHODS …….. 101

RESULTS ……...…….. 103

DISCUSSION ……... 105

REFERENCES ……. 108

TABLE 1 ……..…… 113

TABLE 2 ……...…… 114

LIST OF PUBLICATIONS …… 115

LIST OF FIGURES

INTRODUCTION

Figure 1. Posttranslational modifications in histones……... 13

Figure 2. Classification of transposable elements……. 16

Figure 3. Structures of long terminal repeats (LTRs)……... 17

Figure 4. Transposition mechanisms……. 19

Figure 5. Structures of long interspersed nuclear elements (LINEs)…… 21

Figure 6. RNAi mechanism…….. 22

CHAPTER I

Figure 1. Structures of long terminal repeats (LTRs)……... 34

Figure 2. Biotinylation of carboxylases depends on biotin status in cell culture media…….. 37

Figure 3. Biotin supplementation leads to enrichment of H4K12bio and H3K9me2 at LTRs in human T lymphoma Jurkat cells…… 44

Figure 4. Biotin supplementation leads to enrichment of H4K12bio and H3K9me2 at LTRs in human mammary carcinoma T47D cells……. 45

Figure 5. Biotin supplementation leads to enrichment of H2AK9bio in LTRs in Jurkat cells…….. 46

Figure 6. Methylation status of the promoter region in the progesterone receptor gene before (-) and after (+) treatment with demethylating agent 5-aza-2'-deoxycytidine (AZA) is analyzed to confirm that AZA treatments were successful… 47

Figure 7. Biotin supplementation leads to a reduction in LTR transcription in Jurkat cells……. 48

Figure 8. Biotin supplementation leads to a reduction in LTR transcription in T47D cells…… 49

Figure 9. Production of viral (MMTV) particles in murine mammary carcinoma Mm5MT cells depends on biotin status in cell culture media…… 51

CHAPTER II

Figure 1. Spectrophotometric quantitation of TMB oxidation…….. 69

Figure 2. Histone debiotinylation depends on the amount of nuclear protein and incubation time……... 70

Figure 3. Debiotinylation of histone H1 by nuclear extracts from NCI-H69 cells depended on the pH of the incubation buffer……. 71

Figure 4. Release of biotin by proteolytic degradation of histone H1 was quantitatively minor……... 73

Figure 5. Activities of histone debiotinylases in nuclear extracts from human cells depended on the tissue from which cells originated…….. 74

Figure 6. Activities of histone debiotinylases in proteins fractionated from human plasma……..... 75

CHAPTER III

Figure 1. Biotinylation of K12 co-occurs with acetylation of K5 and K8 and mono-methylation of K16 in human histone H4…….. 88

APPENDIX

Figure 1. Structures of long interspersed nuclear elements (LINEs)…… 99

Figure 2. Biotinylation of carboxylases depends on biotin status in cell culture media…….. 102

Figure 3. Enrichment of H4K12bio, H3K9me2 and RNA Pol II at LINE-1 5'-UTR in human adenocarcinoma HeLa cells……... 104

LIST OF TABLES

CHAPTER III

Table 1. K8, K12, and K16 in histone H4 are targets for biotinylation. Histone biotinylation co-occurs with acetylation and mono-methylation…….. 87

APPENDIX

Table 1. PCR primers used for quantification of DNA from chromatin immunoprecipitation experiments (denoted "ChIP") and transcripts (denoted "Trans")…….. 113

Table 2. PCR primers used for quantification of DNA from chromatin immunoprecipitation experiments (denoted "ChIP")……. 114

