Exploring the miRNA binding network associated with HOTAIRM1 during early neuronal differentiation

Yuqiong (Rachel) Ding

Department of Biochemistry, McGill University, Montréal

November 2023

A THESIS SUBMITTED TO MCGILL UNIVERSITY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF THE DEGREE OF MASTER OF SCIENCE IN BIOCHEMISTRY.

Table of Contents

Abstract (english)

LncRNAs are known to play pivotal roles in various biological processes, including gene regulation. HOTAIRM1, a lncRNA residing in the *HOXA* gene cluster, is extensively studied in cancer research due to its link with various carcinomas. Its aberrant expression has been associated with various carcinomas and is suggested as a diagnostic biomarker for different cancers. Recently, our lab and others have found its essential role in mediating neuronal differentiation. Specifically, it disrupts the pluripotency network maintained in undifferentiated cells and facilitate cellular progression through the early stage of neuronal differentiation. However, the mechanisms that underlie this phenomenon have yet to be fully elucidated. In this research we aimed to explore the microRNA (miRNA) binding network associated with HOTAIRM1 and uncover the mechanisms underpinning early neuronal differentiation. We constructed a NCCIT cell line that contains a tethering system, enabling an efficient pulldown of endogenous HOTAIRM1 from the cytoplasm. This modified cell successfully underwent differentiation upon exposure to retinoic acid, mirroring wildtype NCCIT cells. We also tentatively identified two miRNAs that may interact with HOTAIRM1 during neuronal differentiation. However, it's important to note that these results should be interpreted with caution due to certain sequencing challenges. While we recognize the issues in our sequencing results, we still obtained valuable information on how to optimize our system and pulldown conditions for future experiments. In summary, our cell model serves as a valuable tool to explore the HOTAIRM1 interactome, with the potential to provide insights for neuronal development.

Résumé (français)

Les ARNlncs (ARN longs non-codants) jouent un rôle essentiel dans divers processus biologiques, notamment la régulation des gènes. HOTAIRM1, un ARNlnc situé dans le cluster de gènes *HOXA*, est l'objet d'études approfondies en recherche sur le cancer en raison de ses liens avec différents types de carcinomes. Plus récemment, notre laboratoire et d'autres ont découvert son rôle essentiel dans la médiation de la différenciation neuronale. Plus précisément, il perturbe le réseau de pluripotence présent dans les cellules non-différenciées et facilite la progression cellulaire au cours des premières étapes de la différenciation neuronale. Cependant, les mécanismes sous-jacents à ce phénomène restent à élucider. Dans le cadre de nos recherches, nous avons exploré le réseau des microARNs (miARNs) associé à HOTAIRM1 et élucidé les mécanismes de la différenciation neuronale précoce. Nous avons mis au point une lignée cellulaire de NCCIT dotée d'un système d'ancrage, ce qui nous a permis de réaliser efficacement l'immunoprécipitation de l'HOTAIRM1 endogène situé dans le cytoplasme. Cette lignée cellulaire modifiée a été capable de se différencier avec succès lorsqu'exposée à l'acide rétinoïque, imitant le comportement des cellules NCCIT de type sauvage. Nous avons également identifié provisoirement deux miARNs qui pourraient interagir avec HOTAIRM1 lors de la différenciation neuronale. Il est toutefois important de noter que ces résultats doivent être interprétés avec prudence en raison de certaines difficultés de séquençage. Malgré les problèmes rencontrés dans nos résultats de séquençage, nous avons obtenu des informations précieuses sur la manière d'optimiser notre système et nos conditions d'immunoprécipitation en vue d’expériences futures. En résumé, notre modèle cellulaire constitue un outil précieux

pour explorer l'interactome de HOTAIRM1, avec le potentiel de fournir des informations essentielles pour la recherche sur le développement neuronal.

Acknowledgements

I would like to express my deepest appreciation to the most important person in my academic journey, my supervisor, Dr. Josée Dostie! Her support, guidance, and genuine care have been instrumental in my development during my time at McGill. Under her mentorship, I not only learned the skills of a professional researcher but also developed as a resilient individual. I will forever cherish the opportunity she provided me to learn and grow in her lab. I would also like to thank my committee members: Drs. Sidong Huang and Maria Vera Ugalde for their valuable insights to my project.

A special thanks goes to Raphaëlle Lambert from IRIC for her contribution to miRNA-sequencing and to Patrick Gendron for the miRNA data analysis.

My work would not have been done without helps from my colleagues from Dostie lab (listed below). I want to extend my thanks to Dana for her guidance and training; Ian, for being the steadfast companion through all the experiment highs and lows. And Romane, thanks for all the boundless energy and positivity you have brough to lab everyday.

Lab members:

Xianjin Lian

Dana Segal

Ian Zhao

Romane Monnet

Natalie Schalick (former)

Afrida Ahmed (former)

Theodora Georgakopoulou (former)

Chloe Xu

Ali Shalhoub

Riley McMaster

Laura Vaugeois

Furthermore, I'd like to express my gratitude to people from Sidong Huang's lab who helped me with lentivirus generation (Xianbing Zhu, Fu Zheng, and Paco Jiang). Also, thanks to Hannah and all my friends from sid's and Jerry's lab for being the source of motivation and for sharing countless moments of joy with me outside the lab. I sincerely hope that our friendship will last a lifetime.

Last but not least, I want to acknowledge my family: Haixia Chen, Renhui Ding, Cuitao Yang, and Tonglin Chen. Your unconditional love and support have sustained me through the years. I hope I will forever make you proud.

Contribution of authors

I authored the entire thesis and conducted all the experiments, except where otherwise indicated. Mir-RNA sequencing and data analysis were performed at the Institute for Research in Immunology and Cancer (IRIC). I am responsible for preparing all the Figures and Tables. Feedback was provided by Dr. Josée Dostie and Dr. Marc Fabian.

List of figures and tables

Appendix Figure 3: Gene expression in WT NCCIT and transduced cell lines.

Appendix Figure 4: Electrophoresis run summary of miRNA-sequencing.

Appendix Figure 5: Expression level of *HOXA*2 increases during a 14-day differentiation in RA-treated WT NCCIT.

Appendix Figure 6. Immunofluorescence analysis of the subcellular localization of λN-HOTAIRM1-RNA complexes in transduced cells (NCCIT_LN).

Abbreviations

AGO2----Argonaute II

AMT----4'-aminomethyltrioxalen

ARHGAP24----Rho GTPase-activating protein 24

Cas9----CRISPR-associated protein 9

cDNA----Complementary DNA

ceRNA----Competing endogenous RNA

CERS1----Ceramide synthase 1

CHIRP----Capture hybridization of RNA targets by pulldown

CRISPR----Clustered regularly interspaced short palindromic repeats

DANCR----Differentiation antagonizing non-protein coding RNA

DCP1----decapping protein

DMEM----Dulbecco's modified eagle medium

EC----Embryonal carcinoma cell

eEF2----Eukaryotic elongation factor 2

ESC---- Embryonic stem cell

FBS----Fetal bovine serum

GST----Glutathione S-transferase

H3K27(me)----Histone 3-Lysine 27 (methylated 1, 3, or 3)

H3K4(me)----Histone 3-Lysine 4 (methylated 1, 2, or 3)

HDR----Homology-directed repair

HOTAIRM1----HOXA transcript antisense RNA, myeloid-specific 1

HOTTIP----HOXA distal transcript antisense RNA

HOXA-AS2----HOXA antisense 2

HOX----Homeobox

KI----Knock- in

LARP1----La ribonucleoprotein domain family, member 1

lincRNA----Long intergenic RNA

lncRNA----Long non-coding RNA

MCP----MS2-coat protein

miRNA----Micro-RNA

MLL----Mixed lineage leukemia (complex)

MOI----Multiplicities of infection

mRNA----Messenger RNA

NANOG----Nanog homeobox

ncRNA----Non-coding RNA

NPC----Nuclear pore complex

NT2-D1----NTERA-2 cl.D1

OCT4----Octamer-binding transcription factor 4

PABPN1----Poly-A binding protein nuclear 1

PBS----Phosphate-buffered saline

PCA----Principal Component Analysis

POU5F1----POU class 5 homeobox 1

PRC2----Polycomb repressor complex 2

RaPID----RNA–Protein Interaction Detection

RAR (α or β)----Retinoic acid receptor (alpha or beta)

RARE----Retinoic acid response element

RA----Retinoic acid

RISC----RNA-induced silencing complex

RNA Pol II----RNA polymerase II

RNA-IP----RNA-immunoprecipitation

RPMI 1640----Roswell Park Memorial Institute 1640

RT-qPCR----Real-time quantitative PCR (shortened to qPCR)

RXR-γ ----Retinoic X receptor gamma

SDS-PAGE----Sodium dodecyl sulfate–polyacrylamide gel electrophoresis

SOX2----SRY-box transcription factor 2

SOX3----SRY-box transcription factor 3

TDMD---- Target-directed miRNA degradation

TF----Transcription factor

TINCR----Terminal differentiation–induced noncoding RNA

UTR----Untranslated region

Xist----X inactive-specific transcript

XRN1----exoribonuclease 1

Rationale

Stem cell pluripotency and differentiation rely on the collaboration of various regulators such as transcription factors (TFs), epigenetic modifiers, and signaling pathways. Among these key players, long non-coding RNAs (lncRNAs) have taken center stage as dynamic contributors to these regulatory mechanisms. Through their multifaceted interactions with DNA, RNA, and proteins, lncRNAs exert control over critical cellular activities, encompassing transcription, splicing, translation, and more [1, 2].

Within the diverse landscape of lncRNAs, our focus lies on HOTAIRM1 (*HOXA* Transcript Antisense RNA, Myeloid-Specific 1), a lncRNA residing in the genomic region between *HOXA1* and *HOXA2*. The expression of HOTAIRM1 has been shown to exhibit a dual role in the context of cancer, where it can function both as an oncogene and a tumor suppressor [3]. In previous studies, we unveiled its role in maintaining the precise arrangement of the *HOXA* gene cluster during development and its involvement in regulating pluripotency during neuronal differentiation [4, 5]. Here, we delve deeper in how HOTAIRM1 acts as a competing endogenous RNA (ceRNA) in the context of neuronal differentiation. We aim to explore the mechanisms behind lncRNA-mediated control of differentiation processes and provide a better understanding of how different regulators influence the destiny of stem cells.

Introduction

Overview of lncRNAs

LncRNAs are a family of RNAs with more than 200 nucleotides in length that lack protein coding potential. Long relegated to the realm of 'transcriptional noise' or byproducts of genomic activity, lncRNAs have emerged as crucial players in the regulation of gene expression and diverse cellular processes. Much like their protein-coding counterparts – the messenger RNAs (mRNA) – most lncRNAs are transcribed by RNA Polymerase II (Pol II), which generates precursor transcripts referred to as primary lncRNAs. These primary lncRNAs can be classified into several categories based on their genomic origin, including intergenic lncRNAs, intronic lncRNAs, sense and antisense lncRNAs, each with a unique regulatory potential [6]. As transcripts of Pol II, lncRNAs undergo a range of co-transcriptional modifications, such as capping and polyadenylation. They also undergo alternative splicing, primarily during transcription but also potentially after transcription. These modifications not only enhance lncRNA stability but also diversify their functional implications, thereby broadening their roles in various biological processes [7].

LncRNA functional diversity across subcellular compartments

LncRNAs have been shown to be differentially expressed across various stages of differentiation, which suggests that they may be novel 'fine tuners' of cell fate. Depending on the cellular environment and cell types, lncRNAs exhibit their different impacts by acting as signaling molecules, decoys, guides, or scaffolds (Fig. 1) [2, 8].

Within the nucleus, lncRNAs were shown to guide epigenetic modifiers to specific chromosomal loci, thereby remodeling chromatin structure and lead to either gene silencing or activation. The illustrious lncRNA Xist (X inactive-specific transcript) represents a paradigmatic example of a lncRNA that binds to the polycomb repressive complex PRC2. This binding leads to the induction of H3K27me3 modifications on chromatin, resulting in the transcriptional silencing of one X chromosome in females during development and throughout life [9]. Moreover, lncRNAs such as ES1 and ES2 have been unveiled as key regulators of pluripotency in human embryonic stem cells. These lncRNAs adopt the role of scaffolds to physically interact with *SOX2* and PRC2 to prevent *SOX2* from activating genes that would otherwise lead to the differentiation of ESCs into specific cell types [1]. In addition, nuclear lncRNAs can also regulate gene transcription through transcriptional interference. During this process, lncRNAs often form stable complexes with gene promoters to hinder binding of the preinitiation complex [10, 11] or act as decoys by binding to transcription factors, preventing them from interacting with their intended targets.

In the cytoplasm, lncRNAs participate in regulating mRNA stability, translation, and modulating post-translation. Many lncRNAs exhibit sequence complementary to transcripts arising from the same chromosomal site or distinct loci. This complementarity enables them to interact with their target RNAs through base pairing and ultimately either repress or promote target decay. For instance, during somatic tissue differentiation, the lncRNA TINCR (terminal differentiation-induced ncRNA) forms associations with Staufen1 protein, binding directly to differentiation-related mRNAs via a 25-nucleotide motif, consequently enhancing mRNA stability [12]. While for lncRNAs

bearing microRNA-complementary sites, they function through two mechanisms. Most commonly, they act as competing endogenous RNAs (ceRNAs) or miRNA sponges to sequester microRNAs, thereby reducing their inhibitory effect on target mRNAs. One example is the developmentally regulated lncRNA H19, which is a well-studied lncRNA harboring both canonical and noncanonical binding sites for the let-7 family of microRNAs. Overexpression of H19 decreases let-7 availability, which consequently upregulates the expression of let-7 target genes, leading to various downstream effects on cell proliferation, differentiation, and tumorigenesis [13]. In another scenario, when lncRNAs show extended complementarity to the 3' region of miRNAs along with a central bulge of ≤ 5nt, a mechanism known as target-directed miRNA degradation (TDMD) can occur [14-16]. This process involves specific post-transcriptional modifications of miRNAs, such as adding a tail to their 3′ end and trimming them (miR-shortening). At least in the case of lncRNA CYRANO-directed miR-7 degradation, an enzyme called ZSWIM8, which serves as the substrate receptor of E3 ubiquitin ligases, can recognize this miRNA modifications and promotes AGO polyubiquitination that triggers its degradation by the proteasome. As miRNAs are released from AGO, they are rapidly broken down by cellular exonucleases [17, 18]. It is worth noting that while this is an intriguing mechanism, there are currently only a few established examples of endogenous TDMD involving mammalian transcripts [19].

LncRNAs further contribute to cellular dynamics by coordinating the degradation of existing proteins following their translation from mRNAs. One of the most known

examples is lncRNA HOTAIR, which promotes the degradation of ATXN1 and SNUPN via interactions with E3 ubiquitin ligases DZIP3 and MEX3B in senescent cells [20].

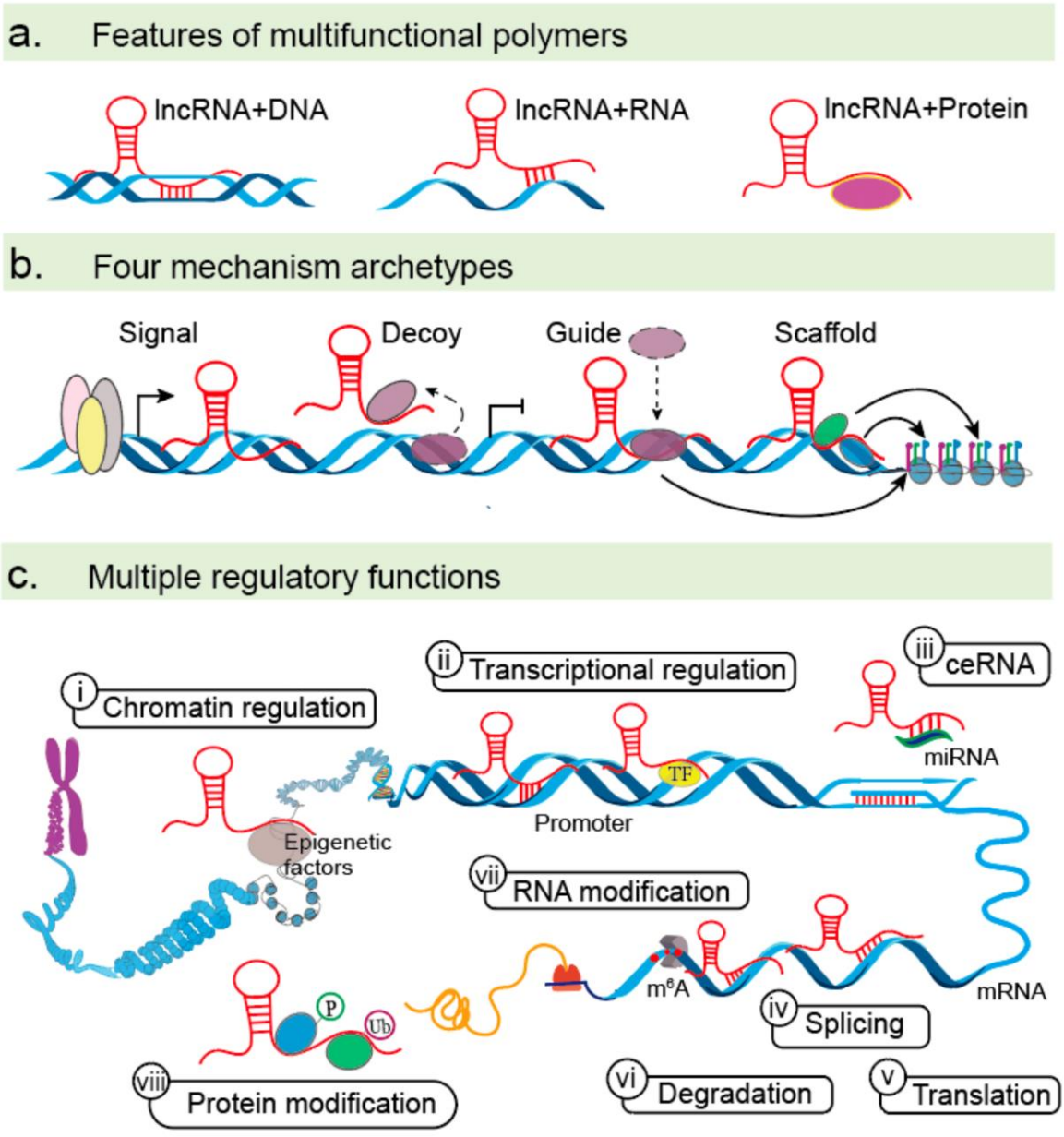


Figure 1. Diagram showing the functions of lncRNAs through different mechanisms. Published by Cheng et al. (2019) in Cells 2019 Vol. 8 Issue 10 Pages 1178 (DOI:10.3390/cells8101178). Reprinted with blanket permission from publisher.

