Tampereen yliopisto

Timo Ahola

THE ROLE OF LONG NON-CODING RNAS IN EPITHELIAL-MESENCHYMAL TRANSITION

Faculty of Medicine and Health Technology
Bachelor's degree
April 2022

TIIVISTELMÄ

Timo Ahola: The role of long non-coding RNA in epithelial-mesenchymal transition
Kandidaatin tutkielma
Tampereen yliopisto
Bioteknologian ja biolääketieteen tekniikan tutkinto-ohjelma
Huhtikuu 2022

Ei-koodaavat RNAt ovat RNA-molekyyleja, jotka lähetti-RNA:iden tavoin transkriptoidaan DNA:sta, mutta ei-koodaavia RNA:ita ei transloida proteiiniksi. Ne osallistuvat useisiin solulle välttämättömiin säätelyprosesseihin. Ei-koodaavat RNA:t jaetaan usein pituutensa mukaan pitkiin ei-koodaaviin RNA:ihin ja lyhyisiin ei-koodaaviin RNA:ihin. Lyhyitä ei-koodaavia RNA:ita on luokiteltu useita eri tyyppejä, mutta yksi eniten tutkituista tyypeistä ovat mikro-RNA:t (miRNA), jotka koostuvat 18–24:stä nukleotidista ja saavat alkunsa niille tyypillisestä syntyprosessista. Useimmiten miRNA:t toimivat geenien hiljentäjinä sitoutumalla niille komplementaarisiin lähetti-RNA:ihin. Pitkät ei-koodaavat RNA:t puolestaan kykenevät vastustamaan miRNA:iden aiheuttamaa hiljentämistä sitomalla niille komplementaarisia miRNA:ita itseensä, ja näin estämällä miRNA-välitteisen hiljennyksen. Ne kykenevät säätelemään geenien ilmentymistä myös sitoutumalla DNA:han ja näin muuttamallaan paikallista kromatiinirakennetta, mikä vaikuttaa kyseisen alueen geenien ilmentymiseen. Tällaisella pitkiin ei-koodaaviin RNA:ihin perustuvalla säätelyllä on huomattu olevan tärkeä rooli monessa biologisessa prosessissa, joista yksi hyvä esimerkki on epiteeli-mesenkyymi -siirtymä (EMT). Siinä epiteelisolut menettävät joitain niille tyypillisiä ominaisuuksia, ja saavat mesenkyymisolujen ominaisuuksia. Vaikka EMT tapahtuukin osana normaalia yksilön kehitystä, on sillä myös synkempi rooli osana syövän kehitystä, sillä EMT:n läpikäynti johtaa syöpäsolujen migraatioon ja etäpesäkkeiden muodostumiseen. Esimerkki pitkästä ei-koodaavasta RNA:sta on *Vimentin antisense 1* (Vim-as1), joka edesauttaa EMT:n syntyä usean eri säätelymekanismin välityksellä. Koska yksi RNA-molekyyli kykenee osallistumaan useaan mekanismiin samanaikaisesti ja yhdellä molekyylillä voi olla monia säätelykohteita, pitkät ei-koodaavat RNA:t muodostavat monimutkaisen säätelyverkoston, jossa suuret muutokset, kuten EMT, tapahtuvat usean säätelijän yhteisvaikutuksesta. Tämän takia pitkiä ei-koodaavia RNA:ita, ja niiden roolia EMT:ssä tulee tutkia useasta näkökulmasta, tarkastellen sekä laajoja muutoksia RNA:iden ilmentymisessä, että molekulaarisia mekanismeja, joiden avulla yksittäinen RNA edesauttaa EMT:tä.

Avainsanat: lncRNA, EMT, ei-koodaava RNA, RNA-säätely

ABSTRACT

Timo Ahola: The role of long non-coding RNA in epithelial-mesenchymal transition
Bachelor's degree
Tampere university
Biotechnology and biomedical engineering
April 2022

Non-coding RNAs (ncRNAs) are RNA-molecules that are transcribed from DNA but are not translated into protein. They have been shown to be important regulators of cellular function acting through their interactions with other RNAs, DNA, and proteins. Non-coding RNAs can be categorized into short ncRNAs which are less than 200 nucleotides in length and long ncRNAs (lncRNAs) which are more than 200 nucleotides in length. Many different types of short non-coding RNAs have been found but one of the most extensively researched types are microRNAs which are 18-24 nucleotides in length and are generated through a characteristic maturation pathway. Micro-RNAs are most often involved in downregulation of gene expression through hybridizing with complementary mRNAs. This repression can be counteracted by lncRNAs which can sequester miRNAs, preventing them from binding to their mRNA targets. Long ncRNAs can also bind to DNA and alter chromatin behavior to influence the expression of genes located in the area. One event where such lncRNAs play a significant role is epithelial-mesenchymal transition (EMT), a process in which epithelial cells lose some of their characteristic properties and gain those of mesenchymal cells. Although EMT occurs as a part of normal embryonic development, it also enables the migration and metastasis of tumor cells during cancer progression. An example of a lncRNA acting during cancerous EMT is vimentin antisense 1 (Vim-as1), which can simultaneously act through several lncRNA-based regulatory mechanisms to drive EMT. The fact that a single lncRNA can regulate gene expression through multiple mechanisms and can have several downstream targets means that the regulatory networks formed by lncRNAs are very complex. Cellular events, such as EMT, occur through the influence of many regulators working in conjunction, instead of being governed by a single master-regulator. For this reason, lncRNAs and their role in EMT require research from several points of view as both global changes in gene expression and the detailed molecular mechanisms underlying single lncRNAs must be investigated. The central role of lncRNAs in EMT suggests that lncRNAs may have potential as a biomarker or therapeutic target for cancer but for now, applications have yet to reach the clinical stage.

Keywords: lncRNA, EMT, non-coding RNA, RNA regulation, epithelial-mesenchymal transition

PREFACE

This bachelor's thesis was conducted in the Faculty of Medicine and Health Technology of Tampere University as a part of the Biotechnology and Biomedical Engineering program studies. The thesis was supervised by associate professor Minna-Liisa Änkö.

Tampere 27.04.2022

Timo Ahola

TABLE OF CONTENTS

1. INTRODUCTION

For decades the central dogma of molecular biology was seen as a two-part process: transcription where DNA was read into messenger RNA (mRNA), and translation where the mRNA code was used as a template to synthesize proteins. Even though transcripts, which did not lead to the translation of a protein were detected, they were largely thought to be transcriptional background noise without a biological function. However, more recently these non-coding RNAs (ncRNAs) have been found to take part in various regulatory mechanisms. The regulatory processes occur during transcription (co-transcriptionally) or between transcription and translation (post-transcriptionally) and include modifications, localization, and translational regulation of mRNA. Non-coding RNAs have been found to be common mediators of these mechanism, working in various ways to alter gene expression. This has opened many new areas of inquiry and revealed that the biological regulation of gene expression is greatly influenced by processes happening outside transcription and translation. (Ratti et al., 2020)

As their roles in cellular regulation were found to be more significant, ncRNAs were also linked to many disease-events where abnormal expression and function of ncRNAs leads to dysfunction of the regulatory mechanisms they mediate, and ultimately, pathogenic cellular behavior. Many cancers and cancerous events have been found to have ncRNA-based regulatory systems behind tumor development and progression. One such event, important for the development and malignancy of many tumors, is epithelial-mesenchymal transition (ECM). ECM is a process which takes place during normal development and cell differentiation, but it can also involve a pathogenic event in adults in diseases such as cancer and preeclampsia. During ECM, epithelial cells gain mesenchymal properties. Since this often involves the loss of cells' capability to tightly bind to the basal lamina and adjacent cells, it can lead to loss of tissue morphology, unwanted cell-migration, and eventually cancer-migration (metastasis). (McCabe and Rasmussen, 2021)

In this thesis I will describe the roles of ncRNAs in EMT, focusing on lncRNAs. The antisense transcript of vimentin, an ECM protein, will be used to illustrate the role of lncRNAs in EMT, thus the focus will be on mechanism relevant to this example.

