Research Paper

Genome-wide methylation and gene-expression analyses in thalassemia

Wei Zhang[1,2,*], Xiaokang Li[3,*], Uet Yu[4], Xin Huang[1], Hongmei Wang[5], Yi Lu[1], Sixi Liu[4], Jian Zhang[1]

[1]School of Medicine, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China
[2]Shenzhen Key Laboratory of Cardiovascular Health and Precision Medicine, School of Public Health and Emergency Management, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China
[3]Center for Reproductive Medicine, University of Hongkong-Shenzhen Hospital, Shenzhen 518053, Guangdong, China
[4]Department of Hematology and Oncology, Shenzhen Children's Hospital, Shenzhen 518038, Guangdong, China
[5]Department of Infectious Diseases, Shenzhen Children's Hospital, Shenzhen 518038, Guangdong, China
*Equal contribution

Correspondence to: Jian Zhang, Sixi Liu, Yi Lu; **email:** zhangjian@sustech.edu.cn; tiger647@126.com, https://orcid.org/0000-0003-1674-2685; luy3@sustech.edu.cn
Keywords: thalassemia, blood, WGBS, RNA-seq
Received: December 5, 2023 **Accepted:** July 11, 2024 **Published:** August 9, 2024

ABSTRACT

Thalassemia is the most common autosomal genetic disorder in humans. The pathogenesis of thalassemia is principally due to the deletion or mutation of globin genes that then leads to disorders in globin-chain synthesis, and its predominant clinical manifestations include chronic forms of hemolytic anemia. However, research on the epigenetics and underlying pathogenesis of thalassemia is in its nascency and not yet been systematically realized. In this study, we compared the results of RNA-seq and the whole-genome bisulfite sequencing (WGBS) on 22 peripheral blood samples from 14 thalassemic patients and eight healthy individuals revealed a genome-wide methylation landscape of differentially methylated regions (DMRs). And functional-enrichment analysis revealed the enriched biological pathways with respect to the differentially expressed genes (DEGs) and differentially methylated genes (DMGs) to include hematopoietic lineage, glucose metabolism, and ribosome. To further analyze the interaction between the transcriptome and methylome, we implemented a comprehensive analysis of overlaps between DEGs and DMGs, and observed that biological processes significantly enriched the immune-related genes (i.e., our hypermethylated and down-regulated gene group). Hypermethylated and hypomethylated regions of thalassemia-related genes exhibited different distribution patterns. We thus, further identified and validated thalassemia-associated DMGs and DEGs by multi-omics integrative analyses of DNA methylation and transcriptomics data, and provided a comprehensive genomic map of thalassemia that will facilitate the exploration of the epigenetics mechanisms and pathogenesis underlying thalassemia.

INTRODUCTION

Thalassemia is a group of autosomally inherited hemolytic disorders and the most prevalent autosomal-recessive hemolytic disease worldwide [1], with approximately one to five percent of the global population estimated to be carriers of thalassemia [2–4]. Molecular biologic studies have revealed that thalassemia is caused by the deletion or mutation of the globin genes [5–7]. Adult hemoglobin is primarily composed of two α-globin subunits and two β-globin subunits [8–10]: the α-globin gene is located on chromosome 16, with each homologous chromosome containing two α-globin genes (i.e., *HBA1* and

HBA2); and the β-globin gene is located on chromosome 11, with each homologous chromosome containing one β-globin gene (i.e., *HBB*). Thalassemia is thus principally segregated into two types in clinical practice, alpha thalassemia and beta thalassemia; although there exist a small number of unusual variants of thalassemia such as delta-beta thalassemia and delta thalassemia [9].

Conventional clinical treatments for thalassemia consist of blood transfusion, iron removal, and splenectomy. Allogeneic hematopoietic stem cell transplantation (HSCT) has more recently been shown to be the most effective therapeutic regimen to treat thalassemia [11–13]. Allogenic HSCT is limited to patients having human leukocyte antigen (HLA) homologous donors, and gene therapy using autologous hematopoietic stem cells provides an alternative treatment for allogenic HSCT [13, 14]. Although there are data supporting HSCT as relatively ideal in the setting of HLA matching, the probability of finding an appropriately histocompatible donor is less than 50% [15]. While previous studies showed no significant difference in the survival of patients receiving peripheral blood-derived stem cells and patients receiving bone marrow-derived stem cells, patients receiving the latter produced a faster engraftment rate and lower rejection rate than those patients receiving the peripheral stem cells [16]. Other studies encompassed peripheral blood, bone marrow, or umbilical cord blood from unrelated donors to adjust the myeloablative regimen and to reduce the toxic effects of the graft, thus significantly augmenting the overall survival rate [17, 18].

To reveal the complex pathogenesis of thalassemia, it is crucial to integrate multi-omics data (e.g., from genomics, epigenomics, proteomics, and metabolomics) to assess their relationships at different molecular levels and their impacts on disease phenotypes [19]. As an important component of epigenomics, DNA methylation is not only related to other epigenetic modifications, but also maintains an important relationship with gene expression [20–23]. Therefore, the integration of DNA methylation and gene expression and the systematic analysis of the relationship between the two comprise a currently exciting area of research. In this study we executed a comprehensive and systematic biologic analysis of 14 pediatric patients with thalassemia and eight age-matched healthy individuals by combining genome-wide transcriptional and global-DNA methylation analyses. By integrating epigenomic and transcriptomic data, we demonstrated an association between DNA methylation and differential gene-expression patterns in thalassemia, and explored the epigenetic mechanisms and pathogenesis underlying thalassemia.

RESULTS

Data from the research subjects

This study comprised a collection of peripheral blood samples from 14 pediatric patients with thalassemia (age range: 4-14 years, Mean±SD: 8±3.0) and eight healthy children (age range: 5-14 years, Mean±SD: 10±2.9) from which the RNA and DNA were sampled for transcriptomic and methylation sequencing, respectively. However, due to problems during the extraction processes in some research subjects, sample Th6 in the thalassemia group was only subjected to methylation sequencing, sample Th14 was only subjected to transcriptome sequencing, and in the control group sample N1 was only subjected to methylation sequencing. All of the remaining samples were subjected to both methylation and transcriptomic sequencing (see Table 1 for details).

RNA-seq and WGBS quality-assessment and alignment summary

All samples in the present study generated high-quality RNA-seq reads, fastp pre-processing was used for raw-read quality, and we obtained over 99% clean data [24] (Supplementary Table 1). The HISAT2-alignment program for mapping sequencing reads was implemented to carry out comparative analysis based on the reference genome; the number of reads and the proportion of effective reads (i.e., the total mapped) for each sample were both ≥96% (Supplementary Table 2). We then calculated the distribution position of the reads in the reference genome according to all the reads that were mapped to the genome (i.e., the total mapped reads). Genome-wide DNA methylation analysis was consequently performed on the samples. In brief, the clean reads obtained from methylation sequencing were filtered to obtain high-quality clean reads for our subsequent data analysis. Each sample showed ≥97% high-quality clean reads (Supplementary Table 3).

Cluster analysis

The expression distribution of different sample genes or transcripts was displayed via the expression-distribution map and based on the fragments per kilobase of transcript per million mapped reads (FPKM) of each gene (Figure 1A). Dimensionality reduction was then used to ascertain the distance relationships between samples using principal component analysis (PCA). In the control group, samples N2 and N5 were outliers that were consistent with the thalassemic group, especially N5 was same with thalassemia group (Figure 1B). In methylation sequencing, the methylation-rate data for CpG sites in each sample were used to perform PCA

Table 1. Table of sample characteristics.

