2019

Argonaute 4 promotes genomic methylation

Kanwalat Chalertpet
Graduate School

ARGONAUTE 4 PROMOTES GENOMIC METHYLATION

Miss Kanwalat Chalertpet

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Biomedical Sciences

Inter-Department of Biomedical Sciences

Graduate School

Chulalongkorn University

Academic Year 2019

โปรตีนอะโกรนอท 4 กระตุ้นการเกิดเมทิลเลชั่นของจีโนม

น.ส.กัญวลัชญ์ ชเลิศเพ็ชร์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาชีวเวชศาสตร์ สหสาขาวิชาชีวเวชศาสตร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2562

Thesis Title	ARGONAUTE 4 PROMOTES GENOMIC METHYLATION
By	Miss Kanwalat Chalertpet
Field of Study	Biomedical Sciences
Thesis Advisor	Professor Apiwat Mutirangura, M.D., Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

.....
Dean of the Graduate School

(Associate Professor Dr. THUMNOON NHUJAK)

DISSERTATION COMMITTEE

.....
Chairman

(Assistant Professor Dr. AMORNPUN SEREEMASPUN, (M.D., Ph.D.))

.....
Thesis Advisor

(Professor Apiwat Mutirangura, M.D., Ph.D.)

.....
Examiner

(Associate Professor Dr. PATTAMAWADEE YANATATSANEEJIT)

.....
Examiner

(Associate Professor Dr. PITHI CHANVORACHOTE)

.....
External Examiner

(Dr. Siwanon Jirawatnotai)

กัญวลัชญ์ ชเลิศเพ็ชร์ : โปรตีนอะโกรนอท 4 กระตุ้นการเกิดเมทิลเลชั่นของจีโนม. (ARGONAUTE 4 PROMOTES GENOMIC METHYLATION) อ.ที่ปรึกษาหลัก : ศ. ดร. นพ.อภิวัฒน์ มุทิรางกูร

RNA-directed DNA methylation หรือ RdDM ถูกพบครั้งแรกในพืช เป็นกระบวนการที่อาร์เอนเอสายสั้นๆ ทำให้เกิดการควบคุมการแสดงออกของยีนในระดับเหนือพันธุกรรม (epigenetics) เช่นเดียวกันกับ CRISPR-Cas9 กระบวนการ RdDM มีส่วนประกอบที่สำคัญคือโปรตีนควบคุม และอาร์เอนเอสายนำ (sgRNA) เมื่อทั้งสองจับกัน คอมเพล็กซ์นี้จะไปจับกับลำดับของดีเอ็นเอที่เป็นคู่สมของอาร์เอนเอสายนำ แล้วชักนำให้เกิดการเติมหมู่เมทิลเลชั่นบริเวณลำดับเบสไซโตซีนที่ติดกับกวานีน (CpG) ขึ้น ดังนั้น การศึกษาครั้งนี้ได้พิสูจน์บทบาทของโปรตีนอะโกรนอท-4 ในกระบวนการ RdDM ของมนุษย์ ผลการศึกษาพบว่าโปรตีนอะโกรนอท-4 มีบทบาทในการกระตุ้นให้เกิดเมทิลเลชั่นของยีนที่จับกับโปรตีนอะโกรนอท-4 ได้ จากผลการศึกษาที่ได้นี้จึงได้ตัดต่อเปปไทด์ที่มีความสามารถในการแทรกผ่านเซลล์ (cell-penetrating peptides หรือ CPP) เข้ากับโปรตีนอะโกรนอท-4 โดยใช้ CPP 2 ชนิดคือ oligoarginines (R9-AGO4) และ xentry (X-AGO4) แล้วนำไปใช้ทรานสเฟ็กส์เซลล์มนุษย์ร่วมกับอาร์เอนเอสายนำ เพื่อใช้เป็นเครื่องมือในการชักนำให้เกิดการเติมหมู่เมทิลเลชั่นของลำดับดีเอ็นเอในบริเวณจำเพาะ ผลการทดลองพบว่า R9-AGO4-sgRNA เท่านั้น ที่สามารถชักนำให้เกิดการเติมหมู่เมทิลเลชั่นของ Alu and LINE-1 ซึ่งเป็น repetitive sequences และบริเวณยีนที่ต้องการได้ การศึกษาครั้งนี้ใช้ยีน *EML2* และ *CCNA1* เป็นตัวอย่าง ในขณะที่ X-AGO4-sgRNA สามารถแทรกผ่านเข้าเซลล์ได้ แต่ไม่สามารถชักนำให้เกิดการเปลี่ยนแปลงของระดับเมทิลเลชั่นได้ การชักนำให้เกิดเมทิลเลชั่นโดย R9-AGO4-LINE-1 หรือ *CCNA1* ยังทำให้เกิดการลดการแสดงออกของยีนทั้งสองเองด้วย นอกจากนี้ เมื่อทรานสเฟ็กส์ R9-AGO4-LINE-1 เข้าสู่เซลล์แล้วเซลล์มะเร็งยังโตช้าลงอีกด้วย ดังนั้น การศึกษาในครั้งนี้จึงเป็นครั้งแรกที่นำโปรตีน R9-AGO4 ทรานสเฟ็กส์เข้าสู่เซลล์มนุษย์ร่วมกับอาร์เอ็นเอสายนำ เพื่อใช้เป็นอีกเครื่องมือทางเลือกในการแก้ไขสภาวะเหนือพันธุกรรมของยีนหรือลำดับดีเอ็นเอที่ต้องการ

สาขาวิชา ชีวเวชศาสตร์ ลายมือชื่อนิสิต .....

ปีการศึกษา 2562 ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

5787850520 : MAJOR BIOMEDICAL SCIENCES

KEYWORD: Argonaute-4, RNA-directed DNA methylation, epigenomic-editing, RdDM, cell-penetrating peptides, guide RNA

Kanwalat Chalertpet : ARGONAUTE 4 PROMOTES GENOMIC METHYLATION . Advisor: Prof. Apiwat Mutirangura, M.D., Ph.D.

RNA-directed DNA methylation (RdDM) is the major small RNA-mediated epigenetic pathway in plants. Similar to genome and epigenome editing with the CRISPR-Cas9 system, RdDM is composed of a regulatory protein and a “guide” RNA (gRNA), which recruits the complex to specific genomic locations containing DNA sequences homologous to the RNA sequence. In humans, we found that Argonaute-4 (AGO4) colocalization has a strong correlation to promoter methylation of AGO4-binding genes. Here, we engineered cell-penetrating tagged oligoarginines (R9) tagged AGO4 (R9-AGO4) or xentry tagged AGO4 (X-AGO4) loaded with single guided-RNA (sgRNA) as a tool to specifically add methylation at genome locations where sgRNA is homologous to. We uncovered that only R9-AGO4-sgRNA complex can methylate both repetitive sequences (Alu and LINE-1) and unique copy gene (*EML2* and *CCNA1*) while X-AGO4-sgRNA can penetrate into cells, but did not provide methylation change. The induction of methylation by R9-AGO4-LINE-1 or R9-AGO4-CCNA1 also contributes to down-regulation of their own gene expression. Moreover, R9-AGO4-LINE-1 can delay proliferation rate of transfected cells compared with their control counterparts. Therefore, our study is the first time reported that the engineered R9-AGO4 protein can be transfected to cells with sgRNA and can be used as an alternative tool for epigenomic editing at any specific genes or DNA sequences in humans.

Field of Study:	Biomedical Sciences	Student's Signature .....
Academic Year:	2019	Advisor's Signature .....

ACKNOWLEDGEMENTS

I would like to express my deep sense of gratitude and profound respect to my thesis advisor Professor Dr. Apiwat Mutirangura who always provided me invaluable helps, excellent advice and constant encouragement though the course of my research in my PhD program with great patience and immense care. Without his guidance and persistent help this thesis would not have been possible.

I would like to thank my committee members: Assistant professor Dr. Amornpun Sereemaspun, Associate Professor Dr. Pattamawadee Yanatatsaneejit, Associate Professor Dr. Pithi Chanvorachote and Dr.Siwanon Jirawatnotai for spending their valuable time on discussing various issues related to my work.

I also thank Associate Professor Dr. Chatchawit Aporntewan, Dr. Piyapat Pin-on, Dr.Maturada Patchsung and Mr. Prakasit Rattanatanyong for their hard working, their helps and encouragement until I got my research publication in Frontiers in Genetics Journal.

I also thank my beloved friends, miss Mananya Techapatiphandee and all members in AM Lab for always being by my side to support me.

Finally, I would like to express my deepest gratitude to my family for their love and understanding all the time.

This thesis was supported by the Thailand Research Fund [DPG5980005, RSA5980060]; a 2011 Research Chair Grant from the National Science and Technology Development Agency (NSTDA), Thailand; the Anantara Siam Bangkok Hotel, Four Seasons Hotel Care for Cancer Fun Run in coordination with the Thai Red Cross Society; the Chulalongkorn University 100th Year Birthday Anniversary Doctoral Degree Scholarship and the Chulalongkorn Academic Advancement in Its 2nd Century Project and was partially supported by Chulalongkorn University under the Higher Education Commission.

Kanwalat Chalertpet

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I
INTRODUCTION

Background and Rationale

RNA-directed DNA methylation (RdDM) is firstly found in plants, described as a process that small non-coding RNA mediated epigenetic modification, resulting in transcriptional gene silencing (1). Later, proteins involved in this process have been extensively studied. One of many regulatory proteins in plant RdDM is Argonaute 4 (AGO4) (2). Mechanistically, RdDM is mainly involved in the biogenesis of 21- to 24-nucleotide small interfering RNAs (siRNAs) and in the loading of single guide RNA (sgRNA) onto AGO4 protein. This sgRNA-bound AGO4 complex recruits domains rearranged methyltransferase (DRM2) to promote *de novo* methylation at sites where sgRNA homology (3, 4). RdDM was further investigated in other organisms whether it exists, including in human cells. They found that small interfering RNA (siRNA) or short hairpin RNA (shRNA) can lead to transcriptional regulation in human cells (5, 6). However, mechanism of human RdDM have not been identified.

In humans, the long AGO family is composed of AGO1 to AGO4, each of which has different functions, abundances and level of expression in cells (7-9) although, their structures are similar and mostly conserved. Hence, to investigate RdDM pathway in humans, the AGO protein that affects RdDM is initially focused on identifying. Aporntewan performed a whole genome analysis between a promoter-methylation dataset (GSE20598) (10) and CLIPZ database (11), which lists all the known binding sites of AGO proteins in the entire genome of human embryonic kidney HEK293 cell, to explore association between human AGO proteins and promoter methylation. The results showed that human AGO4 has the strongest correlation between number of AGO4-binding sites on a promoter and promoter methylation levels (Pearson correlation

coefficient = 0.17 and *p*-value = 1.48×10^{-3}), explaining that the increasing number of AGO4-binding sites on a promoter is proportionate to the amount of methylation on a promoter (12). Together with human AGO4 is grouped into the long AGO family the same as AGO4 from *Arabidopsis* (13), AGO4 protein in that both organisms could share some common roles. Therefore, a possible hypothesis was set and stated that AGO4 might be involved in human RdDM similar to plants.

To verify the hypothesis above: the role of human AGO4 in DNA methylation, Pin-on selected a set of genes that showed a connection between AGO4-binding and promoter methylation from a wide-genome analysis above to be proved. He performed his experiments in tetracycline-regulated AGO4-knocked-down HEK293 cells (Tet+: no AGO4 expression; Tet- : AGO4 expression), the results showed that methylation levels of AGO4-binding genes were not different between Tet+ and Tet- groups. When a combination of tetracycline and 5'-azacytidine (Tet+, Aza+) was added to limit AGO4 mRNA and inactivate DNA methyltransferase activity, the methylation levels of AGO4-binding genes were significantly decreased; however, the methylation levels did not significantly change in Tet-, Aza+ group. The methylation level change did not find in gene lacks AGO4-binding. These results indicated that AGO4 could reintroduce DNA methylation to previously demethylated loci independent to 5'-azacytidine-related mechanism (12).

Moreover, Patsung proved whether siRNA can promote methylation in human cells and this process depends on AGO4, like RdDM in plants, she observed methylation level change of interspersed repetitive sequences (IRS) Alu by transfection Alu-siRNA in tetracycline-regulated AGO4-knocked-down HEK293 cells. The results showed that Alu-siRNA increased Alu methylation level when AGO4 was not limited (Tet); however, when AGO4 was depleted (Tet+), Alu methylation level did not change (12).

All results above bring us to the conclusions that the RdDM process exists in human cells and AGO4 is involved in this mechanism. However, we do not know what the direct action is of AGO4 protein in this pathway related to DNA methylation. In this work, it is aimed to prove how AGO4 localization is related to DNA methylation in the RdDM process: AGO4 either binds to methylated loci or induces methylation. As described above, each of human AGO proteins is different in many factors. It has been reported that AGO4 quantity and its expression in both mRNA and proteins are at the lowest levels compared to its protein family. Owing to these factors to address this objective, AGO4 overexpression will be carried out using the AGO4 plasmid transfection method; chromatin immunoprecipitation (ChIP), quantitative real-time PCR and combined bisulfite restriction analysis (COBRA) of IRS (ALU and LINE-1) will be performed to accomplish the objective above.

Currently, cell-penetrating peptides (CPPs) have been studied for therapeutic applications. CPP are peptides that can facilitate the entry of biologically conjugated cargoes into cells by various mechanisms (14); they can be categorized into many classes depending on their individual properties. Therefore, rigorous selection is considered for each type of CPP to be used. Oligoarginines (R8 – R12), which are grouped into Arginine-rich CPPs, a class of CPPs containing sequences of positively charged amino acid residues, have been used in many studies (15) because of their lower cytotoxicities, broad cell specificities and high uptake efficiencies (16-18).

However, a new class of CPPs, namely xentry, has been recently found, composed of 7-amino acid residues derived from an N-terminal region of the X-protein of the hepatitis B. Like other CPPs, xentry can penetrate plasma membrane; however, the distinct properties are 1) xentry penetrates cells through cell surface protein called syndecan-4, which is mostly found in epithelial cells, unlike the oligoarginines-based CPP mechanism 2) It is limited to permeate syndecan-deficient and non-adherent cells such as circulating blood cells and 3) It is shorter than other CPP classes, which are

normally about 10 to 30-amino acid residues, leading to an increase in capability of delivering cargoes into cells. Regarding these properties, xentry is stabilized for *in vivo* intravenous injection, enhancing a possibility to reach epithelial tissue targets (19).

As postulated, if the AGO4 protein can directly promote *de novo* methylation, CPP-conjugated AGO4 (CPP-AGO4) could be used as an alternative epigenetic editing tool. CPP-AGO4: 1) oligoarginines-tagged AGO4 (R9-AGO4) and 2) Xentry-tagged AGO4 (Xentry-AGO4 or X-AGO4) will be engineered for recombinant CPP-AGO4 protein production. Thereafter, they will be conjugated to single guide RNA (sgRNA) for transfection purposes in order to induce methylation level at specific loci homologous to sgRNA. Targeted methylation level changes, methylation-affected gene expression and cell phenotypic changes and cellular uptake efficiency caused by either of R9-AGO4 or X-AGO4 will be determined. These invented transfection technologies can be exploit in investigations of cellular processes or utilized in clinical treatments, alongside zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and the CRISPR-cas9 system (20, 21).

Hypomethylation of IRS either Alu or LINE-1 has been reported in carcinogenesis (22-24). Moreover, LINE-1 hypomethylation resulted in down-regulation of gene expression similar to gene promoter methylation (25, 26). Therefore, we attempt to use CPP-AGO4 to enhance the methylation level of both IRS and single genes. For single gene experiments, *EML2* and *CCNA1* will be selected; *EML2* is methylated only in brain tissues (27) whereas *CCNA1* is partially methylated in cervical cancer (28).

This study aims to explore how AGO4 localization is related to DNA methylation and to develop and compare new techniques using CPP-AGO4 (R9-AGO4 and X-AGO4) proteins loading sgRNA to induce the methylation of LINE-1, Alu, *EML2* and *CCNA1.* These finding could be important for a better understanding of the human RdDM process and may be used for epigenetic therapies in the future.

Research questions

1. How is human AGO4 localization related to DNA methylation in RdDM process?

2. Can R9-AGO4 or Xentry-AGO4 loading sgRNA specifically induce methylation, resulted in down-regulation in gene expression and phenotypic changes, and which one can provide higher efficiencies?

Hypotheses

1. AGO4 protein localizes to DNA sequences and induce DNA methylation.

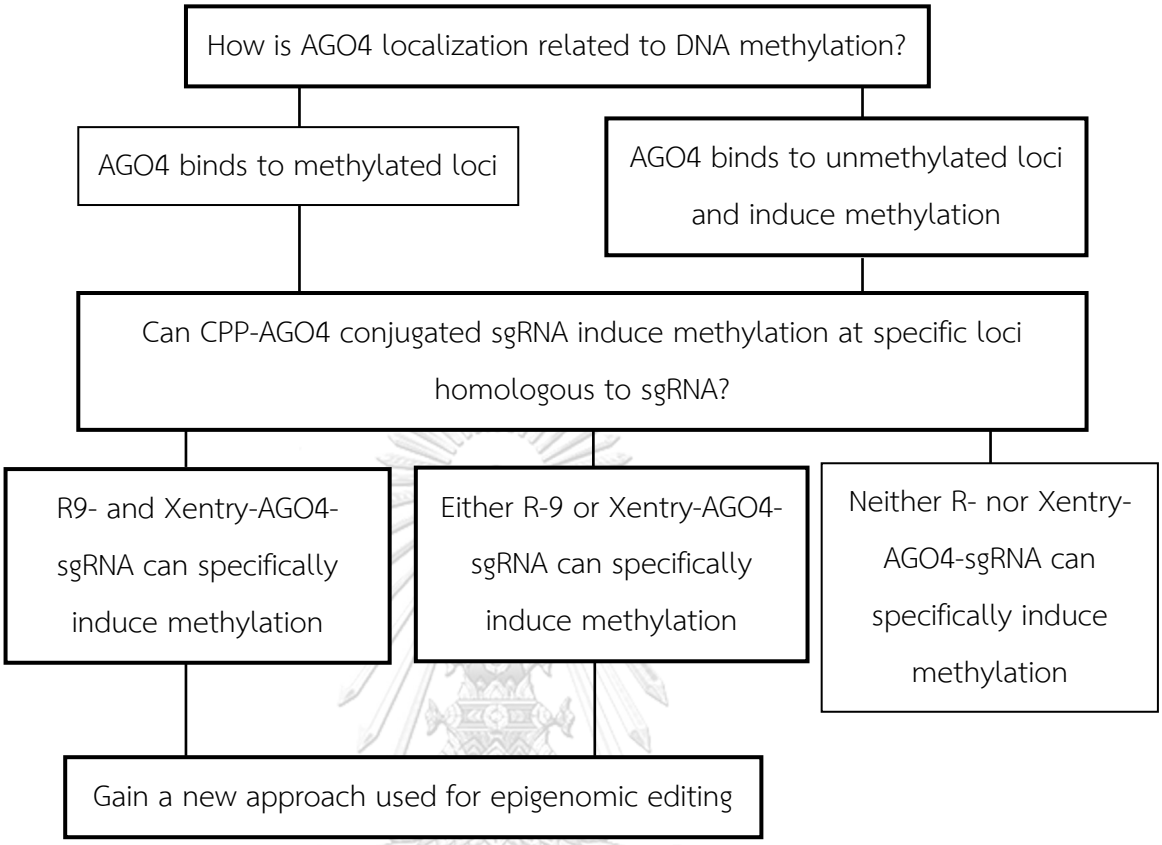
2. CPP-AGO4, (R9-AGO4 and/or Xentry-AGO4), can induce methylation at target sites affected in changes of gene expression and cell phenotype by which xentry-AGO4 could provide better efficiencies than R-AGO4.

Objectives

1. To investigate the association of AGO4 localization and DNA methylation.

2. To verify that CPP-conjugated AGO4 protein-coupling with sgRNA can induce DNA methylation at specific loci that can be used as an epigenomic editing tool.

Conceptual Framework



CHAPTER II
LITERATURE REVIEW

In 2000, the human genome project revealed that our genome contains about 25,000 protein-coding genes which is estimated only about 1-2% of all over 3 billion DNA base pairs in our genome. Although, our genes account for just a small percent of nucleotides compared to all in the genome, they encode proteins which determine underlying cellular phenotypes and functions (29). Due to this importance, our cells have complicated mechanisms to control gene expression at different levels.

The gene expression regulation of human can be occurred in distinct levels, including 1) transcriptional level; 2) RNA processing or post-transcriptional level; 3) protein synthesis or translational level and 4) post-translational level (30). However, each of level is mechanistically related. To illustrate, in transcription process, transcriptional regulatory proteins such as transcriptional activators, and transcriptional repressors not only interact with transcriptional factors or bind to DNA, they can also change chromatin structure which comprises of chromatin de-condensation and condensation through histone modification (31). The histone modification is an epigenetic event which takes part in controlling gene expression at post-translation level. Many previous studies showed that the histone modification has an interplay with DNA methylation (32). Therefore, it can also, in turn, regulate gene expression through RNA silencing at the transcriptional level (transcriptional gene silencing or TGS in short) (Figure 1).

TGS is simply described as inhibition of mRNA transcripts. However, the processes contributing to repression of mRNA synthesis consist of 1) RNA-directed DNA methylation or RdDM; 2) small RNA-independent DNA methylation and 3) chromatin modification (33). Of those mechanisms, only RdDM is classified into TGS. As mentioned earlier in chapter 1, the aim of this research study is to explore whether

argonaute-4 protein plays a role in human RdDM, for this reason, we only focus RdDM in more details.

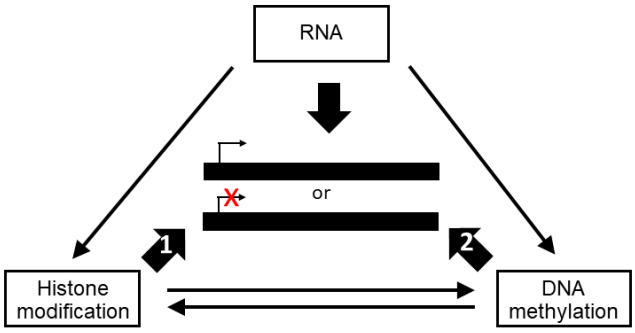


Figure 1 Control of gene expression caused by an interaction between histone modification in post-translational level (1) and DNA methylation in transcriptional level (2).

RNA-directed DNA methylation (RdDM)

Existence of RdDM was first discovered in 1994 when full DNA methylation was found only at specific sequences homologous to potato spindle tuber viroid (PSTVd) which was introduced into tobacco plants (1). After that the experiments were confirmed in other organisms (*Arabidopsis thaliana* and *Drosophila melanogaster*) revealing that non-coding RNA (RNA which is not translated into protein) can cause specific gene silencing through DNA methylation (34-36). Therefore, RdDM is described as the small RNA-mediated epigenetic pathway.