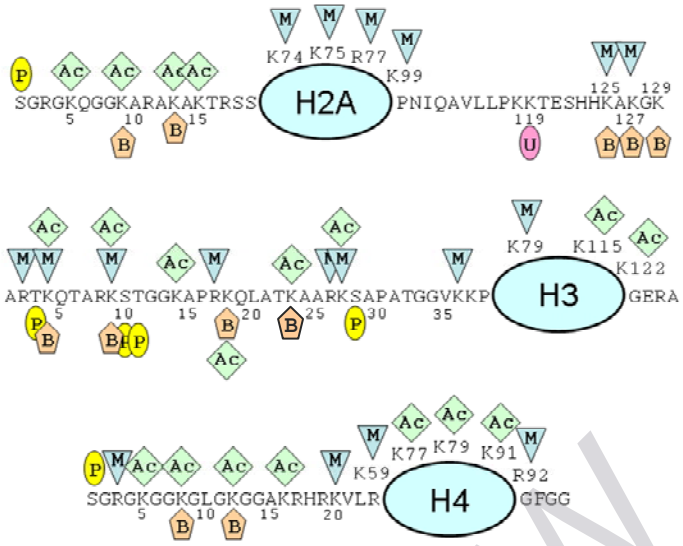
INTRODUCTION

Biotin (vitamin B_7, vitamin H) is an essential, water-soluble vitamin in humans. Biotin serves as a covalently bound coenzyme for acetyl-CoA carboxylases α and β, 3-methylcrotonyl-CoA carboxylase, propionyl-CoA carboxylase, and pyruvate carboxylase (1). These holocarboxylases are involved in fatty acid synthesis, leucine metabolism, and gluconeogenesis (1). Biotin also plays an important role in gene expression and cell signaling. For example, DNA microarray studies revealed that biotin affects the expression of more than 2000 genes in human lymphocytes, and that 28% of these biotin-dependent genes play roles in signal transduction (2). Moreover, biotin is attached covalently to histones (3). Biotinylation of histones plays a crucial role in heterochromatin structures (4) and gene repression (4, 5), and is the focus of this dissertation.

Biotin deficiency is prevalent in the American diet. For example, biotin deficiency has been observed in up to 50% of pregnant women (6-8). Importantly, previous studies demonstrated that mice fed with biotin-deficient diet during pregnancy caused an increase in fetal malformation and mortality (9). These studies support the possibility that biotin deficiency could be a cause of birth defects in humans. Additionally, biotin requirements may be increased during anticonvulsant therapy because anticonvulsants are known to inhibit biotin transport and to increase biotin catabolism (10). On the other hand, biotin supplementation is prevalent in the U.S. For example, about 20% of the U.S. population reports taking biotin supplements (11).

Chromatin Structure

Five major classes of histones have been identified in humans: H1, H2A, H2B, H3, and H4 (12). Histones carry a positive net charge due to the great abundance of arginine and lysine residues in these proteins (12). The binding of negatively charged DNA to positively charged histones is mediated by electrostatic interactions. In chromatin, stretches of 146 base pairs of DNA are wrapped around octamers of core histones (one H3-H3-H4-H4 tetramer and two H2A-H2B dimers) to form nucleosomal core particles. Linker histone H1 associates with the DNA connecting two core particles to complete the nucleosomal assembly (12). Amino acid residues in the N-terminal tails of histones are exposed at the nucleosomal surface. These tails are targets for posttranslational modifications such as acetylation, methylation, ubiquitination, phosphorylation, sumoylation, and poly(ADP-ribosylation) (12, 13) (Fig. 1). Some regions in C-terminal domains (e.g., hinge regions) are also exposed at the nucleosomal surface, and are targets for covalent modifications (12). These covalent modifications play important roles in the epigenetic control of chromatin structure, genome stability, and gene expression (14-16).



Adapted from Hassan & Zempleni (2006), J Nutr.

Figure 1. Posttranslational modifications in histones. Ac = acetylation, B = biotinylation, M = methylation, P = phosphorylation, U = ubiquitination (17).

