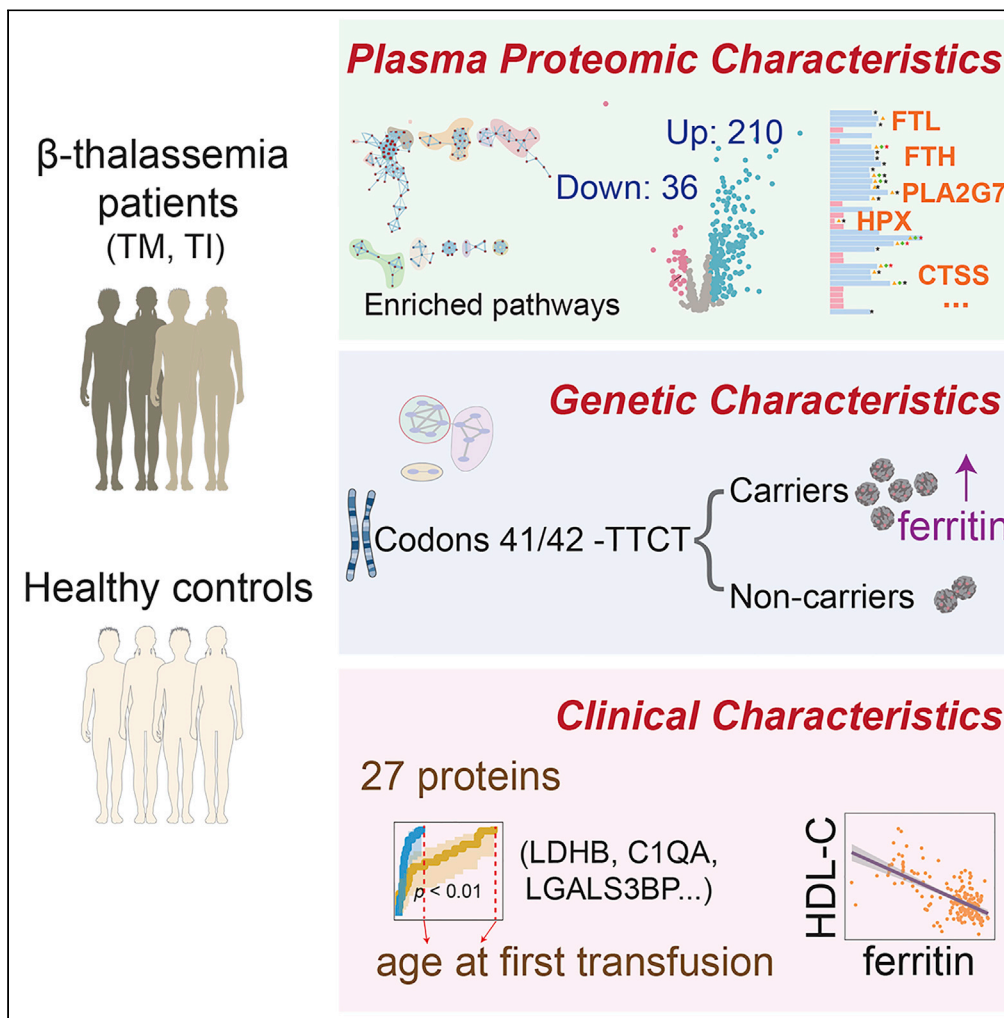


Article

Plasma proteome profiling combined with clinical and genetic features reveals the pathophysiological characteristics of  $\beta$ -thalassemia



Na Li, Peng An, Jifeng Wang, ..., Jianming Luo, Fudi Wang, Fuquan Yang

fqyang@ibp.ac.cn (F.Y.)  
fwang@zju.edu.cn (F.W.)

Highlights

246 dysregulated proteins are detected in plasma of patients with  $\beta$ -thalassemia

13 potential biomarkers and 27 proteins related to disease progression are found

Variations in plasma proteome reveal the disease pathophysiological characteristics

Codons 41/42 -TTCT carriers have higher ferritin levels compared to non-carriers



## Article

Plasma proteome profiling combined with clinical and genetic features reveals the pathophysiological characteristics of  $\beta$ -thalassemia

Na Li,<sup>1,2,6</sup> Peng An,<sup>3,6</sup> Jifeng Wang,<sup>1,6</sup> Tingting Zhang,<sup>1,2</sup> Xiaoqing Qing,<sup>1,2</sup> Bowen Wu,<sup>1,2</sup> Lang Sun,<sup>1,2</sup> Xiang Ding,<sup>1</sup> Lili Niu,<sup>1</sup> Zhensheng Xie,<sup>1</sup> Mengmeng Zhang,<sup>1</sup> Xiaojing Guo,<sup>1,2</sup> Xiulan Chen,<sup>1,2</sup> Tanxi Cai,<sup>1,2</sup> Jianming Luo,<sup>4</sup> Fudi Wang,<sup>5,\*</sup> and Fuquan Yang<sup>1,2,7,\*</sup>

## SUMMARY

**The phenotype of  $\beta$ -thalassemia underlies multigene interactions, making clinical stratification complicated. An increasing number of genetic modifiers affecting the disease severity have been identified, but are still unable to meet the demand of precision diagnosis. Here, we systematically conducted a comparative plasma proteomic profiling on patients with  $\beta$ -thalassemia and healthy controls. Among 246 dysregulated proteins, 13 core protein signatures with excellent biomarker potential are proposed. The combination of proteome and patients' clinical data revealed patients with codons 41/42 -TTCT mutations have an elevated risk of higher iron burden, dysplasia, and osteoporosis than patients with other genotypes. Notably, 85 proteins correlating to fetal hemoglobin (Hb F) were identified, among which the abundance of 27 proteins may affect the transfusion burden in patients with  $\beta$ -thalassemia. The current study thus provides protein signatures as potential diagnostic biomarkers or therapeutic clues for  $\beta$ -thalassemia.**

## INTRODUCTION

$\beta$ -thalassemia is one of the most common inherited single-gene diseases, with a high prevalence in tropical and sub-tropical areas (Marengo-Rowe, 2007; Peters et al., 2012). Epidemiology surveys indicate  $\beta$ -thalassemia is highly pervasive in the provinces of southern China, with a rate of 6.43% in the Guangxi Zhuang Autonomous Region and 3.80% in Guangdong Province (Li et al., 2014; Xiong et al., 2010).  $\beta$ -thalassemia results from mutations in the  $\beta$ -globin gene (*HBB* gene), causing reduced ( $\beta^{++}$  and  $\beta^+$ ) or absent ( $\beta^0$ ) synthesis of the  $\beta$ -globin chain of hemoglobin. The relative excess of  $\alpha$ -globin chains precipitate in the erythroid cell membrane, leading to ineffective erythropoiesis and hemolytic anemia. Subsequent clinical manifestations include splenomegaly, marrow expansion, extramedullary hemopoiesis, and bone marrow dilatation (Taher et al., 2018). Besides, excess iron from blood transfusion and ineffective erythropoiesis deposits in major organs, causing cardiac dysfunction, liver fibrosis, diabetes, and neurological complications (Palmer et al., 2018; Taher and Saliba, 2017).

The severity and clinical presentations of a patient with  $\beta$ -thalassemia are largely determined by the extent to which the mutation impairs the synthesis of  $\beta$ -globin chain. Thus,  $\beta$ -thalassemia classification is based on the heterozygous or homozygous state of the mutation(s).  $\beta$ -thalassemia trait (TT) is usually asymptomatic while homozygosity or compound heterozygosity for  $\beta$ -thalassemia mutations cause  $\beta$ -thalassemia intermedia (TI) and  $\beta$ -thalassemia major (TM) (Shang and Xu, 2017). However, varying levels of disease severity can be observed for patients with the same genotype. This is due to multiple factors contributing to the clinical severity of this disease, such as genetic modifiers involved in the production of fetal hemoglobin (Hb F) (Sripichai and Fucharoen, 2016) and the coexistence with  $\alpha$ -thalassemia (Sripichai et al., 2008). Moreover, complications from standard treatment (i.e., blood transfusion, iron chelation therapy, and splenectomy) also impact the clinical signs. This poor correlation between genotype and phenotype prevents precision clinical stratification. Therefore, exploration of the genotype-phenotype relationship and identification of biomarkers for disease are of paramount importance for precise diagnosis and treatment of the disease.

<sup>1</sup>Key Laboratory of Protein and Peptide Pharmaceuticals & Laboratory of Proteomics, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Department of Nutrition and Health, China Agricultural University, Beijing 100193, China

<sup>4</sup>Department of Pediatrics, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021 China

<sup>5</sup>The Fourth Affiliated Hospital, School of Public Health, State Key Laboratory of Experimental Hematology, Zhejiang University School of Medicine, Hangzhou 310058, China

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead contact

\*Correspondence: fcyang@ibp.ac.cn (F.Y.), fwang@zju.edu.cn (F.W.)

<https://doi.org/10.1016/j.isci.2022.104091>



Despite complicating the classification of  $\beta$ -thalassemia, certain genetic modifiers have been further studied as therapeutic targets of this disease. For example, BCL11A is a well known repressor of Hb F. Frangoul et al. (2021) reported that the production of Hb F was successfully reactivated in one patient with  $\beta$ -thalassemia after receiving autologous CD34<sup>+</sup> cells edited with CRISPR-Cas9 targeting the BCL11A enhancer. As increased Hb F is a well characterized ameliorating factor in patients, identifying proteins potentially affected by Hb F levels may provide new therapeutic approaches of the disease.

To date, extensive research of  $\beta$ -thalassemia has mainly used genomic or transcriptome methods. However, genetic or transcriptional information is not sufficient for the prediction of corresponding protein levels due to the poor correlation between mRNA and protein abundance (Greenbaum et al., 2003). Mass spectrometry (MS)-based proteomics has become a powerful approach for the exploration of biomarkers or pathophysiological characteristics in various diseases (Shen et al., 2020; Tewari et al., 2018; Wewer Albrechtsen et al., 2018). Previous studies of the proteomics profile of patients with  $\beta$ -thalassemia reported dysregulated proteins associated with the key pathological conditions implicated in thalassemia, including oxidative stress (Ponnikorn et al., 2019), hemolysis (Kittivorapart et al., 2018), and the hypercoagulable state (Chanpeng et al., 2019) when compared with healthy individuals. Several proteins correlating with the disease, for example, haptoglobin, hemopexin, and cathepsin S, were proposed as potential clinically relevant biomarkers (Kittivorapart et al., 2018). In addition, although the links to thalassemia are not yet clear, the novel differentially expressed proteins identified may lead to a better understanding of the pathophysiology of thalassemia (Hatairaktham et al., 2013). As differential constituents of plasma proteins can reflect the ongoing physiological and pathological state of an individual, we systematically carried out a deep plasma proteomic profiling study using tandem mass tags (TMT)-labeled quantitative proteomics to delineate protein molecular features in patients with  $\beta$ -thalassemia.

