# Deeper plasma proteome coverage enables identification of novel biomarkers and classification of diseases

**Katrin Hartinger1, Andreas Schmidt2, Zehan Hu1, Claudia Martelli3, Katharina Limm1, Xaver Wurzenberger1, Sebastian Mueller4, Markus Lubeck2, and Nils A. Kulak1,**

[1] PreOmics GmbH, Martinsried, Germany
[2] Bruker Daltonics GmbH & Co. KG, Bremen, Germany
[3] Bruker Switzerland AG, Fällanden, Switzerland.
[4] Biognosys AG, Zuerich, Switzerland

## Introduction

Blood is one of the least invasive biopsies and a valuable specimen for clinical research and patient health monitoring. Since almost all tissues are sustained by the constant blood flow and proteins are constantly being actively secreted or leaked into the blood, plasma provides comprehensive information about health or disease state. However, access to proteome information is limited by the highly dynamic nature of protein abundance in plasma, which spans approximately more than 10 orders of magnitude and with only 22 proteins accounting for 99% of the whole protein content. To address this challenge, we developed a novel workflow for LC-MS-based plasma proteomics that enriches low abundant proteins and enables an improved coverage of the plasma proteome.

## Methods

Plasma samples from a clinical cohort of non-small cell lung cancer patients and healthy donors were obtained from Biognosys. Initially, samples were prepared with the iST sample preparation kit (PreOmics). Additionally, 10 µl of plasma were prepared with the ENRICH-iST kit and 300 ng of peptides were analyzed on the TimsTOF HT mass spectrometer coupled to a nanoElute LC system. A 30 min gradient from 3% to 30% ACN was employed for peptides separation. Data were acquired with a dia-PASEF acquisition program covering precursors in the range from 400 to 1000 m/z. Data were analyzed in Spectronaut 17.4 in directDIA+ mode against a human swissprot database as well as against a spectral library generated by high pH-RP HPLC.

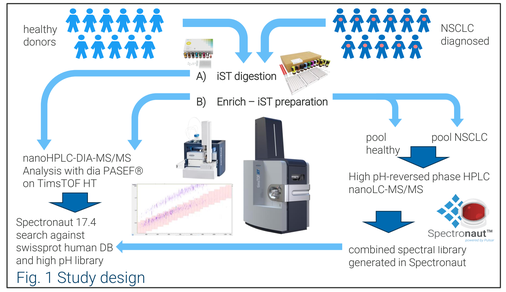


Fig. 1 Study design

## The ENRICH-iST workflow

The ENRICH-iST workflow is optimized to enrich low abundant proteins in biofluids to the surface of magnetic beads. For plasma analysis, 10-20 µls of plasma are mixed with the beads in an optimized sample buffer and incubated for 30 min to bind proteins. After removal of unbound proteins, beads with aggregated proteins are subjected to tryptic digestion using the BCT-iST sample preparation workflow.

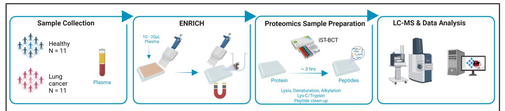


Fig. 2 The ENRICH-iST workflow for manual preparation of plasma sample in 96-well format

## Results library-free DIA vs. high-pH RP HPLC spectral library

More than 1450 protein were identified in ENRICH-iST samples in comparison to 650 protein IDs in neat plasma, increasing the plasma proteome by more than 2x. Moreover, using a sample specific library (~2200 entries) from high-pH fractionation of sample pools, more than 2000 proteins were identified and almost 1500 proteins were quantified in all samples of the dataset (3A). The increase is achieved by reduction of the dynamic range of protein abundance, where the content of high abundant proteins is reduced, which enables detection of proteins in the lower abundance range (3B). Proteins identified in neat plasma generally show an increase of precursors in ENRICH-iST, permitting more accurate quantitation.

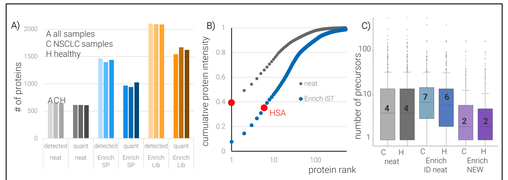


Fig. 3 Overview of the Dataset. A) Number of proteins identified and quantified in neat plasma analysis and ENRICH-iST samples. B) Cumulative relative protein abundance in healthy donors in neat plasma and ENRICH-iST, plotted for the 550 overlapping protein IDs from DIA+. SA is highlighted in red C) Identified precursors per protein, the protein IDs for ENRICH-iST have been split into a group that overlaps with neat plasma (blue) and newly identified proteins (lilac).