The *HOXA g*enes

The homeobox *(HOX)* genes encode an evolutionary conserved family of TFs that regulate the embryo body plan and contribute to cell specification in several adult differentiation processes. Mammals possess 39 *HOX* genes organized into four clusters (A-D), each residing on a distinct chromosome. These clustered *HOX* genes exhibit a phenomenon known as collinearity where they are sequentially expressed in a 3' to 5' direction, mirroring their position along the chromosome and corresponding to their expression pattern along the body axis [21]. Since activation of each *HOX* gene triggers distinct activities with specific functional outcomes, a precise temporal and spatial regulation control on their expression is particularly crucial. Incorrect timing or failed *HOX* gene activation were shown to produce phenotypic alterations [22, 23].

The strict *HOX* genes expression arises through different morphogens including retinoic acid (RA), which is a derivative of vitamin A and functions as a ligand to activate nuclear receptors RARs (retinoic acid receptors) and RXRs (retinoid X receptors). They then form a complex that can bind specific DNA sequences known as retinoic acid response elements (RAREs) in the regulatory regions of target genes. The *HOX* clusters contain a large number of RAREs that give regulatory inputs both locally on nearby *HOX* genes and distantly to coordinately regulate gene expressions [24]. In addition to the right signal, cooperation between many other components such as lncRNAs are thought of as essential to precisely control when and where *HOX* genes activate during development [25].

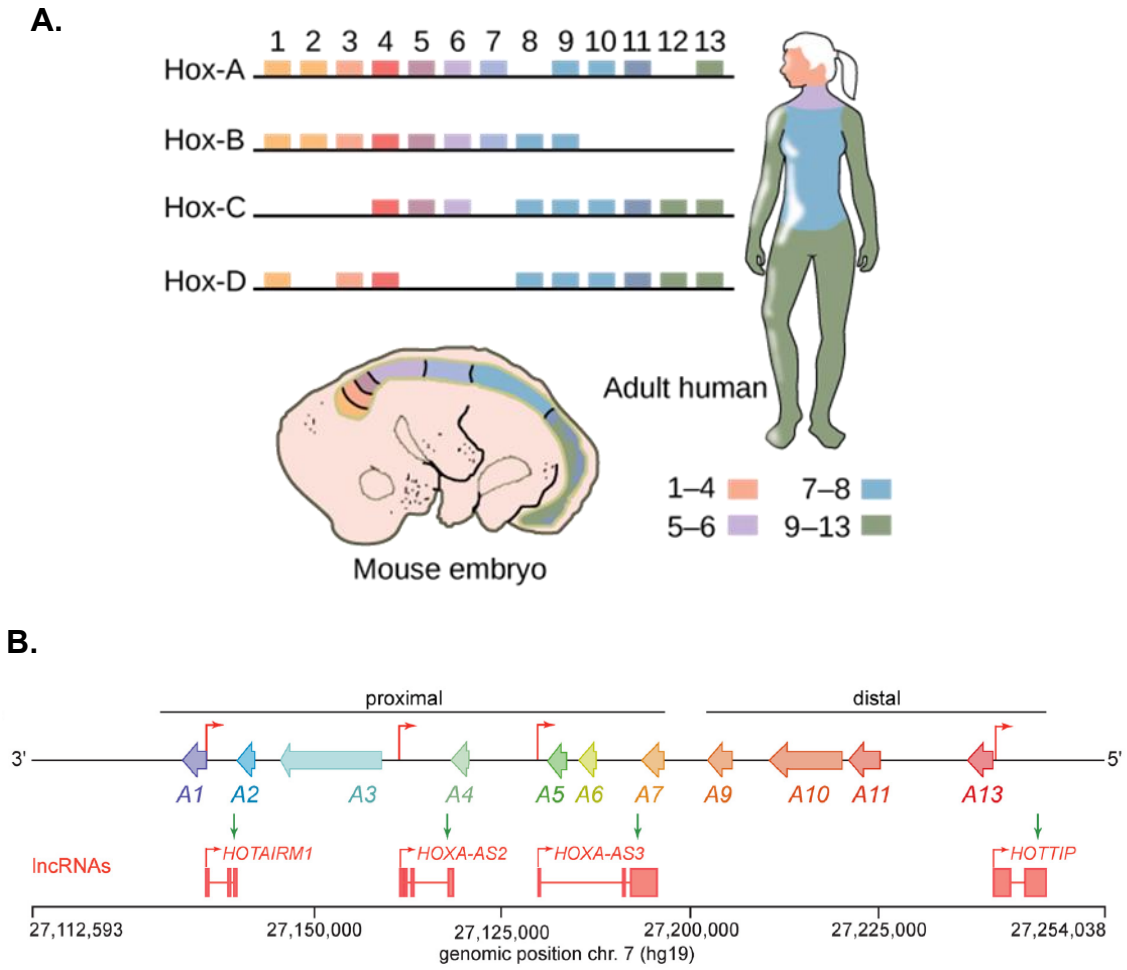


Figure 2. (A) The evolutionarily conserved *HOX* clusters (*HOXA-HOXD*) regulate anterior-to-posterior body plan in human and mouse embryo. Published by Openstax in Biology 2e, Chapter 27: Introduction to Animal Diversity. Retrieved on Aug. 18, 2022, from https://openstax.org/books/biology-2e/pages/27-1-features-of-the-animal-kingdom. **(B)** Gene map of *HOXA* gene cluster together with the lncRNAs along the human chromosome 7. Published by Wang et al. (2017) in Nucleic Acids Research (DOI: 10.1093/nar/gkw966). Reprinted with blanket permission from publisher.

Regulation of *HOXA* genes by *HOXA*-associated lncRNAs

The *HOXA* gene cluster plays a pivotal role in definitive hematopoiesis, and its precise regulation is essential for normal development. Alongside the 13 protein-coding *HOX* genes, this cluster also encodes several lncRNAs, transcribed antisense to the *HOXA* genes (Fig. 2). These lncRNAs have emerged as critical regulators of *HOXA* gene expression and are intimately involved in various developmental processes. One such lncRNA known as HOTTIP (*HOXA* Distal Transcript Antisense RNA) is transcribed from the 5' end of the *HOXA* gene cluster. In 2011, Wang et al. unveiled its role in coordinating the expression of 5' *HOXA* genes by serving as a scaffold [26]. HOTTIP is thought to act as a scaffold to adaptor protein WDR5 and a methyltransferase complex, influencing an existing chromatin loop that brings target genes into contact with each other. This activity then leads to deposition of H3K4me3, a histone mark associated with open and active chromatin. Consequently, the target genes become more accessible to the transcriptional machinery. Accordingly, knockdown of HOTTIP with siRNAs abrogated the activation of *HOXA13* and other 5' genes in descending order [26].

HOTAIRM1 (*HOXA* Transcript Antisense RNA, Myeloid-Specific 1) is another *HOXA* lncRNA positioned at the 3' of the cluster, between the *HOXA1* and *HOXA2* genes. Its expression was first reported in human myeloid cell lineages [27]. Since it shares a promoter with the *HOXA1* gene, it is the first lncRNA expressed at the *HOXA* cluster upon RA-induced differentiation. HOTAIRM1 has been found expressed in both nucleus and cytoplasm, suggesting its involvement in various regulatory processes [5]. In NB4 cells, HOTAIRM1 knockdown quantitatively reduced the expression of *HOXA1* and *HOXA4* during myeloid differentiation, suggesting its role as an activator for proximal

HOXA genes [28]. In contrast, in NT2-D1 cells, Wang et al. demonstrated that HOTAIRM1 can exert opposite regulatory effects on its target genes, depending on the expressed transcript variants [5]. Before RA-induced differentiation, chromatin looping between the *HOXA1/2* and *HOXA5/6/7* genes contributes to the silencing of *HOXA* genes. Upon activation, the accumulation of unspliced HOTAIRM1 is thought to recruit MLL2 to upregulate the expression of *HOXA1-2* through H3K4me3 modification. Simultaneously, the interactions between *HOXA1/2* and *HOXA5/6/7* are disrupted to prevent premature expression of downstream genes. Such loop disruption occurs while spliced HOTAIRM1 binds to PRC2 and trimethylates histone H3 on lysine 27 to repress *HOXA3-7* gene activation. These dynamic interactions together help maintain the spatiotemporal collinearity of the *HOXA* cluster during development [5].

The microRNAs

MicroRNAs (miRNAs) are small RNAs, typically 20~22 nucleotides in length, constituting another class of non-coding RNA widely expressed in the genome. They have a significant influence on gene expression through post-transcriptional gene regulation.

Most of miRNAs are derived from non-protein coding RNA transcripts and their biogenesis can be classified into two pathways.

The main pathway of miRNA production begins with the synthesis of primary miRNAs (pri-miRNAs) by RNA polymerase II. These pri-miRNAs undergo further processing in the nucleus, facilitated by the type III RNAse DROSHA and DGCR8 proteins, resulting in the generation of pre-miRNAs (Fig. 3) [29]. A complex comprising EXPORTIN-5 and RAN-GTP then facilitates nuclear export of pre-miRNAs into the

cytoplasm, where the maturation can be completed. Final miRNAs are cleaved and generated by a different RNase III nuclease called DICER. Single-stranded miRNAs subsequently loaded into Argonaute proteins (AGO) to form effector complexes called RNA-induced silencing complexes (RISCs), which regulate gene expression [30].

Recently, more non-canonical miRNA biogenesis pathways have been identified. These pathways produce miRNAs from sources other than canonical precursors or via different combinations of the proteins involved in the canonical process mainly DROSHA, DICER, EXPORTIN-5, and AGO2 [31, 32] .

Within RISC, the miRNA serves as a molecular guide by pairing with complementary sequences within the 3' untranslated regions (UTRs) of mRNAs. The degree of complementarity between the miRNA seed region (residues 2-7 at the 5’ end) and the target mRNA dictates the mechanism of mRNA silencing. A perfect match induces AGO2 endonuclease activity and leads to mRNA cleavage, effectively preventing its translation into protein. However, binding between miRNA and its target mRNA is not perfect most of the time, leading to the formation of bulge in the binding region. Such imperfect binding typically leads to mRNA decay independently from AGO2-mediated cleavage. This process initiates with the 3’ – 5’ deadenylating of the mRNA targets, facilitated by poly(A)-deadenylase complexes PAN2-PAN3 and CCR4-NOT [33, 34]. Subsequently, the 5’-3’ decapping starts with the help of decapping protein (DCP1). This process exposes the 5’ end of the mRNAs to degradation by the cytoplasmic exoribonuclease 1 (XRN1), which is recruited to the mRNA target through DCP1 [35, 36].

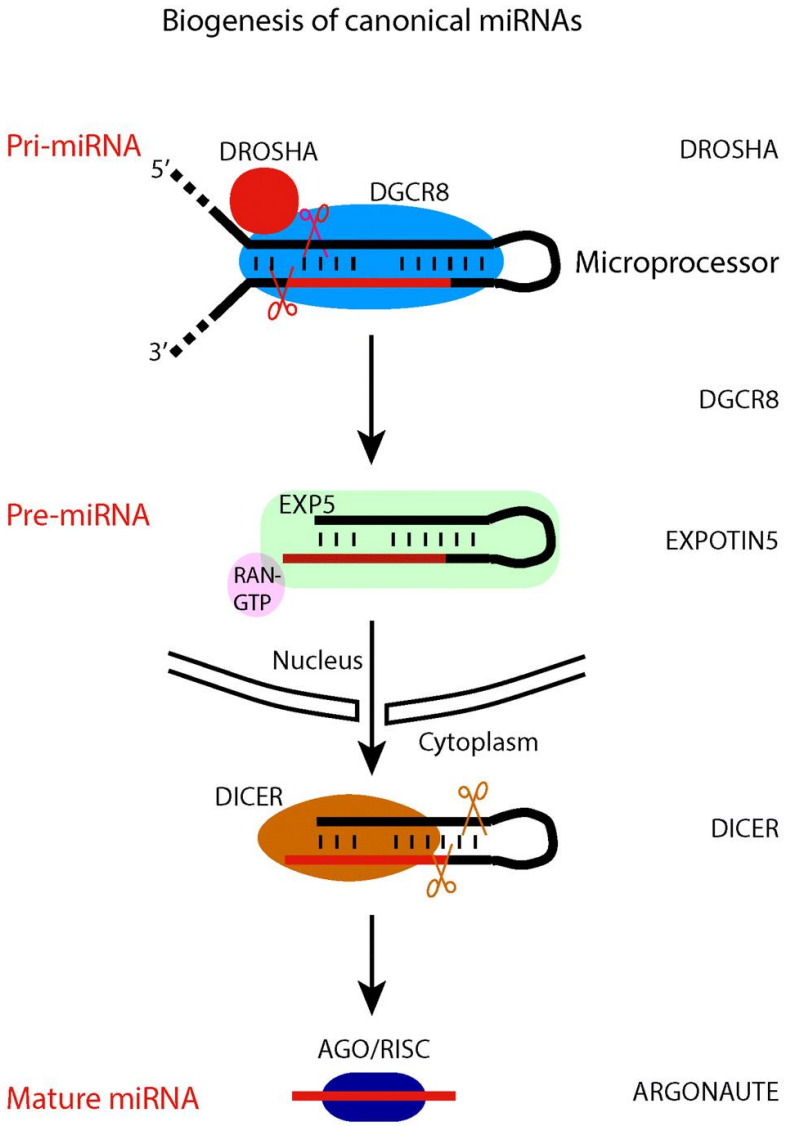


Figure 3. Biogenesis pathway of canonical miRNAs. Published by Guo, W.-T. and Y. Wang (2019) in Cellular and Molecular Life Sciences (DOI:10.1007/s00018-019-03020-9). Reprinted with blanket permission from publisher.

Current project

Beyond its role in the regulation of *HOXA* genes, HOTAIRM1 has emerged as a potentially critical player in many other biological processes, especially in cancers where its expression pattern is altered and is being explored as a potential diagnostic biomarker across diverse carcinoma subtypes.

Functioning as an oncogene, heightened HOTAIRM1 expression has been linked to tumor progression in conditions including glioma [37], endometrial cancer [38] and non-small cell lung cancer [39]. Conversely, reduced levels of HOTAIRM1 have been observed in colorectal cancer, ovarian cancer hepatocellular carcinoma, and other malignancies [40-42]. Many studies suggest that HOTAIRM's dual role in distinct tumor contexts might be achieved through its activity as a ceRNA. For example, in the development of glioblastoma, over-expression of HOTAIRM1 promotes the expression of SP1 protein by sequestering miR-137, which results in cell proliferation and invasion [43]. In ovarian cancer, lowered HOTAIRM1 levels diminishes its capacity to sponge miR-106a-5p. This, in turn, leads to an inhibition on its target mRNA encoding the tumor suppressor ARHGAP24 [41].

While the role of HOTAIRM1 in tumorigenesis has been more widely explored, our research has unveiled its less characterized function as an essential component of the early stages of neuronal differentiation [4]. Our investigations revealed that HOTAIRM1 interacts with the *HOXA1* TF as a ribonucleoprotein (RNP) complex and modulates the expression of a core pluripotent network during neuronal differentiation. Although the mechanisms by which HOTAIRM1 regulates these genes remain to be identified, one possibility lies in its potential to act as a ceRNA, interact with miRNAs,

and play a critical role in modulating this pluripotency networks during neuronal differentiation. In support of this hypothesis, we and others have detected HOTAIRM1 expression in AGO2 pulldown samples (Fig. 4) [44]. Given HOTAIRM1's established role as a ceRNA in other contexts [44-46], it is possible that a similar regulatory mechanism may be at play during neuronal differentiation. Accordingly, a webtool that predicts miRNA binding sites on lncRNAs also revealed 56 miRNAs with binding potential to HOTAIRM1 [47]. This list includes well-known members of the let-7 miRNA family, which have pivotal roles in promoting cellular differentiation [48].

To answer our research question, we decided to characterize the miRNA interactome of HOTAIRM1 in a neuronal differentiation model. To identify endogenous miRNAs interacting with HOTAIRM1 during differentiation, we used the CRISPR/Cas9 editing technique to insert a tethering sequence at the 3' end of HOTAIRM1. This knock-in tethering cell system allowed us to perform native RNA Immunoprecipitation (RIP) to capture the intricate interactions between HOTAIRM1 and miRNAs. After full characterization of our new cell lines, we conducted small-RNA sequencing of RIP samples upon RA-induced differentiation to identify the miRNAs that bind HOTAIRM1 early in neurogenesis. Given the technical challenges associated with genetic modification of NT2-D1 cells – a well-studied neuronal differentiation model – we opted to conduct our analysis in NCCIT, which is another huma cell differentiation model. NCCIT cells exhibit developmental pluripotency and have the capacity to differentiate into derivatives of all three embryonic germ layers upon RA induction [49]. Importantly, previous studies show that it recapitulates sequential *HOX* gene activation upon RA-treatment as seen in NT2-D1 cells [50].