In plants, RdDM is mainly involved in the biogenesis of 24-nucleotide siRNAs and the recruitment of DNA methylation machinery by loading of single guide RNAs (sgRNA) onto the AGO4 protein to target specific loci (3, 4). Moreover, plant RdDM can be divided into 2 pathways: canonical and non-canonical RdDM pathway depending on a specialized transcriptional machinery (37) which involves different types of RNA polymerase, size of sgRNA and effector proteins such as argonate (AGO) protein.

1. Canonical RdDM Pathway

The canonical RdDM pathway is well-characterized compared to non-canonical pathway. It is the pathway which functions to maintain DNA methylation in cells (37). This pathway starts with the transcription of single-stranded RNA (ssRNA) by plant-specific RNA polymerase IV (Pol IV) at the existing DNA methylation loci. The ssRNA is then copied by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) generating long double-stranded RNA (dsRNA) at 26-45 nucleotides (nt) which are processed into small dsRNA (24 nucleotides) by DICER-LIKE 3 (DCL3). The 24-nt siRNAs are loaded into AGO4 or AGO6 protein. If sgRNA-bound AGO4 or AGO6 complex can be complementary to target loci transcripts which are transcribed by RNA polymerase V (Pol V), this complex will recruit DOMAINS REARRANGED METHYLTRASFERASE 2 (DRM2) to establish methylation and this will cause heterochromatin formation in consequence (38) (Figure 2). The canonical RdDM pathway is importance for transposable element (TE) repression (39).

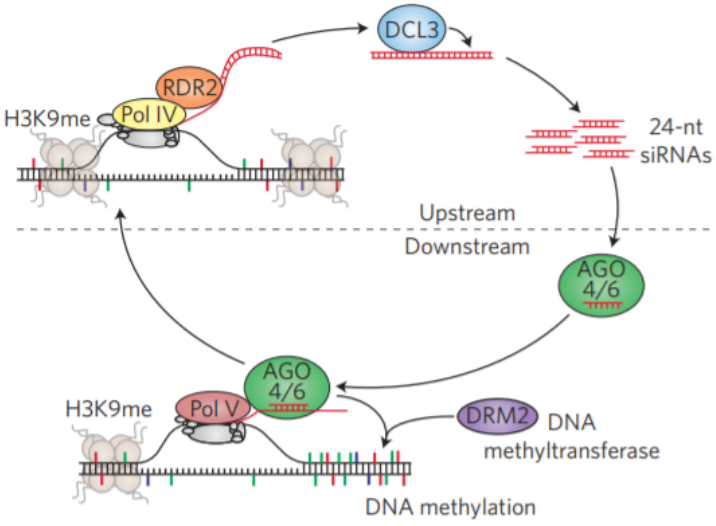


Figure 2 Canonical RdDM pathway (40) consists of 1) Biogenesis of siRNA which is processed by Pol IV, RDR2 and DCL3; 2) Loading of 24-nt siRNAs into AGO4 or AGO6; 3) Scaffold RNA production, processed by Pol V and 4) Protein recruitment for DNA methylation establishment, processed by DRM2.

2. Non-canonical RdDM pathway

The non-canonical RdDM pathway provides establishment of methylation at new target loci. Apart from 24-nt siRNA generated by Pol IV, RDR2 and DCL3 in canonical RdDM pathway, other small RNA can direct RdDM (40). However, the mechanism of this pathway is less understood. This pathway starts from transcription of newly inserted transposons by RNA polymerase II (Pol II) at palindromic TE, or inverted repeat; therefore, this pathway is responsible for host defense mechanism by degradation of viral, or TE mRNA (41). These transcripts are then copied into dsRNA by RNA-DEPENDENT RNA POLYMERASE 6 (RDR6) and then are cleaved into 21-nt siRNA by DCL-LIKE 2 (DCL2) and DCL4. The 21-nt siRNA is loaded into AGO1 and guide cleavage of transposon transcripts or RNAi. (37) (Figure 3). It is noticed that non-canonical RdDM is closely related to post-transcriptional gene silencing (PTGS) more than TGS. However, some of 21-nt siRNA can be also loaded into AGO6, this complex will trigger canonical pathway as described earlier.

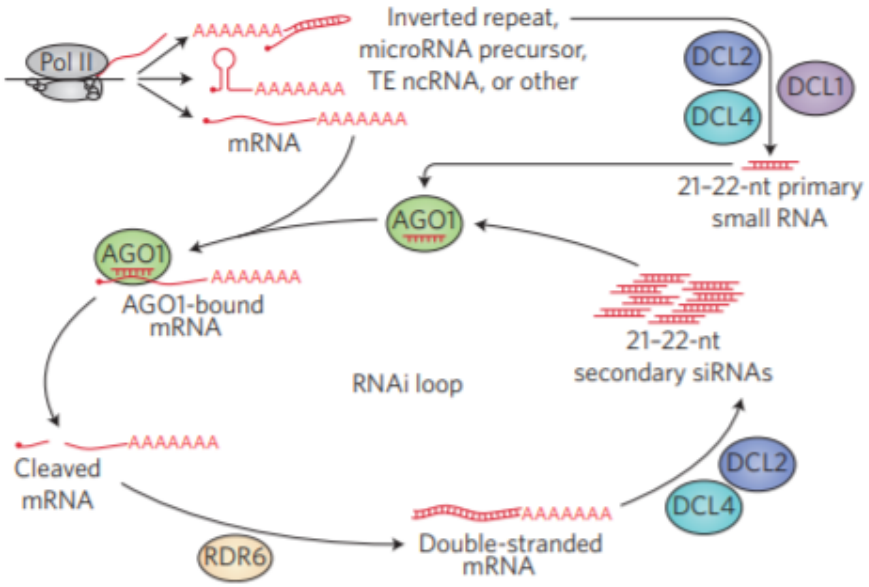


Figure 3 Non-canonical RdDM pathway (40) consist of 1) Biogenesis of siRNA which is processed by Pol II, RDR6 and DCL2 or DCL4; 2) Loading of 21-nt siRNAs into AGO1

and 3) Cleavage of mRNA (RNAi). However, some of 21-nt siRNA can be incorporated to AGO6, canonical RdDM is occurred as a result.

After plant RdDM was emerged, many researches have been widely investigated and found that it plays many roles in biological functions; for example, disease control through viral or TE repression, plant communication, stress response, genome stability, including plant growth and development (42). The question had been arisen whether this mechanism subsists in human. The first study of human RdDM began in 2004 (6).

Evidences of RdDM in Human

The study of small non-coding RNA-directed TGS in human was first reported a decade ago. It demonstrated that

AGO4 is related to promoter methylation

To determine whether any AGO protein is associated with DNA methylation, we first performed whole-genome comparisons between the promoter-methylation dataset (GSE20598) and AGO binding sites (CLIPZ database) (10, 11). The GSE20598 dataset provided the amount of DNA methylation in HEK293 cells; the CLIPZ database provided AGO-binding locations in the same cells. On the basis of these two datasets, the association between DNA methylation and AGO proteins was identified. Among all AGO proteins, AGO4 showed the strongest correlation with DNA methylation, as summarized from multiple probes (Pearson correlation coefficient = 0.17 and *p*-value = 1. $48x10^{-3}$) (Figure 4A to 4D).

The increasing number of AGO4-binding sites on a promoter is proportional to the amount of methylation on the promoter. Moreover, the binding position of AGO4 is correlated with the methylation of a single probe (Figure 4E). Sequence logos and positions in relation with transcriptional start sites (TSS) of AGO4 binding and

methylation are reported (Figure 5). Interestingly, AGO4 binds to the sense strand of methylated DNA located immediately after TSS (Figure 6) (12).

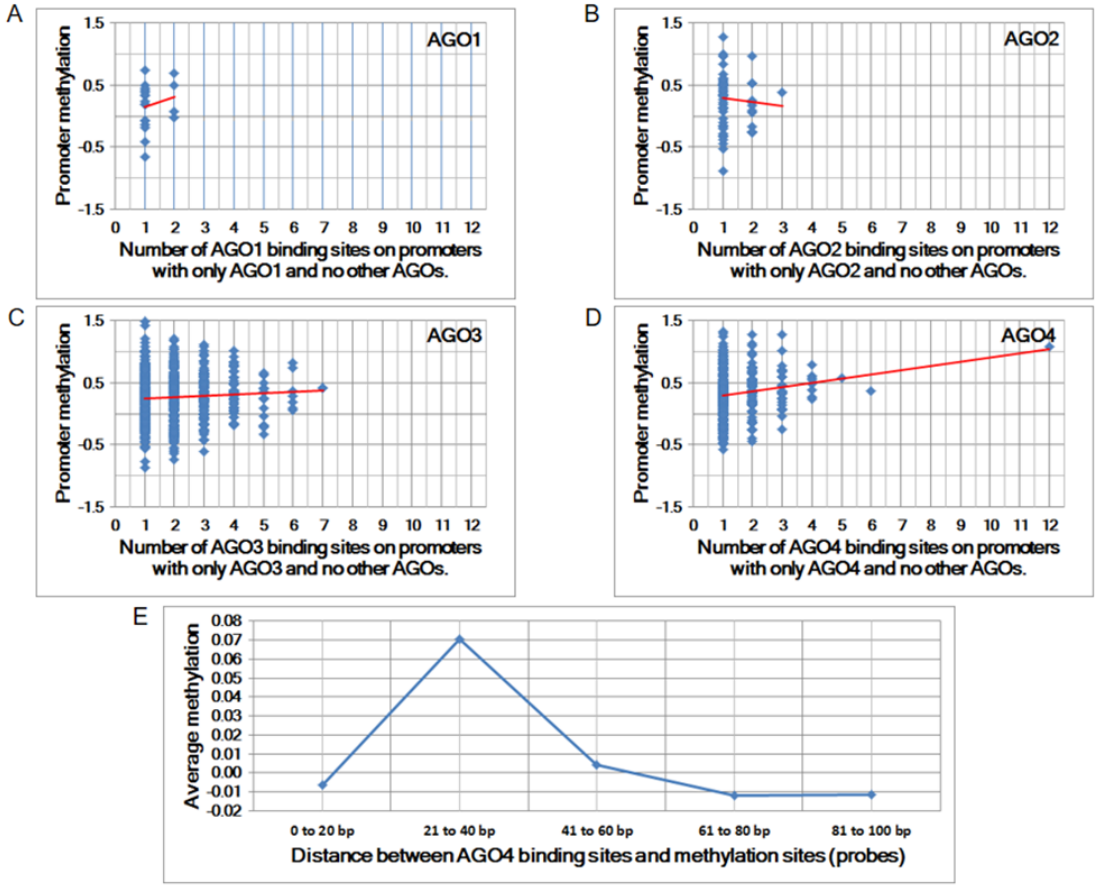


Figure 4 The correlation between AGO proteins and promoter methylation. (A to D) Promoter methylation vs. the number of AGO-binding sites. The Pearson correlation coefficients for A to D are 0.01, -0.06, 0.06, and 0.17, and the corresponding p-values are 8.63E-01, 6.51E-01, 1.11E-01, and 1.48E-03, respectively. (E) Average methylation vs. proximity to AGO binding sites. Each column represents the average amount of methylation at a single probe if the proximity is within a predefined distance (12).

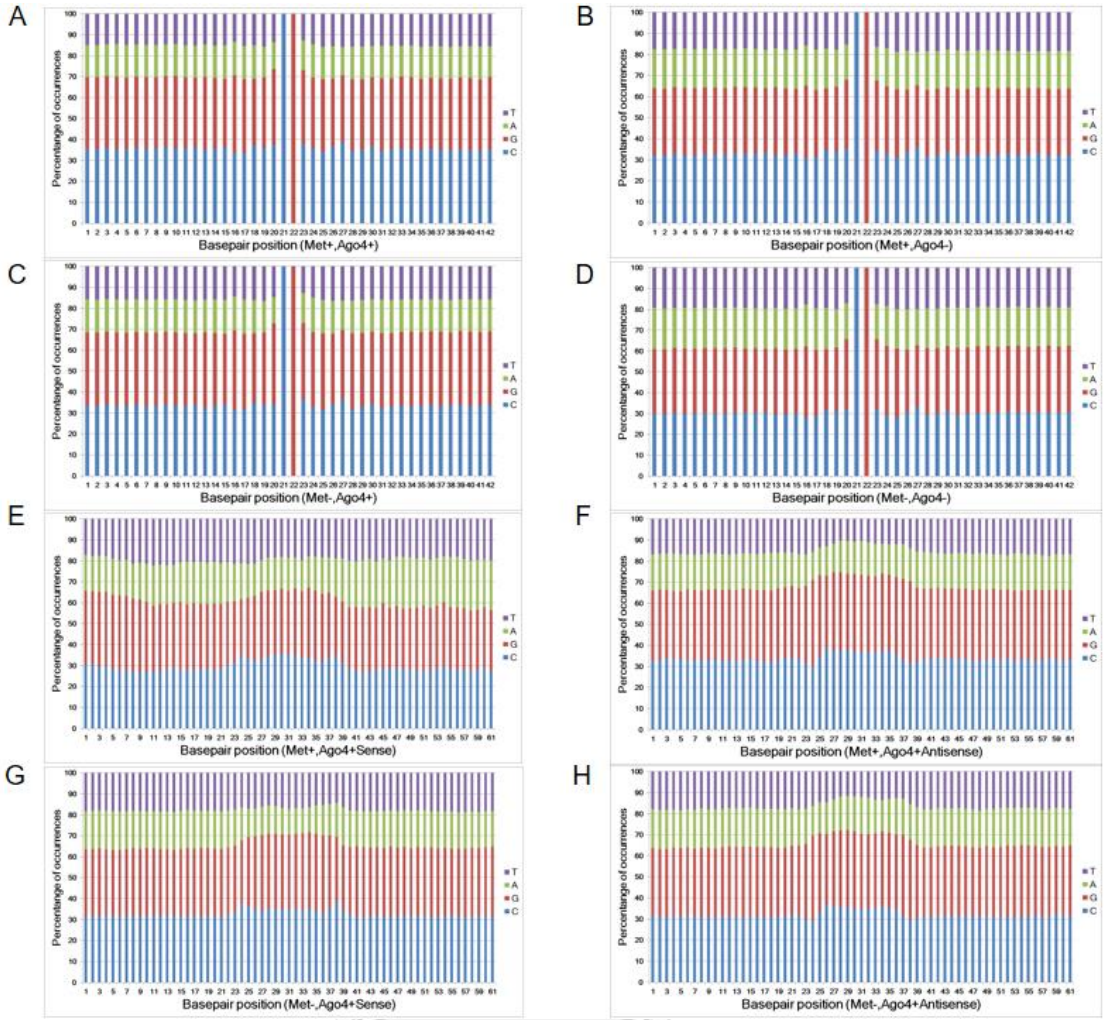


Figure 5 Distribution of DNA bases on methylation and AGO4 binding sites. Raw data of promoter methylation arrays provided both genomic location and methylation levels through 15 probes per one gene promoter. Each probe is 50 bp in length. Met+ is referred to as the methylation level in the 80th percentile. Remaining methylation levels were referred to as Met-. CLIPZ database provided the genomic locations of AGO binding sites. Each binding site is about 15-20 bp in length. AGO4+ is referred to as the overlab between an AGO4 binding site and a methylation probe, while AGO4- is referred to as no AGO4 binding sites on a methylation probe. (A to D) The distribution categorized by (Met+, AGO4+), (Met+, AGO4-), (Met-, AGO4+), and (Met-, AGO4-), respectively. Sense strands were collected from every CpG on the methylation probes by centering CpG between 20-bp flanking sequences on the left and on the right. A total

length was 42 bp. (E to H) The distribution categorized by (Met+, AGO4+sense), (Met+, AGO4+Antisense), (Met+, AGO4+Antisense), (Met-, AGO4+Sense), and (Met-, AGO4+Antisense), respectively where “Sense/Antisense” denotes the orientation of AGO4 binding sites. The sense strands of AGO4 binding sites were centered and then padded with flanking sequences. A total length was 61bp (12).

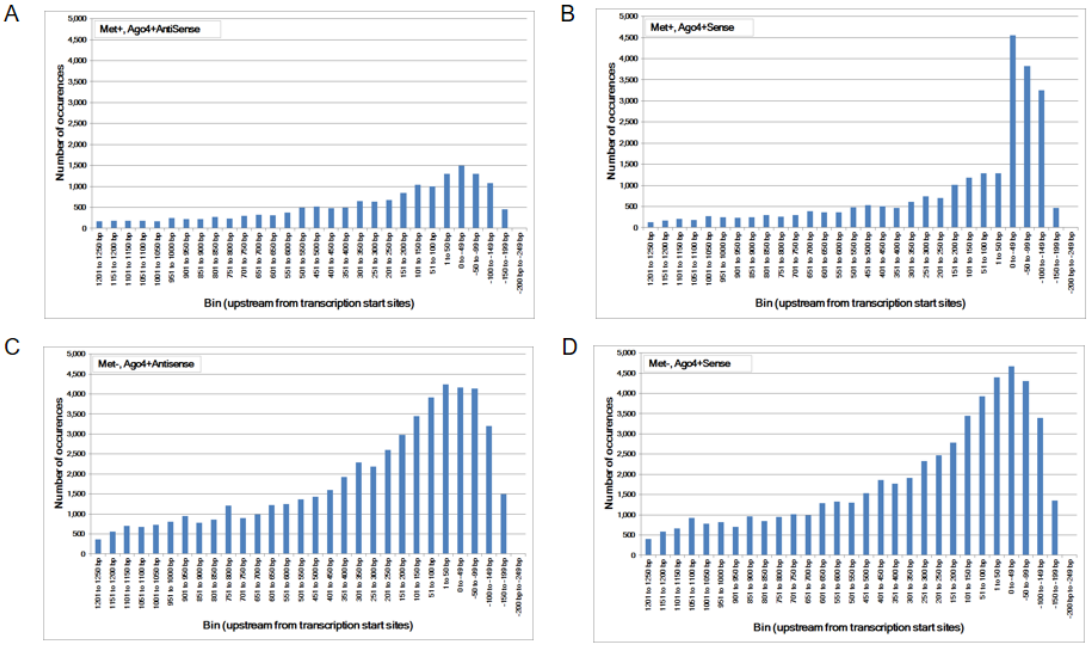


Figure 6 Distribution of AGO4 binding sites on gene promoters. (A to D) Distribution of (Met+, AGO4+antisense), (Met+, AGO4+sense), (Met-, AGO4+antiense), and (Met-, AGO4+sense), respectively. Definitions of Met+, Met-, AGO4+, AGO4-, sense, and antisense are given in figure 5 (12).

Moreover, we found that DNA methylation is associated with AGO-binding sites (Figure 7A and 7B). The highly methylated probes are in proximity with the binding sites of AGO1, AGO2, AGO3, and AGO4 proteins. This is indicated by odd ratio (OR) and 95% confidence interval (CI) > 1. In contrast, number of association studies of the control protein, PUMILIO2, shows OR and CI < 1. In figure 7A, it demonstrated OR and CI when the ChIP sequences that are perfectly matched with the human genome were counted whereas figure 7B reported approximate match. Other similar sequences which

contain mismatch, insertion, and deletion being alternatively targets. All four AGOs (AGO1, AGO2, AGO3 and AGO4) showed OR and CI > 1 in several tests. However, AGO4 showed the higher OR and the lowest *p*-value in most tests (Figure 7 and 8). By allowing approximate match, AGO4 shows the highest odds ratio at binding length ≥ 15 bp, OR = 4.09, 95% CI = 3.73 to 4.50, *p*-value =1.48E-214 (12).

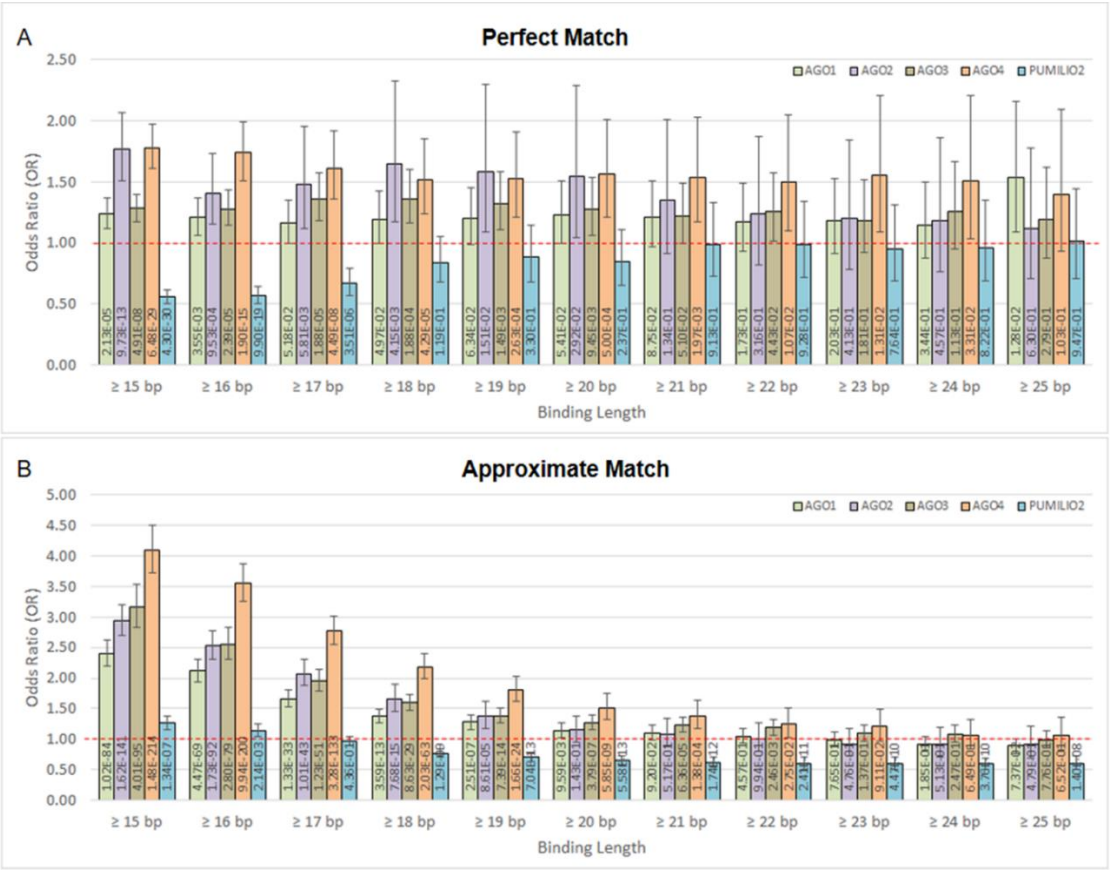


Figure 7 The association between methylation levels and protein-binding sites. The Chi-sqare p-values are calculated from 2x2 contingency tables which count the number of methylation probes. The rows of a contingency table separate highly and lowly methylated probes, whereas the columns separate probes that are in proximity with protein-binding sites and probes that are not. Note that a single probe might be in proximity with multiple protein. (A) Only ChIP DNAs that are perfectly matched with the

human genome are included. (B) Only ChIP DNAs that are approximately matched with the human genome are included. This also includes perfectly matched DNAs (12).

