

Sex-specific transcriptional profiles identified in β-thalassemia patients

β-thalassemia comprises a group of heterogeneous autosomal recessive hereditary anemias characterized by the reduction or absence of β-globin chain synthesis, and it is a highly prevalent disease affecting 1.5% of the global population.¹ Three different clinical conditions are recognized in patients with β-thalassemia minor (trait) being the asymptomatic form, β-thalassemia major (TM) being the most severe form of the disease and β-thalassemia intermedia (TI) presenting with variable severity. Despite extensive characterization of the genetic basis of disease pathogenesis,² currently the classification of patients relies on the severity of symptoms and hemoglobin (Hb) F levels regardless of the underlying genotype. Thus, the aim of the study was to develop an approach for patient stratification based on gene expression, to pinpoint the targets that dictate each phenotype and to provide a framework for the development of therapeutic strategies focused on these targets. To this end, we have analyzed the gene expression profiles of TI, TM and healthy individuals using RNA sequencing (RNAseq) (National Center for Biotechnology Information [NCBI], GSE117221) and we have studied the differentially expressed genes (DEG) and pathways irrespective to patient genotype. Interestingly, after analysis of various confounding factors, we identified sex differences in the patients' expression profiles suggesting that males and females are differentially affected by β-thalassemia. Thus, taking sex into account might benefit prognosis, diagnosis, stratification and therapeutic management of the disease.

In particular, 49 subjects (after exclusion of low quality samples) were included in the analysis and organized in groups of three age- and sex-matched samples within each group (*Online Supplementary Tables S1-2*). RNAseq libraries were generated from erythroid precursor cell cultures after the isolation of peripheral blood mononuclear cells from all participants, as previously described.³ We identified 716 genes with aberrant expression between TI patients and healthy subjects, and 2,885 between TM patients and healthy subjects with most of DEG seen in TI patients being also present in TM patients when compared to healthy subjects, albeit with more pronounced changes (Figure 1A; Table 1; *Online Supplementary Tables S3-5*). However, no significantly DEG were found when TM patients were compared directly to TI patients sug-

gesting that either the global gene expression profile was very similar between the two types of the disease or that substantial variability in expression did not allow the identification of consistent changes, or both (*Online Supplementary Table S6*). In general, only small changes were detected when TM and TI were directly compared (Figure 1B) and the increased gene expression variability seen in TI patients, which did not allow the identification of the same number of significantly DEG as in the case of TM patients, also hindered the identification of DEG between TI and TM patients. Increased gene expression variability in TI potentially reflects the high level of phenotypic heterogeneity for TI patients.

Nonetheless, the expression profiles accurately portrayed the clinical observations of β-thalassemia. The severe TM phenotype was associated with induction of organismal injuries, as well as inhibition of key hematological system genes and inflammatory response molecules compared to the less severe type of the disease (TI) (Figure 1C-D; *Online Supplementary Table S7*). Moderate changes were seen in the expression levels of various globin and other interacting proteins in TI patients, whereas in TM patients the data portrayed the marked repression of β-thalassemia-related proteins with concomitant up-regulation of other globin proteins as a means of compensating for the ineffective erythropoiesis (*Online Supplementary Figure S1*). Focusing on molecular pathways affected by the disease, gene set enrichment analysis (GSEA) revealed very similar pathways in both TI and TM patients as differentially represented when compared to healthy participants, in accordance with the gene expression profiles (*Online Supplementary Figure S2*). Several of the pathways identified have been previously linked to β-thalassemia validating our results, such as the impaired packaging of telomere ends,⁴ impaired unfolded protein response (UPR) pathway⁵ and lipid abnormalities.^{6,7} Nonetheless, the lack of significant changes between TI and TM patients in terms of global gene expression profiles or molecular pathways suggest that a continuous spectrum describes the disease and not distinct conditions.

We then searched for other biological confounders that could affect global expression patterns allowing patient stratification, since the study was designed to limit as much as possible all technical sources of variation (balanced groups in terms of sex and age, standardized cell culture protocol in all centers with the matched samples being cultured at the same time, library construction per-

Table 1. Numbers of significantly differentially expressed genes.

