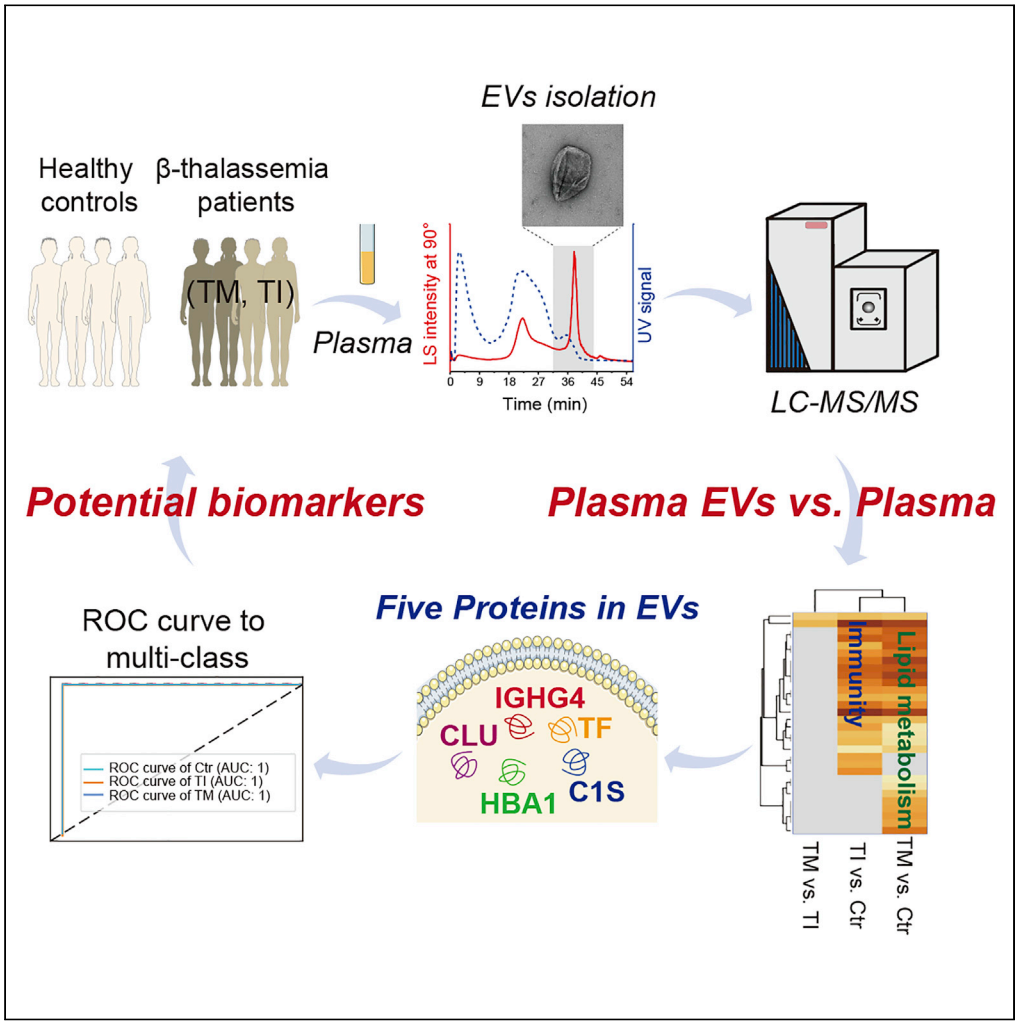
iScience

CellPress
OPEN ACCESS

Article

Differential proteomic patterns of plasma extracellular vesicles show potential to discriminate β-thalassemia subtypes



Na Li, Bowen Wu, Jifeng Wang, ..., Jianming Luo, Fudi Wang, Fuquan Yang

fqyang@ibp.ac.cn

Highlights

Subtypes exhibit significant abnormalities in immunity and lipid metabolism

Plasma EVs showed a superior ability to discriminate subtypes than plasma

A five-protein panel from plasma EVs was proposed for subtype diagnosis

Li et al., iScience 26, 106048
February 17, 2023
https://doi.org/10.1016/j.isci.2023.106048

iScience

CellPress
OPEN ACCESS

Article

Differential proteomic patterns of plasma extracellular vesicles show potential to discriminate β-thalassemia subtypes

Na Li,[1,2,6] Bowen Wu,[1,2,6] Jifeng Wang,[1] Yumeng Yan,[1,2] Peng An,[3] Yuezhen Li,[3] Yuning Liu,[3] Yanfei Hou,[3] Xiaoqing Qing,[1,2] Lili Niu,[1] Xiang Ding,[1,2] Zhensheng Xie,[1,2] Mengmeng Zhang,[1] Xiaojing Guo,[1,2] Xiulan Chen,[1,2] Tanxi Cai,[1,2] Jianming Luo,[4] Fudi Wang,[5] and Fuquan Yang[1,2,7,*]

SUMMARY

The observed specificity of β-thalassemia-subtype phenotypes makes new diagnostic strategies that complement current screening methods necessary to determine each subtype and facilitate therapeutic regimens for different patients. Here, we performed quantitative proteomics of plasma-derived extracellular vesicles (EVs) of β-thalassemia major (TM) patients, β-thalassemia intermedia (TI) patients, and healthy controls to explore subgroup characteristics and potential biomarkers. Plasma quantitative proteomics among the same cohorts were analyzed in parallel to compare the biomarker potential of both specimens. EV proteomics showed significantly more abnormalities in immunity and lipid metabolism in TI and TM, respectively. The differential proteomic patterns of EVs were consistent with but more striking than those of plasma. Notably, we also found EV proteins to have a superior performance for discriminating β-thalassemia subtypes. These findings allowed us to propose a diagnostic model consisting of five proteins in EVs with subtyping potential, demonstrating the ability of plasma-derived EVs for the diagnosis of β-thalassemia patients.

INTRODUCTION

β-Thalassemia is an inherited blood disorder caused by mutations in the β-globin gene (*HBB*), which leads to defects in the synthesis of the β-globin chain. About 1.5% of the global population carries the β-thalassemia variant, with most affected individuals originating from the Indian subcontinent, Southeast Asia, and China.[1] Ineffective erythropoiesis and hemolytic anemia are the hallmarks of β-thalassemia, causing multiple-organ damage and developmental delays. The severity of this disease depends largely on the genotypes of more than 350 reported mutations.[2] In addition, β-thalassemia phenotypes can be modified through the presence of genetic polymorphisms in genes involved in erythropoiesis regulation,[3] further complicating clinical diagnosis.

Qualitative and quantitative analyses of the different types of hemoglobin in conjunction with red blood cell indices represent the first-tier screening strategies for β-thalassemia diagnosis. However, it is first necessary to exclude *a priori* several suspected diseases, such as iron deficiency, structural hemoglobin variants, and anemia of chronic disease, for a diagnosis to be based solely on these detections due to their insufficient sensitivity and specificity.[1] Polymerase chain reaction represents another widely used screening approach for β-thalassemia by identifying the underlying genotypes of individuals. However, only prespecified mutations are detectable, leading to the underdiagnosis of uncommon mutations that are yet to be associated with the disease.[4] Therefore, developing high-efficient screening methods complementary to current approaches will simplify the diagnostic process and decrease the underdiagnosis rate in the future.

β-Thalassemia patients are generally classified as suffering from β-thalassemia intermedia (TI) and major (TM). TM is characterized by severe anemia managed as transfusion-dependent thalassemia, whereas TI includes a more diverse group of clinical manifestations with varying severity, from mild and moderate to severe anemia, and usually requires occasional or intermittent transfusions.[5–7] Over the years, striking

[1]Key Laboratory of Protein and Peptide Pharmaceuticals & Laboratory of Proteomics, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Department of Nutrition and Health, China Agricultural University, Beijing 100193, China

[4]Department of Pediatrics, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021 China

[5]The Fourth Affiliated Hospital, School of Public Health, State Key Laboratory of Experimental Hematology, Zhejiang University School of Medicine, Hangzhou 310058, China

[6]These authors contributed equally

[7]Lead contact

*Correspondence: fqyang@ibp.ac.cn

https://doi.org/10.1016/j.isci.2023.106048

differences in pathophysiological and clinical characteristics were uncovered among different patients, including underlying mechanisms of iron overload, the development of complications with disease progression, and efficient therapeutic targets.[8,9] For example, epidemiological data indicated the incidence of thromboembolic events in TI patients was >4-fold higher than that in TM patients.[10–12] Similarly, both TM and TI patients present with heart failure, the former as a result of left ventricular dysfunction and the latter due to pulmonary hypertension developed from preserved systolic left ventricular function.[13] Therefore, it is essential to develop prompt and accurate diagnostic strategies to distinguish between TI and TM to ensure early intervention as the first step for managing different β-thalassemia forms.[14]

Extracellular vesicles (EVs) have specific advantages for an accurate disease diagnosis due to their key roles in cell-to-cell communication, immune and inflammatory regulation, and cell growth, proliferation, and migration.[15–17] Circulating microparticles from β-thalassemia/HbE patients are able to induce coagulation molecules, pro-inflammatory cytokines, and adhesion molecule expression in endothelial cells,[18] as well as promote cardiac cell proliferation.[19] These characteristics demonstrate how EVs are affected in β-thalassemia. Among the proteins identified in EVs, haptoglobin, hemopexin, and cathepsin S are among potential clinically relevant biomarkers for the levels of hemolysis and inflammation in β-thalassemia.[20,21] Moreover, EV-HSP70 was associated with markers of ineffective erythropoiesis, hemolysis, and β-thalassemia disease severity.[22] Therefore, the proteomic patterns of circulating EVs merit further investigation to improve the diagnosis of β-thalassemia.