Perfect match

Binding Length	Protein	a	b	c	d	OR	95 %CI	*p*-value	Binding Length	Protein	a	b	c	d	OR	95 %CI	*p*-value
15	AGO1	5,562	560	15,866	1,978	1.24	1.12 - 1.37	2.13E-05	21	AGO1	923	91	20,505	2,447	1.21	0.97 - 1.51	8.75E-02
	AGO2	2,496	176	18,932	2,362	1.77	1.51 - 2.07	9.73E-13		AGO2	307	27	21,121	2,511	1.35	0.91 - 2.01	1.34E-01
	AGO3	7,603	762	13,825	1,776	1.28	1.17 - 1.40	4.91E-08		AGO3	1,141	112	20,287	2,426	1.22	1.00 - 1.49	5.10E-02
	AGO4	6,476	497	14,952	2,041	1.78	1.61 - 1.97	6.48E-29		AGO4	719	56	20,709	2,482	1.54	1.17 - 2.03	1.97E-03
	PUMILIO2	2,800	542	18,628	1,996	0.55	0.50 - 0.61	4.30E-30		PUMILIO2	407	49	21,021	2,489	0.98	0.73 - 1.33	9.13E-01
16	AGO1	3,038	306	18,390	2,232	1.20	1.06 - 1.37	3.55E-03	22	AGO1	790	80	20,638	2,458	1.18	0.93 - 1.49	1.73E-01
	AGO2	1,227	105	20,201	2,433	1.41	1.15 - 1.73	9.53E-04		AGO2	260	25	21,168	2,513	1.23	0.82 - 1.87	3.16E-01
	AGO3	4,057	393	17,371	2,145	1.27	1.14 - 1.43	2.39E-05		AGO3	925	88	20,503	2,450	1.26	1.01 - 1.57	4.43E-02
	AGO4	3,314	242	18,114	2,296	1.74	1.51 - 1.99	1.90E-15		AGO4	540	43	20,888	2,495	1.50	1.10 - 2.05	1.07E-02
	PUMILIO2	1,579	314	19,849	2,224	0.56	0.50 - 0.64	9.90E-19		PUMILIO2	383	46	21,045	2,492	0.99	0.72 - 1.34	9.28E-01
17	AGO1	1,929	199	19,499	2,339	1.16	1.00 - 1.35	5.18E-02	23	AGO1	655	66	20,773	2,472	1.18	0.91 - 1.53	2.03E-01
	AGO2	678	55	20,750	2,483	1.48	1.12 - 1.95	5.81E-03		AGO2	232	23	21,196	2,515	1.20	0.78 - 1.84	4.13E-01
	AGO3	2,578	232	18,850	2,306	1.36	1.18 - 1.57	1.88E-05		AGO3	706	71	20,722	2,467	1.18	0.92 - 1.52	1.81E-01
	AGO4	1,934	147	19,494	2,391	1.61	1.36 - 1.92	4.49E-08		AGO4	443	34	20,985	2,504	1.55	1.09 - 2.21	1.31E-02
	PUMILIO2	960	166	20,468	2,372	0.67	0.57 - 0.79	3.51E-06		PUMILIO2	346	43	21,082	2,495	0.95	0.69 - 1.31	7.64E-01
18	AGO1	1,462	147	19,966	2,391	1.19	1.00 - 1.42	4.97E-02	24	AGO1	557	58	20,871	2,480	1.14	0.87 - 1.50	3.44E-01
	AGO2	483	35	20,945	2,503	1.65	1.17 - 2.33	4.15E-03		AGO2	210	21	21,218	2,517	1.19	0.76 - 1.86	4.57E-01
	AGO3	1,927	172	19,501	2,366	1.36	1.16 - 1.60	1.88E-04		AGO3	560	53	20,868	2,485	1.26	0.95 - 1.67	1.13E-01
	AGO4	1,352	108	20,076	2,430	1.52	1.24 - 1.85	4.29E-05		AGO4	367	29	21,061	2,509	1.51	1.03 - 2.21	3.31E-02
	PUMILIO2	686	96	20,742	2,442	0.84	0.68 - 1.05	1.19E-01		PUMILIO2	317	39	21,111	2,499	0.96	0.69 - 1.35	8.22E-01
19	AGO1	1,213	121	20,215	2,417	1.20	0.99 - 1.45	6.34E-02	25	AGO1	464	36	20,964	2,502	1.54	1.09 - 2.16	1.28E-02
	AGO2	398	30	21,030	2,508	1.58	1.09 - 2.30	1.51E-02		AGO2	189	20	21,239	2,518	1.12	0.71 - 1.78	6.30E-01
	AGO3	1,614	147	19,814	2,391	1.32	1.11 - 1.58	1.49E-03		AGO3	440	44	20,988	2,494	1.19	0.87 - 1.62	2.79E-01
	AGO4	1,049	83	20,379	2,455	1.52	1.21 - 1.91	2.63E-04		AGO4	305	26	21,123	2,512	1.40	0.93 - 2.09	1.03E-01
	PUMILIO2	523	70	20,905	2,468	0.88	0.68 - 1.14	3.30E-01		PUMILIO2	299	35	21,129	2,503	1.01	0.71 - 1.44	9.47E-01
20	AGO1	1,037	101	20,391	2,437	1.23	1.00 - 1.51	5.41E-02									
	AGO2	350	27	21,078	2,511	1.54	1.04 - 2.29	2.92E-02									
	AGO3	1,372	129	20,056	2,409	1.28	1.06 - 1.54	9.45E-03									
	AGO4	869	67	20,559	2,471	1.56	1.21 - 2.01	5.00E-04									
	PUMILIO2	439	61	20,989	2,477	0.85	0.65 - 1.11	2.37E-01									

Approximate match

Binding Length	Protein	a	b	c	d	OR	95 %CI	*p*-value	Binding Length	Protein	a	b	c	d	OR	95 %CI	*p*-value
15	AGO1	17,795	1,703	3,633	835	2.40	2.19 - 2.63	1.02E-84	21	AGO1	3,940	432	17,488	2,106	1.10	0.98 - 1.22	9.20E-02
	AGO2	12,918	863	8,510	1,675	2.95	2.70 - 3.21	1.62E-141		AGO2	894	99	20,534	2,439	1.07	0.87 - 1.33	5.17E-01
	AGO3	20,003	2,072	1,425	466	3.16	2.82 - 3.54	4.01E-95		AGO3	5,155	520	16,273	2,018	1.23	1.11 - 1.36	6.36E-05
	AGO4	19,180	1,715	2,248	823	4.09	3.73 - 4.50	1.48E-214		AGO4	1,804	158	19,624	2,380	1.38	1.17 - 1.64	1.38E-04
	PUMILIO2	14,456	1,580	6,972	958	1.26	1.15 - 1.37	1.34E-07		PUMILIO2	1,425	265	20,003	2,273	0.61	0.53 - 0.70	1.74E-12
16	AGO1	16,118	1,496	5,310	1,042	2.11	1.94 - 2.30	4.47E-69	22	AGO1	3,338	381	18,090	2,157	1.04	0.93 - 1.17	4.57E-01
	AGO2	10,357	685	11,071	1,853	2.53	2.31 - 2.77	1.73E-92		AGO2	717	85	20,711	2,453	1.00	0.79 - 1.26	9.94E-01
	AGO3	19,025	1,920	2,403	618	2.55	2.31 - 2.82	2.80E-79		AGO3	4,060	418	17,368	2,120	1.19	1.06 - 1.32	2.46E-03
	AGO4	17,729	1,458	3,699	1,080	3.55	3.26 - 3.87	9.94E-200		AGO4	1,252	121	20,176	2,417	1.24	1.02 - 1.50	2.75E-02
	PUMILIO2	12,752	1,430	8,676	1,108	1.14	1.05 - 1.24	2.14E-03		PUMILIO2	1,125	215	20,303	2,323	0.60	0.51 - 0.70	2.41E-11
17	AGO1	13,188	1,247	8,240	1,291	1.66	1.53 - 1.80	1.33E-33	23	AGO1	2,825	340	18,603	2,198	0.98	0.87 - 1.11	7.65E-01
	AGO2	6,763	462	14,665	2,076	2.07	1.87 - 2.30	1.01E-43		AGO2	589	76	20,839	2,462	0.92	0.72 - 1.17	4.76E-01
	AGO3	16,885	1,663	4,543	875	1.96	1.79 - 2.14	1.23E-51		AGO3	3,166	347	18,262	2,191	1.09	0.97 - 1.23	1.37E-01
	AGO4	14,358	1,074	7,070	1,464	2.77	2.55 - 3.01	3.28E-133		AGO4	967	96	20,461	2,442	1.20	0.97 - 1.49	9.11E-02
	PUMILIO2	9,999	1,205	11,429	1,333	0.97	0.89 - 1.05	4.36E-01		PUMILIO2	976	187	20,452	2,351	0.60	0.51 - 0.71	4.47E-10
18	AGO1	9,490	932	11,938	1,606	1.37	1.26 - 1.49	3.59E-13	24	AGO1	2,395	306	19,033	2,232	0.92	0.81 - 1.04	1.85E-01
	AGO2	3,650	279	17,778	2,259	1.66	1.46 - 1.89	7.68E-15		AGO2	504	65	20,924	2,473	0.92	0.71 - 1.19	5.13E-01
	AGO3	13,149	1,267	8,279	1,271	1.59	1.47 - 1.73	8.63E-29		AGO3	2,462	272	18,966	2,266	1.08	0.95 - 1.23	2.47E-01
	AGO4	9,281	658	12,147	1,880	2.18	1.99 - 2.40	2.03E-63		AGO4	764	86	20,664	2,452	1.05	0.84 - 1.32	6.49E-01
	PUMILIO2	6,505	920	14,923	1,618	0.77	0.70 - 0.84	1.29E-09		PUMILIO2	876	172	20,552	2,366	0.59	0.50 - 0.69	3.76E-10
19	AGO1	6,693	666	14,735	1,872	1.28	1.16 - 1.40	2.51E-07	25	AGO1	2,034	269	19,394	2,269	0.88	0.77 - 1.01	7.37E-02
	AGO2	1,964	173	19,464	2,365	1.38	1.17 - 1.62	8.61E-05		AGO2	436	57	20,992	2,481	0.90	0.68 - 1.20	4.79E-01
	AGO3	9,348	910	12,080	1,628	1.38	1.27 - 1.51	7.39E-14		AGO3	1,897	229	19,531	2,309	0.98	0.85 - 1.13	7.76E-01
	AGO4	5,046	370	16,382	2,168	1.80	1.61 - 2.02	1.66E-24		AGO4	625	70	20,803	2,468	1.06	0.82 - 1.36	6.52E-01
	PUMILIO2	3,739	590	17,689	1,948	0.70	0.63 - 0.77	7.04E-13		PUMILIO2	801	154	20,627	2,384	0.60	0.50 - 0.72	1.40E-08
20	AGO1	4,877	520	16,551	2,018	1.14	1.03 - 1.27	9.59E-03									
	AGO2	1,242	129	20,186	2,409	1.15	0.95 - 1.38	1.43E-01									
	AGO3	6,632	661	14,796	1,877	1.27	1.16 - 1.40	3.79E-07									
	AGO4	2,843	233	18,585	2,305	1.51	1.32 - 1.74	5.85E-09									
	PUMILIO2	2,111	367	19,317	2,171	0.65	0.57 - 0.73	5.58E-13									

Figure 8 The 2x2 table shows the association between methylation levels and protein binding sites. (A) Only ChIP DNAs that are perfectly matched with the human genome are included. (B) Only ChIP DNAs that are approximately matched with the human genome are included (12).

The Argonaute 4 protein is involved in *de* novo methylation of AGO4-binding genes

To investigate AGO4-binding genes and the role of AGO4 in DNA methylation, we obtained target genes from analyses of the GSE20598 and CLIPZ databases. These bioinformatics data provided us with a large number of genes that show a connection between AGO4-binding and promoter methylation. We randomly selected eight loci of three genes: *TRDP, C16ORF89* and *ATAT1* that have high level of DNA methylation and AGO4 binding to study the interaction between AGO4 and DNA methylation; however, *C16ORF89* was showed as an example. *MSN*, which lacks AGO4 binding, was used as a negative control. The experiments were carried out by depleting endogenous AGO4 using a tetracycline-regulated shRNA-expression system in the HEK293 cell line (AGO4sh). The timing of tetracycline (Tet) and 5-AC treatment is shown (Figure 9A). Depletion of AGO4 mRNA expression was confirmed (Figure 9B). After the cells were cultured with Tet (Tet+) or without Tet (Tet-) for 7 days, they were collected to confirm the AGO4-binding genes. All four selected genes bound large quantities of AGO4; *C16ORF89* is provided as an example (Figure 9C), with the exception of *MSN* (Figure 9D). Methylation levels were also evaluated; however, there was no significant difference in methylation levels between the Tet+ and Tet- cells at all nine loci of the four known AGO4-binding genes or *MSN* (Figure 9E and 9F at Tet- and Tet+ bars). Thereafter, the HEK293 cell line was treated with a combination of Tet and 5-AC to limit AGO4 mRNA and inactivate DNA methyltransferase activity for two additional days, until day 9; methylation changes were then investigated. From days 7 to 9, the methylation levels of

all loci were significantly reduced in Tet+ and 5-AC+ shAGO4-induced cells compared with Tet- and 5-AC+ cells (Figure 9E and 9F at Tet- 5-AC+ and Tet+ 5-AC+ bars). This finding suggested a role of AGO4 in the demethylation of AGO4-bound sequences.

The mechanism by which AGO4 prevented demethylation by 5-AC suggested a role of AGO4 in *de novo* DNA methylation. To observe the *de novo* methylation function of endogenous AGO4, cells were cultured continuously with Tet and 5-AC for nine days. Then, the cells were cultured for three days (days 9 to 12) without 5-AC and with or without Tet. The methylation levels of all loci were recovered in cells lacking tetracycline treatment (Tet-, 5-AC+) (Figure 9G and 9H). Moreover, tetracycline and scramble shRNA by themselves did not affect AGO4 expression (Figure10) and did not influence DNA methylation levels (Figure 11). This finding means that AGO4 reintroduces DNA methylation to previously demethylated loci showed in Figure 12 (12).

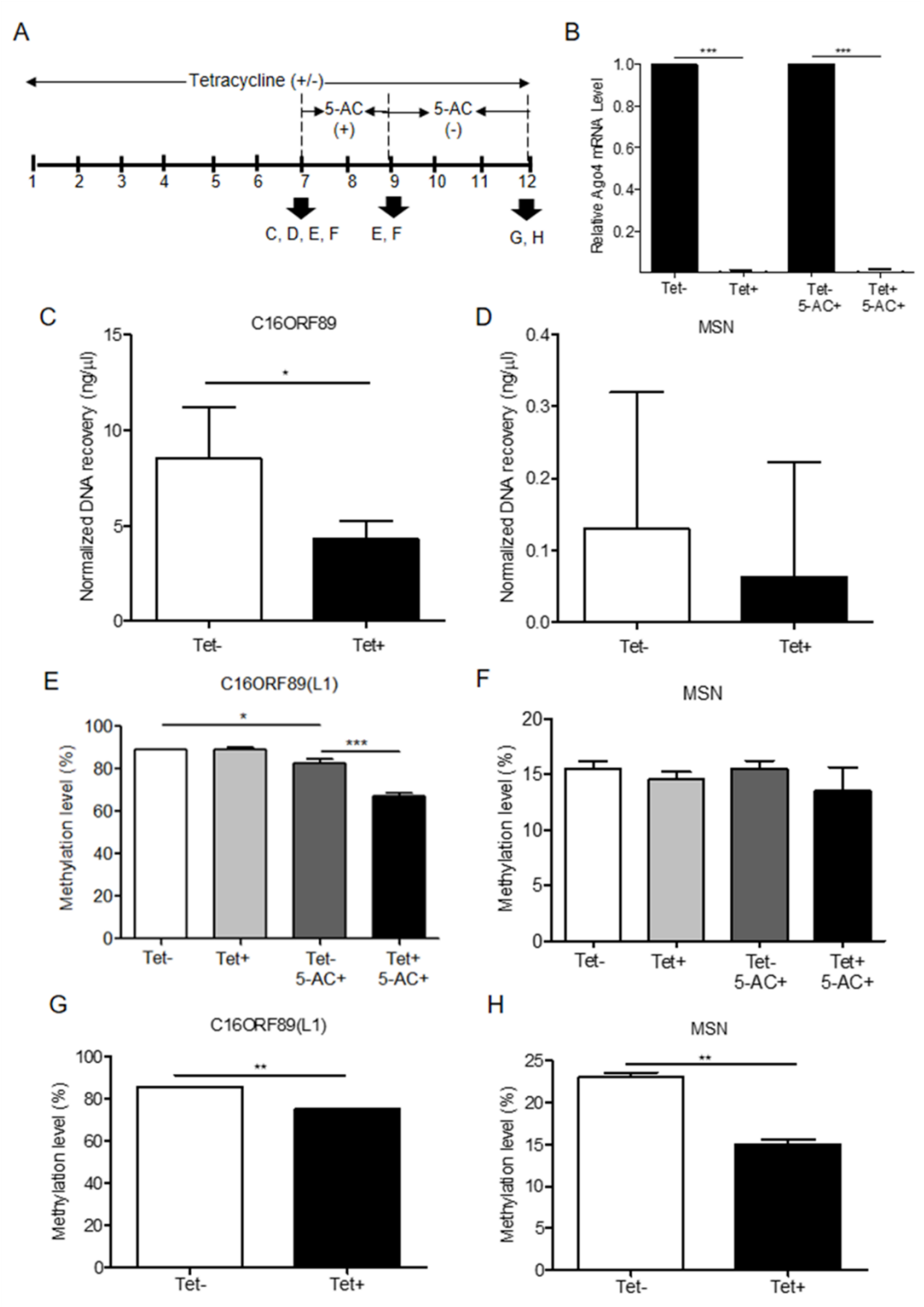


Figure 9 Association of the AGO4-binding genes and methylation levels. (A) Scheme showing that tetracycline-treated cells switch off AGO4 protein expression (Tet+) from

day 1 to 12 and that 5'-azacytidine treatment from day 7 to 9 inhibits DNA methyltransferase (5-AC). (B) Detection of AGO4 mRNA in Tet- (AGO4 expression) and Tet+ (No AGO4 expression) cells; AGO4 was significantly expressed in the absence of Tet. (C) Confirmation of AGO4 binding genes, it showed that AGO4 only bound to the selected gene from bioinformatics data as exemplified by *C16ORF89.* (D) However, AGO4 did not bind to a gene lacking the AGO4-binding site, *MSN*. (E) The AGO4 methylation level of the gene, which containing AGO4-binding sites, was increased due to the presence of the AGO4 protein (Tet-). This increased methylation level was not associated with the presence of DNA methyltransferase (5-AC+). (F) These results were not observed for a gene lacking AGO4-binding sites, *MSN.* (G to H) Recovery of DNA methylation levels is found in cells expressing the AGO4 protein (Tet-) with 5'-azacytidine withdrawal in both of AGO4-binding gene, *C16ORF89* and non AGO4-binding gene, *MSN*; AGO4 is inferred to methylate previously demethylated loci. The above data are the representative dataset from three independent experiments presented as the mean±SEM. Statistical analyses were performed using a paired-sample t-test, where $p<0.05$, $p<0.01$ and $p<0.005$ are represented as *, ** and ***, respectively where Tet+ means tetracycline treatment, 5-AC+ means azacytidine treatment and Tet- and 5-AC- mean untreated groups (12).

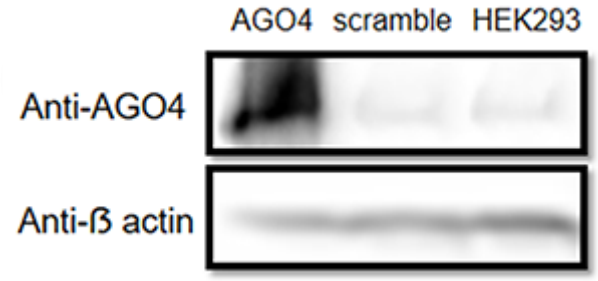


Figure 10 AGO4 protein expression in AGO4 plasmid and scramble shRNA transfected- and untransfected HEK293 cells at 48 hours after transfection. AGO4 protein was upregulated in AGO4 plasmid transfected HEK293 while scramble shRNA transfected HEK293 showed low level of AGO4 protein as same as in untransfected HEK293 cells. β-actin was used to confirm equal protein loading of each lane.

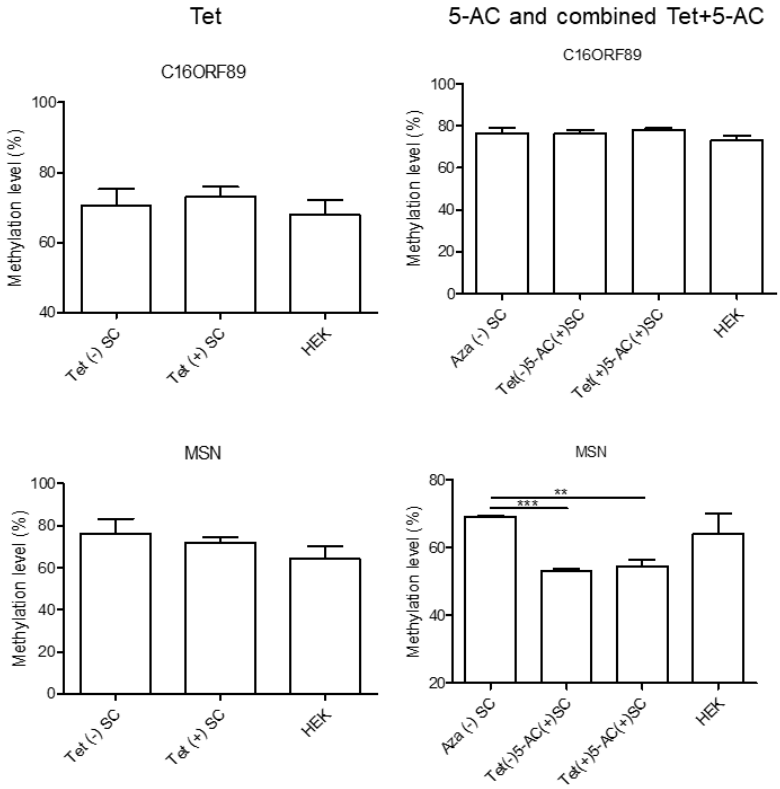


Figure 11 Tetracycline (Tet) (the first panel), and 5'-azacytidine and combined tetracycline and 5'-azacytidine treatment (the second panel) in scramble shRNA transfected HEK293 cells. Methylation levels were observed in AGO4-binding genes: *C16ORF89* and non AGO4-binding gene, *MSN*. Statistical test using an unpaired T-test was shown where $p<0.05$ is represented as **, and SC means scramble shRNA transfected cells (12).