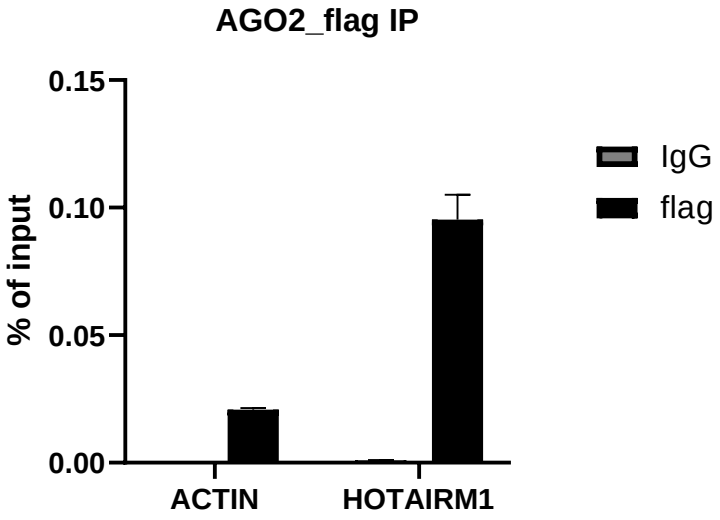


Figure 4. HOTAIRM1 co-immunoprecipitates with Flag-AGO2 when transiently transfected in HEK293T cells. RIP was performed using Flag antibody. Actin was probed as a negative control, and IgG served as a pulldown control. Error bars represent the standard deviation derived from technical triplicates of qPCR. Data kindly provided by Dana Segal (unpublished).

Overview of the λN boxB tethering system

The RNA-protein tethering system consists of boxB RNA stem loops and the N-terminal domain of the bacteriophage λN protein. This established system is renowned for its versatility in probing gene function. By incorporating the λN as a tag onto a protein of interest, such as a transcriptional activator or repressor, researchers can probe the outcome of altered gene expression when the effector is tethered to RNA through boxB interactions [51].

The strong affinity between the boxB stem loop and λN peptide (Kd: 200 ± 56 pM) has attracted the use of this system to investigate RNA-protein interactions and

studying RNA localization. RaPID (RNA–Protein Interaction Detection) is an approach developed by Ramanathan et al.[52], to identify and study RNA–protein interactions in living cells. This technique involves fusing the λN protein with a modified biotin ligase BirA, facilitating the biotinylation of proximal interacting proteins. By strategically placing boxB stem loops adjacent to the RNA of interest, the λN-BirA complex is recruited to the neighboring RNA motif, leading to the biotinylation of proteins in close proximity. Following streptavidin capture and mass spectrometry analysis, RaPID enabled the profiling of proteomes associated with specific RNAs. It is important to note, however, that this approach faces a limitation in distinguishing direct interacting proteins from those merely in the vicinity of λN-BirA.

In 2020, a research team used CRISPR/Cas9 editing to insert 24X MS2 stem loops — another widely employed tethering system — at the 3' end of the NEAT1 lncRNA. By leveraging its binding partner, MCP, they efficiently immunoprecipitated and generated a comprehensive profile of the proteome associated with the endogenous NEAT1 lncRNA [53].

Recognizing the effectiveness of the λN boxB system and its ability to uncover intricate RNA interactomes, we decided to CRISPR/Cas9 knock-in five (5x) boxB stem loops at the 3’ end of the HOTAIRM1 gene to enable its pulldown HOTAIRM1 along with its interactome including the miRNAs interacting with it.

Materials and methods

Cell culture

We received NCCIT cell line as a generous gift from Dr. J. Teodoro (McGill, ATCC; CRL-2073). They are human pluripotent carcinoma cells from a mediastinal mixed germ cell tumour in a Japanese male. NCCIT cells were cultured in Roswell Park Memorial Institute 1640 medium (RPMI; Gibco) supplemented with 10% fetal bovine serum (FBS, Thermo Fisher Scientific). Gene induction and cell differentiation was performed by adding 10 μM RA to the complete medium. All of the modified cell lines, as described in the result, were cultured under the same conditions. Cells were collected after 48 hours of incubation unless otherwise indicated. Cells were grown at 37°C in 5% CO_2 condition. HEK293T (ATCC; CRL-3216) cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Gibco) supplemented with 10% FBS.

Plasmids constructions and transient transfections

PC_GST_3xFM and PC_ λNGST_3XFM were generated by cloning PCR-amplified inserts into a pcDNA3.1 vector (Addgene) featuring a 3XFlag-myc tag sequence in the multiple cloning sites. Specifically, to build PC_GST_3xFM, forward (5'-CATCAGAATTCATGTCCCCTATACTAGGTTATTGG-3') and reverse (5'-CTTGGAATTCTTTTGGAGGATGGTCGCCACC3') PCR primers containing EcoRI restriction sites, were used to amplify GST sequence from a template plasmid encoding GST-Y14. To construct PC_λNGST_3XFM, an annealed duplex sequence of λN was synthesized by IDT, the sense strand (5' Phos-GATCCATGGACGCACAAACACGACGACGTGAGCGTCGCGCTGAGAAACAAGCTCA

ATGGAAAGCTGCAAACG-3'), the antisense strand (5'-/5Phos/AATTTTTGCAGCTTTCCATTGAGCTTTTTTCAGCGCGACGCTCACGTCGTCGTGTTTGTGCGTCCATG-3').

To construct PCNDA3_M1_B, boxB sequences were first amplified from pGL3-BoxB, a gift from Joshua Mendell (Addgene plasmid # 92004; http://n2t.net/addgene:92004 ; RRID:Addgene_92004) using primers forward (5'-CTAAGGGATCCCGCCGTGTAATTCTAGATTCCC -3') and reverse (5'-CACTGGATATCGTATCTTATCATGTCTGCTCGAAG -3') featuring BamHI and EcoRV restriction sites, respectively [54]. The boxB sequences were then flanked by homology arms. The sequences for both homology arms were generated by amplifying NCCIT genomic DNA, reverse-transcribed using oligo-dT, from NCCIT cells. For the upstream homology arm, we used forward primer (5'-GACTAGGTACCGAACCTGCGTTAATTTATAACC -3'), and reverse primer (5'-CGTACGGATCCAATGACTAGGGCTTCTAATAAGTG -3') featuring Acc65I and BamHI restriction sites, respectively. For the downstream homology arm, we used forward primer (5'-CAGATGATATCTTTTGGCTTTCTGTTTTGTTGCTGTCC -3'), and reverse primer (5'- CAGTCGCGGCCGCGCAGGACTTTAGCGTTTTCTCCAC -3') featuring EcoRV and NotI restriction sites, respectively.

For RIP experiments in NCCIT_M1_B (10 million) cells, PC_GST_3xFM or PC_λNGST_3XFM (12.5 µg) were transfected using the Lipofectamine™ STEM transfection reagent as recommended by the manufacturer (Invitrogen™). RIP samples in HEK293T cells were generated from 10 million cells transfected with 6.5 µg of NCCIT_M1_B and

6.5 µg PC_ λNGST_3XFM using Lipofectamine™ 2000 transfection reagent (Invitrogen™).

Insertion of boxB sequences into the endogenous HOTAIRM1 gene by CRISPR/Cas9 editing

HOTAIRM1 guide oligos (top: 5' - CACCTAGAAGCCCTAGTCATTTTT− 3'; bottom: 5' - AAACAAAAATGACTAGGGCTTCTA− 3') were annealed and cloned into lentiCRISPRv2, a gift from Feng Zhang (Addgene plasmid #52961; http://n2t.net/addgene:52961; RRID: Addgene_52961) as per their recommendations [55]. This plasmid was co-transfected with Homology-directed repair (HDR) Donor: NCCIT_M1_B into NCCIT cells using Lipofectamine™ STEM. Cells were treated with puromycin (2 µg/mL) for 3 days to select transfected cells, then plated at approximately 2000 cells/100 mm plate. After 7 days of growth, single colonies were picked and plated into 96-well plates, then expanded. 146 single colonies (2 million cells/colony) were harvested and resuspended in DNAzol® to extract total genomic DNA as recommended (Invitrogen™). Genomic PCR with primers flanking the region (genomic_M1_fwd: 5' – GACGCCCTGCTCCGCGCTGAGCTTGG- 3'; down_5'_rev: 5' - GATTCCTGAGAGCAGTATGAGGC - 3') confirmed the insertion of the 5x boxB sequence at the 3' end of HOTAIRM1. Sanger sequencing confirmed the insertion of 5x boxB sequences (Supplementary Material file).

RNA extraction and quantitative real-time polymerase chain reaction (RT-qPCR)

Total RNA from cells was extracted with the TRIzol® reagent as recommended by the manufacturer (Invitrogen™). For whole cell extraction, the resulting RNA pellets were resuspended in RNAse-free water and quantified using a NanoDrop™ 2000 Spectrophotometer (Thermo Scientific™). 10 µg of total RNA was treated with DNaseI for 13 minutes at 37°C to remove DNA contamination (New England Biolabs Ltd.). DNaseI-treated total RNA was re-extracted with TRIzol®, resupended in RNAse-free water, and re-quantified by NanoDrop™. 1 µg of each DNase-treated total RNA sample was reverse transcribed using SuperScript™ III and oligo(dT)$_{12–18}$ (Invitrogen™) to measure steady-state gene expression with RT-qPCR. Negative controls for RT-qPCR were prepared by omitting SuperScript™ III. The $\Delta\Delta C_t$ method was used to normalize expression values relative to Phosphoglycerate Kinase 1 (*PGK1*) [56]. The sequences of primers used for RT-qPCR are listed found in Supplementary Material file.

Western blotting

Whole cell extracts from modified cells and lysates from RIP assays were denatured for 5 minutes at 95°C in SDS sample buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 7.5% glycerol, 5% β-mercaptoethanol, 0.04% bromophenol blue) and chilled on ice. Protein samples from whole cell extract (1.5 million cells) or 1% of RIP were resolved by SDS-PAGE and transferred onto a 0.45 µm nitrocellulose membrane using a Hoefer TE77X semi-dry transfer unit run at 20 V for 1 hour. 10% methanol was used as transfer buffer. Membranes were blocked in 5% skim milk in PBST (PBS containing 0.1% Tween-20) for 2 hours at room temperature. For RIP sample analysis, λN and GST expression were

probed using anti-Flag antibody (1:1000, Sigma-Aldrich, F1804), α-eEF2 was probed (1:2000, New England Biolads Ltd., 2332S) as control. The whole cell extract from modified NCCIT cell lines were probed with antibodies against HOXA1 (1:700, Abcam, ab230513), NANOG (Abcam, ab109250), SOX2 (Abcam, ab92494), OCT4 (Abcam, ab181557), SOX3 (Abcam, ab183606), or eEF2 as loading control. Primary antibodies were diluted in 5% skim milk in PBST and incubated overnight at 4°C. The membranes were washed 3 times for 5 minutes in PBST before incubating with horseradish peroxidase conjugated AffiniPure goat anti-rabbit IgG (cat.# 111-035-003) or rabbit anti-mouse IgG (cat.# 315-035-003) secondary antibodies (Jackson ImmunoResearch Laboratories Inc.) for 2 hours at room temperature. Membranes were washed 3 times for 10 minutes in PBST before using chemiluminescence to reveal protein bands with the Western Gibco™ Plus-ECL Enhanced Chemiluminescence reagent (PerkinElmer Inc.).

RNA immunoprecipitation (RIP)

Cell pellets of transiently transfected, RA-induced NCCIT_M1_B was lysed in 1 mL lysis buffer (20 mM HEPES pH 7.5, 200 mM NaCl, 5 mM $MgCl_2$, 0.05% NP-40, 0.5 mM DTT, 1X Protease Inhibitor Cocktail (Sigma-Aldrich), 1 mM PMSF, and 0.1 U/mL RiboLock RNase Inhibitor (Thermo Scientific™). While actively growing NCCIT_M1_B_LN and NCCIT_M1_B_GST (control) were lysed in lysis buffer with 0.5% NP-40 and additional 0.1 µg/µl heparin. Samples were incubating on ice for 15 minutes, shearing through a 23 G needle (20 strokes), followed by sonicating on a Covaris® M220 Focused-

ultrasonicator (50W PIP; 10% Duty Factor; 200 Cycles per Burst; 7°C bath temperature; 9s time). Lysates were centrifuged at 15,000 rpm 15 minutes at 4°C to pellet debris.

For NCCIT_M1_B cells, each RIP used the equivalent of 5 million cells in a final volume of 500 µL, with 10% set aside as input (for western blotting or RT-qPCR). Cell lysates were incubated with 5 µg of antibodies against the Flag epitope (Sigma-Aldrich, F1804) or a control IgG (Abcam, ab12073) overnight at 4°C on an end-over-end rotor. Magnetic Protein G beads (Dynabeads™, Themo Scientific™, cat. no. 10003D) were used to pull-down RNP complexes and were washed with 500 µL cold washing buffer 3 minutes, 3 times. For NCCIT_M1_B_LN and NCCIT_M1_B_GST cells, each RIP used the equivalent of 25 million cells. Immunoprecipitation was performed through Anti-FLAG® M2 Magnetic Beads at 10 µg/sample (MilliporeSigma). Other conditions remained the same.

RNA was extracted using TRIzol® and treated with DNaseI (New England Biolabs, Ltd.) for 13 minutes at 37°C. Samples were re-extracted with TRIzol® and resuspended in 22 µl of RNAse-free water. The cDNA pool for total RNA was generated with SuperScript III and Random Hexamer $d(N)_6$ (Thermo Scientific™). The cDNA library for miRNAs was reverse transcribed using the miRCURY LNA™ RT Kit (QIAGEN) followed the manufacturer's instructions. All RNA levels were measured by qPCR using primers listed in Supplementary Material file.

Lentivirus production and stable cell line construction

Individual transgenes were cloned into pLX317 lentiviral vectors, kindly provided by Sidong Huang, originally from William Hahn (Addgene plasmid # 115446; http://n2t.net/addgene:115446 ; RRID:Addgene_115446) [57]. The λN sequence was amplified using the LN_fwd primer (5'- CATCGCTAGCGATCCATGGACGCACAAACAC -3') and LN_rev primer (5' GTACACTAGTTTAAAGATCTTCTTCAGAAATAAG -3') featuring BmtI and SpeI restriction sites, respectively. GST was cloned using the GST_fwd primer (5'- CTGAGCTAGCATGTCCCCTATACTAGGTTATTGG -3)', and GST_rev primer (5'- GTACACTAGTTTAAAGATCTTCTTCAGAAATAAG -3') with BmtI and SpeI restriction sites for ligation, respectively.

Lentiviral transduction was performed following the protocol from http://www.broadinstitute.org/rnai/public/resources/protocols. Briefly, 3×10^6 HEK293T cells were seeded in six-well plates with 2 mL DMEM medium per well and grown overnight. On the second day, cells were transfected with a lentiviral construct, the packaging (psPAX2) and envelope (pCMV-VSV-G) plasmids using the jetPRIME® reagent following the manufacture's instructions (Polyplus transfection). Virus-containing media was collected 24 hours after transfection and stored at −80 °C. Infected cells (~8 hours for infection and ~20 hours for recovery) were selected in medium containing puromycin (2 μg/mL) for 2–3 days and collected for cell line development.

Immunofluorescence (IF)

NCCIT_M1_B_GST, NCCIT_M1_B_LN, and NCCIT_LN cells were grown on glass coverslips for 2 days with or without RA treatment. Cells then were rinsed with PBS and fixed with 2% formaldehyde for 30 minutes. Permeabilization was performed in 0.5% Triton X-100 followed by 30 minutes blocking with 2% BSA. The cells were incubated with anti-flag antibody (1:2000, Sigma-Aldrich, F1804) in blocking buffer at RT for 45 mins, and then washed three times with PBS. Alexa Fluor 586 anti-mouse Red (Invitrogen™) was next incubated in blocking buffer for 30 mins in the dark. After 5 minutes, the coverslips were washed with PBS before incubating with DAPI for 5-10 mins. The cells were then washed again with PBS for 20 min before a final wash with distilled water for 10 minutes. Coverslips were then mounted and visualized at 100x using a ZEISS AxioCam MRm and its imaging software AxioVision Rel.4.8.

Results

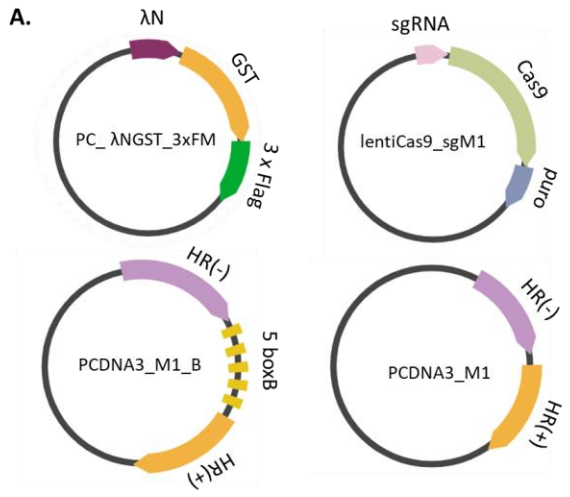
Aim 1. Establishment of NCCIT cell lines featuring an HOTAIRM1 tethering system using CRISPR/Cas9 editing.