Mass spectrometry (MS)-based quantitative proteomic approaches are versatile, systematic, and scalable, facilitating a cost-effective and rapid evaluation. In this study, we used a tandem mass tags (TMT)-based quantitative proteomics to identify differences in plasma-derived EV proteins among healthy controls and β-thalassemia patients with different levels of disease severity. These differential proteins were validated by parallel reaction monitoring (PRM), a targeted proteomic technology. To explore the most promising biomarkers, paired plasma proteomic data were also analyzed in this study. Our study provides further insights into the biological characteristics of different clinical subtypes of β-thalassemia and potential biomarkers to improve the effective screening of this disease.

RESULTS

Characterization of plasma-derived EVs from β-thalassemia patients and healthy controls

To compare with the plasma proteomics, we used the same clinical cohort as our previous study on plasma proteomics of β-thalassemia.[23] In the discovery stage of this study, plasma samples from seven random individuals in each group were pooled together to reduce individual differences. The pooled plasma was then separated using an asymmetrical field-flow fractionation (AF4) to obtain EVs. A total of six EV samples were used for downstream TMT-6plex-based quantitative proteomic analysis. The differential proteins identified in the first stage were sequentially submitted to targeted quantification in the verification stage. The overall workflow is shown in Figure 1A.

The representative elution peak of EVs in AF4, referred to as P4, was characterized in our previous methodological works.[24,25] Isolated EVs from plasma through AF4 were also characterized in this study by transmission electron microscopy, nanoparticle tracking analysis, and Western blot (Figures 1B and S1). The elution profiles of six pooled plasma were similar, whereas the elution profiles of individual samples represented the biological heterogeneity (Figures 1C and S2–S4). More importantly, there were no significant differences in particle number and density of P4 among different groups or sexes, as determined from multiangle light scattering signals (Figures 1D and 1E).

Proteomic profile of plasma-derived EVs from β-thalassemia patients measured by TMT-based quantitative proteomics

TMT-based quantitative proteomic analysis of the six EV samples identified 375 proteins, of which 362 were quantified with high confidence. The quantified proteins with a p value ≤ 0.05 and fold change ≥ 1.2 or ≤ 0.8 were regarded as differential proteins at the discovery stage. Overall, 57 proteins were significantly dysregulated in TI vs. Ctr, 61 proteins were altered between TM vs. Ctr, and 10 differential proteins were identified in TM vs. TI (Figure 2A, Table S2). Hierarchical clustering and principal component analysis (PCA) (Figures 2B and 2C) of these differential proteins showed that β-thalassemia patients (TI and TM) and healthy volunteers can be clearly distinguished. In addition, there were significant differential protein

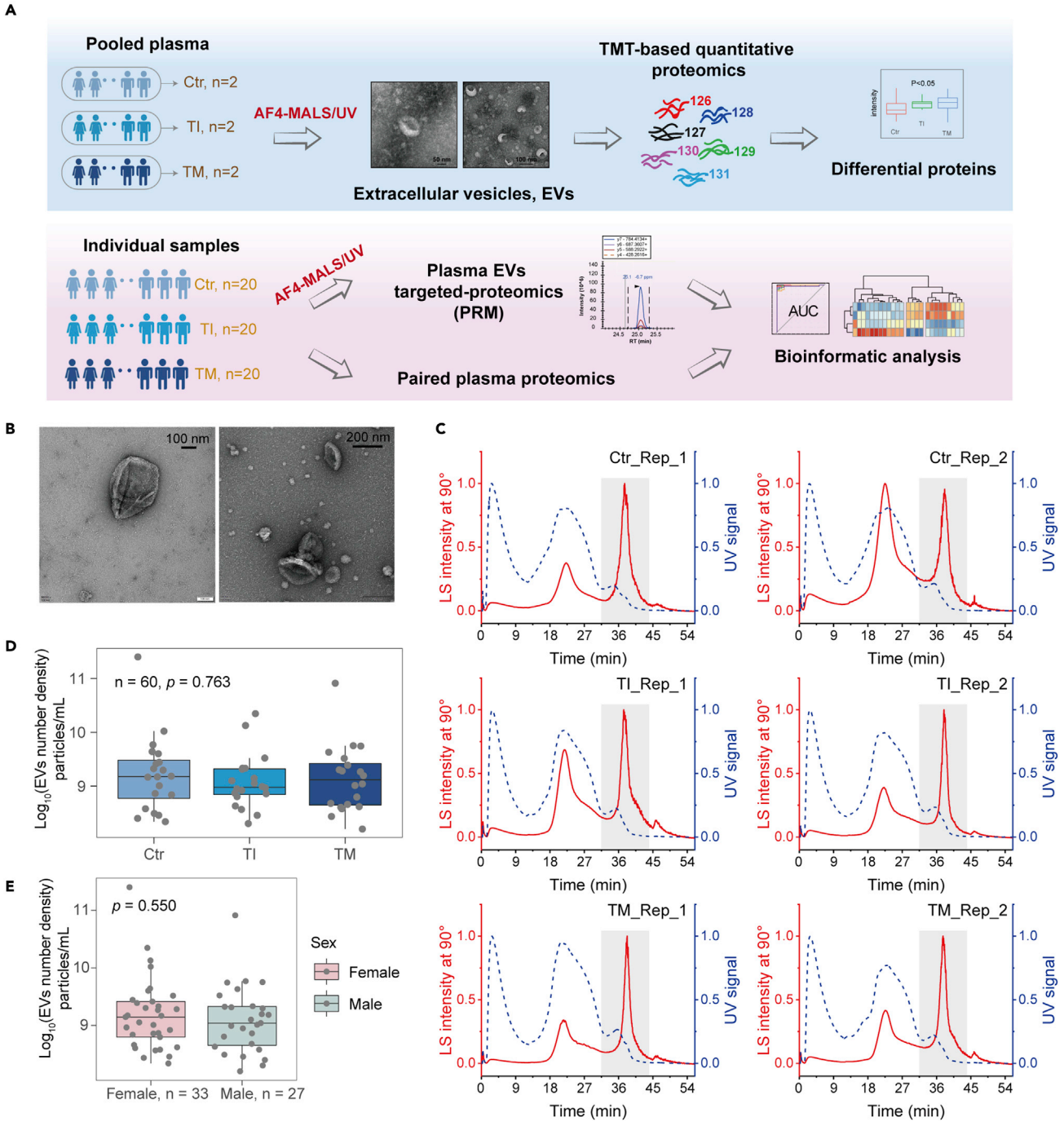


Figure 1. Plasma EV proteomic analyses of β-thalassemia patients and healthy controls

(A) Overview of the plasma EV proteomic workflow for analysis of TM and TI patients and healthy controls (Ctr). Paired plasma proteomic data were analyzed in parallel to reveal subgroup characteristics and potential biomarkers.

(B) Transmission electron microscopy micrographs of EVs. Scaled bars: 100 nm, left; 200 nm, right.

(C) AF4 separation fractograms obtained from six pooled samples. The red solid lines and blue dashed lines represent the light scattering (LS) intensity at 90°C and the UV signal, respectively.

Comparison of individual EV particle number density from different (D) groups (each group, n = 20) or (E) genders (female, n = 33; male, n = 27). 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 interquartile range (IQR). Statistical analyses were performed using one-way ANOVA for the comparison of three groups and a two-sample *t*-test for the comparison of genders. See also Figures S1–S4.