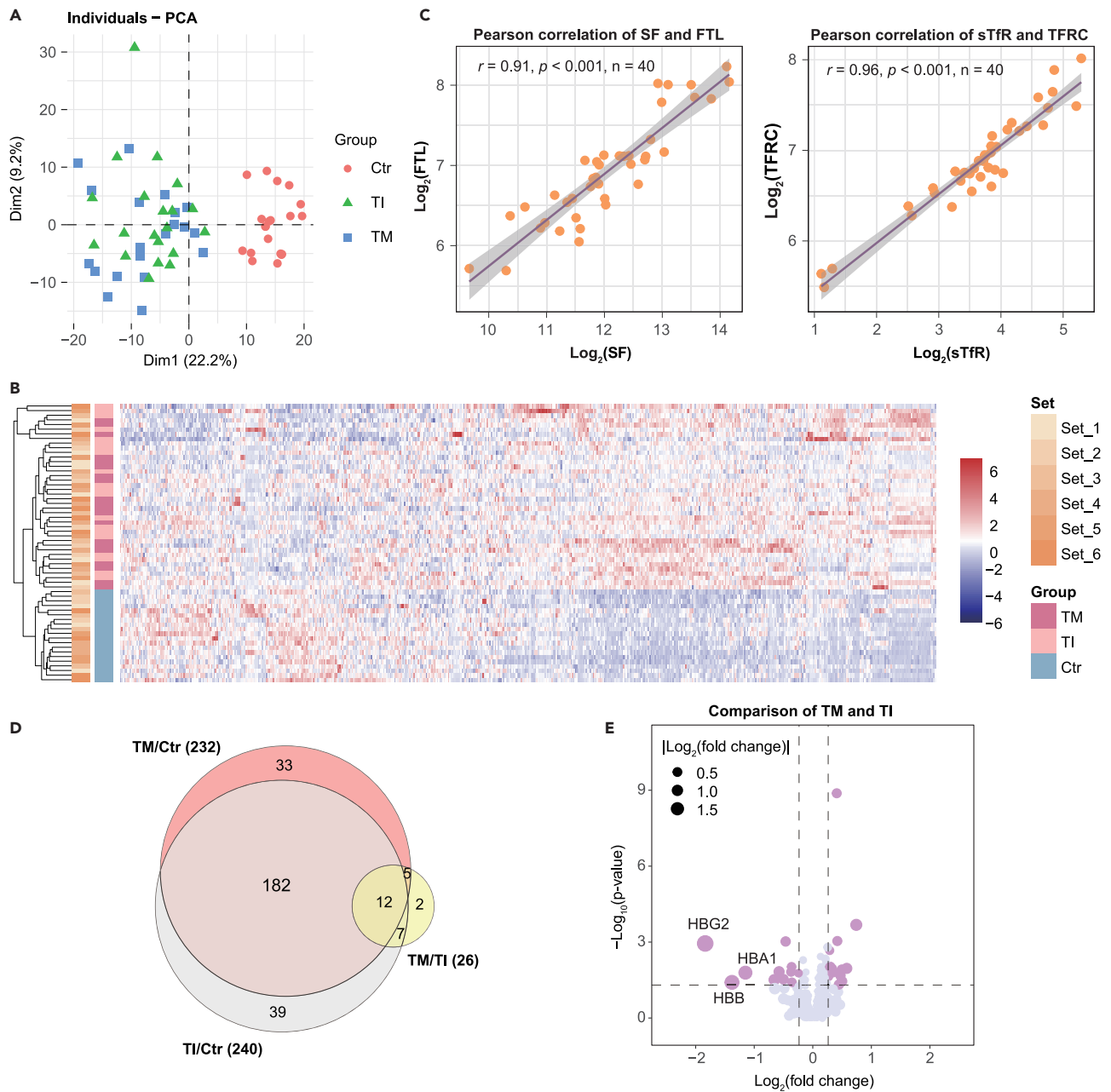
## RESULTS

### Overall synopsis of the plasma proteomic profiling of patients with $\beta$ -thalassemia

Plasma samples from 20 patients with TM, 20 patients with TI, and 20 healthy controls were used for the quantitative proteomic analysis. The overall workflow of this study is exhibited in Figure S1A. To monitor the LC-MS/MS platform instrument stability, 1  $\mu$ g of a tryptic digest of 293T whole-cell lysates was analyzed between batches. Similar distribution of peptide spectrum matches (PSMs) in all quality control samples and strong positive correlations between samples allowed for data acquisition of TMT-labeled samples (Figures S1B and S1C). We then performed unsupervised principal component analysis (PCA) and hierarchical clustering of quantified proteins. The samples were clustered by clinical groups but not TMT sets, indicating no significant batch effect or measurable bias (Figures 1A and 1B, Figure S2). Although the patients with  $\beta$ -thalassemia were clearly separated from the controls by PCA and the heatmap, these measures could not statistically separate patients with TM and TI. A correlation was noted between serum ferritin (SF) quantified by immunoassay methods and ferritin light chain (FTL) quantified by our proteomics ( $r = 0.91$ ,  $p < 0.001$ ). Transferrin receptor protein 1 (TFRC) and serum TfR (sTfR) also exhibited a high degree of correlation ( $r = 0.96$ ,  $p < 0.001$ ), as measured by TMT-based proteomics and immunoturbidimetry-based methods using a biochemistry analyzer, respectively (Figure 1C).

In the multiplexed TMT-based quantitative proteomics, a total of 868 proteins were identified with high confidence, of which 592 proteins were quantified across all samples. The following data analysis was performed on these 592 proteins and collected clinical data for 337 recruited individuals (Table S1). First, we calculated the protein fold changes in pairwise comparison between experimental groups and checked their correlation between TMT sets. Even with inter-individual proteomic heterogeneity, positive correlations indicated the consistency in proteomic features of patients with  $\beta$ -thalassemia, and the similarity between proteomic profiling of patients with TI and TM (Figures S3A and S3B).

Next, prior to differential protein analysis, we analyzed the contribution of quantitative variances from TMT sets and experimental groups. Quantitative variances across TMT sets were smaller than those brought by experimental groups (Figure S3D), indicating negligible batch effects. Therefore, one-way ANOVA was used for comparison between experimental groups. Of 280 significantly altered proteins in the three groups ( $p < 0.05$  and fold change  $>2$  or  $<0.85$ ), a comparison of patients with TM and healthy controls showed 193 upregulated proteins and 39 downregulated proteins in patients with TM. Similarly, a total of 199 upregulated and 41 downregulated proteins were identified in TI versus healthy control patients. However, only slight changes in protein abundance were found when comparing TM and TI



**Figure 1. Plasma proteomic profiling of patients with  $\beta$ -thalassemia and healthy controls**

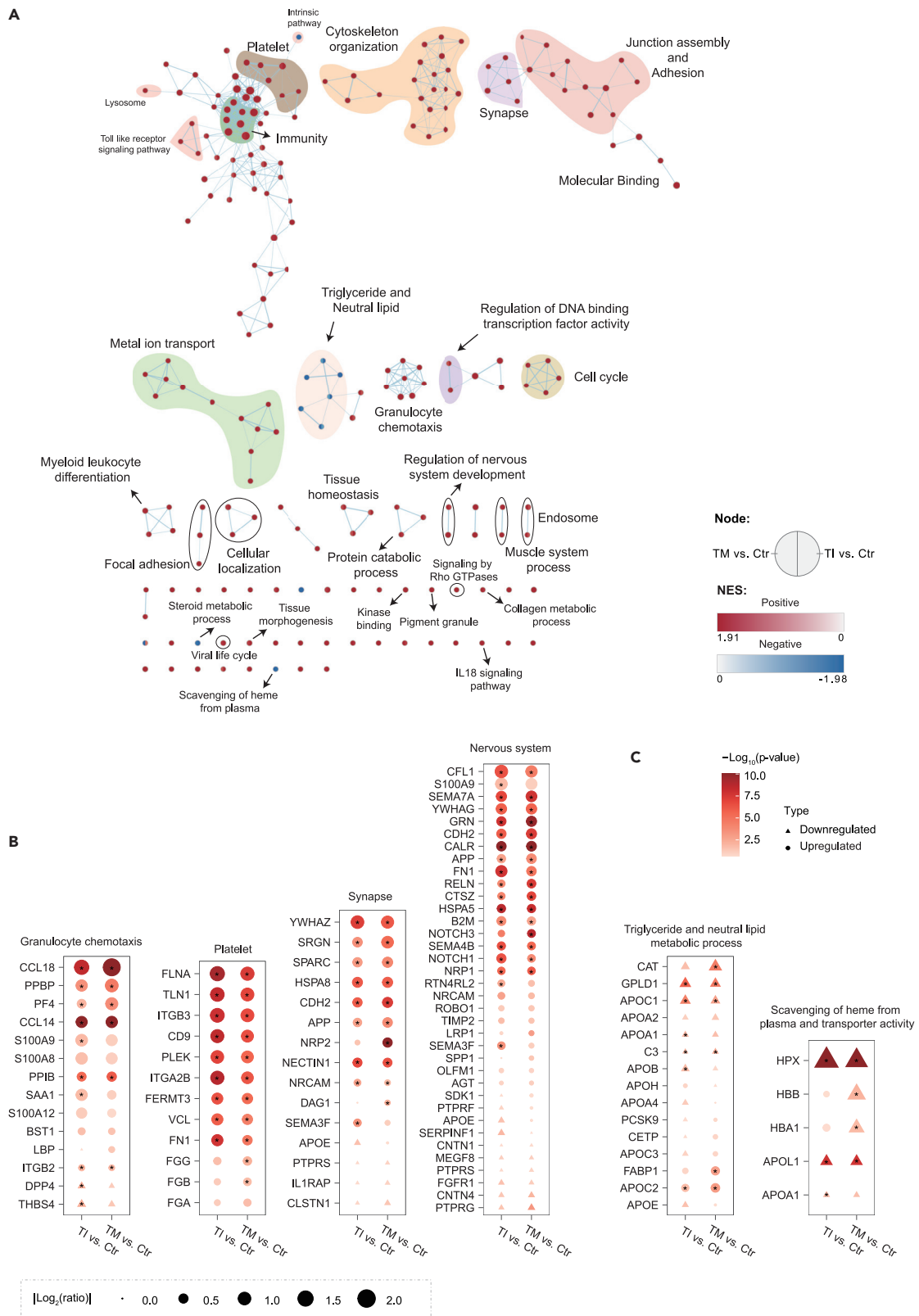
(A) PCA on proteins quantified in all samples of the investigated individuals. Patients with  $\beta$ -thalassemia are distinctly separated from healthy controls (Ctr), while TM and TI are mixed together.

(B) Heatmap of quantified proteins illustrating larger intergroup differences than intragroup variations. Set one to six represents six TMT sets in the experiment.

(C) Pearson correlation analyses of two pairs of protein indicators.

(D) The number of significantly altered proteins in each pairwise comparison. There are a similar number of dysregulated proteins in TM/Ctr (232 proteins) and TI/Ctr (240 proteins).

(E) Volcano plot of differential protein profiles in the comparison of TM and TI. Most of the significantly altered proteins were distinctly associated with mutated globin genes, such as HBG2, HBB, and HBA1. Colored dots represent proteins with p value < 0.05 and fold change < 0.85 or > 1.2. See also Figures S1–S3 and Table S2.



**Figure 2. Enrichment map analysis of altered pathways and protein processes**

(A) The significantly enriched pathways and protein processes of TM vs. control or TI vs. control. Each gene set is a node, of which the size is scaled to the number of genes in the gene set and the degree of overlap between gene sets is scaled to edge thickness. Quadrants delineate the comparison groups, in which the color is mapped to the normalization enrichment scores (NES). Red refers to the upregulated gene sets, whereas blue refers to the downregulated gene sets.

(B) The shared proteins in upregulated gene-set modules including platelet function, synapse, and nervous system.

(C) The shared proteins in downregulated gene-set modules including the triglyceride and neutral lipid metabolic process and scavenging of heme from plasma and transporter activity. (\* $p < 0.05$ ). Statistical analysis was performed using ANOVA, followed by post-hoc analysis using TukeyHSD with 95% of confidence. See also [Figure S4](#) and [Table S3](#).

([Figures 1D–1E](#) and [S3C](#), [Table S2](#)). Interestingly, a large number of proteins were upregulated in patients with  $\beta$ -thalassemia when compared with healthy controls ([Figure S3C](#)).

**Enrichment map analysis reveals biological pathways involved in  $\beta$ -thalassemia pathophysiology**

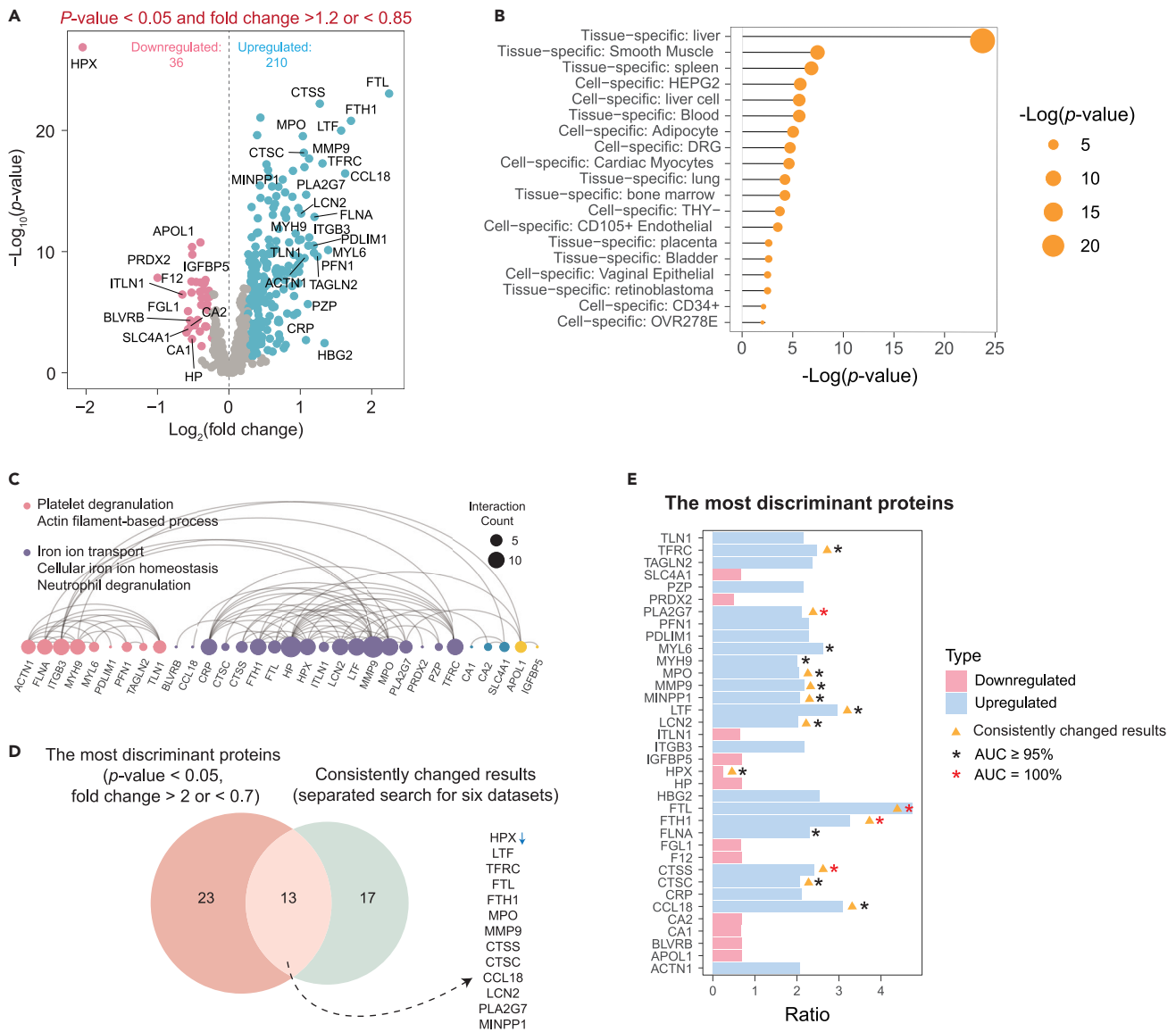
The plasma proteome can represent relevant characteristics of diseases ([Pernemalm et al., 2019](#); [Sun et al., 2019](#); [Tewari et al., 2018](#); [Wewer Albrechtsen et al., 2018](#)). To explore the altered pathways or protein processes caused by  $\beta$ -thalassemia, we conducted comparative enrichment map analysis on differential protein profiles in TM vs. control and TI vs. control ([Table S3](#)). The significantly enriched gene sets were organized into higher-level modules based on shared components to minimize the redundant gene-sets information ([Figure 2A](#)).