Histone Biotinylation

Recently, a novel posttranslational modification of histones has been identified: biotinylation of distinct lysine residues (3). The following biotinylation sites have been identified: K9, K13, K125, K127, and K129 in histone H2A (18); K4, K9, K18, and perhaps K23 in histone H3 (19, 20); and K8 and K12 in histone H4 (21). Importantly, K12-biotinylation of histone H4 (H4K12bio) is a mark for repeat regions and heterochromatin, and plays a role in gene repression (4); H4K12bio co-localizes with the repression marker K9-dimethylation of histone H3 (H3K9me2) (4).

Biotinylation of histones is one of the very few chromatin modifications where dietary nutrient supply might become a limiting factor. The covalent attachment of biotin to histones is mediated by holocarboxylase synthetase (HCS) in an ATP-dependent

reaction (22, 23). HCS localizes primarily to the nuclear compartment, consistent with its role in chromatin structure (18, 22). Importantly, evidence has been provided that HCS binds to chromosomes (23).

Originally, it has been proposed that biotinidase also serves as a histone biotinyl by utilizing biocytin (biotinyl-ε-lysine) as a biotin donor (24). Previous studies (18, 25) demonstrated nuclear localization of biotinidase, consistent with its role in chromatin structure. However, the cellular distribution of biotinidase is controversial; Stanley et al. suggested that biotinidase localizes primarily to the cytoplasm (26). Recent studies suggest that biotinidase is not the dominant enzyme for histone biotinylation, based on the following lines of evidence (23). First, knockdown of biotinidase by siRNA has minor reduction on histone biotinylation in *Drosophila* when compared to HCS knockdown flies, and this reduction might be a secondary effect due to impaired biotin recycling. Second, reduction of biotinylation histones in biotinidase-deficient flies do not consistently display phenotypes such as decreased lifespan and heat tolerance. I propose that biotinylation of histones by biotinidase is a true phenomenon, but that biotinylation is driven by unphysiologically high concentrations of biocytin *in vitro*. I further propose that biotinidase might act as a histone debiotinylase *in vivo* (see below).

Chromatin modifications occur at multiple yet specific target sites, and the combinatorial modification profiles alter the interactions of histone tails with DNA or with chromatin-associated proteins such as transcription factors. This “histone code” is an important regulator of gene expression (27). It has been proposed that a given modification of a specific histone residue will affect subsequent additional modifications of the same histone molecule (28). Interactions of various modifications in both the same

and distinct molecules play important roles in the epigenetic control of chromatin structure. For example, S10-phosphorylated histone H3 inhibits H3K9me2 but is synergistically coupled with K9- and K14- acetylated histone H3 during mitogenic and hormonal stimulation in mammalian cells (13, 27). Moreover, various epigenetic mechanisms play crucial roles in the regulation of transcription. For example, H3K9me2 and methylation of cytosine residues in DNA are associated with transcriptionally repressed genes and repeat regions, whereas K4-trimethylation of histone H3 (H3K4me3) is associated with active genes (13, 27, 29). These mechanisms of gene regulation have been adopted in the repression of transposable elements as discussed below.

Transposable Elements

Transposable elements (TEs) constitute about 45% of the human genome (30-32) (Fig. 2). TEs are also known as transposons or “jumping genes”. They are mobile, repetitive DNA sequences that can move from one location of the genome to another, a mechanism called transposition (31). Transpositions cause mutations, impair genomic stability, and increase the amount of DNA in the genome, leading to human diseases such as cancer and autoimmunity (30, 33). For example, the insertion of a LINE-1 element into the blood clotting factor VIII gene on the X chromosome caused hemophilia A (30).

