REVIEW ARTICLE

A link between long non-coding RNA (lncRNA) and thalassaemia: A review

Afshan SUMERA1*, Ammu K RADHAKRISHNAN2, Abdul AZIZ BABA1, Elizabeth GEORGE3

1School of Medicine, International Medical University, Bukit Jalil, 57000 Kuala Lumpur, Malaysia. 2Jeffrey Cheah School of Medicine and Health Sciences, Monash University Malaysia, 47500 Bandar Sunway, Malaysia. 3Assunta Hospital, Petaling Jaya, Selangor, Malaysia.

Abstract

The long non-coding RNAs (lncRNAs) are the most prevalent and functionally diverse member of the non-coding RNA (ncRNA). The lncRNA has previously been considered to be a form of transcriptional “noise” but recent studies have found that the lncRNA to be associated with various disease conditions. It has also been found to play important roles in various physiological processes such as haemopoiesis, where lncRNA is reported to act as a fine-tuner of this very important process. To date, the effects of dysregulated lncRNA in thalassaemia has not been fully explored. This review article focuses on the possible roles of dysregulated lncRNAs in the pathogenesis of thalassaemia.

Keywords: Thalassaemia, long non-coding RNA (lncRNA), erythropoiesis

INTRODUCTION

Thalassaemia is a progressive global disorder due to single gene defects. It is expected that there would be around 900,000 births of clinically significant thalassaemia disorders in the next 20 years.1 Thalassaemia is a quantitative disease, where the production of globin chains that are required to produce haemoglobin is decreased due to mutations in the human globin gene. Mutated genes fail to produce adequate amounts of haemoglobin, which causes the patient to present with several clinical symptoms related to lack of oxygen in the body. There are many types of thalassaemia depending on the site of mutation in the human globin gene. More recently, thalassaemic individuals are classified into two categories, which are based on the need for blood transfusion, i.e. (i) transfusion-dependent and (ii) transfusion independent.1 At present, a diagnosis of thalassaemia is made by clinical history, physical examination, complete blood count (CBC), haemoglobin (Hb) analysis by electrophoresis or high-performance liquid chromatography (HPLC). However, there is a need to improve the detection and diagnosis of thalassaemia. One approach that could be used to achieve this would be to study the irregular transcription patterns in erythrocytes from thalassaemic individuals, which could help the clinicians and scientist to understand better the pathogenesis of this disease, which is associated with defects in DNA sequences.2 This review article focuses on possible roles of dysregulated long-chain non-coding RNAs (lncRNAs) in the pathogenesis of thalassaemia.

RNA types and their function

In the early 1970s, it was discovered that the genome contains protein-coding micro-molecules known as ribonucleic acid (RNA), which are single-stranded biopolymers that form an essential element in all living cells.3 The RNA carries genetic information from the DNA stored in the nucleus of a cell, which is then translated into proteins. Initially, there were only three types of RNA were known; i.e. (i) messenger RNA (mRNA); (ii) transfer RNA (tRNA); and (iii) ribosomal RNA (rRNA).3 The tRNA and rRNA are present in an organelle found in the cytoplasm of cells, known as the ribosome. The tRNA serves as an adapter molecule as such has two functional ends. One end reads and interacts with the triplet code present in the mRNA while the other end interacts with the right amino acid. The main function of the rRNA is to attach new amino acids to the peptide chain that is
being newly synthesised.\(^4\) The mRNA carries a sequence, which serves as a template to encode amino acid sequences that form a peptide chain, required to form a protein.

The pre-mRNA or unprocessed mRNA, which is known as the “heterogeneous nuclear RNAs” (hnRNAs) changes into mature mRNA after extensive processing.\(^5\) Previously it was widely reported that half of hnRNAs are limited to the nucleus and do not contain coding sequences.\(^6\) However, in 1977, it was discovered that introns, represents just a small portion of the non-coding sequences.\(^9\) Soon after, in the 1980s, small nuclear RNA (snRNA) and small nucleolar RNAs (snoRNAs) were identified as essential players in RNA post-transcriptional processing.\(^11\) The snRNA can only be found in eukaryotes.\(^12\)

Non-coding RNA

Approximately two-thirds of the human genome is reported to be composed of “junk” or non-coding DNA\(^13\), which includes transposons, pseudogenes etc. and only the remaining one-third make up the protein-coding regions that range from 20,000-30,000 genes.\(^15\) In addition, there is a heterogeneous class of non-coding RNA (ncRNA), which forms the vast majority of the human genome was considered as “junk” or “waste” matter because these were not directly involved in the translation process.\(^11\) What these findings mean is that only a small portion of the human genome, which is transcribed into mRNA results in protein-encoding.