By capturing the entire enriched results of the gene set enrichment analysis (GSEA), we found the variation tendency of gene sets in TM vs. control and TI vs. control to be strikingly similar. In total, there were 115 overlapping gene sets (over 50%) significantly enriched in the two pairs, thereby strengthening the similarity of altered pathways and protein processes between TM and TI ([Figure S4A](#)). Compared with healthy controls, both patients with TM and TI were accompanied by dozens of upregulated functional modules. For instance, driven by C-C motif chemokine 18/14 (CCL18/CCL14), platelet basic protein (PPBP) and platelet factor 4 (PF4), abnormal granulocyte chemotaxis was enriched. Besides, platelet functions were prominently affected, including platelet activation, platelet aggregation, platelet degranulation, and response to elevated platelet cytosolic  $Ca^{2+}$ . Platelet activation ([Fayed et al., 2018](#)) and aggregation ([Winichagoon et al., 1981](#)) observed in patients with  $\beta$ -thalassemia are risk factors for hypercoagulopathy, resulting in thromboembolic events ([Cappellini et al., 2012](#); [Chanpeng et al., 2019](#)). In the above-mentioned platelet modules, we also observed a significant upregulation of pleckstrin (PLEK), whose intronic polymorphism is an independent genetic risk factor for venous thromboembolism ([Lindstrom et al., 2019](#)). Some other modules, such as regulation of nervous system development and synapse, consisting of 14-3-3 protein zeta/delta (YWHAZ), 14-3-3 protein epsilon (YWHAE), and amyloid-beta precursor protein (APP), emphasized the underlying neurodegenerative disorders linked to iron accumulation in the brain ([Manfre et al., 1999](#)) ([Figure 2B](#)).

Downregulated gene sets in patients included the triglyceride and neutral lipid metabolic process, steroid metabolic process, and scavenging of heme from plasma. The downregulated phosphatidylinositol-glycan-specific phospholipase D (GPLD1), apolipoprotein C-I (APOC1), and catalase (CAT) are the leading proteins in the triglyceride and neutral lipid metabolic process. Impaired lipid metabolism in patients with  $\beta$ -thalassemia is further supported by the higher triglyceride (TG) levels found in patients with  $\beta$ -thalassemia compared with healthy controls ([Figure S4B](#)). Significant downregulation of hemopexin (HPX), HBB, hemoglobin subunit alpha (HBA1), apolipoprotein L1 (APOL1), and apolipoprotein A-I (APOA1) was identified in patients with  $\beta$ -thalassemia, contributing to the enrichment of the scavenging of heme from plasma pathway ([Figure 2C](#)). Heme is a toxic and proinflammatory molecule that activates the Toll-like receptor signaling pathway, thus inducing the production of reactive oxygen species and inflammatory cytokines ([Belcher et al., 2014](#)). As the main plasmatic scavenger of heme, HPX transports heme to the liver to be broken down. The deficiency of HPX reflects the severity of hemolysis in thalassemia ([Smith and McCulloh, 2015](#)).

**Differential proteomic analysis identifies potential diagnostic biomarkers of  $\beta$ -thalassemia**

As mentioned above, patients with TM and TI have quite similar plasma proteomic profiles. The discrepancies in differential proteins between TM and TI were small; however, the majority of these proteins were directly related to the pathophysiology of the disease, such as hemoglobin subunit gamma-2



**Figure 3. Comparison of patients with  $\beta$ -thalassemia and healthy controls**

(A) Volcano plot showing the downregulated proteins (pink) and upregulated proteins (blue) in comparison of patients with  $\beta$ -thalassemia and healthy individuals.

(B) Altered proteins were submitted to Metascape for tissue-specific enrichment analysis using the PaGenBase database.

(C) Interaction network analysis of proteins with  $p$  value < 0.05 and fold change > 2 or < 0.7 using STRING. The enriched biological functions in the GO database for each cluster are marked.

(D) Venn diagram showing the intersection of the most discriminant proteins and the consistently changed results from six datasets. The 13 core protein signatures include TFRC, PLA2G7, MPO, MMP9, MINPP1, LTF, LCN2, HPX, FTL, FTH1, CTSS, CTSC, and CCL18. Except for HPX, all of the core proteins are upregulated in patients.

(E) List of the most discriminant proteins. The proteins with  $AUC \geq 95\%$  or presented consistent results are respectively labeled. See also [Figure S5](#) and [Tables S4](#) and [S5](#).

(HBG2), HBA1, and HBB ([Figure 1E](#)). Therefore, differential protein analysis was concentrated on patients with  $\beta$ -thalassemia and healthy controls to discover potential diagnostic biomarkers of  $\beta$ -thalassemia.

Significant dysregulation was observed in 246 proteins, accounting for approximately 40% of all quantified proteins ([Figure 3A](#), [Table S4](#)). As a result of cell death or damage, tissue proteins are proposed to leak into the blood. To explore the sources of these altered proteins, they were queried using the PaGenBase

(Pan et al., 2013), a database repository for the collection of tissue- and time-specific pattern genes (Figure 3B). As plasma proteins are largely secreted by the liver (Anderson and Anderson, 2002), it is no surprise that we found that the liver is the most significant tissue source for differential proteins. In addition, proteins from muscle, spleen, adipocyte, cardiac myocyte, bone marrow, and CD34<sup>+</sup> were also enriched. This observation is consistent with current knowledge about organ damage brought on by iron deposition and ineffective erythropoiesis in  $\beta$ -thalassemia (Taher et al., 2021).

The differentially expressed proteins between patients with  $\beta$ -thalassemia and healthy individuals then underwent a protein-protein interaction enrichment analysis using Metascape (Zhou et al., 2019) (Figure S5). Densely connected protein complexes detected by the Molecular Complex Detection (MCODE) clustering algorithm (Bader and Hogue, 2003) were also highly enriched in the pathways or processes revealed by the enrichment map analysis shown in Figure S5. These include the complement system, immune cell response, platelet degranulation, and actin-myosin-associated movement. Moreover, glycolysis and gluconeogenesis, post-translational protein phosphorylation, and regulation of IGF transport and uptake by IGFs were markedly enriched.

To further screen the most discriminant proteins, a stricter criterion  $p$  value  $< 0.05$  and fold change  $> 2$  or  $< 0.7$  was used. A total of 36 proteins were identified, which were divided into two main clusters by network interaction analysis (Table S5). One cluster annotated by gene ontology (GO) corresponds to platelet degranulation and the actin filament-based process; the other cluster relates to iron metabolism (Figure 3C). Furthermore, we assessed the differential proteins consistently changed across all six TMT sets by separately processing the raw data to avoid the impact from batches. At the general significance levels ( $p$  value  $< 0.05$  and fold change  $> 2$  or  $< 0.7$ ), the intersection of differential proteins in the six TMT sets resulted in the identification of 30 proteins. The core protein signatures of 13 proteins were identified by the combination of both the stricter criterion and consistently changed results from the separated analyses, as demonstrated in Figure 3D.

With respect to clinical application, diagnostic biomarker candidates should meet the requirements of sensitivity and specificity. Thus, receiver operating characteristic (ROC) curve analysis was performed for the most discriminant proteins and the corresponding area under curve (AUC) values were calculated. With the highest AUC, FTL, ferritin heavy chain (FTH1), cathepsin S (CTSS), and platelet-activating factor acetylhydrolase (PLA2G7) were identified to have the best diagnostic potential (Figure 3E). FTL and FTH1 are the light and heavy chains of ferritin, respectively. The increase of ferritin directly reflects the iron overload status in patients with  $\beta$ -thalassemia. CTSS is a member of the family of cysteine cathepsin. It acts as a lysosomal protease, which promotes the degradation of damaged or harmful proteins via the lysosome. The fourth protein, PLA2G7, is reported to be associated with cardiovascular disease risk (Jensen et al., 2014).

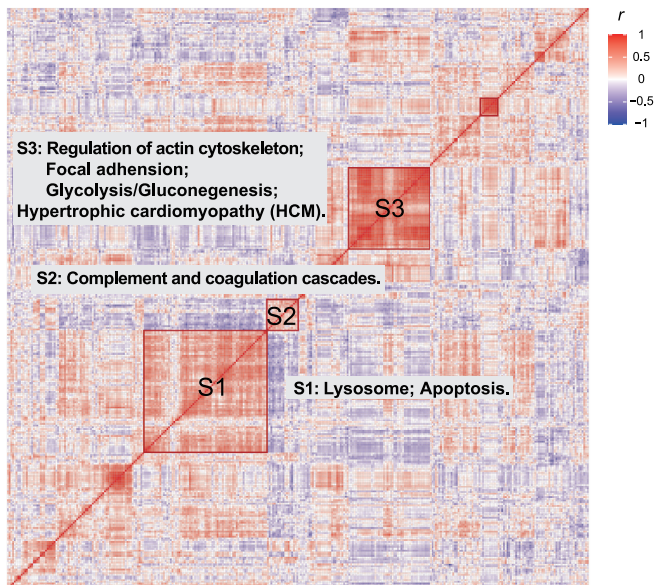
### Global correlation analysis on altered proteins and clinical data

Global correlation analysis was performed to discover co-regulated protein functions and the links between altered proteins and patients' clinical data. Three high correlation squares (S1, S2, and S3) with different enriched pathways are shown in Figure 4A (Table S6). Four of these pathways were significantly upregulated in the GSEA (Figure 4B), including the lysosome, regulation of actin cytoskeleton, focal adhesion, and hypertrophic cardiomyopathy pathways. The upregulated hypertrophic cardiomyopathy in S3 recapitulates iron-overload-induced cardiomyopathy in  $\beta$ -thalassemia, which is the leading cause of mortality in patients with  $\beta$ -thalassemia (Shah et al., 2019). Interestingly, compared with healthy controls, the lysosome pathway is upregulated more significantly in TM than T1 (Figure 4C). Here, iron overload in  $\beta$ -thalassemia causes large amounts of iron to accumulate in the lysosome, which may jeopardize lysosomal membrane integrity (Terman and Kurz, 2013). In addition, one major cause of ineffective erythropoiesis is accelerated apoptosis, a pathway involving the lysosomal degradation of cytoplasmic components (Rund and Rachmilewitz, 2005). Therefore, the lysosome pathway may be directly related to the severity of the mutations.

