Overall Experimental Design

Protocol development for fractionation and SELDI-TOF MS analysis of plasma. EDTA- and heparin-plasma specimens obtained from 4 patients with SCD were denatured with urea-CHAPS and separated into six fractions by step pH elution following anion-exchange chromatography. Fractions were each applied at different dilutions (1:20, 1:10, 1:5, and 1:4) to four solid-phase chromatographic surface types (BioRad ProteinChips, formerly Ciphergen ProteinChips), and analyzed by SELDI-TOF MS. Qualitative comparisons across chromatographic surfaces and concentrations resulted in an optimized set of chromatographic surface types and plasma concentrations per fraction that yielded the most comprehensive panel of peaks with the least redundancy of peaks (Table S1).

Overview of Sample Processing. Plasma samples (standardization controls plus 55 patients) were denatured and fractionated as above through the ion exchange chromatography to as above reduce sample complexity (Fig. S2). Seven selected fractions were bound to specific affinity matrix chip surfaces as optimized above (total of 17 fraction-chip combinations) submitted to SELDI-TOF MS in duplicate, yielding 34 chip sets. Each chip was read twice, under separate conditions to optimize resolution of low and high mass/charge peptides, yielding 68 data sets for each of the 55 patients, or 3740 total spectra. These 3740 spectra were analyzed by the CiphergenExpress software package. Standardization controls were derived from pooled reference plasma from healthy subjects and used to monitor run to run consistency. Originally, samples from 56 patients were analyzed, but one was disqualified upon DNA sequencing diagnosis of sickle-β-thalassemia in a patient previously believed to have homozygous sickle cell anemia.

Plasma denaturation. Plasma samples were transferred into randomized wells in 96 well plates (20 μL/aliquot), denatured for 20 min in 6 M urea/2% CHAPS and diluted to a final urea concentration of 3 M. Samples were separated into six fractions by step pH elution (buffers of pH 3–9, and an organic isopropanol-based buffer) by anion exchange chromatography (BioRad).

ProteinChip application. Fractionated samples were applied using an automated system (Biomek2000, Beckman-Coulter), in duplicate, to solid-phase protein chip surfaces (BioRad ProteinChips) as per developed protocol. One spot per chip was allotted to a reference sample of plasma pooled from 30 control subjects processed similarly for experimental quality control. Two layers of matrix [50% sinapic acid in acetonitrile/trifluoroacetic acid (50%/0.5% v/v)] were applied over bound samples.

SELDI-TOF MS. Chips were read by a Protein Biological System II mass spectrometer (BioRad), and time-of-flight spectra were each generated by an average of 130 laser shots. Mass spectra were obtained using 2 different instrument settings to optimize detection of peaks within either the low m/z range (2–10 kDa) or high m/z range (5–50 kDa). Low m/z runs were set with a high mass of 50 kDa, optimized from 2 to 10 kDa. High m/z runs had a high mass of 200 kDa, optimized from 5 to 50 kDa. An all-in-one protein-peptide molecular mass calibrator (BioRad) was used to externally calibrate mass accuracy.

Spectrum processing and peak detection. CiphergenExpress™ 3.0 was used to process and detect peaks in the 3740 resulting spectra. For each of 34 data sets for each experimental and control specimen, the baseline was adjusted and spectra were normalized to the total ion current, excluding matrix peaks (*m/z* 0 kDa to 2 kDa). Mass range was 10 to 200 kDa for the high energy spectral output, and 2 to 20 kDa for the low energy spectra. Autodetection of peaks was performed with a first pass signal to noise (S/N) ratio of 3 and min peak threshold of 5% of all spectra. Cluster completion by second pass detection was set at a S/N of 2 with a cluster window of peak width 0.8 for low energy spectra and 2.0 for the high energy spectra. Acceptable experimental quality was judged by average coefficients of variation (%CV) of intensities <30% from pooled reference data sets. Percent CV ranged from 17–29%.

Statistical analysis. Univariate analysis was performed through p-value and ROC area calculation of individual peaks by CiphergenExpress™ 3.0 software. Multivariate analysis was performed by logistic regression (JMP 5.1) and by Random Forests classification/regression (Salford Systems, San Diego, CA).[1]

Peak identification. Ten peaks of interest were selected from the collective results of the Random Forest and stepwise logistical regression models for identification. Two plasma samples from each group were selected for analysis and identification of proteins, based on qualitative examination of SELDI-TOF spectra to represent high intensity of the peaks of interest.

SDS polyacrylamide gel electrophoresis resolution of proteins. In duplicate, plasma samples were depleted of albumin and IgG (Vivapure Anti-HSA/IgG spin columns, Vivascience), and concentrated by centrifugal filter devices (Centricon; Ultracel YM-3, MWCO 3000) to a final volume of approximately 60 µL each. NuPAGE LDS buffer (4×; Invitrogen) was added to the purified samples in a 1:4 ratio. Samples underwent denaturation at 95°C for 5 min. Denatured samples were run in triplicate on precast 4–12% Bis-tris polyacrylamide gels (NuPAGE; Invitrogen) in a sodium dodecyl sulfate buffer for 105 min at 110V. Gels were stained with Coomassie blue dye for 8–12 hr, then destained with distilled deionized water washes every 30 min for 2 hr.