2. NONCODING RNAS, A DIVERSE GROUP WITHOUT CODING POTENTIAL

Non-coding RNAs are a very diverse group of molecules, differing greatly in length and function. Often the first level of classification of ncRNAs is by length: ncRNAs are divided into long non-coding RNAs (lncRNAs) that are more than 200 nucleotides in length and short non-coding RNAs that are less than 200 nucleotides in length. Non-coding RNAs can partake in a wide range of regulatory processes through their molecular properties. They can use their ability to hybridize with DNA through base-pairing, they can bind with many RNA-binding proteins (RBPs), and they can form secondary structures by base-pairing with themselves. This plurality of mechanisms makes ncRNAs an extremely diverse group with only a few common properties. (Ratti et al., 2020)a

2.1 microRNAs are short ncRNAs that inhibit complementary mRNAs

Short ncRNAs, simply defined as RNAs less than 200 nucleotides in length, are also a very diverse group, consisting of many different types of RNAs involved in a plethora of mechanisms, essentially in every facet of biological action. They can be further classified into different functional groups such as microRNAs (miRNAs), small interfering RNAs, small nuclear RNAs, and circular RNAs. Micro-RNAs are one of the most well-studied and –mechanistically understood group of ncRNAs. They are RNA-molecules 18-24 nucleotides in length and most often partake in downregulation of gene expression at the translational level through miRNA-induced silencing. Micro-RNAs hybridize with complementary mRNA sequences, inducing mRNA degradation or translational repression. Micro-RNAs are also significant actors in other forms of RNA-regulation as they can interact with other ncRNAs, such as lncRNAs, in addition to mRNAs. (Ratti et al., 2020)

Canonically, miRNAs are transcribed from specific miRNA-genes, which can produce either mono- or polycistronic transcripts. Even though miRNAs usually act in their short form in the cytosol, they originate in the nucleus as primary-miRNAs (pri-miRNAs), which are longer RNAs transcribed by RNA-polymerase-II or -III, and contain one or more hairpin structures, and often a poly-A tail, and a 5'-cap. Pri-miRNAs are processed inside the nucleus by DROSHA (which is called Pasha in flies) into a simpler hairpin structure, known as precursor-miRNA (pre-miRNA). Pre-miRNAs are transferred into the cytoplasm by exportin 5, where an RNAse called DICER cleaves the pre-miRNA into the final, functional, length of the miRNA. Finally, a single miRNA-strand is loaded to an RNA-induced silencing (RISC) complex, which works as a mediator of miRNA-induced silencing. (Treiber et al., 2019) The biogenesis of miRNAs is presented in Figure 1.

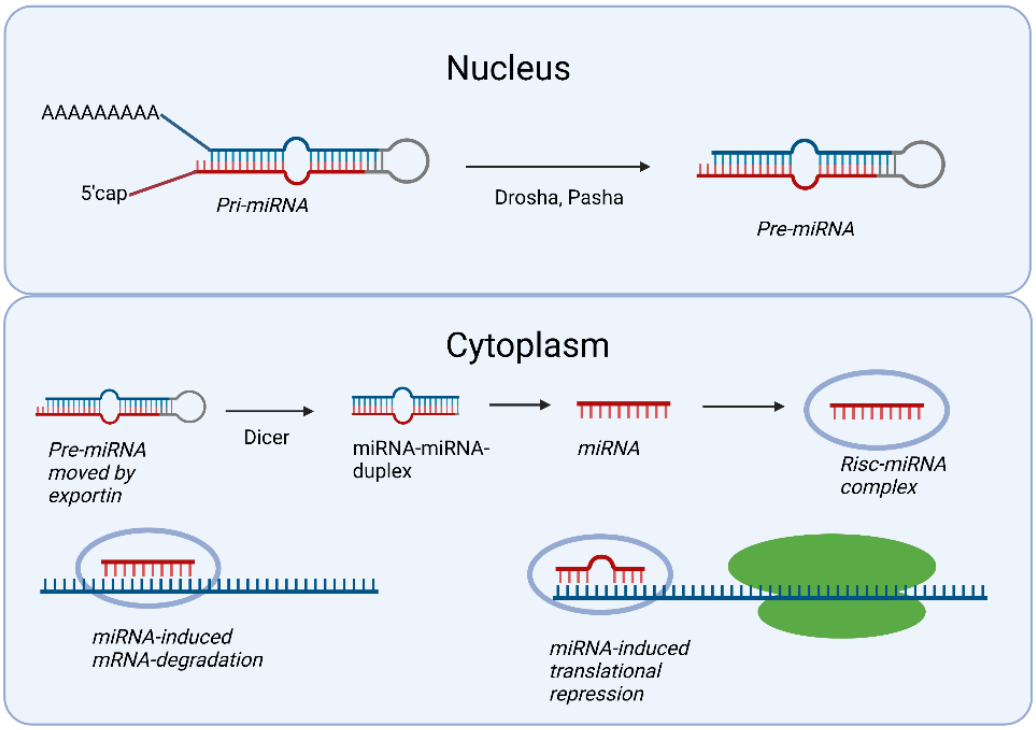

Figure 1: The transcription and maturation of miRNAs. Created with BioRender.com

Most miRNAs induce silencing by binding to the 3' untranslated regions (UTRs) of mRNAs. Depending on the degree of complementarity between the miRNA and mRNA, miRNA-binding can either lead to degradation of mRNA, or translational repression, where the bound miRNA acts as a translational block. (Ratti et al., 2020) The molecular mechanism of silencing works through the components of the RISC-complex. The minimal components of a RISC-complex are the guide strand (miRNA) and an argonaute (AGO) -protein. When the interaction between miRNA and mRNA is fully complementary, the endonucleic activity of AGO2 targets the mRNA, leading to cleavage. However, most interactions between miRNA and mRNA are not fully complementary, as there are often at least minor discrepancies between RNA sequences. In these cases, the endonuclease remains inactive. Instead, AGO will recruit proteins capable of perturbating the poly-A tail and 5' cap of mRNAs to the RISC-complex. The poly-A tail and 5'-cap are important regulators of mRNA, and their removal often leads to translational repression and decreased stability of the mRNA, also exposing the mRNA to traditional exonucleases. (O'Brien et al., 2018)