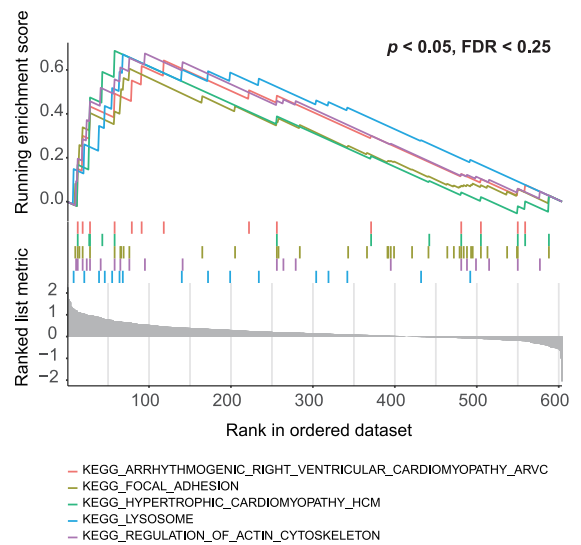
We observed that ferritin in S1 correlated negatively with high-density lipoprotein cholesterol (HDL-C). Thus, we analyzed the relationship between ferritin and lipid profiles in 170 patients with  $\beta$ -thalassemia. HDL-C showed significantly negative correlations with ferritin after adjusting for age and body mass index ( $r = -0.64$ ), while there was no correlation between ferritin and TG or between ferritin and total cholesterol (T-CHO) (Figure 4D, Figure S6).



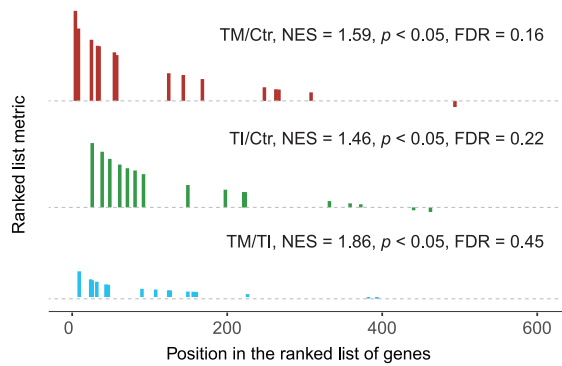
**A Global correlation analysis**



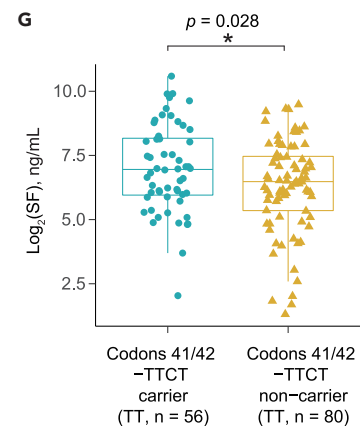
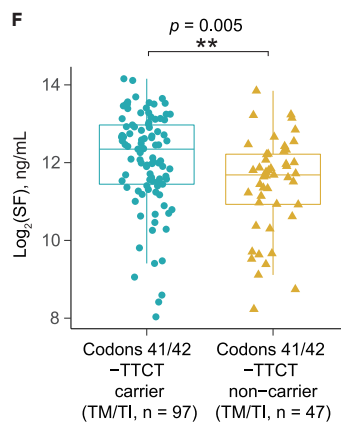
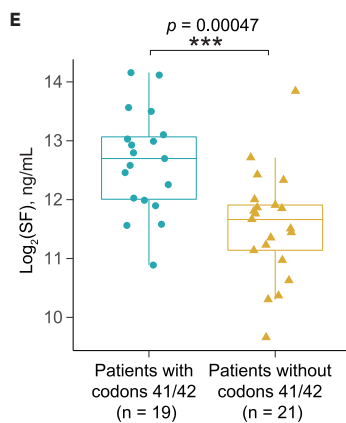
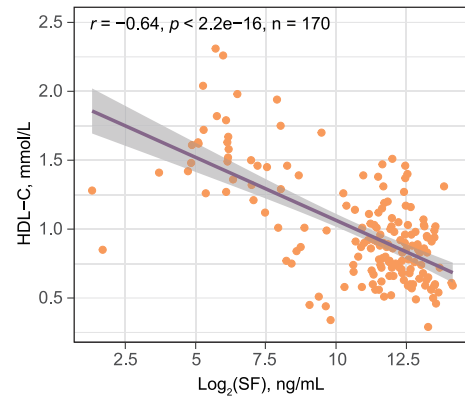
**B**



**C KEGG\_LYSOSOME**



**D Pearson correlation of SF and HDL-C**



**Figure 4. Correlation analysis of quantified proteins and clinical data**

(A) Three high correlation squares were found by global correlation analysis of altered proteins and clinical data.  
(B) The pathways revealed in high correlation squares were examined in GSEA.

**Figure 4. Continued**

(C) Poster of the upregulated lysosome pathway in the two comparison groups.

(D) Validation for the negative correlation of ferritin and HDL-C (n = 170).

(E–G) Ferritin levels of codons 41/42 -TTCT mutation carriers and non-carriers who participated in proteomic analysis. Validation for the higher ferritin levels of codons 41/42 -TTCT mutation carriers than that of non-carriers in (F) TM/TI and (G) TT. The center line in the boxplot depicts the median, the box limits depict the upper and lower quartiles, and the whiskers encompass 1.5 times the IQR. (\*p < 0.05, \*\*p < 0.01, and \*\*\*p < 0.001). The significance was calculated using two-sided unpaired t-test. See also [Figures S6–S9](#) and [Table S6](#).

In addition, ferritin also significantly correlated with the codons 41/42 -TTCT mutation. As one of the most common mutation types in China, codons 41/42 -TTCT accounts for about 30% of the genotypes ([Kountouris et al., 2014](#)). For further analysis, all of the patients with  $\beta$ -thalassemia were divided into two groups: codons 41/42 -TTCT carriers and non-carriers. Unlike other types of high-frequency mutations, the ferritin levels of codons 41/42 -TTCT carriers were significantly higher than that of non-carriers among all patients with TM/TI and TT ([Figures 4E–4G](#) and [S7](#)). The same results were also observed in all patients with  $\beta^0/\beta^0$  or  $\beta^0/\beta^N$  ([Figure S8](#)). These results indicate that patients with TM with codons 41/42 -TTCT have significantly higher iron burden than those with other mutations. Likewise, even though TT does not require blood transfusion, such carriers may require more attention to monitor their iron content.

We also comparatively analyzed the differential proteins between codons 41/42 -TTCT mutation carriers and non-carriers. These altered proteins were mainly enriched into three significant functional components: the regulation of IGF transport and uptake by IGF-BPs, the lysosome, and low-density lipoprotein particle ([Figure S9A](#)). Among them, the proteins associated with the insulin-like ternary complex were all downregulated ([Figure S9B](#)), while the proteins associated with the lysosome were all upregulated.

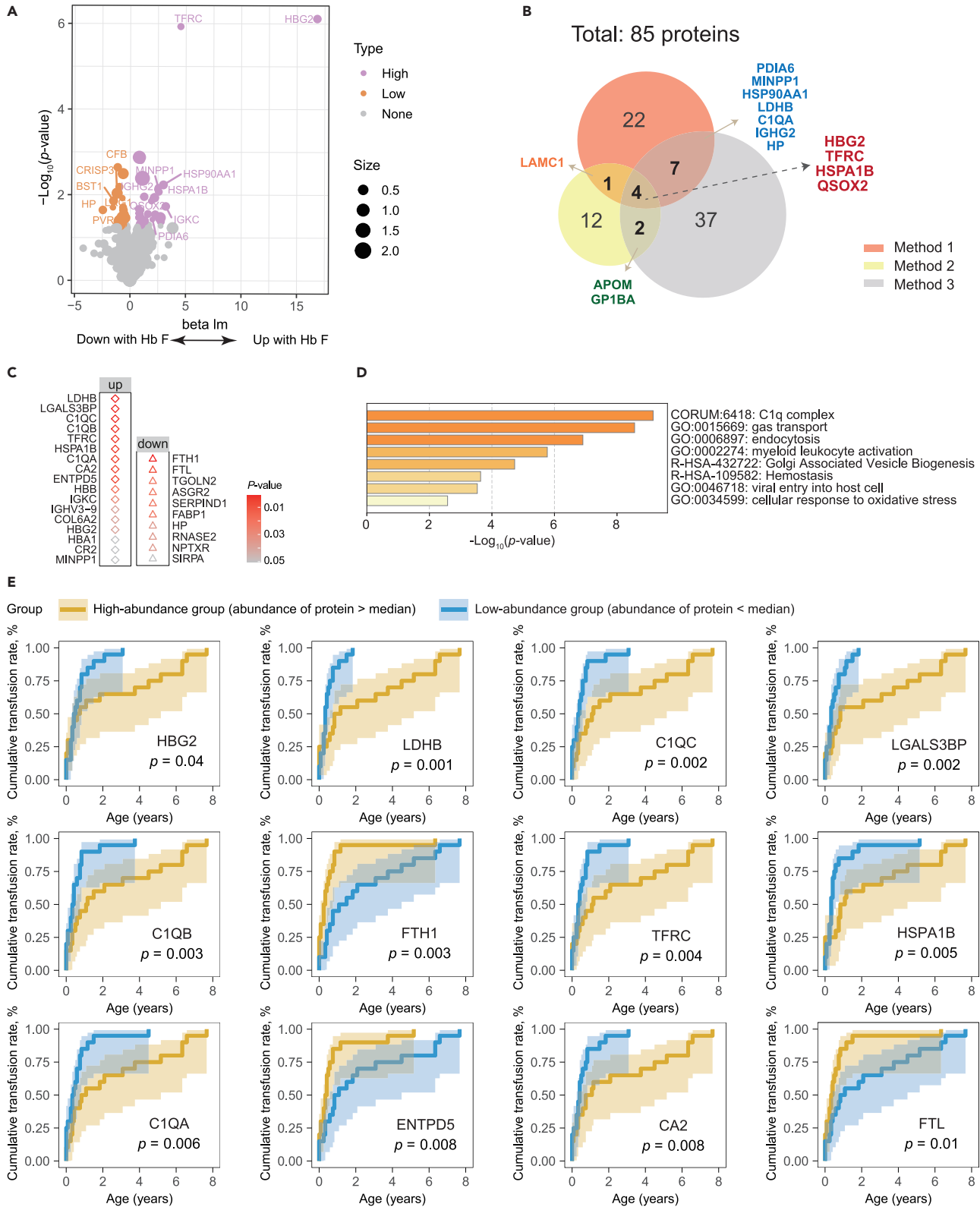
**Plasma proteins related to the fetal hemoglobin level and the progression of  $\beta$ -thalassemia**

Increased Hb F levels in patients with  $\beta$ -thalassemia can compensate for the loss of  $\beta$ -globin, reduce the redundant  $\alpha$ -globin chain in red blood cells, and ameliorate the ineffective erythropoiesis and anemia. To identify proteins potentially affected by Hb F levels, three methods were used to distinguish proteins associated with Hb F levels. For the first and second methods, patients were categorized into high Hb F and low Hb F groups according to the upper and lower quartile of Hb F levels (first method) or Hb F levels  $\geq 10$  g/dL and  $< 10$  g/dL (second method). This is because Hb F levels  $\geq 10$  g/dL have been reported to alleviate transfusion burden in patients ([CRISPR gene therapy trial on hold, 2018](#)). Comparative proteome analyses were then applied and the number of differential proteins identified using the two methods were 34 and 19, respectively. In the third method, Hb F levels were directly correlated with plasma proteins after adjusting for disease status, and 50 proteins were significantly associated with Hb F level ([Figure 5A](#)). In total, 85 proteins related to Hb F levels were identified using the three methods ([Figure 5B](#)).

We then wanted to determine whether these proteins are associated with the progression of  $\beta$ -thalassemia. The age at first transfusion represents the severity of ineffective hematopoiesis ([Danjou et al., 2012](#)). Based on the age at first transfusion, cumulative incidence rate of patients with  $\beta$ -thalassemia receiving transfusion therapy (referred to as cumulative transfusion rate here) were compared in patients with higher or lower levels of Hb F-related proteins, as shown in [Figure 5B](#). Kaplan-Meier curve analyses showed that the levels of 27 proteins were associated with cumulative transfusion rates, with these proteins relating to the C1q complex, gas transport, and myeloid leukocyte activation ([Figures 5C](#) and [5D](#)). Using a more stringent significance level (p value  $\leq 0.01$ ), [Figure 5E](#) shows proteins with higher abundance may alleviate the transfusion burden of patients with  $\beta$ -thalassemia, including L-lactate dehydrogenase B chain (LDHB), C1QC, galectin-3-binding protein (LGALS3BP), C1QB, TFRC, heat shock 70 kDa protein 1B (HSPA1B), C1QA, and carbonic anhydrase 2 (CA2); conversely, lower abundance of FTH1, ectonucleoside triphosphate diphosphohydrolase 5 (ENTPD5), or FTL was potentially associated with alleviated transfusion burden. These proteins in patients displayed higher significance levels than HBG2 (the subunit of Hb F; p value = 0.04) with respect to the cumulative transfusion rate.

