RESEARCH ARTICLE

Long Noncoding RNAs *TYMSOS*, *VASH1-AS1*, and *LINC01001* Expressions as Biomarkers of β-thalassaemia Severity Among the Malaysian Patients

Norafiza Mohd Yasin[1], Nur Fatin Aqilah Raman[1], Siti Aishah Sulaiman[2,*], Nurhairulamri Kosnan[1], Nor Azian Abdul Murad[2], Ermi Neiza Mohd Sahid[1], Yuslina Yusoff[1], Hishamshah Mohd Ibrahim[3], Zulaiha Muda[3], Ezalia Esa[1]

[1]Haematology Unit, Cancer Research Centre, Institute for Medical Research (IMR), National Institutes for Health, Persiaran Setia Murni, Setia Alam, 40170 Shah Alam, Selangor, Malaysia
[2]UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Jalan Yacob Latif, Bandar Tun Razak, 56000 Cheras, Wilayah Persekutuan Kuala Lumpur, Malaysia
[3]Paediatric Department, Women and Children Hospital, Tunku Azizah Hospital (HTA), Lot 25, Jalan Raja Muda Abdul Aziz, Kampung Baru, 50300 Kuala Lumpur, Malaysia

*Corresponding author. Email: sitiaishahsulaiman@ukm.edu.my

Received date: Jul 21, 2025; Revised date: Sep 2, 2025; Accepted date: Sep 11, 2025

Abstract

BACKGROUND: β-thalassaemia has heterogeneous disease severities ranging from mild to trait, and major. Long noncoding RNAs (lncRNAs) are known to regulate microRNAs (miRNAs) and genes, possibly modifying the disease phenotypes. However, limited data exist on the lncRNAs in β-thalassaemia, and no study has yet addressed this gap in Malaysia. Therefore, this study aimed to identify the expression profile of lncRNAs in β-thalassaemia major and trait among Malaysians.

METHODS: Case-control study was conducted in two phases at Tunku Azizah Women and Children's Hospital and the Institute of Medical Research, from September 2019 to November 2021. Total of 141 individuals were recruited, comprising β-thalassemia major (MAJOR, n=11), β-thalassemia trait (TRAIT, n=17), and healthy controls (CON, n=113). All participants were genotyped for thalassaemia and assessed for their haemoglobin and red blood cells (RBC) indices. In the first phase, discovery of lncRNA was performed using microarray, and differential expression of lncRNAs (DEL) in MAJOR, TRAIT, and CON groups was identified. Significant lncRNAs were subjected to lncRNA-miRNA prediction, and gene ontology and biological pathway were analyzed. In validation phase, six potential lncRNAs were further validated via using lncRNA polymerase chain reaction (PCR) custom array.

RESULTS: Total of 364 DELs were identified in MAJOR group, and 128 DELs were identified in TRAIT group. Between the MAJOR and TRAIT groups, 100 DELs were dysregulated in MAJOR group. Two molecular networks comprising the lncRNA interactions with miRNAs were identified and associated with traits and major phenotypes, resulting in six potential lncRNAs for validation. Among these six lncRNAs, three lncRNAs (*TYMSOS*, *VASH1-AS1*, and *LINC01001*) were reduced in the MAJOR group (fold change (FC)=-6.67, p=0.026; FC=-8.33, p=0.022; and FC=-8.33, p=0.021, respectively).

CONCLUSION: Expressions of *TYMSOS*, *VASH1-AS1*, and *LINC01001* lncRNAs were altered differently between β-thalassaemia major and trait patients. Therefore these lncRNAs may serve as novel biomarkers for β-thalassaemia disease severity in Malaysian population.

KEYWORDS: β-thalassaemia, lncRNAs, miRNAs, severity, major, trait, molecular network

Indones Biomed J. 2025; 17(5): 436-46

Introduction

β-thalassaemia is an autosomal recessive disorder caused by mutations in the β-globin (*HBB*) gene, resulting in abnormal haemoglobin production, ineffective erythropoiesis, and a shortened lifespan of red blood cells (RBC).(1) Until now, more than 950 mutations have been reported within the *HBB* gene and its flanking regions.(2) In Malaysia, β-thalassaemia is among the most prevalent hereditary diseases (~5% carriers) and a major public health problem.(3) Therefore, significant efforts have been made to identify the molecular changes within the *HBB* gene among Malaysians in order to reduce the burden and improve patient outcomes.

Previous studies have identified mutations associated with β-thalassaemia patients in Malaysia.(3,4) Among them, the most common mutations in Malaysian Malays are Hb Malay (codon 19 (A>G) *HBB*:c.59A>G, IVS-I-1 (G>T) *HBB*:c.92+1G>T, IVS-I-5 (G>C) *HBB*:c.92+5G>C, polyadenylation signal (polyA) (AATAAA>AATAGA) *HBB*:c.*112A>G, and Hb E thalassaemia, (*HBB*:c.79G>A). On the other hand, Malaysian Chinese patients are often presented with codons 41/42 (–TTCT) *HBB*:c.126_129delCTTT, IVS-II-654 (C>T) *HBB*:c.316-197C>T, -28 (A>G) *HBB*:c.-78A>G, and codon 17 (A>T) *HBB*:c.52A>T.(3,4) The overall data also showed that the most commonly known form of β-thalassaemia in Malaysia is the combination of Hb E and β-thal (Hb E/β-thal), which accounted for 34.37% of the patients (4), and is quite common among South-East Asians (5-7). However, the clinical manifestations of β-thalassaemia are highly variable, ranging from the mild thalassaemia trait to severe thalassaemia major.(8) This phenotype heterogeneity poses significant challenges in prevention strategies, prenatal screening, clinical diagnosis, and treatment planning.