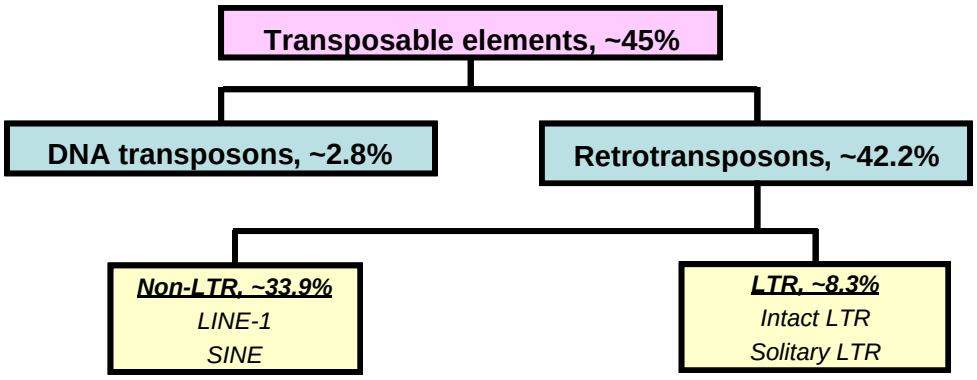


Figure 2. Classification of transposable elements.

There are two major classes of TEs, the retrotransposons (Type I) and DNA transposons (Type II); type classification is based on the mechanism of transposition (30, 31). Retrotransposons contain coding information for reverse transcriptase and their retrotransposition involves an RNA template as an intermediate (34). There are two major subclasses of retrotransposons, LTR and non-LTR, depending on the presence and absence of repeat elements called long terminal repeats (LTRs) (30, 31). Human endogenous retroviral elements (HERVs) are a subclass of retrotransposons and account for more than 10% of the human genome.

Long Terminal Repeats (LTRs)

Mammalian genomes contain two types of LTRs, i.e., intact LTR and solitary LTR (Fig. 3).

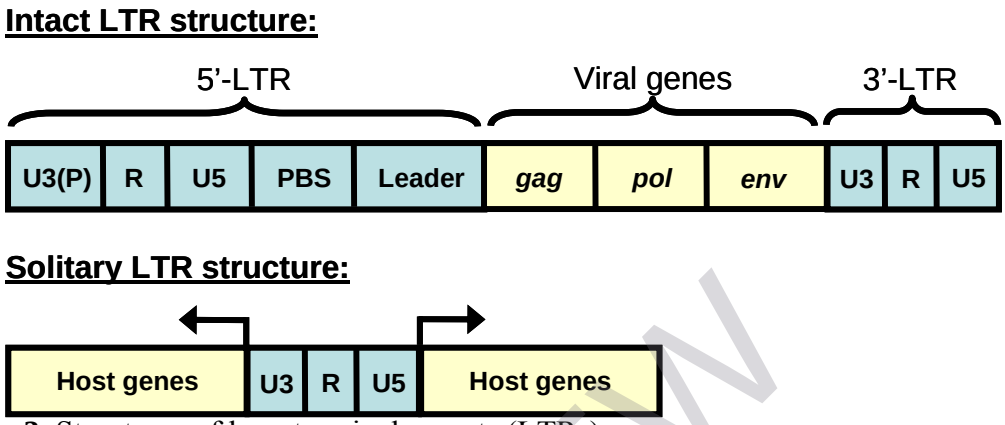


Figure 3. Structures of long terminal repeats (LTRs).

Intact LTRs contain the following regions (35):

(i) U3 = 200-1200 bp; the U3 region in the 5'-LTR contains the transcriptional promoter (P) elements responsible for transcription of both viral genes and host genes;

(ii) R = 18-250 bp; a repeat sequence containing the transcription start site and the polyadenylation site;

(iii) U5 = 75-250 bp; U5 is the site where reverse transcription of the transposon originates;

(iv) Primer binding site (PBS) = 16-19 bp; a complementary sequence of the 3'-end of the specific cellular tRNA molecules that serves as primer for the initiation of the reverse transcription reaction; and

(v) Leader = 90-500 bp; a non-translated region downstream of the transcription start site.

In intact LTR, the viral genes *gag* (group-specific antigen), *pol* (DNA polymerase), and *env* (envelop proteins) are flanked by two repeat regions: 5'-LTR and 3'-LTR. The *pol* gene contains open reading frames for reverse transcriptase, RNase H, and integrase or endonuclease, which are required for reverse transcription of viral RNA template into negative stranded-DNA and integration of viral genes into the host genome (34). The expression of retroviral genes is regulated by promoters in the 5'-LTR (36).