**DISCUSSION**

Exploration of proteins closely associated with  $\beta$ -thalassemia phenotypes can illuminate the underlying mechanisms involved in disease pathogenesis, ultimately improving the diagnosis and clinical management of  $\beta$ -thalassemia. To this end, our study utilized a TMT-based quantitative proteomics approach to characterize the plasma differences between 40 patients with  $\beta$ -thalassemia and 20 healthy controls.



**Figure 5. Proteins associated with Hb F expression**

- (A) The relationship between Hb F levels and other proteins was estimated by linear regression analysis. Volcano plots representing changes of the plasma proteome with Hb F level.
- (B) Venn diagram of identified proteins in the three methods. In total, 85 proteins were found to be related to Hb F level.
- (C) Proteins that significantly associated with cumulative incidence rate of patients with  $\beta$ -thalassemia receiving transfusion therapy. For Kaplan-Meier curve analysis, each patient was classified into high/low-abundance subgroups based on the median protein abundance. The log rank test was used to assess the statistical significance of differences between subgroups and the corresponding p value is scaled to the color. Up: proteins in patient with higher levels correspond to a later time for first transfusion. Down: proteins in patient with lower levels correspond to a later time for first transfusion.
- (D) Bar graph of enriched terms in Metascape across proteins related to cumulative transfusion rates.
- (E) Kaplan-Meier curve analysis presenting significant difference of overall age at first transfusion between the high-abundance and low-abundance groups (high-abundance group: patients with protein abundance > median; low-abundance group: patients with protein abundance < median). HBG2 and proteins with  $p < 0.01$  are shown. The statistical analysis was performed using the log rank test.

With the combination of high-abundance protein removal and prefractionation techniques, we achieved an extensive identification of the plasma proteome and detected a plethora of dysregulated proteins.

Of the 246 differential proteins detected in patients with  $\beta$ -thalassemia and healthy controls, 13 core protein signatures exhibited significant biomarker potential for  $\beta$ -thalassemia diagnosis (Table S7), of which most have direct or indirect links to  $\beta$ -thalassemia. For instance, the upregulation of TFRC, lactotransferrin (LTF), FTH1, and FTL is directly related to iron overload and the impaired hemoglobin synthesis in patients. Myeloperoxidase (MPO) is an oxidative stress biomarker, and upregulation of neutrophil gelatinase-associated lipocalin (LCN2) was suspected to decrease ROS or iron in  $\beta$ -thalassemia (Roudkenar et al., 2008). Excess heme produced by hemolysis in  $\beta$ -thalassemia depletes endogenous HPX (Muller-Eberhard et al., 1968), the most prominent downregulated protein in this study. Administration of exogenous HPX has been shown to prevent heme-iron loading in the cardiovascular system (Vinchi et al., 2013) and revert the heme-induced proinflammatory state of macrophages (Vinchi et al., 2016). Importantly, significant changes in the abundance of HPX and CTSS were also observed in the thalassemic plasma extracellular vesicles by Kittivorapart et al. (2018), and evaluated plasma proteins including PLA2G7, CCL18, and LCN2, have been reported separately in other comparative studies of  $\beta$ -thalassemia (Dimitriou et al., 2005; Tselepis et al., 2010), reinforcing their potential to be clinical biomarkers in plasma. However, Gumus et al. (2016) did not find a significant difference in plasma MMP9 levels between patients with  $\beta$ -thalassemia and healthy individuals. Notably, this is the first report of elevated multiple inositol polyphosphate phosphatase 1 (MINPP1) in the plasma of patients with  $\beta$ -thalassemia. As elevated expression of MINPP1 involved in pontocerebellar hypoplasia (Appelhof et al., 2020; Uuncu et al., 2020), it may be a modulator of neurodegenerative syndrome in  $\beta$ -thalassemia.

The diverse and complex clinical manifestations observed in patients with  $\beta$ -thalassemia prompted us to further investigate the perturbed pathways based on the previous identification of differential proteins. Our analysis of differential protein profiles using an enrichment map revealed a series of pathways affected by  $\beta$ -thalassemia, the majority of which are confirmed or proposed to be related to the disease. It is worth mentioning that these upregulated pathways, such as myeloid leukocyte differentiation, regulation of DNA binding transcription factor activity, and the cell cycle may be caused by the compensatory increase in hematopoiesis. We also observed that the synapse and regulation of nervous system development were upregulated in patients, which implies the pathophysiological change in the nervous system. Hemoglobin was reported to be expressed in neurons (Biagioli et al., 2009). The impaired production of hemoglobin and iron deposition in the nerve system may be involved in neurodegenerative diseases such as Alzheimer disease (Altinoz et al., 2019; Bagwe-Parab and Kaur, 2020).

Global correlation analysis found the lysosome pathway is upregulated in turn among patients with TM, patients with T1, and healthy controls. Concurrently, hydrolytic enzymes in the lysosome (cathepsin B, cathepsin C, cathepsin D, cathepsin S, cathepsin H, cathepsin Z, NPC intracellular cholesterol transporter 2, lysosome-associated membrane glycoprotein 2, and prosaposin) showed the same trend and high detection accuracy (Figure S10). This suggests that these hydrolytic enzymes may serve as indicators for the severity of  $\beta$ -thalassemia. A possible explanation for the correlation of the lysosome with disease severity is iron accumulation in the intracellular compartments. Transferrin-bound or ferritin-bound iron can be transported into lysosome for recycling (Li et al., 2010). Accumulation of large amounts of iron in the lysosome yields strong reactive oxygen species (ROS), which may impair lysosomal activity. Impaired lysosomal activity by ROS accumulation has also been reported to induce ferroptotic cell death triggered

by HSP90, along with HSC70 and Lamp-2a (Wu et al., 2019). In our plasma proteomics, increased levels of HSP90 and Lamp-2a proteins underline an important role of the lysosome in  $\beta$ -thalassemia (Table S2). In addition, hypertrophic cardiomyopathy, a highly lethal complication of iron overload, was enriched in high correlation square S3.

In contrast to the upregulated signaling pathways, the triglyceride and neutral lipid metabolic pathway was significantly downregulated (Figure 2C). This implies a dysregulation of lipid metabolism, as abnormal TG, HDL-C, and low-density lipoprotein cholesterol (LDL-C) levels were seen in patients (Figure S4B). Moreover, high levels of plasma lipids and lipid peroxidation were commonly found in previous studies investigating the lipid profiles of patients with  $\beta$ -thalassemia (Madani et al., 2011; Nasr et al., 2008; Rachmilewitz et al., 1976). The observation of altered lipid metabolic pathways in the plasma further validates our proteomic strategy.

Although patients with TM and TI displayed some obvious distinctions in their clinical manifestations, hematological parameters, and complications (Karimi et al., 2014), we observed a minimal difference between patients with TM and TI in their plasma proteome. The reasons may be as follows: first, TI and TM share same pathological mechanisms in the erythroid cell; second, TI and TM were categorized based on the genotypes, but the disease severity of patients with same genotype are always affected by genetic or environmental modifiers. However, several pronounced changes were still identified between patients with TI and TM, especially in proteins directly related to disease pathophysiology. For example, the incremental abundances of FTL and FTH1 among Ctr, TI, and TM directly support clinical distinguish of TI and TM by transfusion dependence and degree of ineffective erythropoiesis. Moreover, the incremental abundances of lysosome-associated proteins among three groups were also apparent, indicating a continuous but not distinct variation spectrum of the disease from TI to TM (Figure S10).

Despite the minimal difference between TI and TM in their plasma proteome, we noticed that ferritin levels of codons 41/42 -TTCT carriers are higher than that of non-carriers, no matter which type of  $\beta$ -thalassemia (TM, TI, or TT) the person has. This is consistent with evidence that more severe clinical symptoms are observed in patients with  $\beta^0$ -thalassemia with either the homozygous or heterozygous codons 41/42 -TTCT mutation (Laosombat et al., 2001). It also reinforces that TT with the codons 41/42 -TTCT mutation may require more strict iron status monitoring than other  $\beta$ -thalassemia genotypes. Differential proteins between codons 41/42 -TTCT carriers and non-carriers also highlighted the downregulation of the insulin-like ternary complex, including IGF1, IGFBP3 (insulin-like growth factor-binding protein 3), IGFBP5 (insulin-like growth factor-binding protein complex acid labile subunit), and IGFBP3 (Figure S9). As reported previously, patients with  $\beta$ -thalassemia with growth retardation had lower IGF1 and IGFBP3 levels than those without growth retardation (Wu et al., 2003). Furthermore, ferritin levels have been reported to be significantly higher in patients with  $\beta$ -thalassemia with IGF1 at less than 0.1 percentile compared to patients with normal IGF1 (Roth et al., 1997). IGF1 is a major effector of bone development and IGFBPs mediate the bioavailability of IGFs (Clemmons, 2018). Disruption of both IGF1 and IGFBP3 genes in mice exhibited a dramatic retardation in growth (Yakar et al., 2002). In addition, a positive correlation between bone mineral density and the levels of IGF-1 and IGFBP3 in patients with  $\beta$ -thalassemia was observed (Lasco et al., 2002; Soliman et al., 1998). Therefore, patients with the codons 41/42 -TTCT mutation are more likely to acquire iron overload compared to patients with other mutations. This in turn leads to increased risk of growth retardation, dysplasia, and osteoporosis diseases. Patients with the codons 41/42 -TTCT mutation thus also need to monitor growth and development. To the best of our knowledge, this is the first time these associations in patients with  $\beta$ -thalassemia with the codons 41/42 -TTCT mutation have been reported.

Plasma proteins associated with Hb F levels could potentially be utilized as therapeutic targets of  $\beta$ -thalassemia. Our analysis revealed 85 candidate proteins. Furthermore, cumulative transfusion rate analyses in patients suggested 27 of these proteins are potentially related to the progression of  $\beta$ -thalassemia (Table S7), such as LDHB, C1QC, and LGALS3BP (Figure 5C). For example, LGALS3BP is a negative regulator of the NF- $\kappa$ B signaling pathway (Hong et al., 2019). Evidence indicates that NF- $\kappa$ B plays a crucial role in erythropoiesis (Jeong et al., 2011; Zhang et al., 1998) while the induction of Hb F by decitabine may be through the suppression of NF- $\kappa$ B activity (Theodorou et al., 2020). Therefore, increased LGALS3BP protein expression may suggest restricted NF- $\kappa$ B activity and increased Hb F production in the erythroid cells. Further work is required to delineate the roles of these proteins in  $\beta$ -thalassemia. To our knowledge,

this is the first report showing significant associations between these proteins and patients' cumulative transfusion rates.

In summary, this study revealed protein signatures potentially involved in the pathophysiology of  $\beta$ -thalassemia utilizing deep plasma proteomics, coupling genetic and clinical information. We further discovered plasma proteins correlated with Hb F levels and that could affect transfusion burden in patients with  $\beta$ -thalassemia. These proteins can be further explored as potential diagnostic biomarkers or therapeutic targets of  $\beta$ -thalassemia.

### Limitations of study

Owing to the relatively small number of participants, the diagnostic nature of the 13 core protein signatures awaits further verification in larger independent populations. Besides, co-inheritance of  $\beta$ -thalassemia and other hemoglobinopathies like  $\alpha$ -thalassemia complicates clinical manifestations. Whether these proteins could be used in such conditions needs to be further explored. Another important finding of this study is that codons 41/42 carriers have higher ferritin levels, implying a higher risk or level of iron overload. However, the lack of long-term follow-up studies limits the assessment of growth and development in these carries. In addition, functional investigations of these proteins related with Hb F levels and cumulative transfusion rates are needed.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Patients and healthy participants
- METHOD DETAILS
  - $\beta$ -thalassemia genotyping
  - Hematological and serum parameters
  - High-abundance protein removal and protein digestion
  - TMT 10-plex labeling
  - High pH reversed-phase HPLC fractionation
  - LC-MS/MS analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - MS data analysis
  - Bioinformatics analysis

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104091>.

### ACKNOWLEDGMENTS

This study was supported by a grant from the National Key R&D Program of China (nos. 2018YFA0507801 and 2018YFA0507802). We thank Dr. Yuezhen Li, Ms. Yanfei Hou and Ms. Yuning Liu for assisting in blood collection.