Various factors contributed to genotype-phenotype heterogeneity in β-thalassaemia. The degree of chain imbalance and the mechanism of β-thalassaemia mutations are primary factors in determining the β-thalassemia phenotypes.(4,8,9) Due to these complex genetic factors and differential phenotypes, many studies investigated the additional genetic modifiers of β-thalassaemia.(10) One such example is the long noncoding RNAs (lncRNAs), which bind to microRNAs (miRNAs) and act as sponges to prevent them from exerting their function.(11) This lncRNA-miRNA-gene network, also known as competitive endogenous RNA (ceRNA), which is a system of RNA interactions and co-regulatory networks, where both lncRNAs and genes compete for binding to specific miRNAs, resulting in an increase or decrease in gene expression.(10) One of the earliest studies of lncRNAs in β-thalassaemia reported that 605 lncRNAs were upregulated, and 257 lncRNAs were downregulated in β-thalassemia minor compared to controls.(12) In the previous study, one miRNA, miR-486-3p (a well-known regulator of γ-globin expression (12), was found to be associated with six lncRNAs in a negative relationship. Another study compared β-thalassemia major, β-thalassemia minor, and healthy controls and found that more than 3,000 lncRNAs were dysregulated in β-thalassemia major patients, and over 1,000 lncRNAs were dysregulated in β-thalassemia minor patients.(13) Among these dysregulated lncRNAs, lncRNA *XIST*, *lincRNA-TPM1*, and *lincRNA-RUNX2-2* were upregulated, and lncRNA *DQ583499* and *MRFS16P* were downregulated in β-thalassemia individuals.(13) These findings suggest that lncRNAs could play a role in determining the severity of β-thalassemia. Unfortunately, none of these studies were performed on Malaysian patients. Therefore, this pilot study aimed to determine the lncRNA expression profile and the potential molecular network of lncRNAs and miRNAs in Malaysian β-thalassemia patients, whose preliminary clinical and genotypic data had been published previously.(9) Identification of these lncRNAs and their potential roles in determining disease severity may help unravel the underlying molecular landscape of β-thalassemia and its associated phenotypes, thereby enabling the discovery of new biomarkers for disease severity assessment.

Methods

Study Design and Participants

This case-control study was a pilot study designed for two phases of investigation, and was conducted at the Tunku Azizah Women and Children's Hospital and the Institute of Medical Research, Malaysia, from September, 2019 to November, 2021. A total of 141 individuals were recruited, including β-thalassemia major (MAJOR, n=11), β-thalassemia trait (TRAIT, n=17), and healthy controls (CON, n=113). All participants were genotyped for thalassaemia, and details on their clinical phenotypes, including age, haemoglobin (Hb) analysis, transfusion requirements, spleen size, and extramedullary changes, were collected from their clinical records. The patients were considered β-thalassemia major when they were transfusion-dependent (requiring more than six transfusions yearly), had the β0 mutation, and had an enlarged spleen. Whereas, β-thalassemia trait patients were diagnosed based

on Hb analysis and their β-globin mutations. Patients with other systemic diseases, such as liver disorders (cirrhosis), chronic renal failure, hypothyroidism, and hyperthyroidism, were excluded from this study. Healthy controls are included in the study if they have a normal complete blood count (CBC), normal Hb, and no α- and β-thalassemia genotypes, as well as no other existing genetic diseases and a family history of thalassemia. This study was conducted in accordance with the guidelines of the Declaration of Helsinki and was approved by the Medical Research and Ethics Committee of the IMR (Ethics No.: NMRR-19-2280-48428, dated September 20, 2019). Written informed consent was obtained from all subjects involved in the study. The first phase of this study was the lncRNA discovery phase that included a total of nine participants comprising the healthy controls (CON, n=3), β-thalassemia major with homozygous Codon 41/42 (-TTCT) mutation (MAJOR, n=3), and β-thalassemia trait patients with heterozygous Codon 41/42 (-TTCT) mutation (TRAIT, n=3) from the all 141 participants. The second phase involved validating the significant lncRNAs identified in the first phase in the remaining participants, comprising the MAJOR (n=8), TRAIT (n=14), and CON (n=110) groups.

Analysis of Blood Parameters

All participants underwent analysis of their CBC and Hb levels. The CBC was measured via the automated haematology analyser. The Hb analysis was performed via the peripheral blood film, capillary electrophoresis (SEBIA, Lisses, France), via the HPLX (Bio-Rad Laboratories, Hercules, CA, USA), and Hb electrophoresis (SEBIA), according to the standard protocol for thalassemia diagnosis in our centre.(14) The cut-off values for mean corpuscular haemoglobin (MCH) and mean corpuscular volume (MCV) were <25 pg and <80 fL, respectively.

DNA and RNA Isolation

The whole blood was used to isolate DNA for thalassemia genotyping. The DNA was isolated using the QIAamp DNA Blood Mini Kit (Qiagen Biotechnology, Hilden, Germany) as recommended by the manufacturer. The thalassemia genotyping was done as previously described (14) as part of the standard protocol in thalassemia diagnosis. The total RNA was isolated from whole blood samples using miRNeasy Serum/Plasma Kit (Qiagen Biotechnology) according to the manufacturer's protocol. The RNA quality and RNA integrity number (RIN) were measured using the Agilent RNA 6000 Pico kit on the Agilent 2100 Bioanalyzer System (Agilent Technologies, Santa Clara, CA, USA). Only RNA samples with a RIN score of ≥8 were used for the downstream analysis.

Differential lncRNA Expression

The discovery phase measured the expression of lncRNAs via the Agilent SurePrint G3 Human Gene Expression v3 8x60K microarray kit (Agilent Technologies) as recommended by the manufacturer. Raw data were extracted using the Agilent Feature Extraction Software (Agilent Technologies), and the raw data were normalised using the quantile normalization method via GeneSpring v.14 (Agilent Technologies), as recommended by the manufacturer. The differential expression analysis was performed using the normalized expression data via the online analysis tool ExAtlas (http://lgsun.grc.nia.nih.gov/exatlas). Three comparison analyses were performed: 1) MAJOR *vs.* CON, 2) TRAIT *vs.* CON, and 3) MAJOR *vs.* TRAIT groups. Differentially expressed lncRNAs (DELs) were selected based on the cut-off values of fold change (FC) >4 and adjusted *p*-value (Benjamin-Hochberg false discovery rate)<0.05.