There are two major types of ncRNA, which are (i) small RNA (sRNA) and (ii) long non-coding RNAs (lncRNAs).\(^11\) The lncRNAs are reported to be the most prevalent and functionally diverse member of the ncRNA.\(^11\) At present, there is no single definition for lncRNAs based on their biological functions. The most widely used definition of IncRNA is based on their location and size; where an ncRNA that is more than or equal-to \(\geq\) 200 nucleotides of length is known as lncRNA while those that are less than 200 nucleotides are defined as sRNA.\(^18\) As shown in Table 1, the IncRNA can be classified into five mutually non-exclusive groups based on their genomic position while the sRNA can be classified into three groups. One way of differentiating IncRNA from sRNA is based on their function as a primary or spliced transcript.\(^17\)

The IncRNA is reported to be present in almost all species including yeast\(^19\), prokaryotes\(^20\), viruses\(^21\), plants\(^22\) and animals.\(^23,24\)

What is long non-coding RNA?

The IncRNAs are reported to be inadequately moderated among various species when compared to other well-studied RNAs\(^25\), which raises the question of whether the IncRNA has functional roles in all species; or does it have species-particular attributes. Due to its weak expression, some researchers have proposed IncRNA to be transcriptional noise.\(^27-29\) However, to date, there are several reports on their biological role in processes that range from transcription to translation\(^27-35\), protein localisation\(^36,37\), cellular structure integrity\(^38,39\), imprinting\(^40,42\), cell cycle\(^43,44\), apoptosis\(^21,45\), stem cell pluripotency\(^46\), stem cell reprogramming\(^47\) and heat shock response.\(^48,49\)

In the recent year, studies on IncRNAs appear to be gaining momentum; especially their role in diseases that occur following their up or down-regulation. Many of these studies have provided some insights to understand the biological roles of IncRNA. In the early 1970s, the IncRNA was considered to be “junk”.\(^16\) This concept was found to be untrue as more studies were carried out to investigate the role of IncRNA. For instance,

<table>
<thead>
<tr>
<th>TABLE 1: Classification of IncRNA and sRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IncRNA</strong></td>
</tr>
<tr>
<td>1. Stand-alone IncRNAs</td>
</tr>
<tr>
<td>2. Natural antisense transcripts</td>
</tr>
<tr>
<td>3. Pseudogenes</td>
</tr>
<tr>
<td>4. Long intronic ncRNAs</td>
</tr>
<tr>
<td>5. Divergent transcripts, promoter-associated transcripts, and enhancer RNAs</td>
</tr>
</tbody>
</table>

IncRNA: long non-coding RNA; sRNA: small RNA; RNA: ribonucleic acid; [Adapted from]\(^16\)
in the early 1990s, an association of lncRNA and gene expression was reported in H19 and Xist.23,51 More recent studies using advanced technologies such as microarray hybridisation and deep sequencing have found that these non-coding sequences to be an important part of the genome that are transcribed at some point during development.52-54 So, what are the known biological roles of lncRNA?

Role of lncRNA in cellular processes

To date, there is not much known about the biological functions of lncRNAs. However, their role in various cellular processes cannot be denied. For instance, lncRNA can provide potential and target recognition by folding itself into higher-order structures.55-57 Other important roles of lncRNA can be seen in gene expression, genomic imprinting, nuclear organisation and gene dosage compensation.58 Some studies have shown that the lncRNA can act as recruiters, scaffolders and tethers in gene expression.59,60 Genome-wide studies show that lncRNA can recruit chromatin-modifying factors such as polycomb repressive complex 2 (PRC2), which are needed for regulation of chromatin states by interaction.61 Another important role is to silence ribosomal DNA (rDNA) by directing the methylation process.62 In the genome, lncRNA are transcribed from a single locus, and they can control allele and locus in cis.63 During transcription, lncRNA attaches to chromatin via DNA:RNA hybridisation and can act as cis-acting molecular tethers.63 In addition, during transcription, lncRNA can function as decoyers, inhibitors and co-regulators.64 The lncRNA can also inhibit polymerase II (Pol II) directly by inhibiting the activity of the dihydrofolate reductase (DHFR) during transcription.64 Alternatively, the lncRNA could decoy transcription factors, co-regulate or co-activated several nuclear steroid receptors.65 Another role for lncRNA is in the regulation of the nuclear compartment by interacting with nuclear bodies.66 The lncRNAs can also affect mRNA destabilisation by interfering the micro-RNA (miRNA)-mediated process by hindering spliceosome formation or masking the binding sites miRNA.67