Proteolysis. Coomassie blue stained bands were excised from the gels, diced into 1–2 mm^2 pieces, and destained with 25 mM NH_4HCO_3/50% acetonitrile. Gel pieces were completely desiccated in a vacuum centrifuge, and resuspended in 10 mM dithiothreitol/25 mM NH_4HCO_3. Gel samples underwent reduction for 1 hr at 56°C, before being alkylated with 55 mM iodoacetamide in a dark room for 45 min at room temperature. Gel samples were then rinsed for 10 min with 25 mM NH_4HCO_3, dehydrated with two washes of 25 mM NH_4HCO_3/50% acetonitrile, and redesiccated in the vacuum centrifuge. Samples were rehydrated in 20–30 µL trypsin (12.5 ng/µL in 25 mM NH_4HCO_3) at 4°C for 30 min. Volume of added trypsin solution was three times the volume of the dessicated gel pieces. To reduce the amount of self-digested trypsin peptides in the final samples, remaining trypsin solution was then removed from the rehydrated samples, and samples were washed briefly with 25 mM NH_4HCO_3. Samples were subsequently incubated at 37°C for 8–12 hr, kept hydrated in 50 µL NH_4HCO_3 (25 mM), to achieve proteolysis of gel-contained protein. The resulting aqueous extract was reserved. Peptides were extracted from the gel pieces by sonicating samples on ice for 5 min in 30 µL of a 5% formic acid/50% acetonitrile solution. This was repeated and supernatants were pooled with

the first aqueous extract. Volume was reduced to approximately 5–10 μL by vacuum centrifugation. Additional 0.1% formic acid (15 μL) was added to each sample for volume expansion.

Passive gel elution for SELDI confirmation. Bands excised from polyacrylamide gels were diced into 1- to 2-mm^2 cubes and washed with 50% methanol/10% acetic acid for 30 min on a vortex mixer. Gel slices were destained with 50% acetonitrile in ddH2O for 15 min, and dehydrated with 100% acetonitrile for an additional 15 min. Samples were desiccated in the vacuum centrifuge at 45°C for 3 min. Passive elution of protein from the gel pieces was achieved by vigorous shaking of dried gel samples in 50 μL 50% formic acid/25% acetonitrile/15% isopropanol for 2 hr at room temperature. Extracts (2 μL each) were applied to NP20 protein chip surfaces, air dried, and coated with two layers of 50% sinapic acid in 50% acetonitrile/0.5% trifluoroacetic acid. Eluted proteins were confirmed by SELDI-TOF analysis to be the same peaks identified from the screening analysis. Protein chips were read at the same settings used in biomarker discovery.

High-pressure liquid chromatography mass spectrometry (LC-MS/MS). Samples were purified using Vivaspin Vivapure MALDI-prep C18 microcolumns. LC-MS/MS analysis of tryptic peptides was carried out using an ion trap mass spectrometer (Deca XP plus, Thermo Electron, San Jose, CA) according to the procedure described previously.[2] Briefly, samples were injected, using an autosampler, and trapped in one of the two trapping columns (C18, Agilent, Santa Clara, CA). After washing with solvent A (0.1% formic acid), trapped peptides were eluted and fractionated through a reversed-phase PicoFrit column (BioBasic C18, 75 μm × 10 cm, tip = 15 μm, New Objective, Woburn, MA) using 0–60% solvent B (acetonitrile) in solvent A for 30 min (flow rate ≈200 nl/min). The PicoFrit column was placed directly in front of the ion transfer tube of LCQ DecaXP mass spectrometer equipped with a nanospray ionization source. The *m*/*z* ratios of peptides and their fragmented ions were recorded using a method which acquired three MS^2 scans following each full MS scan. The raw data files were searched against the Swiss-Prot human database by using the BioWorks program (Thermo Electron, San Jose, CA).

MALDI TOF-TOF. Digested and eluted samples were mixed with CHCA matrix (5 mg/ml in 50% ACN, 0.1% TFA) and spotted on steel targets. These were analyzed in a Bruker UltraFlex TOF-TOF (Billerica, MA) in positive ion reflectron mode over a range of 0.9-5 kDa. Prominent peptide peaks were chosen for fragmentation analysis using the LIFT module. Fragment peak analysis was conducted manually and peak lists were submitted to Mascot for comparison to the human database (Matrix Science, ver. 3/13/07).

REFERENCES

1. Breiman L. Random Forests. *Machine Learning*. 2001;45:5-32.
2. Li T, Evdokimov E, Shen RF, et al. Sumoylation of heterogeneous nuclear ribonucleoproteins, zinc finger proteins, and nuclear pore complex proteins: a proteomic analysis. *Proc Natl Acad Sci U S A*. 2004;101:8551-8556.