LTRs transpose using two different mechanisms: (*i)* replicative transposition, and (*ii*) non-replicative transposition (33) (Fig. 4). Replicative transposition increases the amount of DNA in the genome due to the underlying "copy-and-paste" mechanism; this mechanism plays a role in the transposition of endogenous retroviral elements. In replicative transposition, elements are first transcribed into viral mRNAs. Cellular tRNA binds to the PBS in the 5'-LTR region of viral RNA for the initiation of complementary DNA synthesis forming a linear double-stranded DNA with LTRs being present at both ends (33, 35). This double-stranded DNA is named proviral DNA. The RNA template and tRNA primer are removed by RNase H. Finally, the proviral DNA is inserted into the host genome by an enzyme called integrase. Retrotranspositions occur in germline cells, the newly generated LTR becomes permanently fixed in the host genome.

Non-replicative transposition is a "cut-and-paste" mechanism that does not increase the amount of DNA in the genome. First, the transposase makes a double-stranded cut in the donor DNA at the ends of the LTR transposon and makes a staggered cut in the recipient DNA. Each end of the donor DNA is then joined to an overhanging end of the recipient DNA.

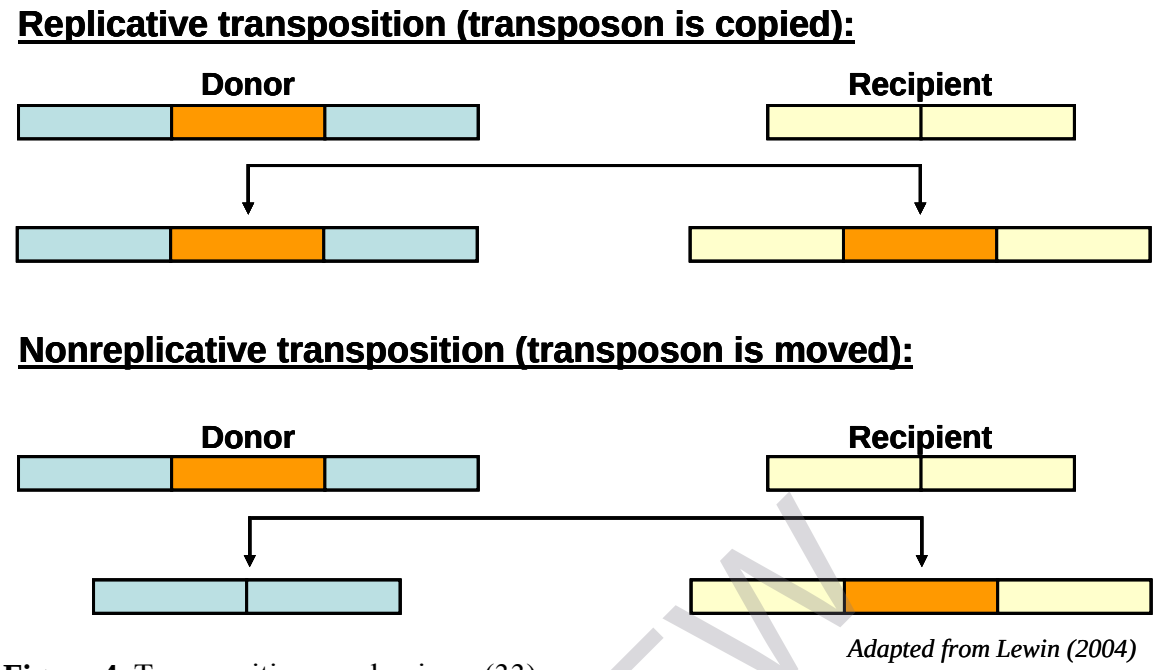


Figure 4. Transposition mechanisms (33).