Analysis		All	Males	Females
TI vs. H	Samples	16 TI vs. 17 H	7 TI vs. 8 H	9 TI vs. 9 H
	Up-regulated	147	315	5
	Down-regulated	569	1,244	9
	Total	716	1,559	14
TM vs. H	Samples	16 TM vs. 17 H	8 TM vs. 8 H	8 TM vs. 9 H
	Up-regulated	939	40	100
	Down-regulated	1,946	401	210
	Total	2,885	441	310
TM vs. TI	Samples	16 TI vs. 16 TM	8 TM vs. 7 TI	8 TM vs. 9 TI
	Up-regulated	0	0	0
	Down-regulated	0	0	1
	Total	0	0	1

TI: β-thalassemia intermedia; TM: β-thalassemia major; H: healthy. Numbers of significantly differentially expressed genes are shown for all comparisons performed. The analysis is produced by DESeq2 and differentially expressed genes were defined as significant when $P_{adj}<0.1$.

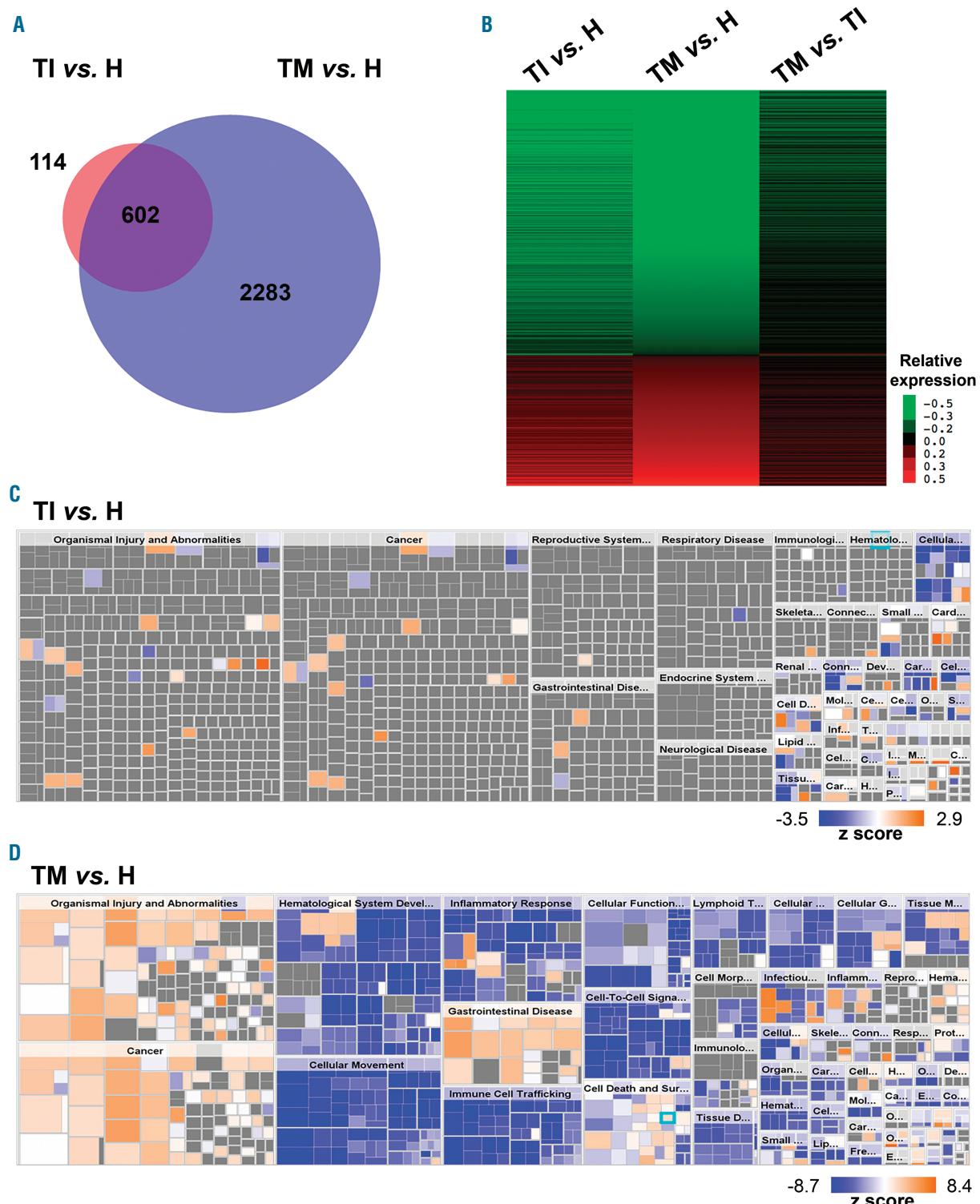


Figure 1. Differential gene expression analysis of β -thalassemia intermedia or β -thalassemia major patients against healthy participants. (A) Venn diagram depicting common significantly differentially expressed genes (DEG) when β -thalassemia intermedia (TI) (n=16) or β -thalassemia major (TM) (n=16) patients were compared to healthy (H) participants (n=17). (B) Heatmap depicting relative normalized gene expression levels (z score) of all 2,999 genes that were found significantly differentially expressed in TI or TM patients when compared to healthy participants. The log₂fold change values of the genes used range from -3.0 to 3.0. (C-D) Mosaic graphs produced by Ingenuity Pathway Analysis (IPA) depicting enriched terms regarding diseases and biological functions when TI patients were compared to healthy participants (C) or when TM patients were compared to healthy participants (D). The z score depicts predicted inhibition or activation of disease/function, whereas the size of the box represents the significance of each identified term (-log₁₀ P-value). Due to visualization purposes, category labels are not shown in full, but detailed enrichment terms can be found in the *Online Supplementary Table S7*.

formed by a single user and balanced sequencing runs). Distinct analyses performed per research center, age group and other clinical characteristics of the patients (such as HbF levels and presence or absence of splenectomy, hepatomegaly, extramedullary hematopoiesis or bone deformities) were unable to detect any major differences in the expression profiles. In order to further explore the data, we performed Principal Component

Analysis (PCA), which visualizes strong patterns in a dataset by reducing the dimensionality of the dataset and clustering of samples based on their similarity. PCA did not yield clear clustering when taken into account all the different patient characteristics, with the striking exception of sex, where all samples were clustered into two distinct groups representing males and females irrespective of the disease status (Figure 2A). Although a clear dis-