Recent research has revealed that the full picture of miRNAs is more than originally thought, as many miRNAs have been noted to diverge from the canonical pathway in origin and mechanism of action. Many miRNAs do not originate from the canonical pathway of miRNA biogenesis, but instead can originate from introns of non-miRNA genes or from other ncRNAs, such as transfer-RNAs (tRNAs), which can be cleaved into miRNAs by RNases. Some miRNAs are also never transported into the cytoplasm, but instead go through their maturation process in the nucleus,

some mature in the cytoplasm but are later transported into the nucleus, and some can go through continuous shuttling between the two spaces. In the nucleus, these non-canonical miRNAs can regulate the transcription of their target genes, resulting in gene silencing or even activation in some cases. (Stavast and Erkeland, 2019) Interestingly miRNAs can also partake in other systems besides regulating gene expression. Extracellular circulating miRNAs are miRNAs which can be found in bodily fluids, for example blood or urine. They can end up in the fluids as a result of cell death, but cells can also excrete certain miRNAs into vesicles through active transport, indicating that they have a role in intercellular signaling. (Condrat et al., 2020)

With their significant effect on gene expression and multi-step biogenesis, miRNAs present a very potential window for the development of pathogenicity. Micro-RNAs have been associated with several diseases, including cancers. In many cases, they have also been proposed to serve as biomarkers or therapeutic targets. However, the inherent complexity of the regulatory systems they are involved in makes the evaluation of side effects and off-targets of potential therapeutics difficult. (Condrat et al., 2020) The expression of miRNAs can be perturbated on many different levels, starting from transcriptional regulation of the miRNA genes, and including alterations in the proteins involved in miRNA maturation. A single miRNA can also have several target mRNAs since sufficient similarities in sequence can be found between several genes. This means that several factors, such as other ncRNAs, can influence the function of miRNAs, while alterations in a single miRNA can have multiple downstream effects.

2.2 Long ncRNAs regulate cellular function through various mechanisms

Long ncRNAs share many properties with mRNA, often containing exons and introns, a poly-A tail, and a 5’cap, but importantly, lncRNAs have no open reading frame (ORF) and thus, cannot be translated into protein. Non-coding RNA genes are located around the genome and ncRNAs can be classified by origin as follows: sense lncRNAs, antisense lncRNAs, bidirectional lncRNAs, intronic lncRNAs, and intergenic lncRNAs. Long ncRNAs can have regulatory functions both in *cis* (locally) and in *trans* (distally). (Dhanoa et al., 2018)

Long non-coding RNAs can regulate gene expression by hybridizing with DNA or other RNAs, or by base-pairing with themselves in order to form secondary structures. Antisense lncRNAs are transcribed from the same locus as a protein-coding gene, but from the opposite strand. Long non-coding RNAs can also regulate the expression of their adjacent gene through interactions with the chromatin in its site of transcription. More distal mechanisms involve binding to proteins or other RNAs. Long non-coding can bind to proteins, working as scaffolds that aid their conformation or preventing them from binding to other nucleic acid targets. Their ability to bind other RNAs also

enables them to competitively bind miRNAs, removing miRNAs from their mRNA targets and resulting in the loss of miRNA-induced silencing. (Dhanoa et al., 2018)

Due to their involvement in a plethora of gene regulatory mechanisms, lncRNAs have been associated with several diseases. Since different tissues and diseases, cancer specifically, have been shown to have unique lncRNA signatures, they have been suggested as potential biomarkers for various diseases. This is further supported by the fact that lncRNAs are often released within apoptotic cells or exosomes and can be bound to proteins making them resistant to RNases, and therefore more reliably detectable. (Ratti et al., 2020)

2.2.1 Competitive endogenous RNAs can counteract miRNAs

Since miRNAs are frequently only partially complementary to their target mRNAs, leading to translational repression instead of mRNA-degradation, a single mRNA can be regulated by multiple miRNAs, while a single miRNA can also have several targets. This leads to the existence of the competitive endogenous RNA (ceRNA) -mechanism, where one RNA binds to a miRNA, sequestering it from other mRNA targets of the miRNA, preventing miRNA-induced silencing. A ceRNA-relationship can exist between any two RNAs, for example two mRNAs. Importantly, many lncRNAs act as ceRNAs, often called miRNA “sponges”. Long non-coding ceRNAs have been identified as key-regulators of many developmental processes, such as muscle differentiation, which gives them a more tissue-specific expression signature than mRNAs. Rogue function of lnc-ceRNAs can also be a key regulatory event in the development of malignancies, such as cancer. (Wang et al., 2016) The ceRNA-mechanism is presented in Figure 2.

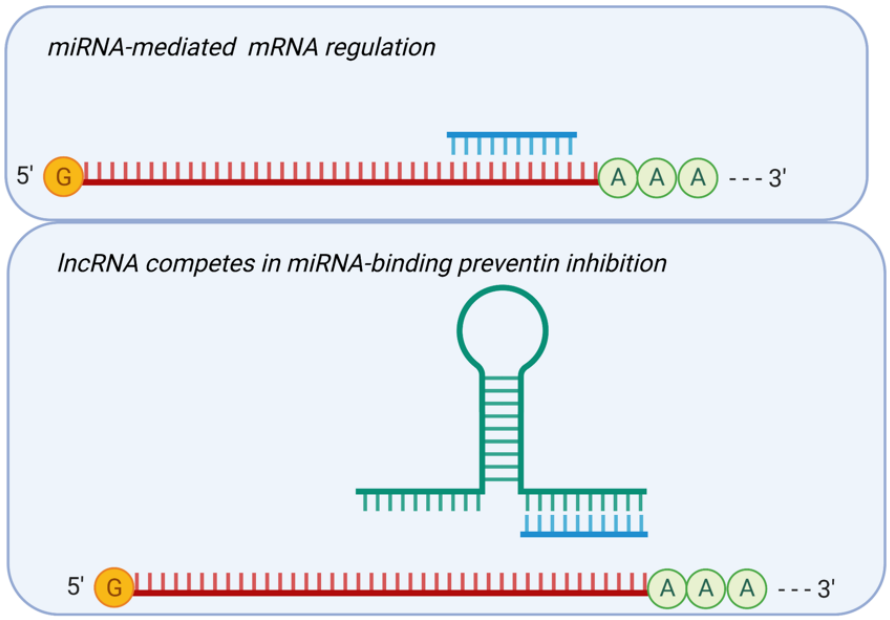

Figure 2: Long non-coding RNAs can sequester miRNAs, preventing them from binding to their mRNA targets. Created with BioRender.com

2.2.2 Long ncRNA-mediated gene regulation through R-loops

Prominent examples of lncRNAs regulating gene expression through chromatin-binding are the DNA-RNA hybrids, termed R-loops. The essence of an R-loop is a three-stranded structure, with a double strand between one of the DNA-strands and an invading RNA strand, and a displaced DNA strand. R-loops are most commonly formed as a byproduct of ongoing transcription, forming behind the elongating RNA-polymerase. Their formation is aided by the fact that negative supercoiling of the DNA, occurring behind the polymerase, often results in strand separation, giving space for hybrid-formation. Even though RNA-DNA binding happens within the transcription bubble, R-loops form during a later re-invasion of the RNA, as the initial hybridization is broken when the DNA and RNA are ejected from different parts of the polymerase. R-loops can be formed by all kinds of RNAs, including mRNAs, but often they are formed by lncRNAs, allowing them to influence gene regulation through chromatin-binding. (Niehrs and Luke, 2020)