The majority of mammalian LTR transposons are solitary LTRs. In solitary LTR, the retroviral genes have been deleted by recombination events between LTRs (30, 36-38). Solitary LTRs no longer flank viral genes but 3'-LTR or 5'-LTR remnants are maintained and may affect the transcription of adjacent host genes (37, 38).

Most LTRs are inactive due to deletions, mutations or hypermethylations, but 54 promoter-active LTRs were identified in human testes (36). Active LTRs cause genomic instability due to insertion of transposons into the genome, which may cause chromosome breakage and genome rearrangement. Thus, repression of both intact and solitary LTRs is important to prevent abnormal activity of viral and host genes, and to decrease the incidence of retrotransposition events.

Epigenetic mechanisms are known to play critical roles in the regulation of LTRs. For example, methylated cytosine residues, H3K9me2, K27-monomethylated histone H3, and

K20-trimethylated histone H4 are enriched at transposons (31, 39). Drug-induced hypomethylation of DNA (40) and hyperacetylation of histones (41, 42) cause transcriptional activation of LTRs. Notwithstanding the importance of the above modifications for chromosomal stability and human health, evidence has been provided that other, yet unknown, chromatin modifications are also critical for silencing of retroviral elements (42).

Long Interspersed Nuclear Elements-1 (LINE-1s or L1s)

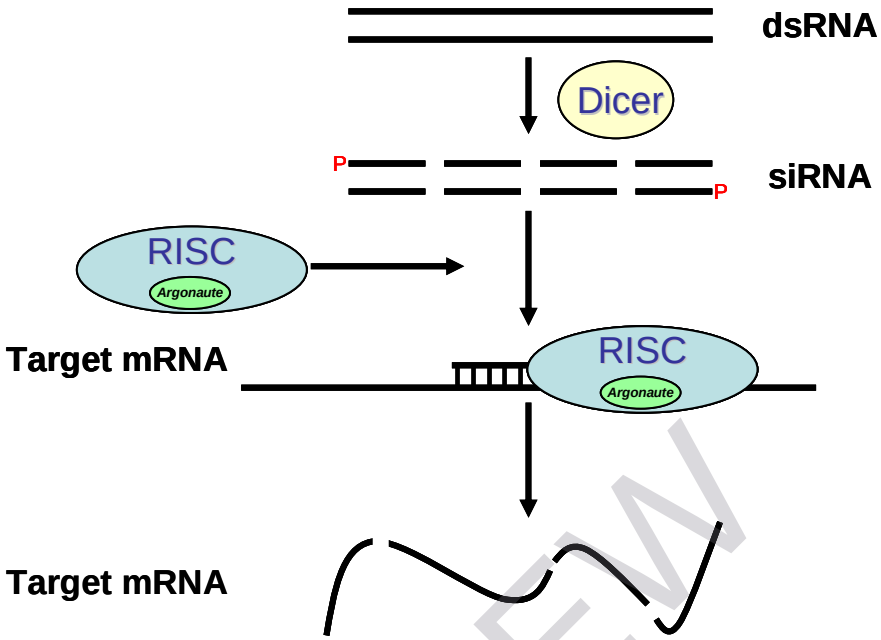
LINE-1 is a non-LTR retrotransposon that constitutes about 17% of the human genome (43, 44). A typical full-length LINE-1 is 4000-6000 bp long and contains the following regions (Fig. 5) (33): (*i*) the 5'-untranslated region (5'-UTR) containing an internal promoter; (*ii*) two non-overlapping open reading frames (ORFs) that are required for retrotransposition: ORF1 encodes a nucleic acid binding protein, and ORF2 encodes a endonuclease and reverse transcriptase; and (*iii*) the 3'-untranslated region (3'-UTR) containing a poly(A) tail. LINE-1 utilizes an RNA template and a free 3'-OH group at the endonuclease cut site on the genomic DNA as a primer to initiate transcription and retrotransposition. This process is known as target-primed reverse transcription (44).