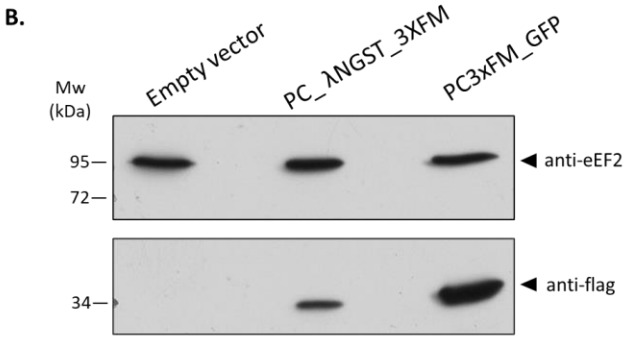
To establish the cell lines, we first cloned the λN peptide sequence into the PCDNA3.1 plasmid containing a 3XFlag-myc tag (Fig. 5A). However, due to the small size of the λN protein, I failed to detect its expression (6.88kDa) following the plasmid transfection in HEK293T cells (data not shown). To avoid problems associated with the peptide's small size, I inserted a glutathione S-transferase (GST) sequence between λN and flag tag (PC_ λNGST_3xFM). This GST tag increased the protein's size to 32.51 kDa, stabilizing it and facilitating its detection on western blot, in addition to possibly acting as an alternative purification tag in future experiments (Fig. 5A and B). A vector containing GST protein alone was also constructed (PC_GST_3xFM) as negative control for future experiments such as RIPs.

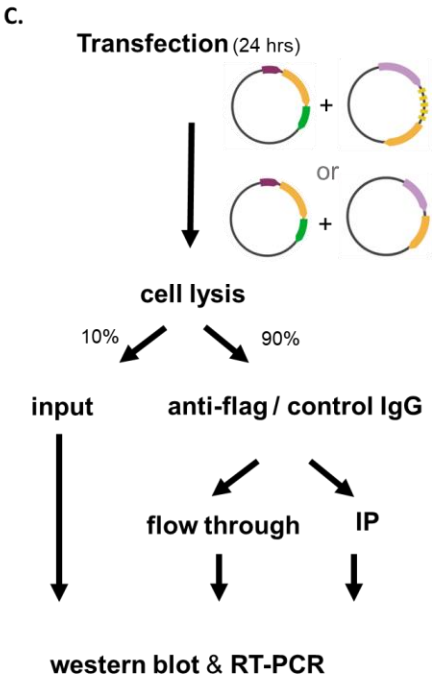
To introduce a double-stranded DNA break at the HOTAIRM1 3' end, we used the Cas9 enzyme. A single-guide RNA (sgRNA) sequence was thoughtfully designed to recognize 3' HOTAIRM1 sequences before the polyadenylation sequence motif as described in the Appendix (Appendix, Figure 1A) [58]. This sequence was then cloned into the lentiCRISPRv2 plasmid (Fig. 5A). Precise CRISPR/Cas9-mediated gene knock-in usually relies on a repair template with homology arms, the size of which varies from dozens of base pairs for smaller insertions, to several hundred for larger fragments. Given the 5x boxB (~290nt) sequence size, I constructed a plasmid (PCDNA3_M1_B) containing upstream (+) and downstream (-) homology arms, each spanning

approximately 700 nucleotides to flank the boxB repeats. This design aimed to increase the likelihood of a successful and precise insertion.

After cloning, binding of the λN boxB system was tested by transiently transfecting HEK293T cells with both PC_ λNGST_3xFM and PCDNA3_M1_B plasmids. A template only containing two homology arms (PCDNA_M1) served as control (Fig. 5A and C). Western blot analysis confirmed a success pulldown of λN_GST_3xFlag with the flag antibody in our RIP samples (Fig. 5D, *top*). Also, endpoint PCR validated the specific pulldown of boxB via the flag-fused λN, as compared to the control without boxB repeats (Fig. 5D, *bottom*).


A.
λN
GST
PC_ λNGST_3xFM
3 x Flag
sgRNA
Cas9
lentiCas9_sgM1
puro
HR(-)
PCDNA3_M1_B
5 boxB
HR(+)
HR(-)
PCDNA3_M1
HR(+)


B.
Mw
(kDa)
Empty vector
PC_ λNGST_3XFM
PC3xFM_GFP
95
72
anti-eEF2
34
anti-flag


C.
Transfection (24 hrs)
+
or
+
cell lysis
10%
90%
input
anti-flag / control IgG
flow through
IP
western blot & RT-PCR

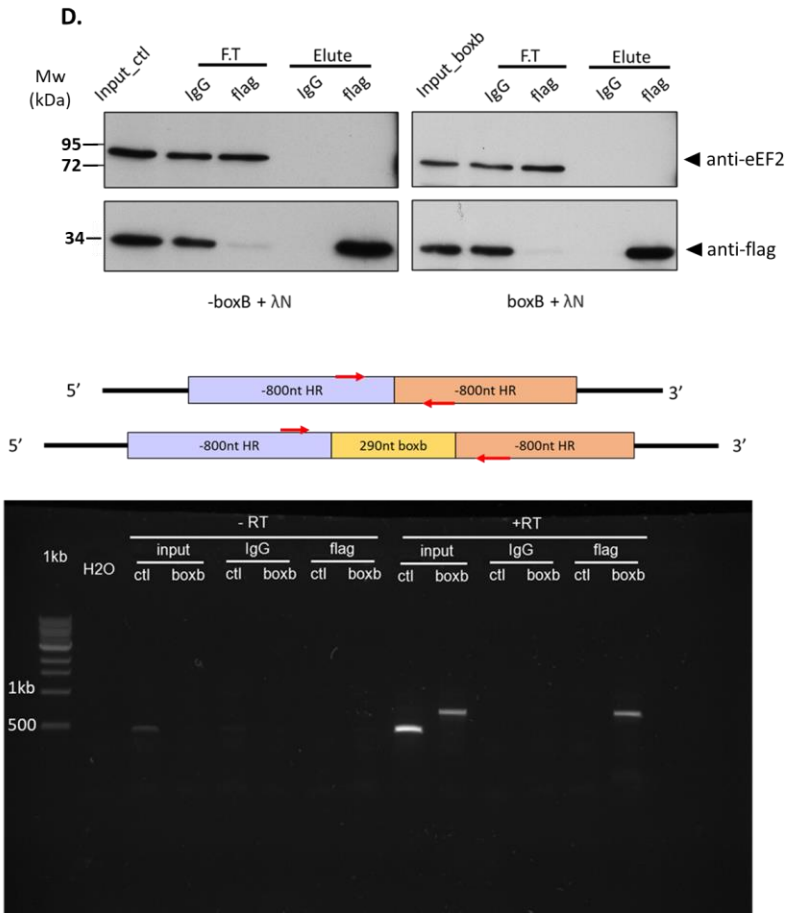
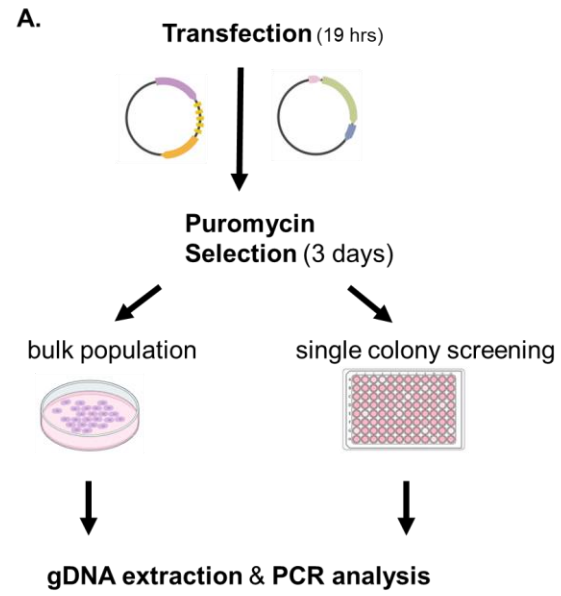

D.
Mw
(kDa)
Input_ctl
F.T
Elute
IgG
flag
IgG
flag
Input_boxb
F.T
Elute
IgG
flag
IgG
flag
95
72
anti-eEF2
34
anti-flag
-boxB + λN
boxB + λN
5′
-800nt HR
-800nt HR
3′
5′
-800nt HR
290nt boxb
-800nt HR
3′
- RT
+RT
1kb
H2O
input
IgG
flag
input
IgG
flag
ctl
boxb
ctl
boxb
ctl
boxb
ctl
boxb
ctl
boxb
ctl
boxb
1kb
500

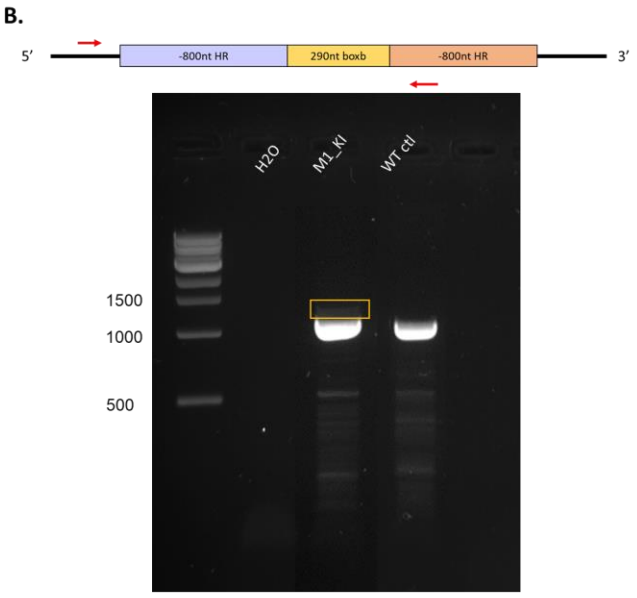
Figure 5. Validation of the boxB tethering system in transient transfection. **(A)** Schematic representation of cloned plasmids: λN is cloned in PCDNA3.1 with GST with a 3XFlag-myc tag at the C terminus (*top left*). Lentiviral expression vector for Cas9 enzyme and sgRNA of HOTAIRM1 with puromycin selection marker (*top right*). The repair template features two homology arms (+667 nt and -715 nt) flanking the 5 boxB sequence (*bottom left*). The control vector expresses sequences for both homology arms (*bottom right*). **(B)** Western blot showing λN_GST_3xFlag protein expression at 32.51 kDa, with eukaryotic translation elongation factor 2 (eEF2) as an internal control. **(C)** RIP Workflow: HEK293T cells were transfected with PC_λNGST_3xFM and PCDNA3_M1_B or PCDNA3_M1 (ctl.) for 24 hours, followed by cell lysis. 10% of the cell lysate served as input, while the rest was equally distributed for anti-flag immunoprecipitation or IgG control. Samples from input, flowthrough, and IP were analyzed by western blot and PCR. **(D)** Western blot showing the successful pulldown of λN_GST_3xFlag through the flag antibody, with no band in the IgG group, confirming pulldown specificity (*top*). PCR tests the efficiency of RIP from control and boxB group using protein G dyna beads. The red arrows indicate regions used for PCR (*bottom*).

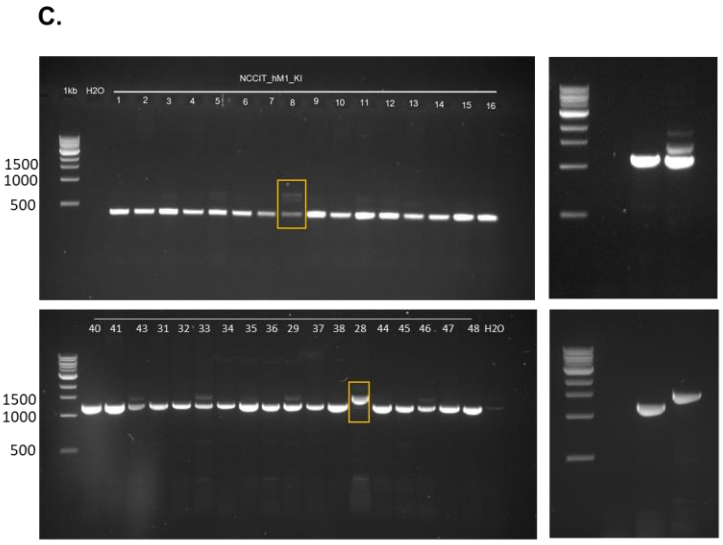
To engineer NCCIT cell lines where endogenous HOTAIRM1 can be tethered through the boxB system, we designed a workflow for efficient CRISPR/Cas9 knock-in of boxB repeats to the targeted gene site (Fig. 6A). Following transfection of the donor template and sgRNA construct, rounds of colony screening were performed. Genotyping PCR was first carried out with genomic PCR primers to confirm a successful KI of boxB repeats within a bulk population. A PCR product of the expected

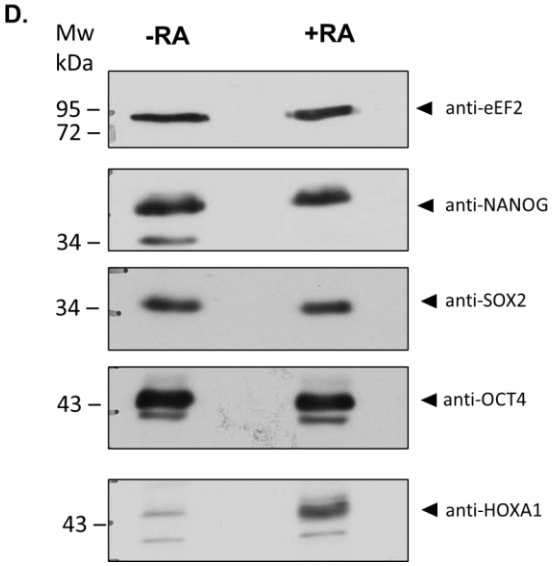
size was obtained, indicating that at least a subset of NCCIT cells had successfully undergone boxB insertion (Fig. 6B). Given this positive result, cells were seeded at low density for single colony isolation and transferred to 96-well-plates for single clone expansion. One of the 146 colonies analyzed showed successful homozygous insertion of boxB repeats (NCCIT_M1_B), while a second one showed heterozygous boxB insertion (NCCIT_M1_0.5B) (Fig. 6C). We confirmed the accuracy of boxB repeat sequences by Sanger sequencing. NCCIT_M1_B was chosen for further experiments. Since the PAM site of sgRNA was not mutated, we also ensured the absence of Cas9 expression to avoid any potential further cutting or other artifacts in our modified cell line (Appendix, Figure 1B).

To validate the integrity of the newly established cell line, we analyzed the expression of pluripotency marker genes, *SOX3* and *HOXA*1 during RA-induced differentiation. NCCIT_M1_B cells were treated with 10 μM RA for two days to reach an early differentiation state, and pluripotency markers were subsequently measured via western blotting and RT-qPCR. As expected, we found that both *NANOG* and *SOX2* displayed lower expression both at the protein and RNA levels upon RA induction of the modified cell line (Fig. 6D and E). The other component *OCT4*/*POU5F1* in this core pluripotency network remained relatively unchanged as previously reported in wild type NCCIT cells [4] (Fig. 6D and E). As anticipated, the *HOXA1* expression level increased after 2 days of differentiation in NCCIT_M1_B cells (Fig. 6D). Together, these results confirmed the maintenance of stem cell identity and validated the integrity of this CRISPR/Cas9-engineered NCCIT cell line.


A.
Transfection (19 hrs)
Puromycin
Selection (3 days)
bulk population
single colony screening
gDNA extraction & PCR analysis


B.
5'
-800nt HR
290nt boxb
-800nt HR
3'
H2O
M1_KI
WT ctl
1500
1000
500


C.
1500
1000
500


D.
Mw
kDa
-RA
+RA
95 –
72 –
anti-eEF2
34 –
anti-NANOG
34 –
anti-SOX2
43 –
anti-OCT4
43 –
anti-HOXA1

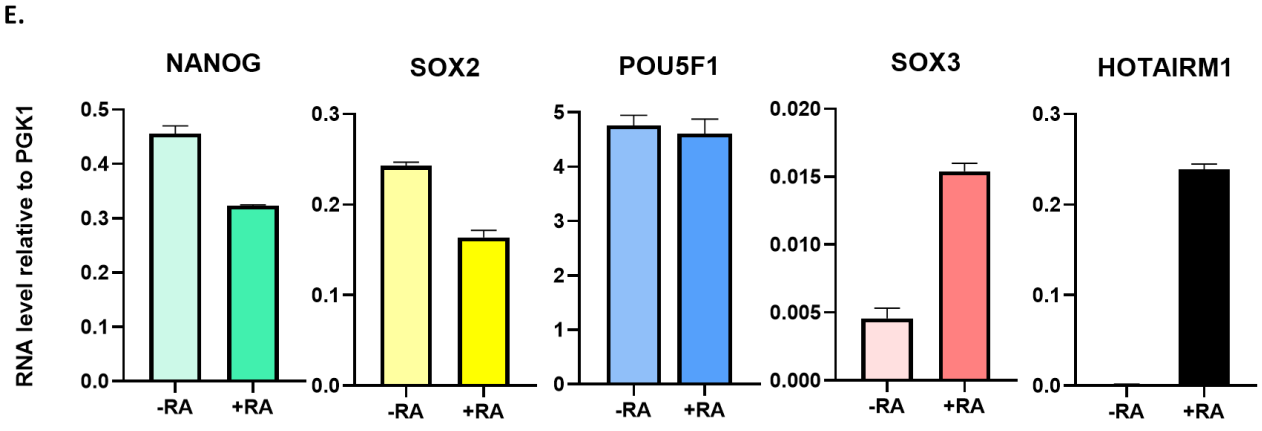

E.
NANOG
SOX2
POU5F1
SOX3
HOTAIRM1
RNA level relative to PGK1
-RA
+RA

Figure 6. Development of NCCIT cell line with tagged HOTAIRM1 using CRISPR KI**.** **(A)** Workflow of CRISPR/Cas9 KI of boxB repeats into NCCIT cells. Cells were transfected with sgRNA (lentiCas9_sgM1) and the repair template (PCDNA3_M1_B) for 19 hours, then underwent puromycin selection for 3 days. gDNA extraction and PCR analysis were first performed for both bulk population then for single cell colonies. **(B)** PCR result for KI condition using the bulk population. A genomic forward primer (red arrow) was used to avoid the amplification of residual plasmid signals from the cells. The yellow box (~1.4 kb) indicates a subset of NCCIT cells that successfully underwent boxB insertion. **(C)** PCR results of single colonies showing heterogeneous boxB insertion (NCCIT_M1_0.5B) (*top*) and homologous insertion of boxB repeats (NCCIT_M1_B) (*bottom*). **(D)** Western blot analysis of pluripotent factors and *HOXA1* protein levels in NCCIT_M1_B cells under RA-induced and non-induced conditions, eEF2 was probed as a control. **(E)** Corresponding RNA levels of each gene measured by RT-qPCR, with error bars indicating standard deviations from technical triplicates of qPCR (n=1).

