HUPO CONFERENCE 2023 - PP03.100

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

Katrin Hartinger1, Christopher Below4, Andreas Schmidt2, Zehan Hu1, Claudia Martelli3, Katharina Limm1, Xaver Wurzenberger1, Verena Tellstroem2, Roland Bruderer4, Markus Lubeck2, Lukas Reiter4, Nils A. Kulak1, Jonathan Moss5

[1] PreOmics GmbH, Martinsried, Germany, [2] Bruker Daltonics GmbH & Co. KG, Bremen, Germany, [3] Bruker Switzerland AG, Fällanden, Switzerland, [4] Biognosys AG, Zürich, Switzerland, [5] Bruker Pty Ltd. Melbourne Australia

Introduction

Blood is one of the least invasive biopsies and a valuable specimen for clinical research. 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 challenging because the highly dynamic nature of protein abundance in plasma, which spans more than 10 orders of magnitude and with only 22 proteins accounting for>95% 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.

Plasma cohorts preparation by the ENRICH-iST

The ENRICH-iST technology is optimized to enrich low abundant proteins in plasma/serum to the surface of magnetic beads (EN-beads). For plasma ENRICH-iST workflow, 10-20 µls of plasma are mixed with the EN-beads in an optimized buffer and incubated for 30 min to bind proteins, unbound proteins are removed and EN-beads with proteins are subjected to tryptic digestion using the iST-BCT sample preparation workflow.

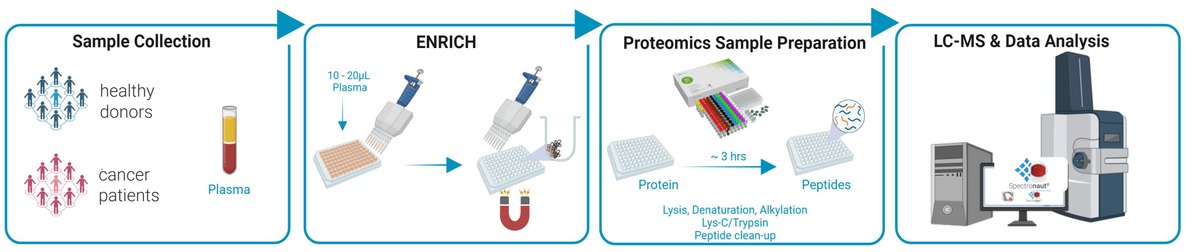


Fig. 1 Complete ENRICH-iST workflow for proteomics preparation of plasma sample in 96-well format

Lung cancer cohort analysis set-up

Plasma samples from a clinical cohort of non-small cell lung cancer (NSCLC) patients and healthy donors were obtained from Biognosys. Samples of 2 µl starting volume were prepared with the iST sample preparation kit (neat plasma, PreOmics) and 10 µl with the ENRICH-iST kit (ENRICH plasma, PreOmics). 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 Swiss-Prot database as well as against a spectral library generated by high pH-RP HPLC.

Results lung cancer cohort

More than 1450 protein groups were identified in ENRICH plasma compared to 650 protein group IDs in neat plasma, increasing the plasma proteome by more than 2-fold. Moreover, using a sample specific library (~2200 entries) from high-pH fractionation of plasma pools, more than 2000 proteins were identified and almost 1500 proteins were quantified in all samples (2A). The increased protein coverage is achieved by reducing the dynamic range of protein abundance, where the content of high abundant proteins is reduced, enabling detection of proteins in the lower abundance range while maintaining excellent quantitation with a median CV ~20% over all healthy donors (2BC). The overlap of significantly changed proteins between neat and enriched plasma is more than 80% (blue, 6/8), however we observed a 3fold increase in significant proteins upon enrichment as the green markers (2D) demonstrated. Selecting 21 protein profile from the enriched dataset, we are able to determine one patient group reliably from the healthy donors and from a second patient group that shows a more diverse pattern of protein abundances (2E)