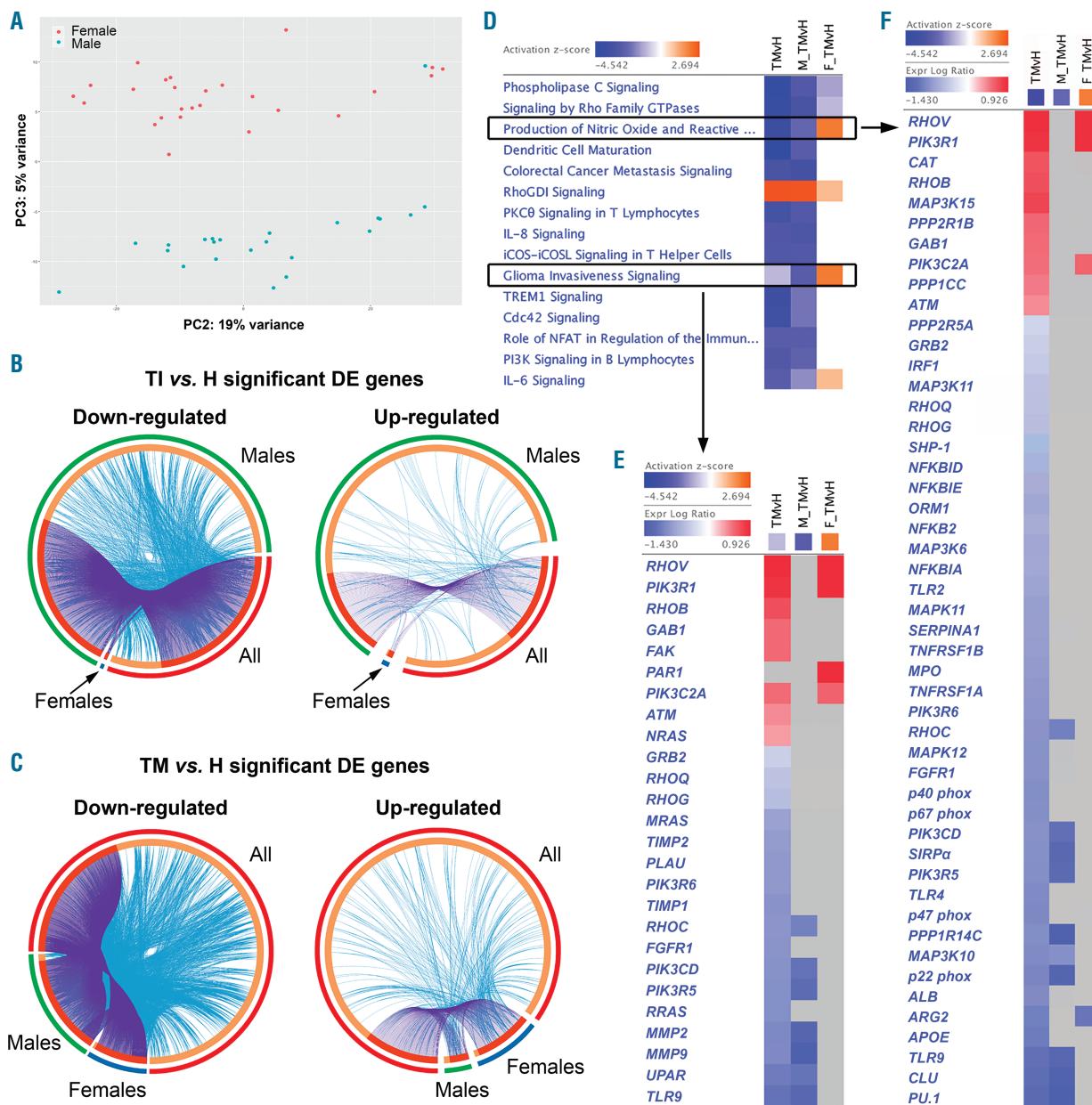


Figure 2. Differential gene expression analysis of β-thalassemia patients against healthy participants according to sex. (A) Principal Component Analysis (PCA) graph showing clustering of samples according to sex irrespective of disease status. (B-C) Circos plots depicting the down-regulated (left) and up-regulated (right) significantly differentially expressed genes (DEG) when β-thalassemia intermedia (TI) (B) or β-thalassemia major (TM) (C) patients were analysed against healthy participants. Outer circle represents the type of analysis with red depicting analysis of all samples (sixteen TI or sixteen TM samples), green depicting analysis of male samples only (seven TI or eight TM samples) and blue depicting analysis of female samples only (nine TI or eight TM samples). Inner circle represents the overlap with dark orange depicting genes that exist in multiple lists and light orange depicting genes that are unique to that particular list. Purple lines link the same genes when shared by multiple lists, whereas blue lines link different genes that fall into the same ontology term. (D) Heatmap produced by Ingenuity Pathway Analysis (IPA) depicting enriched terms regarding canonical pathways when all TM patients were analyzed against healthy participants, male only TM patients were analyzed against healthy males or female only TM patients were analyzed against healthy females. The activation z-score depicts predicted inhibition or activation of the pathway. (E-F) Heatmaps produced by IPA for two example canonical pathways; glioma invasiveness signaling (E) and production of nitric oxide and reactive oxygen species in macrophages (F). The heatmaps depict gene expression levels (expression log ratio, i.e., \log_2 fold change), whereas the boxes above the heatmaps depict the z-score corresponding to panel D and showing the predicted inhibition or activation of the pathway.