Aim 2. Small RNA sequencing of the endogenous RNA pulldown.

2A. Investigate the pulldown efficiency of NCCIT_M1_B.

To further validate our NCCIT_M1_B cell line, RIP was conducted to evaluate the pulldown efficiency and specificity. NCCIT_M1_B cells were transfected with the protein-expressing vector PC_λNGST_3xFM for 6 hours. The early stage of differentiation was then induced by treating cells with 10μM RA for 48 hours. Several control conditions were included to distinguish between miRNAs specifically bound to

HOTAIRM1 and those interacting non-specifically: Wildtype NCCIT cells were transfected with PC_λNGST_3xFM to exclude miRNAs attracted by the fusion protein (WT+LN), while the KI cells were transfected with PC_GST_3xFM, a plasmid lacking the λN protein, to serve as a control for miRNAs that might non-specifically bind to the GST protein during the RIP process (M1_B+GST). For each of these conditions, 10 million cells were collected and lysed. A 10% fraction of the lysate was set aside as input, while the remaining lysate was equally distributed. Half of lysate were subjected to λN-HOTAIRM1-RNA complexes pulldown using flag antibody, and the other half underwent pulldown with IgG as a negative IP control. Following RNA extraction and DNaseI treatment, we quantified the pulldown efficiency of the target lncRNA HOTAIRM1 by RT-qPCR.

Western blot analysis revealed a successful transfection and protein expression in the cells based on signals observed in input samples while the IP indicates a robust affinity of the flag antibody for a specific pulldown of the flag-tagged λN and GST proteins (Fig. 7A). Meanwhile, the qPCR analysis shows that flag-fused λN specifically recognized boxB-tagged HOTAIRM1 (Fig. 7B), with a pulldown efficiency of approximately 4% relative to the input.

2B. Small RNA-sequencing.

After validating our cell line, we prepared RNA from input and IP samples from all three conditions for small RNA sequencing. To ensure that miRNAs remained intact during sample preparation, quality control and small-RNA isolation were performed at the Institute for Research in Immunology and Cancer (IRIC). Principal Component Analysis

(PCA) of our results revealed an unexpected pattern (Fig. 7C). The input samples, initially thought of as similar, exhibited substantial dispersion from each other. Also, IP samples from the KI cell line did not display the anticipated distinct clustering when compared to the control IP samples. These results led us to delve further in the data to identify potential underlying factors contributing to these differences.

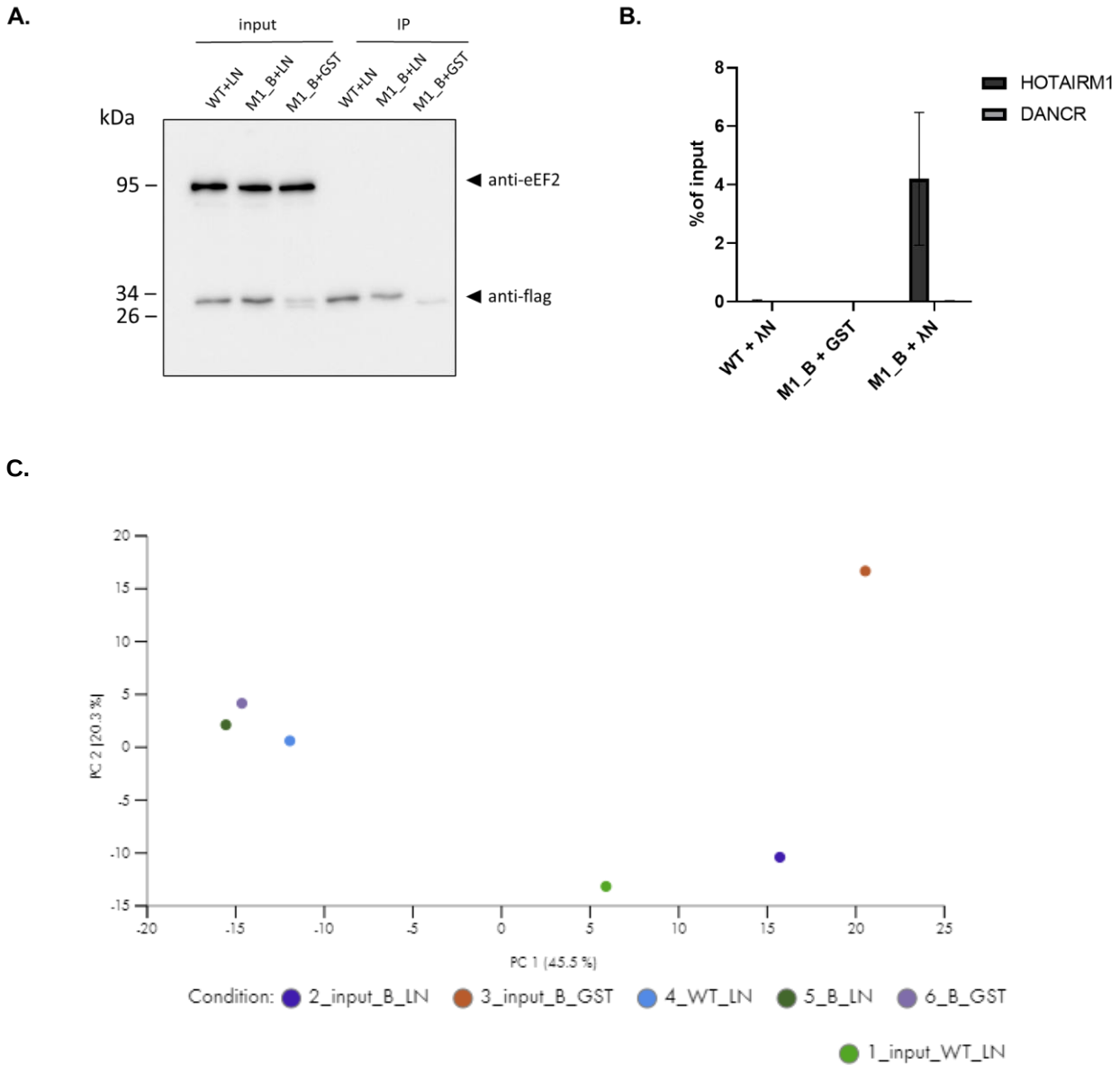


Figure 7. Validation and distribution patterns of immunoprecipitated components in RA-induced NCCIT_M1_B cells. **(A)** Western blot confirms specific immunoprecipitation of transfected PC_λNGST_3xFM and PC_GST in RA-induced NCCIT_M1_B cells. Input (1%) and RIP (5%) samples are from one biological replicate in (B), with eEF2 as a specificity control. **(B)** HOTAIRM1 is specifically pulled down in RA-induced NCCIT_M1_B cells through PC_λNGST_3xFM (M1_B+λN). Bar plots show the relative amount of immunoprecipitated HOTAIRM1 using flag antibody, with DANCR as an internal control. Values are averages from ≥2 biological replicates, and error bars indicate standard deviation between biological replicates. **(C)** Principal Component Analysis (PCA) plot reveals distribution patterns: inputs (1-3) exhibit dispersion, while the IP sample (5) from the KI cell line clusters with control IPs (4 and 6).

Aim 3. Optimization of the pulldown system.

While refining our RIP protocol and optimizing our system, we noticed an inconsistent pulldown efficiency possibly due to unstable protein expression. To address this issue and increase the pulldown efficiency, we decided to introduce a stable expression of the λN protein into our modified cell line. This modification serves a dual purpose: it ensures sustained and more uniform levels of protein expression within cells over extended durations in addition to increasing reproducibility between experiments by eliminating the need for recurrent transfections.

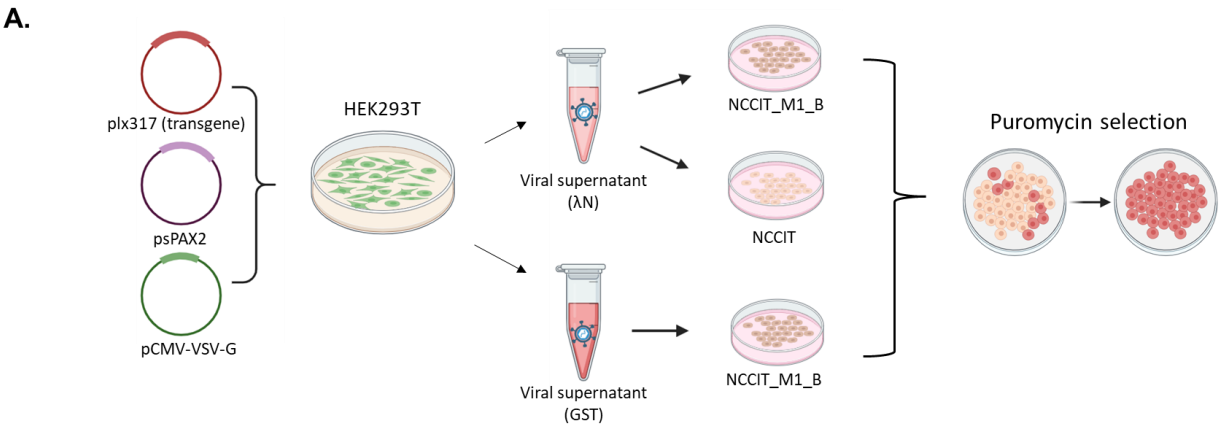
During the construction of lentiviral expression vectors and the subsequent generation of viral particles, we decided to first establish two distinct cell line conditions: NCCIT_M1_B cells transduced with lentivirus expressing λN_GST_flag, and a control

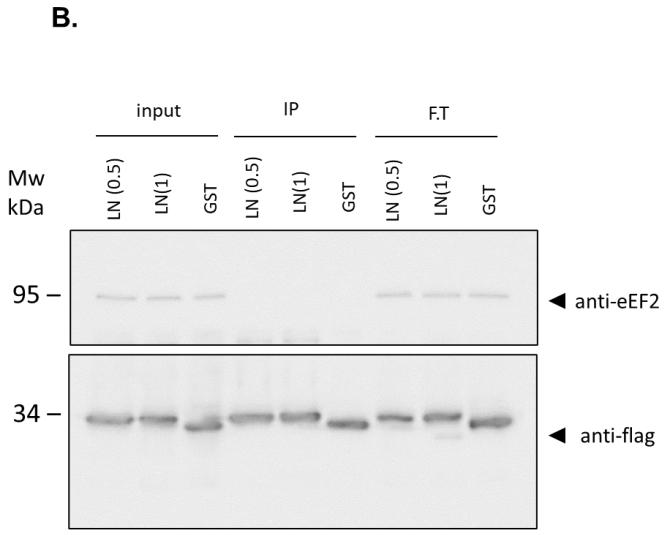
group receiving lentiviral particles expressing GST_flag. Notably, cells transduced with λN_GST_flag were subjected to transduction at two different Multiplicities of Infection (MOI) in order to maximize the number of cells featuring a single integration. Subsequent puromycin selection allowed the development of a stable cell population harboring integrated and expressed transgenes (Fig. 8A). Validation of successful gene integrations was confirmed through western blot analysis affirming the presence of the corresponding proteins (Appendix Figure 2).

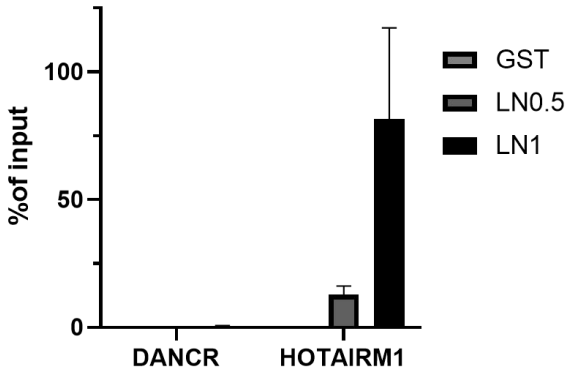
With our modified cells demonstrating the capacity for proper differentiation under RA induction (Appendix Figure 3C and D), we proceeded with the RIP experiment to assess whether the heightened λN levels would correspondingly amplify the pulldown efficiency of HOTAIRM1. Quantified through RT-qPCR, we observed that cells transduced with a low MOI (LN0.5) showed approximately 13% of the total HOTAIRM1 pulldown, whereas cells transduced with a high MOI (LN1) exhibited a significantly higher IP efficiency of around 80% (Fig. 8B). The absence of a signal from our control cell line, NCCIT_M1_B_GST, and the negative control lncRNA DANCR further confirmed the specificity of our system. Given the higher pulldown efficiency observed in the high MOI transduced cells, all subsequent experiments were conducted using these cells, referred to as NCCIT_M1_B_LN.

To further investigate where λN-HOTAIRM1-RNA complexes might localize in our modified cells, we performed immunofluorescence (IF) to observe the cellular localization of the fusion proteins. In each cell line, nuclei were stained by DAPI, while the λN or GST control proteins were stained through their flag tag using the anti-flag primary antibody and an Alexa Fluor 568 secondary antibody. Immunofluorescence

analysis revealed that the λN and GST control proteins were each expressed mainly in the cytoplasm (Fig. 8C). This result indicates that our system likely captures HOTAIRM1 from the cytoplasm, precisely where miRNAs predominantly reside. Such confirmation reinforces the relevance and accuracy of our approach in elucidating the miRNA interactions we seek to explore.


A.
plx317 (transgene)
psPAX2
pCMV-VSV-G
HEK293T
Viral supernatant
(λN)
Viral supernatant
(GST)
NCCIT_M1_B
NCCIT
NCCIT_M1_B
Puromycin selection


B.
input
IP
F.T
Mw
kDa
LN (0.5)
LN(1)
GST
LN (0.5)
LN(1)
GST
LN (0.5)
LN(1)
GST
95 –
34 –
anti-eEF2
anti-flag


%of input
100
50
0
GST
LN0.5
LN1
DANCR
HOTAIRM1

C.

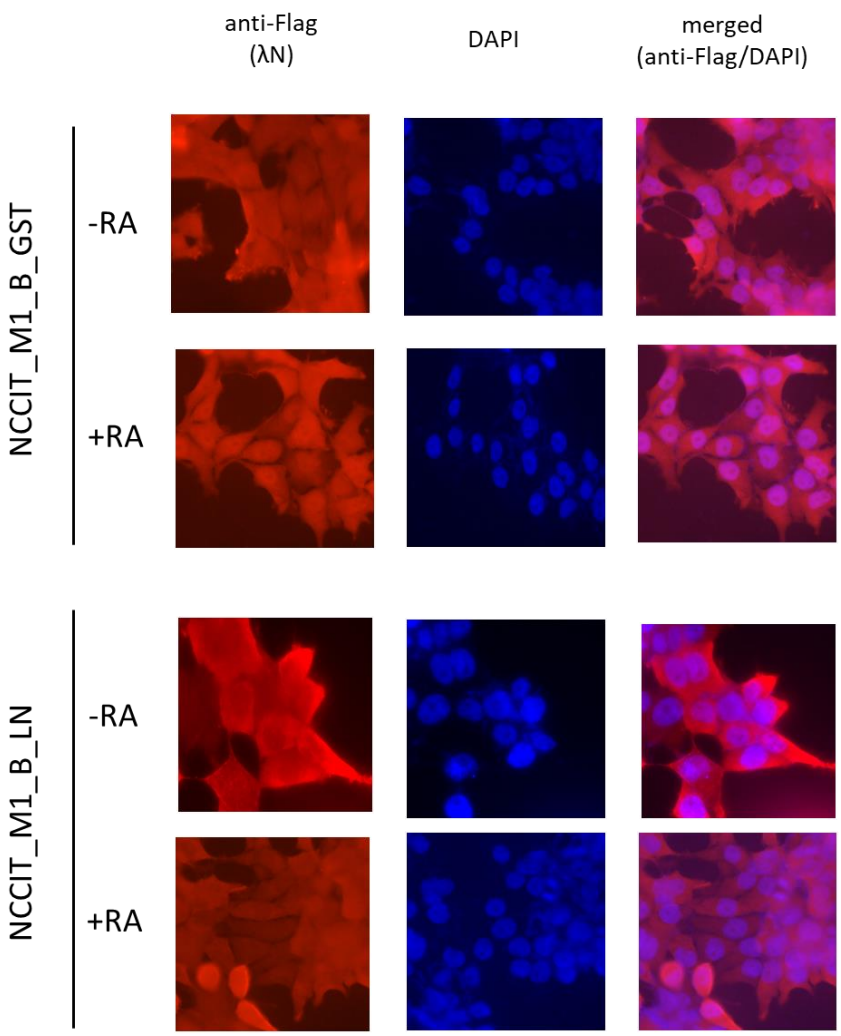


Figure 8. Development of transduced cell lines and subcellular localization analysis. **(A)** Lentiviral vector creation and transduction process. Lentiviral vectors are created by co-transfection of a packaging cell line (HEK293T) with the transgene plasmid and two packaging plasmids which encode the structural and envelope proteins. The packaging cells produce infectious particles, which were then used to transduce the target cells,

followed by puromycin selection at 2ug/mL. **(B)** Western blot confirmed the expression of λN_flag and GST_flag protein in both input and IP samples from transduced cell lines. eEF2 was a specificity control for flag IP (*top*). ~80% of HOTAIRM1 was pulled down in high MOI transduced cells, while around 20% was pulled down in low MOI transduced cells (*bottom*). N=1, error bars indicate standard deviation from technical triplicates. **(C)** Immunofluorescence analysis showing the subcellular localization of λN-HOTAIRM1-RNA complexes in transduced cells (NCCIT_M1_B_GST and NCCIT_M1_B_LN). Blue DAPI staining represents nuclei, and red staining depicts anti-Flag antibodies with Alexa Fluor 568 secondary antibodies.