Prediction lncRNA-miRNAs Relationship

All DELs were confirmed with the official lncRNA identification based on two lncRNA databases, LNCipedia (15) and NONCODE (16). Using lncRNA-miRNA database LncBook v2 (17) and miRNet (18), the predicted miRNA targets of each significant DELs were determined. These predicted miRNAs were then subjected to a Venn diagram analysis via the EVenn tool (19) to identify the commonly predicted miRNAs among the DELs. The final list of predicted miRNAs was subjected to Gene Ontology (GO) enrichment and biological pathway analyses using g:Profiler (20), based on three known databases: the Kyoto Encyclopedia of Genes and Genomes (KEGG), REACTOME, and WikiPathways. An adjusted $p<0.05$ was used to select the most significant GO and biological pathways.

Construction of the lncRNA-miRNA Network

The list of DELs and associated predicted mRNAs was also used to identify the co-regulatory network between the molecules via miRNet (19) and visualized using Cytoscape v.3.10.3 (21). The cytoHubba plugin was used to calculate the degrees, betweenness, and closeness centralities of the constructed networks.(20)

Validation of the Significant DELs

The expression of selected DELs was confirmed using Real-Time PCR array panel via the QuantiNova LNA

lncRNA PCR custom panel (Catalogue No.: SBCA96110 & SBCA96404 Kit, Qiagen Biotechnology, Hilden, Germany) on a StepOne Plus real-time PCR system (Applied Biosystems, Massachusetts, USA), as recommended by the manufacturer. Briefly, the custom array plate was run at the recommended cycle conditions (95°C of initial heat activation for 2 min, followed by 45 cycles of two-step cycling (95°C of denaturation for 5 seconds and 60°C of annealing/extension for 10 seconds). The raw data were analysed via the GeneGlobe data analysis tool (https://geneglobe.qiagen.com/us/), last accessed on January 10, 2024. The differential expression was determined using the $2^{-\Delta\Delta CT}$ method, and the *p*-value significance was set at <0.05 for specific group comparisons, as recommended by the manufacturer. The Glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), Beta-actin (*ACTB*), and Beta-2 microglobulin (*B2M*) genes were used as endogenous reference genes.

Statistical Analysis

Statistical analysis was performed using SPSS Statistics v.22.0 (IBM Corporation, Armonk, New York, USA) using the ANOVA analysis with specific comparisons between 1) MAJOR *vs.* CON, 2) TRAIT *vs.* CON, and 3) MAJOR *vs.* TRAIT groups. Statistical significance was set at a $p<0.05$. All other graphs and plots were generated using Prism v.8.2.0 (GraphPad Software, San Diego, USA).

Results

Observed DELs in β-thalassaemia

A summary of the patients involved in the discovery of lncRNA microarray analysis was listed in Table 1. Although the discovery sample size was small, significant differences were observed in age and RBC indices (RBC, Hb, MCV, MCH, and hemoglobin analysis (HbA, HbA2, and HbF) between the groups ($p<0.05$). The summary of differentially expressed lncRNAs (DELs) with the top five lncRNAs in each comparison was listed in Table 2. A total of 93 upregulated and 271 downregulated DELs were identified in the MAJOR group compared to the CON group. In a comparison between the TRAIT and CON groups, there were 10 upregulated and 118 downregulated DELs in the

Table 1. Summary of the clinical information of the individuals used in the microarray analysis.

Clinical Information	CON (n=3)	TRAIT (n=3)	MAJOR (n=3)	*p*-value
Age (mean±SD)	35.67±5.03	42.67±7.51	18.50±4.95	0.020
Gender, n (%)				
Male	1 (11.11)	-	1 (11.11)	n.s.
Female	2 (22.22)	3 (33.33)	2 (22.22)	
Ethnicity, n (%)				
Malay	3 (33.33)	2 (22.22)	2 (22.22)	n.s.
Chinese	-	1 (11.11)	1 (11.11)	
RBC ($\times10^6$/uL)	4.62±0.60	5.97±0.47	1.90±0.00	0.007
Hb (g/dL)	13.63±1.30	10.03±1.86	4.10±0.00	0.002
Hct (%)	40.10±2.88	37.13±4.02	-	n.s.
MCV (fL)	87.33±4.73	62.00±5.57	68.00±0.00	0.009
MCH (pg)	29.20±1.11	16.80±1.75	22.40±0.00	0.001
MCHC (g/dL)	33.47 ±1.02	27.10±4.26	32.80±0.00	n.s.
RDW-CV (%)	13.73±0.51	21.30±7.07	30.30±0.00	n.s.
CE/HPLX (%)				
HbA	97.00±0.36	94.90±0.85	0.40±0.00	0.000
HbA2	2.87±0.25	5.20±0.71	2.80±0.00	0.022
HbF	0.04±0.00	0.05±0.00	101.60±1.27	0.013

Abbreviation: SD = Standard deviation, NS = Not significant, RBC = Red blood cell, Hb = Hemoglobin, Hct = Hematocrit, MCV = Mean corpuscular volume, MCH = Mean corpuscular hemoglobin, MCHC = Mean corpuscular hemoglobin concentration, RDW = Red cell distribution width, HbA = Hemoglobin A, HbA2 = Hemoglobin A2, HbF = Hemoglobin F.

Table 2. List of the top five most dysregulated lncRNAs identified in β-thalassaemia major, β-thalassaemia trait, and healthy controls comparisons during the discovery lncRNA microarray phase.

Comparison Analysis	LncRNAs	Log_2 Fold Change
MAJOR vs. CON		
93 DELs are upregulated	*LNC-PDCD11-2*	4.93
	LINC01907	4.93
	CDRT3	4.91
	LINC00999	4.64
	LNC-PTPRU-1	4.61
271 DELs are down-regulated	*LINC00963*	-11.3
	LNC-WWC2-1	-9.71
	FAM74A4	-9.43
	FAM239B	-8.59
	TEX41	-8.34
TRAIT *vs.* CON		
10 DELs are upregulated	*LOC101927556*	2.45
	LNC-SEMA6A-5	2.43
	LNC-C8A-1	2.37
	LNC-IGIP-1	2.36
	LNC-TTC26-2	2.33
118 DELs are downregulated	*LINC00963*	-7.26
	FAM74A4	-6.05
	TEX41	-6.02
	LOC284454	-5.91
	LNC-WWC2-1	5.68
MAJOR *vs.* TRAIT		
78 DELs are upregulated	*LNC-GPR27-1*	5.83
	CDRT3	5.10
	LINC01907	4.95
	POT1-AS1	4.87
	LINC00999	4.61
22 DELs are downregulated	*PCBP1-AS1*	-3.14
	LNC-DNAJC8-1	-3.09
	CASC15	-3.04
	LNC-PFKP-11	-2.99
	C1orf220	-2.89

Abbreviation: DELs = Differentially expressed lncRNAs, MAJOR = β-thalassaemia major, TRAIT = β-thalassaemia trait, and CON = healthy controls.