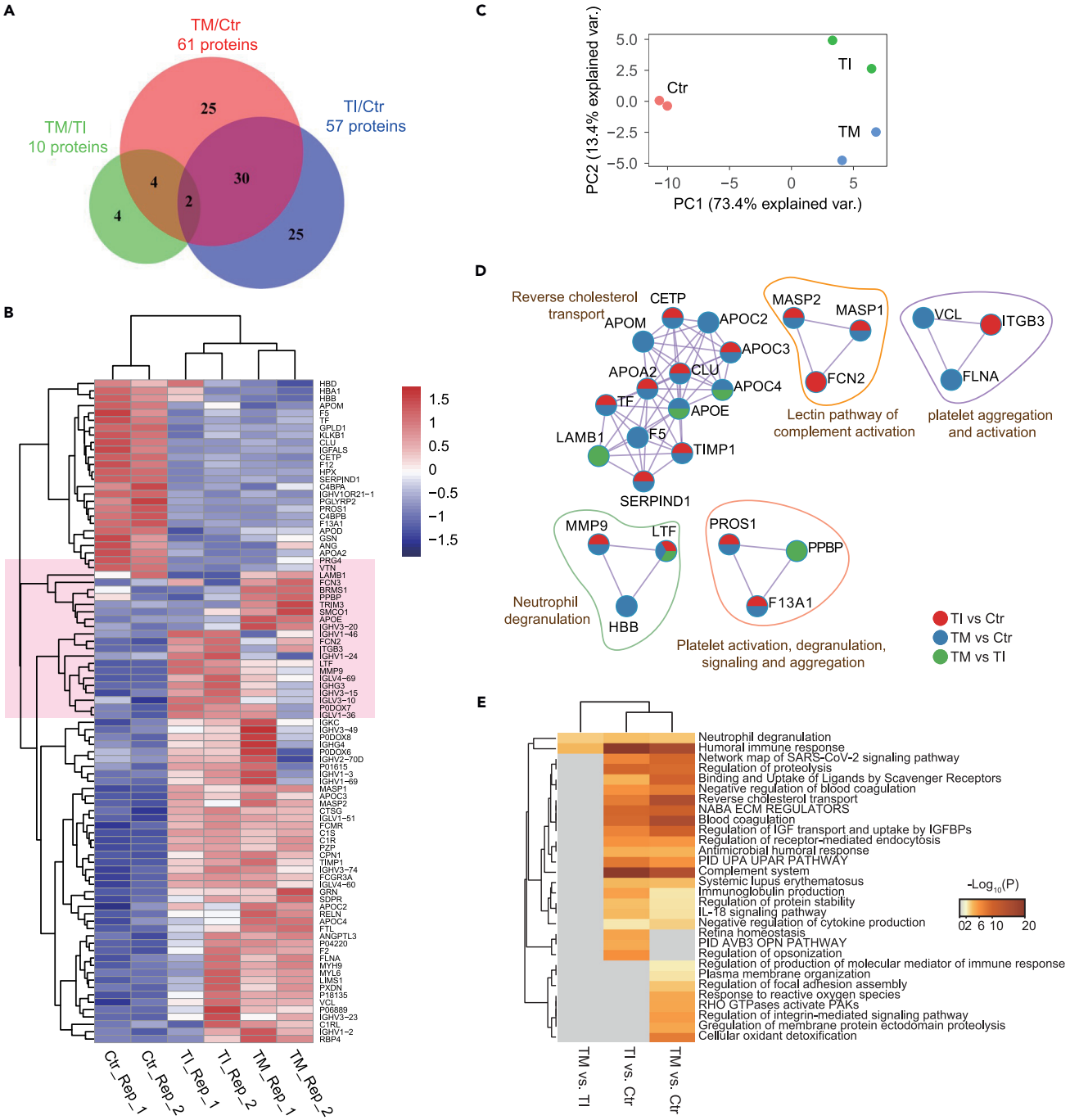


Figure 2. TMT-based quantitative proteomic analyses of EVs isolated from pooled samples

(A) Overlap of differential proteins in each pairwise comparison. There was a certain similarity and specificity in the number of dysregulated proteins when comparing TM/Ctr and TI/Ctr.

(B) Heatmap showing the altered proteins after comparing the three groups. Subtype-specific differential proteins are highlighted in the pink areas.

(C) Six pooled samples from three groups were clearly separated on these altered proteins.

(D) Protein-protein interaction analysis of these altered proteins using Metascape. The enriched biological functions of each cluster are marked.

(E) Enrichment analysis on the differential proteins observed in each pairwise comparison was performed using Metascape. Representative enriched terms with top p values are displayed in the dendrogram. The gradient color refers to the p values (lack of enrichment for the corresponding terms are colored in gray).

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2. See also Table S2.

clusters in the heatmap between TI and TM patients, who could also be separated in the PC2 dimension of the PCA.

A protein-protein interaction enrichment analysis showed these differential proteins were mainly involved in reverse cholesterol transport, lectin pathway of complement activation, neutrophil degranulation, and platelet-related pathways (Figure 2D). In order to test if there were functional changes among the three groups, we performed a functional assessment of pairwise differential proteins using Metascape. Our results showed that differentially expressed proteins in TI vs. Ctr were highly enriched in humoral immune response, complement system pathways, and retina homeostasis, while differential proteins in TM vs. Ctr were significantly enriched in reverse cholesterol transport, blood coagulation,binding and uptake of ligands by scavenger receptors, regulation of insulin-like growth factor (IGF) transport and uptake by insulin-like growth factor binding proteins (IGFBPs), or cellular oxidant detoxification (Figure 2E). The affected signaling pathways displayed different degrees of significance in the different subtypes.

Verification of the differential proteins by PRM

To verify the differential proteins identified in the discovery stage, a total of 216 tryptic peptides from 46 proteins were submitted for monitoring using PRM in 60 individual samples. Among the monitored proteins, 36 were successfully quantified, with 30 showing statistically significant differences between the three groups (Figure 3A and Tables S3 and S4). We found strong interactions across all proteins, except for immunoglobulins that were not matched in the STRING database (Figure 3B).

Compared with TMT-6plex experiments, we found a total of 17 differential and 3 new differential proteins between TM and Ctr that were validated or identified in the PRM experiments. In addition, we were able to validate 17 differential proteins and identify 5 new differential proteins between TI and Ctr. Importantly, we also identified but could not validate 5 differential proteins between TM and TI, with a total of 12 new differential proteins identified (Figures 3C and S4). Among the proteins that were uniquely identified as differential proteins, most are immunoglobulins (86%) in TI vs. Ctr; while in TM vs. Ctr, we found a predominance of lipoproteins and hemoglobin, which prompted us to further investigate the pathway features of the different types of β-thalassemia.

To investigate the changes in signaling pathways, gene set variation analysis (GSVA) and gene set enrichment analysis (GSEA) were separately performed. First, GSVA on untargeted proteomics revealed that the cellular oxidant detoxification and the immune system of TI were upregulated compared to TM, whereas the lipid metabolic process was downregulated in the former (Figure 3D). In addition, the result of GSEA on the plasma proteomic data (Table S5) was concordant with this observation and showed both immune response and antioxidant-activity-related gene sets were downregulated in TM, while the regulation of lipid metabolic process was upregulated (Figure 3E). The proteomic results of plasma and corresponding EVs revealed a more pronounced degree of immune dysfunction in TI than in TM patients. Furthermore, the proportion of differential immunoglobulins in TM vs. TI in plasma EVs (30.8% vs. 7.7%) was significantly higher than that of plasma (Figure S5), confirming the dysregulated levels of immunoglobulins are more striking in EVs.

Diagnostic performance evaluation of potential biomarkers

The discriminatory accuracy of each differential protein identified in the PRM analysis was compared using the area under the curve (AUC) in receiver operating characteristic (ROC) analysis. As shown in Figure 4A, the top six proteins with the largest AUCs for the diagnosis of β-thalassemia patients are complement C1s subcomponent (C1S), clusterin (CLU), lactotransferrin, complement C1r subcomponent, C4b-binding protein beta chain, and C4b-binding protein alpha chain. C1S and CLU were identified with the best diagnostic potential for the prediction of β-thalassemia patients.

For discriminating between TM and TI patients, we tested the diagnostic efficacy of these proteins on each group of patients. The top six proteins with the largest AUCs were serotransferrin (TF), hemoglobin subunit alpha (HBA1), immunoglobulin heavy constant gamma 4 (IGHG4), HBB, plasma kallikrein (KLKB1), and apolipoprotein M (APOM) (Figure 4B). Analysis of the corresponding plasma proteomic data showed the AUCs of these proteins in plasma EVs were higher than those in the plasma. In addition, the largest AUCs of the top six proteins in plasma were also lower than those of the aforementioned proteins, showing plasma EV-derived proteins are better at discriminating between TM and TI than those derived from

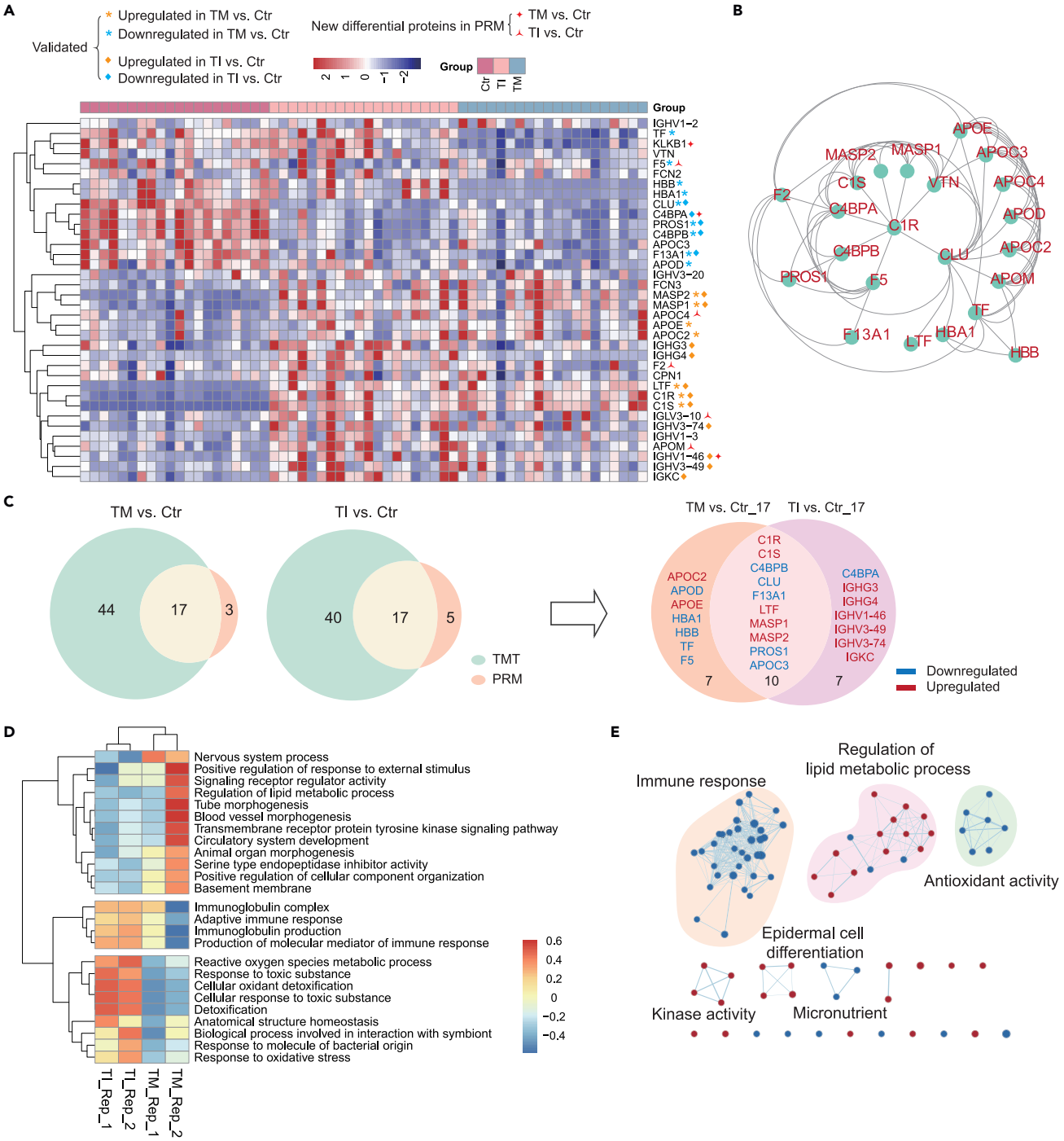


Figure 3. Differential proteins revealed at the discovery stage were further analyzed by PRM technology

(A) Validated and new differential proteins in the verification stage of our study.