Samples	Gender	Age (years)	Type	Sequencing
Th1	M	8	CD41-42/ CD41-42	WGBS/RNA-seq
Th2	F	10	41-42M/17M	WGBS/RNA-seq
Th3	F	8	CD41-42/IVS-2-654	WGBS/RNA-seq
Th4	M	14	CD41-42/ CD41-42	WGBS/RNA-seq
Th5	F	7	SEA	WGBS
Th6	F	13	CD41-42/IVS-I-654	WGBS/RNA-seq
Th7	F	5	IVS-II-654/βE	WGBS/RNA-seq
Th8	M	8	654M/17M	WGBS/RNA-seq
Th9	M	6	IVS-II-654/ IVS-II-654	WGBS/RNA-seq
Th10	F	11	CD41-42/βE	WGBS/RNA-seq
Th11	M	4	-SEA/ cs	WGBS/RNA-seq
Th12	F	7	CD41-42/IVS-2-	WGBS/RNA-seq
Th13	F	6	IVS-1-128/-28	WGBS/RNA-seq
Th14	F	5	CD41-42/IVS-1-1	RNA-seq
N1	M	12	Normal	WGBS
N2	F	11	Normal	WGBS/RNA-seq
N3	M	12	Normal	WGBS/RNA-seq
N4	F	7	Normal	WGBS/RNA-seq
N5	F	9	Normal	WGBS/RNA-seq
N6	M	14	Normal	WGBS/RNA-seq
N7	F	10	Normal	WGBS/RNA-seq
N8	M	5	Normal	WGBS/RNA-seq

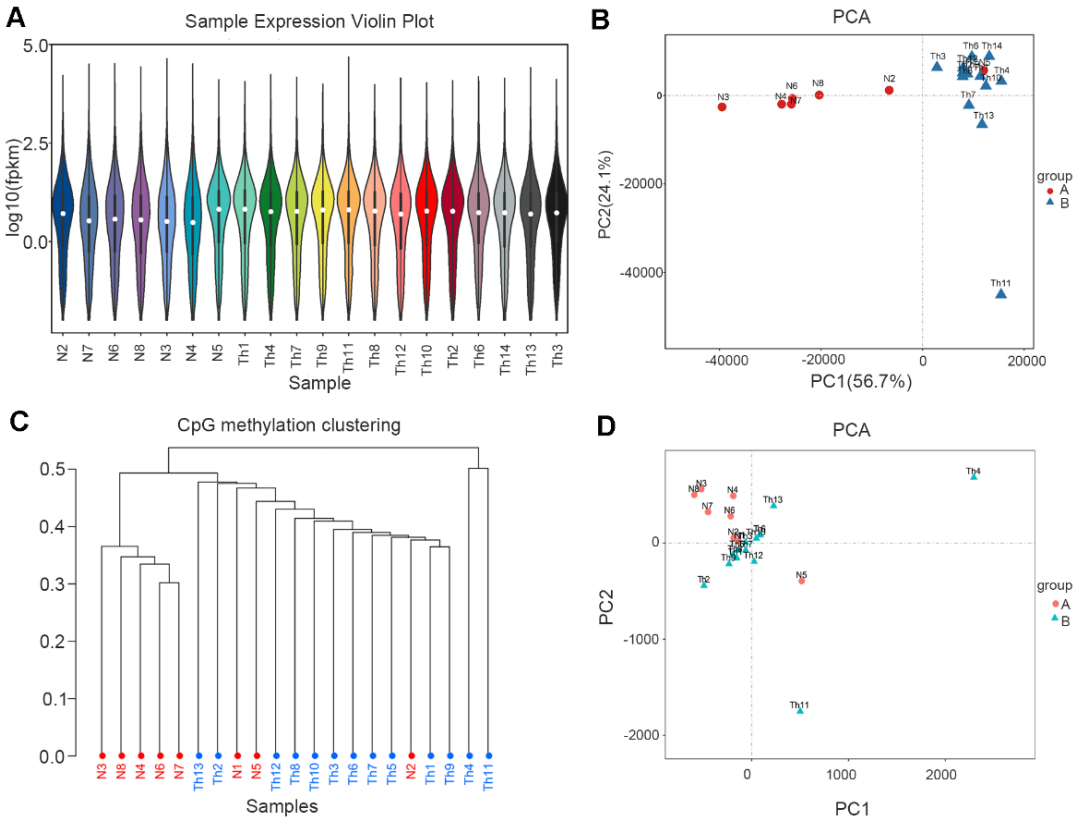

Figure 1. RNA-seq and WGBS data cluster analysis. (**A**) Sample expression violin plot. White dot represents the median Q2. The black rectangle represents the range from the lower quartile to the upper quartile. (**B**) Principal Component Analysis. Red dots represent normal. Blue dots represent thalassemia. (**C**) Principal Component Analysis (WGBS). Red dots represent normal. Blue dots represent thalassemia. (**D**) CpG methylation clustering. Red represents normal. Blue represents thalassemia.

and cluster analysis on all samples, and our results showed that three samples (N1, N5, and N2) were outliers in the control group (Figure 1C, 1D).

Analysis of DEGs

Via statistical analysis of the differences among genomes, genes with an FDR6 (six percent) <0.05 and |log2 fold-change| >1 were screened as significantly differentially expressed genes (DEGs); and in the thalassemic group, we uncovered 2,155 upregulated and 657 downregulated genes (Figure 2A). We generated volcano plots based on the DEGs in the comparison group (Figure 2B), executed hierarchical clustering of differential gene-expression patterns, and created a heatmap to present the clustering results (Figure 2C). Our results confirmed that samples N2 and N5 deviated from the control group.

Functional-enrichment analysis of DEGs

To better understand the biologic functions related to DEGs, we executed Gene Ontology (GO)-enrichment analysis, and noted that molecular function was significantly enriched in haptoglobin binding, cytokine binding, structural constituent of ribosome, oxygen binding, and immunoglobulin receptor binding—indicating that in addition to the changes in globin binding, hemoglobin binding, and oxygen binding, many other related molecular functions were also significantly different in the thalassemia group (Figure 3A). We also found enriched biological processes such as immune system process, cell activation, ribosome biogenesis, leukocyte activation, and lymphocyte activation—indicating that the immune system of thalassemic patients was significantly different from that of healthy individuals (Figure 3B).

Significantly enriched pathways were established among DEGs through further GO- enrichment analysis, and our results showed that among the top 20 most significant differences, the second-highest ranking was reflected by hematopoietic cell lineage—indicating significant changes in the hematopoietic cell lineage-related genes in thalassemia patients (Figure 3C). We also observed significant changes in the expression of multiple transcription factors and epigenetic-modification enzymes in the DEGs, including DNA methyltransferases and deacetylases. Quantitative PCR was used to verify the results and to confirm that they were consistent with sequencing (Figure 3D).

DNA-methylation changes

Analysis of differentially methylated CpG sites (DMCs) revealed that there were 5,793 upregulated and 2,629 downregulated genes with regard to CpG site methylation in the thalassemia group (Data were not shown), indicating an overall effect of thalassemia on genome-wide methylation (Figure 4A). We also determined the genomic distribution of DNA methylation changes and the distribution of DMCs associated with CpG islands, and demonstrated that 5.92% of DMCs were located in promoter regions, 10.95% in exons, 23.39% in introns, and 59.73% of DMCs were located in intergenic regions (Figure 4B). In addition, most of the DMCs were located in "non-CpG-rich regions" (denoted as "open sea" in Figure 4C).

Functional-enrichment analysis of differentially methylated genes (DMGs)

We executed GO and Kyoto Encyclopedia of Genes and Genomes (KEGG)-enrichment analyses on the biological functions and signaling pathways of DMGs

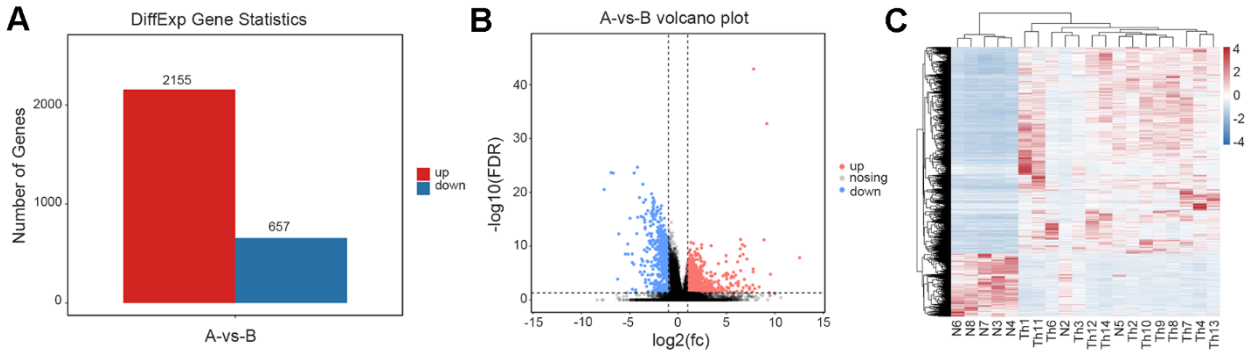

Figure 2. Analysis of differentially expressed genes. (**A**) The number of differentially expressed genes between normal and thalassemia group. Red represents up expression; Blue represents down expression. (**B**) The volcano plot shows the differentially expressed genes between normal and thalassemia group. Each dot corresponds to a gene. Red dots represent up expression; Gray dots represent not significant; Blue dots represent down expression. (**C**) Heatmap of differentially expressed genes in two groups. Red represents high expression; Blue represents low expression.

linked with thalassemia. Through KEGG-enrichment analysis of genes associated with the DMCs on the CpG sites, we obtained pathways of type II diabetes mellitus, platelet N1tivation, metabolic pathway, and ECM-receptor interaction (Figure 5A). Analysis of differentially methylated regions (DMRs) revealed that there were 82 upregulated genes and seven downregulated genes (Data were not shown). KEGG pathway-enrichment analysis identified glycolysis/ gluconeogenesis, pyruvate metabolism, fatty acid degradation, type II diabetes mellitus, MAPK signaling pathway, and T cell receptor signaling pathway (Figure 5B).