TRAIT group. In the last comparison of MAJOR *vs.* TRAIT groups, a total of 78 upregulated and 22 downregulated DELs were identified in the MAJOR group compared to the TRAIT group.

Multiple overlapping DELs exist between the comparisons, but none are common for all three comparisons (Figure 1A). Among these overlapped lncRNAs, 92 lncRNAs were shared between the TRAIT *vs.* CON and MAJOR *vs.* CON comparisons. Since these lncRNAs were significant in both thalassaemia groups compared to healthy controls, they may be associated with thalassaemia, regardless of disease severity. On the other hand, 28 lncRNAs were only present in the TRAIT group compared to the CON group, but not in the other comparisons, which may indicate that these lncRNAs are specifically associated with the trait phenotype. Between the MAJOR *vs*. CON and MAJOR *vs.* TRAIT comparisons, we identified 51 common, significantly expressed lncRNAs. Whereas 40 lncRNAs were only specific to the MAJOR group compared to the TRAIT group, which were not present in other comparisons. These 91 lncRNAs (51 + 40) may explain the major phenotype since they are altered in the MAJOR group only, even when compared to the TRAIT group. Eight lncRNAs were common between TRAIT *vs.* CON and MAJOR *vs.* TRAIT comparisons. From these eight lncRNAs, seven of them (*LOC101927556*, *LNC-SEMA6A-5*, *LNC-C8A-1*, *LNC-IGIP-1*, *EXTL3-AS1*, *LNC-ACACA-1*, and *LNC-HMCN1-2*) were upregulated in the TRAIT group (Figure 1B) and yet were downregulated in the MAJOR group. A similar pattern was also observed in the last lncRNA, *LNC-INPPL1-3*, which was reduced in the TRAIT group, yet its expression was higher in the MAJOR group. These unique differential expression patterns may explain the differences in severity, as these lncRNAs were altered differently in the TRAIT group compared to the MAJOR group.

Significant lncRNAs and Their Predicted miRNA Targets

Due to the limited information available on the function of lncRNA and existing databases, a prediction of the target miRNAs that could bind to DELs was performed. The lncRNA-miRNA relationship is a well-established method to identify the biological functions of these DELs.(22) From the 92 lncRNAs that are significant in both comparisons between MAJOR *vs*. CON and TRAIT *vs.* CON, 303 miRNAs were predicted to bind to these 92 lncRNAs. These 303 miRNAs were then subjected to Gene Ontology (GO) and biological pathway analyses to determine their functions (Figure 2A). A total of 24 biological processes (BP) were significant, with the top three processes being miRNA-mediated post-transcriptional gene silencing, regulatory ncRNA-mediated post-transcriptional gene silencing, and post-transcriptional gene silencing (Supplementary Table 1). For the cellular component (CC) analysis, four CCs were significant: the RISC complex, RNAi effector complex, ribonucleoprotein complex, and protein-containing complex. No molecular function (MF) was enriched for this set of miRNAs. The biological pathway analysis

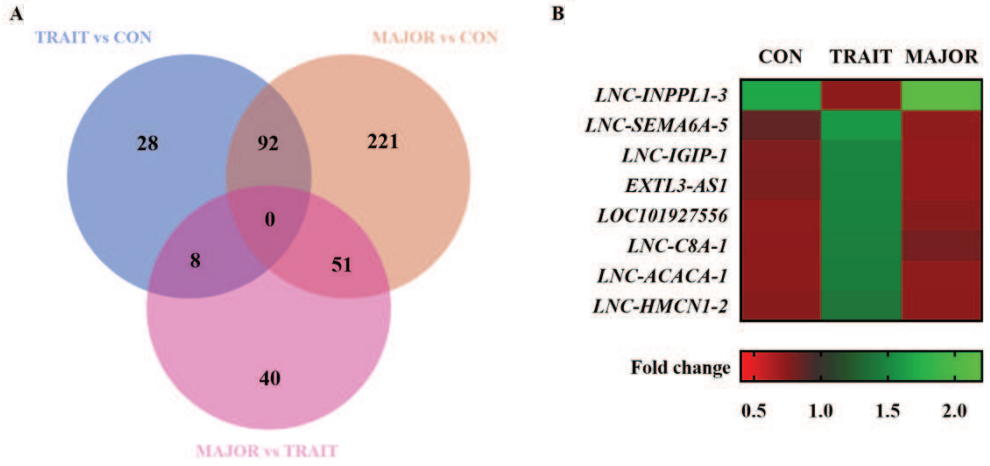

Figure 1. Summary of the DELs among the three comparison analyses. A: The Venn diagram shows the DELs in three comparisons: TRAIT *vs.* CON, MAJOR *vs.* CON, and MAJOR *vs.* TRAIT. B: Heatmap showing the expressions of eight DELs shared between the TRAIT *vs.* CON and MAJOR *vs.* TRAIT comparisons.

was performed based on three known databases: KEGG, REACTOME, and WikiPathways. Only one biological pathway, miRNAs in cancer, was significant based on the KEGG database, whereas two biological pathways were significant based on WikiPathways, which are the miRNAs involved in DNA damage response and 2q21.1 copy number variation syndrome (Supplementary Table 1).

For the 28 significant lncRNAs specific to the trait phenotype, 27 miRNAs were predicted as common targets for these lncRNAs. However, no GO and biological pathways were enriched for these predicted miRNAs. Similarly, from the eight lncRNAs (Figure 1B) that may explain the severity, seven common miRNAs (hsa-miR-6878-3p, hsa-miR-1205, hsa-miR-10394-5p, hsa-miR-6849-3p, hsa-miR-1224-3p, hsa-miR-5584-5p, and hsa-miR-6873-3p) were predicted to bind to these lncRNAs. However, no GO enrichment and biological pathways were enriched for these miRNAs.