(B) Strong protein-protein interactions between these proteins were observed.

(C) A high proportion of selected proteins were validated for testing. Venn diagrams showing the overlap in proteins validated in TM/Ctr and TI/Ctr.

(D) GSVA was performed on the TMT-based MS data of EVs for subgroup pathway-specific analysis.

(E) GSEA on differential plasma protein profiling of TM vs. TI.

See also Figure S5 and Tables S3–S5.

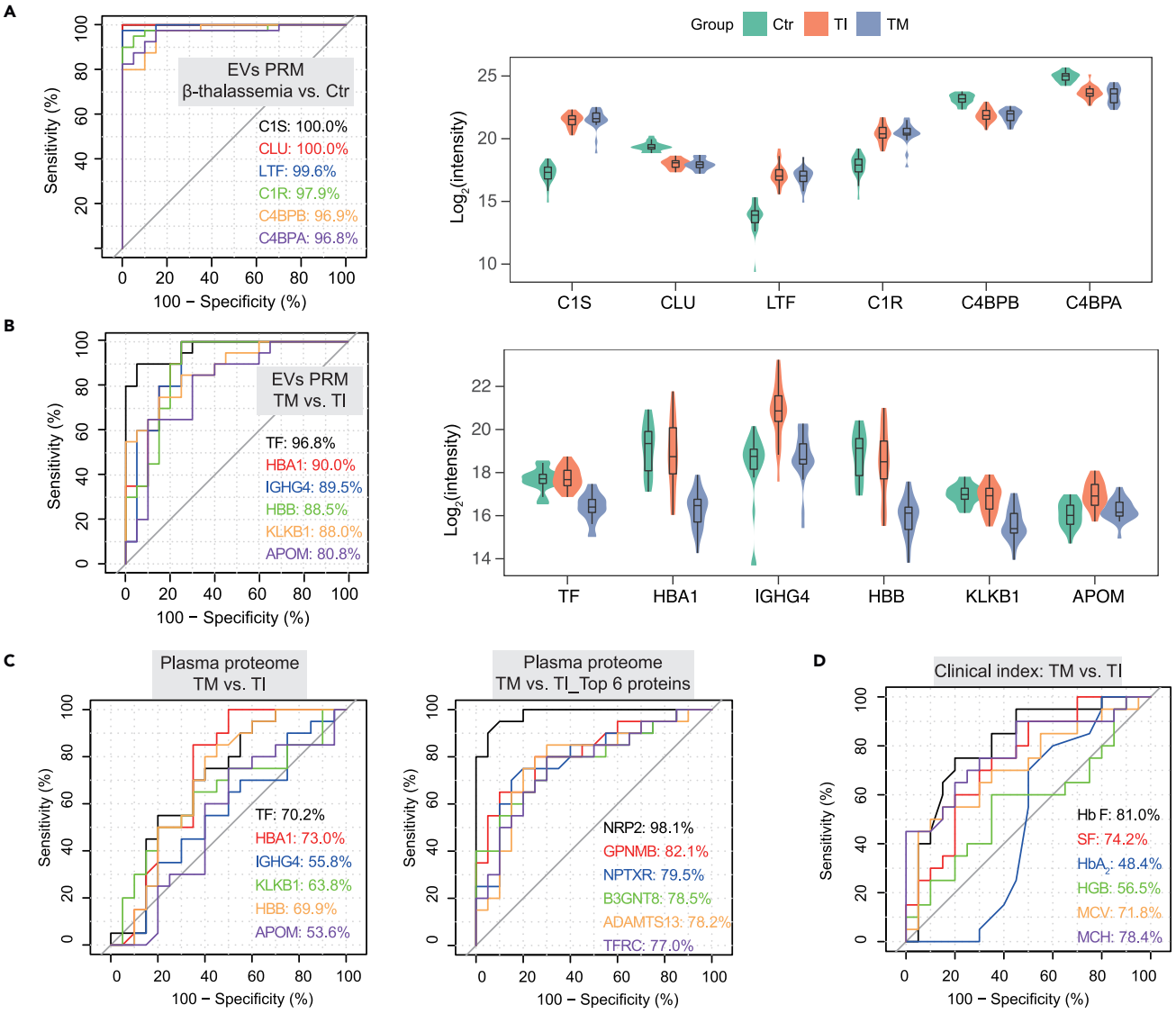


Figure 4. ROC curve evaluating the ability of proteins from plasma and plasma EVs to differentiate between the three groups

ROC curves of the top six proteins with the highest AUCs in EVs differentiating.

(A) β-Thalassemia patients from Ctr.

(B) TM patients from TI patients. The intensity of the corresponding proteins in the three groups is shown on the right.

(C) Plasma proteins were used for ROC curve analysis. The EV-derived top six proteins with the highest AUCs were analyzed on the level of plasma (right); the top six plasma proteins with the highest AUCs are shown on the left for comparison.

(D) ROC curve analysis of clinically diagnostic indices. 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 interquartile range (IQR).

plasma (Figure 4C). Furthermore, the AUCs of clinically acceptable Hb F and SF proteins were 81% and 74.2%, respectively, which are still lower than that of TF (Figure 4D). With an AUC of 96.8%, the TF in plasma EVs has the greatest diagnostic potential to differentiate between TI and TM patients.

A minimal set of variables is often set to build a prediction model, whereby we used the Boruta algorithm to measure the importance of all variables as this algorithm offers the most powerful approach based on previous reports.[26] The top 20 most important proteins ranked by Boruta for classification of the three groups are shown in Figure 5A. The aforementioned proteins (with high AUCs) are ranked top, and it was possible to observe high individual heterogeneity in patients with TI. This feature makes it significantly more difficult to differentiate between TI and TM than to distinguish between healthy controls and β-thalassemic

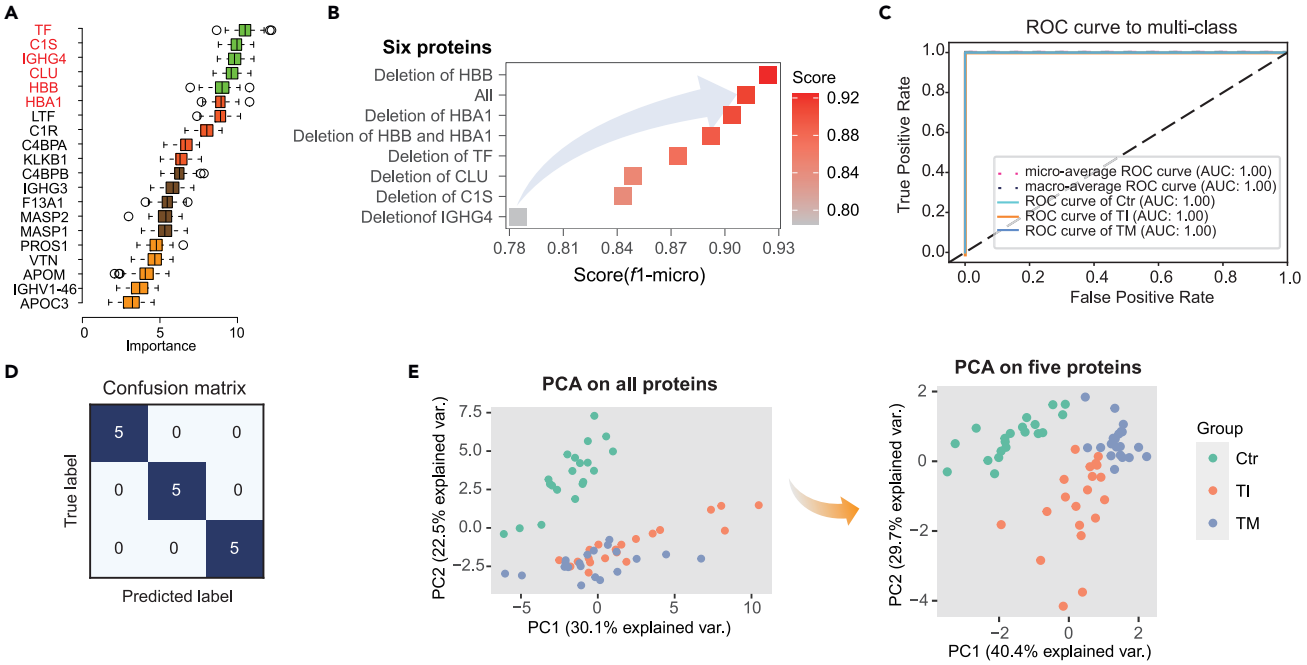


Figure 5. Extraction of characteristic differential proteins and construction of a machine learning model

(A) Boruta was used to measure the importance of each protein quantified in the PRM. The EV-derived top six proteins with the highest AUCs were ranked in the top. Only the top 20 ranked proteins are shown.

(B) Feature trimming strategy uncovering the importance of each protein as measured by *f*1-micro.

(C) ROC curve of the multiclass classifier trained using five proteins indicated a superior performance of these five EV proteins in subtype classification.

(D) Confusion matrix visualizing the performance of the classification of the machine learning model on the test set.

(E) Unsupervised PCA depicting the three groups were better defined on the protein panel than on all differential proteins. 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 interquartile range (IQR).