Analysis of overlap between DMGs and DEGs

To understand the relationship between epigenetic regulation and transcriptomic changes in the blood of thalassemic patients, we undertook an integrated analysis of DMGs and DEGs (after removing three outliers: samples N1, N5, and N2), and noted that the difference in the positional relationship between DMGs and coding genes affected the regulatory effect of DNA methylation. Figure 6A depicts the analysis of changes in DMG locations and DEG-expression levels, and further overlap analysis between the 2,323 DMGs and 4,442 DEGs showed overlap in 779 genes (Data were

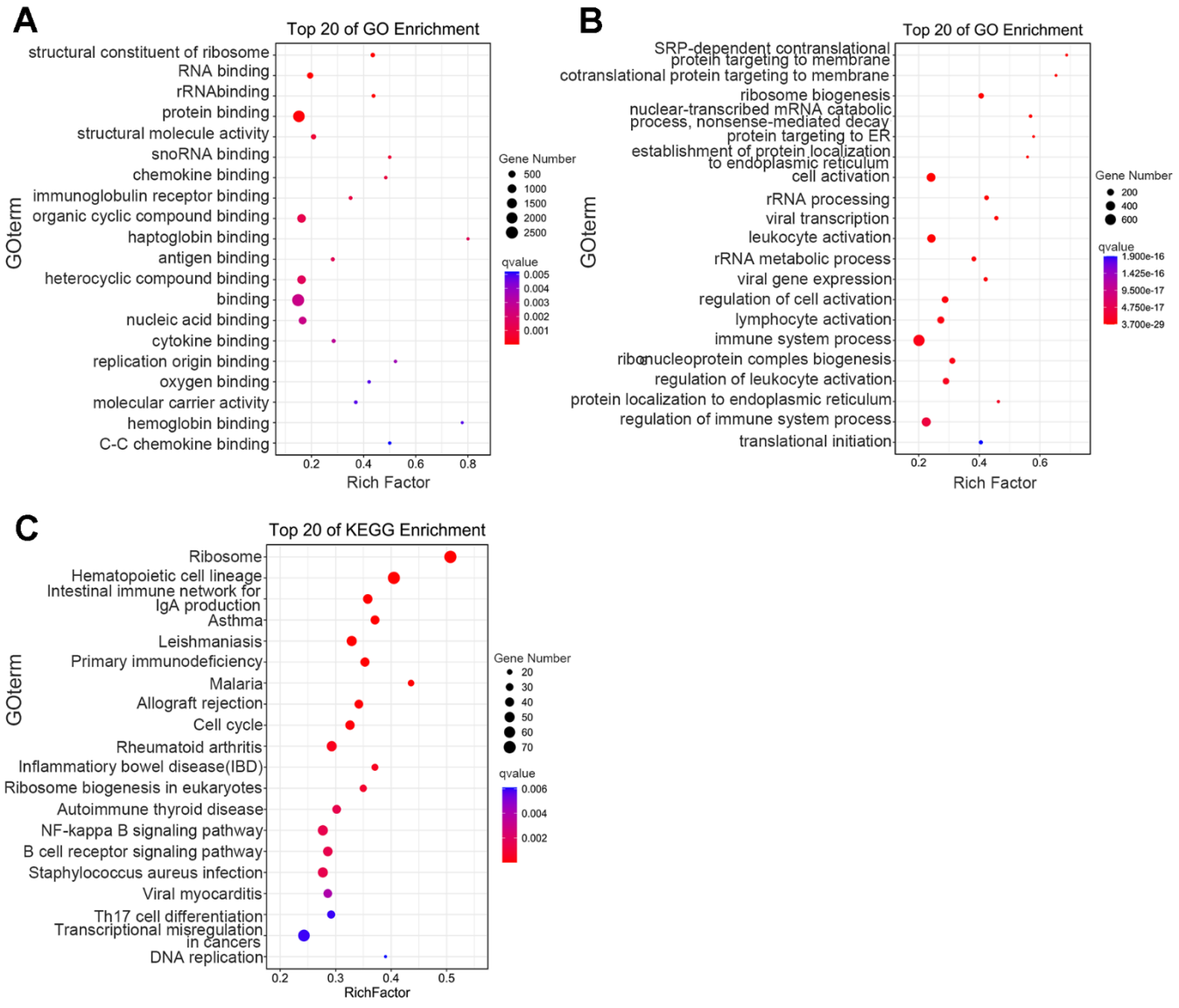

Figure 3. Functional-enrichment analysis of differentially expressed genes by GO and KEGG. The 20 most significantly enriched biological functions using GO (**A**, **B**) and KEGG (**C**) are illustrated in dot plots. Rich factor refers to the proportion of DEGs belonging to a specific term. Node size (gene number) refers to the number of DEGs within each term and node color indicates the level of significance (−log10 p-value).

not shown). We divided these 779 genes into four categories N1cording to the direction of DNA methylation and gene expression: "Hypo-Down" (for hypomethylated and downregulated genes); "Hypo-Up" (for hypomethylated and upregulated genes); "Hyper-Down" (for hypermethylated and down-regulated genes); and "Hyper-Up" (for hyper-methylated and up-regulated genes). We then selected the genes with the top 30-fold differences in gene expression in eN1h category (Figure 6A), and GO-enrichment analysis indicated that immune-related genes were principally enriched in biological processes (i.e., our Hyper-Down group, Figure 6B).

DISCUSSION

We herein collected peripheral blood samples from pediatric patients with thalassemia and performed genome-wide DNA methylation and transcriptomics analyses to elucidate the molecular changes occurring within the blood cells of thalassemic patients. Via the GO-enrichment analysis of DEGs, cell part, cell surface, and side of membrane were significantly enriched in the cellular component. Unbalanced globin-chain synthesis results in the precipitation of excess globin chains in erythrocyte precursors, and results in structural changes to the cell membrane. Athanasiou et al. used a micropipette aspiration technique to measure the ratio of the elastic-shear moduli of red blood cells in thalassemia-model mice and patients with thalassemia, and found that the rigidity of the thalassemic red blood cells was significantly greater than that of normal red blood cells [25]. A previous study showed that the reduction in erythrocyte deformability and membrane stability resulted in the destruction of erythrocytes as they passed through bone marrow-cavity blood vessels, the splenic sinus, and capillary networks—shortening their lifespan in the circulation [26]. We also performed KEGG-enrichment analysis, and noted that the second highest-ranked object of the top 20 most significant differences was hematopoietic lineage, indicating robust changes in hematopoietic lineage-related genes in thalassemic patients (Figure 3C). We additionally uncovered significant changes in the expression of