For the last set of 91 lncRNAs associated with the major phenotype, 245 target miRNAs were predicted and subjected to GO and biological pathways analyses to determine their functions (Figure 2B). No GO MF was enriched for this set of miRNAs, though 24 BP were significant, with the top three processes being the post-transcriptional gene silencing, miRNA-mediated post-transcriptional gene silencing, and regulatory ncRNA-mediated post-transcriptional gene silencing (Supplementary Table 2). For the CC analysis, four CCs were significant: the RISC complex, RNAi effector complex, ribonucleoprotein complex, and protein-containing complex. The biological pathway analysis revealed that only one biological pathway, specifically miRNAs in cancer, was significant according to the KEGG database. Whereas, five biological pathways were significant based on WikiPathways: the miRNAs involved in DNA damage response, cell differentiation expanded index, glial

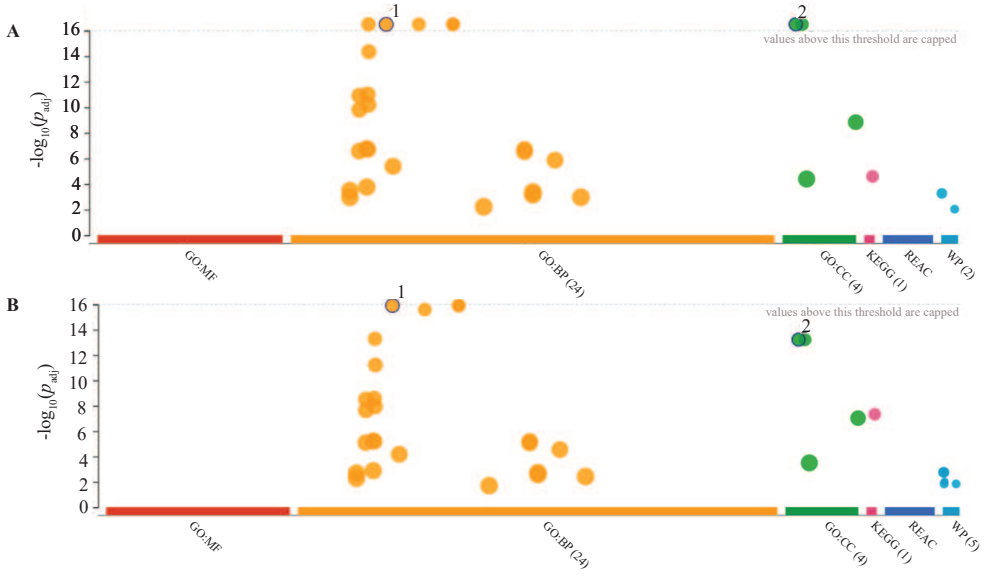

Figure 2. Summary of the GO and biological pathway analyses of the predicted microRNA targets of the DELs generated by the go:Profiler tool. A: GO and biological pathway analyses of the predicted microRNAs from the significant DELs shared between the TRAIT *vs.* CON and MAJOR *vs.* CON comparisons. B: GO and biological pathway analyses of the predicted microRNAs from the significant DELs specific to the MAJOR group.

cell differentiation, cell differentiation index, and 2q21.1 copy number variation syndrome (Supplementary Table 2).

LncRNA-miRNA Network

To further identify the DELs that could explain the disease severity differences between the TRAIT and MAJOR groups, their DELs and predicted miRNAs were analyzed for their interactions in a molecular network using miRNet and visualized via Cytoscape. Focusing on the DELs that are altered in the TRAIT group, 45 miRNAs and seven lncRNAs were identified in a molecular network (Figure 3A). This molecular network was further analyzed for a biological

Figure 3. Molecular networks showing the interactions between the key lncRNAs and their predicted target miRNAs. The molecular network is specific to lncRNAs and microRNAs associated with Trait phenotypes (A) and Major phenotypes (B). The blue colour represents lncRNAs, and the orange colour represents predicted miRNAs.

pathway analysis to confirm its interactions in disease (Table 3). Fifteen biological pathways were enriched for this TRAIT molecular network, with the top three pathways being Cell Proliferation, Granulopoiesis, and Anti-Cell Proliferation. Interestingly, one pathway, Hematopoiesis, was enriched by this network, with the involvement of 15 miRNAs, including the let-7 family, miR-1, miR-15a, and the five key lncRNAs of *VASH1-AS1*, *PCBP1-AS1*, *TYMSOS*, *EXTL3-AS1*, and *LINC00707* (Figure 3A).

Focusing on the DELs altered in the MAJOR group, 76 miRNAs and 25 lncRNAs were constructed in a molecular network (Figure 3B). A further assessment of this molecular network biological pathway revealed that 11 pathways were significant (Table 3). The top three significant pathways were Cell Differentiation, Anti-Cell Proliferation, and Cell Division. Interestingly, one pathway, Glucose metabolism, was only enriched by this network, involving seven miRNAs, including the let-7 family, miR-34a-5p, and three key lncRNAs: *PCBP1-AS1*, *LINC01001*, and *VASH1-AS1* (Figure 3B).

Validation of the Significant lncRNAs in β-thalassaemia

Six potential lncRNAs (*VASH1-AS1*, *PCBP1-AS1*, *EXTL3-AS1*, *TYMSOS*, *LINC01001*, and *LINC00707*) from the molecular networks were further validated in the validation cohort.(23) In the first comparison (TRAIT *vs.* CON), all the selected lncRNAs were upregulated in the TRAIT individuals ($p<0.05$) (Table 4). Only one lncRNA, TYMSOS, was significant in the MAJOR individuals compared to healthy controls (FC=4.71, $p=0.043$). A specific comparison between the MAJOR and TRAIT individuals showed that three lncRNAs (*VASH1-AS1*, *TYMSOS*, and *LINC01001*) were significant, but their expressions were down-regulated in MAJOR individuals compared to TRAIT individuals.