See also Figures S6 and S7.

patients, for which C1S and CLU are excellent. Based on these considerations, we selected the top six important proteins (TF, C1S, IGHG4, CLU, HBB, and HBA1) as variables for the subsequent analysis (Figure S6). In the modeling process trained by GridSearchCV, we used these proteins as candidates to build a support vector machine classifier and stratify the three groups. By trimming the number of features, we tested the importance of each feature against the mean score (scoring by F1-micro) in the 5-fold cross-validation. The best score was achieved when HBB was deleted from the 6 proteins (Figure 5B, score: 0.924), and Spearman's rank correlation analyses of these six proteins revealed high correlation between HBA1 and HBB (Figure S7). Thus, we retained TF, C1S, IGHG4, CLU, and HBA1 in the model. Finally, the three groups could be well stratified using these five proteins and the trained parameter settings (Figures 5C and 5D, AUC of each group = 1) on the test set (n = 15). Unsupervised clustering analysis of these five proteins in different individuals allowed for the separation of the three groups (Figure 5E).

DISCUSSION

The characterization and diagnosis of β-thalassemia subtypes will enhance clinical treatment and management of distinctly associated complications. For this purpose, the current study used quantitative proteomics to characterize proteome differences between β-thalassemia subgroups by comparing the differential proteomic patterns of plasma-derived EVs and plasma. We revealed β-thalassemia subgroup-specific patterns of dysregulated proteins and demonstrated that plasma EVs are better for subtype classification in clinical diagnosis.

We have previously developed an AF4-based EV isolation method[24] tailored for practical clinical application. The rapid, simple, and highly reproducible advantages of this method allow for characterizing EV proteins with biomarker potential in β-thalassemia. Although a sharp increase in EV concentration has been reported in thalassemic patients,[21,22] our results showed no significant differences in the density of

plasma EVs among the three groups under study (TM, TI, and Ctr). Previous studies reported that older age and splenectomy are independently associated with an increased risk of disease-related complications, thrombosis and pulmonary hypertension in particular.[27,28] We did not observe such patterns, which might reflect the different sampling strategy that included younger patients in our study.

We envisioned that the potential protein biomarkers found in plasma, a simple, easily accessible sample type, could assess disease severity. However, our previous plasma proteomics study of β-thalassemia patients[23] revealed that patients with TM and TI present with remarkably similar plasma proteomic profiles, and the differential proteins were not significant and not able to distinguish between TM and TI subtypes, which propelled us to perform the current proteomic study on plasma EVs. Based on the triangular MS-based biomarker mining strategy,[29] a total of 30 dysregulated plasma EV proteins were reported here in response to varying degrees of impaired β-globin synthesis. When comparing TM and TI groups, about 40% (12 of 30) of plasma EV proteins showed statistically significant differences, while only 9% (26 of 280) of plasma proteins were dysregulated, implying a distinct EV proteomic profile exists in different β-thalassemia subtypes that also have different degrees of pathway dysfunction. Patients with TI showed enhanced immune dysregulation, compared to lipid transport and metabolic process dysregulation in TM patients. Lipids are the most important biological targets of oxidative stress. In β-thalassemia, the accumulation of free α-globin leads to the formation of α-hemichromes and reactive oxygen species (ROS) that damage red blood cells. Further generation of ROS could also be promoted by the presence of iron.[30] TM patients exhibited stronger lipid transport and metabolic processes than TI patients probably due to both greater imbalance of the α/β globin subunits as well as more iron overload.

In addition to lipid abnormalities, we found immune response and immunoglobulin production were upregulated in TI vs. TM although increased immunoglobulin levels in β-thalassemia patients have long been reported.[31] On the one hand, ''overworking'' of the reticuloendothelial system in the clearance of hemolytic red cells may reduce its capacity for antigen clearance.[32] In the absence of transfusion therapy, TI patients typically have more ineffective erythropoiesis than well-transfused TM patients, which is probably the main reason for the underlying pathophysiological mechanisms and clinical complications usually observed in TI patients. On the other hand, increasing evidence indicates a higher rate of alloimmunization among TI patients than among TM patients.[33–35] It has been previously hypothesized that the recipient's immunity is a main contributing factor to alloimmunization against red blood cell antigens,[36] one of the side effects of transfusion. This observation suggests that the immune response of TI patients after transfusion therapy merits further attention in future studies. Furthermore, immune abnormalities are responsible for the increased susceptibility of β-thalassemia patients to infectious diseases,[37] which thus represents an important cause of morbidity and mortality associated with the disease.

Despite the similarity in signaling pathways between plasma and plasma EVs, the proteins present in the former are less distinguishable in different β-thalassemia subtypes than those present in plasma EVs. ROC analysis on TM and TI patients showed higher sensitivity and specificity for EV proteins than for plasma proteins to distinguish different subtypes. We used a machine-learning-based pipeline and developed a set of biomarker combinations to distinguish between different subtypes of β-thalassemia and healthy controls. The biomarker panel was composed of five proteins, including TF, C1S, IGHG4, CLU, and HBA1, all of which are functionally strongly associated with β-thalassemia. HBA1 represents α-globin in the adult hemoglobin, and an imbalance between this chain and the β-globin chain represents the molecular pathogenesis of β-thalassemia. C1S is responsible for the activation and proteolytic activity of the C1 complex of the complement system, which is activated by high levels of C1S. In addition, the urinary levels of CLU are a promising biomarker for kidney injury and early renal damage associated with β-thalassemia in children.[38] IGHG4 lies within a conserved region of the immunoglobulin heavy chains, mediating the effector phase of humoral immunity and helping to eliminate bound antigens. Based on our findings, the differences in immunity abnormalities should drive the subtyping of β-thalassemia patients. As expected, IGHG4 played a crucial role in this panel (Figure 5B). Finally, TF is involved in iron metabolism and β-thalassemia. Interestingly, TF was downregulated in TM patients only, while no differences were observed in TI patients compared to healthy controls. Inadequate circulating transferrin is postulated to be the reason for the maldistribution of iron in the bone marrow, liver, and/or spleen in cases of β-thalassemia.[39] Li et al. reported that exogenous transferrin normalized labile plasma iron concentrations in β-thalassemic mice and represented a limiting factor contributing to anemia.[40] Specifically, low levels of transferrin in TM patients were observed in the plasma but mainly in plasma EVs (Figure S8). Hence, reduced transferrin in EVs may be

largely responsible for disturbed iron utilization in the bone marrow, whereby the administration of transferrin via EVs may be more effective for treatment than administration via plasma. A plausible explanation for why only TM patients present with low transferrin levels is that liver-derived transferrin may be disturbed by severe iron overload in the hepatocytes, a pattern consistently observed in TM patients.

In conclusion, our comprehensive plasma and plasma-derived EV proteomic analysis of β-thalassemia patients demonstrated differential proteomic patterns exist between β-thalassemia subtypes. Our study broadened the knowledge on the pathophysiological features of β-thalassemia subtypes and, more importantly, demonstrates that plasma-derived EV proteins can be utilized as new biomarkers to differentiate between healthy individuals and patients with TI and TM forms of the disease. Therefore, we suggest that plasma EVs should be more potential than plasma itself in the search for protein biomarkers for the definition or diagnosis of β-thalassemia disease severity. The findings in this study may benefit accurate subtyping and precision diagnosis of β-thalassemia in the future.

Limitations of the study

In this study, blood samples were collected two weeks after the last transfusion and just before the next transfusion according to the patient's frequency of transfusion. The interference caused by stored blood transfusion should be concerned, even if its contribution has been minimized. Moreover, although we found significant differences in the abundance of many immunoglobulins in different β-thalassemia patients, the underlying mechanisms of abnormal immune function between subgroups and the extent to which alloimmunization contributes to immunoglobulin dysregulation are unclear. This warrants further exploration to better serve the clinical management of different patients. Furthermore, we believe that the inclusion of ROC analyses of EV-depleted plasma proteins could be almost a proof of the superiority of EV protein composition as biomarkers, even though plasma EV proteins represent a very small percentage of total plasma protein content. Given the diagnostic potential of the five proteins of plasma-derived EVs for accurate subtyping and precise diagnosis of β-thalassemia, further verification of these five or fewer proteins in the panel with larger independent populations and more clinical conditions will be the follow-up work.

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
  - Clinical specimens
- METHOD DETAILS
  - Plasma-derived extracellular vesicles (EVs) isolation and characterization
  - In-solution digestion
  - TMT-based LC-MS/MS analysis
  - Spectral library construction and PRM analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bioinformatic analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.106048.

ACKNOWLEDGMENTS

This study was supported by the National Key Research and Development Program of China under grants 2018YFA0507801 and 2018YFA0507802.

AUTHOR CONTRIBUTIONS

N.L., B.W., and F.Y. conceptualized and designed the study. Clinical samples were collected and provided by F.W., J.L., P.A., Y. Li, Y. Liu, and Y.H. N.L. performed the data analysis, generated figures for the

manuscript, and wrote the manuscript. B.W. performed the experimental work and contributed to the data analysis. J.W. was responsible for LC-MS/MS data collection. F.Y. supervised the project. Y.Y. and Q.Q. contributed to the experimental work. T.C., X.C., X.D., L.N., Z.X., M.Z., and X.G. provided helpful inputs on MS-based proteomic data collection and contributed with comments to the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 27, 2022
Revised: December 1, 2022
Accepted: January 20, 2023
Published: January 25, 2023

REFERENCES

1. Origa, R. (2017). beta-Thalassemia. Genet. Med. *19*, 609–619. https://doi.org/10.1038/gim.2016.173.

2. 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.

3. Hariharan, P., and Nadkarni, A. (2021). Insight of fetal to adult hemoglobin switch: genetic modulators and therapeutic targets. Blood Rev. *49*, 100823. https://doi.org/10.1016/j.blre.2021.100823.