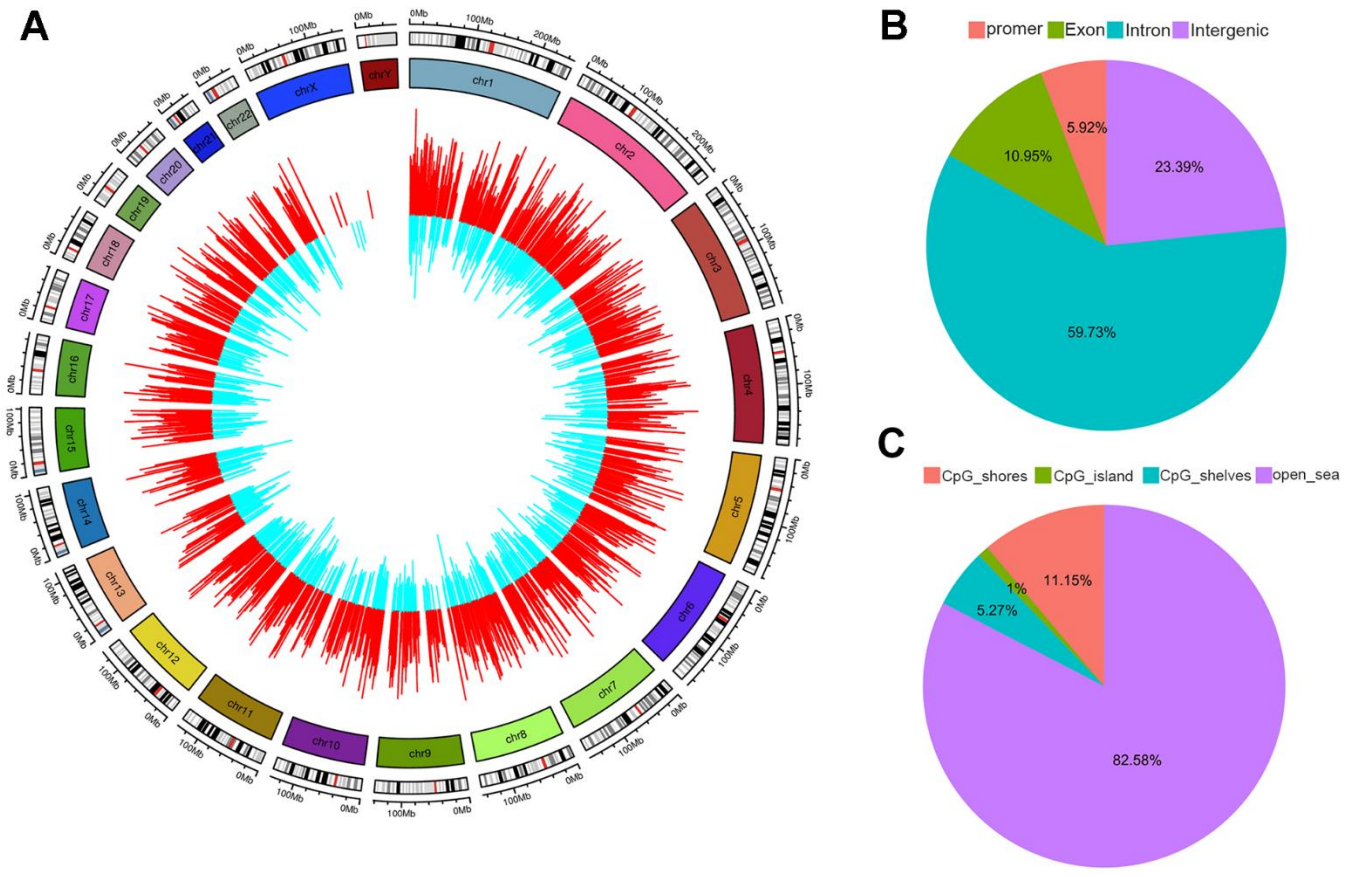

Figure 4. Distribution of differentially methylated CpGs (DMR). (**A**) Outer circle represents hypermethylated CpGs colored in red. Inner circle represents hypomethylated CpGs colored in blue. The height of each bar indicates the methylation change between thalassemia and normal. The distributions of DMR summarized based on genomic location (**B**) and relative to CpG islands (CpGi) (**C**).

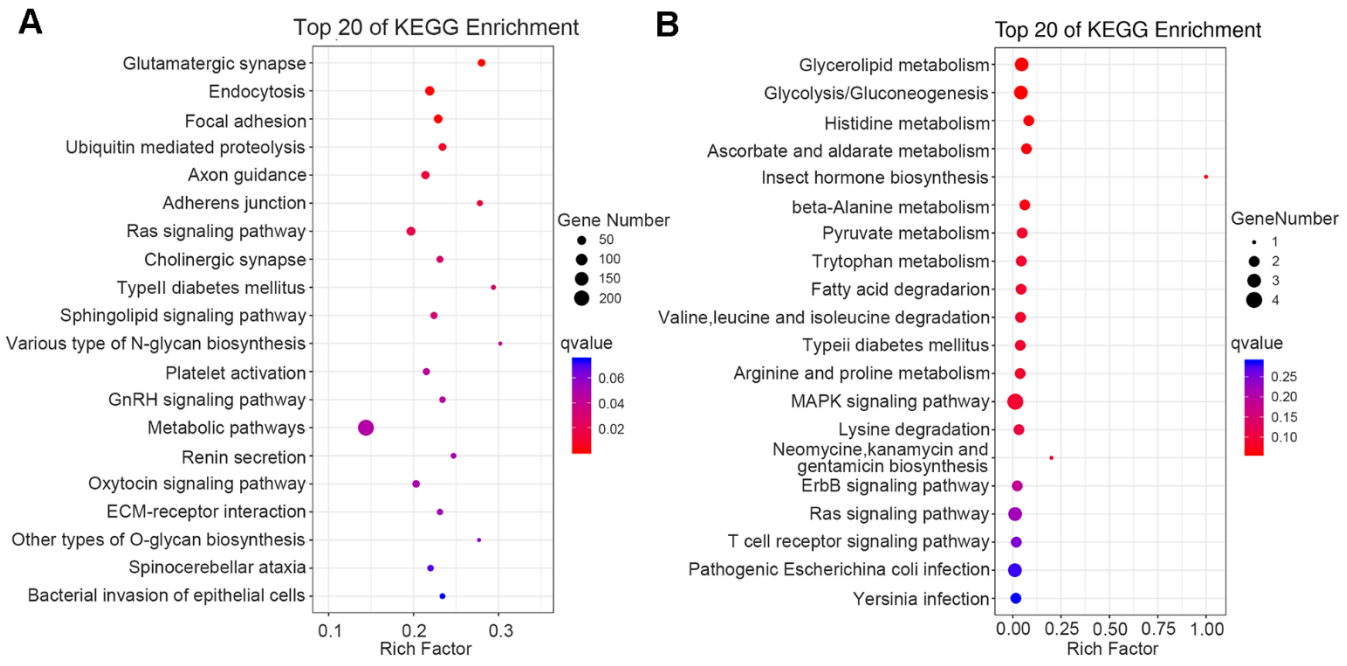

Figure 5. Functional enrichment analysis of DMGs by KEGG. The 20 most significantly enriched biological functions in DMC (**A**) and DMR (**B**) using KEGG are illustrated in dot plots. Rich factor refers to the proportion of DMGs belonging to a specific term. Node size (gene number) refers to the number of DMGs within each term and node color indicates the level of significance (−log10 p-value). DMGs mean differentially methylated genes.