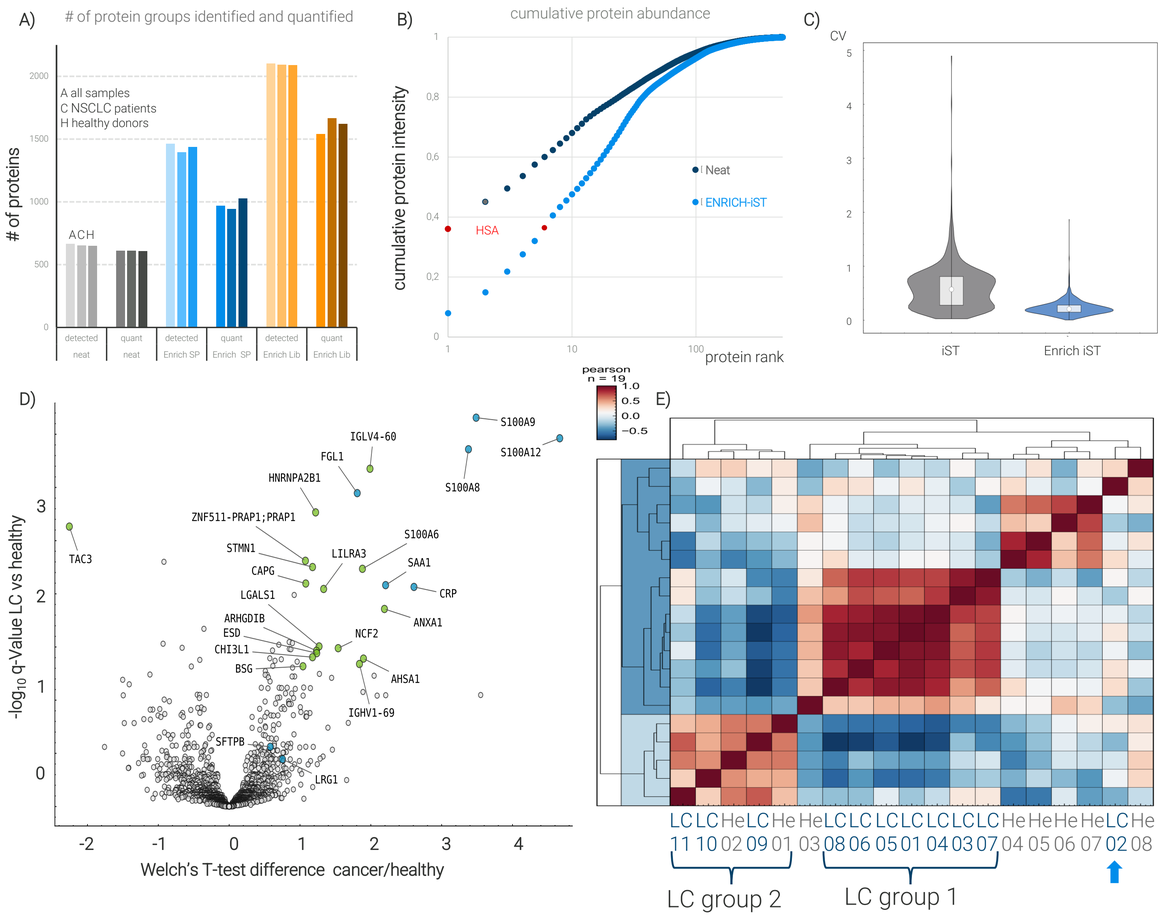


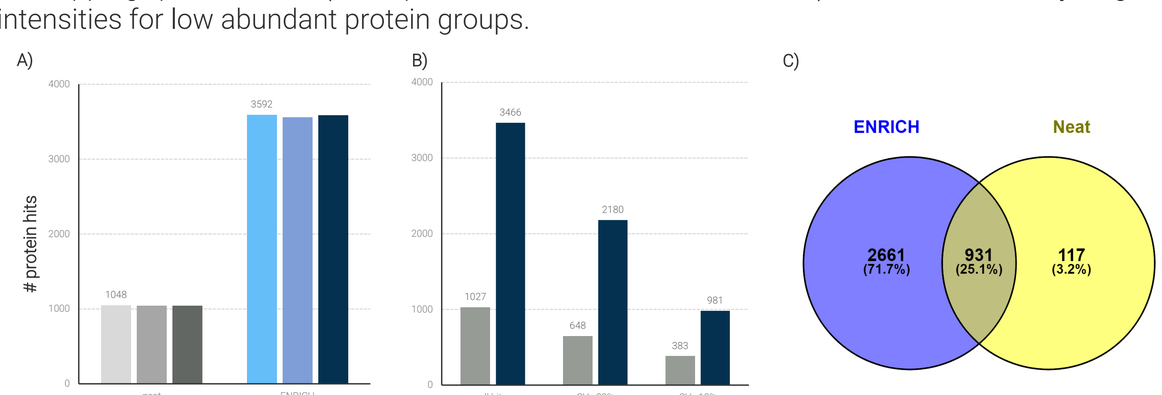
Fig. 2 In depth analysis of lung cancer cohort
A) Number of proteins identified and quantified in neat plasma analysis and ENRICH plasma. B) Cumulative relative protein abundance in healthy donors in neat plasma and ENRICH plasma, plotted for the 550 overlapping protein IDs from directDIA+. Human serum albumin (HSA) is highlighted in red. C) CV distribution for all proteins quantified in healthy donors processed by iST and ENRICH-iST workflows. D) 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). E) Patient stratification based on protein abundance values of 21 significantly changed markers provides reliable identification of LC group 1 (center) from LC group2 and healthy. LC group 2 has a more diverse protein abundance pattern.

Colorectal cancer cohort analysis set-up

Biognosys provided plasma samples of 6 colorectal cancer patients and 6 age-matched healthy donors. Three additional pooled plasma samples (Sigma) were added. Samples for neat plasma were prepared at Biognosys, ENRICH plasma was perpared at PreOmics. All samples (0.8 ug/injection) were analyzed with a 60 min gradient on an Easy-nanoLC equipped with a IonOpticks Aurora column (reversed-phase C18, 25cmx75um). The dia-PASEF acquisition program covered precursors from 325-1150 m/z with 25Da windows in 11 tims ramps. Data were analyzed in Spectronaut 18.2 in directDIA+ mode against a human Swiss-Prot database.

Findings from the colorectal cancer cohort

Plasma samples prepared with the ENRICH-iST technology identified almost 3600 protein groups over all samples, representing a ~3.5-fold increase over the comparable neat plasma dataset. A similar number of protein hits was observed in cancer and healthy donors showing a high reproducibility of the ENRICH-iST sample preparation (Fig 3A). From all observed 3466 protein groups with sufficient quantitative datapoints, 2180 were quantified with a CV < 20% from all donors representing a robust protein quantitation (3B). Neat and ENRICH plasma datasets showed an 89% overlap in protein IDs (2C). Interestingly, both datasets cover a similar abundance range of proteins covering about 5 orders of magnitude, however 3.5 x more proteins were identified after enrichment, demonstrating an efficient reduction of the dynamic range that allows deep proteome coverage (2D). Principle component analysis was performed on all proteins and all samples. Controls of plasma pools (yellow dots) cluster tightly in the PCA plot, whereas samples from cancer (dark lilac) and healthy donors (turquoise), show a wider spread demonstrating substantial individual differences of protein abundance in plasma (2E). Similarly, datasets from neat and ENRICH plasma showed only a moderate correlation for overlapping protein IDs (2F, Spearman correlation, rho =0.63) with considerably higher intensities for low abundant protein groups.