Role of lncRNA in Haemopoiesis

Haemopoiesis is an important process that must happen throughout the life of a human. This is a life-long process that continuously produces new blood cells. One of the important regulators of this process is a transcription factor known as PU.1, which is considered to be a master haemopoietic regulator.68 This protein directs myeloid and lymphoid stem cell differentiation and suppresses leukaemic proliferation. One of the lncRNA, known as anti-sense to PU.1 is reported to affect the expression of this protein.69 The presence of this lncRNA can result in haematological malignancies as it has been found to negatively regulate the translation of the PU.1 mRNA.60

The lncRNA can also act as fine-tuners of cellular processes at various stages of cell differentiation and lineage commitment; thereby regulating the fate of a cell by the positive or negative feedback loop.23,69-71 For instance, the lncRNAES1 and lncRNAES2 are known as stem cell pluripotency-associated lncRNAs as these lncRNAs can silence the expression of the Sox2 gene in human embryonic stem cells (ESC).72 The Sox2 gene is essential for ESC to maintain the pluripotent state.72 Silencing of the Sox2 gene is shown to compromise the pluripotent state of both mouse and human ESC.73 This is one of the ways that lncRNA may be able to control the fate of the cell i.e. by deciding between cell differentiation and proliferation.

One of the important lncRNA that has been studied in haematopoietic stem cells is the lncRNA-EPS (erythroid pro-survival).73 This lncRNA can activate transcription by indirect repression of apoptosis-associated speck-like protein (ASC) (also known as Pycard), which is a key modulator of caspase-mediated apoptosis.74 Indirect repression is achieved by producing transcriptional repressors.75 Over-expression of Pycard will inhibit proliferation of erythroid cells; thereby promoting apoptosis in these cells as well as interfere with their terminal differentiation and enucleation.56 The exact mechanism of repression has not yet been elucidated.

The lncRNAs have also been implicated to play a role in lymphopoiesis, where these molecules facilitate development and function of adaptive immunity by modulating differentiation of memory T-call and lymphocyte activation.76

Role of lncRNA in Erythropoiesis

Haemopoiesis is a well-regulated complicated process. This process starts when the haematopoietic stem cells (HSC) in the bone marrow are stimulated with various growth factors to differentiate into the different types of mature blood cells. Several studies have found
that differentiation of erythroid progenitors and megakaryocytes start early in the hierarchy from HSC compared to other cells.\textsuperscript{77,78}

Erythropoiesis starts with early committed precursors, which are the burst-forming unit-erythroid (BFU-E) and colony-forming unit-erythroid (CFU-E) cells. Following this, these cells will be differentiated into erythroid precursors that possess more distinct morphology. The different stages of erythroid differentiation include pro-erythroblast, basophilic, polychromatophilic and orthochromatic erythroblast. The nucleus ejects at orthochromatic erythroblast and changes into reticulocyte to accommodate more Hb within the cell.\textsuperscript{79}

To date, approximately 1100 potential IncRNA genes, which may be involved in the regulation of erythropoiesis have been identified.\textsuperscript{80} In adults, most of these IncRNAs are particularly seen in the late stages of this process, which range from accommodation of Hb within a cell to ejection of the nucleus. Among these IncRNAs, a total of 96 are found to be exclusively expressed during erythropoiesis.\textsuperscript{81} Most of these IncRNAs are associated with regulation of final stages of erythropoiesis.

All the different types of IncRNA such as intergenic, antisense, intronic, and enhancer loci, pseudogenes have been reported to play a role in erythropoiesis.\textsuperscript{81} However, the level of expression of the IncRNA in these cells varies according to different studies.\textsuperscript{80,81} Majority of foetal liver erythroid IncRNAs were expressed in adult human erythroblasts and also in foetal and adult mouse erythroblasts.\textsuperscript{78,81} The IncRNA binds to the promoter region by GATA and other transcription factors in erythroblast and this causes developmental stage-specific expression.\textsuperscript{81} The importance of inter-species orthologous region expression of IncRNA has also been studied. Researchers found that almost 85\% of coding genes expressed in mouse were also expressed in human erythroblast.\textsuperscript{80} However, the expression of IncRNA was found to be strictly species-specific because only 15\% of mouse IncRNA was observed in humans.\textsuperscript{80}

The IncRNA-EPS is a 2531 nucleotides IncRNA, which is expressed during differentiation of foetal erythroblast in mouse.\textsuperscript{74} The IncRNA-EPS is located intergenically in the spleen, bone marrow and foetal liver.\textsuperscript{74} Its expression is observed during the transition of CFU-Es to haemoglobin synthesising cells.\textsuperscript{82} In normal erythropoiesis, the levels of IncRNA-EPS are usually in an inverse relationship with Pycard, which allow cells to differentiate more.\textsuperscript{83} However, IncRNA-EPS can control apoptosis in an erythroid generation. Downregulation of IncRNA-EPS results in the increased apoptosis of erythroid progenitors by targeting Pycard, which significantly reduce the number of cells (Figure 1).\textsuperscript{82} Furthermore, knockdown of IncRNA-EPS resulted in higher levels of Pycard, thereby favouring cell death through unknown mechanisms.\textsuperscript{82} These findings strongly suggest that IncRNA-EPS nurtures erythropoiesis by inhibiting apoptosis (Fig. 1a and b).