Most LINE-1s are defective due to 5'-truncation, inversion, or mutations, but 80-100 active LINE-1s are found in average human diploid genome (44, 45). Additionally, diseases such as hemophilia, breast cancer, colon cancer, and Duchenne muscular dystrophy were reported to be caused by LINE-1 retrotranspositions (44-49).

Figure 5. Structures of long interspersed nuclear elements (LINEs). ORF = open reading frame, UTR = untranslated region.

Previous studies demonstrated that RNAi plays a role in the suppression of LINE-1 retrotranspositions. 5'UTR region in full length LINE-1 contains both sense and antisense promoter activities that drive the bi-directional transcription (50). Antisense mRNA produced by the antisense promoter in the 5'UTR is processed into LINE-1-specific siRNAs and are degraded by RISC (50). Additionally, previous studies showed that mutations in argonaute- and dicer-family proteins result in the reactivation of TEs (31).

RNA interference (RNAi) is the most common and well-studied RNA silencing mechanism that inhibits gene expression at the post-transcriptional level (Fig. 6). It is an evolutionarily conserved mechanism mediated by, and targeted against, RNA. This mechanism is initiated when an endonuclease called dicer cleaves long, double-stranded RNA (dsRNA) into 21-25 bp small-interfering RNAs (siRNAs) (31, 51). These siRNAs are incorporated into the RNA-induced silencing complex (RISC), which contains the protein argonaute. The siRNA-loaded RISC complex then targets mRNAs with complementary sequences to induce cleavage of the mRNA; cleavage is mediated by argonaute (31, 51). Degradation of mRNA by argonaute prevents the translation of the target genes.



Adapted from Kawasaki, Taira & Morris (2005), Cell Cycle

Figure 6. RNAi mechanism. dsRNA = double-stranded RNA, RISC = RNA-induced silencing complex, siRNA = small-interfering RNA.

Additionally, RNAi pathways also play a role in the recruitment of chromatin modifications to distinct genomic loci. Dicer-dependent siRNAs target the nascent mRNA while still being attached to RNA polymerase II and DNA; subsequently, siRNAs are incorporated into a complex called RNA-induced transcriptional silencing (RITS) complex (31). The cleavage of these nascent mRNAs recruits enzymes involved in chromatin modifications such as histone methyltransferases, DNA methyltransferases, and other chromatin-modifying factors (31), ultimately repressesing target genes.

DNA transposons do not utilize an RNA intermediate to integrate into the genome. A transposase enzyme recognizes the terminal inverted repeats (TIRs) which are present at both ends of DNA transposons. It excises the DNA sequences from one location of the

genome and then integrates it into another location. This process is known as cut-and-paste transposition (31).

My dissertation work focused on investigating novel roles of histone biotinylation in the repression of retrotransposons. I have divided my dissertation into four chapters as follows:

- Chapter I**: A diet-dependent epigenetic mechanism that represses transposable elements.**
- Chapter II: **An avidin-based assay for histone debiotinylase activity in human cell nuclei.**
- Chapter III: **Identification of novel histone biotinylation sites by mass spectrometry.**

CHAPTER I

Chapter I, the primary project of my dissertation, focuses on determining whether histone biotinylation contributes to repressing LTRs and retrotransposition events. Here, I show that H4K12bio and other species of biotinylated histones are enriched at LTR in human cells; that knockdown of HCS disrupts the accumulation of H4K12bio at LTR; that biotin depletion increases LTR transcript abundance and production of viral particles in human and murine cell lines and primary cells; that HCS knockdown and biotin depletion increase the frequency of chromosomal abnormalities and the incidence of transposition events in Jurkat cells and *Drosophila melanogaster*, respectively; and that histone biotinylation at LTR depends on DNA methylation, whereas DNA methylation does not depend on biotin.