3A. Evaluating the specificity of miR-4516 and miR-3960 interaction using NCCIT_M1_B_LN.

Although what can be concluded from our mir-sequencing results is rather limited given that more sequencing depth appears required, two miRNAs (hsa-miR-4516 and hsa-miR-3960) nonetheless caught our attention as they were noticeably increased in our KI IP group (Table 1). However, it is worth noting that the fold change for miR-4516 and miR-3960 in our KI IP condition were only 1.6 and 2.1 times higher than that in the WT IP condition, respectively. This prompted us to investigate whether these two miRNAs were genuinely associated with HOTAIRM1 or simply a non-specific binding event due to the presence of λN, which is an RNA binding protein.

We therefore decided to use our transduced cell lines to assess the expression of both miRNAs. To enhance pulldown specificity, this time we increased stringency of our washing buffers and added heparin to minimize non-specific interaction. We also

introduced non-RA treated cell conditions to explore whether miR-4516 and miR-3960 could be pulled down in the absence of HOTAIRM1 expression. The results obtained from miRNA RT-qPCR (Fig. 9B) revealed that both miRNAs are consistently expressed in our cell lines, with their levels remaining unaffected by RA-induced differentiation. Although low levels were still captured in our control IP groups, higher quantities of miR-4516 and miR-3960 was pulled down in non-differentiated NCCIT_M1_B_LN cells compared to the differentiated cellular state, which was inversely correlating with HOTAIRM1 expression. However, it's difficult to draw strong conclusions at this stage about whether these two miRNAs interact with HOTAIRM1 or simply engage in non-specific binding to λN. The outcome could also depend on the levels of miRNA expression within the cell.

Name	1_input_WT_LN	3_input_B_GST	2_input_B_LN	4_IP_WT_LN	6_IP_B_GST	5_IP_B_LN
hsa-miR-3960	107	162	217	570	40	**1208**
hsa-miR-4516	259	244	378	717	62	**1155**
hsa-miR-4488	23	30	55	13	1	**34**

Table 1. Top3 miRNAs enriched in the IP group from the small-RNA sequencing result. (Full dataset can be found in Supplementary Material file)

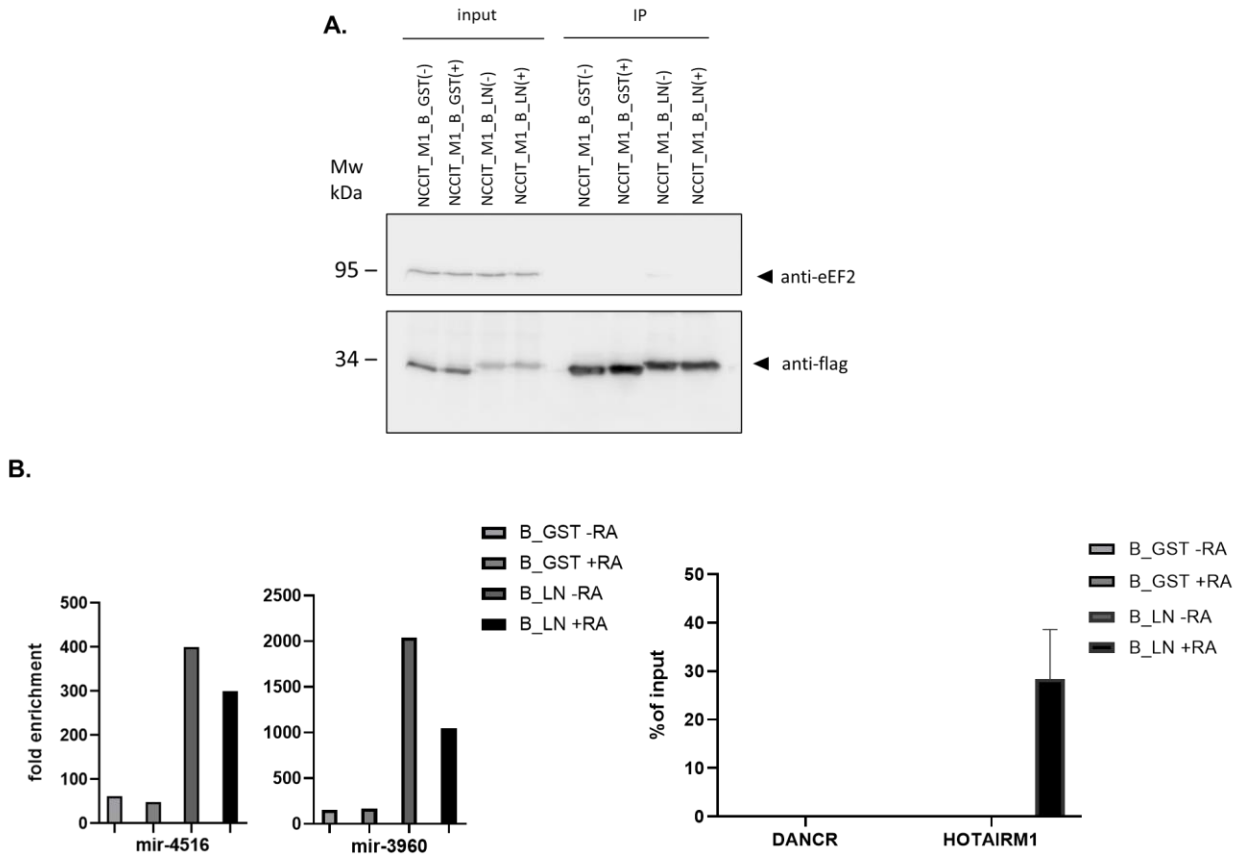


Figure 9. Evaluation of the binding specificity of miR-4516 and miR-3960. **(A)** Western blot showing a success pulldown through M2 flag beads with the transduced cell lines for both RA treated and non-treated conditions. **(B)** miRNA RT-qPCR results illustrating the pulldown efficiency of each miRNA under different conditions (*left*). Data are presented as fold changes relative to input and U6 snRNA, serving as an internal standard, and are compared to the expression with HOTAIRM1 (*right*). DANCR is utilized as a negative control for assessing pulldown specificity. Error bars indicate standard deviation from technical triplicates of qPCR.

Discussion

HOTAIRM1 is thought to participate in proper progression through the early stages of neuronal differentiation but the mechanism behind this phenomenon remains elusive. Here we aimed to explore the miRNA interactions with HOTAIRM1 during the neuronal differentiation process. Using human EC line NCCIT, we applied CRISPR/Cas9 KI technique to construct a cell line (NCCIT_M1_B) featuring the integration of five boxB aptamers at the 3' end of HOTAIRM1 to enable native HOTAIRM1 pulldown through the λN protein. Since the length of five (5x) boxB repeats is relatively short compared to other widely used aptamers, this allowed us to minimize the possible impacts on HOTAIRM1's three-dimensional folding and cell distribution during the differentiation process [53, 59, 60].

When validating the NCCIT_M1_B cell line, we measured the behavior of pluripotency markers upon RA-induced differentiation. We found that expression of *NANOG* and *SOX2* showed noticeable decrease after RA treatment at both protein and RNA level (Fig. 6D and E), which resembled what we saw in wildtype NCCIT (Appendix, Figure 2A) [4]. This phenomenon can be attributed to the rapid expression of HOTAIRM1 and *HOXA1* during ES differentiation as it has been established that HOTAIRM1 and the *HOXA1* TF participate in regulating expression of both human and mouse *NANOG* and *SOX2* [4, 61]. A third component in this core pluripotency network *OCT4*/*POU5F1* showed a different expression pattern (Fig. 6E). Previous findings from our lab suggest that *POU5F1* expression typically increases after the first day of RA induction and experiences a more substantial decrease after three days of induction (Appendix, Figure 3B). Therefore, the relatively mild change in *POU5F1* expression in

our modified cell line may indicate that it is in a transitional phase. Further experiments can be conducted to explore whether this anticipated trend holds true.

The unexpected findings from our PCA analysis (Fig. 7C) prompted us to closely examine our bioanalyzer data, which helped us identify several contributing factors. The significant dispersion observed in our input samples may be attributed to an insufficient number of reads. As we treated our first miRNA sequencing as a test run, we only had 1.5 M-2 M reads per sample instead of the recommended ~5 M reads [62]. While analyzing the electropherogram peaks (Appendix, Figure 4), we also noticed that most material was centered around 170 bp (with a total adaptor size of 160 bp), indicating potential fragmentation and miRNA loss due to RNA degradation [63]. On the other hand, the IP sample from the KI cell line did not exhibit a distinct clustering from the control IPs. This discrepancy might be due to a failure to retain an adequate quantity of HOTAIRM1 and the proportion of HOTAIRM1 that interact with miRNAs. The limited pulldown percentage (4%) could result in the loss of HOTAIRM1 molecules actively interacting with miRNAs, potentially leading to an underrepresentation of this interaction in our data [64]. Furthermore, the incubation and washing conditions may have not been ideal, causing the loss of low-abundance or weak interactors [65]. Non-specific bindings of miRNAs to HOTAIRM1 or λN during the RIP process are also possible.

To address these issues, we optimized our system by integrating stable λN expression into our KI cell line via viral vectors, leading to a heightened pulldown efficiency. While cells transduced with high MOI exhibited a significantly high pulldown efficiency of HOTAIRM1 at 81% (Fig. 8B), we encountered a subsequent decrease to 28% in our second RIP experiment (Fig. 9B). This decrease may be attributed to

several factors. First, our use of freshly transduced cells in the initial IP was expected to lead to more efficient λN protein expression post-puromycin selection. In the subsequent RIP, the freeze-thaw process and cell division may have resulted in partly λN silencing due to its exogenous nature. Though we expressed λN under the control of the EF-1a promoter, which is a widely recognized robust expression driver in lentiviral gene delivery systems for mammalian ES cells [66-68], recent studies have raised concerns about the susceptibility of the EF-1a promoter to silencing in transgenic animals due to hypermethylation issues [69, 70]. On the other hand, since we only performed this IP two times, this can simply be a variation between biological replicates. However, it is important to mention that both IP efficiencies are within the expected range and are consistent with findings from other studies. Chu et al. demonstrated that over 60 % of the target ncRNA could be retrieved using biotinylated antisense oligos (ChIRP), while Chen et al. reported an elution efficiency of 20 % for MS2-tagged lncRNA through MS2 binding partner [53, 71].

We also verified the subcellular location of integrated λN protein through IF (Fig. 8C). This experiment confirmed the expression of λN in the cytoplasm, which validated our potential ability to efficiently pulldown HOTAIRM1 from the cytoplasm during the RIP procedure.

To decrease the non-specific binding of miRNAs during the RIP process, we increased stringency of our washing buffer and added heparin, which mimics the high negative charge of nucleic acids, acting as an RNA competitor [72].

From our miRNA sequencing results, we observed significant enrichments of miR-4516 and miR-3960 in our IP group, underscoring their potential as binding

partners for HOTAIRM1. While miR-4516 has been more involved in cancers, miR-3960 showed more relevant role in cell differentiation and *HOXA* gene regulation.

One of the cellular targets of miR-3960 is CERS1 (ceramide synthase 1) [73], which is an enzyme involved in ceramide synthesis. The expression of CERS1 was found to influence the differentiation of both neurons and stem cells, especially in mice [74-76].

Study by Xin et al. has suggested a close interaction between HOTAIRM1 and miR-3960. This interaction results in the downregulation of both HOTAIRM1 and HOXA1 at the mRNA level, influencing dendritic differentiation [46]. Therefore, one potential mechanism can be miR-3960 exerting a regulatory influence on HOTAIRM1 before RA-induced differentiation begins. At this stage, miR-3960 expression levels may be relatively high, suppressing HOTAIRM1. With the introduction of RA, the inhibitory effect on HOTAIRM1 could gradually diminish. However, miR-3960 may still interact with HOTAIRM1 during the early stages of neuronal differentiation as observed in the 2-day RA induction phase, preceding HOTAIRM1's expression peak on the 5th day (Fig. 9B, Appendix, Figure 3B).

Another layer of complexity unfolds as we draw insights from the study conducted by Hu et al. Their research identified the role of miR-3960 in a regulatory feedback loop during mouse osteoblast differentiation, where it directly interacts with *Hoxa2* and inhibits bone formation by repressing Runx2 expression [77]. Given the high degree of sequence similarity between *HOXA2* in humans and mice, it's possible that miR-3960 also regulate human *HOXA2* expression. As demonstrated by Wang et al., HOTAIRM1 regulates the temporal collinear activation of *HOXA* genes, including

HOXA2 [5]. Based on these previous findings, it is possible that HOTAIRM1 acts as a ceRNA and titrate miR-3960 away from *HOXA2,* thus releasing its inhibitory function on *HOXA2* and leading to its expression. It's important to note that substantial *HOXA*2 expression doesn't occur until day7 (Appendix, Figure 5).

As both potential mechanisms can explain our results in Fig. 8B which showed an inverse relationship between miR-3960 and HOTAIRM1 expression after 2 days of differentiation. The intricate interactions among HOTAIRM1/*HOXA1*, miR-3960, and *HOXA2* need to be explored further through techniques such as miRNA inhibition, overexpression in an extended time course study.

Limitations & future directions

During the RIP process, we employed both shearing and sonication to break down cellular structures and release cellular contents, allowing access to RNA molecules and protein complexes. It is essential to acknowledge that these procedures might inadvertently introduce fragmentation to HOTAIRM1, potentially leading to the loss of specific miRNA interactions at particular regions. Though we quantified the total HOTAIRM1 expression in our IP samples using qPCR with primer pairs targeting the third exon of HOTAIRM1, this analysis did not provide insights into the integrity of the HOTAIRM1 transcript. To address this, future experiments such as Northern blotting or qPCR can be conducted to assess different regions of HOTAIRM1, ensuring that we have captured the complete microRNAome of HOTAIRM1. The shearing step could also be reevaluated for potential omission, which may decrease the risk of unwanted fragmentation on lncRNAs.

In this study, we strategically introduced five boxB aptamers at the 3' end of HOTAIRM1 in order to maintain the lncRNA's natural structure and function while minimizing disruptions to its transcriptional regulation and initiation. However, it's important to acknowledge that miRNA regulatory elements tend to cluster toward the middle regions and 3' ends of lncRNA transcripts [78]. As a result, there is a possibility that our approach inadvertently interfered with a few miRNA interactions with HOTAIRM1 due to its insertion point. Within the context of this study, the choice of KI site remains the most suitable one as it helps better maintain the biological characteristics of HOTAIRM1. But future studies are essential to thoroughly explore the potential implications of this limitation and its effects on the overall miRNA interaction network of HOTAIRM1.

While our initial miRNA sequencing results offered valuable insight and direction, the limitations of a small sample size and inadequate read depth necessitate caution regarding its reliability. With the aid of our refined cell model (NCCIT_M1_B_LN), I anticipate a more accurate and robust dataset from the second mir-sequencing results. I also constructed a WT NCCIT cell line expressing the λN protein (NCCIT_LN) to better control our result (Fig. 8A; Appendix, Figure 6). We hope the updated results will contribute to a deeper understanding of the endogenous microRNA species intricately linked to HOTAIRM1 during neuronal differentiation.

Upon obtaining the enhanced miRNA sequencing results, one potential avenue of research would involve the utilization of antagomirs, enabling the exploration of specific miRNA targets that involve in regulating the differentiation process and gene expression [79]. Intriguingly, I am also curious whether miR-4488 will appear in our IP

sample this time, as it exhibited a minor increase in our initial sequencing, though much lower reads compared with miR-4516 and miR-3960 (Table 1).

As part of our ongoing efforts to refine our result, the introduction of cross-linking stands out as another method to enhance the precision of our investigations [80]. One promising candidate for in vivo cross-linking is 4'-aminomethyltrioxalen (AMT), a psoralen-derivative crosslinker. AMT is particularly attractive for our purposes because it creates inter-strand crosslinks between uridine bases in RNA while remaining unreactive towards proteins [81]. This approach holds the potential to shed light on the miRNAs with low binding affinities or transient associations to HOTAIRM1.

In addition to exploring the miRNA network, the development of our modified cell line, NCCIT_M1_B_LN, also provides a platform to explore HOTAIRM1's functions and regulatory mechanisms beyond the scope of our study. It also introduces a fresh perspective on how we can decipher the functions and roles of lncRNAs, broadening our understanding of their intricate involvement in cellular processes.