Another important IncRNA expressed in erythropoiesis is the IncRNA Fas-antisense 1 (Fas-AS1 or Saf), which is needed for the maintenance of erythrocyte production\textsuperscript{84}. The GATA-1 binds to the promoter region of IncRNA

![FIG.1: LincRNA-EPS and erythropoiesis. (a) LincRNA-EPS represses Pycard gene to decrease apoptosis of precursors of red blood cells. (b) RNAi knockdown of LincRNA to activate intercellular caspases and cause apoptosis. (Adapt from (75))](image-url)
Saf during transcription. It is repressed during early stages of erythropoiesis, but expression increases during late stages. Other important lncRNAs have been identified in erythropoiesis such as AlncRNA-EC7, which regulates expression of erythrocyte membrane gene (SLC4A1).85

Role of lncRNA in beta-thalassaemia
To date, the diversity in the phenotypic presentation of beta-thalassaemia (beta-thal) with same mutations in a beta-globin gene is not well understood. However, the role of epigenetic regulators and modifiers including lncRNA on haemoglobin synthesis have been studied.86–88 There have not been many studies reported on the role of dysregulated lncRNA in beta-thal in humans. Many tissue-specific lncRNAs are expressed during haemopoiesis and erythropoiesis80–82,89,90. The lncRNA that may play a role in changing the expression of the human globin gene has not been studied.91 The probable mechanisms proposed so far on the role of lncRNA on beta-thal include (i) recruitment of chromatin enzymes during epigenetic process69,92; (ii) scaffolding of transcription factors99,93; and (iii) enhancing the interaction between enhancers and promoters,84,95

It has been proposed that lncRNAs may act by regulating protein-coding genes in such pathways in the emergence of histanoxia, which is caused by elevated foetal haemoglobin (HbF) levels and dysontogenesis in thalassaemia major patients.94 Literature suggests that lncRNAs may be involved in the pathogenesis of beta-thal by affecting the functions of microRNAs (miRNA).96 In normal cells, lncRNA prevents binding of miRNA, thereby suppressing the function of miRNA.70 Hence, by studying the functions of well-characterised miRNAs, it might be possible to identify the role of lncRNA in thalassaemia. Further research is warranted to study the role of lncRNA in Hb disorders.

High levels of HbF were reported to be induced by dysregulated lncRNAs and miRNAs in hereditary persistence of foetal Hb (HPFH) and beta-thal carriers.96 One of the possible mechanisms proposed for induction of high levels of HbF levels by lncRNA is through the activation of Hb-subunit epsilon-1 (HBE1) and haemopoietic cell lineage-inducible molecules.96 The lncRNA can also inhibit expression of apoptosis-inducible molecules.96 Another mechanism could be through regulation of gamma-globin gene (HBG) expression by the HBS1L-MYB intergenic long noncoding RNA (HMI-lncRNA).91 The HMI-lncRNA is studied as an enhancer RNA (eRNA) and functions by facilitating the interactions between the distal enhancer and the myeloblastosis (MYB) gene promoter.91 Downregulation of this lncRNA resulted in a significant increase of HbF levels through increased expression of HBG.91 However, the exact mechanism needs to be elucidated by future studies. Furthermore, elucidating the function of lncRNA in inducing HbF levels may help researchers to design therapeutic agents that can target this pathway in sickle cell disease and beta-thal.

Other important lncRNAs that have been recently studied in beta-thal include (i) metastasis-associated lung adenocarcinoma transcript 1 (MALAT1); (ii) myocardial infarction associated transcript (MIAT) and (iii) antisense non-coding RNA in the inhibitors of CDK4 (INK) locus (ANRIL). Previously, MALAT1, which is expressed in normal cells and tissues, was studied mainly in lung cancers metastasis.97 MIAT is involved in RNA splicing and regulation of gene expression98, is implicated in disorders of the microvasculature99,100 while ANRIL’s role is reported to be epigenetic silencing of other genes in its’ cluster.101 A recent study showed that differential expression of these three important lncRNAs (MALATI, MIAT and ANRIL) in beta-thal and healthy individuals, strongly suggests a possible role for these lncRNAs in disease pathophysiology and phenotypic features.101 Some of the important lncRNAs associated with the pathogenesis of thalassaemia are summarised in Table 2.