tinction between sexes was anticipated, it raised the question whether males and females were affected in distinctive ways by the disease. In order to examine such a possibility, same sex analysis between healthy participants and patients was performed expecting fewer changes due to the lower number of samples used per differential expression analysis (from 49 participants, 23 were males and 26 were females) and similar DEG between males and females, as β-thalassemia presents with similar phenotype in both sexes and has not been linked to sex-defining genes. Interestingly, very different results were obtained for males and females suffering from TI against healthy participants, with 1,559 DEG in males, but only 14 DEG in females (Table 1; *Online Supplementary Tables S8-S9*). The very low number of DEG in females highlights the increased biological variability seen in females and suggests that other factors might play an important role in determining the disease outcome. When comparing the significantly DEG identified in both male and female TI patients with the genes identified only in male TI patients, a significant overlap was seen in down-regulated genes, not only in terms of specific genes, but also in terms of gene functionality through Gene Ontology (GO) terms (Figure 2B; *Online Supplementary Figure S3A-B*; *Online Supplementary Table S5*). In contrast, in up-regulated genes, fewer common genes were identified and fewer similarities in their functionality suggesting less conserved changes. Different results were also found for males and females suffering from TM when compared to healthy participants, albeit less prominent, with 441 DEG identified in males and 310 DEG in females (Table 1; *Online Supplementary Tables S10-S11*). Furthermore, the overwhelming majority of genes identified in either male or female TM patients were also yielded when all TM patients were analysed against healthy subjects irrespective of their sex (Figure 2C; *Online Supplementary Figure S3C-D*; *Online Supplementary Table S5*). The limited number of deregulated genes in common between female and male TM patients could demonstrate sex-specific differences, further supported by the association of different terms related to diseases and body functions in male and female patients when compared to same sex healthy subjects (*Online Supplementary Figure S4*). Dissection of the molecular pathways involved through pathway and GO analysis revealed pathways with opposing status between males and females, such as the production of nitric oxide and reactive oxygen species in macrophages, and glioma invasiveness signalling (Figure 2D-F; *Online Supplementary Figure S4*). Per sex, all the significant DEG identified exhibited a unanimous direction of transcription, but different members of the pathway were differentially expressed in males and females. The DEG identified in male or female TI and TM patients could be potentially invaluable for the development of sex-specific treatment options and stratification strategies (*Online Supplementary Tables S5, S12-S13*; *Online Supplementary Figure S5*).

To our knowledge no other studies comparing gene expression profiles in males and females suffering from β-thalassemia currently exist, however, there have been reports of correlations of disease symptoms or complications related to sex. For instance, HbF levels have been found significantly higher in the female population of TM patients and this difference became more apparent after the age of 30 years.⁸ When considering complications of the disease, male TM patients have shown a strong association with diabetes⁹ and although no clear reason currently exists for such an association, it can be partly attributed to increased sensitivity of males to iron over-

load.¹⁰ Better survival rate has also been reported in females rather than males with fewer occurrences of cardiac complications and cardiac-based morbidities.¹¹ In terms of development of osteoporosis and osteopenia in TM patients, a sex difference was seen in the prevalence and the severity of the disorder with males being more frequently and severely affected than females.¹² In general, various pathways have been found to exhibit sex-related differences, many of which are linked to β-thalassemia, such as oxidative stress defense,¹³ lipid metabolism¹⁴ and erythropoietin activity.¹⁵ The present study, besides the identification of sex-specific transcriptional profiles in β-thalassemia through public availability of our data, represents a novel resource for meta-analyses and follow-up studies. In conclusion, our data highlight the need for considering sex as an important variable of the disease, which should be taken into account when developing differential diagnostic and therapeutic strategies.

Aikaterini Nanou,¹ Chrisavgi Toumpaki,¹ Pavlos Fanis,² Nicoletta Bianchi,³ Lucia Carmela Cosenza,³ Cristina Zuccato,³ George Sentis,¹ Giorgos Giagkas,¹ Coralea Stephanou,² Marios Phylactides,² Soteroula Christou,⁴ Michalis Hadjigavriel,⁵ Maria Sitarou,⁶ Carsten W. Lederer,² Roberto Gambari,³ Marina Kleanthous² and Eleni Katsantoni¹

¹Basic Research Center, Biomedical Research Foundation, Academy of Athens, Athens, Greece; ²Molecular Genetics Thalassaemia Department, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; ³Department of Life Sciences and Biotechnology, Ferrara University, Ferrara, Italy; ⁴Thalassaemia Clinic, Archbishop Makarios III Hospital, Nicosia, Cyprus; ⁵Limassol General Hospital, Department of Internal Medicine, Limassol, Cyprus and ⁶Thalassemia Clinic Larnaca, Larnaca General Hospital, Larnaca, Cyprus

Correspondence:

ELENI KATSANTONI - ekatsantoni@bioacademy.gr

doi:10.3324/haematol.2020.248013

Disclosures: no conflicts of interest to disclose.