### AUTHOR CONTRIBUTIONS

Li, N., An, P., Wang, F., and Yang, F. conceptualized and designed the study. Clinical samples were collected by An, P. and Luo, J. An, P. designed the patient questionnaire and finished all the clinical data detection and collection. Li, N. performed the experimental work and the data analysis, generated figures for the manuscript, and wrote the first draft of the manuscript. An, P. contributed to the data analysis and edited the manuscript. Wang, J. optimized the DDA workflow for data acquisition and was responsible for LC-MS/MS data collection. Yang, F. and Wang, F. supervised the project. Zhang, T., Wu, B., Qing, Q., and Sun, L. contributed to the experimental work and data analysis. Cai, T., Chen, X., Ding, X., Niu, L.,

Xie, Z., Zhang, M., and Guo, X. provided helpful inputs on MS-based proteomic data collection and contributed comments to the manuscript. All authors read and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 3, 2022

Revised: February 17, 2022

Accepted: March 14, 2022

Published: April 15, 2022

## REFERENCES

- Altinoz, M.A., Guloksuz, S., Schmidt-Kastner, R., Kenis, G., Ince, B., and Rutten, B.P.F. (2019). Involvement of hemoglobins in the pathophysiology of Alzheimer's disease. *Exp. Gerontol.* *126*, 110680. <https://doi.org/10.1016/j.exger.2019.110680>.
- Anderson, N.L., and Anderson, N.G. (2002). The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* *1*, 845–867. <https://doi.org/10.1074/mcp.r200007-mcp200>.
- Appelhof, B., Wagner, M., Hoefele, J., Heinze, A., Roser, T., Koch-Hogrebe, M., Roosendaal, S.D., Dehghani, M., Mehrjardi, M.Y.V., Torti, E., et al. (2020). Pontocerebellar hypoplasia due to bi-allelic variants in MINPP1. *Eur. J. Hum. Genet.* <https://doi.org/10.1038/s41431-020-00749-x>.
- Bader, G.D., and Hogue, C.W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* *4*, 2. <https://doi.org/10.1186/1471-2105-4-2>.
- Bagwe-Parab, S., and Kaur, G. (2020). Molecular targets and therapeutic interventions for iron induced neurodegeneration. *Brain Res. Bull.* *156*, 1–9. <https://doi.org/10.1016/j.brainresbull.2019.12.011>.
- Belcher, J.D., Chen, C., Nguyen, J., Milbauer, L., Abdulla, F., Alayash, A.I., Smith, A., Nath, K.A., Hebbel, R.P., and Vercellotti, G.M. (2014). Heme triggers TLR4 signaling leading to endothelial cell activation and vaso-occlusion in murine sickle cell disease. *Blood* *123*, 377–390. <https://doi.org/10.1182/blood-2013-04-495887>.
- Biagioli, M., Pinto, M., Cesselli, D., Zaninello, M., Lazarevic, D., Roncaglia, P., Simone, R., Vlachouli, C., Plessy, C., Bertin, N., et al. (2009). Unexpected expression of alpha- and beta-globin in mesencephalic dopaminergic neurons and glial cells. *Proc. Natl. Acad. Sci. U S A* *106*, 15454–15459. <https://doi.org/10.1073/pnas.0813216106>.
- Cappellini, M.D., Poggiali, E., Taher, A.T., and Musallam, K.M. (2012). Hypercoagulability in beta-thalassemia: a status quo. *Expert Rev. Hematol.* *5*, 505–511. quiz 512. <https://doi.org/10.1586/ehm.12.42>.
- Chanpeng, P., Svasti, S., Paiboonsukwong, K., Smith, D.R., and Leecharoenkit, K. (2019). Platelet proteome reveals specific proteins associated with platelet activation and the hypercoagulable state in beta-thalassemia/HbE patients. *Sci. Rep.* *9*, 6059. <https://doi.org/10.1038/s41598-019-42432-2>.
- Clemmons, D.R. (2018). Role of IGF-binding proteins in regulating IGF responses to changes in metabolism. *J. Mol. Endocrinol.* *61*, T139–T169. <https://doi.org/10.1530/JME-18-0016>.
- CRISPR gene therapy trial on hold. *Nat. Biotechnol.* *36*, 562. <https://doi.org/10.1038/nbt0718-562>.
- Danjou, F., Anni, F., Perseu, L., Satta, S., Dessi, C., Lai, M.E., Fortina, P., Devoto, M., and Galanello, R. (2012). Genetic modifiers of beta-thalassemia and clinical severity as assessed by age at first transfusion. *Haematologica* *97*, 989–993. <https://doi.org/10.3324/haematol.2011.053504>.
- Dimitriou, E., Verhoeck, M., Altun, S., Karabatsos, F., Moraitou, M., Youssef, J., Boot, R., Sarafidou, J., Karagiorga, M., Aerts, H., and Michelakakis, H. (2005). Elevated plasma chemokine CCL18/PARC in beta-thalassemia. *Blood Cells Mol. Dis.* *35*, 328–331. <https://doi.org/10.1016/j.bcmd.2005.07.006>.
- Fayed, M.A., Abdel-Hady, H.E., Hafez, M.M., Salama, O.S., and Al-Tonbary, Y.A. (2018). Study of platelet activation, hypercoagulable state, and the association with pulmonary hypertension in children with beta-thalassemia. *Hematol. Oncol. Stem Cell Ther.* *11*, 65–74. <https://doi.org/10.1016/j.hemonc.2017.05.028>.
- Frangoul, H., Altshuler, D., Cappellini, M.D., Chen, Y.S., Domm, J., Eustace, B.K., Foell, J., de la Fuente, J., Grupp, S., Handgretinger, R., et al. (2021). CRISPR-Cas9 gene editing for sickle cell disease and beta-thalassemia. *N. Engl. J. Med.* *384*, 252–260. <https://doi.org/10.1056/NEJMoa2031054>.
- Greenbaum, D., Colangelo, C., Williams, K., and Gerstein, M. (2003). Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* *4*, 117. <https://doi.org/10.1186/gb-2003-4-9-117>.
- Gumus, P., Kahraman-Ceneli, S., Akcali, A., Sorsa, T., Tervahartiala, T., Buduneli, N., and Ozcaka, O. (2016). Association of thalassemia major and gingival inflammation: a pilot study. *Arch. Oral Biol.* *64*, 80–84. <https://doi.org/10.1016/j.archoralbio.2015.12.008>.
- Hatairaktham, S., Srisawat, C., Siritanaratkul, N., Chiangjong, W., Fucharoen, S., Thongboonkerd, V., and Kalpravidh, R.W. (2013). Differential plasma proteome profiles of mild versus severe beta-thalassemia/Hb E. *Ann. Hematol.* *92*, 365–377. <https://doi.org/10.1007/s00277-012-1629-5>.
- Hong, C.S., Park, M.R., Sun, E.G., Choi, W., Hwang, J.E., Bae, W.K., Rhee, J.H., Cho, S.H., and Chung, I.J. (2019). Gal-3BP negatively regulates NF-kappaB signaling by inhibiting the activation of TAK1. *Front. Immunol.* *10*, 1760. <https://doi.org/10.3389/fimmu.2019.01760>.
- Jensen, M.K., Bertoia, M.L., Cahill, L.E., Agarwal, I., Rimm, E.B., and Mukamal, K.J. (2014). Novel metabolic biomarkers of cardiovascular disease. *Nat. Rev. Endocrinol.* *10*, 659–672. <https://doi.org/10.1038/nrendo.2014.155>.
- Jeong, J.Y., Silver, M., Parnes, A., Nikiforow, S., Berliner, N., and Vanasse, G.J. (2011). Resveratrol ameliorates TNFalpha-mediated suppression of erythropoiesis in human CD34(+) cells via modulation of NF-kappaB signalling. *Br. J. Haematol.* *155*, 93–101. <https://doi.org/10.1111/j.1365-2141.2011.08800.x>.
- Karimi, M., Cohan, N., De Sanctis, V., Mallat, N.S., and Taher, A. (2014). Guidelines for diagnosis and management of beta-thalassemia intermedia. *Pediatr. Hematol. Oncol.* *31*, 583–596. <https://doi.org/10.3109/08880018.2014.937884>.
- Kittivorapart, J., Crew, V.K., Wilson, M.C., Heesom, K.J., Siritanaratkul, N., and Teye, A.M. (2018). Quantitative proteomics of plasma vesicles identify novel biomarkers for hemoglobin E/beta-thalassemic patients. *Blood Adv.* *2*, 95–104. <https://doi.org/10.1182/bloodadvances.2017011726>.
- Kountouris, P., Lederer, C.W., Fanis, P., Feleki, X., Old, J., and Kleanthous, M. (2014). IthaGenes: an interactive database for haemoglobin variations and epidemiology. *PLoS One* *9*, e103020. <https://doi.org/10.1371/journal.pone.0103020>.
- Laosombat, V., Wongchanchailert, M., Sattayasevana, B., Wiriyasateinkul, A., and Fucharoen, S. (2001). Clinical and hematologic features of beta0-thalassemia (frameshift 41/42 mutation) in Thai patients. *Haematologica* *86*, 138–141.
- Lasco, A., Morabito, N., Gaudio, A., Crisafulli, A., Meo, A., Denuzzo, G., and Frisina, N. (2002). Osteoporosis and beta-thalassemia major: role of the IGF-I/IGFBP-III axis. *J. Endocrinol. Invest.* *25*, 338–344. <https://doi.org/10.1007/BF03344015>.
- Li, L., Fang, C.J., Ryan, J.C., Niemi, E.C., Lebron, J.A., Bjorkman, P.J., Arase, H., Torti, F.M., Torti,