Although some physiological roles of R-loops have been known for decades, they were originally thought to be mainly involved in damaging events. Unaccounted R-loops can function as blockers of transcription, preventing the gene expression of nearby genes. The structure also leaves the misplaced ssDNA without the protection of the double-stranded structure, making it susceptible to surrounding biochemical stress, which can result in a damaged DNA-strand. R-loops are mostly associated with double-strand breaks (DSBs), due to their ability to generate genomic instability. In the presence of an R-loop, genomic damage can happen during replication as a result of breakage of the replication fork since the R-loop can act as a physical obstacle, preventing the progression of the replication fork. This phenomenon is assisted by ongoing transcription at the R-loop, as

it can also halt the progression of RNA-polymerases, creating increased blockage. Due to their damaging nature, cells have safeguards to prevent R-loops from forming and resolving already formed R-loops. Heterochromatin is a major factor that prevents R-loops, as markers of it have been associated with diminished numbers of R-loops, whereas in open chromatin an increase in R-loops has been observed. More local mechanisms of prevention have also been suggested, in the form of RNA-binding proteins (RBPs), which can bind to RNA co-transcriptionally, blocking hybrid-formation. Topoisomerases also contribute by removing the negative supercoiling of DNA, which diminishes strand-separation. Ways in which cells resolve R-loops include RNA-dependent ATPases, which have been shown to exhibit helicase activity to DNA-RNA hybrids. Another mechanism involves RNase H proteins, a set of enzymes capable of degrading the RNA of DNA-RNA hybrids. (García-Muse and Aguilera, 2019)

R-loops are, however, much more than harmful by-products of transcription. R-loop formation has been shown to be an important local regulator, and some RNAs have even been shown to form regulatory R-loops *in trans*, in a spatially distant locus. Regulatory roles are also indicated by the fact that R-loops are distributed non-randomly across the genome within regions of the DNA that are rich in GC-nucleotides. This makes R-loops common in mammalian promoters since they often contain GC-islands. The fact that R-loops often form in promoter regions gives them great potential for transcriptional regulation. R-loops have been noted to regulate transcriptional activation by chromatin accessibility allowing for the binding of regulatory factors, and transcriptional termination, preventing transcriptional read-through. Transcriptional activity also increases the incidence of R-loops, meaning that the regulation can work in both directions: R-loops can promote transcription while transcription can promote R-loop formation. This highlights the importance of balancing the creation and resolution of R-loops, in order to maintain normal cellular behavior. (Niehrs and Luke, 2020)

3. EPITHELIAL-MESENCHYMAL TRANSITION

Epithelial-mesenchymal transition (EMT) is a process where epithelial cells lose some of their characteristic properties and gain mesenchymal properties. Hallmarks of EMT are considered to be the cells disassociating from the epithelial sheet and the downregulation of E-cadherin. Many transcription factors such as ZEB1, SNAIL, and SLUG are strongly associated with EMT, driving EMT by repressing the transcription of E-cadherin. EMT is also characterized by the loss of other epithelial markers, such as cadherins and claudins, and the gain of mesenchymal markers, for example vimentin and N-cadherin. During embryogenesis, EMT and its reverse process mesenchymal-epithelial transition (MET) occur several times during tissue differentiation and organ development. For example, during neural development the neural plate undergoes EMT, further developing into neurons and the surrounding epidermis. EMT can also occur in adults during tissue regeneration and wound-healing. In these cases, EMT is tied to inflammatory reaction and will cease when the inflammation is removed. EMT also occurs in adults during cancer development and progression. Cells that undergo EMT have been shown to exhibit increased stemness, which includes the augmentation of proliferation potential. More mesenchymal-type properties also mean that the cells gain invasive capabilities and can break away from their tissue structure. Therefore, cells undergoing EMT can lead to the cancer cells invading surrounding tissues. Motile mesenchymal-like cancer cells can also form more distal metastasis, leading to catastrophic spreading of the tumor. Notably, metastasis usually requires both EMT and MET, as after cells that have undergone EMT have invaded other tissues, they have to go through MET in order to settle down and form metastasis. This means that the spreading of cancer demands flexible regulation of EMT so that the cells are able to spread and form metastasis. (McCabe and Rasmussen, 2021) The progression of EMT is presented in Figure 3.

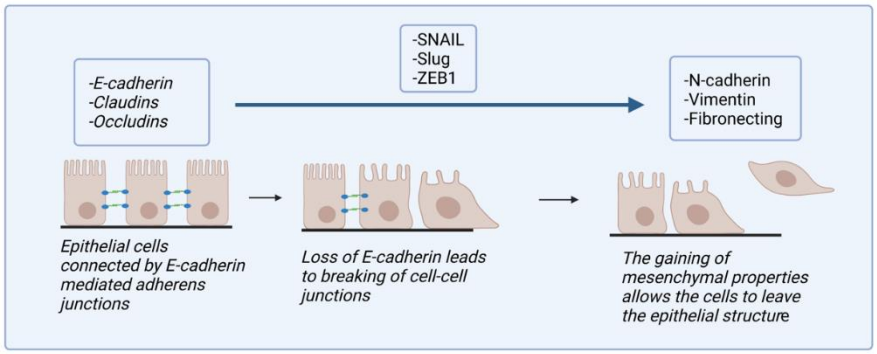

Figure 2: The loss of epithelial cell-adhesion can result in the cells exiting the tissue structure. Created with BioRender.com

4. THE ROLE OF LNCRNAS IN EMT

4.1 Identification of lncRNAs involved in EMT

Soon after lncRNAs had sparked the interest of researchers, they were associated with diseases, especially cancer. They were found to be involved in many different cancerous events across a number of cancer-types. One of the most significant events was found to be EMT, where several lncRNAs can function as facilitators of the transition. Regulatory relationships were discovered between many well-characterized lncRNAs and known EMT-promoting genes. For example, Metastasis Associated Lung Adenocarcinoma Transcript 1 (Malat1) was found to regulate ZEB1 through sponging, while simultaneously partaking in chromatin remodeling that facilitates EMT. Similar examples of lncRNAs interacting with other regulatory factors of EMT can be found across cancer-types. (Dhamija and Dieterich's, 2016)

Frequently, the induction of EMT by lncRNAs occurs because lncRNAs disturb an existing regulatory balance. An interesting example of this is lncRNAs perturbating the regulatory relationship between p53, a known tumor-suppressor, and EMT-inducing transcription factors (reviewed by Parfenyev, et al. in 2021). Under normal circumstances, the p53-protein suppresses many known EMT-inducing transcription factors, such as *ZEB1* and *SNAIL,* by enhancing the expression of several miRNA-genes which inhibit the mentioned transcription factors. In turn, the EMT-inducing transcription factors can regulate the function of p53, creating a two-way regulatory relationship. The transcription factors, for example the Zeb1-protein, can bind to p53, preventing its function as a transcription factor for EMT-suppressing miRNA-genes. The result is a feedback loop where p53 and EMT-inducing transcription factors suppress each other's activity. However, this balance can be disturbed by lncRNAs, for instance the aforementioned Malat1. By spongin the miRNAs responsible for repressing EMT-inducing transcription factors, the lncRNAs can break the feedback loop by preventing the repression of the transcription factors. Increased expression of the EMT-transcription factors due to sponging of the miRNAs leads to increased inhibition of p53. (Parfenyev et al., 2021) The mechanism is illustrated in Figure 4.