4. Luo, S., Chen, X., Zeng, D., Tang, N., Yuan, D., Zhong, Q., Mao, A., Xu, R., and Yan, T. (2022). The value of single-molecule real-time technology in the diagnosis of rare thalassemia variants and analysis of phenotype-genotype correlation. J. Hum. Genet. *67*, 183–195. https://doi.org/10.1038/s10038-021-00983-1.

5. Viprakasit, V., and Ekwattanakit, S. (2018). Clinical classification, screening and diagnosis for thalassemia. Hematol. Oncol. Clin. North Am. *32*, 193–211. https://doi.org/10.1016/j.hoc.2017.11.006.

6. Taher, A.T., Viprakasit, V., Musallam, K.M., and Cappellini, M.D. (2013). Treating iron overload in patients with non-transfusion-dependent thalassemia. Am. J. Hematol. *88*, 409–415. https://doi.org/10.1002/ajh.23405.

7. M.D. Cappellini, A. Cohen, J. Porter, A. Taher, and V. Viprakasit, eds. (2014). Guidelines for the Management of Transfusion Dependent Thalassaemia (TDT), 3rd edition.

8. Saliba, A.N., and Taher, A.T. (2016). Morbidities in non-transfusion-dependent thalassemia. Ann. N. Y. Acad. Sci. *1368*, 82–94. https://doi.org/10.1111/nyas.13083.

9. Cappellini, M.D., and Motta, I. (2017). New therapeutic targets in transfusion-dependent and -independent thalassemia. Hematology. Am. Soc. Hematol. Educ. Program *2017*, 278–283. https://doi.org/10.1182/asheducation-2017.1.278.

10. Taher, A., Isma'eel, H., Mehio, G., Bignamini, D., Kattamis, A., Rachmilewitz, E.A., and Cappellini, M.D. (2006). Prevalence of thromboembolic events among 8,860 patients with thalassaemia major and intermedia in the Mediterranean area and Iran. Thromb. Haemost. *96*, 488–491.

11. Cappellini, M.D., Grespi, E., Cassinerio, E., Bignamini, D., and Fiorelli, G. (2005). Coagulation and splenectomy: an overview. Ann. N. Y. Acad. Sci. *1054*, 317–324. https://doi.org/10.1196/annals.1345.039.

12. Taher, A.T., Musallam, K.M., and Inati, A. (2009). The hypercoagulable state in thalassemia intermedia. Hemoglobin *33* (*Suppl 1*), S160–S169. https://doi.org/10.3109/03630260903351619.

13. Aessopos, A., Farmakis, D., Deftereos, S., Tsironi, M., Tassiopoulos, S., Moyssakis, I., and Karagiorga, M. (2005). Thalassemia heart disease: a comparative evaluation of thalassemia major and thalassemia intermedia. Chest *127*, 1523–1530. https://doi.org/10.1378/chest.127.5.1523.

14. 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.

15. Rontogianni, S., Synadaki, E., Li, B., Liefaard, M.C., Lips, E.H., Wesseling, J., Wu, W., and Altelaar, M. (2019). Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. Commun. Biol. *2*, 325. https://doi.org/10.1038/s42003-019-0570-8.

16. Shah, R., Patel, T., and Freedman, J.E. (2018). Circulating extracellular vesicles in human disease. N. Engl. J. Med. *379*, 958–966. https://doi.org/10.1056/NEJMra1704286.

17. Marar, C., Starich, B., and Wirtz, D. (2021). Extracellular vesicles in immunomodulation and tumor progression. Nat. Immunol. *22*, 560–570. https://doi.org/10.1038/s41590-021-00899-0.

18. Kheansaard, W., Phongpao, K., Paiboonsukwong, K., Pattanapanyasat, K., Chaichompoo, P., and Svasti, S. (2018). Microparticles from beta-thalassaemia/HbE patients induce endothelial cell dysfunction. Sci. Rep. *8*, 13033. https://doi.org/10.1038/s41598-018-31386-6.

19. Atipimonpat, A., Siwaponanan, P., Khuhapinant, A., Svasti, S., Sukapirom, K., Khowawisetsut, L., and Pattanapanyasat, K. (2021). Extracellular vesicles from thalassemia patients carry iron-containing ferritin and hemichrome that promote cardiac cell proliferation. Ann. Hematol. *100*, 1929–1946. https://doi.org/10.1007/s00277-021-04567-z.

20. Kittivorapart, J., Crew, V.K., Wilson, M.C., Heesom, K.J., Siritanaratkul, N., and Toye, 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.

21. Chaichompoo, P., Kumya, P., Khowawisetsut, L., Chiangjong, W., Chaiyarit, S., Pongsakul, N., Sirithanaratanakul, N., Fucharoen, S., Thongboonkerd, V., and Pattanapanyasat, K. (2012). Characterizations and proteome analysis of platelet-free plasma-derived microparticles in beta-thalassemia/hemoglobin E patients. J. Proteomics *76*, 239–250. https://doi.org/10.1016/j.jprot.2012.06.004.

22. Levin, C., Koren, A., Rebibo-Sabbah, A., Koifman, N., Brenner, B., and Aharon, A. (2018). Extracellular vesicle characteristics in beta-thalassemia as potential biomarkers for spleen functional status and ineffective erythropoiesis. Front. Physiol. *9*, 1214. https://doi.org/10.3389/fphys.2018.01214.

23. Li, N., An, P., Wang, J., Zhang, T., Qing, X., Wu, B., Sun, L., Ding, X., Niu, L., Xie, Z., et al. (2022). Plasma proteome profiling combined with clinical and genetic features reveals the pathophysiological characteristics of beta-thalassemia. iScience *25*, 104091. https://doi.org/10.1016/j.isci.2022.104091.

24. Wu, B., Chen, X., Wang, J., Qing, X., Wang, Z., Ding, X., Xie, Z., Niu, L., Guo, X., Cai, T., et al. (2020). Separation and characterization of extracellular vesicles from human plasma by asymmetrical flow field-flow fractionation. Anal. Chim. Acta *1127*, 234–245. https://doi.org/10.1016/j.aca.2020.06.071.

25. Cai, T., Zhang, Q., Wu, B., Wang, J., Li, N., Zhang, T., Wang, Z., Luo, J., Guo, X., Ding, X., et al. (2021). LncRNA-encoded microproteins: a new form of cargo in cell culture-derived and circulating extracellular vesicles. J. Extracell. Vesicles *10*, e12123. https://doi.org/10.1002/jev2.12123.

26. Degenhardt, F., Seifert, S., and Szymczak, S. (2019). Evaluation of variable selection methods for random forests and omics data sets. Brief. Bioinform. *20*, 492–503. https://doi.org/10.1093/bib/bbx124.

27. Taher, A.T., Musallam, K.M., Karimi, M., El-Beshlawy, A., Belhoul, K., Daar, S., Saned, M.S., El-Chafic, A.H., Fasulo, M.R., and Cappellini, M.D. (2010). Overview on practices in thalassemia intermedia management aiming for lowering complication rates across a region of endemicity: the OPTIMAL CARE study. Blood *115*, 1886–1892. https://doi.org/10.1182/blood-2009-09-243154.

28. Amin, S.S., Jalal, S.D., Ali, K.M., Mohammed, A.I., Rasool, L.K., and Osman, T.J. (2020). Beta-thalassemia intermedia: a single thalassemia center experience from Northeastern Iraq. BioMed Res. Int. *2020*, 2807120. https://doi.org/10.1155/2020/2807120.

29. Geyer, P.E., Holdt, L.M., Teupser, D., and Mann, M. (2017). Revisiting biomarker discovery by plasma proteomics. Mol. Syst. Biol. *13*, 942. https://doi.org/10.15252/msb.20156297.

30. Mettananda, S., Gibbons, R.J., and Higgs, D.R. (2015). alpha-Globin as a molecular target in the treatment of beta-thalassemia. Blood *125*, 3694–3701. https://doi.org/10.1182/blood-2015-03-633594.

31. Wasi, C., Wasi, P., and Thongcharoen, P. (1971). Serum-immunoglobulin levels in thalassaemia and the effects of splenectomy. Lancet *2*, 237–239. https://doi.org/10.1016/s0140-6736(71)92573-6.

32. Tovo, P.A., Miniero, R., Barbera, C., Sacchetti, L., and Saitta, M. (1981). Serum immunoglobulins in homozygous beta-thalassemia. Acta Haematol. *65*, 21–25. https://doi.org/10.1159/000207144.

33. Azarkeivan, A., Ansari, S., Ahmadi, M.H., Hajibeigy, B., Maghsudlu, M., Nasizadeh, S., Shaigan, M., Toolabi, A., and Salahmand, M. (2011). Blood transfusion and alloimmunization in patients with thalassemia: multicenter study. Pediatr. Hematol. Oncol. *28*, 479–485. https://doi.org/10.3109/08880018.2011.568595.

34. Abdelrazik, A.M., Elshafie, S.M., El Said, M.N., Ezzat Ahmed, G.M., Al-Gamil, A.K.A., El Nahhas, M.G.M., and Sady, A.A.B. (2016). Study of red blood cell alloimmunization risk factors in multiply transfused thalassemia patients: role in improving thalassemia transfusion practice in Fayoum, Egypt. Transfusion *56*, 2303–2307. https://doi.org/10.1111/trf.13695.

35. Ang, A.L., Lim, C.Y., Ng, W.Y., and Lam, J.C.M. (2021). Non-transfusion dependent thalassemia is independently associated with higher alloimmunization risk than transfusion dependent thalassemia and would benefit the most from extended red cell antigen-matching. Transfusion *61*, 2566–2577. https://doi.org/10.1111/trf.16590.