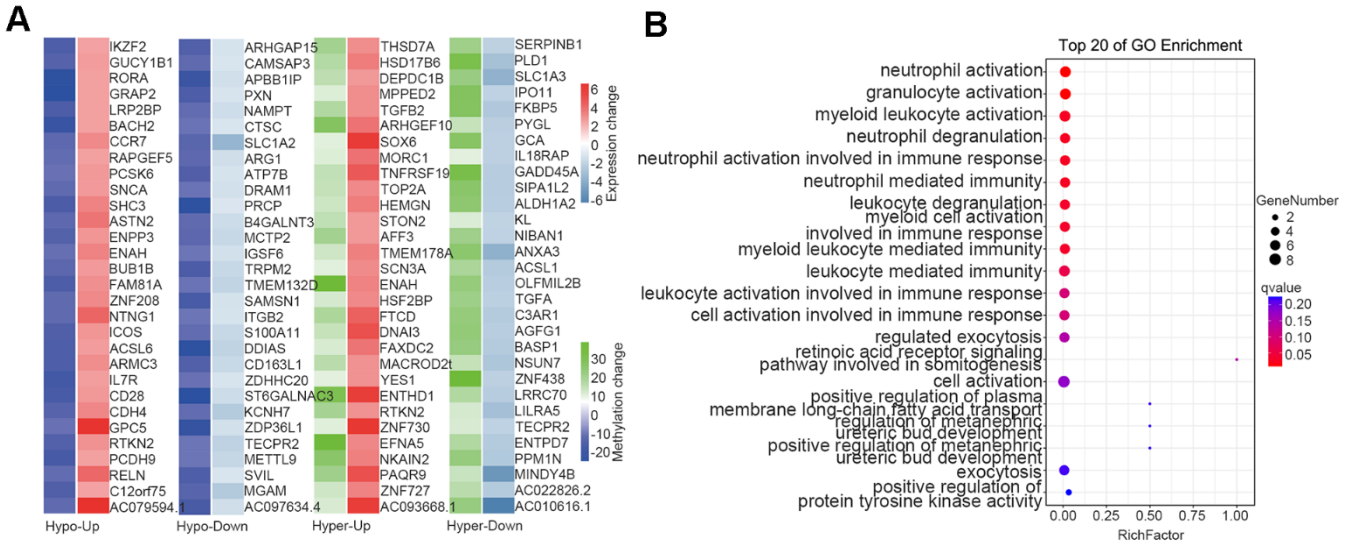

Figure 6. Overlap between DMGs and DEGs. (**A**) Heatmap of methylation and expression changes of the overlapping DMGs and DEGs. The first column of each group corresponds to the methylation change (Blue: Hypomethylated, Green: Hypermethylated), while the second column represents the gene expression change (Red: Upregulated, Right blue: downregulated). (**B**) Pathway enrichment analysis of overlapping genes in methylation and mRNA datasets. The 20 most significantly enriched pathways are illustrated in dot plots. Gene ratio refers to the proportion of DEGs belonging to a specific term. Node size (count) refers to the number of DEGs within each term and the color indicates the level of significantly (− log10 P-value). DMGs differentially methylated genes, DEGs differentially expressed genes.

multiple transcription-factor genes and epigenetic-modifying enzymes among the DEGs.

The KEGG-enrichment analysis of DMGs showed that type II diabetes mellitus and platelet activation were most enriched, and our analysis of DMRs showed that KEGG was enriched in glycolysis/gluconeogenesis, type II diabetes mellitus, and T cell receptor signaling pathway—among others (Figure 5B). Numerous studies have revealed that iron overload was correlated with diabetes. For example, Chen et al. showed that patients with type II diabetes manifested higher serum ferritin than healthy individuals [27] and Liang et al. demonstrated that children with high serum ferritin levels exhibited a higher prevalence of impaired fasting glucose than children with low levels [28]. Ansari et al. showed that nearly one-third of β-thalassemia patients who exhibited iron overload were insulin resistant, and that the insulin-resistance index increased with age and the elevation in serum ferritin, suggesting a close relationship between iron overload and islet resistance [29]. Luo et al., in a study of 79 patients with thalassemia, 114 patients with hemoglobin-H disease, and 18 patients with hemoglobin E-beta thalassemia, observed 33 cases of hypoglycemia, 25 cases of impaired glucose tolerance, and four cases of diabetes. In that study patients with thalassemia demonstrated symptoms that ranged from impaired glucose tolerance to symptomatic diabetes [30]. Although thalassemia patients who are dependent upon continual blood transfusions without adherence to iron chelators are at high risk of developing diabetes, there exists a broad therapeutic window that spans the spectrum from abnormal glucose metabolism to symptomatic diabetes. Intensive removal of iron thus reduces insulin resistance and is expected to delay the onset of diabetes.

With the development of biotechnology, gene insertion and gene editing have become strategies to correct and replace ineffective β globin in patients with β thalassemia [31–34]. Considerable efforts have also been made to study pharmacological drugs stimulating the production of γ globin and HbF [35, 36]. Allogeneic hematopoietic stem cell transplantation (HSCT) has been used successfully in the past few decades to provide curative treatment for transfusion-dependent patients, but only for a small number of patients with compatible donors [37, 38]. However, blood transfusion is more affordable as a traditional therapy. A wide spectrum of immune abnormalities has been described in numerous studies involving β-thalassemia patients with multiple transfusions—with iron overload a major cause of immunodeficiency in β-thalassemia. Since humans lack an efficient mechanism with which to excrete excess iron, chronic blood transfusions may lead to iron N1cumulation and thus generate reactive oxygen species that can trigger lipid, protein, DNA, and subcellular organellar damage, which can then cause cellular dysfunction, apoptosis, and necrosis, and precipitate target organ toxicity and dysfunction [39]. Our analysis of the overlap between DMGs and DEGs revealed that the immune-related genes (Hyper-Down group) were enriched in biological processes, indicating alterations to the immune system of patients with thalassemia relative to healthy individuals.

With the treatment of thalassemia has developed significantly, resulting in an increase in life expectancy. At the same time, a number of new onset and chronic diseases have emerged, including cancer. Over the years, several cases of solid and hematologic malignancies in patients with thalassemia have been reported in the literature [40–42]. In our study, KEGG results showed that in addition to immune-related diseases being enriched, cancer-related signaling pathways such as cell cycle, DNA replication, and NF-κB signaling pathways were also significantly enriched. Although the mechanisms underlying cancer development in patients with thalassemia are not well understood, studies have shown that patients with TDT and β thalassemia major have a higher risk of developing malignancies compared with the general population [43].

We herein analyzed the transcriptomes and methylomes of thalassemic patients to determine the relationship between epigenetics and gene-expression levels in thalassemia and uncovered enrichments in biological pathways such as hematopoietic lineage, immunity, glucose metabolism, and ribosomes. Furthermore, tumor-associated signaling pathways were enriched, such as NF-κB signaling pathway. Although the exact associations between these pathways and thalassemia remain unclear, they may be of great significance in understanding the molecular underpinnings of thalassemia, and should facilitate the discovery of novel genes related to its epigenetic regulation.

MATERIALS AND METHODS

Samples collection

This subject has been approved by the Medical Ethics Committee of Southern University of Science and Technology (Project number: SUSTC-JY2017041). The study was conducted in accordance with the ethical standards as laid down in the 2018NL-106-02 Declaration of Helsinki and its later amendments or comparable ethical standards. We collected peripheral blood samples from 14 pediatric patients with thalassemia (age range: 4-14 years, Mean±SD: 8±3.0) and 8 healthy children (age range: 5-14 years,

Mean±SD: 10±2.9). Patient characteristics are given in Table 1. RNA and DNA were extracted for transcriptomic and methylation sequencing, respectively. However, sample Th6 and N1 were only subjected to methylation sequencing, and Th14 was only subjected to transcriptome sequencing.

Genome-wide gene expression and methylation profiling

Total RNA Trizol reagent kit (Invitrogen, CA, USA) according to the manufacturer's protocol. RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and checked using RNase free agarose gel electrophoresis. Eukaryotic mRNA was enriched by Oligo(dT) beads, while prokaryotic mRNA was enriched by removing rRNA by Ribo-ZeroTM Magnetic Kit (Epicentre, WI, USA). Then the enriched mRNA was fragmented into short fragments and second-strand cDNA were synthesized by DNA polymerase I, RNase H, dNTP and buffer. QiaQuick PCR extraction kit (Qiagen, The Netherlands) was used to purify the cDNA fragments, and ligated to Illumina sequencing adapters. RNA-seq was used Illumina HiSeq2500 by Gene Denovo Biotechnology Co. (Guangzhou, China).

Genomic DNA (gDNA) was extracted from whole blood using DNeasy Blood & Tissue Kit (Qiagen, CA, USA) according to the manufacturer's protocol. DNA concentration and integrity were detected by NanoPhotometer® spectrophotometer (Implen, CA, USA) and Agarose Gel Electrophoresis respectively. For library construction, genomic DNAs were fragmented into 100-300bp by Sonication (Covaris, Massachusetts, USA) and purified with MiniElute PCR Purification Kit (Qiagen, MD, USA). After purification, a single adenosine was added to the 3′ends of the fragmented DNA. And then adapter-ligated DNA fragments were treated. Fragment with adapters were bisulfite converted using Methylation-Gold kit (Zymo, CA, USA), unmethylated cytosine is converted to uracil during sodium bisulfite treatment. Finally, the converted DNA fragments were PCR amplified and sequenced using Illumina HiSeqTM 2500 by Gene Denovo Biotechnology Co. (Guangzhou, China).

