Multicenter evaluation of label-free quantification in human plasma on a high dynamic range benchmark set

Info

PUBLISHED DATE
2025-10-02

JOURNAL
Nature Communications

LINK
https://doi.org/10.1038/s41467-025-64501-z

Authors

Ute Distler Han Byul Yoo Oliver Kardell Dana Hein Malte Sielaff Marian Scherer Anna M. Jozefowicz Christian Leps David Gomez-Zepeda Christine von Toerne Juliane Merl-Pham Teresa K. Barth Johanna Tüshaus Pieter Giesbertz Torsten Müller Georg Kliewer Karim Aljakouch Barbara Helm Henry Unger Dario L. Frey **Dominic Helm** Luisa Schwarzmüller Oliver Popp Di Qin Susanne I. Wudy Ludwig Roman Sinn Julia Mergner Christina Ludwig Axel Imhof Bernhard Kuster Stefan F. Lichtenthaler Jeroen Krijgsveld **Ursula Klingmüller** Philipp Mertins Fabian Coscia Markus Ralser Michael Mülleder Stefanie M. Hauck Stefan Tenzer

Abstract

Human plasma is routinely collected during clinical care and constitutes a rich source of biomarkers for diagnostics and patient stratification. Liquid chromatography-mass spectrometry (LC-MS)-based proteomics is a key method for plasma biomarker discovery, but the high dynamic range of plasma proteins poses significant challenges for MS analysis and data processing. To benchmark the quantitative performance of neat plasma analysis, we introduce a multispecies sample set based on a human tryptic plasma digest containing varying low level spike-ins of yeast and E. coli tryptic proteome digests, termed PYE. By analysing the sample set on state-of-the-art LC-MS platforms across twelve different sites in data-dependent (DDA) and data-independent acquisition (DIA) modes, we provide a data resource comprising a total of 1116 individual LC-MS runs. Centralized data analysis shows that DIA methods outperform DDA-based approaches regarding identifications, data completeness, accuracy, and precision. DIA achieves excellent technical reproducibility, as demonstrated by coefficients of variation (CVs) between 3.3% and 9.8% at protein level. Comparative analysis of different setups clearly shows a high overlap in identified proteins and proves that accurate and precise quantitative measurements are feasible across multiple sites, even in a complex matrix such as plasma, using state-of-the-art instrumentation. The collected dataset, including the PYE sample set and strategy presented, serves as a valuable resource for optimizing the accuracy and reproducibility of LC-MS and bioinformatic workflows for clinical plasma proteome analysis.