Discussion

This study is the first pilot investigation of the long non-coding RNA (lncRNA) expression profile in Malaysian β-thalassaemia major and trait patients. Previous studies have shown that various lncRNAs are implicated in β-thalassaemia patients (13,24,25), although these studies were performed in different populations. Despite the limited information available on lncRNAs and their roles in β-thalassaemia, the regulatory roles of lncRNAs have been widely reported in other diseases, such as cancer, diabetes, and other blood diseases.(24-26) Therefore, there is a need to identify the lncRNAs associated with β-thalassaemia, particularly in the population with a high prevalence of β-thalassaemia.

In this study, we identified six potential lncRNAs (*VASH1-AS1*, *PCBP1-AS1*, *EXTL3-AS1*, *TYMSOS*, *LINC01001*, and *LINC00707*) associated with β-thalassaemia in Malaysian patients. Among them, lncRNA *TYMSOS* was significantly upregulated in the TRAIT individuals, compared to MAJOR and CON individuals. LncRNA *TYMSOS* is derived from the opposite strand of the thymidylate synthetase (*TYMS*) gene that encodes a critical enzyme involved in DNA synthesis and repair. A previous study showed that genetic variants of this TYMS gene contributed to variations in red blood cell folate and homocysteine concentrations in Northwestern European

Table 3. Summary of the significant biological pathways in the molecular networks of trait and major phenotypes.

Biological Pathway	Network Hit	*p*-value	FDR
TRAIT molecular network			
Cell Proliferation	21	0.0000	0.0000
Granulopoiesis	7	0.0001	0.0028
Anti-Cell Proliferation	6	0.0001	0.0046
Cell Division	8	0.0003	0.0050
Hematopoiesis	15	0.0003	0.0050
Cell Differentiation	15	0.0003	0.0050
Aging	15	0.0009	0.0128
Adipocyte Differentiation	11	0.0010	0.0128
DNA Damage Response	7	0.0013	0.0140
Angiogenesis	14	0.0016	0.0147
Cardiotoxicity	7	0.0016	0.0147
Cardiac Remodeling	5	0.0019	0.0162
Tumor Suppressor MiRNAs	14	0.0031	0.0235
T-Cell Differentiation	6	0.0066	0.0468
Cell Cycle	13	0.0074	0.0468
MAJOR molecular network			
Cell Differentiation	17	0.0000	0.0000
Anti-Cell Proliferation	7	0.0000	0.0000
Cell Division	8	0.0000	0.0004
T-Cell Differentiation	7	0.0001	0.0023
Tumor Suppressor MiRNAs	13	0.0001	0.0023
Innate Immunity	10	0.0003	0.0048
Aging	12	0.0006	0.0081
Adipocyte Differentiation	9	0.0007	0.0089
Hematopoiesis	11	0.0010	0.0108
Glucose Metabolism	7	0.0027	0.0253
Cell Death	11	0.0029	0.0253

Abbreviation: MAJOR = β-thalassaemia major, TRAIT = β-thalassaemia trait, and FDR = False discovery rate *p*-value.

Table 4. Summary of the selected differentially expressed lncRNA expressions in the validation cohort.

LncRNA	TRAIT *vs.* CON		MAJOR *vs.* CON		MAJOR *vs.* TRAIT	
	Fold Change	*p*-value	Fold Change	*p*-value	Fold Change	*p*-value
VASH1-AS1	34.8	0.000	4.33	NS	-8.33	0.022
PCBP1-AS1	3.53	0.017	5.97	NS	1.69	NS
EXTL3-AS1	2.72	0.018	7.17	NS	2.63	NS
TYMSOS	31.4	0.000	4.71	0.043	-6.67	0.026
LINC01001	34.0	0.000	4.23	NS	-8.33	0.021
LINC00707	5.86	0.035	5.97	NS	1.02	NS

Abbreviation: NS = Not significant, MAJOR = β-thalassaemia major, TRAIT = β-thalassaemia trait, and CON = healthy controls.

adults.(27) Another study on the Chinese population also showed that polymorphism of the *TYMS* gene was significantly associated with a higher risk of leukemia in individuals with low folate intake (28), suggesting the important role of *TYMS* in blood regulation and disorders. Unfortunately, no previous evidence was available on the role of the lncRNA *TYMSOS* in thalassaemia and blood disorders. One transcriptome study of healthy human tissues reported that lncRNA *TYMSOS* expression was highest in bone marrow tissues, followed by colon and kidney tissues (29), implying a possible role similar to that of its host gene, *TYMS*, in blood regulation. Moreover, in the molecular network analysis, lncRNA *TYMSOS* was predicted to interact with the miR-214-3p, a known miRNA that regulates oxidative stress in red blood cells.(29) An *in vitro* study of oxidative stress in erythroid cells showed that miR-214-3p expression was dose-dependent following hemin treatment (an oxidative stress inducer).(30) Elevated miR-214-3p promoted oxidative stress in the erythroid cells by suppressing the Activating Transcription Factor 4 (*ATF4*), a transcriptional activator of various genes and a key player in cellular stress response.(30,31) Based on the known role of miR-214-3p in oxidative stress, the elevated expression of lncRNA *TYMSOS* observed in the TRAIT group could suggest that lncRNA *TYMSOS* sponges miR-214-3p, acting as a negative feedback mechanism to protect cells from oxidative stress. Furthermore, this negative regulation of lncRNA *TYMSOS* towards miR-214-3p may also partly explain the major phenotype, as the expression of this lncRNA *TYMSOS* was reduced in the MAJOR group, implying that this negative regulation may be absent or diminished. However, further studies are needed to confirm this relationship between lncRNA *TYMSOS* and miR-214-3p, particularly their interactions that could explain the differences between major and trait phenotypes.