- S.V., Nakamura, M.C., and Seaman, W.E. (2010). Binding and uptake of H-ferritin are mediated by human transferrin receptor-1. *Proc. Natl. Acad. Sci. U S A* 107, 3505–3510. <https://doi.org/10.1073/pnas.0913192107>.
- Li, B., Zhang, X.Z., Yin, A.H., Zhao, Q.G., Wu, L., Ma, Y.Z., Luo, M.Y., and Yu, S.Y. (2014). High prevalence of thalassemia in migrant populations in Guangdong Province, China. *BMC Public Health* 14, 905. <https://doi.org/10.1186/1471-2458-14-905>.
- Li, J., Chen, X., Yi, J., Liu, Y., Li, D., Wang, J., Hou, D., Jiang, X., Zhang, J., Wang, J., et al. (2016). Identification and characterization of 293T cell-derived exosomes by profiling the protein, mRNA and MicroRNA components. *PLoS One* 11, e0163043. <https://doi.org/10.1371/journal.pone.0163043>.
- Lindstrom, S., Wang, L., Smith, E.N., Gordon, W., van Hylckama Vlieg, A., de Andrade, M., Brody, J.A., Pattee, J.W., Haessler, J., Brumpton, B.M., et al. (2019). Genomic and transcriptomic association studies identify 16 novel susceptibility loci for venous thromboembolism. *Blood* 134, 1645–1657. <https://doi.org/10.1182/blood.2019000435>.
- Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., Li, K., Zhang, G., Jin, Z., He, F., et al. (2019). iProX: an integrated proteome resource. *Nucleic Acids Res.* 47, D1211–D1217. <https://doi.org/10.1093/nar/gky869>.
- Madani, H., Rahimi, Z., Manavi-Shad, M., Mozafari, H., Akramipour, R., Vaisi-Raygani, A., Rezaei, M., Malek-Khosravi, S., Shakiba, E., and Parsian, A. (2011). Plasma lipids and lipoproteins in children and young adults with major beta-thalassemia from western Iran: influence of genotype. *Mol. Biol. Rep.* 38, 2573–2578. <https://doi.org/10.1007/s11033-010-0397-3>.
- Manfre, L., Giarratano, E., Maggio, A., Banco, A., Vaccaro, G., and Lagalla, R. (1999). MR imaging of the brain: findings in asymptomatic patients with thalassemia intermedia and sickle cell-thalassemia disease. *AJR Am. J. Roentgenol.* 173, 1477–1480. <https://doi.org/10.2214/ajr.173.6.10584785>.
- Marengo-Rowe, A.J. (2007). The thalassemias and related disorders. *Proc. (Bayl. Univ. Med. Cent.)* 20, 27–31. <https://doi.org/10.1080/08998280.2007.11928230>.
- Merico, D., Isserlin, R., Stueker, O., Emili, A., and Bader, G.D. (2010). Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5, e13984. <https://doi.org/10.1371/journal.pone.0013984>.
- Muller-Eberhard, U., Javid, J., Liem, H.H., Hanstein, A., and Hanna, M. (1968). Plasma concentrations of hemopexin, haptoglobin and heme in patients with various hemolytic diseases. *Blood* 32, 811–815.
- Nasr, M.R., Abdelmaksoud, A.M., Abd El-Aal, K.S., Mabrouk, N.A., and Ismael, W.M. (2008). Plasma lipid profile and lipid peroxidation in beta-thalassemic children. *J. Clin. Lipidol.* 2, 405–409. <https://doi.org/10.1016/j.jacl.2008.10.008>.
- Palmer, W.C., Vishnu, P., Sanchez, W., Aqel, B., Riegert-Johnson, D., Seaman, L.A.K., Bowman, A.W., and Rivera, C.E. (2018). Diagnosis and management of genetic iron overload disorders. *J. Gen. Intern. Med.* 33, 2230–2236. <https://doi.org/10.1007/s11606-018-4669-2>.
- Pan, J.B., Hu, S.C., Shi, D., Cai, M.C., Li, Y.B., Zou, Q., and Ji, Z.L. (2013). PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. *PLoS One* 8, e80747. <https://doi.org/10.1371/journal.pone.0080747>.
- Pernemalm, M., Sandberg, A., Zhu, Y., Boekel, J., Tamburro, D., Schwenk, J.M., Bjork, A., Wahren-Herlenius, M., Amark, H., Ostenson, C.G., et al. (2019). In-depth human plasma proteome analysis captures tissue proteins and transfer of protein variants across the placenta. *Elife* 8. <https://doi.org/10.7554/eLife.41608>.
- Peters, M., Heijboer, H., Smiers, F., and Giordano, P.C. (2012). Diagnosis and management of thalassaemia. *BMJ* 344, e228. <https://doi.org/10.1136/bmj.e228>.
- Ponnikorn, S., Mongkolrob, R., Klongthay, S., Roytrakul, S., Srisanga, K., Tungpradabkul, S., and Hongeng, S. (2019). Comparative proteome-wide analysis of bone marrow microenvironment of beta-thalassemia/hemoglobin E. *Proteomes* 7. <https://doi.org/10.3390/proteomes7010008>.
- Rachmilewitz, E.A., Shohet, S.B., and Lubin, B.H. (1976). Lipid membrane peroxidation in beta-thalassemia major. *Blood* 47, 495–505.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Muller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77. <https://doi.org/10.1186/1471-2105-12-77>.
- Roth, C., Pekrun, A., Bartz, M., Jarry, H., Eber, S., Lakomek, M., and Schroter, W. (1997). Short stature and failure of pubertal development in thalassaemia major: evidence for hypothalamic neurosecretory dysfunction of growth hormone secretion and defective pituitary gonadotropin secretion. *Eur. J. Pediatr.* 156, 777–783. <https://doi.org/10.1007/s004310050711>.
- Roudkenar, M.H., Halabian, R., Oodi, A., Roushandeh, A.M., Yaghamai, P., Najar, M.R., Amirzadeh, N., and Shokrgozar, M.A. (2008). Upregulation of neutrophil gelatinase-associated lipocalin, NGAL/Lcn2, in beta-thalassemia patients. *Arch. Med. Res.* 39, 402–407. <https://doi.org/10.1016/j.arcmed.2007.12.002>.
- Rund, D., and Rachmilewitz, E. (2005). Beta-thalassemia. *N. Engl. J. Med.* 353, 1135–1146. <https://doi.org/10.1056/NEJMra050436>.
- Shah, F.T., Sayani, F., Trompeter, S., Drasar, E., and Piga, A. (2019). Challenges of blood transfusions in beta-thalassemia. *Blood Rev.* 37, 100588. <https://doi.org/10.1016/j.blre.2019.100588>.
- Shang, X., and Xu, X. (2017). Update in the genetics of thalassemia: what clinicians need to know. *Best Pract. Res. Clin. Obstet. Gynaecol.* 39, 3–15. <https://doi.org/10.1016/j.bpobgyn.2016.10.012>.
- Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72.e15. <https://doi.org/10.1016/j.cell.2020.05.032>.
- Smith, A., and McCulloh, R.J. (2015). Hemopexin and haptoglobin: allies against heme toxicity from hemoglobin not contenders. *Front. Physiol.* 6, 187. <https://doi.org/10.3389/fphys.2015.00187>.
- Soliman, A.T., El Banna, N., Abdel Fattah, M., ElZalabani, M.M., and Ansari, B.M. (1998). Bone mineral density in prepubertal children with beta-thalassemia: correlation with growth and hormonal data. *Metabolism* 47, 541–548. [https://doi.org/10.1016/s0026-0495\(98\)90237-2](https://doi.org/10.1016/s0026-0495(98)90237-2).
- Sripichai, O., and Fucharoen, S. (2016). Fetal hemoglobin regulation in beta-thalassemia: heterogeneity, modifiers and therapeutic approaches. *Expert Rev. Hematol.* 9, 1129–1137. <https://doi.org/10.1080/17474086.2016.1255142>.
- Sripichai, O., Munkongdee, T., Kunkhaek, C., Svasti, S., Winichagoon, P., and Fucharoen, S. (2008). Coinheritance of the different copy numbers of alpha-globin gene modifies severity of beta-thalassemia/Hb E disease. *Ann. Hematol.* 87, 375–379. <https://doi.org/10.1007/s00277-007-0407-2>.
- Subramanian, A., Kuehn, H., Gould, J., Tamayo, P., and Mesirov, J.P. (2007). GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* 23, 3251–3253. <https://doi.org/10.1093/bioinformatics/btm369>.
- Sun, Z., Liu, X., Wu, D., Gao, H., Jiang, J., Yang, Y., Wu, J., Gao, Q., Wang, J., Jiang, Z., et al. (2019). Circulating proteomic panels for diagnosis and risk stratification of acute-on-chronic liver failure in patients with viral hepatitis B. *Theranostics* 9, 1200–1214. <https://doi.org/10.7150/thno.31991>.
- Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 47, D607–D613. <https://doi.org/10.1093/nar/gky1131>.
- Taher, A.T., and Saliba, A.N. (2017). Iron overload in thalassemia: different organs at different rates. *Hematol. Am. Soc. Hematol. Educ. Program* 2017, 265–271. <https://doi.org/10.1182/asheducation-2017.1.265>.
- Taher, A.T., Weatherall, D.J., and Cappellini, M.D. (2018). Thalassaemia. *Lancet* 391, 155–167. [https://doi.org/10.1016/S0140-6736\(17\)31822-6](https://doi.org/10.1016/S0140-6736(17)31822-6).
- Taher, A.T., Musallam, K.M., and Cappellini, M.D. (2021). beta-thalassemias. *N. Engl. J. Med.* 384, 727–743. <https://doi.org/10.1056/NEJMra2021838>.
- Terman, A., and Kurz, T. (2013). Lysosomal iron, iron chelation, and cell death. *Antioxid. Redox Signal.* 18, 888–898. <https://doi.org/10.1089/ars.2012.4885>.
- Tewari, S., Renney, G., Brewin, J., Gardner, K., Kirkham, F., Inusa, B., Barrett, J.E., Menzel, S., Thein, S.L., Ward, M., and Rees, D.C. (2018). Proteomic analysis of plasma from children with



sickle cell anemia and silent cerebral infarction. *Haematologica* 103, 1136–1142. <https://doi.org/10.3324/haematol.2018.187815>.

Theodorou, A., Phylactides, M., Katsantoni, E., Vougas, K., Garbis, S.D., Fanis, P., Sitarou, M., Thein, S.L., and Kleanthous, M. (2020). Proteomic studies for the investigation of gamma-globin induction by decitabine in human primary erythroid progenitor cultures. *J. Clin. Med.* 9. <https://doi.org/10.3390/jcm9010134>.

Tselepis, A.D., Hahalis, G., Tellis, C.C., Papavasiliou, E.C., Mylona, P.T., Kourakli, A., and Alexopoulos, D.C. (2010). Plasma levels of lipoprotein-associated phospholipase A(2) are increased in patients with beta-thalassemia. *J. Lipid Res.* 51, 3331–3341. <https://doi.org/10.1194/jlr.M007229>.

Ucuncu, E., Rajamani, K., Wilson, M.S.C., Medina-Cano, D., Altin, N., David, P., Barcia, G., Lefort, N., Banal, C., Vasilache-Dangles, M.T., et al. (2020). MINPP1 prevents intracellular accumulation of the chelator inositol hexakisphosphate and is mutated in Pontocerebellar Hypoplasia. *Nat. Commun.* 11, 6087. <https://doi.org/10.1038/s41467-020-19919-y>.

Vinchi, F., De Franceschi, L., Ghigo, A., Townes, T., Cimino, J., Silengo, L., Hirsch, E., Altruda, F., and Tolosano, E. (2013). Hemopexin therapy improves cardiovascular function by preventing heme-induced endothelial toxicity in mouse

models of hemolytic diseases. *Circulation* 127, 1317–1329. <https://doi.org/10.1161/CIRCULATIONAHA.112.130179>.

Vinchi, F., Costa da Silva, M., Ingoglia, G., Petrillo, S., Brinkman, N., Zuercher, A., Cerwenka, A., Tolosano, E., and Muckenthaler, M.U. (2016). Hemopexin therapy reverts heme-induced proinflammatory phenotypic switching of macrophages in a mouse model of sickle cell disease. *Blood* 127, 473–486. <https://doi.org/10.1182/blood-2015-08-663245>.

Wewer Albrechtsen, N.J., Geyer, P.E., Doll, S., Treit, P.V., Bojsen-Moller, K.N., Martinussen, C., Jorgensen, N.B., Torekov, S.S., Meier, F., Niu, L., et al. (2018). Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after roux-En-Y gastric bypass surgery. *Cell Syst.* 7, 601–612.e3. <https://doi.org/10.1016/j.cels.2018.10.012>.

Winichagoon, P., Fucharoen, S., and Wasi, P. (1981). Increased circulating platelet aggregates in thalassaemia. *Southeast Asian J. Trop. Med. Public Health* 12, 556–560.

Wu, K.H., Tsai, F.J., and Peng, C.T. (2003). Growth hormone (GH) deficiency in patients with beta-thalassemia major and the efficacy of recombinant GH treatment. *Ann. Hematol.* 82, 637–640. <https://doi.org/10.1007/s00277-003-0712-3>.

Wu, Z., Geng, Y., Lu, X., Shi, Y., Wu, G., Zhang, M., Shan, B., Pan, H., and Yuan, J. (2019). Chaperone-mediated autophagy is involved in the execution of ferroptosis. *Proc. Natl. Acad. Sci. U S A* 116, 2996–3005. <https://doi.org/10.1073/pnas.1819728116>.

Xiong, F., Sun, M., Zhang, X., Cai, R., Zhou, Y., Lou, J., Zeng, L., Sun, Q., Xiao, Q., Shang, X., et al. (2010). Molecular epidemiological survey of haemoglobinopathies in the Guangxi Zhuang Autonomous Region of Southern China. *Clin. Genet.* 78, 139–148. <https://doi.org/10.1111/j.1399-0004.2010.01430.x>.

Yakar, S., Rosen, C.J., Beamer, W.G., Ackert-Bicknell, C.L., Wu, Y., Liu, J.L., Ooi, G.T., Setser, J., Frystyk, J., Boisclair, Y.R., and LeRoith, D. (2002). Circulating levels of IGF-1 directly regulate bone growth and density. *J. Clin. Invest.* 110, 771–781. <https://doi.org/10.1172/JCI15463>.

Zhang, M.Y., Sun, S.C., Bell, L., and Miller, B.A. (1998). NF-kappaB transcription factors are involved in normal erythropoiesis. *Blood* 91, 4136–4144.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* 10, 1523. <https://doi.org/10.1038/s41467-019-09234-6>.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Whole blood, serum and plasma from adults and children	The First Affiliated Hospital of Guangxi Medical University	This paper (Table S1)
<b>Chemicals, peptides, and recombinant proteins</b>		
Iodoacetamide (IAM)	Sigma-Aldrich	Cat #I1149
DL-Dithiothreitol (DTT)	Sigma-Aldrich	Cat #D9163
Urea	Sigma-Aldrich	Cat #51456
Trypsin	Promega	Cat #V5113
LYSYL ENDOPEPTIDASE	WAKO	Cat #125-05061
Hydroxylamine solution	Sigma-Aldrich	Cat #467804
Methanol	J.T.Baker	Cat #9093-68
Acetonitrile	J.T.Baker	Cat #9017-03
<b>Critical commercial assays</b>		
High Select™ Top-14 Abundant Protein Depletion Mini Spin Columns	Thermo Fisher Scientific	Cat #A36370
BCA Protein Assay Kit	Thermo Fisher Scientific	Cat #23227
TMT 10-plex reagents	Thermo Fisher Scientific	Cat #90111
<b>Deposited data</b>		
Mass spectrometry data	<a href="http://www.proteomexchange.org">http://www.proteomexchange.org</a>	PXD032179
<b>Software and algorithms</b>		
Proteome Discoverer Version 2.2.0.388	Thermo Fisher Scientific	N/A
R V3.6.2	R Project	<a href="https://www.r-project.org">https://www.r-project.org</a>
RStudio Version 1.2.1335	Open source	<a href="https://www.rstudio.com/">https://www.rstudio.com/</a>
GSEA software V4.1.0	N/A	<a href="http://www.gsea-msigdb.org/gsea/">http://www.gsea-msigdb.org/gsea/</a>
String	N/A	<a href="https://cn.string-db.org">https://cn.string-db.org</a>
Metascape	N/A	<a href="https://metascape.org/">https://metascape.org/</a>
Enrichment Map V3.1.0	N/A	<a href="https://www.baderlab.org/Software/EnrichmentMap#Downloads">https://www.baderlab.org/Software/EnrichmentMap#Downloads</a>
Cytoscape 3.7.2	Open source	<a href="https://cytoscape.org">https://cytoscape.org</a>
UniProt	UniProt Consortium	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Fuquan Yang ([fqyang@ibp.ac.cn](mailto:fqyang@ibp.ac.cn)).