Data filtering

Reads obtained from the sequencing machines include raw reads containing adapters or low quality bases which will affect the following assembly and analysis. To get high quality clean reads, reads were further filtered by fastp (version 0.18.0) [24]. Briefly, we removed reads containing adapters, containing more than 10% of unknown nucleotides (N), and low quality reads containing more than 50% of low quality (Q-value ≤ 20) bases for RNA-seq. We removed reads containing more than 10% of unknown nucleotides (N), and low quality reads containing more than 40% of low quality (Q-value ≤ 20) bases for methylation profiling.

Principal component analysis

R package Seurat v3.1.2 was used to process the single-cell data expression matrix. The data were first normalized by 'NormalizeData'. 'FindVariableGenes' was then used to identify 2000 highly variable genes. Principal component analysis (PCA) was performed with R package gmodelsin this experience. PCA is a statistical procedure that converts hundreds of thousands of correlated variables (gene expression/ methylation level) into a set of values of linearly uncorrelated variables called principal components. PCA is largely used to reveal the structure/relationship of the samples/datas.

Differentially expressed genes (DEGs)

RNAs differential expression analysis was performed by DESeq2 software between two different groups [44]. The genes/transcripts with the parameter of false discovery rate (FDR) below 0.05 and absolute fold change≥2 were considered differentially expressed genes/transcripts.

Functional enrichment analysis

To identify and compare the overrepresented biological functions, enrichment analysis was performed using a hypergeometric test with our in-house R analysis package richR (http://github.com/hurlab/richR). Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, Gene Ontology (GO) terms were used in the enrichment analysis, and the calculated p-value were gone through FDR Correction, taking FDR ≤ 0.05 as a threshold. GO terms meeting this condition were defined as significantly enriched GO terms in DEGs. KEGG is the major public pathway-related database [45, 46]. Pathway enrichment analysis identified significantly enriched metabolic pathways or signal transduction pathways in DEGs comparing with the whole genome background.

Methylation level analysis

The obtained clean reads were mapped to the species reference genome using BSMAP software (version: 2.90) [47]. Then a custom Perl script was used to call methylated cytosines and the methylated cytosines were tested with the correction algorithm described in Lister R. et al. (2009) [48]. The methylation level was calculated from the percentage of methylated cytosines

per sequence context (CG, CHG and CHH) across the whole genome, each chromosome and different regions of the genome. To assess different methylation patterns in different genomic regions, the methylation profile of the flanking 2 kb regions and gene body (or transposable elements) was plotted based on the average methylation levels for each window.

Differentially methylated cytosines (DMCs) and differentially methylated regions (DMRs) analysis

Differential DNA methylation between the two groups at each locus was determined using Pearson's chi-square test ($\chi 2$) in methyl Kit (version: 1.7.10) [49]. To identify differentially methylated cytosines (DMCs), the minimum read coverage to call a methylation status for a base was set to 4. Differentially methylated cytosines for each sequence context (CG, CHG and CHH) between two groups were identified according to different criteria. And the differentially methylated regions (DMRs) also were identified according to different criteria.

Consent for publication

All authors consent to publish the work.

Availability of data and materials

The whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) data from this study have been deposited in the Genome Sequence Archive in BIG Data Center (http://bigd.big.ac.cn/), Chinese Academy of Sciences, under the accession number: HRA002227; HRA002236.

AUTHOR CONTRIBUTIONS

JZ, SXL and YL designed research; UY and HMW collected samples; XKL and XH performed research; WZ and XH analyzed data; WZ and XKL wrote the paper.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

ETHICAL STATEMENT AND CONSENT

The study was approved by the medical ethics committee of Southern University of science and technology (No. SUSTC-JY2017041) followed the tenets of the Declaration of Helsinki (2018NL-106-02). Written informed consents were obtained from all study subjects or relatives.

FUNDING

This research was supported by Shenzhen Science and Technology Innovation Commission, JCYJ20170412152943794, JCYJ20170412154619484, ZDSYS20200810171403013, 20220815153635001; NSFC 82173336, 81773146, 81972766, 81972420, 81802949. We appreciate Xinyu Ye and Ming Chang for their technical help.

REFERENCES

1. McLean E, Cogswell M, Egli I, Wojdyla D, de Benoist B. Worldwide prevalence of anaemia, WHO Vitamin and Mineral Nutrition Information System, 1993-2005. Public Health Nutr. 2009; 12:444–54. https://doi.org/10.1017/S1368980008002401 PMID:18498676

2. Weatherall DJ. The inherited diseases of hemoglobin are an emerging global health burden. Blood. 2010; 115:4331–6. https://doi.org/10.1182/blood-2010-01-251348 PMID:20233970

3. Vichinsky E. Complexity of alpha thalassemia: growing health problem with new approaches to screening, diagnosis, and therapy. Ann N Y Acad Sci. 2010; 1202:180–7. https://doi.org/10.1111/j.1749-6632.2010.05572.x PMID:20712791

4. Modell B, Darlison M. Global epidemiology of haemoglobin disorders and derived service indicators. Bull World Health Organ. 2008; 86:480–7. https://doi.org/10.2471/blt.06.036673 PMID:18568278

5. Clegg JB, Weatherall DJ. Molecular basis of thalassaemia. Br Med Bull. 1976; 32:262–9. https://doi.org/10.1093/oxfordjournals.bmb.a071373 PMID:788836

6. Rund D, Rachmilewitz E. Beta-thalassemia. N Engl J Med. 2005; 353:1135–46. https://doi.org/10.1056/NEJMra050436 PMID:16162884

7. Cao A, Galanello R. Beta-thalassemia. Genet Med. 2010; 12:61–76. https://doi.org/10.1097/GIM.0b013e3181cd68ed PMID:20098328

8. Paikari A, Sheehan VA. Fetal haemoglobin induction in sickle cell disease. Br J Haematol. 2018; 180:189–200. https://doi.org/10.1111/bjh.15021 PMID:29143315

9. Viprakasit V, Ekwattanakit S. Clinical Classification, Screening and Diagnosis for Thalassemia. Hematol Oncol Clin North Am. 2018; 32:193–211.

https://doi.org/10.1016/j.hoc.2017.11.006 PMID:29458726

10. Weatherall DJ, Clegg JB. Molecular genetics of human hemoglobin. Annu Rev Genet. 1976; 10:157–78. https://doi.org/10.1146/annurev.ge.10.120176.001105 PMID:797307

11. Baronciani D, Angelucci E, Potschger U, Gaziev J, Yesilipek A, Zecca M, Orofino MG, Giardini C, Al-Ahmari A, Marktel S, de la Fuente J, Ghavamzadeh A, Hussein AA, et al. Hemopoietic stem cell transplantation in thalassemia: a report from the European Society for Blood and Bone Marrow Transplantation Hemoglobinopathy Registry, 2000-2010. Bone Marrow Transplant. 2016; 51:536–41. https://doi.org/10.1038/bmt.2015.293 PMID:26752139

12. La Nasa G, Giardini C, Argiolu F, Locatelli F, Arras M, De Stefano P, Ledda A, Pizzati A, Sanna MA, Vacca A, Lucarelli G, Contu L. Unrelated donor bone marrow transplantation for thalassemia: the effect of extended haplotypes. Blood. 2002; 99:4350–6. https://doi.org/10.1182/blood.v99.12.4350 PMID:12036861

13. Hongeng S, Pakakasama S, Chaisiripoomkere W, Chuansumrit A, Sirachainan N, Ungkanont A, Jootar S. Outcome of transplantation with unrelated donor bone marrow in children with severe thalassaemia. Bone Marrow Transplant. 2004; 33:377–9. https://doi.org/10.1038/sj.bmt.1704361 PMID:14676781

14. Cappellini MD, Porter JB, Viprakasit V, Taher AT. A paradigm shift on beta-thalassaemia treatment: How will we manage this old disease with new therapies? Blood Rev. 2018; 32:300–11. https://doi.org/10.1016/j.blre.2018.02.001 PMID:29455932

15. Amjad F, Fatima T, Fayyaz T, Khan MA, Qadeer MI. Novel genetic therapeutic approaches for modulating the severity of β-thalassemia (Review). Biomed Rep. 2020; 13:48. https://doi.org/10.3892/br.2020.1355 PMID:32953110