Potential roles of lncRNA in other diseases
Dysregulated lncRNAs have been linked with numerous human diseases, including malignancies, such as neoplasms arising from liver, ovaries and kidney breast cancer.88 It appears that dysregulation of lncRNA favours tissue invasion and metastasis by neoplastic progression and can cause malignant transformation. One example is lncRNA p15AS, which silences p15 gene by initiating chromatin condensation, which results in acute haematological malignancies such as acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL).102 Recently it was reported that dysregulated lncRNAs are involved in the development of malignant T-lymphocytes in Mycosis fungoides and Sézary syndrome.103 Differential expression of lncRNA
<table>
<thead>
<tr>
<th>LncRNA</th>
<th>Normal function</th>
<th>Effect /role in thalassaemia</th>
<th>Mechanism</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR_001589, NR_120526 and T315543</td>
<td>Facilitating the interactions between the distal enhancer and the myeloblastosis (MYB) gene promoter</td>
<td>High levels of HbF</td>
<td>Activation of haemoglobin subunit epsilon-1 (HBE1) and haemopoietic cell lineage-inducible molecules</td>
<td>Lai K et al. 96</td>
</tr>
<tr>
<td>HMI-lncRNA (MYB enhancer RNA)</td>
<td>Regulates erythroid cell proliferation, maturation, and foetal haemoglobin (HbF) expression.</td>
<td>High levels of HbF</td>
<td>Downregulation increased gamma-globin gene (HBG) expression in erythroid cells</td>
<td>Morrison TA et al. 91</td>
</tr>
<tr>
<td>Metastasis-associated lung adenocarcinoma transcript 1 (MALAT1)</td>
<td>Splicing regulation</td>
<td>MALAT1 over-expression in thalassaemia as an adaptive response to increased DNA damage, one of the major players in thalassaemia phenotype and complications</td>
<td>MALAT1 expression is part of an adaptive response to hypoxia</td>
<td>Fakhr-Eldeen A et al. 101</td>
</tr>
<tr>
<td>myocardial infarction associated transcript (MIAT)</td>
<td>RNA splicing and regulation of gene expression</td>
<td>Upregulated expression may be associated to the increased severity of β-thalassaemia,</td>
<td>Increased MIAT expression in β-thal is associated with endothelial dysfunction</td>
<td>Tsuiji H et al. 98</td>
</tr>
<tr>
<td>antisense non-coding RNA in the INhibitors of CDK4 (INK) locus (ANRIL)</td>
<td>Epigenetic silencing of other genes in its’ cluster</td>
<td>Haematopoietic differentiation, proliferation, and stress response</td>
<td>Not known</td>
<td>Fakhr-Eldeen A 101</td>
</tr>
<tr>
<td>DQ583499, X-inactive specific transcript (Xist), lincRNA-TPM1, MRFS16P, and lincRNA-RUNX2-2</td>
<td>Not known</td>
<td>Epigenetic regulators and modifiers on Hb synthesis</td>
<td>Abnormal expression levels of lncRNAs and mRNA in β-thalassaemia cases may be correlated with its various clinical phenotypes</td>
<td>Ma J et al. 88</td>
</tr>
</tbody>
</table>

Intronic RNA: long non-coding RNA
HBF: Foetal haemoglobin
was also observed in several in rare inherited diseases like facioscapulohumeral muscular dystrophy, Prader-Willi and haemolytic anaemia, elevated liver enzymes, low platelet count (HELLP) syndromes.

**Conclusion and Future work**

The roles played by lncRNAs as regulators of multiple cellular processes, biomarkers for disease prognosis and potential therapeutic targets cannot be denied. In humans, many types of anaemia are associated with an increased rate of apoptosis of erythrocyte precursors in the bone marrow. If research can find suitable ways to silence the expression of these pro-apoptotic pathways, it would be possible to design and develop better drugs to treat anaemia secondary to thalassaemia, which is resistant to EPO therapy. Exploring into greater depth about the LncRNA-EPS–Pycard axis could also aid the development of pharmacological agents that can manipulate this pathway to impact erythropoiesis.

**Authors’ contribution:** Sumera authored the manuscript, Radhakrishnan, Aziz Baba and George reviewed the manuscript.

**Conflict of interest:** The authors declare no conflict of interest.

**REFERENCES**