36. Al-Riyami, A.Z., and Daar, S. (2019). Red cell alloimmunization in transfusion-dependent and transfusion-independent beta thalassemia: a review from the Eastern Mediterranean Region (EMRO). Transfus. Apher. Sci. *58*, 102678. https://doi.org/10.1016/j.transci.2019.102678.

37. Farmakis, D., Giakoumis, A., Polymeropoulos, E., and Aessopos, A. (2003). Pathogenetic aspects of immune deficiency associated with beta-thalassemia. Med. Sci. Monit. *9*, RA19-22.

38. Abd El-Khalik, S.R., Sharaby, R.M., Nasif, E., Hamza, M.B., and Ibrahim, R.R. (2021). Netrin-1 and clusterin: innovative potential diagnostic biomarkers for early renal damage in beta-thalassemia major children. IUBMB Life *73*, 800–810. https://doi.org/10.1002/iub.2464.

39. Ginzburg, Y.Z., Rybicki, A.C., Suzuka, S.M., Hall, C.B., Breuer, W., Cabantchik, Z.I., Bouhassira, E.E., Fabry, M.E., and Nagel, R.L. (2009). Exogenous iron increases hemoglobin in beta-thalassemic mice. Exp. Hematol. *37*, 172–183. https://doi.org/10.1016/j.exphem.2008.10.004.

40. Li, H., Rybicki, A.C., Suzuka, S.M., von Bonsdorff, L., Breuer, W., Hall, C.B., Cabantchik, Z.I., Bouhassira, E.E., Fabry, M.E., and Ginzburg, Y.Z. (2010). Transferrin therapy ameliorates disease in beta-thalassemic mice. Nat. Med. *16*, 177–182. https://doi.org/10.1038/nm.2073.

41. 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.

42. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf. *12*, 77. https://doi.org/10.1186/1471-2105-12-77.

43. 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.

44. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

45. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.

46. 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.

47. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504. https://doi.org/10.1101/gr.1239303.

48. Zhang, R., He, Y., Yi, J., Zhang, L., Shen, C., Liu, S., Liu, L., Liu, B., and Qiao, L. (2020). Proteomic and metabolic elucidation of solar-powered biomanufacturing by bio-abiotic hybrid system. Chem *6*, 234–249. https://doi.org/10.1016/j.chempr.2019.11.002.

49. Sun, Y., Guo, Z., Liu, X., Yang, L., Jing, Z., Cai, M., Zheng, Z., Shao, C., Zhang, Y., Sun, H., et al. (2022). Noninvasive urinary protein signatures associated with colorectal cancer diagnosis and metastasis. Nat. Commun. *13*, 2757. https://doi.org/10.1038/s41467-022-30391-8.

50. Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. *49*, D605–D612. https://doi.org/10.1093/nar/gkaa1074.

51. Kursa, M.B., and Rudnicki, W.R. (2010). Feature selection with the Boruta package. J. Stat. Softw. *36*, 1–13. https://doi.org/10.18637/jss.v036.i11.

52. Kursa, M.B., Jankowski, A., and Rudnicki, W.R. (2010). Boruta - a system for feature selection. Fundam. Inform. *101*, 271–285. https://doi.org/10.3233/FI-2010-288.

53. Fabian Pedregosa, G.V., Gramfort, A., Michel, V., Bertrand, T., Grisel, O., Blondel, M., Peter, P., Weiss, R., Vincent, D., Vanderplas, J., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD9	Abcam	ab92726; RRID: AB_10561589
CD63	Abcam	ab134045; RRID: AB_2800495
CD81	Abcam	ab109201; RRID: AB_10866464
TSG101	Abcam	ab125011; RRID: AB_10974262
HSP90	Abcam	ab34909; RRID: AB_733041
APOA1	Abcam	ab52945; RRID: AB_2056661
APOB	Abcam	ab20737; RRID: AB_2056954
Albumin	Abcam	ab207327; RRID: AB_2755031
Biological samples		
Whole blood	The First Affiliated Hospital of Guangxi Medical University	This paper (Table S1)
Chemicals, peptides, and recombinant proteins		
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
Critical commercial assays		
TMT 6-plex reagents	Thermo Fisher Scientific	Cat #90061
The indexed retention time standards (iRT) peptides kit	Biognosys	Cat #Ki-300-2
Deposited data		
Mass spectrometry data	http://www.proteomexchange.org	PXD034548 PXD032179
Software and algorithms		
Proteome Discoverer Version 2.2.0.388	Thermo Fisher Scientific	N/A
SpectroDive 10.4	Biognosys	N/A
R V3.6.2	R Project	https://www.r-project.org
RStudio Version 1.2.1335	Open source	https://www.rstudio.com/
Python 3.9.12	N/A	https://www.python.org
GSEA software V4.1.0	N/A	http://www.gsea-msigdb.org/gsea/
String	N/A	https://cn.string-db.org
Metascape	N/A	https://metascape.org/
Cytoscape 3.7.2	Open source	https://cytoscape.org
Astra Version 5.3.4.20	Wyatt Technology	N/A
ZetaView software 8.04.02 SP2	Particle Metrix	N/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).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The mass spectrometry proteomics data have been deposited into the ProteomeXchange Consortium via the iProX[41] partner repository with dataset identifiers ProteomeXchange: PXD034548, PXD032179.
- 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

Clinical specimens

Peripheral blood samples of 20 TI ($\beta^+\beta^+/\beta^+\beta^0$, 9.2 ± 2.7 years, male: 10, female: 10) patients, 20 TM ($\beta^0\beta^0$, 8.9 ± 2.3 years, male: 9, female: 11) patients, and 20 age- and sex-matched healthy individuals (9.9 ± 2.3 years, male: 8, female: 12) were provided by The First Affiliated Hospital of Guangxi Medical University (Guangxi Zhuang Autonomous Region, China). Detailed clinical information is available in Table S1 and from a previous plasma proteomics study.[23] Written informed consent was obtained from the parents or legal guardians of patients under the age of consent. All patients received regular blood transfusion, iron chelation therapy, and none underwent splenectomy. A large proportion of these patients require regular blood transfusions (> once a month), whereby sampling was performed two weeks after the last transfusion to eliminate the impact of blood transfusion for the results. A questionnaire including the first time of transfusion, treatment strategies and medical conditions was answered by each patient. Briefly, whole blood was collected into EDTA-anticoagulated tubes and centrifuged at 3,000 × *g* for 10 min to remove debris and platelets, after which the supernatant was aliquoted and stored at −80°C until use. This study was approved by the Ethics Committee of The First Affiliated Hospital of Guangxi Medical University and performed in accordance with the Declaration of Helsinki.

For TMT 6plex-based quantitative proteomics, we randomly selected plasma from 14 individuals in each group and pooled it into two samples (7 individuals/sample). For PRM validation, an individual plasma sample was used.

METHOD DETAILS

Plasma-derived extracellular vesicles (EVs) isolation and characterization

After diluted with the equal volume of PBS, the plasma samples were further centrifugated at 4200 × *g* for 20 min at 4°C to remove dead cell debris, apoptotic bodies, and platelets. EVs were isolated from the collected supernatant using asymmetrical flow field-flow fractionation (AF4) technology with an Eclipse 3 system (Wyatt Technology Europe GmbH, Germany).[24] The system contains an HPLC isocratic pump, a manual injection valve with a 500 μL loop, a UV detector at a wavelength of 280 nm (Agilent Technologies, California) and a MALS detector with 18 scattering angles from 14.4° to 163.3° at 658 nm. The separation channel was a trapezoidal shaped Mylar spacer with a thickness of 350 μm above a regenerated cellulose membrane (10 kDa, SUPERON). The tip-to-tip length of the channel was 152 mm and the initial triangle channel breadth of 21.5 mm was decreased to a final breadth of 3 mm. PBS buffer was filtered with a 0.22 μm filter membrane and used as mobile phase for separation. The Astra software (Version 5.3.4.20, Wyatt Technology) was used for data acquisition and processing. The total flow rate was maintained at 2.7 mL/min for 9 min for sample injection and simultaneous focusing/relaxation. After this, the two valves were rotated so that all flows entered the channel inlet for separation at flow rate conditions (outflow rate: 4 mL/min; crossflow rate: 3 mL/min). EVs fractions were collected at Peak 4 during AF4 separation and concentrated using Millipore Amicon® Ultra-0.5 3 kDa MWCO filters. For transmission electron microscopy analysis, one drop of concentrated EVs fractions was applied to glow discharged carbon-layered copper

grids for 1 min, and negatively stained with 2% uranyl acetate solution for 1 min. For nanoparticle tracking analysis, the EVs samples (n = 3) were diluted in PBS to obtain the optimal detection concentration of 10^5–10^7 particles/mL, then analyzed using ZetaView to determine vesicle concentration and size distribution. The data were analyzed using ZetaView software 8.04.02 SP2. For Western blot analysis, the protein concentration was determined by BCA assay and a 10 μg protein loading volume of EVs sample or whole plasma (each n = 2) was used for protein characterization.

In-solution digestion

After solubilized in lysis buffer (8 M urea, 50 mM NH_4HCO_3), EVs proteins were reduced with 20 mM dithiothreitol for 1 h at 37°C and alkylated with 40 mM iodoacetamide for 45 min at room temperature in the dark. For protein digestion, Lys C was first added at an enzyme/protein ratio of 1:50 (w/w) for 3 h at 37°C. After diluting with 25 mM NH_4HCO_3 to the final concentration of <1.5 M urea, the samples were further digested at 37°C overnight with trypsin at an enzyme/protein ratio of 1:50 (w/w). The digestion was then stopped by adding formic acid to a final concentration of 0.1% and desalted with a C18 Oasis HLB cartridge (Waters, USA). All peptide samples were dried in a centrifugal vacuum concentrator (Labconco, USA).