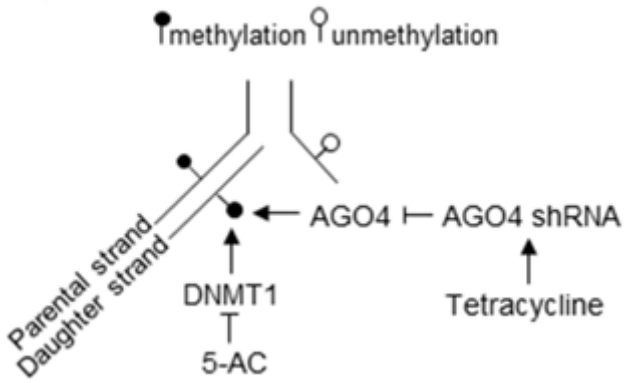


Figure 12 Schematic overview of the role of AGO4 in *de novo* methylation. After DNA replication, DNA methylation is maintained by DNMT1 by the recognition of hemimethylated CpG on the parental strand as a template and the addition of newly methylated CpG to the daughter strand. DNMT1 is inhibited by the action of 5-AC, and the methylation level should be decreased by half under 5-AC treatment. In our study, when tetracycline was added, AGO4 shRNA was manipulated, resulting in AGO4 expression repression. Thus, without tetracycline, AGO4 is upregulated and binds to DNA loci, and DNA methylation is maintained although 5-AC is added, suggesting that AGO4 involved in *de novo* methylation.

Transfection of Alu siRNA increases Alu methylation when AGO4 is presented

Previously, we demonstrated that Alu element siRNA (Alu siRNA) transfection increased Alu methylation (43). To determine whether siRNA can promote methylation in human cells and whether this process is dependent on AGO4, Alu siRNA was used to transfect the AGO4sh-induced HEK293 cell line, and the cells were cultured under Tet+ or Tet- conditions. We found that Alu siRNA increased the Alu methylation level when AGO4 was not limited; on the other hand, Alu siRNA did not promote methylation when AGO4 was depleted in Tet+ cultured HEK293 cells (Figure 13). Therefore, Alu siRNA may form an RdDM complex with AGO4 to methylate Alu sequences.

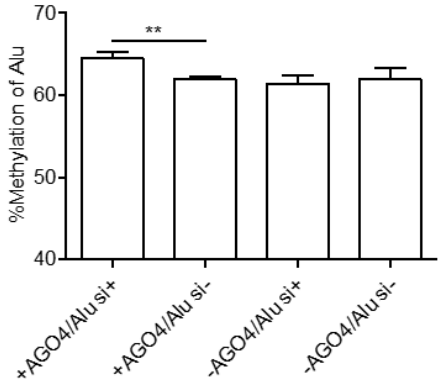


Figure 13 Alu methylation upon Alu siRNA transfection in Tet-controlled AGO4-expressing cells. The Alu methylation level was significantly increased when AGO4 was upregulated under Alu siRNA transfection. The above data are the representative dataset from six independent experiments presented as the mean±SEM. Statistical analysis was performed using an unpaired t-test where $p<0.01$ is represented as **.

Human Argonaute Proteins

One of the effector proteins involved in TGS or PTGS is AGO proteins. AGO can be divided into two subclasses according to sequence homology. The first group is AGO subfamily which falls into *Arabidopsis thaliana* AGO1 homology. Another is PIWI subfamily, it is closely related to *Drosophila melanogaster* (44). AGO protein found to be conserved through evolution from prokaryotes to eukaryotes, it composes of 3 structural domains which are PAZ domain, MID-domain and PIWI-domain (Figure 14).

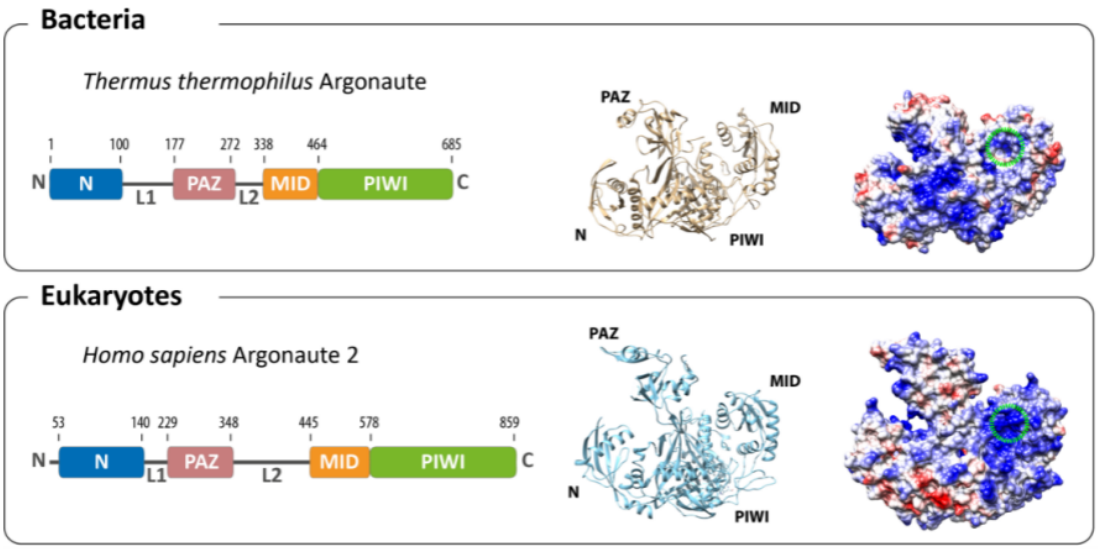


Figure 14 Argonaute structure in prokaryotes and eukaryotes (modified from (45)).

1. **PAZ domain:** Function to bind with 3' of the small RNA.
2. **MID domain:** It is a conserved region which binds to 5' of the small RNA.
3. **PIWI domain:** Contain RNase H-like active sites for RNA cleavage (13).

In human, sequences analysis was performed to identify human AGO proteins. The result revealed that human contains eight putative genes which are 1). AGO subfamily: AGO1, AGO2, AGO3 and AGO4. and 2). PIWI subfamily: PIWIL1, PIWIL2, PIWIL3 and PIWIL4. As previously described, AGO proteins play roles in both of TGS and PTGS by binding to siRNA or miRNA resulting in gene expression repression, RNA degradation, including translational inhibition (46). There are less studies for each AGO proteins' function, the most extensive study is human AGO2 related small RNA-mediated TGS.

For human AGO4, it differs by a few amino acid sequences from the N-terminal and PIWI-domains of AGO2, which has been well described with respect to its structure and function. AGO4 does not have catalytic cleavage activity (47), and the quantity of AGO4 and its expression in both mRNA and protein are at the lowest levels compared to its protein family. Moreover, this finding also found that AGO4 expression inhibition is

mainly found in nucleus at the level of mRNA synthesis (48, 49). Due to these limits, we aimed to overexpress AGO4 using plasmid transfection and cell-penetrating peptide (CPP)-tagged AGO4 protein to observe the action of AGO4 in the RdDM pathway in this study.

Cell-penetrating Peptides (CPPs)

At present, drug delivery has been widely studied in part of targeted therapy for human diseases. However, the most important obstacle is that those drugs which can be oligonucleotides, active peptides or proteins cannot translocate into cells due to membrane barrier which has a selectively semi-permeable hydrophobic property. The discovery of CPP provides many benefits to deliver those cargoes into cells without membrane damage or causing cytotoxicity in cells, but giving high efficiency in cellular uptake and specificity to targets (50) when appropriate CPP is selected to use.

CPPs or also known as protein transduction domains (PTDs) is defined as short peptides that facilitate biologically active cargoes to across cellular membrane due to an attachment of CPP to cargo molecules through covalent conjugation or electrostatic interaction (51) (Figure 15). CPPs are categorized into many classes according to their individua properties, including peptide sequences and binding properties as described below (14).

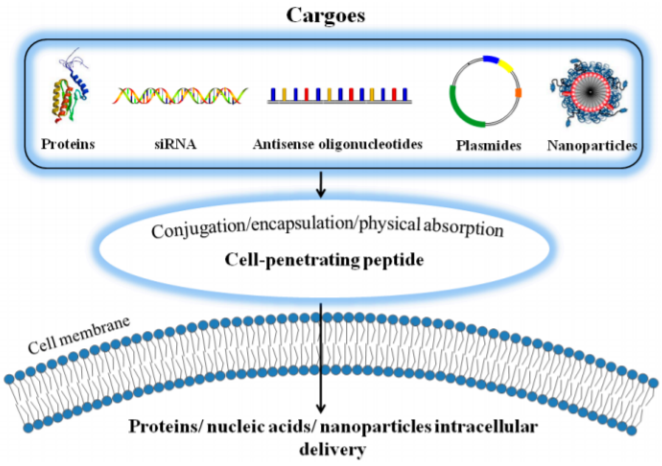


Figure 15 Interaction of CPP and biologically active cargo molecules (52).

1. **Primary amphipathic CPPs**: They mostly contain more than 20 amino acids which can be either positively or negatively charge such as MPG or TP10.They can translocate into cells by direct penetration through pore formation or inverted micelles formed to the membrane.

2. **Secondary amphipathic CPPs**: They typically contain less than 20 amino acids, they can be α-helical, β-sheet or proline-rich CPPs such as M918 or penetratin. Their amphipathic property is revealed when they form interaction with a phospholipid membrane resulting in an increase of translocation efficiency.

3. **Non-amphipathic CPPs**: They mostly are cationic amino acid sequences such as oligoarginines (R9) and TAT and bind to anionic lipid membrane.

Due to advantages of CPPs using in therapeutic areas, more CPPs are discovered and modified to potentially enhance the capability in systemic drug delivery utilization. An established CPP database showed many types of CPP being used (Figure 16), owning to this, CPP should be selected for utilization with cautions. One factor should be considered is an uptake mechanism.

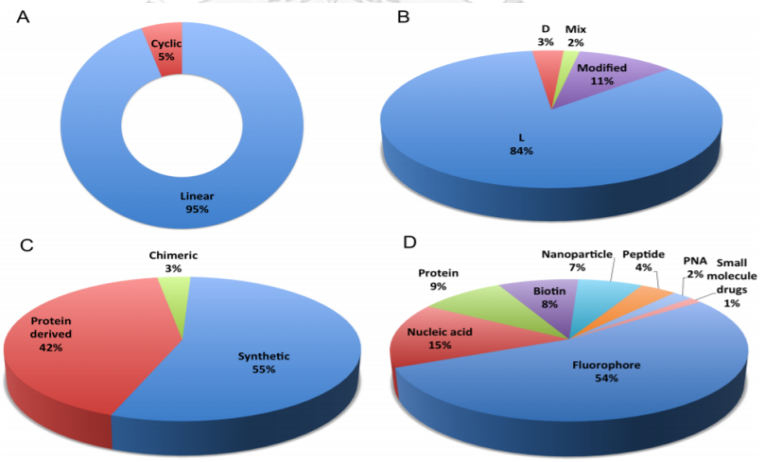


Figure 16 Available CPPs found in CPP database in 2002 based on **(A)** conformation **(B)** modification **(C)** origin and **(D)** delivery types (53).

CPP uptake mechanisms are not clarified so far. However, based on their properties, the internalization mechanisms can be categorized into 2 groups (Figure 17) (54).

1. **Energy-independent penetration**: Direct penetration (membrane transduction) falls into this groups. It consists of pore formation, the carpet-like model and an inverted micelle formation. This pathway occurs owing to the interaction of positively charged CPPs and negatively charged of cellular membrane contributing to membrane destabilization.

2. **Energy-dependent penetration**: Endocytosis falls to this group; however, it is involved micropinocytosis, and receptor-mediated endocytosis. After endocytosis occurs, cargoes must escape from endosomes to cytosol and reach the targets.

Which mechanism the CPPs is used depending on pH and temperature in experiments, cell type, nature of CPP and CPP concentration (50).

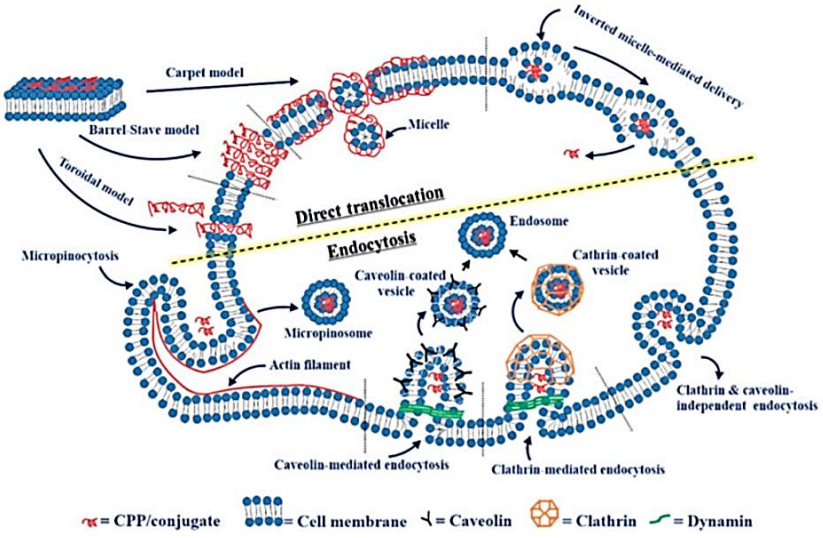


Figure 17 Mechanisms of CPP translocation (55).

However, we selected 9x-oligoarginines and xentry as CPPs in our study.

1. **Oligoarginines or arginine-rich CPP**: It is derived from synthesis of homopolyarginine, here 9 oligoarginines (R9) are used. The oligoarginines translocate through cellular membrane via a direct penetration; however, some studies showed the endocytosis depending on conjugated cargo or even the two uptake processes may share together (56-58). In the study of using oligoarginines, the result also showed that it

has more efficiency to enter cells compared to other positively charged amino acid, lysine according to its guanidium head group. Moreover, increase of using arginine number also develop the translocation efficiency (59, 60). The R9 is applied in many studies with good achievements with lower cytotoxicities, broad cell specificities and high uptake efficieincies (61-63), the R9 was selected in this study.

2. **Xentry**: It is seven amino acid peptides (LCLRPVG) derived from X-protein of hepatitis B virus. It penetrates into cells by the binding of heparan sulphate moieties on syndecan-4 which belong to clathrin pathway and can be applied to use in many cell types (19, 64). Even it is just a new class of CPP which was found, owing to its short sequences, it may offer good results when conjugate it with large cargo molecules like AGO4. Xentry is selected for this study.

CHAPTER III

MATERIALS AND METHODS

Cells and culture

Human cervical adenocarcinoma HeLa, human cervical retinoblastoma C33a, proximal tubule epithelial HK2, and embryonic kidney HEK293 cell lines were purchased from American Type Culture Collection (ATCC, VA, USA). WSU-HN31 cell line was kindly provided by Dr. Silvio Gutkind (National Institutes of Health/NIIDCR, MD, USA). They were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Thermo Fisher Scientific, MA, USA) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific) and 1% antibiotic-antimycotic (Thermo Fisher Scientific) in a 5% humidified incubator at 37°C.

Argonaute-4 plasmid transfection and overexpression

pIRESneo-FLAG/HA-AGO4 (AGO4) and pcDNA3.1/HIS A (PC) plasmids (Addgene, MA, USA) which was used as a control group were transformed into DH5α competent cells (Promega, WI, USA). Both AGO4 and PC-transformed cells was cultured in 100 μg/ml ampicillin (Merck Millipore, NJ, USA) containing Luria-Bertani (LB) broth for plasmid extraction using the plasmid DNA miniprep kits (Thermo Fisher Scientific) according to the manufacturer's protocols. Sequence fidelity was confirmed by Sanger sequencing (data not shown).

One day before transfection, HeLa cells were seeded at 2 x, 10^5 cells/ml in 6-well plate in 2-ml DMEM. The next day, plasmid transfection complexes were prepared; 2 μg of AGO4 and PC plasmid was transfected into HeLa using 10 μl of Lipofectamine®2000 (Thermo Fisher Scientific) mixed with Opti-MEM® reduced serum media (Thermo Fisher Scientific) in a total volume of 500 μl according to the manufacturer's protocols. The transfection duration was 72 h for chromatin immunoprecipitation (ChIP) and the observation and quantification of LINE-1 or Alu methylation; overexpression of AGO4 mRNA and protein was confirmed.

Protein preparation and Western blot assay

AGO4- and PC- overexpressing HeLa cells were collected as whole cell lysate using RIPA cell lysis buffer (Amresco, OH, USA). Cells were further sonicated at a 30% amplitude for 5 s, which was repeated three times. The protein concentration was determined by using a BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instruction. Twenty-five μg of whole cell lysate was subsequently used for 6% SDS-PAGE and Western blot analysis as previously described (65). The antibodies used in these assay were 1:1000 eLF2C4 (C-12) (SC-32655) (Santa Cruz, CA, USA), 1:5000 goat anti-mouse IgG HRP (SC-2005) (Santa Cruz), and anti- beta-actin antibody [AC-15] (HRP) (Ab49900) (Abcam, Cambridge, UK). Each antibody was diluted in the Bløk® Noise Cancelling Reagent (Merck Millipore). Finally, the signal was visualized using the SuperSignal® west femto chemiluminescent substrate (Thermo Fisher Scientific) on a C-Digit® blot scanner (LI-COR Biosciences, NE, USA).

Chromatin immunoprecipitation (ChIP)

To detect AGO4 protein-binding genes, ChIP was performed as previously described with some modifications (66). Briefly, cells were fixed with 1% final concentration of formaldehyde for 30 min at room temperature, and then 0.125 M final concentration of glycine was added for 5 min to stop the crosslink between DNA-protein reaction. After that cells were washed twice with 1X protease inhibitor (Thermo Fisher Scientific) containing cold PBS and collected using scrapers. The cells were sonicated at a 30% amplitude on 15 s and off 30 s, repeated for eight times. The supernatant was then collected by centrifugation and diluted with ChIP dilution buffer. Protein A or G plus-agarose (Santa Cruz) was added to the samples with agitation for 1 h at 4°C, and the supernatant was harvested. Following this, immunoprecipitation was performed using 10 μg of each antibody: HA-probe (Y-11) (SC-805) (Santa Cruz) and normal rabbit IgG (#2729) (Cell Signaling, MA, USA). Protein A or G-agarose was added again to capture the antigen-antibody complexes. The precipitated complexes were then washed with

150 and 500 mM sodium chloride and lithium chloride buffer. After that the complexes were decrosslinked, and the DNA was precipitated using ethanol and purified by phenol-chloroform DNA extraction as previously described. Finally, the DNA was used for further experiments consisting of real-time PCR and conventional PCR.

RNA preparation and cDNA synthesis

Cells were collected using trypsinization and washed with phosphate-buffered saline (PBS), Trizol™ (Thermo Fisher Scientific) was then added to the cell pellet for RNA extraction. 500-100 ng of total RNA of each sample was used to synthesize cDNA using the RevertAid™ first-strand cDNA synthesis kit (Thermo Fisher Scientific) according to the manufacturer's specification.

Real-time PCR

To observe AGO4 mRNA expression, for quantification of AGO4-binding LINE-1 or ALU DNA, or to investigate whether AGO4-protein-conjugated single guide RNA (sgRNA) has an impact on LINE-1 or *MKI67* expression in five biological replications, real-time PCR was carried out using Power SYBR® Green PCR Master Mix (Applied Biosystems, CA, USA) with AGO4 FW and RW primers (GAPDH FW and RW primers were used as an internal control) or copy LINE-1 FW and RW primers, copy ALU FW and RW primers, LINE-1 exp FW and RW primers, or MKI FW and RW primers (shown in table). The amplifications were performed in a 7500 Fast Real-Time PCR System (Applied Biosystems). The $\triangle\triangle$CT method (67) was used to calculate the amount of AGO4-binding LINE-1 or ALU in AGO4 plasmid-overexpressed cells compared to pcDNA3.1 empty-plasmid-transfected cells. In the case of LINE-1 or *MKI67* expression, fold enrichment of expression was calculated from CPP-AGO4-sgRNA transfected cells versus untransfected cells. Lastly, sample paired T-tests were performed for statistical analysis at confidence intervals of 95%.

DNA preparation and sodium bisulfite treatment

Cells were harvested by trypsinization for DNA extraction using 10% sodium dodecyl sulfate (SDS) (Sigma-Aldrich, MO, USA) containing lysis buffer II (0.75 M NaCl, 0.024 M EDTA at pH 8), and 20 mg/ml proteinase K (USB, OH, USA). Cells were incubated in a 50°C water bath for three nights. Phenol-chloroform extraction and ethanol precipitation were then performed as previously described (68). After this 750 ng of DNA was used to perform sodium bisulfite treatment using the EZ DNA methylation-Gold™ kit (Zymo Research, CA, USA) according to the manufacturer's instructions. The eluted DNA was further applied for combined bisulfite restriction analysis (COBRA) or methylation-specific PCR (MS-PCR).

LINE-1 or Alu-combined Bisulfite Restriction Analysis (COBRA) and methylation analysis

To observe methylation levels of LINE-1 or ALU in transfected cells, the sodium-bisulfite-treated DNA in each sample was amplified by a PCR containing 1X PCR buffer (Qiagen), 0.2 mM deoxynucleotide triphosphate (dNTP) (Biotechrabbit, Hennigsdorf, Germany), 1 mM magnesium chloride (Qiagen), 25 U of HotStarTaq DNA Polymerase (Qiagen) and 0.3 μM LINE-1 met FW and RW primers or Alu met FW and RW primers, as shown in table1. For L1 amplification, the cycling program was set at 95°C for 15 minutes, followed by 30 cycles of 95°C for 45 seconds, 55°C for 45 seconds, 72°C for 45 seconds, with a final extension at 72°C for 7 minutes. LINE-1 PCR products were then put through COBRA by using 2U of TaqI (Thermo Scientific) with 0.5X TaqI buffer (Thermo Scientific) and incubating at 65°C overnight. For ALU amplification, the program was set as follows: 95°C 15 minutes for one cycle, followed by 40 cycles of 95°C for 45 seconds, 57°C for 45 seconds, 72°C for 45 seconds, with a final extension at 72°C for 7 minutes. ALU PCR products were subjected to COBRA using 2U of TaqI (Thermo Scientific) and 2U of TasI (Thermo Scientific) with 5X NEB3 buffer (New England Biolabs, MA, USA) and 1 μg/μl bovine serum albumin (BSA) (New England

Biolabs) and incubated at 65°C overnight. Then, cut PCR products were analysed by 8% acrylamide and SYBR (Lonza, Basel, Switzerland) gel stain. The band intensity of LINE-1 or ALU methylation was observed and measured by Strom840 and ImageQuanNT Software (Amersham Biosciences, Little Chalfont, UK).

To analyze LINE-1 methylation level, the band intensities of each LINE-1 products sizes of 92, 60, 50, 42 and 32 base pair (bp) were applied for the calculation followed the formula (23): [(A+2C+F)x100]/(2A+2B+2C+2F) where A = band intensity at 92 bp/92, B = 60/56, C = 50/48, D = 42/40, E = 32/28, and F = [(D+E)-(B+C)]/2. For Alu methylation level analysis, the band intensities of each Alu products sizes of 133, 90, 75, 58, and 43 bp were used for calculation following this formula (23): [(2F+D+C)x100]/(2A+2B+2D+2F), where A = band intensity at 133 bp/133, B = 58/58, C = 75/75, D = 90/90, E = 43/43, F = [(E+B)-(C+D)]/2. The global methylation of LINE-1 or Alu measured in each replication was statistically tested using a paired sample *T*-test.