## Comparison of neat plasma and ENRICH-iST samples

From neat plasma samples, only a small number of significant proteins (q-value 0.05) could be retrieved (Fig 4A). Most were also significantly higher after the ENRICH-iST preparation to be significantly altered (Fig. 4B). Interestingly, S100A8 and S100A9, two of the most differentially abundant hits in neat plasma analysis, have almost identical regulation in ENRICH-iST, corroborating our assumption, that the enrichment step can preserve the inherent protein abundance ratios. Overall, 20 proteins were significantly altered between the study groups in the ENRICH-iST dataset.

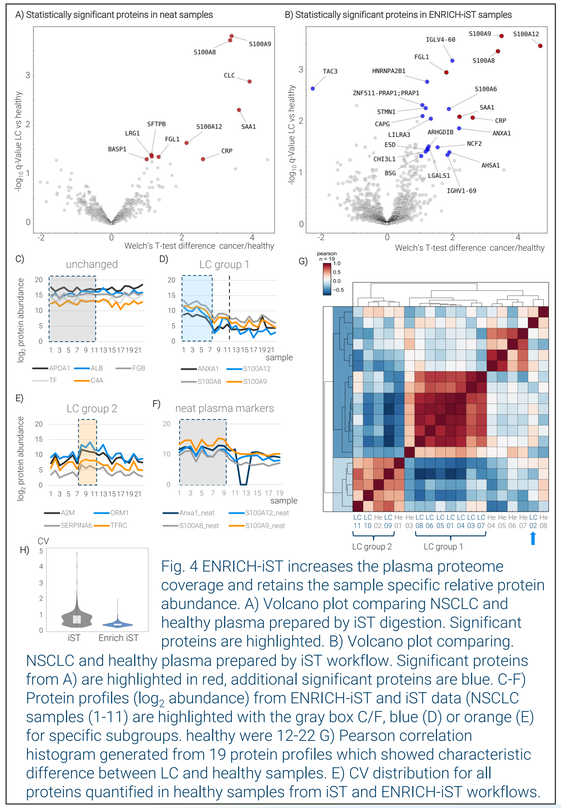


Fig. 4 ENRICH-iST increases the plasma proteome coverage and retains the sample specific relative protein abundance. A) Volcano plot comparing NSCLC and healthy plasma prepared by iST digestion. Significant proteins are highlighted. B) Volcano plot comparing. NSCLC and healthy plasma prepared by iST workflow. Significant proteins from A) are highlighted in red, additional significant proteins are blue. C-F) Protein profiles ($log_2$ abundance) from ENRICH-iST and iST data (NSCLC samples (1-11) are highlighted with the gray box C/F, blue (D) or orange (E) for specific subgroups. healthy were 12-22 G) Pearson correlation histogram generated from 19 protein profiles which showed characteristic difference between LC and healthy samples. E) CV distribution for all proteins quantified in healthy samples from iST and ENRICH-iST workflows.

Plotting normalized protein abundances for all cancer samples (Fig. 4C-F) demonstrates, that common plasma proteins are not changed by ENRICH-iST. However, we also observe that significant lung cancer markers are not upregulated in all patient plasma samples. Based on the current data, we can distinguish 2 patient groups. Characteristic protein for group 1 are S100A8, S100A9, S100A12, and Anxa1, of which the first are common markers for inflammation and immune response (Fig. 4D). These proteins are unaltered in group 2, for which TFRC, ORM1, A2M, and SerpinA6 had the highest difference (Fig. 4E). Since this group comprises only 4 samples, none of the hits are found to be upregulated in volcano plots. All markers for group 1 showed a similar behavior in the neat plasma data (Fig. 4F). Based on 23 most altered protein profiles, we can distinguish most patient groups and healthy donors (Fig 4G), however, we lack any additional patient data that would allow further refinement. Calculating the coefficient of variation for all proteins in the healthy group, demonstrates a high reproducibility of protein abundance with a median of 20% (Fig. 4H).

## Summary

ENRICH is a fast and untargeted protein capture technology, that reduces the dynamic range of protein abundance in biofluids and requires low starting volumes (10-20 ul). With ENRICH-iST, we observed more than 2000 proteins in human plasma samples and quantified more than 1000 reliably. Only from ENRICH-iST, we were able to determine 2 different patient groups in the lung cancer study.

## Citations

Plots have been created with instant clue (vs 11.3, instantclue.uni-koeln.de). Figure 2 was created with Biorender.

Data from neat plasma analysis were reanalyzed from App. Note LCMS 197, Bruker Daltonics 2022.

## Conclusion

- ENRICH-iST efficiently reduces the dynamic range in plasma increasing the number of detected proteins by at least 2x
- Deeper plasma proteome coverage and more significant proteins are observed with ENRICH-iST at high reproducibility (CV ~20%)
- Sample inherent protein abundances are retained upon ENRICH-iST sample preparation

Technology