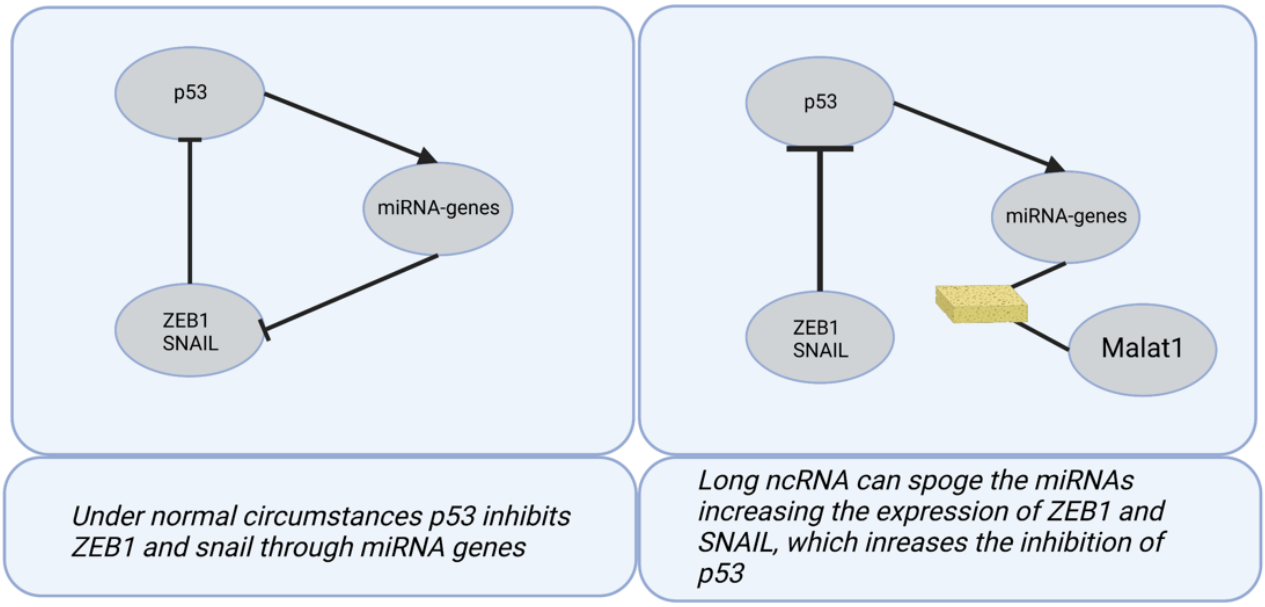

Figure 4: Through sponging, lncRNAs can disturb the regulatory balance between p53 and EMT-inducing transcription factors. Many lncRNAs and transcription factors are involved in the process, but Malat1, *ZEB1,* and *SNAIL* are presented as examples. Created with BioRender.com

This example illustrates how lncRNAs can have profound downstream effects, even when their total expression levels are low. After the initial sponging event the existing regulatory mechanisms can further enhance the effect. Increased expression of the EMT-inducing transcription factors will lead to increased inhibition of p53, which in turn will lead to further increase of the expression of the transcription factors, creating a feedback loop through which EMT can ultimately be unleashed. This is supported by the fact that p53 inhibition, which is present in almost all metastatic cancers, has in many cases been attributed to the decrease in the expression of miRNAs that normally inhibit EMT-inducing transcription factors (Parfenyev et al., 2021).

4.2 Transcriptome-wide analyses define global changes during EMT

The advancements in sequencing technologies have allowed for transcriptome-wide analysis of lncRNA expression. Liao et al (2017) used RNA-sequencing before and after cells had undergone EMT to identify lncRNAs that change in their expression during EMT. They treated MCF10a-cells, cells from a well-established breast-cancer cell line, with Transforming growth-factor β, which had been shown to induce EMT in the cells. RNA-sequencing demonstrated that the total number of lncRNAs changing in expression was 3404. Interestingly, this number was higher than the number of altered coding mRNAs (2394). The findings suggested that global changes in lncRNA expression may play a role in EMT. In order to identify regulatory relationships between altered lncRNAs and mRNAs, Liao et al searched for all mRNAs whose expression correlated with any lncRNAs level (p-value <0.05). Long ncRNAs and mRNAs with correlated relative expression were proposed to

have a regulatory relationship. Of the differentially expressed lncRNAs, 208 were found to correlate with the expression of 1428 mRNAs, suggesting the ability of a single lncRNA to influence the expression of several genes. To verify the correlative relationship, Liao, et al. induced EMT by other methods, first by overexpressing Twist in HMLE-cells, and then by overexpressing SNAIL in MFC10a-cells. Both methods resulted in considerable changes in lncRNA expression, with over 2000 altered lncRNAs. However, when comparing lncRNAs whose expression was altered by differently induced EMTs, 89.6 percent were found to be unique to only one induction agent. (Liao et al., 2017) The fact that global alterations of lncRNA-expression occur in all tested EMTs suggests that they have a considerable role in its development. However, it seems that the global changes are highly cell-type specific with certain key-regulators commonly altered in them. This shows that an important step in investigating the role of lncRNAs in EMT, and perhaps other cancer-events, is identifying the most prominent regulators from the total pool of differentially expressed lncRNAs.

4.3 Meta-analyses reveal EMT-specific lncRNA signatures

The development of statistical and computational analysis methods, combined with more comprehensive transcriptomics databases, has allowed researchers to investigate the relationship between lncRNAs and EMT using previously published data. For example, the cancer genome atlas (TCGA) contains 20,000 cancer transcriptomes along with normal samples across 33 different cancer-types (Hutter and Zenklusen, 2018). The datasets can be used to analyze the changes in lncRNA expression between tumor and healthy samples. Using large sample sizes and global expression profiles allows for the identification of lncRNA signatures characteristic of e.g., cancer subtypes. A signature refers to a cohort of event-specific, for example EMT, or tissue-specific, for example mesenchymal tissue, expressions of lncRNAs. Using this meta-analysis approach, Du, et al. (2021) were able to analyze 430 samples, 411 tumor samples, and 19 healthy, from TCGA to determine the lncRNA signature of EMT in bladder cancer. They then determined stromal and immune scores for each of the samples. This was done by determining the expression levels of stromal- and immune-related genes in the samples. The samples were categorized into high and low stromal groups by comparing them to the median stromal score of all samples. Differentially expressed genes (DEGs) were then determined between the high and low stromal groups. Differences that had a log fold change greater than one, meaning the change in expression value is more than two-fold, were considered significant EMT-related genes. To find lncRNAs related to EMT, they obtained 200 EMT-related genes from the Molecular Signatures Database (Liberzon et al., 2015). Using Pearson correlation, they determined which lncRNAs were co-expressed with the EMT-related genes, with significant relationships having a correlation greater than 0.4 and a p-value under 0.001. Four hundred and twenty-one lncRNAs were found, which were then intersected with the DEGs determined previously. 82 lncRNAs were found to be both correlated with EMT-

related genes and differentially expressed. To further narrow down the list of relevant lncRNAs they also constructed a list of immune-related lncRNAs by cross-referencing DEGs to immune-related genes. The immune-related EMT-lncRNAs were investigated since crosstalk between EMT and immune processes had previously been reported to influence the survival of cancer cells. When intersecting the EMT- and immune-related lncRNAs, they were left with 72 lncRNAs. Since TCGA accompanies the sample data with clinical information it also allows for the validation of these signatures by investigating the correlation between the expression of the lncRNAs of interest and patient survival and prognosis. Du et al. found that five of the 72 lncRNAs were able to predict response to treatment and survival of the patients. (Du et al., 2021) This example shows how existing data, when assembled into a database, can be used to find lncRNAs which undergo changes in their expression during EMT. These RNAs can then be further analyzed using other approaches to determine whether they have an active role in driving EMT or have predictive potential with regards to the malignancy of cancer cases.