TMT-based LC-MS/MS analysis

Desalted peptides were reconstituted in 100 mM HEPES (pH 8.5) and processed according to the manufacturer's protocol with the TMT 6-plex kit (Thermo Fisher Scientific). Briefly, one unit of TMT reagent was equilibrated and resuspended with acetonitrile, and each peptide solution was incubated with a specific tag (tags 126, 127 used for Ctr samples; tags 128, 129 used for TI samples and 130, 131 used for TM samples, respectively) for 1h at room temperature. 5% hydroxylamine was then added to each sample for 15 min to quench the reaction. Samples were mixed at equal amounts, vacuum-dried and kept at −80°C. Subsequently, TMT-labeled pooled peptide sample was then dissolved with 0.1% FA, desalted using a C18 Oasis HLB cartridge, and dried by vacuum centrifugation.

The peptide mixtures were dissolved in buffer A (2% ACN, pH 10, adjusted with ammonium hydroxide) and then injected onto a C18 column (130 Å, 5 μm, 4.6 × 250 mm, Waters, USA) and eluted using buffers A and B (98% ACN, pH 10) with a 76 min gradient of buffer B (5–8% for 5 min, 8–18% for 35 min, 18–32% for 22 min, 32–95% for 2 min, 95–95% for 4 min, and 95–5% for 1 min, 5–5% for 7 min) at a flow rate of 0.7 mL/min on a RIGOL L-3000 Series HPLC system. The eluted peptide fractions were collected every 90s and combined into 10 fractions by pooling fraction 1, 11, 21; fraction 2, 12, 22; and so on. All combined fractions were dried using speed vacuum and stored at −80°C.

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). The lyophilized peptides were resuspended in 0.1% FA, loaded onto an in-house packed C18 trap column (100 μm × 2 cm, Reprosil-Pur C18 AQ, 5 μm, Dr. Maisch GmbH, Germany) and then separated on an in-house packed C18 capillary column (75 μm × 20 cm, Reprosil-Pur C18 AQ, 3 μm, Dr. Maisch GmbH, Germany). The peptides were eluted using mobile phase A (0.1% FA) and B (ACN/0.1% FA) with a 78-min gradient (4–10% B, 5 min; 10–22% B, 50 min; 22–32% B, 15 min; 32–90% B, 1 min; 90% B, 7 min) at a flow rate of 310 nL/min. MS analysis was operated on a data-dependent acquisition (DDA) mode. MS data were acquired at a high resolution 70,000 (m/z 200) across the mass range of 300–1600 m/z with an AGC of 3e6 and a maximum injection time of 60 ms. The top 20 precursor ions were selected from each MS full scan with an isolation width of 2 m/z for fragmentation with normalized collision energy of 27%. MS/MS spectra were acquired at a resolution of 17,500 at m/z 200. The target value was 5e4 with a maximum injection time of 80 ms. Ions selected for MS/MS were dynamically excluded for a duration of 50 s.

The MS/MS data were processed with Proteome Discoverer 2.2 (Thermo Scientific) and searched against the Uniprot human protein database (downloaded on February 2018) together with a common contaminant list using the Sequest HT search engine for protein identification. Trypsin was selected as the enzyme and two miss cleavages were allowed. The mass tolerance of precursor and product ions were set to 10 ppm and 0.02 Da, respectively. Carbamidomethylation on cysteine and TMT-labeling (lysine and peptide N-terminus) were set as fixed modifications. Methionine oxidation and acetylation of the N-terminus were specified as the variable modifications. The length of minimal peptide was specified to 6 amino acid residues. The false discovery rate (FDR) was set to 1% for both protein and peptide identifications. Contaminant proteins were removed from the datasets before downstream analysis was performed.

Spectral library construction and PRM analysis

PRM-MS/MS analysis was performed using the Orbitrap Eclipse Tribrid mass spectrometer coupled with an EASY-nLC 1200 HPLC system. Peptide sample was loaded onto an in-house packed C18 trap column (100 μm × 2 cm, Reprosil-Pur C18 AQ, 5 μm, Dr. Maisch GmbH) and separated on a C18 capillary column (75 μm × 20 cm, Reprosil-Pur C18 AQ, 3 μm, Dr. Maisch GmbH) using mobile phases A (0.1% FA) and B (ACN/0.1% FA) with a gradient (5–10% phase B, 3 min; 10–20% B, 22 min; 20–30% B, 22 min; 30–40% B, 13 min; 40–99% B, 4 min; 99% B, 9 min) at a flow rate of 300 nL/min. The indexed retention time standards (iRT) peptides kit (Biognosys, Switzerland) was used for retention time calibration.

For the development of the PRM assay, the peptides from each sample were equally mixed together, and 1 μg of the mixed peptides spiked with 10 × iRT standard peptides were used for data-dependent MS/MS analysis. Full MS scans acquired at a resolution of 60,000 at m/z 200 across the mass range of 350–1500 *m/z* with an AGC of 4e5 and maximum injection time of 50 ms. MS/MS spectra were acquired at a resolution of 15,000 at m/z 200, an isolation window of 1.6, a collision energy of 30% and an AGC target value of 5e4 with a maximum injection time of 22 ms. The SpectroDive™ 10.4 (Biognosys) software was used for PRM panel generation. Specifically, the DDA raw data were searched using the Pulsar engine against a human Swiss-Prot FASTA database (20,601 entries) for spectral library generation. Pulsar search parameters were set as follows: the enzyme was selected as trypsin/P; carbamidomethyl (C) was set to a fixed modification, while acetyl (Protein N-term) and oxidation (M) were set to variable modifications; ion types included b, y; other settings were set as default.

Differential proteins identified in TMT-based quantitative proteomics were filtered using the following settings for the assay panel generation: precursor *m/z*, 350–1500; precursor charge, 2–3; missed cleavages, 0; peptide length, 8–25; Only proteotypic; allowed modification, carbamidomethyl (C); fragment *m/z*, 300–1800; max fragment charge, 2; Ion types, b and y; min fragment length, 3; Top N fragments, 6. Furthermore, precursors with poor signal and peak shape were discarded. After retention time (RT) calibration, the RT window was set for ±5 min.

For PRM validation, full MS scans were acquired at a resolution of 60,000 at m/z 200 with an AGC of 4e5 and a maximum injection time of 20 ms. The MS/MS spectra were acquired at a resolution of 30,000 at m/z 200, an isolation window of 1, a collision energy of 30%, and an AGC target value of 2e5 with a maximum injection time of 80 ms. An unscheduled PRM was first performed to ensure reasonable retention times were achieved. The PRM raw files obtained were processed using SpectroDive v10.4. A q-value cutoff of 0.01 was set for peptide selection. All selected peaks were manually checked after automated peak detection by SpectroDive. Product ion signal peaks with significant interfering signal around the peak apex were excluded, and more than three transitions per precursor were selected for quantification in SpectroDive. In addition, to improve the confidence of peptide identification, selected peaks were also checked by dotp value in Skyline v21.2.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bioinformatic analysis

The data were processed using an in-house R (version 3.6.2) script. For differential protein analysis, a two-sample *t*-test was performed using $\log_2$-transformed protein abundance to compare two groups. A p value of 0.05 was considered as statistically significant (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). Specifically, a One-way ANOVA was used for comparison of three groups and Tukey's Honest Significant Difference (TukeyHSD) statistical test was performed for the post-hoc analysis. We implemented a receiver operating characteristic (ROC) curve analysis using the R package pROC for statistical assessment.[42]

For TMT-based proteomics, proteins with a p value <0.05 and a fold change >1.2 or <0.8 were considered as up or downregulated proteins, respectively. The differential proteins identified in each pairwise comparison were submitted to Metascape[43] for functional and protein-protein interaction (PPI) enrichment analyses using the Molecular Complex Detection (MCODE) clustering algorithm. Gene Set Variation Analysis (GSVA) was performed in the R package GSVA[44] and differential gene-sets (p value <0.05) were calculated by lmFit function in the R package limma.[45]

Ranked plasma proteins in TM vs. TI were analyzed using the GSEA (V4.1.0) software.[46] The enriched gene-sets with p value <0.05 and FDR <0.25 were visualized in Cytoscape,[47] 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.

For targeted proteomics, a normalization was performed to correct the sample loading amount and MS signal intensity using the total ionic chromatograph (TIC) signal of MS1.[48,49] Specifically, the abundance of each peptide of each sample was normalized with the TIC of the respective sample. Proteins with a p value <0.05 were considered as statistically significant differential proteins. Network interaction and functional enrichment analyses were performed by STRING.[50] The R package Boruta (version 7.0.0),[51] a feature selection algorithm that works as a wrapper around random forest,[52] was used for the feature selection. The program attempts to capture all the important and interesting features from a dataset with respect to an outcome variable.

After feature selection, the python Scikit-learn library[53] was applied to machine learning classification. To develop the diagnostic model using the PRM data, a support vector machine (SVM) classifier was employed for binary classification and OnevsRestclassifier using SVM was performed to deal with multi-class classification problems. First, the GridSearchCV function was used to find the optimal values for the SVC model hyperparameter (kernel, C, and gamma). GridSearchCV uses all the combinations of the predefined parameter values and evaluates a model's performance for each combination using k-fold (here, k = 5) cross-validation. SVM using linear kernel with C = 1 was selected as the final classification model, with the best cross-validation accuracy of 0.98.

Before modeling, we test the features again using the k-fold (here, k = 5) cross-validation approach. F1 score is a weighted average of precision and recall. The importance of these six features to the diagnostic model was confirmed by removing a feature at a time, which was evaluated using the micro F1 score. Finally, the samples were randomly separated into training/testing sets with 75% of samples as training and 25% as testing. After this, confusion matrix and ROC analysis were used to assess the performance of our classification model.