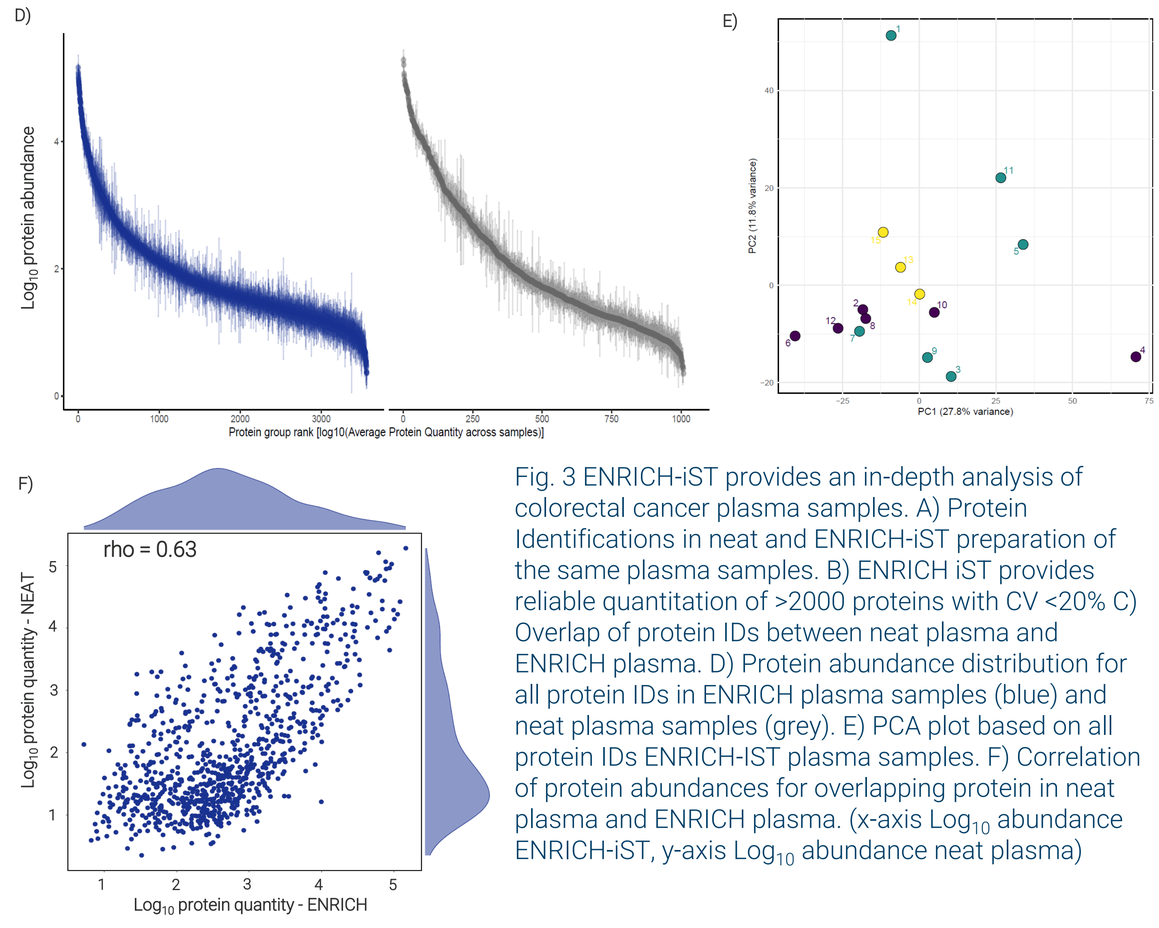


Fig. 3 ENRICH-iST provides an in-depth analysis of colorectal cancer plasma samples. A) Protein Identifications in neat and ENRICH-iST preparation of the same plasma samples. B) ENRICH iST provides reliable quantitation of >2000 proteins with CV <20% C) Overlap of protein IDs between neat plasma and ENRICH plasma. D) Protein abundance distribution for all protein IDs in ENRICH plasma samples (blue) and neat plasma samples (grey). E) PCA plot based on all protein IDs ENRICH-IST plasma samples. F) Correlation of protein abundances for overlapping protein in neat plasma and ENRICH plasma. (x-axis Log_{10} abundance ENRICH-iST, y-axis Log_{10} abundance neat plasma)

Summary

ENRICH-iST is a fast and untargeted protein capture technology, that reduces the dynamic range of protein abundance in plasma/serum and requires low starting volumes (10-20 µl). With ENRICH-iST, we observed between 2000 to 3500 proteins in small scale studies of lung cancer and colorectal cancer cohorts. We quantified more than 1500 protein groups reliably in both studies and could demonstrate that differences between groups are contained upon ENRICH-iST preparation. Moreover, we distinguished 2 cancer patient groups in the NSCLC study.

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

© 2023 Bruker

For Research Use Only. Not for use in clinical diagnostic procedures.

Innovation with Integrity