Contributions: AN performed experiments, analyzed results and wrote the paper; CT, PF, NB, LCC, CZ and CS performed experiments; GS analyzed results and performed experiments; GG analyzed results; SC, MH and MS provided patient samples and evaluated the clinical picture of the patients; MP, CWL, RG and MK designed the research; EK designed the research, performed experiments, analyzed results and wrote the paper

Acknowledgments: the authors would like to thank Dr. Sjaak Philipsen for critical reading of the manuscript, GeneCore/EMBL for sequencing support and Panayiota Papasavva for helpful discussions.

Funding: this work was supported by the European Union's FP7 THALAMOSS (Project no. 306201 to E.K., R.G., M.K.), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813091 (E.K.) and by the Republic of Cyprus through the Research Promotion Foundation under grants agreements YTEIA/BIOΣ0609 (BE)/01 (E.K., M.K.) and YTEIA/BIOΣ/0311(BE)/20 (M.K.).

References

1. Colah R, Gorakshakar A, Nadkarni A. Global burden, distribution and prevention of β-thalassemias and hemoglobin E disorders. *Expert Rev Hematol.* 2010;3(1):103-117.
2. Thein SL. Genetic basis and genetic modifiers of β-thalassemia and sickle cell disease. *Adv Exp Med Biol.* 2017;1013:27-57.
3. Cosenza LC, Breda L, Breveglieri G, et al. A validated cellular biobank for β-thalassemia. *J Transl Med.* 2016;14(1):255.
4. Chaichompoo P, Pattanapanyasat K, Winichagoon P, Fucharoen S, Svasti S. Accelerated telomere shortening in β-thalassemia/HbE patients. *Blood Cells Mol Dis.* 2015;55(2):173-179.
5. Litanatudom P, Leecharoenkiat A, Wannatung T, Svasti S,

- Fucharoen S, Smith DR. A mechanism of ineffective erythropoiesis in β -thalassemia/Hb E disease. *Haematologica*. 2010;95(5):716-723.
6. Amendola G, Danise P, Todisco N, D'Urzo G, Di Palma A, Di Concilio R. Lipid profile in β -thalassemia intermedia patients: correlation with erythroid bone marrow activity. *Int J Lab Hematol*. 2007; 29(3):172-176.
 7. Livrea MA, Tesoriere L, Maggio A, D'Arpa D, Pintaudi AM, Pedone E. Oxidative modification of low-density lipoprotein and atherogenic risk in beta-thalassemia. *Blood*. 1998;92(10):3936-3942.
 8. el-Hazmi MA, Warsy AS, Addar MH, Babae Z. Fetal haemoglobin level-effect of gender, age and haemoglobin disorders. *Mol Cell Biochem*. 1994;135(2):181-186.
 9. Pes GM, Tolu F, Dore MP. Anti-thyroid peroxidase antibodies and male gender are associated with diabetes occurrence in patients with β -thalassemia major. *J Diabetes Res*. 2016;2016:1401829.
 10. Marsella M, Borgna-Pignatti C, Meloni A, et al. Cardiac iron and cardiac disease in males and females with transfusion-dependent thalassemia major: a $T2^*$ magnetic resonance imaging study. *Haematologica*. 2011;96(4):515-520.
 11. Marsella M, Pepe A, Borgna-Pignatti C. Better survival and less cardiac morbidity in female patients with thalassemia major: a review of the literature. *Ann NY Acad Sci*. 2010;1202:129-138.
 12. Kyriakou A, Savva SC, Savvides I, et al. Gender differences in the prevalence and severity of bone disease in thalassaemia. *Pediatr Endocrinol Rev*. 2008;6(Suppl 1):S116-122.
 13. Kander MC, Cui Y, Liu Z. Gender difference in oxidative stress: a new look at the mechanisms for cardiovascular diseases. *J Cell Mol Med*. 2017;21(5):1024-1032.
 14. Link JC, Reue K. Genetic basis for sex differences in obesity and lipid metabolism. *Annu Rev Nutr*. 2017;37:225-245.
 15. Soliz J, Khemiri H, Caravagna C, Seaborn T. Erythropoietin and the sex-dimorphic chemoreflex pathway. *Adv Exp Med Biol*. 2012; 758:55-62.