4.4 The different approaches to investigating lncRNAs in EMT

The examples above illustrate how the role of lncRNAs in EMT can be studied through different approaches. The first is to examine individual relationships between RNAs and protein through *in vitro* studies. This type of research is able to determine the molecular mechanisms underlying the influence of lncRNAs and can identify causal relationships between regulatory factors. However, since large events, like EMT, induce global transcriptomic changes, individual study of every factor is challenging. For this reason, transcriptome-wide analyses of lncRNAs can identify key-regulators in the complex regulatory networks. Identification can be done from databases using meta-analyses to find associations between expressions of lncRNAs and cancer-events. However, associations alone tell little about the actual role of the lncRNA in the cancer-event. Therefore, validation of these results is an important step. Validation can be done by using patient data, as demonstrated by Du, et al., (2021) or experimentally as done by Liao, et al. (2017) Even after validation, the association still does not tell about the detailed regulatory mechanisms underlying the event. Thus, the different research approaches are best used in tandem. After finding lncRNAs with strong associations with EMT or predictive diagnostic potential, *in vitro* studies can be used to determine which are actively driving the event and which alter their expression as a result of EMT. This identification process as a whole can potentially identify lncRNAs with therapeutic potential due to their active role in the induction of EMT.

5. VIMENTIN AND VIM-AS1, A CASE STUDY

5.1 Vimentin and vimentin antisense 1

The complexity of lncRNA regulation and the different research approaches are well illustrated by examining the different functions of a single lncRNA. Vimentin antisense 1 (Vim-as1) is the antisense transcript of vimentin. The vimentin (*VIM*) gene encodes an intermediate filament and is a known mesenchymal marker. Increased expression of *VIM* is strongly associated with EMT. The *VIM* promoter is capable of initiating transcription to both directions, meaning it is bi-directional. This bi-directionality allows for head-to-head transcription, where two transcripts originate from the same promoter. In the case of *VIM*, the other transcript is Vim-as1. The transcription of both Vim-as1 and vimentin is initiated by the same transcription factors, meaning that transcription of vimentin will also result in Vim-as1 being transcribed, even though the expression level of Vim-as1 is 2-3-fold lower. (Boque-Sastre et al., 2015) The bi-directional promoter is presented in Figure 5. Vim-as1 is a good example of lncRNA-regulation and its involvement in EMT as after its discovery it has been found to take part in several EMT-related events.

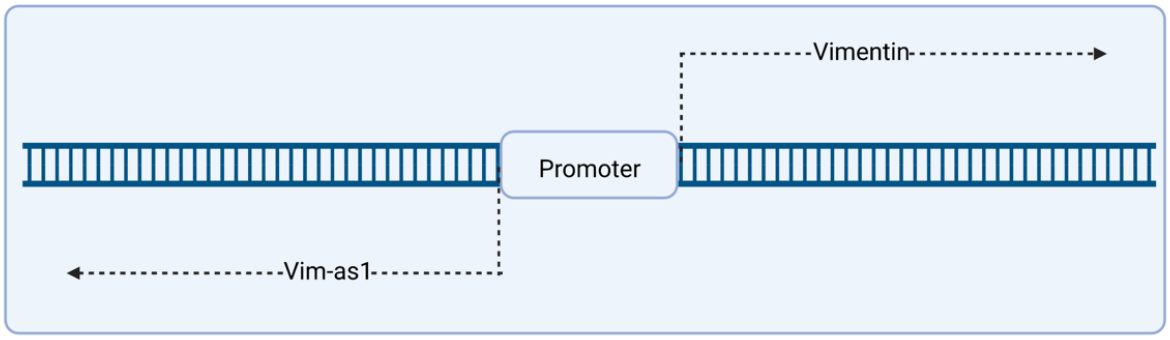

Figure 5: The bi-directional promoter of vimentin and Vim-as1. Created with BioRender.com

5.2 RNA-regulation of vimentin

Vim-as1 can exert local regulation through R-loop formation. In 2016 Boque-Sastre and colleagues showed that Vim-as1 can form R-loops and that knockdown of the transcription of Vim-as1 or disruption of the R-loop structure both lead to lowered expression of vimentin. The mechanism behind their relationship is the R-loop's ability to decondense local chromatin, making *VIM* more accessible to the transcription machinery. This relationship was verified by showing that Vim-as1 knockdown also reduces the binding of the NF-κB-pathway transcription factors and increases nucleosome occupancy. (Boque-Sastre et al., 2015) The R-loop formed by Vim-as1 is shown in Figure 6.

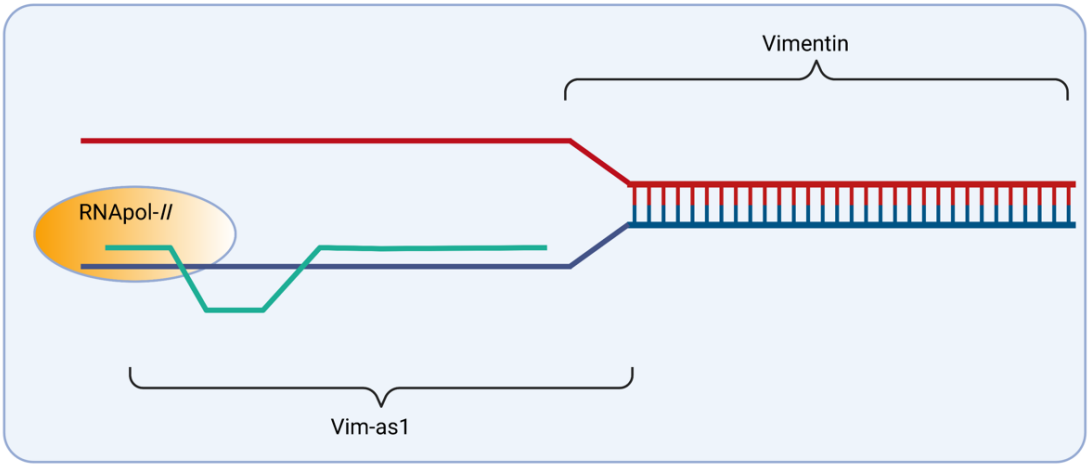

Figure 6: A stable R-loop formed by Vim-as1 makes VIM more accessible to transcription factors. Created with BioRender.com

Similar results were found in 2020 by Mehdi, et al. as they showed that the expression of Vim-as1 correlates with the expression of vimentin in lung cancer cells. However, they noted that the expression of vimentin is also governed by another lncRNA, called AGAP2-as1 (Mohebi et al., 2020), which had earlier been shown to promote metastasis in hepatocellular carcinoma by sponging mir-16-5p. (Liu et al., 2019). Interestingly, overexpression of mir-16-5p had previously been shown to downregulate the expression of vimentin in bone cancer cells (Zhang et al., 2018). These findings suggest a regulatory mechanism between AGAP2-as1 and vimentin and illustrate how the same gene can simultaneously be upregulated by two different lncRNAs, through different mechanisms.