Methylation-specific PCR (MSP)

Seven hundred and fifty nanograms of sodium-bisulfite-treated DNA was used in this experiment to observe levels of methylated and unmethylated *EML2* and *CCNA1*. Methylated and unmethylated primers (Table 1), both at concentrations of 0.3 μM, were added to the PCR master mix. The program was set as follows: 95°C for 15 minutes for one cycle, followed by 35 cycles of 95°C for 1 minute, 56°C (for *EML2*; or 59°C for *CCNA1*) for 1 minute, and 72°C for 1 minute, and with a final cycle at 72°C for 7 minutes. PCR products were visualized on an 8% acrylamide gel stained by SYBR (Lonza). An EpiTect® PCR Control DNA Set (Qiagen) was used to represent actual methylated and unmethylated band positions. Methylated and unmethylated band intensities were measured using Storm 840 and ImageQuaNT software (Amersham Biosciences). Changes in *EML2* sequences due to CPP-AGO4-EML2 were also confirmed by Sanger sequencing.

Table 1 Oligonucleotide sequences and conditions for PCR and qPCR analyses

Primer	Sequences (5'-3')	Annealing temperature (°C)
LINE-1 met FW	GTTAAAGAAAGGGGTGAYGGT	55°C
LINE-1 met RW	AATACRCCRTTTCTTAAACCRATCTA	
ALU met FW	GGYGYGGTGGTTTAYGTTTGTAA	57°C
ALU met RW	CTAACTTTTTATATTTTTAATAAAAACRAAATTTCACCA	
AGO4-FW	CAGGAATTCAGGGAACCAGCCG	56°C
AGO4-RW	CTGCCTTCCGCACTGTCATGATC	
GAPDH-FW	TGGAAGGACTCATGACCACAG	56°C
GAPDH-RW	TTCAGCTCAGGGATGACCTT	
Copy LINE-1 FW	CTCCCAGCGTGAGCG	56°C
Copy LINE-1 RW	ACTCCCTAGTGAGATGAACCCG	
LINE-1 exp FW	GGCCAGTGTGTGTGCGCACCG	60°C
LINE-1 exp RW	CCAGGTGTGGGATATAGTCTCGTGG	
MKI FW	CCACACTGTGTCGTCGTTTG	60°C
MKI RW	CCGTGCGCTCATCCATTC	
EML2 met FW	CGGTTTCGAAGTTTTGTTATTCGTC	56°C
EML2 met RW	CGTATCCCCACGCCGA	
EML2 unmet FW	CCAAATTGGTTTTGAAGTTTTGTTATTTGTT	56°C
EML2 unmet RW	TTATTAATCCCATATCCCCACACCAA	
EML2 seq FW	TCGTCGTTATGAGTAGTTTTGG	57°C
EML2 seq RW	CACGCCGATCTAAATCCC	

Primer	Sequences (5'-3')	Annealing temperature (°C)
CCNA1 met FW	TTTCGAGGATTTCGCGTCGT	59°C
CCNA1 met RW	CTCCTAAAAACCCTAACTCGA	
CCNA1 exp FW	GCCTGGCAAACTATACTGTGAAC	60°c
CCNA1 exp RW	GTGCAGAAGCCTATGACGATTA	
CCNA1 unmet FW	TTAGTGTGGGTAGGGTGTT	59°C
CCNA1 unmet RW	CCCTAACTCAAAAAAACAACA	
CCNA1 unmet FW	TTAGTGTGGGTAGGGTGTT	59°C
CCNA1 unmet RW	CCCTAACTCAAAAAAACAACA	

Cell-penetrating peptide-tagged AGO4 plasmid construction and protein production

1. Oligoarginines-tagged AGO4 (R-AGO4) and EGFP (R-EGFP) plasmids

To produce R-AGO4 and R-EGFP recombinant proteins, 9X arginines (R9) as cell-penetrating peptide (CPP), and AGO4 mRNA (NM_017629.3) or EGFP mRNA (AB909457.1) were in-frame inserted into the pRSET A vector (Thermo Scientific) between *Bam*HI and *Eco*RI sites. Both vectors were constructed by the GeneArt™ Plasmid Construction Service (Thermo Scientific). The fidelities of the sequences were confirmed by Sanger sequencing.

2. Xentry-tagged AGO4 (Xentry-AGO4) and EGFP (Xentry-EGFP) plasmids

To produce Xentry-AGO4 and Xentry-EGFP recombinant proteins, Xentry sequences (5'-ATGCTGTGCCTGCGTCCGGTGGGC-3') was added to 5' of AGO4 mRNA (NM_017629.3) or or EGFP mRNA (AB909457.1) were in-frame inserted into the pRSET A vector (Thermo Scientific) between *Bam*HI and *Eco*RI sites. Both vectors were constructed by the GeneArt™ Plasmid Construction Service (Thermo Scientific). The

fidelities of the sequences were confirmed by Sanger sequencing. All vector constructions were shown in figure 18.

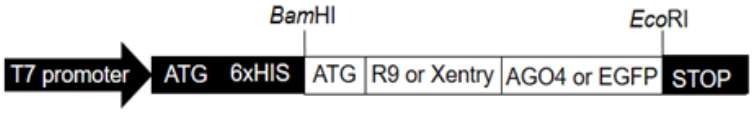


Figure 18 R9- or Xentry-tagged AGO4 or EGFP plasmid construction.

Each vector was transformed into BL21(DE3)pLysS competent cells (Promega, WI, USA) for protein production. The transformed cells were then cultured in 2YT medium, and overexpression was induced by adding 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) (Sigma-Aldrich) until the OD600 reached 0.8. The cells were cultured further for 20 h at 20°C in a 250-rpm incubator shaker. Then, the cells were collected by centrifugation at 10,000 rpm for 15 minutes at 10°C. Cell pastes were lysed by lysis buffer [50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole, 10% glycerol, 10 mg/ml lysozyme, 1% Triton X-100, and 1X protease inhibitor], and sonicated and centrifuged at 10,000 rpm at 10°C for 20 min to harvest the crude protein supernatant. R9- or X-AGO4 and R9- or X-EGFP, which was soluble in supernatant, was purified on a HisTrap™ HP (GE Healthcare) column according to the manufacturer's instructions. The purified R9- or X-AGO4 and R9- or X-EGFP proteins were confirmed using Coomassie®-Brilliant-Blue-R-250 (C.I.42660) (Merck Millipore)-stained 6% SDS-PAGE for AGO4 and 12% SDS-PAGE for EGFP proteins. Furthermore, Western blot analysis was performed as indicated above using 1:2,500 Anti-6X HIS tag® antibody [HIS.H8] (Ab18184) (Abcam) and 1:5,000 Goat anti-mouse IgG-HRP) (SC-2005) (Santa Cruz). After that the fraction that contains purified proteins was taken to perform buffer exchange by using Vivaspin® protein concentrator spin columns (GE Healthcare) to collect the purified proteins in PBS. Concentrations of both purified proteins were determined by Pierce BCA protein assay kit (Thermo Fisher Scientific).

Cell-penetrating peptide AGO4-conjugated single guided RNA (sgRNA) transfection

Purified R9- or X-AGO4 was conjugated to sgRNA of LINE-1, ALU, *EML2* or *CCNA1* (Table 2) using a slicing assay as previously describe (69), with some modifications. Briefly, the desired amount (or other protein amount in dose-dependent experiment) of the R9- or X-AGO4 protein was mixed with 200 nM of each sgRNA in conjugation buffer [1.5 mM $MgCl_2$, 5% glycerol, 50 mM KCl and 20 mM Tris-HCl (pH 7.0)] and incubated in a 37°C water bath for 30 min. Then, the conjugated complexes of R9- or X-AGO4-sgRNA were dropped into 500 µl DMEM-cultured $7.5x10^4$ cells in a 24-well plate to allow the complexes to be internalized into the cells and induce methylation at specific targets homologous to the sgRNA; non-induced R9- or X-AGO4 represented as buffer was used as a control. After transfection at defined time, cells were collected to measure changes of methylation levels of target genes by COBRA or MSP.

For purified R9- or X-AGO4 -EGFP, the protein was transfected to observe whether CPP can facilitate protein entry into cells. One microgram of R9- or X-AGO4 -EGFP was transfected in $7.5x10^4$ cells/well in a 24-well plate containing DMEM and observed immediately under a Zeiss LSM 800 confocal laser scanning microscope (Carl Zeiss Microscope, Oberkochen, Germany). Hoechst 33342 (Cell Signaling), as a cell-permeant nuclear counterstain, at a final concentration of 1 µg/µl, was used as a positive control dye to identify the cell position.

Table 2 Single guided strand RNA using in combination with CPP-AGO4 protein

Single guided-strand RNA	Sequences (5'- 3')
LINE-1 antisense Tas	GGAGUGACCCGAUUUUCCAGGUGCG
LINE-1 antisense Taq	AGCGCAAUAUUCGGGUGGGAGUGAC
ALU-25 antisense	AUCCGCCCGCCUCGGCCUCCCAAAG
ALU-21 antisense	UCCGCCCGCCUCGGCCUCCCA
ALU-18 antisense	UCCGCCCGCCUCGGCCUC
EML2 antisense R.1	CUACGAAAAAAUCUCACCCAACUCC
EML2 antisense R.2	AUCUAAACCCCGAAAACGCGAAAAA
CCNA1 antisense	CCCGAGGGCGCGGCUCGC
CCNA1 sense	GCGAGCCGCGCCCUCGGG

Cell proliferation assay

After transfection, cells were seeded for $4x10^3$ cells/well in 96-well plate containing 100 µl DMEM to assess cell proliferation. A day later, DMEM was removed, MTT reagent was added to every well according to the manufacturer's protocol. Cells was incubated at 37°C for 4 hours, cell proliferation was later determined using Biochrom Asys UVM 340 microplate reader at 562 nm. Experiments were repeated every day, continuously for 4 days in nine technical replications. Line graph was generated to observe cell proliferation trend.

CHAPTER IV

RESULTS

1. Localization of AGO4 induced *de novo* methylation

To investigate the connection of AGO4 localization and DNA methylation in human RdDM process, we firstly observed whether AGO4 localizes to LINE-1 and Alu. We transfected a hemagglutinin (HA)-tagged AGO4 plasmid into HeLa cells at various time point to find the appropriated transfection duration. We also used an empty plasmid (pcDNA, or PC more briefly) as a control plasmid. Overexpression of AGO4 after transfection was detected at 72 h at the mRNA level and at 24, 36, 48 and 72 h at the protein level. The results revealed that HA-AGO4 protein reached the highest expression level within 48 h and gradually decreased. While AGO4 mRNA was still upregulated around 100,000 folds compared to PC transfection group at 72 h post-transfection (Figure 19A and 19B). Moreover, the genome-wide interspersed repetitive sequence (IRS) methylation of both LINE-1 and Alu was observed. The results showed that LINE-1 and Alu methylation was significantly increased in AGO4-overexpressing cells (Figure 19C and 19D).

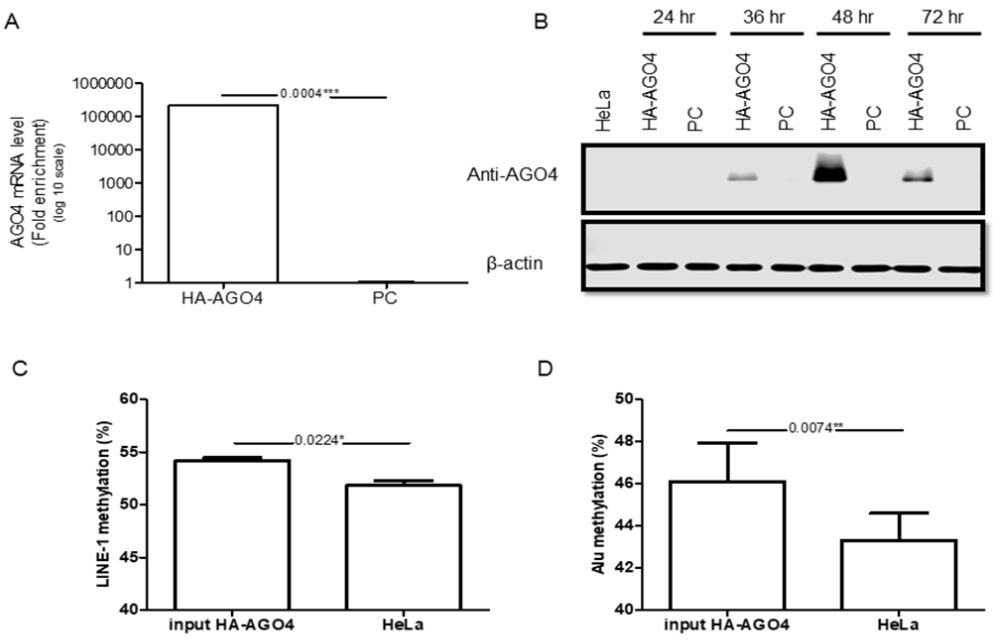


Figure 19 Confirmation of AGO4 expression in HA-AGO4 and PC plasmid overexpressed HeLa. (A) AGO4 mRNA expression was confirmed by real-time PCR. Overexpression of HA-AGO4 resulted in a significant increase in IRS methylation. (B) AGO4 protein expression in time-course of plasmid transfection was detected by Western blot analysis. (C) LINE-1 and (D) Alu in the input (cell lysate) of HA-AGO4 compared with untransfected HeLa cells. The experiment was performed in triplicate, a paired sample *t*-test was used for statistical analysis where $p < 0.05$, $p < 0.01$ and $p < 0.005$ represented as *, **, ***, respectively and error bar showed the standard error (SE).

Then, AGO4 localization to IRS was explored using HA antibody in ChIP experiment; normal rabbit IgG was used as an isotype control antibody to determine the background of non-specific binding of antibody. We discovered that HA-AGO4 preferentially bound to LINE-1 and Alu sequences as the fold enrichment of HA-AGO4 binding to LINE-1 and Alu was higher than its control counterpart (Figure 20). In addition

to IRS methylation, AGO4-overexpressing cells also revealed higher methylation levels compared to PC-overexpressing cells (Figure 21).

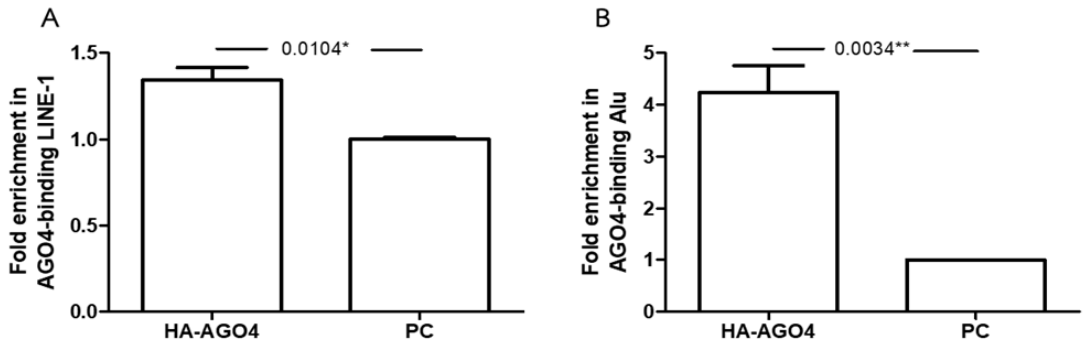


Figure 20 Confirmation of AGO4 binds (A) LINE-1 and (B) Alu by ChIP real-time PCR. Fold enrichment in AGO4-binding IRS with standard error (SE) shown is the average of 5 biological replicates, a paired sample *t*-test was used to detect significant differences at the $p<0.05$ and $p<0.01$ level which is represented as * and **, respectively.

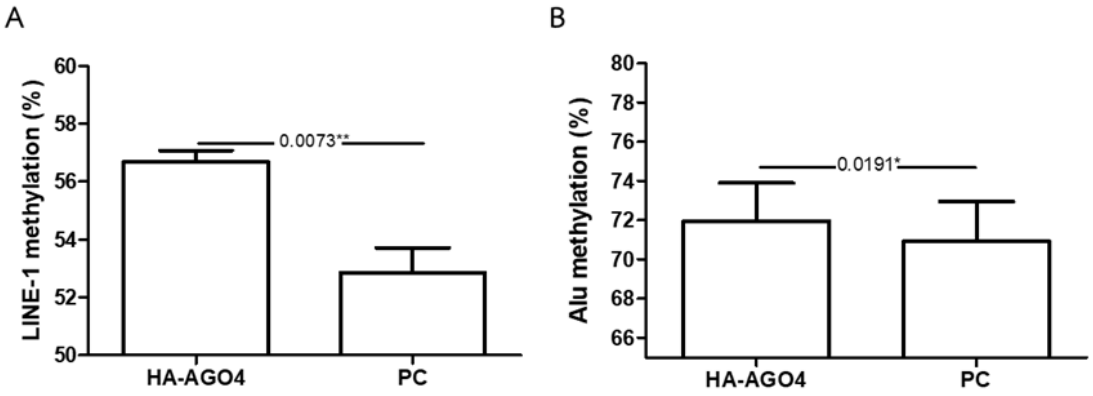


Figure 21 AGO4 localization is involved in RdDM by introducing IRS methylation (A) LINE-1 and (B) Alu in human cells. The experiment was done for 5 biological replicates. A paired sample *t*-test was used for statistical analysis where $p < 0.05$ and $p < 0.01$ and $p < 0.005$ represented as * and **, respectively and error bar showed the standard error (SE).

2. Optimal conditions using in CPP-AGO4-sgRNA transfection

2.1 Recombinant CPP-AGO4 and CPP-EGFP protein production and purification

After we proved that AGO4 localization can induce *de novo* methylation, we engineered CPP portion (R9 or Xentry) which was added at 5' of AGO4 or EGFP mRNA sequences (R9-AGO4, R9-EGFP, Xentry-AGO4 and Xentry-EGFP) for plasmid construction and recombinant AGO4 or EGFP protein production, aimed for transfection to induce specific methylation by incorporation of single guided RNA (sgRNA) into CPP-AGO4 protein. Each plasmid was transformed into *Escherichia coli* strain BL(21)DE3pLysS for those protein productions. Optimal conditions including culture media, IPTG concentration and temperature for bacterial cell culture were investigated to obtain the highest yield of cell paste. We found that the transformed cells were grown when cultured in 2YT medium at 20°C and induced by 1 mM IPTG for protein production (Figure 22A to 22C). Moreover, R9-AGO4 transformed cells which cultured followed optimal conditions above were collected in time-course (Figure 22D).

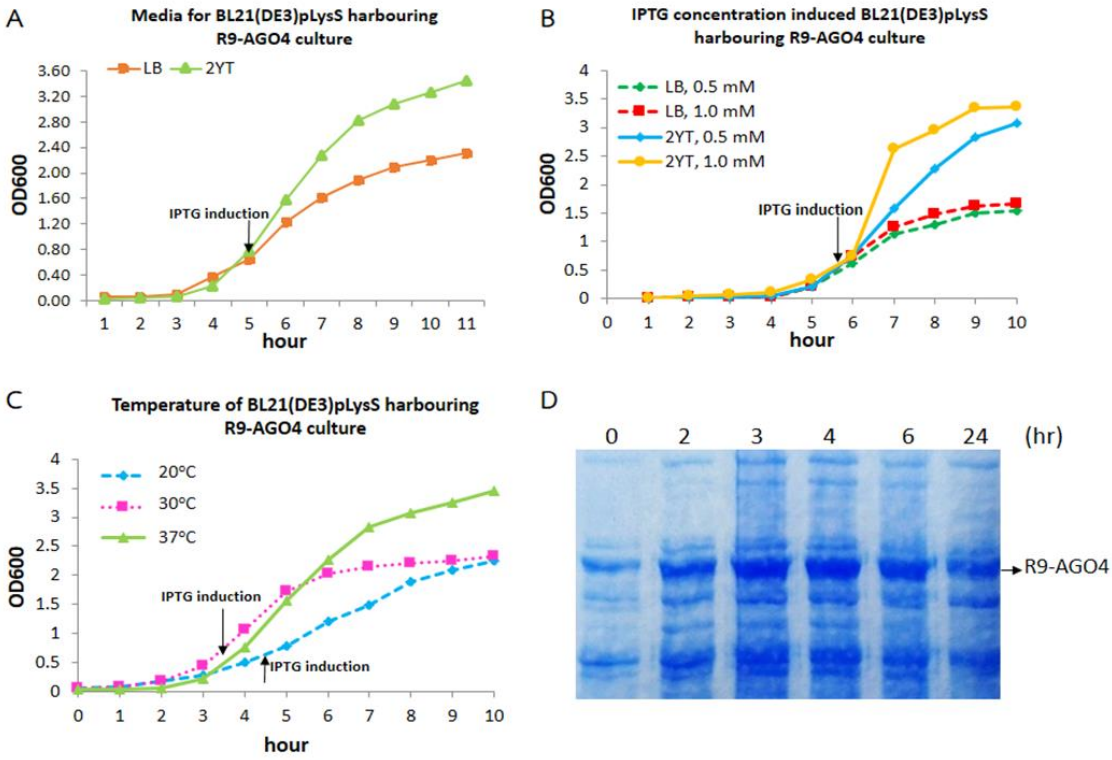


Figure 22 Investigation of optimal conditions for R9-AGO4 transformed BL21(DE3)pLysS cells. (A) culture media between LB and 2YT, (B) IPTG concentration between 0.5 mM and 1.0 mM, (C) temperature between 20°C, 30°C and 37°C. (D) R9-AGO4 protein levels in time-course collection were observed on 6%SDS-PAGE, the R9-AGO4 was shown at about 100 kDa protein and was overexpressed until 24 h culture.

Transformed bacterial cell culture was harvested by ultracentrifugation, supernatant fraction and cell pellet were separately collected and used to explore whether R9-AGO4 or R9-EGFP is soluble. The result showed that both of R9-AGO4 and R9-EGFP was soluble in supernatant (Figure 23A); moreover, soluble EGFP has been reported (70). Hence, supernatant fraction from R9-AGO4 and R9-EGFP, including X-AGO4 and X-EGFP transformed cell cultures can be straight for protein purification. According to N-terminal 6xHIS tagged-pRSET A plasmid, protein purification can be facilitated by using HisTrap™ HP (GE Healthcare) columns. All purified proteins were detected by Coomassie blue R-250 staining 6% and 12% SDS-PAGE for R9- and X-

AGO4, and R- and X-EGFP, respectively (Figure 23B and 23D). The result of AGO4 purification in figure 23B indicated that our R9- and X-AGO4 protein cannot bind well to $Ni2^{+}$-bead trapped inside column, leading to the elution of desired proteins in early fractions. However, the purified R9- and X-AGO4 purified proteins was confirmed by Western blot analysis using HIS antibody (Figure 23D). Thereafter, the fractions that contained purified R9- and X-AGO4 proved by Western blot analysis were applied for human cell line transfection. On the other hands, we did not perform Western blot analysis for R9- and X-EGFP as the fractions that contained those proteins was shown in green color (Figure 23E).