16. Ghavamzadeh A, Kasaeian A, Rostami T, Kiumarsi A. Comparable Outcomes of Allogeneic Peripheral Blood versus Bone Marrow Hematopoietic Stem Cell Transplantation in Major Thalassemia: A Multivariate Long-Term Cohort Analysis. Biol Blood Marrow Transplant. 2019; 25:307–12. https://doi.org/10.1016/j.bbmt.2018.09.026 PMID:30266673

17. Sun L, Wang N, Chen Y, Tang L, Xing C, Lu N, Shi Y, Ma Y, Lin F, Yu K, Feng J. Unrelated Donor Peripheral Blood Stem Cell Transplantation for Patients with β-Thalassemia Major Based on a Novel Conditioning Regimen. Biol Blood Marrow Transplant. 2019; 25:1592–6. https://doi.org/10.1016/j.bbmt.2019.03.028 PMID:30951841

18. Shenoy S, Thompson AA. Unrelated donor stem cell transplantation for transfusion-dependent thalassemia. Ann N Y Acad Sci. 2016; 1368:122–6. https://doi.org/10.1111/nyas.13019 PMID:26999376

19. Zaina S, Lund G. Integrating genomic and epigenomic information: a promising strategy for identifying functional DNA variants of human disease. Clin Genet. 2012; 81:334–40. https://doi.org/10.1111/j.1399-0004.2011.01840.x PMID:22292420

20. van Eijk KR, de Jong S, Strengman E, Buizer-Voskamp JE, Kahn RS, Boks MP, Horvath S, Ophoff RA. Identification of schizophrenia-associated loci by combining DNA methylation and gene expression data from whole blood. Eur J Hum Genet. 2015; 23:1106–10. https://doi.org/10.1038/ejhg.2014.245 PMID:25424713

21. Li Y, Xu J, Ju H, Xiao Y, Chen H, Lv J, Shao T, Bai J, Zhang Y, Wang L, Wang X, Ren H, Li X. A network-based, integrative approach to identify genes with aberrant co-methylation in colorectal cancer. Mol Biosyst. 2014; 10:180–90. https://doi.org/10.1039/c3mb70270g PMID:24317156

22. Shen H, Laird PW. Interplay between the cancer genome and epigenome. Cell. 2013; 153:38–55. https://doi.org/10.1016/j.cell.2013.03.008 PMID:23540689

23. van Eijk KR, de Jong S, Boks MP, Langeveld T, Colas F, Veldink JH, de Kovel CG, Janson E, Strengman E, Langfelder P, Kahn RS, van den Berg LH, Horvath S, Ophoff RA. Genetic analysis of DNA methylation and gene expression levels in whole blood of healthy human subjects. BMC Genomics. 2012; 13:636. https://doi.org/10.1186/1471-2164-13-636 PMID:23157493

24. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018; 34:i884–90. https://doi.org/10.1093/bioinformatics/bty560 PMID:30423086

25. Athanasiou G, Zoubos N, Missirlis Y. Erythrocyte membrane deformability in patients with thalassemia syndromes. Nouv Rev Fr Hematol (1978). 1991; 33:15–20. PMID:1945820

26. Schrier SL, Rachmilewitz E, Mohandas N. Cellular and membrane properties of alpha and beta thalassemic erythrocytes are different: implication for differences in clinical manifestations. Blood. 1989; 74:2194–202. https://doi.org/10.1182/blood.V74.6.2194.2194 PMID:2804358

27. Chen L, Li Y, Zhang F, Zhang S, Zhou X, Ji L. Elevated serum ferritin concentration is associated with incident type 2 diabetes mellitus in a Chinese population: A prospective cohort study. Diabetes Res Clin Pract. 2018; 139:155–62. https://doi.org/10.1016/j.diabres.2018.03.001 PMID:29524483

28. Liang Y, Bajoria R, Jiang Y, Su H, Pan H, Xia N, Chatterjee R, Lai Y. Prevalence of diabetes mellitus in Chinese children with thalassaemia major. Trop Med Int Health. 2017; 22:716–24. https://doi.org/10.1111/tmi.12876 PMID:28544032

29. Ansari AM, Bhat KG, Dsa SS, Mahalingam S, Joseph N. Study of Insulin Resistance in Patients With β Thalassemia Major and Validity of Triglyceride Glucose (TYG) Index. J Pediatr Hematol Oncol. 2018; 40:128–31. https://doi.org/10.1097/MPH.0000000000001011 PMID:29227325

30. Luo Y, Bajoria R, Lai Y, Pan H, Li Q, Zhang Z, Yang P, Chatterjee R, Liang Y. Prevalence of abnormal glucose homeostasis in Chinese patients with non-transfusion-dependent thalassemia. Diabetes Metab Syndr Obes. 2019; 12:457–68. https://doi.org/10.2147/DMSO.S194591 PMID:31114275

31. Soni S. Gene therapies for transfusion dependent β-thalassemia: Current status and critical criteria for success. Am J Hematol. 2020; 95:1099–112. https://doi.org/10.1002/ajh.25909 PMID:32562290

32. Magrin E, Miccio A, Cavazzana M. Lentiviral and genome-editing strategies for the treatment of β-hemoglobinopathies. Blood. 2019; 134:1203–13. https://doi.org/10.1182/blood.2019000949 PMID:31467062

33. Antoniani C, Meneghini V, Lattanzi A, Felix T, Romano O, Magrin E, Weber L, Pavani G, El Hoss S, Kurita R, Nakamura Y, Cradick TJ, Lundberg AS, et al. Induction of fetal hemoglobin synthesis by CRISPR/Cas9-mediated editing of the human β-globin locus. Blood. 2018; 131:1960–73. https://doi.org/10.1182/blood-2017-10-811505 PMID:29519807

34. Frangoul H, Altshuler D, Cappellini MD, Chen YS, Domm J, Eustace BK, Foell J, de la Fuente J, Grupp S, Handgretinger R, Ho TW, Kattamis A, Kernytsky A, et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. N Engl J Med. 2021; 384:252–60. https://doi.org/10.1056/NEJMoa2031054 PMID:33283989

35. Ren Q, Zhou YL, Wang L, Chen YS, Ma YN, Li PP, Yin XL. Clinical trial on the effects of thalidomide on hemoglobin synthesis in patients with moderate thalassemia intermedia. Ann Hematol. 2018; 97:1933–9. https://doi.org/10.1007/s00277-018-3395-5 PMID:29931453

36. Chen J, Zhu W, Cai N, Bu S, Li J, Huang L. Thalidomide induces haematologic responses in patients with β-thalassaemia. Eur J Haematol. 2017; 99:437–41. https://doi.org/10.1111/ejh.12955 PMID:28850716

37. Angelucci E, Matthes-Martin S, Baronciani D, Bernaudin F, Bonanomi S, Cappellini MD, Dalle JH, Di Bartolomeo P, de Heredia CD, Dickerhoff R, Giardini C, Gluckman E, Hussein AA, et al, and EBMT Inborn Error and EBMT Paediatric Working Parties. Hematopoietic stem cell transplantation in thalassemia major and sickle cell disease: indications and management recommendations from an international expert panel. Haematologica. 2014; 99:811–20. https://doi.org/10.3324/haematol.2013.099747 PMID:24790059

38. Mohamed SY. Thalassemia Major: Transplantation or Transfusion and Chelation. Hematol Oncol Stem Cell Ther. 2017; 10:290–8. https://doi.org/10.1016/j.hemonc.2017.05.022 PMID:28651066

39. Brittenham GM. Iron-chelating therapy for transfusional iron overload. N Engl J Med. 2011; 364:146–56. https://doi.org/10.1056/NEJMct1004810 PMID:21226580

40. Hodroj MH, Bou-Fakhredin R, Nour-Eldine W, Noureldine HA, Noureldine MH, Taher AT. Thalassemia and malignancy: An emerging concern? Blood Rev. 2019; 37:100585. https://doi.org/10.1016/j.blre.2019.06.002 PMID:31253373

41. Picardo E, Mitidieri M, Minniti E, Ambroggio S, D'Addato F, Benedetto C, Gregori G, Baù MG. The first case of breast cancer in thalassemic patient: case report and review of literature. Gynecol Endocrinol. 2015; 31:345–8. https://doi.org/10.3109/09513590.2014.998646 PMID:25578420