Besides lncRNA *TYMSOS*, two lncRNAs (*VASH1-AS1* and *LINC01001*) were altered in the MAJOR group compared to the TRAIT group. Similar to *TYMSOS*, both lncRNAs have no previous publication of their roles in blood disorders, particularly thalassaemia. However, in the molecular network analysis, both lncRNAs were shown to co-regulate the let-7 miRNA family. A previous study of pediatric thalassaemia in the Chinese population showed that let-7 miRNAs, hsa-let-7f-1-3p and hsa-let-7a-3p, were upregulated, whereas hsa-let-7b-5p, hsa-let-7i-5p, and hsa-let-7g-5p were downregulated in patients with thalassaemia compared to healthy controls.(32) Moreover, the same study also showed that two of these miRNAs, hsa-let-7b-5p and hsa-let-7i-5p, were significantly reduced in adolescent and adult patients.(32) Importantly, these let-7 miRNAs were shown to regulate the B-cell lymphoma/leukemia 11A (BCL11A) gene, a critical regulator of foetal-to-adult hemoglobin switching, and have been implicated in β-thalassaemia.(33,34) Since a high level of fetal hemoglobin (HbF) (up to 90%) is often observed in β-thalassaemia major compared to β-thalassaemia trait, the co-regulation of let-7 miRNAs by *VASH1-AS1* and *LINC01001* lncRNAs may partly suggest that they are involved in hemoglobin switching, thus contributing to the differences seen in the two phenotypes. Therefore, further studies are needed to explore the relationship between *VASH1-AS1* and *LINC01001* lncRNAs with the let-7 family, particularly in contributing to the dysregulation of foetal-to-adult hemoglobin switching.

Unfortunately, the sample size for the discovery and validation was small, and there was an imbalance between the groups. These issues may affect the statistical power to identify truly significant lncRNAs and increase the risk of Type II errors, thus potentially impacting the reproducibility of the results. Furthermore, some of these lncRNAs exhibit

opposite expression between the major and trait individuals, which may be due to actual biological changes in disease severity; however, these changes could also be influenced by the small sample size.(35) Moreover, there are potential confounding factors that could contribute to changes in lncRNA expression, such as age, gender, ethnicity, presence of other co-morbidities, and environmental factors, which may affect the differences observed in the findings. Previous studies of Malaysian β-thalassaemia patients reported that factors such as age and BTB and CNC homology 1 (BACH1) expression were associated with clinical phenotypes (36,37); however, whether these factors affect lncRNA expression requires further investigation with a larger cohort with a much more diverse population. Another limitation is that all the regulatory networks of lncRNAs and miRNAs are based on prediction analysis and have not been confirmed through direct binding experiments, such as luciferase assays. We also did not measure the expressions of target miRNAs in the validation phase, thus limiting the interpretations of these molecular networks and their contributions to the differences seen in MAJOR and TRAIT phenotypes. Therefore, future studies should incorporate the validation of miRNA expression in the same cohort and confirm their direct molecular binding.

Conclusion

Three key lncRNAs (*TYMSOS*, *VASH1-AS1*, and *LINC01001*) were significantly altered between MAJOR and TRAIT individuals, which may partly explain the phenotypes seen in the Malaysian β-thalassemia patients. However, the current findings would require further validation with a larger and more diverse cohort to confirm the significance of these lncRNAs, as well as functional studies to evaluate the direct interactions and regulation of these molecules. Nevertheless, the identification of these lncRNAs may serve as new biomarkers to distinguish disease severity in β-thalassaemia patients and contribute to the future development of a diagnostic and prognostic panel for these patients.

Acknowledgments

We thank the Director General of Health Malaysia for his permission to publish this article. This manuscript is funded by the Institute for Medical Research (IMR) of Malaysia (Grant code: NMRR-19-2280-48428). The funders have no role in the drafting, writing, and publishing of the manuscript.

Authors Contribution

NMY and SAS proposed the study design and concept, performed data analyses, results interpretation and wrote the manuscript. NMY secured the grant. NFAR and NK performed sample collection, microarray, and Real-Time PCR experiments, as well as analysed the data. NFAR wrote the first draft of the manuscript, while NAAM, ENS, YY, HMI, ZM, EE reviewed and edited the final draft. All authors took part in a critical revision of the manuscript.

Conflict of Interest

All authors declare that they have no competing interests.

References

1. Kattamis A, Kwiatkowski JL, Aydinok Y. Thalassaemia. Lancet. 2022; 399(10343): 2310-24.
2. Alwi ZB, Syed-Hassan SR. Thalassemia in Malaysia. Hemoglobin. 2022; 46(1): 45-52.
3. Giardine B, Borg J, Viennas E, Pavlidis C, Moradkhani K, Joly P, *et al.* Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. Nucleic acids research. 2014; 42(Database issue): D1063-9.
4. Mohd Ibrahim H, Muda Z, Othman IS, Mohamed Unni MN, Teh KH, Thevarajah A, *et al.* Observational study on the current status of thalassaemia in Malaysia: a report from the Malaysian Thalassaemia Registry. BMJ Open. 2020; 10(6): e037974. doi: 10.1136/bmjopen-2020-037974.
5. Nur Handayani NS, Husna N, Rahmil G, Ghifari RA, Widyawati L, Lesmana I. Splice-site and frameshift mutations of β-globin gene found in thalassemia carrier screening in Yogyakarta Special Region, Indonesia. Indonesian Biomed J. 2021; 13(1): 55-60.
6. Organs GaLPiAwTMaIAwIOiS. Glucose and lipid profiles in adolescents with thalassemia major and its association with iron overload in specific organs. Indones Biomed J. 2019; 11(2): 188-93.
7. Tesio N, Bauer DE. Molecular basis and genetic modifiers of thalassemia. Hematol Oncol Clin North Am. 2023; 37(2): 273-99.
8. Abdullah UYH, Ibrahim HM, Mahmud NB, Salleh MZ, Teh LK, Noorizhab M, *et al.* Genotype-phenotype correlation of β-thalassemia in Malaysian population: Toward effective genetic counseling. Hemoglobin. 2020; 44(3): 184-9.
9. Hariharan P, Gorivale M, Sawant P, Mehta P, Nadkarni A. Significance of genetic modifiers of hemoglobinopathies leading towards precision medicine. Sci Rep. 2021; 11(1): 20906.doi: 10.1038/s41598-021-00169-x.
10. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, *et al.* Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023; 24(6): 430-47.