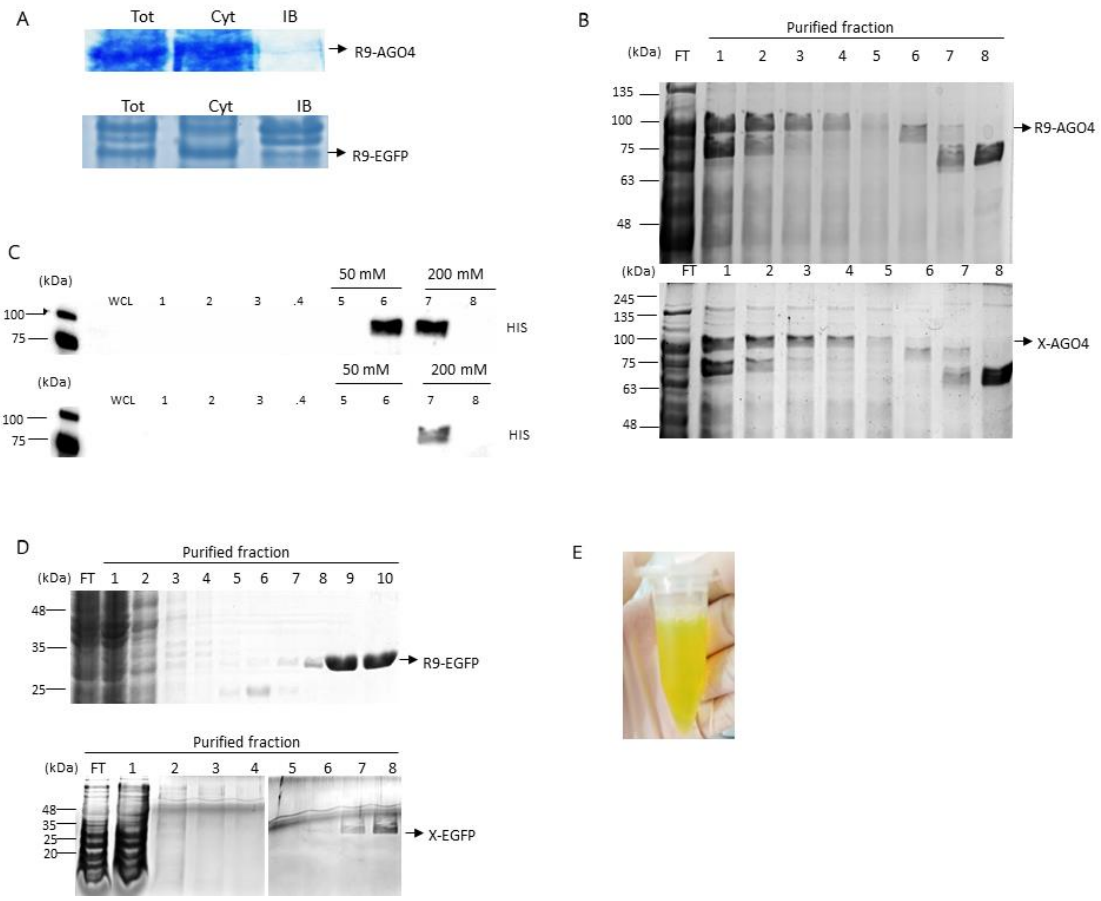


Figure 23 Purification of the R9- and X-AGO4 and R9- and X-EGFP proteins. (A) Detection of R9-AGO4 (upper) and R9-EGFP (lower) solubility indicated that both of proteins which found in cytoplasmic fraction is soluble. Tot, total lysate; Cyt, Cytoplasmic fraction (supernatant); and IB, inclusion body (found in cell paste). (B) The

purified R9-AGO4 (upper) and X-AGO4 (lower) were detected at about 100 kDa on 6% SDS-PAGE. FT, unbound protein; 1 and 2, (20 mM imidazole); 3 and 4, (50 mM imidazole); 5 and 6, (100 mM imidazole); 7 and 8, (200 mM imidazole)-containing washing buffer. (C) The Western blot analysis using 1:2,500 anti-6X HIS antibody showed that purified R9-AGO4 (upper) found in fraction 6 and 7 while purified X-AGO4 (lower) found in fraction 7. Definitions of FT and number 1 to 8 are given above. (D) The purified R9-EGFP (upper) and X-EGFP (lower) were detected at about 30 kDa on 12% SDS PAGE. Definitions of FT and number 1 to 8 are given above, number 9 and 10 represent 250 mM imidazole-containing washing buffer. (F) The elution of both purified R9- and X-EGFP came out as green color.

2.2 Optimal time for transfection determined by R9- and X-EGFP

To determine transfection duration, the R9- and X-EGFP was used to transfect HeLa cells; the penetration of those proteins was captured by a confocal microscope every 3 min until 12 h. The results found that R9-EGFP started to penetrate most cells in a time-dependent manner (Figure 24). Moreover, we expanded the transfection duration to compare transfection efficiency between R9- and Xentry-EGFP, the results found that at 72 h post-transfection X-EGFP can translocate into cells more than using R9-EGFP (Figure 25). We also compared transfection system in Hanks' Balanced Salt solution (HBSS) buffer and DMEM and found that both of R9-EGFP and X-EGFP can translocate into cells in HBSS buffer better than in DMEM (Figure 26). Consequently, we decided to transfect our engineer proteins in HBSS buffer for 24 h after that HBSS was removed, DMEM was added. Cells were cultured for 72 h and finally harvested.

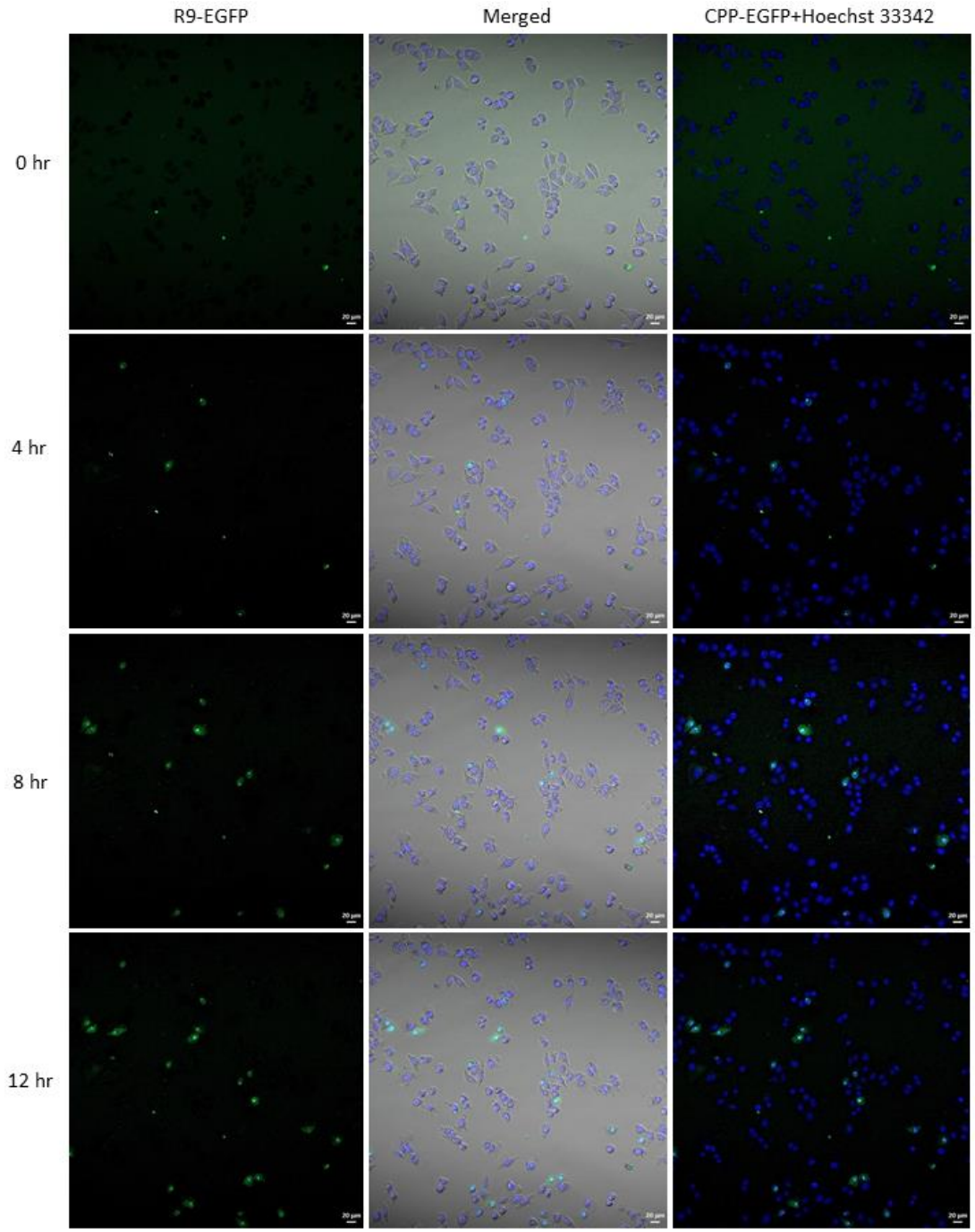


Figure 24 The confocal images showed the translocation of R9-EGFP (green color) in HeLa cells in time-course at 0, 4, 8 and 12 h posttransfection. Hoechst 3332 (blue color) was used for the nuclei labeled. Scale denotes 20 μm.

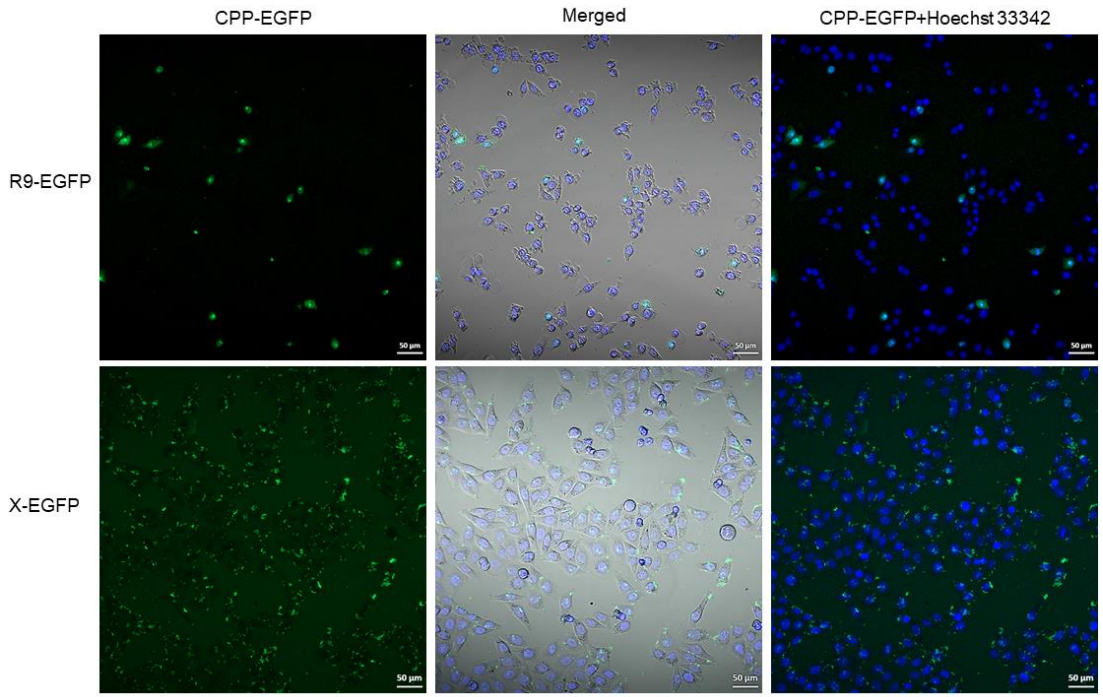


Figure 25 The confocal images showed the translocation of CPP-EGFP (green): R9-EGFP (upper panel) and Xentry-EGFP (lower panel) in HeLa cells after transfection for 72 h. Hoechst 3332 (blue) was used for the nuclei labeled. Scale denotes 50 µm.

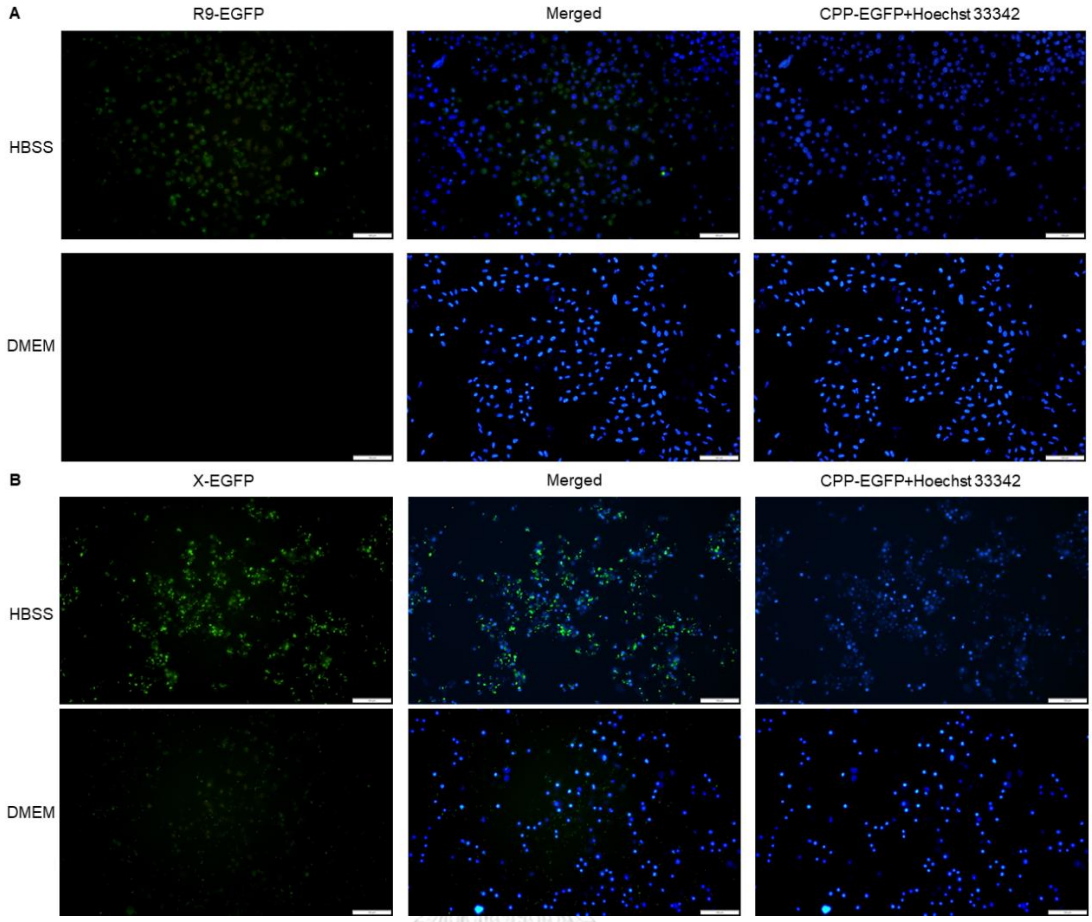


Figure 26 The confocal images showed the translocation of (A) R9-EGFP and (B) X-EGFP in HeLa cells by transfection in HBSS buffer (upper panel) compared with transfection in DMEM (lower panel) for 72 h. Hoechst 3332 (blue color) was used for the nuclei labeled. Scale denotes 100 μm.

2.3 Optimal length of single guided RNA (sgRNA) to induce methylation of target sites

To determine appropriate sgRNA length to be used for incorporation with CPP-AGO4 in inducing methylation of target sequences, sgRNA complementary to Alu sequences at 18, 21 and 25 bp length with or without 5'-phosphate were used for transfection as a previous study reported that 5'-phosphate-siRNA is required to bind at the MID domain of the AGO2 protein. After 72 h of transfection, cells were harvested to observe Alu methylation change. The result showed that 18 bp length sgRNA provided

the highest Alu methylation level among other lengths; however, this was not significant (Figure 27). Moreover, 5'-phosphate was not related to an increase of Alu methylation. We decided to use 25 bp sgRNA without 5'-phosphate for further experiments.

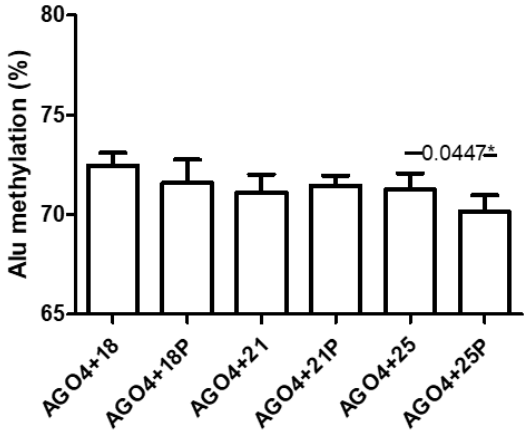


Figure 27 Alu methylation level when R9-AGO4 protein was conjugated with different length of Alu sgRNA, including with or without 5'-phosphate. The experiment was done in 10 replicates and scale bar represented a standard error (SEM).

3. AGO4 as a master protein regulates human RdDM

3.1 IRS (LINE-1 or Alu)

A. CPP-AGO4-sgRNA increased methylation levels of IRS (LINE-1 and Alu)

To validate the role of AGO4 in human RdDM and to invent a new epigenomic-editing tool, we engineered the AGO4 protein for *in vitro* transfection. Due to the AGO protein was reported to automatically form a complex with single guide RNA (sgRNA) (50). We initially tried to induce the methylation of Alu or LINE-1 by using Alu or LINE-1 sgRNAs to incorporate with CPP-AGO4. We discovered R9-AGO4-Alu sgRNA increased Alu methylation significantly, but not LINE-1 methylation (Figure 28A and 28B), indicating the absence of off-target effects. However, X-AGO4-Alu did not provide an augmentation of Alu methylation, surprisingly, it non-specifically induced LINE-1 methylation (Figure 28C and 28D).

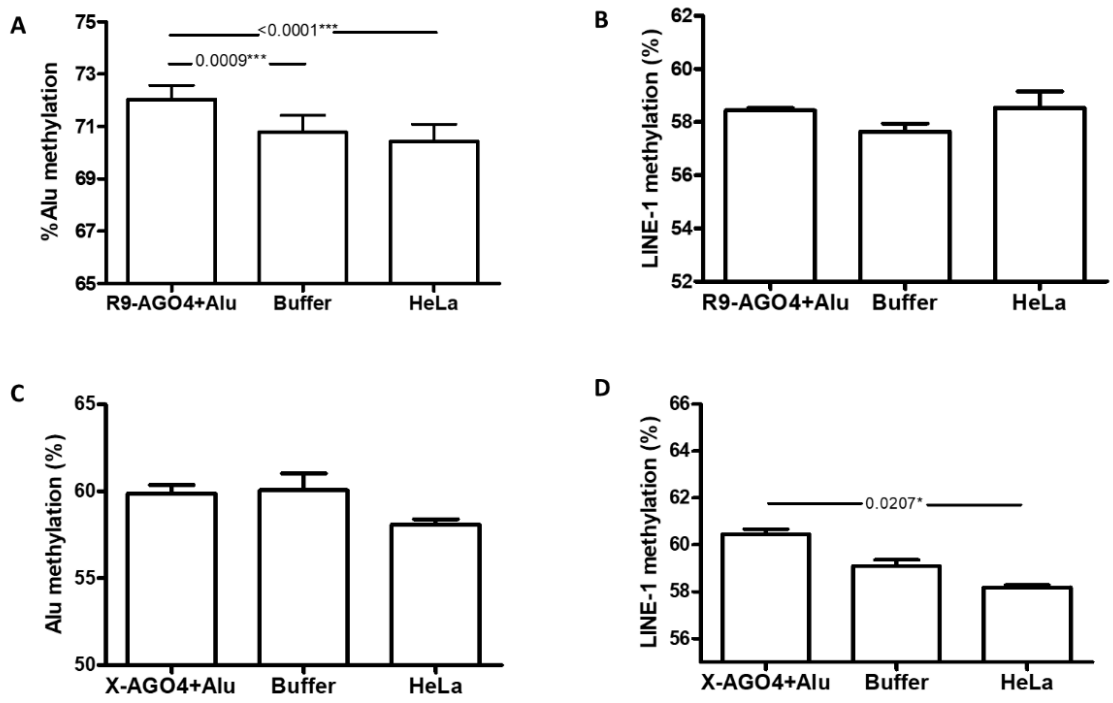


Figure 28 Change in Alu methylation level when using (A) R9-AGO4-Alu transfected HeLa cells, by this transfection, it did not induce LINE-1 methylation (B). (C) X-AGO4-Alu transfection did not affect Alu methylation, but LINE-1 methylation (D). The experiment was performed in 5 replicates, scale bar represented a standard error (SEM).

Due to the results above, we tried to prove why X-AGO4-Alu did not give rise in increasing of Alu methylation, even the transfection efficiency of using xentry portion was better than using R9. As the previous study reported that xentry permeates cells using syndecan-4, we performed Western blot analysis to check syndecan-4 protein expression in various cells. The result revealed that syndecan-4 protein expression found in all cell lines we used with different levels (Figure 29A). However, when we tried to observe whether xentry-tagged protein can translocate into nucleus in HeLa cells using Z-stack confocal microscopy images, we found that less percentage of our protein reside in nucleus (Figure 29B). This might be the reason why methylation levels at targets did not significantly change.

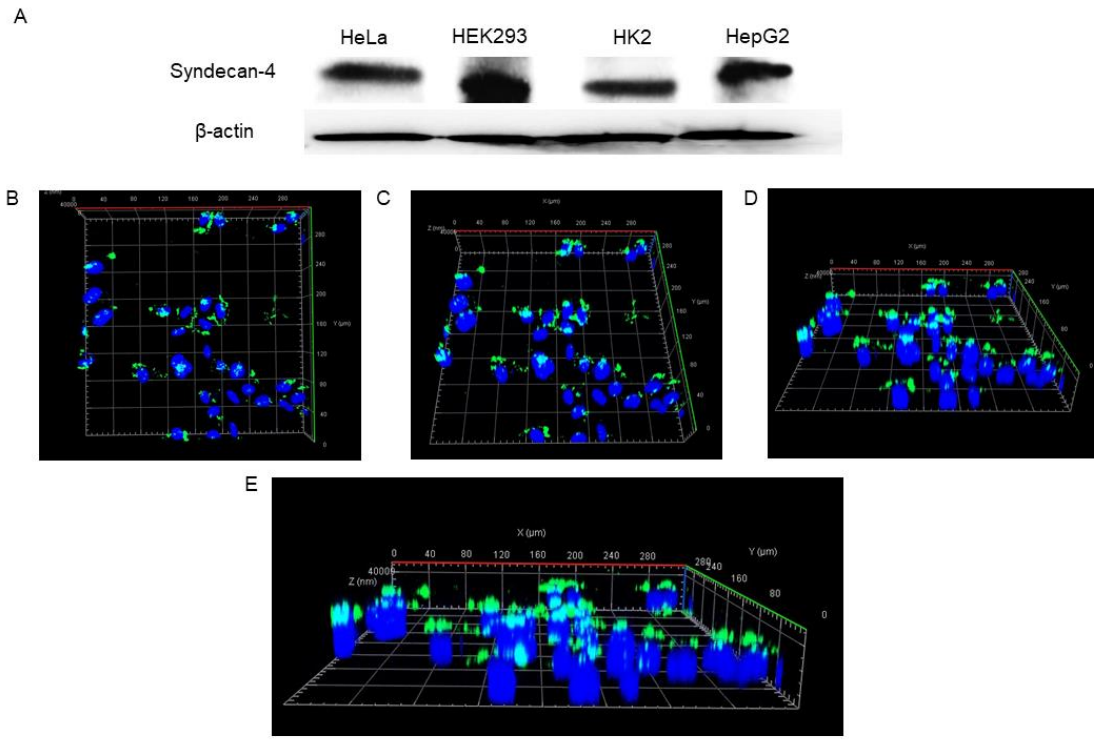


Figure 29 Prove of xentry-tagged protein translocation. (A) Expression of syndecan-4 which is xentry-binding protein was observed using Western blot analysis. β-actin was used as used as a loading control. (B to E) Z-stack images of xentry-EGFP (green channel), Hoechst (blue channel) and xentry-EGFP residing in nucleus represented in the overlays of the related green and blue channel images.