Conclusion

Our work provide a poweful cellular tool for efficient pulldown of HOTAIRM1 from its intricate cellular surroundings, achieved through strategic tagging of HOTAIRM1 with boxB aptamers. We validated the RA-inducible behavior of our modified NCCIT cell lines by comparing the expression of pluripotent markers and HOTAIRM1 levels with those of WT NCCIT cells. This supports the validity of our engineered cell lines for in-depth exploration of HOTAIRM1's regulatory role in neuronal differentiation. Furthermore, confirmation of the subcellular localization of the integrated λN protein within the cytoplasm reinforces that our system is appropriate to unveil HOTAIRM1's role as a ceRNA interacting with miRNAs. Despite the recognized limitations of our initial miRNA sequencing, we have seen several intriguing miRNAs displaying potential for binding to HOTAIRM1.

Looking ahead, our study has laid a strong foundation for future investigations into the miRNA network of HOTAIRM1. Our modified cell systems also hold great promise for revealing additional facets of HOTAIRM1's functions across various biological contexts.

References

1. Ng, S.Y., R. Johnson, and L.W. Stanton, *Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors.* Embo j, 2012. **31**(3): p. 522-33.
2. Bridges, M.C., A.C. Daulagala, and A. Kourtidis, *LNCcation: lncRNA localization and function.* J Cell Biol, 2021. **220**(2).
3. Zhao, Y., et al., *Long noncoding RNA HOTAIRM1 in human cancers.* Clin Chim Acta, 2020. **511**: p. 255-259.
4. Segal, D., et al., *A conserved HOTAIRM1-HOXA1 regulatory axis contributes early to neuronal differentiation.* RNA Biology, 2023. **20**(1): p. 1523-1539.
5. Wang, X.Q. and J. Dostie, *Reciprocal regulation of chromatin state and architecture by HOTAIRM1 contributes to temporal collinear HOXA gene activation.* Nucleic Acids Res, 2017. **45**(3): p. 1091-1104.
6. Ma, L., V.B. Bajic, and Z. Zhang, *On the classification of long non-coding RNAs.* RNA Biology, 2013. **10**(6): p. 924-933.
7. Quinn, J.J. and H.Y. Chang, *Unique features of long non-coding RNA biogenesis and function.* Nature Reviews Genetics, 2016. **17**(1): p. 47-62.
8. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Mol Cell, 2011. **43**(6): p. 904-14.
9. Brockdorff, N., et al., *Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome.* Nature, 1991. **351**(6324): p. 329-31.
10. Martens, J.A., L. Laprade, and F. Winston, *Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene.* Nature, 2004. **429**(6991): p. 571-4.
11. Martianov, I., et al., *Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript.* Nature, 2007. **445**(7128): p. 666-670.
12. Kretz, M., et al., *Control of somatic tissue differentiation by the long non-coding RNA TINCR.* Nature, 2013. **493**(7431): p. 231-235.
13. Kallen, A.N., et al., *The imprinted H19 lncRNA antagonizes let-7 microRNAs.* Mol Cell, 2013. **52**(1): p. 101-12.
14. Ameres, S.L., et al., *Target RNA–Directed Trimming and Tailing of Small Silencing RNAs.* Science, 2010. **328**(5985): p. 1534-1539.
15. Jones, B.T., et al., *Target-directed microRNA degradation regulates developmental microRNA expression and embryonic growth in mammals.* bioRxiv, 2023.
16. Ghini, F., et al., *Endogenous transcripts control miRNA levels and activity in mammalian cells by target-directed miRNA degradation.* Nature Communications, 2018. **9**(1): p. 3119.
17. Kleaveland, B., et al., *A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain.* Cell, 2018. **174**(2): p. 350-362.e17.
18. Shi, C.Y., et al., *The ZSWIM8 ubiquitin ligase mediates target-directed microRNA degradation.* Science, 2020. **370**(6523).
19. Clyde, D., *Target-directed microRNA degradation in Drosophila.* Nature Reviews Genetics, 2022. **23**(12): p. 713-713.

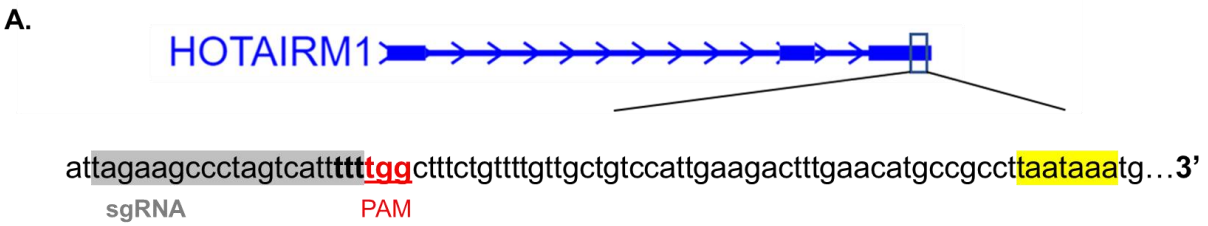
20. Yoon, J.H., et al., *Scaffold function of long non-coding RNA HOTAIR in protein ubiquitination.* Nat Commun, 2013. **4**: p. 2939.
21. Mallo, M. and C.R. Alonso, *The regulation of Hox gene expression during animal development.* Development, 2013. **140**(19): p. 3951-63.
22. Zákány, J., et al., *Deletion of a HoxD enhancer induces transcriptional heterochrony leading to transposition of the sacrum.* Embo j, 1997. **16**(14): p. 4393-402.
23. Juan, A.H. and F.H. Ruddle, *Enhancer timing of Hox gene expression: deletion of the endogenous Hoxc8 early enhancer.* Development, 2003. **130**(20): p. 4823-34.
24. Afzal, Z. and R. Krumlauf, *Transcriptional Regulation and Implications for Controlling Hox Gene Expression.* Journal of Developmental Biology, 2022. **10**(1): p. 4.
25. Nolte, C., B. De Kumar, and R. Krumlauf, *Hox genes: Downstream "effectors" of retinoic acid signaling in vertebrate embryogenesis.* Genesis, 2019. **57**(7-8): p. e23306.
26. Wang, K.C., et al., *A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression.* Nature, 2011. **472**(7341): p. 120-4.
27. Zhang, X., S.M. Weissman, and P.E. Newburger, *Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells.* RNA Biol, 2014. **11**(6): p. 777-87.
28. Zhang, X., et al., *A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster.* Blood, 2009. **113**(11): p. 2526-34.
29. Ha, M. and V.N. Kim, *Regulation of microRNA biogenesis.* Nature Reviews Molecular Cell Biology, 2014. **15**(8): p. 509-524.
30. Dong, H., et al., *MicroRNA: function, detection, and bioanalysis.* Chem Rev, 2013. **113**(8): p. 6207-33.
31. Stavast, C.J. and S.J. Erkeland, *The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation.* Cells, 2019. **8**(11).
32. O'Brien, J., et al., *Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.* Front Endocrinol (Lausanne), 2018. **9**: p. 402.
33. Yi, H., et al., *PABP Cooperates with the CCR4-NOT Complex to Promote mRNA Deadenylation and Block Precocious Decay.* Mol Cell, 2018. **70**(6): p. 1081-1088.e5.
34. Eisen, T.J., et al., *The Dynamics of Cytoplasmic mRNA Metabolism.* Mol Cell, 2020. **77**(4): p. 786-799.e10.
35. Behm-Ansmant, I., et al., *mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes.* Genes Dev, 2006. **20**(14): p. 1885-98.
36. Naeli, P., et al., *The intricate balance between microRNA-induced mRNA decay and translational repression.* Febs j, 2023. **290**(10): p. 2508-2524.
37. Xia, H., et al., *Long Noncoding RNA HOTAIRM1 Maintains Tumorigenicity of Glioblastoma Stem-Like Cells Through Regulation of HOX Gene Expression.* Neurotherapeutics, 2020. **17**(2): p. 754-764.

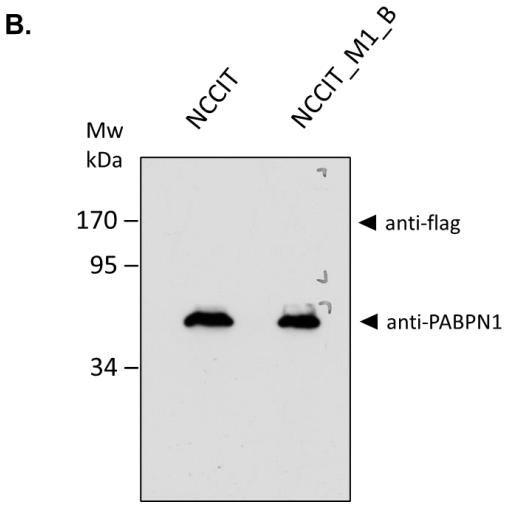
38. Li, X., et al., *LncRNA HOTAIRM1/HOXA1 Axis Promotes Cell Proliferation, Migration And Invasion In Endometrial Cancer.* Onco Targets Ther, 2019. **12**: p. 10997-11015.
39. Xiong, F., et al., *Clinicopathologic Features and the Prognostic Implications of Long Noncoding RNA HOTAIRM1 in Non-Small Cell Lung Cancer.* Genet Test Mol Biomarkers, 2020. **24**(1): p. 47-53.
40. Wan, L., et al., *HOTAIRM1 as a potential biomarker for diagnosis of colorectal cancer functions the role in the tumour suppressor.* J Cell Mol Med, 2016. **20**(11): p. 2036-2044.
41. Chao, H., et al., *HOTAIRM1 suppresses cell proliferation and invasion in ovarian cancer through facilitating ARHGAP24 expression by sponging miR-106a-5p.* Life Sci, 2020. **243**: p. 117296.
42. Zhang, Y., et al., *LncRNA HOTAIRM1 inhibits the progression of hepatocellular carcinoma by inhibiting the Wnt signaling pathway.* Eur Rev Med Pharmacol Sci, 2018. **22**(15): p. 4861-4868.
43. Hao, Y., et al., *Over-expression of long noncoding RNA HOTAIRM1 promotes cell proliferation and invasion in human glioblastoma by up-regulating SP1 via sponging miR-137.* NeuroReport, 2020. **31**(2): p. 109-117.
44. Chen, Z.-H., et al., *The lncRNA HOTAIRM1 regulates the degradation of PML-RARA oncoprotein and myeloid cell differentiation by enhancing the autophagy pathway.* Cell Death & Differentiation, 2017. **24**(2): p. 212-224.
45. Liang, Q., et al., *Long non-coding RNA, HOTAIRM1, promotes glioma malignancy by forming a ceRNA network.* Aging (Albany NY), 2019. **11**(17): p. 6805-6838.
46. Xin, J., et al., *Downregulation of long noncoding RNA HOTAIRM1 promotes monocyte/dendritic cell differentiation through competitively binding to endogenous miR-3960.* OncoTargets and Therapy, 2017. **10**(null): p. 1307-1315.
47. Soutschek, M., et al., *scanMiR: a biochemically based toolkit for versatile and efficient microRNA target prediction.* Bioinformatics, 2022. **38**(9): p. 2466-2473.
48. Boyerinas, B., et al., *The role of let-7 in cell differentiation and cancer.* Endocr Relat Cancer, 2010. **17**(1): p. F19-36.
49. Damjanov, I., B. Horvat, and Z. Gibas, *Retinoic acid-induced differentiation of the developmentally pluripotent human germ cell tumor-derived cell line, NCCIT.* Lab Invest, 1993. **68**(2): p. 220-32.
50. Ahmed, A., *Characterization of the HOXA-Antisense 2 lncRNA during cellular differentiation*, in *Department of Biochemistry*. 2023, McGill University.
51. Baron-Benhamou, J., et al., *Using the lambdaN peptide to tether proteins to RNAs.* Methods Mol Biol, 2004. **257**: p. 135-54.
52. Ramanathan, M., et al., *RNA-protein interaction detection in living cells.* Nat Methods, 2018. **15**(3): p. 207-212.
53. Chen, B., et al., *Live cell imaging and proteomic profiling of endogenous NEAT1 lncRNA by CRISPR/Cas9-mediated knock-in.* Protein & Cell, 2020. **11**(9): p. 641-660.
54. Golden, R.J., et al., *An Argonaute phosphorylation cycle promotes microRNA-mediated silencing.* Nature, 2017. **542**(7640): p. 197-202.

55. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-784.
56. Rousseau, M., et al., *Classifying leukemia types with chromatin conformation data.* Genome Biology, 2014. **15**(4): p. R60.
57. Li, J., et al., *An alternative splicing switch in FLNB promotes the mesenchymal cell state in human breast cancer.* Elife, 2018. **7**.
58. Labun, K., et al., *CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing.* Nucleic Acids Research, 2019. **47**(W1): p. W171-W174.
59. Said, N., et al., *In vivo expression and purification of aptamer-tagged small RNA regulators.* Nucleic Acids Research, 2009. **37**(20): p. e133-e133.
60. Iioka, H., et al., *Efficient detection of RNA–protein interactions using tethered RNAs.* Nucleic Acids Research, 2011. **39**(8): p. e53-e53.
61. De Kumar, B., et al., *Dynamic regulation of Nanog and stem cell-signaling pathways by Hoxa1 during early neuro-ectodermal differentiation of ES cells.* Proceedings of the National Academy of Sciences, 2017. **114**(23): p. 5838-5845.
62. Campbell, J.D., et al., *Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data.* Rna, 2015. **21**(2): p. 164-71.
63. Ibberson, D., et al., *RNA degradation compromises the reliability of microRNA expression profiling.* BMC Biotechnol, 2009. **9**: p. 102.
64. Ramanathan, M., D.F. Porter, and P.A. Khavari, *Methods to study RNA-protein interactions.* Nat Methods, 2019. **16**(3): p. 225-234.
65. Gräwe, C., et al., *RNA-Centric Methods: Toward the Interactome of Specific RNA Transcripts.* Trends Biotechnol, 2021. **39**(9): p. 890-900.
66. *Efficiency of the Elongation Factor-1α Promoter in Mammalian Embryonic Stem Cells Using Lentiviral Gene Delivery Systems.* Stem Cells and Development, 2007. **16**(4): p. 537-546.
67. Qin, J.Y., et al., *Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.* PLoS One, 2010. **5**(5): p. e10611.
68. *Promoter-Dependent EGFP Expression during Embryonic Stem Cell Propagation and Differentiation.* Stem Cells and Development, 2008. **17**(2): p. 279-290.
69. Battulin, N., et al., *The human EF1a promoter does not provide expression of the transgene in mice.* Transgenic Res, 2022. **31**(4-5): p. 525-535.
70. Seita, Y., et al., *Comprehensive evaluation of ubiquitous promoters suitable for the generation of transgenic cynomolgus monkeys†.* Biology of Reproduction, 2019. **100**(6): p. 1440-1452.
71. Chu, C., et al., *Systematic discovery of Xist RNA binding proteins.* Cell, 2015. **161**(2): p. 404-16.
72. Bendak, K., et al., *A rapid method for assessing the RNA-binding potential of a protein.* Nucleic Acids Res, 2012. **40**(14): p. e105.
73. Huang, S., et al., *A new microRNA signal pathway regulated by long noncoding RNA TGFB2-OT1 in autophagy and inflammation of vascular endothelial cells.* Autophagy, 2015. **11**(12): p. 2172-2183.
74. Bieberich, E., *It's a Lipid's World: Bioactive Lipid Metabolism and Signaling in Neural Stem Cell Differentiation.* Neurochemical Research, 2012. **37**(6): p. 1208-1229.

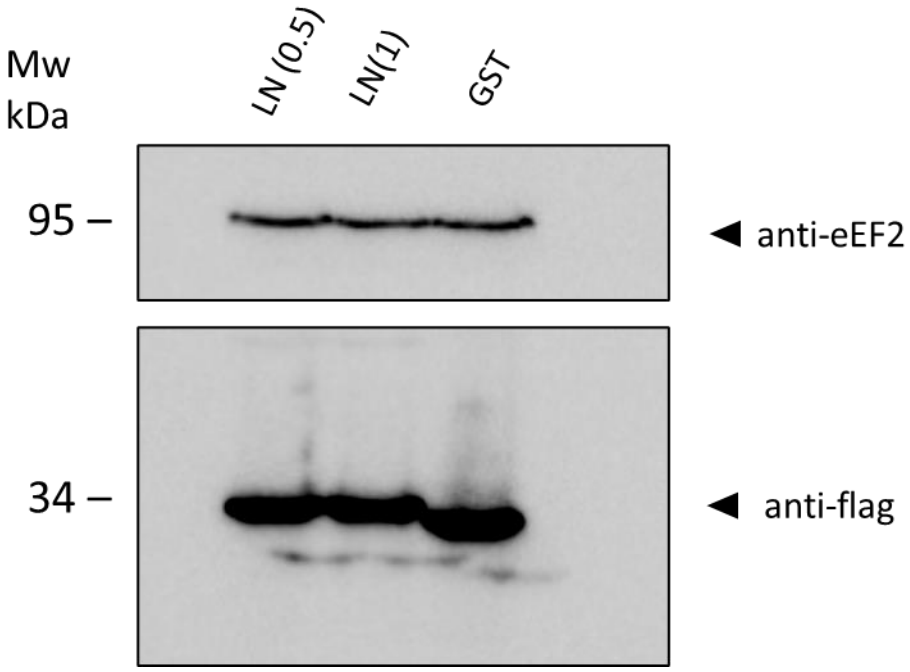
75. Park, H., et al., *Transcript profiling and lipidomic analysis of ceramide subspecies in mouse embryonic stem cells and embryoid bodies.* J Lipid Res, 2010. **51**(3): p. 480-9.
76. Zhao, L., et al., *A deficiency of ceramide biosynthesis causes cerebellar purkinje cell neurodegeneration and lipofuscin accumulation.* PLoS Genet, 2011. **7**(5): p. e1002063.
77. Hu, R., et al., *A Runx2/miR-3960/miR-2861 regulatory feedback loop during mouse osteoblast differentiation.* J Biol Chem, 2011. **286**(14): p. 12328-39.
78. Jalali, S., et al., *Systematic transcriptome wide analysis of lncRNA-miRNA interactions.* PLoS One, 2013. **8**(2): p. e53823.
79. Ling, H., M. Fabbri, and G.A. Calin, *MicroRNAs and other non-coding RNAs as targets for anticancer drug development.* Nat Rev Drug Discov, 2013. **12**(11): p. 847-65.
80. Marranci, A., et al., *Systematic evaluation of the microRNAome through miR-CATCHv2.0 identifies positive and negative regulators of BRAF-X1 mRNA.* RNA Biol, 2019. **16**(7): p. 865-878.
81. Engreitz, J.M., et al., *RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites.* Cell, 2014. **159**(1): p. 188-199.