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

The mass spectrometry data have been deposited to the ProteomeXchange Consortium via the iProX partner repository (Ma et al., 2019). Accession number is listed in the [key resources table](#). This paper does not

report original code. Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Patients and healthy participants

Patients and healthy volunteers were recruited at The First Affiliated Hospital of Guangxi Medical University (Guangxi Zhuang Autonomous Region, China). Written informed consent was provided by all participants. For patients under the age of 18 years, consent was provided by the parents or legal guardian. Two weeks after receiving their last blood transfusion, patients offered their blood sample, and completed a questionnaire including the first time for transfusion, treatment conditions, and other medical conditions. A total of 286  $\beta$ -thalassemia patients (mean  $\pm$  SD,  $23.3 \pm 14.6$  years) and 51 healthy controls ( $8.7 \pm 2.1$  years) were recruited for this study and 20 TI ( $\beta^+\beta^+/\beta^+\beta^0$ ,  $9.2 \pm 2.7$  years), 20 TM ( $\beta^0\beta^0$ ,  $8.9 \pm 2.3$  years), and 20 age- and gender-matched healthy controls ( $9.9 \pm 2.3$  years) were enrolled for the TMT-based plasma proteomic study. The detailed clinical information of these participants was list in [Table S1](#). The patients included in the plasma proteomic study were all received blood transfusions and iron chelation therapy, and none underwent splenectomy. Whole blood was collected into an EDTA Vacutainer and centrifuged at  $3,000 \times g$  for 10 min. Harvested plasma from the supernatant were stored at  $-80^\circ\text{C}$  until use. This study was designed and conducted in accordance with the Declaration of Helsinki. The Ethics Committee of The First Affiliated Hospital of Guangxi Medical University approved this study.

## METHOD DETAILS

### $\beta$ -thalassemia genotyping

DNA was extracted from the blood samples and genotyping was performed at BGI Clinical Laboratories (Shenzhen, China) using Gap-PCR and SNP detection.  $\beta^+$  mutations include  $-28\text{A} > \text{G}$  (*HBB*: c.-78A > G),  $-29\text{A} > \text{G}$  (*HBB*: c.-79A > G), codon 26 G > A (Hb E, *HBB*: c.79G > A), IVS-II-654C654 > T (*HBB*: c.316-197C > T), and IVS-II-5 G > C (*HBB*: c.315+5G > C).  $\beta^0$  mutations include codons 41/42 -TTCT (*HBB*: c.126\_129delCTTT), codon 17A > T (*HBB*: c.52A > T), codons 71/72 + A (*HBB*: c.216\_217insA), IVS-I-1 G > T (*HBB*: c.92+1G > T), codon 43 G > T (*HBB*: c.130G > T), IVS-I-130 G > C (*HBB*: c.93-1G > C), codon 37 G > A (*HBB*: c.114G > A), codons 27/28 + C (*HBB*: c.84\_85insC), and codon 30A > G (*HBB*: c.91A > G).

### Hematological and serum parameters

Hematological and serum iron parameters were measured from fasting blood samples collected two weeks after the most recent blood transfusion. Hematological values were measured using a CELL-DYN automated hematology analyzer (Abbott Diagnostics). Fetal hemoglobin (Hb F) was measured using high-performance liquid chromatography (VARIANT II, Bio-Rad). Serum ferritin (SF) was measured using an electrochemiluminescence-based immunoassay (Cobas e601, Roche). Serum soluble transferrin receptor (sTfR) was measured by immunoturbidimetry utilizing an automatic biochemistry analyzer (Abbott C8000).

### High-abundance protein removal and protein digestion

To remove the top 14 high-abundance proteins, 10  $\mu\text{L}$  of plasma was loaded on High-Select™ Top14 Abundant Protein Depletion Mini Spin Columns (Thermo Fisher Scientific, USA) according to the manufacturer's protocols. The plasma and resin slurry was incubated in the column with gentle end-over-end mixing for 10 min at room temperature. After incubation, the mini column was placed into a 2 mL collection tube and underwent centrifugal concentration at  $1,000 \times g$  for 2 min to collect the low-abundance components. The collected sample was concentrated and buffer exchanged with 8 M Urea/100 mM  $\text{NH}_4\text{HCO}_3$  using a Millipore Amicon® Ultra-0.5 3 kDa MWCO filter according to the manufacturer's protocol. The protein concentration was quantified using a BCA Protein Assay Kit (Thermo Fisher Scientific, USA).

For digestion, the concentrated proteins were reduced with 20 mM DTT (Sigma-Aldrich) at  $37^\circ\text{C}$  for 1 h, followed by alkylation with 40 mM iodoacetamide (IAM, Sigma-Aldrich) for 30 min at room temperature in the dark. The urea in the sample solution was then diluted to less than 2 M with 50 mM  $\text{NH}_4\text{HCO}_3$ . The sample was digested with Lys C at an enzyme:protein ratio of 1:50 (w/w) for two hours at  $37^\circ\text{C}$  and further digested at  $37^\circ\text{C}$  overnight with trypsin at a trypsin:protein ratio of 1:50 (w/w). Digestion was quenched by addition of formic acid to a final concentration of 1% (v/v). All peptide samples were desalted

with Oasis HLB solid phase extraction columns (Bedford, MA, USA) and then dried under vacuum for TMT labeling.

### TMT 10-plex labeling

Desalted peptides were resuspended in 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES, pH 8.5). TMT 10-plex reagents (0.4 mg) in acetonitrile were then added to 30  $\mu$ g of peptides to a final acetonitrile concentration of ~30% (v/v) and incubated for one hour at room temperature. The reaction was finally quenched with hydroxylamine to a final concentration of 0.1% (v/v). To test the labeling efficiency and optimize the mixing ratio, 2  $\mu$ g of TMT-labeled peptides from each sample were pooled together. After the labeling check, the remaining labeled peptides were mixed together at a ratio of 1:1:1:1:1:1:1:1:1:1 and vacuum centrifuged to dryness. The pooled peptide sample was then resolved with 0.1% FA/H<sub>2</sub>O, subjected to HLB solid-phase desalination, and dried.

### High pH reversed-phase HPLC fractionation

Prior to subsequent liquid chromatography-tandem mass spectrometry (LC-MS/MS) processing, peptides were dissolved in buffer A (2% ACN/98% H<sub>2</sub>O, pH=10) and fractionated using an XBridge peptide BEH C18 column (130 Å, 3.5  $\mu$ m, 2.1 mm  $\times$  150 mm, Waters) on a Rigol L-3000 HPLC system with a binary buffer system. The gradient of buffer B (98% ACN/2% H<sub>2</sub>O, pH = 10) was set as follows: 5–8% for 5 min, 8–18% for 35 min, 18–32% for 22 min, 32–95% for 2 min, and 95% for 4 min at a constant flow rate of 0.23 mL/min. All fractions were collected at 90 s intervals and concatenated into 12 post-fractions as described (Li et al., 2016). The combined fractions were lyophilized and stored at –80°C for further LC-MS/MS analysis.

### LC-MS/MS analysis

All mass spectrometry data were collected on a Q Exactive mass spectrometer (Thermo Fisher Scientific) coupled with an EASY-nLC 1000 HPLC system (Thermo Fisher Scientific). Each sample was resuspended in 0.1% FA and loaded onto an in-house packed C18 trap column (100  $\mu$ m  $\times$  2 cm, Reprosil-Pur C18 AQ, 5  $\mu$ m, Dr. Maisch GmbH) and separated on a PFSelf-Pack C18 analytical column (75  $\mu$ m  $\times$  20 cm, Reprosil-Pur C18 AQ, 3  $\mu$ m, Dr. Maisch GmbH) with mobile phase A (0.1% FA) and mobile phase B (ACN/0.1% FA) with the following gradient: 4–9% phase B, 5 min; 9–22% B, 80 min; 22–32% B, 25 min; 32–90% B, 1 min; 90% B, 9 min at a constant flow rate of 310 nL/min. Full scan spectra were acquired with a resolution of 70,000 (*m/z* 200), and the dynamic exclusion time was 50 s. The AGC value was 3e6 and maximum fill time was 60 ms in full scan mode. MS/MS spectra were acquired at AGC 5e4 with maximum injection time of 80 ms. The collision energy of HCD was 32%.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### MS data analysis

Raw MS/MS data were searched against the UniProt human protein database (released 24 July 13, 2018), combined with commonly observed contaminants (245 sequences), in Proteome Discoverer 2.2 (PD2.2, Thermo Scientific) using the Sequest HT search engine for protein identification. Reporter ion intensities were corrected according to the distribution of TMT reporter ion isotope specified by the manufacturer. Fully digested with trypsin was selected and two missed cleavages were allowed during database search. Mass tolerance for precursor and fragment ions were 10 ppm and 0.02 Da, respectively. Cysteine carbamidomethylation (+57.021) and TMT labeling of lysine and peptide N-terminus (+229.163) were specified as fixed modifications, while methionine oxidation (+15.995) and N-terminal protein acetylation (+42.011) were selected as variable modifications. False discovery rate (FDR) was determined by searching a reverse sequence decoy database and set to 1% for both protein and peptide level with a minimum length of 7 amino acids. Proteins matching to contaminants were filtered out.

To control for inconsistencies in sample loading, the protein abundance was normalized by setting the Normalization Mode parameter in PD 2.2 as total peptide amount. The average of normalized abundance for each protein was then scaled to 100 to enable the comparison across experiments. Thus, the scaled abundance of proteins was used in the following bioinformatics analysis.

### Bioinformatics analysis

The data were processed using R (version 3.6.2). Log<sub>2</sub>-transformed abundance of proteins was used for statistical tests.

A two-way Analysis of Variance (ANOVA) was performed to assess the effects of TMT sets and experimental groups on each protein abundance. The resulting p value was adjusted using the Benjamini-Hochberg method, and an adjusted p value < 0.05 was marked as a significant result. The magnitude of studied effects was estimate by partial eta squared (partial  $\eta^2$ ).

Differential protein analyses were performed using a One-way ANOVA for comparison of three groups (TM, TI, and control) and a two-sample t-test for comparison of two groups ( $\beta$ -thalassemia patients and healthy controls). Tests for two-sample comparison were two-sided. For multiple comparisons among the three groups, Tukey's Honest Significant Difference (TukeyHSD) statistical test was used for the post-hoc analysis. For general significance levels, proteins with p value < 0.05 and fold change >1.2 or <0.85 were considered as up- or downregulated proteins. The stricter criterion is p value < 0.05 and fold change >2 or <0.7.

The global correlation analysis (Wewer Albrechtsen et al., 2018) was finished by the Corrplot package. Functional enrichment analysis of proteins in high correlation squares was performed in STRING (version 11.0) (Szkarczyk et al., 2019). For Gene Set Enrichment Analysis (GSEA), all quantified proteins were ranked by  $\log_2$  (fold change) and analyzed using the GSEA (V4.1.0) software (Subramanian et al., 2007).

Differential proteins identified during the comparison of  $\beta$ -thalassemia patients and healthy controls were submitted to Metascape (Zhou et al., 2019) (version 3.5) for tissue-specific enrichment analysis using PaGenBase database and for protein-protein interaction enrichment analysis using the Molecular Complex Detection (MCODE) clustering algorithm. Network interaction and functional enrichment analyses were performed by STRING or Metascape.

The Pearson's correlation coefficient was used to measure the strength of a linear association between two variables. The receiver operating characteristic (ROC) curve analysis was performed in the pROC package (Robin et al., 2011). The relationship between the first time of patients' transfusion and proteins associated with Hb F was estimated by Kaplan-Meier curves, which was finished by survival and survminer packages.

For enrichment map analysis, ranked proteins in TM vs. control and TI vs. control were separately analyzed using the GSEA (V4.1.0) software (Subramanian et al., 2007). The enriched gene-sets with p value < 0.05 and FDR < 0.25 were visualized using plug-in EnrichmentMap 3.3 of Cytoscape (Merico et al., 2010), where nodes represented gene-sets and colors represented Normalized Enrichment Scores (NES) (red: positive; blue: negative). Gene-sets were grouped and scaled according to shared components (edge >0.5) and their similarity (size).

To determine the effect of Hb F at the protein level, the following linear regression model was used:

$$\text{Protein level} \sim \alpha + \beta_1 \times \text{Hb F level} + \beta_2 \times \text{group} + \varepsilon$$