42. Pellegrino C, Dragonetti G, Chiusolo P, Rossi M, Orlando N, Teofili L. Acute Promyelocytic Leukemia in a Woman with Thalassemia Intermedia: Case Report and

Review of Literature on Hematological Malignancies in β-Thalassemia Patients. Hematol Rep. 2022; 14:310–21. https://doi.org/10.3390/hematolrep14040045 PMID:36278522

43. Chung WS, Lin CL, Lin CL, Kao CH. Thalassaemia and risk of cancer: a population-based cohort study. J Epidemiol Community Health. 2015; 69:1066–70. https://doi.org/10.1136/jech-2014-205075 PMID:25922472

44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550. https://doi.org/10.1186/s13059-014-0550-8 PMID:25516281

45. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. https://doi.org/10.1093/nar/28.1.27 PMID:10592173

46. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–40. https://doi.org/10.1093/bioinformatics/btp616 PMID:19910308

47. Xi Y, Li W. BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics. 2009; 10:232. https://doi.org/10.1186/1471-2105-10-232 PMID:19635165

48. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–22. https://doi.org/10.1038/nature08514 PMID:19829295

49. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol. 2012; 13:R87. https://doi.org/10.1186/gb-2012-13-10-r87 PMID:23034086

SUPPLEMENTARY MATERIALS

Supplementary Tables

Supplementary Table 1. Table of transcriptome sequencing raw data filtering statistics.

Sample	RawDatas	CleanData(%)	Adapter(%)	LowQuality(%)	polyA(%)	N(%)
Th1	50006004	49902462 (99.79%)	18724 (0.04%)	84640 (0.17%)	0 (0.00%)	178 (0.00%)
Th2	55443846	55318884 (99.77%)	18758 (0.03%)	106046 (0.19%)	0 (0.00%)	158 (0.00%)
Th3	40425602	40328798 (99.76%)	14374 (0.04%)	82310 (0.20%)	0 (0.00%)	120 (0.00%)
Th4	50783392	50666806 (99.77%)	15442 (0.03%)	100984 (0.20%)	0 (0.00%)	160 (0.00%)
Th6	41425714	41330790 (99.77%)	11046 (0.03%)	83752 (0.20%)	0 (0.00%)	126 (0.00%)
Th7	48321188	48208000 (99.77%)	17604 (0.04%)	93980 (0.19%)	0 (0.00%)	1604 (0.00%)
Th8	49510754	49381966 (99.74%)	19880 (0.04%)	107278 (0.22%)	0 (0.00%)	1630 (0.00%)
Th9	41641166	41549312 (99.78%)	17202 (0.04%)	72980 (0.18%)	0 (0.00%)	1672 (0.00%)
Th10	50783392	50666806 (99.77%)	15442 (0.03%)	100984 (0.20%)	0 (0.00%)	160 (0.00%)
Th11	42764640	42666194 (99.77%)	16834 (0.04%)	79844 (0.19%)	0 (0.00%)	1768 (0.00%)
Th12	49234158	49125144 (99.78%)	18574 (0.04%)	88456 (0.18%)	0 (0.00%)	1984 (0.00%)
Th13	52239378	52129862 (99.79%)	14880 (0.03%)	93242 (0.18%)	0 (0.00%)	1394 (0.00%)
Th14	46151816	46051722 (99.78%)	14316 (0.03%)	84262 (0.18%)	0 (0.00%)	1516 (0.00%)
N2	44249462	44147832 (99.77%)	22396 (0.05%)	77442 (0.18%)	0 (0.00%)	1792 (0.00%)
N3	40172616	40101378 (99.82%)	11700 (0.03%)	57826 (0.14%)	0 (0.00%)	1712 (0.00%)
N4	47446602	47361862 (99.82%)	11574 (0.02%)	71458 (0.15%)	0 (0.00%)	1708 (0.00%)
N5	47509608	47425176 (99.82%)	12592 (0.03%)	70254 (0.15%)	0 (0.00%)	1586 (0.00%)
N6	47154318	47062864 (99.81%)	15354 (0.03%)	74538 (0.16%)	0 (0.00%)	1562 (0.00%)
N7	40852226	40773584 (99.81%)	13550 (0.03%)	63426 (0.16%)	0 (0.00%)	1666 (0.00%)
N8	53202306	53087908 (99.78%)	16950 (0.03%)	95996 (0.18%)	0 (0.00%)	1452 (0.00%)

Supplementary Table 2. Table of transcriptome data comparison reference statistics table.

Sample	Total	Unmapped(%)	Unique_Mapped(%)	Multiple_Mapped(%)	Total_Mapped(%)
Th1	49316178	1192760 (2.42%)	46786713 (94.87%)	1336705 (2.71%)	48123418 (97.58%)
Th2	54484068	1539106 (2.82%)	51421781 (94.38%)	1523181 (2.80%)	52944962 (97.18%)
Th3	39512696	1134702 (2.87%)	37355273 (94.54%)	1022721 (2.59%)	38377994 (97.13%)
Th4	49564172	1434264 (2.89%)	45467523 (91.73%)	2662385 (5.37%)	48129908 (97.11%)
Th6	39645906	1080551 (2.73%)	37507972 (94.61%)	1057383 (2.67%)	38565355 (97.27%)
Th7	47612628	1366704 (2.87%)	44619278 (93.71%)	1626646 (3.42%)	46245924 (97.13%)
Th8	48803368	1467442 (3.01%)	46079180 (94.42%)	1256746 (2.58%)	47335926 (96.99%)
Th9	41023672	1122839 (2.74%)	38860892 (94.73%)	1039941 (2.53%)	39900833 (97.26%)
Th10	48552004	1397544 (2.88%)	45777749 (94.29%)	1376711 (2.84%)	47154460 (97.12%)
Th11	42289604	1276988 (3.02%)	39704597 (93.89%)	1308019 (3.09%)	41012616 (96.98%)
Th12	47466072	1457269 (3.07%)	44642002 (94.05%)	1366801 (2.88%)	46008803 (96.93%)
Th13	50973230	1472993 (2.89%)	47266662 (92.73%)	2233575 (4.38%)	49500237 (97.11%)
Th14	44509112	1179953 (2.65%)	41900120 (94.14%)	1429039 (3.21%)	43329159 (97.35%)
N2	43472084	1180433 (2.72%)	41033506 (94.39%)	1258145 (2.89%)	42291651 (97.28%)
N3	39785296	1022730 (2.57%)	37570857 (94.43%)	1191709 (3.00%)	38762566 (97.43%)
N4	46871142	1163005 (2.48%)	44295823 (94.51%)	1412314 (3.01%)	45708137 (97.52%)
N5	46406672	1292575 (2.79%)	43923147 (94.65%)	1190950 (2.57%)	45114097 (97.21%)
N6	46543924	1191423 (2.56%)	43894398 (94.31%)	1458103 (3.13%)	45352501 (97.44%)
N7	40434784	1040819 (2.57%)	38290611 (94.70%)	1103354 (2.73%)	39393965 (97.43%)
N8	52532404	1543542 (2.94%)	49433409 (94.10%)	1555453 (2.96%)	50988862 (97.06%)

Supplementary Table 3. Table of methylation data filter.

Sample	Clean reads num	HQ Clean reads num (%)
Th1	655011516	643133678 (98.19%)
Th2	578756378	571110174 (98.68%)
Th3	591190230	581830674 (98.42%)
Th4	633080368	621180388 (98.12%)
Th5	623730348	620351826 (99.46%)
Th6	633309176	618384592 (97.64%)
Th7	611371898	597365680 (97.71%)
Th8	602652242	589629610 (97.84%)
Th9	620137130	609657726 (98.31%)
Th10	617834478	613911016 (99.36%)
Th11	589514406	580554920 (98.48%)
Th12	597078888	586488056 (98.23%)
Th13	594013150	587207694 (98.85%)
N1	590001510	583815932 (98.95%)
N2	653835408	641331770 (98.09%)
N3	586415144	578223570 (98.6%)
N4	581657012	573956604 (98.68%)
N5	621917978	615017628 (98.89%)
N6	648774992	636028996 (98.04%)
N7	640678680	627502560 (97.94%)
N8	583698310	575715568 (98.63%)