5.3 Vimentin as1 as a ceRNA

The relation between EMT and Vim-as1 was again shown by Hajar, et al. in 2018 when they linked its increased expression to considerably lower survival of colon cancer patients. They also showed that the migration and proliferation potential of colon cancer cells was heightened with increases in the expression of Vim-as1. A direct link with EMT was established as knockdown of Vim-as1 decreased both the migration potential of the cells and decreased the expression of several mesenchymal markers, including *VIM*, *SNAIL*, and *ZEB1*. (Bardaji et al., 2018) The relationship between Vim-as1 and EMT is partially explained by Vim-as1's ability to increase the transcription of vimentin. However, Vim-as1 can have additional roles in actively driving EMT by acting as a ceRNA. In 2021, Xiong, et al. showed that Vim-as1 overexpression plays a role in the malignancy of bladder cancer. Instead of chromatin regulation, Vim-as1 was shown to promote EMT by sponging mir-665, which functions as a repressor of *ZEB1*. (Xiong et al., 2021) The ceRNA function of Vim-as1 during EMT may not be limited to regulation of ZEB1, as mir-665 has also been shown to inhibit the expression of *SNAIL*. (Wang et al., 2021) The above examples show how a single lncRNA can be involved in several mechanisms which lead to EMT. In addition, each driver of EMT can be influenced by multiple lncRNAs. This creates a tremendously complex regulatory network, through which EMT is ultimately controlled. The complexity of the system means that major events, like EMT, occur as a result of several regulatory factors working together instead of through the influence of a single factor. The arising complexity is illustrated in Figure 7 by summarizing the mentioned regulation of Vim-as1 and vimentin.

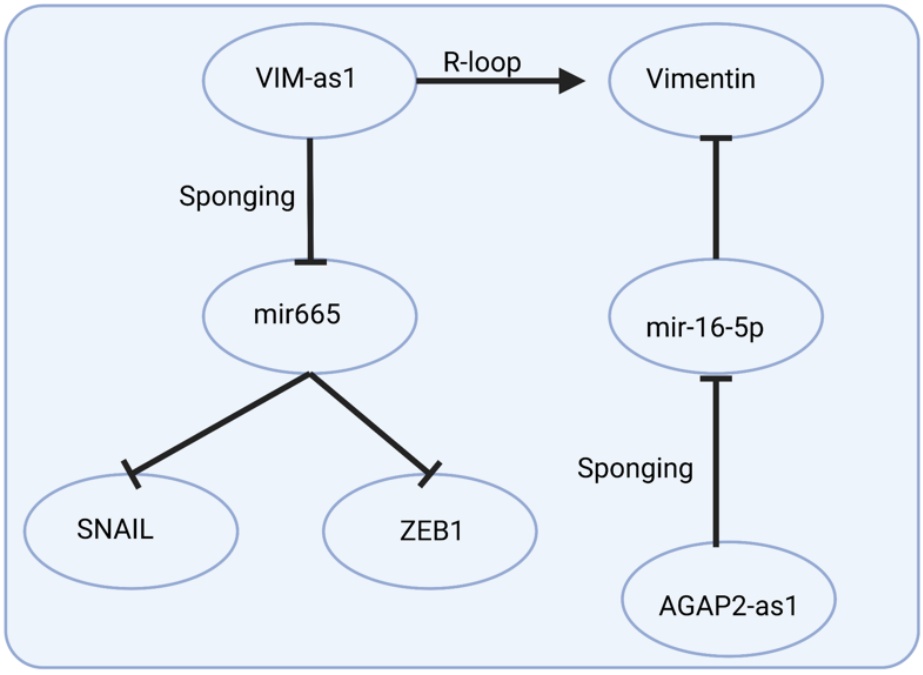

Figure 7: Vimentin can be influenced by several lncRNAs, which themselves can have many regulatory targets. Created with BioRender.com

6. DISCUSSION

As seen with Vim-as1, lncRNA-mediated regulation is a complex interplay between several actors that together can have vast effects on cancer-related cellular events, such as EMT. When considering the molecular mechanisms underlying EMT, lncRNAs are an important factor whose role must not be ignored. Detecting and preventing EMT is a major goal for cancer treatment as EMT is a central process during cancer invasion and metastasis. The effect lncRNAs have on EMT makes them a potent biomarker or even a therapeutic target for cancer. As discussed in this thesis, the complexity of the regulatory networks involving lncRNAs means that the analysis of individual lncRNAs will likely tell little about the biological reality. Instead of searching for individual indicators of EMT, it may be more feasible to assemble panels consisting of several regulators which can be analyzed simultaneously when determining the degree of EMT from samples. For this approach, the panel-assembly is a critical step. EMT-related signatures can be found through meta-analysis, as done by Du, et al. in 2021. Several lncRNA-panels have been shown to have diagnostic potential, for example a four lncRNA panel which included Vim-as1, was shown to predict the survival of lung cancer patients (Zhang et al., 2021). However, as shown by Liao, et al. in 2017, changes in lncRNA expression profile can be very different depending on the cell type and EMT induction agent. For this reason, the biomarker potential of panels assembled through statistical analysis is likely limited to certain cancer subtypes, and poor in identifying therapeutic targets. For identifying broader biomarkers, the key is identifying common changes in the lncRNA profiles across different cell-types and induction agents, as done by Liao, et al. (2017).

To be used as therapeutic targets, the molecular mechanisms underlying the changes in lncRNA must be uncovered. Only the molecular mechanism will tell whether a specific lncRNA drives EMT or whether the changes in its expression are the result of EMT. This is why detailed experiments, such as the ones presented by Parfenyev, et al. (2021), are needed in addition to more broad signature identification. Also, the function of a lncRNA must be carefully investigated since, as shown by Vim-as1, a single molecule can have effects through various mechanisms, which means that therapeutic alterations of lncRNA can alter the function of all its targets which may result in far-reaching side effects. Overall, the identification of lncRNA as biomarkers and therapeutic targets is a multi-step process and demands both global, and detailed analysis. Even though clinical cancer treatments that target EMT through lncRNAs do not exist yet, they show great potential. Importantly, lncRNAs may be used complementarily in conjunction with other methods of treatment and diagnosis.

7. CONCLUSIONS

Non-coding RNAs are RNAs that do not code for a protein. Although they were originally thought to be transcriptional noise without biological function, they have been found to have many important regulatory functions. An important event non-coding RNAs are involved in is EMT, in which epithelial cells lose their characteristic properties and gain those of mesenchymal cells. Although EMT happens as a part of normal embryonic development, it often occurs during cancers, allowing the cancer cells to break tissue morphology and invade surrounding tissues. Non-coding RNAs can be classified by their length into short and long ncRNAs. Many different types of short ncRNAs exist, but one of the most well-studied are micro-RNAs (miRNAs), 18-24 nucleotides in length. Micro-RNAs can downregulate gene expression by hybridizing with complementary mRNAs. Long non-coding RNAs (lncRNAs) can counteract miRNAs by sequestering them, preventing them from binding to mRNA. They can also regulate gene expression through chromatin structure by forming DNA-RNA hybrids, a good example of which are R-loops. As shown by vimentin antisense 1, long non-coding RNAs can have several simultaneous regulatory roles, with many RNAs often regulating the same target. Thus, the regulatory systems formed by long non-coding RNAs should be studies on several different levels, examining broad systematic changes along with more detailed research. Their involvement in epithelial-mesenchymal transition makes long non-coding RNAs a prominent possible biomarker or therapeutic target in cancer, but their complex nature means that vigorous research is needed for identification and validation.