According to xentry-tagged protein permeating cells did not provide impressive results, we were turning to reveal efficiency of using R9-tagged AGO4 with Alu sgRNA by expanding the experiments in more cell lines. Alu methylation levels were significantly increased in all cell lines when R9-AGO4-Alu was transfected, but not in its control counterparts (Figure 30). Moreover, we examined further whether R9-AGO4-conjugated LINE-1 can alter LINE-1 methylation.

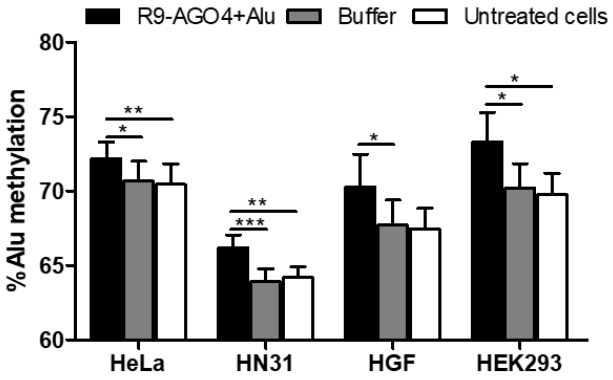


Figure 30 R9-AGO4-Alu transfection increased Alu methylation level in HeLa, HN31, HGF and HEK293 cell lines.

In general, hypomethylation of LINE-1 in cancer cells increases LINE-1 transcription, and LINE-1 RNA is involved in cancer development (25). Therefore, we evaluated if CPP-AGO4-LINE-1 results in these consequences due to LINE-1 methylation. As X-AGO4 did not provide positive results for Alu methylation, we decided to use only R9-AGO4 for LINE-1 experiments. The results showed that R9-AGO4-LINE-1 transfection elevated LINE-1 methylation with no off-target in Alu methylation (Figure 31A and 31B).

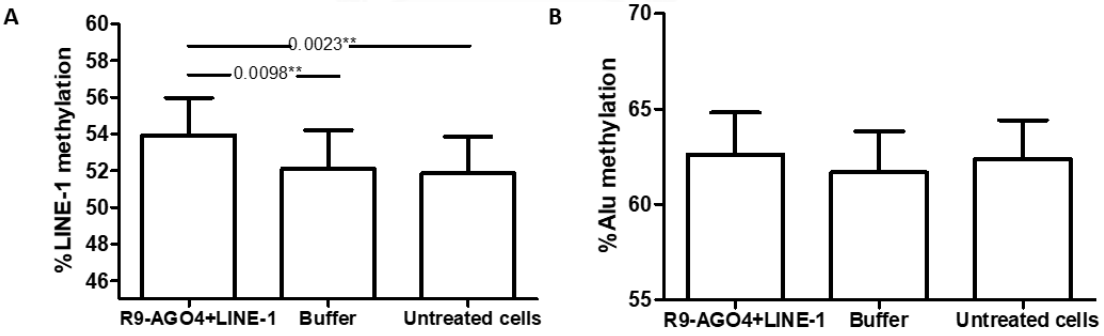


Figure 31 Transfection of R9-AGO4-LINE-1 increased LINE-1 methylation (A) but not in Alu methylation (B). The experiment was performed in 5 replicates, scale bar represented a standard error (SEM).

Moreover, an increase of LINE-1 methylation also diminished LINE-1 RNA levels (Figure 32A). Furthermore, we found that decrease in LINE-1 expression also contributed to the reduction of HeLa cell growth observed by the 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT) assay (Figure 32B); marker of proliferation Ki-67 (*MKI67*) gene expression (Figure 32C) and cell numbers were observed under phase-contrast microscopy (Figure 33).

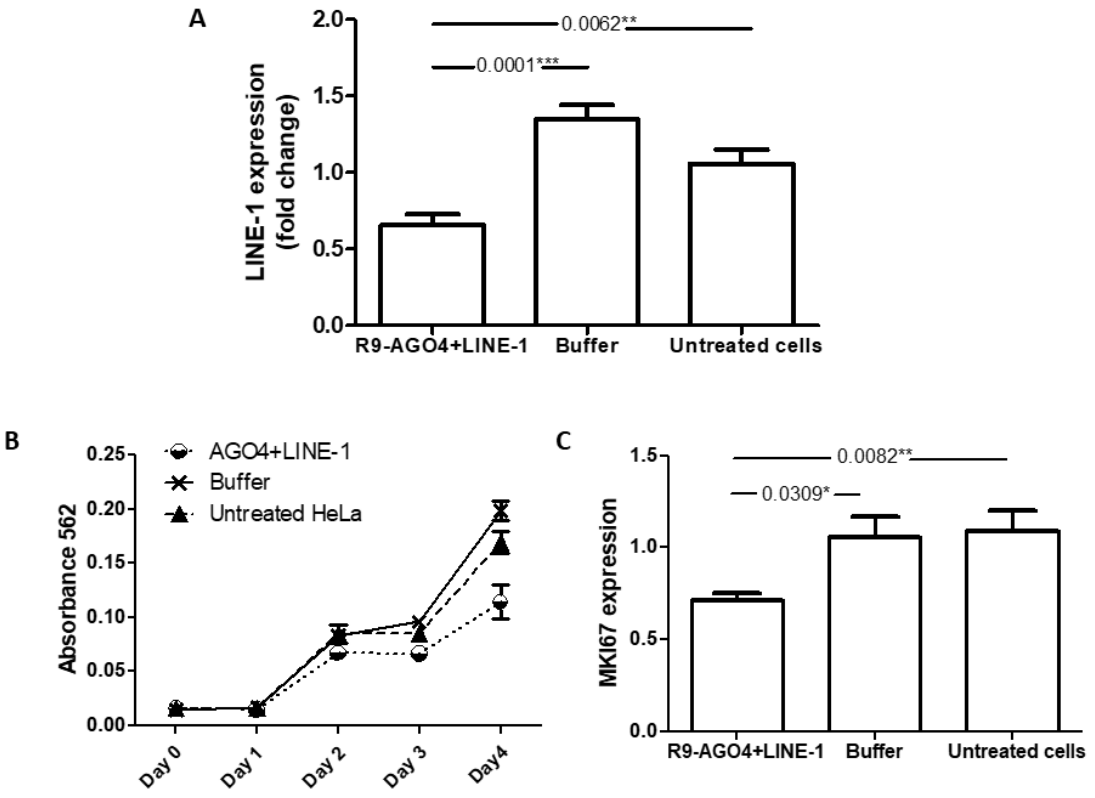


Figure 32 The increasing of LINE-1 methylation due to R9-AGO4-LINE-1 also resulted in an alleviation of LINE-1 mRNA (A). Moreover, cells with R9-AGO4-LINE-1 transfection showed the delay in cell proliferation compared to their control counterparts test by MTT (B) and MKI67 expression (C). The results were performed in 5 replicates, scale bar represented a standard error (SEM).

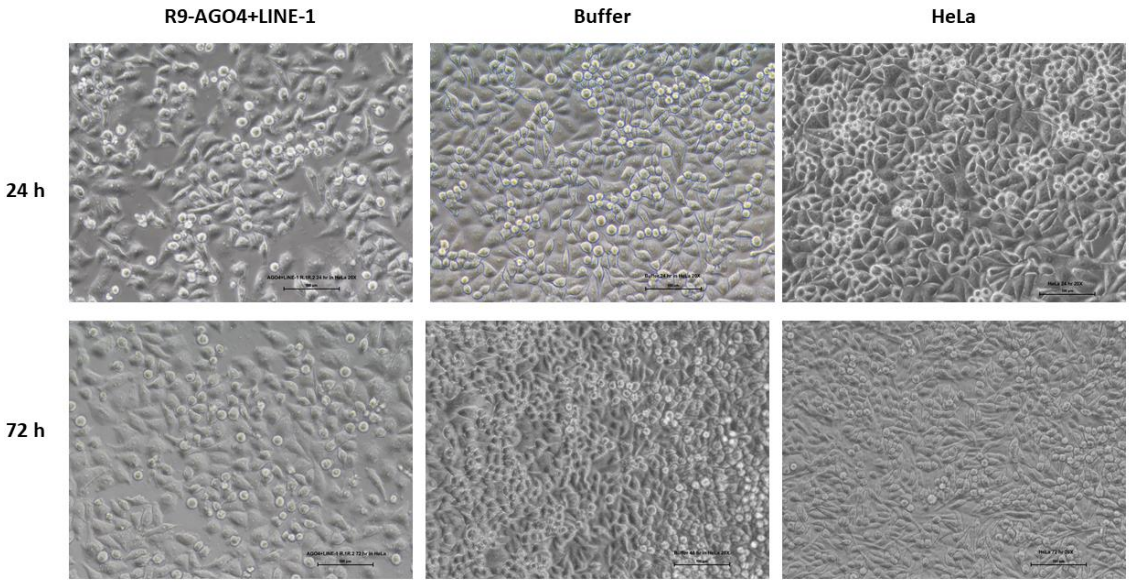


Figure 33 Transfection of R9-AGO4-LINE-1 delayed in HeLa cells proliferation compared to its control counterparts, observed at 24 h (upper) and 72 h (lower) after transfection using a phase contrast microscope. Scale denotes 100 μm.

3.1.2 Single gene

A. *EML2*

Finally, we evaluated whether CPP-AGO4-sgRNA can methylate unique sequences. We selected *EML2* and *CCNA1* for this study. Immediate 3' to the *EML2* transcriptional start site was methylated only in nerve cells and completely unmethylated in most epithelial cell lines (32). We tried to induce the methylation of *EML2* using CPP-AGO4-EML2 in HeLa and HEK293 cell lines. The results showed that R9-AGO4-EML2 was able to promote *EML2* methylation in HeLa and HEK293 cells, while X-AGO4-EML2 could not induce EML2 methylation (Figure 34A and 34B). However, R9-AGO4-EML2 provided a wide range of methylation percentages (Figure 34C). The induction of methylation of *EML2* at CpG dinucleotides complementary to the sgRNA was confirmed by Sanger sequencing (Figure 34D). Thereafter, we attempted to improve the transfection method to enhance *de novo* methylation levels of target genes. Multiple transfections and dose-dependent protein amount using for transfections were

performed; however, these did not ameliorate *EML2* methylation level. (Figure 35A). Additionally, CPP-AGO4-EML2 did not affect the genome-wide methylation observed for Alu (Figure 35B).

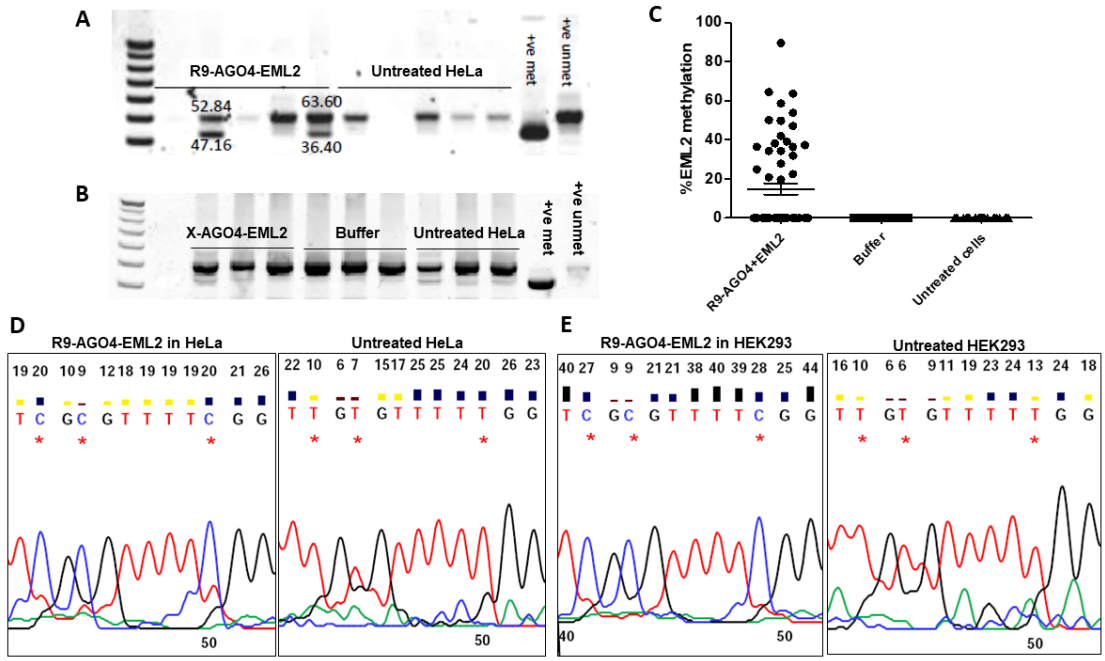


Figure 34 The increasing of EML2 methylation when (A) R9-AGO4-EML2 was transfected to HeLa cells. (B) However, X-AGO4-EML2 transfection could not augment EML2 methylation level. (C) R9-AGO4-EML2 transfection caused *EML2* methylation in a wide range; moreover, *EML2* methylation was not observed every time after transfection. (D and E) The results of *EML2* methylation induction were confirmed by sodium bisulfite-combined Sanger sequencing in HeLa and HEK293 cells, respectively.

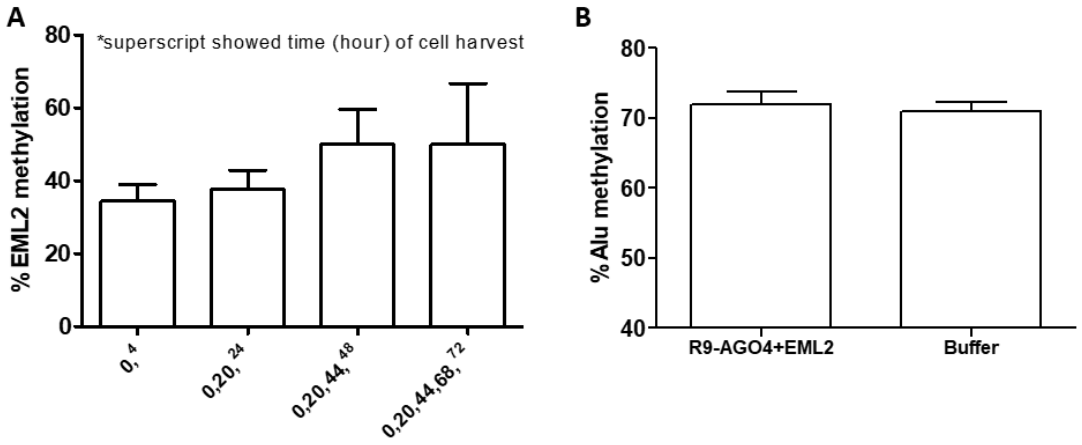


Figure 35 Multiple transfections using R9-AGO4-EML2 were attempted. (A) Improvement of *EML2* methylation levels was found. (B) Higher *EML2* methylation caused by R9-AGO4-EML2 could not lead to off-target Alu methylation. The results were performed in 5 replicates, scale bar represented a standard error (SEM).

B. *CCNA1*

The promoter of *CCNA1* is partially methylated in the C33a cervical cancer cell line (31). When we induced methylation using R9-AGO4-CCNA1, the results demonstrated that *CCNA1* methylation was increased (Figure 36A). Interestingly, the sense strand sgRNA provided a higher level of *CCNA1* methylation than the antisense strand without any off-target effects detected by ALU methylation (Figure 36B), while the antisense strand sgRNA resulted in a higher level of *EML2* methylation than the sense strand. It is possible that either the sense or antisense strand sgRNA could induce methylation in complex with the AGO4 protein. Moreover, the methylation level caused by the R9-AGO4-CCNA1 sense strand also resulted in the repression of *CCNA1* expression (Figure 36C).

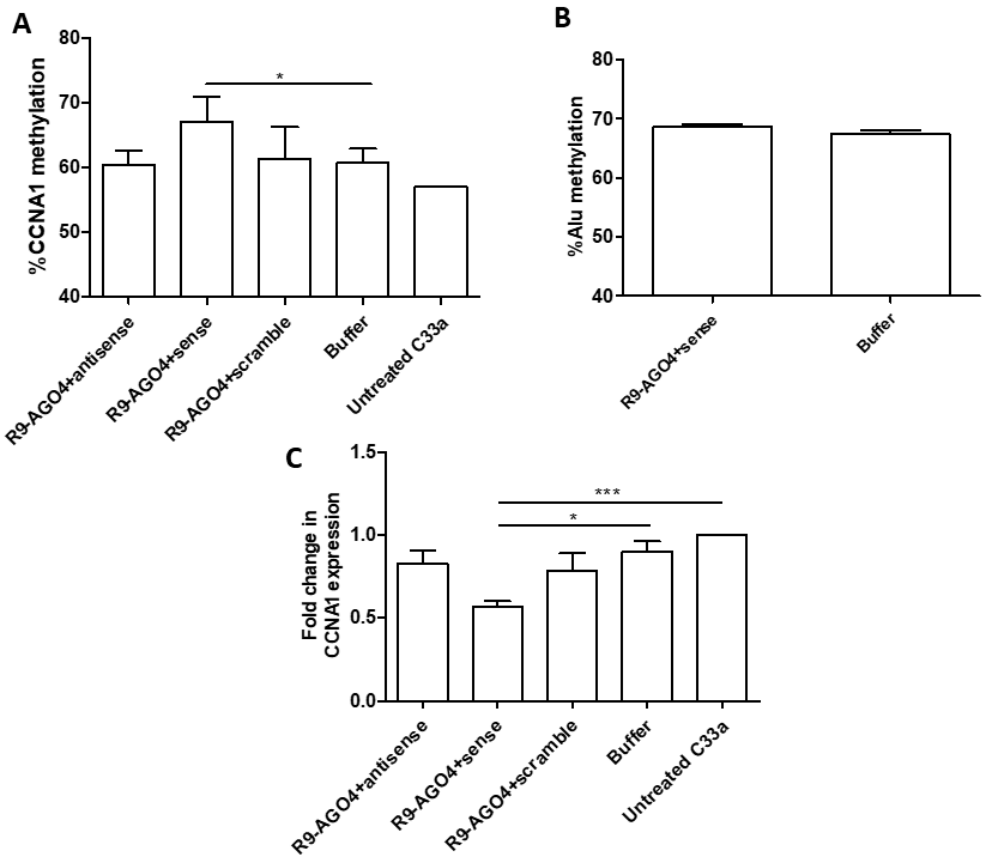


Figure 36 Induction of *CCNA1* promoter methylation by R9-AGO4-CCNA1. (A and B) By using sense strand *CCNA1* sgRNA incorporated with R9-AGO4 provided a higher methylation level of *CCNA1* than other groups, without off-target of Alu methylation. (C) Accumulation of *CCNA1* methylation resulted in *CCNA1* expression inhibition. The results were performed in 5 replicates, scale bar represented a standard error (SEM).

CHAPTER V
DISCUSSION

In this study, we performed several experiments to determine that human AGO4 is a crucial protein in the RdDM complex. First, we found AGO protein localization with DNA methylation. Second, AGO4 overexpression increases IRS methylation of the genome and AGO4-bound sequences. Finally, DNA methylation of DNA sequences homologous to sgRNA was increase when exogenous R9-AGO4 was introduced implying that the purified R9-AGO4 can be developed further to be used as an alternative epigenomic editing protein.

Due to our recent publication in bioinformatics study, Pearson correlation coefficients were used to evaluate the correlation of promoter methylation and the number of AGO-binding sites. A positive correlation was found for AGO4. Therefore, we hypothesized that human AGO4 plays a role in RdDM. However, this function might be redundant with other proteins. As the supported evidences were found from chi-square test of the association of methylation levels and protein-binding sites (Table 1), the results showed that AGO1 to AGO3 are also statistically significant; however, with lesser degree of that AGO4. Consequently, we postulated that the function RdDM of human AGO4 might be redundant with other AGO proteins. This may explain why our results did not show huge significances in methylation changes either endogenous or exogenous AGO4 overexpression or AGO4 depletion.

In addition to confirm whether AGO4 is involved in RdDM, overexpression of exogenous AGO4 provided the results that revealed significant relation between AGO4 binding and higher methylation levels of LINE-1 and Alu. Moreover, CPP-AGO4-sgRNA transfection can specifically promote LINE-1, Alu, *EML2* and *CCNA1* methylation. Some of those methylation changes like LINE-1 or *CCNA1* also contributed to their expression

and cell phenotypic changes. Therefore, all data obtained is in good agreement that AGO4 protein is one of the proteins processed RdDM in human cells. However, there are some points that we would like to draw attention to.

For human CPP-AGO4 protein production, we selected to manipulate our proteins in *Escherichia coli* as it provided the high cell density cultures, meaning that we will obtain high recombinant protein yields in short times. Yet, proteins produced from this platform are not post-translated (71), the previous studies only found that AGO2 possess phosphorylation for mRNA binding (72) and sumoylation for control its stability (73). However, more research in structure of each human AGO protein including AGO4 is desirable before obtaining a definitive answer of human RdDM mechanism or to apply them for epigenomic editing tool. For example, we can engineer only DNA methyltransferase binding sequences of AGO4 with CPP fusion and use it for specific methylation induction.

In order to use CPP-AGO4 as an epigenetic editing tool, when R9-EGFP was transfected to cells, we observed that cells start to die after transfection for 24 h. This might be the reason why R9-AGO4-LINE-1 or Alu transfection provided just 3 to 5 percent increase in methylation and R9-AGO4-EML2 transfection caused *EML2* methylation in wide ranges. However, Alu and LINE-1 account for about a million copy in human genome (74, 75), change in methylation at 3 to 5 percentage of LINE-1 or Alu is composed of 30,000 to 50,000 copies. This may be enough to affect human genome stability or the variable methylation at different LINE-1 or Alu loci may give rise to different cellular phenotypes (23). For xentry-AGO4-sgRNA, it cannot increase any methylation status, this might be because our recombinant protein cannot translocate into nucleus and process RdDM. The most likely explanation of the negative result is observed from z-stack, most xentry-EGFP resided at cell surface or outside nucleus. To use xentry-AGO4 as an epigenomic editing tool, nuclear localization sequences (NLS)

should be engineered into the recombinant protein to ensure that our desired protein will penetrate to nucleus. Until that time, xentry-AGO4-sgRNA should be exercised again as well as other cell penetrating peptides should be intensively studied, the one that causes less harmful to cells and provides the most specific at targets should be selected.