11. Lai K, Jia S, Yu S, Luo J, He Y. Genome-wide analysis of aberrantly expressed lncRNAs and miRNAs with associated co-expression and ceRNA networks in β-thalassemia and hereditary persistence of fetal hemoglobin. Oncotarget. 2017; 8(30): 49931-43.
12. Lulli V, Romania P, Morsilli O, Cianciulli P, Gabbianelli M, Testa U, *et al.* MicroRNA-486-3p regulates γ-globin expression in human erythroid cells by directly modulating BCL11A. PloS One. 2013; 8(4): e60436. doi: 10.1371/journal.pone.0060436.
13. Ma J, Liu F, Du X, Ma D, Xiong L. Changes in lncRNAs and related genes in β-thalassemia minor and β-thalassemia major. Front Med. 2017; 11(1): 74-86.
14. Mohd Yasin N, Tukiman NI, Raman NFA, Kosnan N, Mohd Sahid EN, Mat Yusoff Y, *et al.* Genotype spectrum among children with transfusion dependent beta-thalassemia (TDT) at Woman and Children Hospital (WCH), Tunku Azizah Hospital (HTA), Malaysia. J Med Biomedical Sci. 2021; 6(2): 3-4.
15. Yasin NM, Abdul Hamid FS, Hassan S, Sudin A, Yassim H, Mohd Sahid EN, *et al.* Molecular and hematological studies in a cohort of beta zero South East Asia deletion (β°-thal SEA) from Malaysian perspective. Front Pediatr. 2022: 10: 974496. doi: 10.3389/fped.2022.974496.
16. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, *et al.* LNCipedia 5: Towards a reference set of human long non-coding RNAs. Nucleic Acids Res. 2019; 47(D1): D135-9.
17. Zhao L, Wang J, Li Y, Song T, Wu Y, Fang S, *et al.* NONCODEV6: An updated database dedicated to long non-coding RNA annotation in both animals and plants. Nucleic Acids Res. 2021; 49(D1): D165-71.
18. Li Z, Liu L, Feng C, Qin Y, Xiao J, Zhang Z, *et al.* LncBook 2.0: integrating human long non-coding RNAs with multi-omics annotations. Nucleic Acids Res. 2023; 51(D1): D186-91.
19. Chang L, Xia J. MicroRNA Regulatory Network Analysis Using miRNet 2.0. Methods Mol Biol. 2023: 2594: 185-204.
20. Kolberg L, Raudvere U, Kuzmin I, Adler P, Vilo J, Peterson H. g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 2023; 51(W1): W207-12.
21. Yang M, Chen T, Liu YX, Huang L. Visualizing set relationships: EVenn's comprehensive approach to Venn diagrams. iMeta. 2024; 3(3): e184. doi: 10.1002/imt2.184.
22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13(11): 2498-504.
23. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014; 8(Suppl 4): S11. doi: 10.1186/1752-0509-8-S4-S11.
24. Karagkouni D, Karavangeli A, Paraskevopoulou MD, Hatzigeorgiou AG. Characterizing miRNA-lncRNA interplay. Methods Mol Biol. 2021; 2372: 243-62.
25. Jia W, Jia S, Chen P, He Y. Construction and analysis of a long non-coding RNA (lncRNA)-associated ceRNA network in β-thalassemia and hereditary persistence of fetal hemoglobin. Med Sci Monit. 2019: 25: 7079-86.
26. Fakhr-Eldeen A, Toraih EA, Fawzy MS. Long non-coding RNAs MALAT1, MIAT and ANRIL gene expression profiles in beta-thalassemia patients: A cross-sectional analysis. Hematology. 2019; 24(1): 308-17.
27. Guan X, Sun Y, Zhang C. LncRNAs in blood cells: Roles in cell development and potential pathogenesis in hematological malignancies. Crit Rev Oncol Hematol. 2022: 180: 103849. doi: 10.1016/j.critrevonc.2022.103849.
28. Sumera A, Radhakrishnan AK, Aziz Baba A, George E. A link between long non-coding RNA (lncRNA) and thalassaemia: A review. Malays J Pathol. 2020; 42(3): 323-32.
29. Tsagakis I, Douka K, Birds I, Aspden JL. Long non-coding RNAs in development and disease: conservation to mechanisms. J Pathol. 2020; 250(5): 480-95.
30. Kealey C, Brown KS, Woodside JV, Young I, Murray L, Boreham CA, *et al.* A common insertion/deletion polymorphism of the thymidylate synthase (TYMS) gene is a determinant of red blood cell folate and homocysteine concentrations. Hum Genet. 2005; 116(5): 347-53.
31. Liu P, Zhang M, Xie X, Jin J, Holman CD. Polymorphisms of 5,10-methylenetetrahydrofolate reductase and thymidylate synthase, dietary folate intake, and the risk of leukemia in adults. Tumour Biol. 2016; 37(3): 3265-75.
32. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Mol Cell Proteomics. 2014; 13(2): 397-406.
33. Li Y, Zhang H, Hu B, Wang P, Wang W, Liu J. Post-transcriptional regulation of erythropoiesis. Blood Sci. 2023; 5(3): 150-9.
34. Penglong T, Saensuwanna A, Jantapaso H, Phuwakanjana P, Jearawiriyapaisarn N, Paiboonsukwong K, *et al.* miR-214 aggravates oxidative stress in thalassemic erythroid cells by targeting ATF4. PloS One. 2024; 19(4): e0300958. doi: 10.1371/journal.pone.0300958.
35. Tang H, Kang R, Liu J, Tang D. ATF4 in cellular stress, ferroptosis, and cancer. Arch Toxicol. 2024; 98(4): 1025-41.
36. Wang H, Chen M, Xu S, Pan Y, Zhang Y, Huang H, *et al.* Abnormal regulation of microRNAs and related genes in pediatric β-thalassemia. J Clin Lab Anal. 2021; 35(9): e23945. doi: 10.1002/jcla.23945.
37. Sankaran VG, Xu J, Ragoczy T, Ippolito GC, Walkley CR, Maika SD, *et al.* Developmental and species-divergent globin switching are driven by BCL11A. Nature. 2009; 460(7259): 1093-7.
38. Huang P, Peslak SA, Shehu V, Keller CA, Giardine B, Shi J, *et al.* let-7 miRNAs repress HIC2 to regulate BCL11A transcription and hemoglobin switching. Blood. 2024; 143(19): 1980-91.