8. REFERENCES

Bardaji, H.R., Asadi, M.H., Yaghoobi, M.M., 2018. Long noncoding RNA VIM-AS1 promotes colorectal cancer progression and metastasis by inducing EMT. European Journal of Cell Biology 97, 279–288. https://doi.org/10.1016/j.ejcb.2018.04.004

Boque-Sastre, R., Soler, M., Oliveira-Mateos, C., Portela, A., Moutinho, C., Sayols, S., Villanueva, A., Esteller, M., Guil, S., 2015. Head-to-head antisense transcription and R-loop formation promote transcriptional activation. Proceedings of the National Academy of Sciences 112, 5785–5790. https://doi.org/10.1073/pnas.1421197112

Condrat, C.E., Thompson, D.C., Barbu, M.G., Bugnar, O.L., Boboc, A., Cretoiu, D., Suciu, N., Cretoiu, S.M., Voinea, S.C., 2020. miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. Cells 9, 276. https://doi.org/10.3390/cells9020276

Dhamija, S., Diederichs, S., 2016. From junk to master regulators of invasion: lncRNA functions in migration, EMT, and metastasis. International Journal of Cancer 139, 269–280. https://doi.org/10.1002/ijc.30039

Dhanoa, J.K., Sethi, R.S., Verma, R., Arora, J.S., Mukhopadhyay, C.S., 2018. Long non-coding RNA: its evolutionary relics and biological implications in mammals: a review. Journal of Animal Science and Technology 60, 25. https://doi.org/10.1186/s40781-018-0183-7

Du, Y., Wang, B., Jiang, X., Cao, J., Yu, J., Wang, Y., Wang, X., Liu, H., 2021. Identification and Validation of a Stromal EMT Related LncRNA Signature as a Potential Marker to Predict Bladder Cancer Outcome. Frontiers in Oncology 11.

García-Muse, T., Aguilera, A., 2019. R Loops: From Physiological to Pathological Roles. Cell 179, 604–618. https://doi.org/10.1016/j.cell.2019.08.055

Hutter, C., Zenklusen, J.C., 2018. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data. Cell 173, 283–285. https://doi.org/10.1016/j.cell.2018.03.042

Liao, J.-Y., Wu, J., Wang, Y.-J., He, J.-H., Deng, W.-X., Hu, K., Zhang, Y.-C., Zhang, Y., Yan, H., Wang, D.-L., Liu, Q., Zeng, M.-S., Phillip Koeffler, H., Song, E., Yin, D., 2017. Deep sequencing reveals a global reprogramming of lncRNA transcriptome during EMT. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1864, 1703–1713. https://doi.org/10.1016/j.bbamcr.2017.06.003

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., Tamayo, P., 2015. The Molecular Signatures Database Hallmark Gene Set Collection. Cell Systems 1, 417–425. https://doi.org/10.1016/j.cels.2015.12.004

Liu, Z., Wang, Y., Wang, L., Yao, B., Sun, L., Liu, R., Chen, T., Niu, Y., Tu, K., Liu, Q., 2019. Long noncoding RNA AGAP2-AS1, functioning as a competitive endogenous RNA, upregulates ANXA11 expression by sponging miR-16-5p and promotes proliferation and metastasis in hepatocellular carcinoma. J Exp Clin Cancer Res 38, 194. https://doi.org/10.1186/s13046-019-1188-x

McCabe, E.M., Rasmussen, T.P., 2021. lncRNA involvement in cancer stem cell function and epithelial-mesenchymal transitions. Seminars in Cancer Biology 75, 38–48. https://doi.org/10.1016/j.semcancer.2020.12.012

Mohebi, M., Ghafouri-Fard, S., Modarressi, M.H., Dashti, S., Zekri, A., Kholghi-Oskooei, V., Taheri, M., 2020. Expression analysis of vimentin and the related lncRNA network in breast cancer. Experimental and Molecular Pathology 115, 104439. https://doi.org/10.1016/j.yexmp.2020.104439

Niehrs, C., Luke, B., 2020. Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol 21, 167–178. https://doi.org/10.1038/s41580-019-0206-3

O'Brien, J., Hayder, H., Zayed, Y., Peng, C., 2018. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Frontiers in Endocrinology 9.

Parfenyev, S., Singh, A., Fedorova, O., Daks, A., Kulshreshtha, R., Barlev, N.A., 2021. Interplay between p53 and non-coding RNAs in the regulation of EMT in breast cancer. Cell Death Dis 12, 1–16. https://doi.org/10.1038/s41419-020-03327-7

Ratti, M., Lampis, A., Ghidini, M., Salati, M., Mirchev, M.B., Valeri, N., Hahne, J.C., 2020. MicroRNAs (miRNAs) and Long Non-Coding RNAs (lncRNAs) as New Tools for Cancer Therapy: First Steps from Bench to Bedside. Target Oncol 15, 261–278. https://doi.org/10.1007/s11523-020-00717-x

Stavast, C.J., Erkeland, S.J., 2019. The Non-Canonical Aspects of MicroRNAs: Many Roads to Gene Regulation. Cells 8, 1465. https://doi.org/10.3390/cells8111465

Treiber, T., Treiber, N., Meister, G., 2019. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. Nat Rev Mol Cell Biol 20, 5–20. https://doi.org/10.1038/s41580-018-0059-1

Wang, W., Ying, Y., Xie, H., Li, J., Ma, X., He, L., Xu, M., Chen, S., Shen, H., Zheng, X., Liu, B., Wang, X., Xie, L., 2021. miR-665 inhibits epithelial-to-mesenchymal transition in bladder cancer via the SMAD3/SNAIL axis. Cell Cycle 20, 1242–1252. https://doi.org/10.1080/15384101.2021.1929677

Wang, Y., Hou, J., He, D., Sun, M., Zhang, P., Yu, Y., Chen, Y., 2016. The Emerging Function and Mechanism of ceRNAs in Cancer. Trends Genet 32, 211–224. https://doi.org/10.1016/j.tig.2016.02.001

Xiong, Y., Zu, X., Wang, L., Li, Y., Chen, M., He, W., Qi, L., 2021. The VIM-AS1/miR-655/ZEB1 axis modulates bladder cancer cell metastasis by regulating epithelial–mesenchymal transition. Cancer Cell International 21, 233. https://doi.org/10.1186/s12935-021-01841-y

Zhang, B., Wang, R., Li, K., Peng, Z., Liu, D., Zhang, Y., Zhou, L., 2021. An Immune-Related lncRNA Expression Profile to Improve Prognosis Prediction for Lung Adenocarcinoma: From Bioinformatics to Clinical Word. Front Oncol 11, 671341–671341. https://doi.org/10.3389/fonc.2021.671341

Zhang, H., Yang, K., Ren, T., Huang, Y., Tang, X., Guo, W., 2018. miR-16-5p inhibits chordoma cell proliferation, invasion and metastasis by targeting Smad3. Cell Death Dis 9, 1–13. https://doi.org/10.1038/s41419-018-0738-z