Epigenomic editing could be beneficial in *ex vivo* or *in vitro* therapy. To date, ZFNs, TALENs, CRISPR-cas9 and Dnmt3a with a catalytically inactive Cas9 have been developed for future personalized medicine at both the genomic and epigenomic levels (76-78). CPP-AGO4-sgRNA is different from the other epigenetic editing technology. While CPP-AGO4-sgRNA used protein transfection, the others used transgenes. Transgene technology is more effective. Unfortunately, plasmid transfection has a drawback. Transgene technology can generate an insertion of recombinant DNA into host genomes that might lead to unwanted mutations (79, 80). Epigenomic editing using protein transfection tools such as CPP-AGO4 can prevent this unwanted complication. The CRISPR-cas9 protein complex is 3 times larger than CPP-AGO4 (7, 81). Moreover, CRISPR-cas9 are not human proteins (81). Therefore, CPP-AGO4 will be a more appropriate choice for therapies using protein transduction technology in the future.

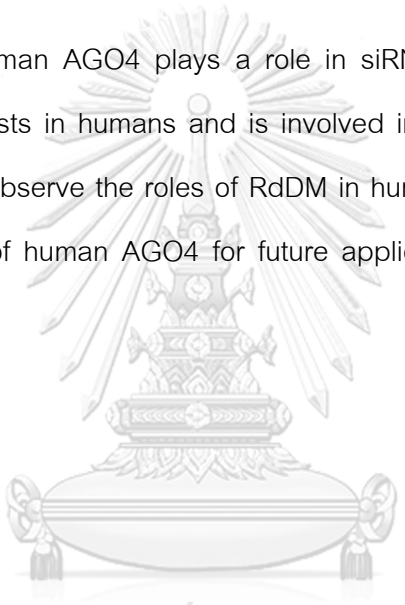
Additionally, human RdDM complexes should be further explored in detail. In plants, there are two effector AGO proteins that are involved in RdDM, one of which is AGO4 (82). AGO4/siRNA complexes are selectively transported into the nucleus and then recruit *de novo* methylase DRM2 to methylate target DNA (83, 84). As noted before, DRM2 plays a role in plant RdDM and is an ortholog of human DNMT3, which is in accordance with the results we obtained showing that AGO4 participates in *de novo* methylation. Therefore, the association of RdDM effector protein complexes, including AGO4, each DNMT (DNMT3A/3B, and DNMT1), and other proteins that serve as transcription factors should be investigated. Another issue that might be a piece of the

puzzle is that plant RdDM is processed by RNA polymerase IV and V (Pol IV and Pol V). These enzymes have conserved amino acid sequences similar to Pol II, but their largest domain is different from Pol II, suggesting that this domain might have evolved to function in RdDM (85). However, Pol I, II and III are present in humans (86), and the specialized function of Pol involved in RdDM should therefore be identified. Moreover, evidence from *S. pombe* and human cells showed the presence of AGO1-mediated heterochromatin formation during transcriptional gene silencing, together with small RNAs (87, 88); similarly, AGO3 is loaded by piRNAs derived from transposon mRNAs in *Drosophila* leading to the hypothesis that AGO3-generating piRNA could play a role in chromatin formation at transposon loci (89). Thus, it is possible that human AGO1 and AGO3 could have interrelated effects on the methylation levels of histones and DNA. All of these observations give rise to one last question that should be addressed: Is there any interplay between DNA methylation and chromatin modification caused by cohesion of AGO proteins? Notably, our genome-wide correlation study showed that the correlation between AGO2 and methylation was far less significant than that of the other AGO protein; thus AGO2 may not play a role in RdDM and may only function in RNAi processes only. Futhermore, a study in mice indicated by generating individual AGO-deficient cells, each AGO possesses miRNA silencing activity. However, AGO2 yielded the greatest effect (90). Consistent with the results of PR-9 anti-gene transfection in AGO-depleted cells, AGO2-lacking cells showed the largest decrease in PR-9 gene expression, while other AGOs had little effect (91). As the greatest quantity of AGO2 was found in humans, all of these lines of evidence led us to address whether the amount of AGO protein in cells could be a factor determining the pathway being used for gene silencing.

DNA methylation plays roles in several processes, such as genomic imprinting, x-inactivation, cell differentiation, aging and adaptation to the environment. Our genome-

wide study of colocalization between AGO4 binding and DNA methylation revealed that human AGO4 methylates DNA located immediately downstream of the TSS. This location was also identified in our previous tissue-specific DNA methylation study (27), suggesting that the tissue specificity might be modulated by RdDM. However, as AGO4 can also bind both LINEs and SINEs and contribute to the control of gene expression and prevention of genomic instability (92), other functions of AGO4 should also be explored.

In summary, human AGO4 plays a role in siRNA-mediated DNA methylation, implying that RdDM exists in humans and is involved in *de novo* methylation. Further studies are needed to observe the roles of RdDM in human cells, and this study might shed light on the use of human AGO4 for future applications in epigenomic therapy.



REFERENCES

1. Wassenegger M, Heimes S, Riedel L, Sänger HL. RNA-directed de novo methylation of genomic sequences in plants. Cell. 1994;76(3):567-76.

2. He X-J, Hsu Y-F, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, et al. An effector of RNA-directed DNA methylation in Arabidopsis is an ARGONAUTE 4-and RNA-binding protein. Cell. 2009;137(3):498-508.

3. Xie M, Yu B. siRNA-directed DNA methylation in plants. Current genomics. 2015;16(1):23-31.

4. Zilberman D, Cao X, Jacobsen SE. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science. 2003;299(5607):716-9.

5. Castanotto D, Tommasi S, Li M, Li H, Yanow S, Pfeifer GP, et al. Short hairpin RNA-directed cytosine (CpG) methylation of the RASSF1A gene promoter in HeLa cells. Molecular Therapy. 2005;12(1):179-83.

6. Morris KV, Chan SW-L, Jacobsen SE, Looney DJ. Small interfering RNA-induced transcriptional gene silencing in human cells. Science. 2004;305(5688):1289-92.

7. Sasaki T, Shiohama A, Minoshima S, Shimizu N. Identification of eight members of the Argonaute family in the human genome☆. Genomics. 2003;82(3):323-30.

8. Schürmann N, Trabuco LG, Bender C, Russell RB, Grimm D. Molecular dissection of human Argonaute proteins by DNA shuffling. Nature structural & molecular biology. 2013;20(7):818-26.

9. Wang D, Zhang Z, O'Loughlin E, Lee T, Houel S, O'Carroll D, et al. Quantitative functions of Argonaute proteins in mammalian development. Genes & development. 2012;26(7):693-704.

10. Komashko VM, Farnham PJ. 5-azacytidine treatment reorganizes genomic histone modification patterns. Epigenetics. 2010;5(3):229-40.

11. Khorshid M, Rodak C, Zavolan M. CLIPZ: a database and analysis environment for experimentally determined binding sites of RNA-binding proteins. Nucleic acids research. 2010;39(suppl_1):D245-D52.

12. Chalertpet K, Pin-on P, Aporntewan C, Maturada P, Ingrungruanglert P, Israsena N, et al. Argonaute-4 as an effector protein in RNA-directed DNA methylation in human cells. Frontiers in genetics. 2019;10:645.

13. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin EV, et al. The evolutionary journey of Argonaute proteins. Nature structural & molecular biology. 2014;21(9):743.

14. Madani F, Lindberg S, Langel Ü, Futaki S, Gräslund A. Mechanisms of cellular uptake of cell-penetrating peptides. Journal of biophysics. 2011;2011.

15. Schmidt N, Mishra A, Lai GH, Wong GC. Arginine-rich cell-penetrating peptides. FEBS letters. 2010;584(9):1806-13.

16. Herce H, Garcia A, Litt J, Kane R, Martin P, Enrique N, et al. Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides. Biophysical journal. 2009;97(7):1917-25.

17. Jo J, Hong S, Choi WY, Lee DR. Cell-penetrating peptide (CPP)-conjugated proteins is an efficient tool for manipulation of human mesenchymal stromal cells. Scientific reports. 2014;4.

18. Tünnemann G, Ter-Avetisyan G, Martin RM, Stöckl M, Herrmann A, Cardoso MC. Live-cell analysis of cell penetration ability and toxicity of oligo-arginines. Journal of Peptide Science. 2008;14(4):469-76.

19. Montrose K, Yang Y, Sun X, Wiles S, Krissansen GW. Xentry, a new class of cell-penetrating peptide uniquely equipped for delivery of drugs. Scientific reports. 2013;3:1661.

20. Laufer BI, Singh SM. Strategies for precision modulation of gene expression by epigenome editing: an overview. Epigenetics & chromatin. 2015;8(1):34.

21. Nemudryi A, Valetdinova K, Medvedev S, Zakian S. TALEN and CRISPR/Cas genome editing systems: tools of discovery. Acta Naturae (англоязычная версия). 2014;6(3 (22)).

22. Bae JM, Shin SH, Kwon HJ, Park SY, Kook MC, Kim YW, et al. ALU and LINE-1 hypomethylations in multistep gastric carcinogenesis and their prognostic implications.

International journal of cancer. 2012;131(6):1323-31.

23. Kitkumthorn N, Mutirangura A. Long interspersed nuclear element-1 hypomethylation in cancer: biology and clinical applications. Clinical epigenetics. 2011;2(2):315-30.

24. Luo Y, Lu X, Xie H. Dynamic Alu methylation during normal development, aging, and tumorigenesis. BioMed research international. 2014;2014.

25. Aporntewan C, Phokaew C, Piriyapongsa J, Ngamphiw C, Ittiwut C, Tongsima S, et al. Hypomethylation of intragenic LINE-1 represses transcription in cancer cells through AGO2. PloS one. 2011;6(3):e17934.

26. Moarii M, Boeva V, Vert J-P, Reyal F. Changes in correlation between promoter methylation and gene expression in cancer. BMC genomics. 2015;16(1):873.

27. Muangsub T, Samsuwan J, Tongyoo P, Kitkumthorn N, Mutirangura A. Analysis of methylation microarray for tissue specific detection. Gene. 2014;553(1):31-41.

28. Chalertpet K, Pakdeechaidan W, Patel V, Mutirangura A, Yanatatsaneejit P. Human papillomavirus type 16 E7 oncoprotein mediates CCNA1 promoter methylation. Cancer science. 2015;106(10):1333-40.

29. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. science. 2001;291(5507):1304-51.

30. Naumova OY, Lee M, Rychkov SY, Vlasova NV, Grigorenko EL. Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics. Child development. 2013;84(1):76-88.

31. Cooper GM, Hausman RE, Hausman RE. The cell: a molecular approach: ASM press Washington, DC; 2000.

32. Vaissière T, Sawan C, Herceg Z. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research/Reviews in Mutation Research. 2008;659(1-2):40-8.

33. Vaucheret H, Fagard M. Transcriptional gene silencing in plants: targets, inducers and regulators. TRENDS in Genetics. 2001;17(1):29-35.

34. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-

stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Current Biology. 2001;11(13):1017-27.

35. Aufsatz W, Mette MF, van der Winden J, Matzke AJ, Matzke M. RNA-directed DNA methylation in Arabidopsis. Proceedings of the National Academy of Sciences. 2002;99(suppl 4):16499-506.

36. Mathieu O, Bender J. RNA-directed DNA methylation. Journal of cell science. 2004;117(21):4881-8.

37. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nature Reviews Genetics. 2014;15(6):394-408.

38. Wierzbicki AT, Haag JR, Pikaard CS. Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell. 2008;135(4):635-48.

39. Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005;308(5718):118-20.

40. Cuerda-Gil D, Slotkin RK. Non-canonical RNA-directed DNA methylation. Nature plants. 2016;2(11):16163.

41. Liu L, Chen X. RNA quality control as a key to suppressing RNA silencing of endogenous genes in plants. Molecular plant. 2016;9(6):826-36.

42. Mahfouz MM. RNA-directed DNA methylation: mechanisms and functions. Plant signaling & behavior. 2010;5(7):806-16.

43. Patchsung M, Settayanon S, Pongpanich M, Mutirangura D, Jintarith P, Mutirangura A. Alu siRNA to increase Alu element methylation and prevent DNA damage. Epigenomics. 2018;10(2):175-85.

44. Hutvagner G, Simard MJ. Argonaute proteins: key players in RNA silencing. Nature reviews Molecular cell biology. 2008;9(1):22-32.

45. Willkomm S, Zander A, Gust A, Grohmann D. A prokaryotic twist on argonaute function. Life. 2015;5(1):538-53.

46. Höck J, Meister G. The Argonaute protein family. Genome biology. 2008;9(2):210.

47. Hauptmann J, Kater L, Löffler P, Merkl R, Meister G. Generation of catalytic human

Ago4 identifies structural elements important for RNA cleavage. Rna. 2014;20(10):1532-8.

48. Turchinovich A, Surowy H, Tonevitsky A, Burwinkel B. Interference in transcription of overexpressed genes by promoter-proximal downstream sequences. Scientific reports. 2016;6:30735.

49. Valdmanis PN, Gu S, Schüermann N, Sethupathy P, Grimm D, Kay MA. Expression determinants of mammalian argonaute proteins in mediating gene silencing. Nucleic acids research. 2011;40(8):3704-13.

50. Habault J, Poyet J-L. Recent advances in cell penetrating peptide-based anticancer therapies. Molecules. 2019;24(5):927.

51. Derakhshankhah H, Jafari S. Cell penetrating peptides: A concise review with emphasis on biomedical applications. Biomedicine & Pharmacotherapy. 2018;108:1090-6.

52. Ye J, Liu E, Yu Z, Pei X, Chen S, Zhang P, et al. CPP-assisted intracellular drug delivery, what is next? International journal of molecular sciences. 2016;17(11):1892.

53. Agrawal P, Bhalla S, Usmani SS, Singh S, Chaudhary K, Raghava GP, et al. CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. Nucleic acids research. 2015;44(D1):D1098-D103.

54. Guidotti G, Brambilla L, Rossi D. Cell-penetrating peptides: from basic research to clinics. Trends in pharmacological sciences. 2017;38(4):406-24.

55. Singh T, Murthy AS, Yang H-J, Im J. Versatility of cell-penetrating peptides for intracellular delivery of siRNA. Drug delivery. 2018;25(1):1996-2006.

56. Allolio C, Magarkar A, Jurkiewicz P, Baxová K, Javanainen M, Mason PE, et al. Arginine-rich cell-penetrating peptides induce membrane multilamellarity and subsequently enter via formation of a fusion pore. Proceedings of the National Academy of Sciences. 2018;115(47):11923-8.

57. Nakase I, Niwa M, Takeuchi T, Sonomura K, Kawabata N, Koike Y, et al. Cellular uptake of arginine-rich peptides: roles for macropinocytosis and actin rearrangement. Molecular therapy. 2004;10(6):1011-22.

58. Wender PA, Mitchell DJ, Pattabiraman K, Pelkey ET, Steinman L, Rothbard JB. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake:

peptoid molecular transporters. Proceedings of the National Academy of Sciences. 2000;97(24):13003-8.

59. Mitchell DJ, Steinman L, Kim D, Fathman C, Rothbard J. Polyarginine enters cells more efficiently than other polycationic homopolymers. The Journal of Peptide Research. 2000;56(5):318-25.

60. Park J, Ryu J, Kim K-A, Lee HJ, Bahn JH, Han K, et al. Mutational analysis of a human immunodeficiency virus type 1 Tat protein transduction domain which is required for delivery of an exogenous protein into mammalian cells. Journal of general virology. 2002;83(5):1173-81.

61. Jo J, Hong S, Choi WY, Lee DR. Cell-penetrating peptide (CPP)-conjugated proteins is an efficient tool for manipulation of human mesenchymal stromal cells. Scientific reports. 2014;4:4378.

62. Rejinold NS, Han Y, Yoo J, Seok HY, Park JH, Kim Y-C. Evaluation of cell penetrating peptide coated Mn: ZnS nanoparticles for paclitaxel delivery to cancer cells. Scientific reports. 2018;8(1):1899.

63. Trenner A, Godau J, Sartori AA. A short BRCA2-derived cell-penetrating peptide targets RAD51 function and confers hypersensitivity toward PARP inhibition. Molecular cancer therapeutics. 2018;17(7):1392-404.

64. Montrose K, Yang Y, Krissansen GW. The tetrapeptide core of the carrier peptide Xentry is cell-penetrating: novel activatable forms of Xentry. Scientific reports. 2014;4:4900.

65. Mahmood T, Yang P-C. Western blot: technique, theory, and trouble shooting. North American journal of medical sciences. 2012;4(9):429.

66. Carey MF, Peterson CL, Smale ST. Chromatin immunoprecipitation (chip). Cold Spring Harbor Protocols. 2009;2009(9):pdb. prot5279.

67. Rao X, Huang X, Zhou Z, Lin X. An improvement of the 2ˆ (–delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. Biostatistics, bioinformatics and biomathematics. 2013;3(3):71.

68. Glick BR, Pasternak JJ, Patten CL. Molecular biotechnology: principles and

applications of recombinant DNA: Washington, DC: ASM Press; 2010.

69. De N, MacRae IJ. Purification and assembly of human Argonaute, Dicer, and TRBP complexes. Argonaute Proteins: Springer; 2011. p. 107-19.

70. Davis SJ, Vierstra RD. Soluble, highly fluorescent variants of green fluorescent protein (GFP) for use in higher plants. Plant molecular biology. 1998;36(4):521-8.

71. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Frontiers in microbiology. 2014;5:172.

72. Huberdeau MQ, Zeitler DM, Hauptmann J, Bruckmann A, Fressigné L, Danner J, et al. Phosphorylation of Argonaute proteins affects mRNA binding and is essential for microRNA-guided gene silencing in vivo. The EMBO journal. 2017;36(14):2088-106.

73. Sahin U, Lapaquette P, Andrieux A, Faure G, Dejean A. Sumoylation of human argonaute 2 at lysine-402 regulates its stability. PLoS One. 2014;9(7):e102957.

74. Deininger P. Alu elements: know the SINEs. Genome biology. 2011;12(12):236.

75. Rodić N, Burns KH. Long interspersed element–1 (LINE-1): passenger or driver in human neoplasms? PLoS genetics. 2013;9(3):e1003402.

76. Gaj T, Gersbach CA, Barbas III CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in biotechnology. 2013;31(7):397-405.

77. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, et al. Editing DNA methylation in the mammalian genome. Cell. 2016;167(1):233-47. e17.

78. Vojta A, Dobrinić P, Tadić V, Bočkor L, Korać P, Julg B, et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. Nucleic acids research. 2016;44(12):5615-28.

79. Bachu R, Bergareche I, Chasin LA. CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. Biotechnology and bioengineering. 2015;112(10):2154-62.

80. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013;153(4):910-8.

81. Horvath P, Barrangou R. CRISPR/Cas, the immune system of bacteria and

archaea. Science. 2010;327(5962):167-70.

82. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell. 2005;123(4):607-20.

83. Ketting RF. The many faces of RNAi. Developmental cell. 2011;20(2):148-61.

84. Hegge JW, Swarts DC, van der Oost J. Prokaryotic Argonaute proteins: novel genome-editing tools? Nature reviews Microbiology. 2017.

85. Huang L, Jones AM, Searle I, Patel K, Vogler H, Hubner NC, et al. An atypical RNA polymerase involved in RNA silencing shares small subunits with RNA polymerase II. Nature structural & molecular biology. 2009;16(1):91.

86. Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Molecular cell. 2012;45(4):439-46.

87. Kim DH, Villeneuve LM, Morris KV, Rossi JJ. Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. Nature structural & molecular biology. 2006;13(9):793.

88. Qi Y, He X, Wang X-J, Kohany O, Jurka J, Hannon GJ. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. Nature. 2006;443(7114):1008.

89. Ye R, Wang W, Iki T, Liu C, Wu Y, Ishikawa M, et al. Cytoplasmic assembly and selective nuclear import of Arabidopsis Argonaute4/siRNA complexes. Molecular cell. 2012;46(6):859-70.

90. Su H, Trombly MI, Chen J, Wang X. Essential and overlapping functions for mammalian Argonautes in microRNA silencing. Genes & development. 2009;23(3):304-17.

91. Chu Y, Yue X, Younger ST, Janowski BA, Corey DR. Involvement of argonaute proteins in gene silencing and activation by RNAs complementary to a non-coding transcript at the progesterone receptor promoter. Nucleic acids research. 2010;38(21):7736-48.

92. Mutirangura A. A Hypothesis to Explain How the DNA of Elderly People Is Prone to Damage: Genome-Wide Hypomethylation Drives Genomic Instability in the Elderly by

Reducing Youth-Associated Gnome-Stabilizing DNA Gaps. Epigenetics: IntechOpen; 2019.

VITA

NAME	กัญวลัชญ์ ชเลิศเพ็ชร์
DATE OF BIRTH	26 ธันวาคม 2528
PLACE OF BIRTH	สงขลา
INSTITUTIONS ATTENDED	ปริญญาตรี พันธุศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปริญญาโท พันธุศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ปริญญาเอก ชีวเวชศาสตร์ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
HOME ADDRESS	354 ซอย 33 หมู่บ้านธงชัย ถนนรังสิต-นครนายก ตำบลประชาธิปัตย์ อำเภอธัญบุรี จังหวัดปทุมธานี
PUBLICATION	1. Chalertpet K, Pin-on P, Aporntewan C, Maturada P, Ingrungruanglert P, Israsena N, et al. Argonaute-4 as an effector protein in RNA-directed DNA methylation in human cells. Frontiers in genetics. 2019;10:645. 2. Puttipanyalears C, Arayataweegool A, Chalertpet K, Rattanachayoto P, Mahattanasakul P, Tangjaturonsasme N, Kerekhanjanarong V, Mutirangura A, Kitkuthorn N. TRH site-specific methylation in oral and oropharyngeal squamous cell carcinoma. BMC Cancer. 2018; 18(1): 786. 3. Chalertpet K, Pakdeechaidan W, Patel V, Mutirangura A, Yannatatsaneejit P. Human Papillomavirus type 16 E7 oncoprotein mediates CCNA1 promoter methylation. Cancer Sci. 2015; 106(10): 1333-40.
AWARD RECEIVED	2014-2016 Doctoral Degree Chulalongkorn University 100th

Year Birthday

2015 รางวัลชมเชย สาขาวิทยาศาสตร์ ผลงานวิจัยในการประชุมวิชาการครั้งที่52 เรื่อง กลไกของโปรตีน E7 ของ Human Papillomavirus ไทป์ 16 ใน การเกิดเมทิลเลชัน และการแสดงออกของยีน CCNA1ในมะเร็งปากมดลูก

2014 รางวัลศาสตราจารย์กิตติคุณ ดร. ถาวร วัชราภัย ผลการสอบวิทยานิพนธ์ ขั้นดีมาก หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาพันธุศาสตร์ ปีการศึกษา 2556

2014 รางวัลผู้สอบได้คะแนนยอดเยี่ยม หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาพันธุศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จากมูลนิธิศาสตราจารย์ดร.แถบ นีละนิธิ