Appendices



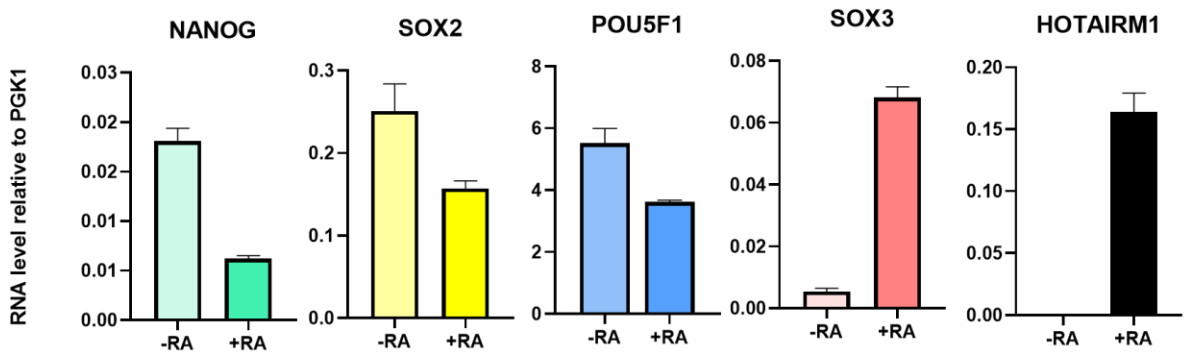


Appendix Figure 1. Short guide RNA design for HOTAIRM1 and Cas9 enzyme validation. **(A)** Overview of the knock-in region at the 3rd exon of the HOTAIRM1 gene. The sgRNA sequence is highlighted in gray, the PAM (NGG) sequence is shown in red, and the cut site is emphasized in bold. The PolyA signal is highlighted in yellow. **(B)** Western blot showing the absence of the Cas9 enzyme (~161 kDa) in our modified cell line (NCCIT_M1_B), with PABPN1 used as an internal control.

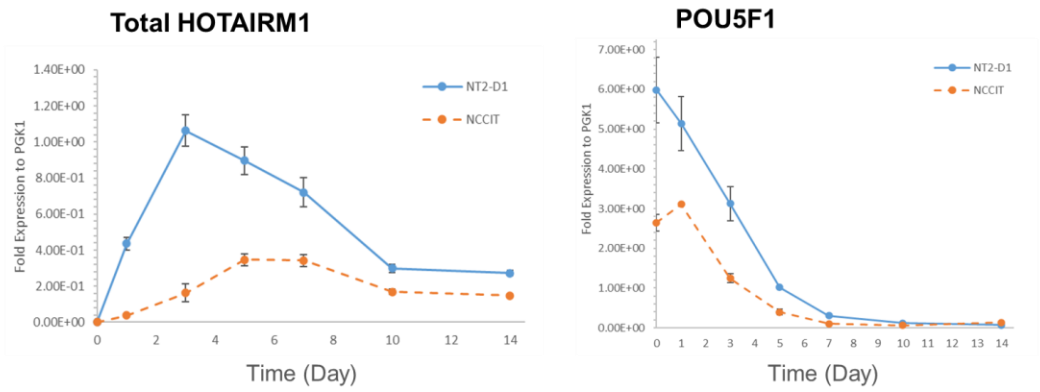


Appendix Figure 2. Western blot displaying λN and GST expression in transduced cell lines. LN (0.5) represents the cells transduced with low MOI; LN (1) represents the cells received high MOI. EEF2 acts as an internal control.

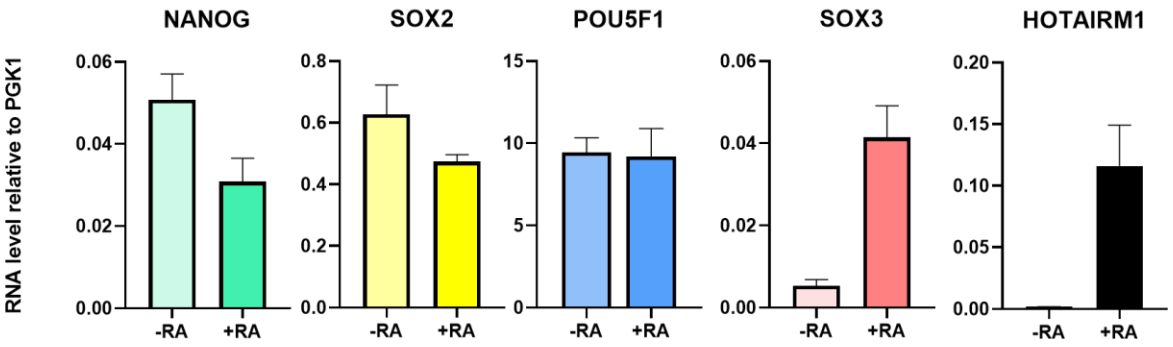
A. WT NCCIT



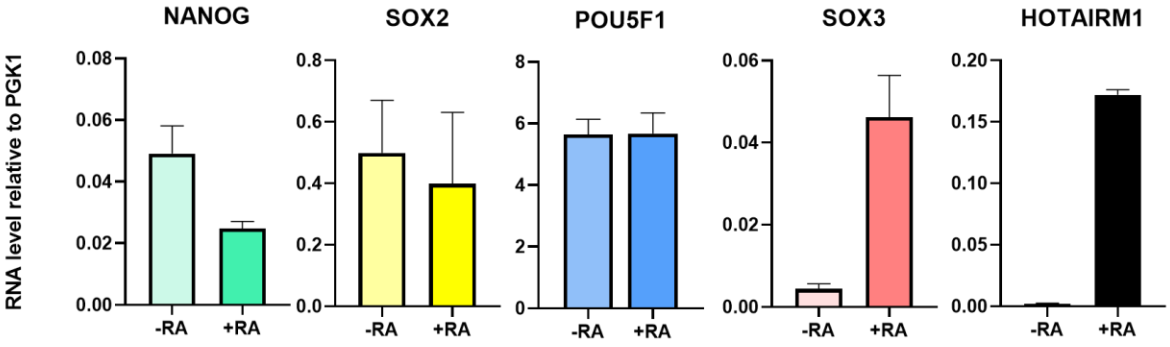
B.



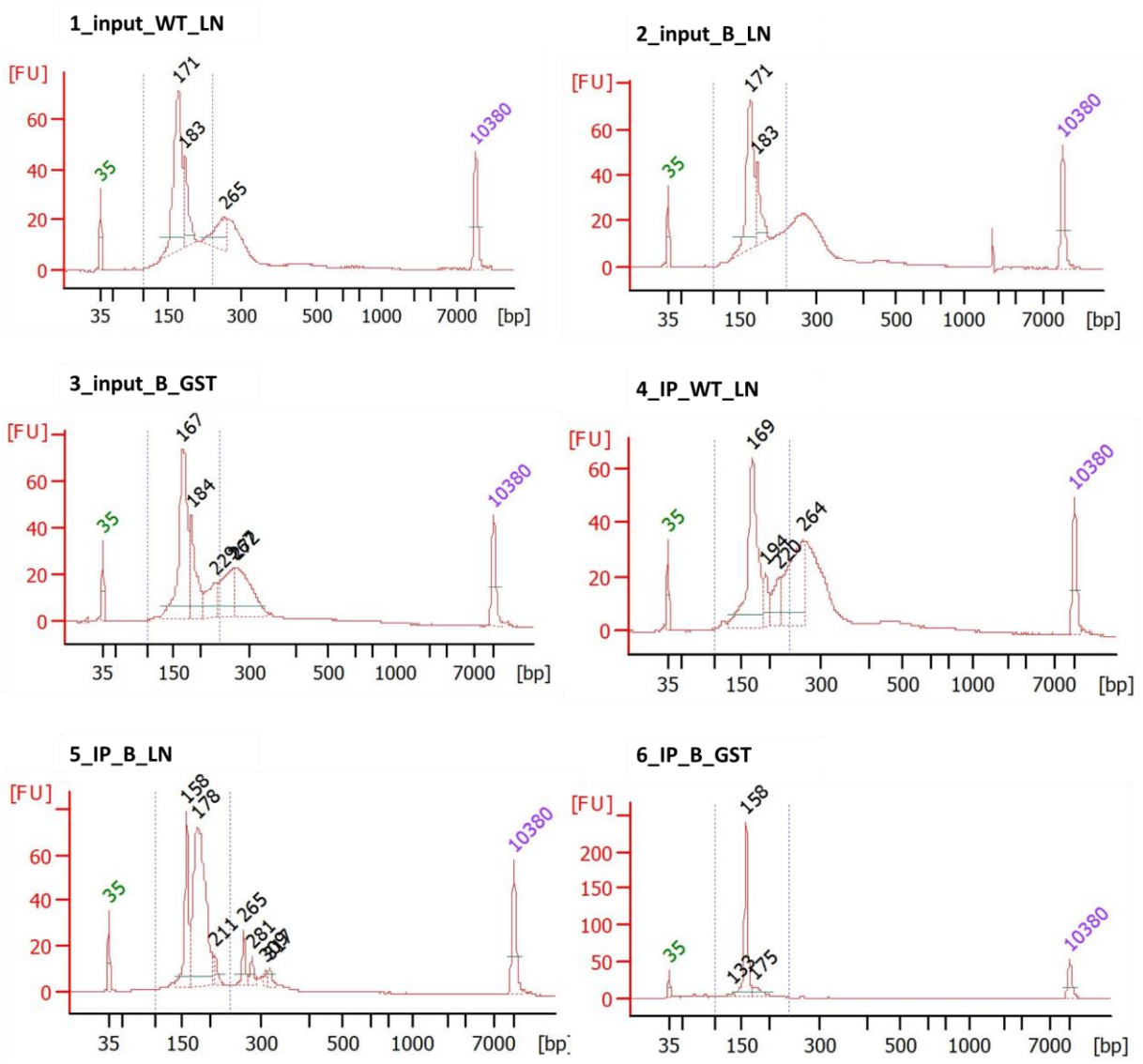
C. NCCIT_M1_B_GST



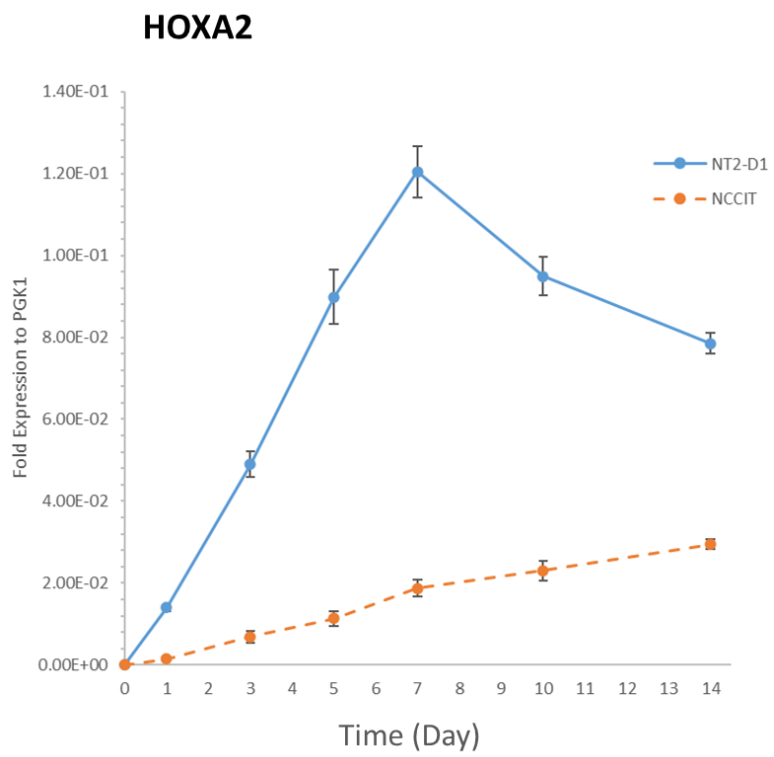
D. NCCIT_M1_B_LN



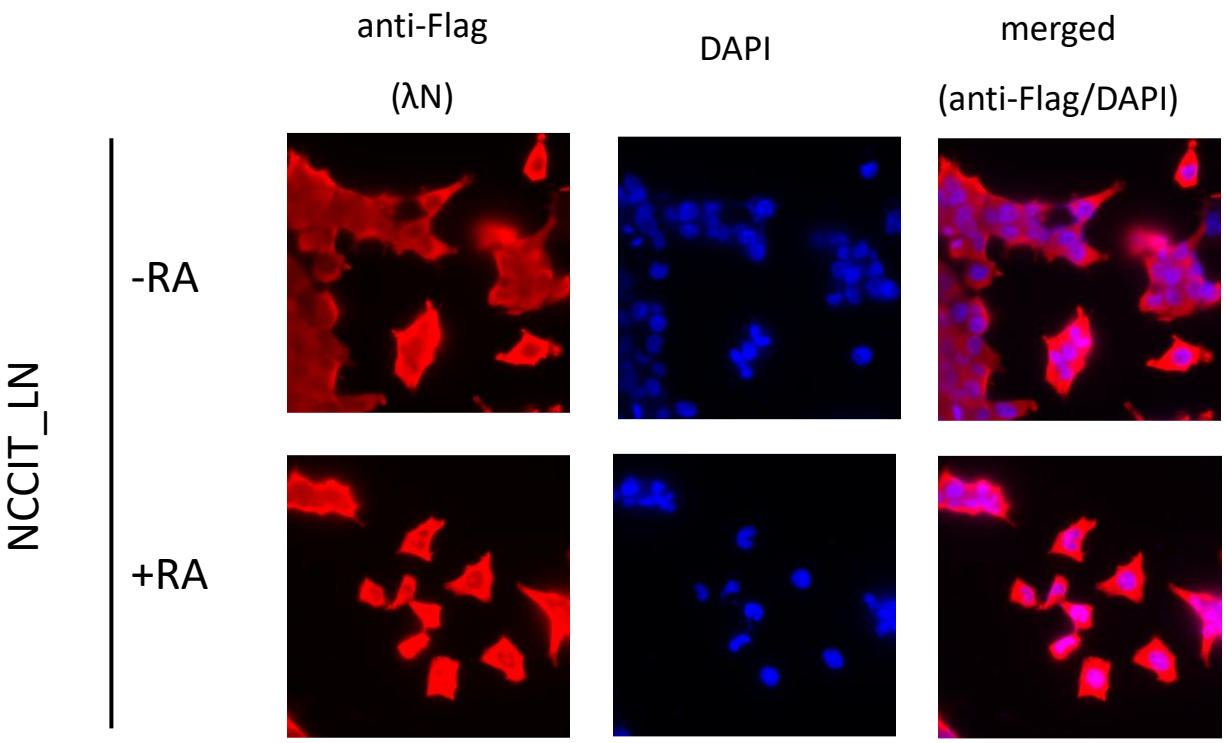
Appendix Figure 3. Gene expression in WT NCCIT and transduced cell lines. **(A)** RT-qPCR results showing the decreased RNA levels for pluripotency markers (*NANOG*, *SOX2*, *POU5F1*) and increased expression of SOX3 and HOTAIRM1 in wildtype NCCIT cells after RA induction. PGK1 serves as an internal control. Error bars represent standard deviations from technical triplicates (n=1). **(B)** Expression levels of HOTAIRM1 and *POU5F1* during a 14-day differentiation in RA-treated NCCIT and NT2-D1 cell lines. The data is adapted from an unpublished master's dissertation (Ahmed A: Characterization of the *HOXA*-Antisense 2 lncRNA during cellular differentiation. McGill University, Montreal). **(C)** RNA quantification for NCCIT_M1_B_GST cell line under RA-induced and un-induced cellular states. *POU5F1* showed a very mild change after RA treatment. **(D)** Same as (C) but for NCCIT_M1_B_LN cell line. Values are averages from 2 biological replicates; error bars indicate standard deviation from biological replicates.



Appendix Figure 4. Electrophoresis run summary showing a high sensitivity DNA assay performed by the Institute for Research in Immunology and Cancer (IRIC). Multiple peaks were observed in the electrophoresis run, with most peaks appearing at approximately 170 base pairs (bp) for samples 1 to 6. The data source is copyrighted material from Agilent Technologies, Inc. (Copyright 2003 - 2009)



Appendix Figure 5. Expression level of *HOXA*2 increases during a 14-day differentiation in RA-treated NCCIT and NT2-D1 cell lines. The data is adapted from unpublished master's dissertation (Ahmed A: Characterization of the *HOXA*-Antisense 2 lncRNA during cellular differentiation. McGill University, Montreal).



Appendix Figure 6. Immunofluorescence analysis showing the subcellular localization of λN-HOTAIRM1-RNA complexes in transduced cells (NCCIT_LN). Blue DAPI staining represents nuclei, and red staining depicts anti-Flag antibodies with Alexa Fluor 568 secondary antibodies.