A High-Throughput and Automated Depletion Pipeline that Enables Precise and Reproducible Deep Profiling of Human Plasma

Marco Tognetti, Sebastian Müller, Edina Kishazi, Roland Bruderer, Lukas Reiter

Introduction

Blood is the most frequently used sample in diagnostic processes, and plasma is its liquid component. Plasma contains molecular clues coming from all the tissues that make up an organism, thus enabling the holistic analysis of an individual's health state. The major challenge in plasma protein analysis is the >10 orders of magnitude dynamic range and the strong prevalence of few, very abundant proteins. Depletion of the most abundant proteins increases coverage of the plasma proteome. Here, we compare the precision and quantitative accuracy effects of depletion and show a high-throughput, automated depletion pipeline that enables comprehensive quantification of the plasma proteome on a biological cohort while assessing all levels of variability.

Methods

Plasma samples (20 ul) were filtered, diluted, and depleted using the Multiple Affinity Removal Column Human 14 (Agilent) on an automated HPLC system. Samples were reduced, alkylated, digested with trypsin and LsyC, cleaned-up, dried down, dissolved in LC solvent A with the iRT kit (Biognosys) spiked in, peptide concentrations determined, and 1 ug analyzed with a 2-hour gradient on an Orbitrap mass spectrometer using data independent acquisition (DIA).

The controlled quantitative experiment (CQE) sample set was generated from 20 healthy human plasma samples spiked with fixed ratios of Escherichia coli and Saccharomyces cerevisiae, leading to a 1:2 and 4:3-nominal ratio. The resulting 40 samples were processed either with or without depletion. Data analysis was performed using SpectroMine and Spectonaut (Biognosys).

Preliminary Data

In the CQE, depletion led to a significant increase in the number of identifications: from 572 to an average of 1,471 proteins (+257%, n=40, p-value=1e-98). Given the controlled nature of the experiment, we could assess the conservation of the quantities post depletion and compare the quality of the candidate lists created by statistical testing. Overall, depletion led to an increase of 362% in true hits (from 170 to 615) while controlling the actual error rate (<1%). Furthermore, we

found that only 3 proteins were co-depleted. In summary, the automated depletion more than tripled the number of proteins identified and the number of true hits while maintaining quantitative accuracy in the CQE and reducing manual work to about half a day per 96 samples.

We analyzed a cohort comprising 15 healthy, 14 lung cancer, and 6 pancreatic cancer plasma samples using two depletion column batches to research large cohorts' feasibility. Altogether, we processed 70 samples (plus 14 quality controls) within two weeks and quantified 1,650 proteins (825 proteins/hour), of which 1,413 in at least 50% of the runs. Furthermore, based on quality control samples, we could characterize variance introduced on each level: injection (coefficient of variation (CV)=9%), digestion (CV=11%), depletion (CV=12%), and column (CV=16%), all much lower than the inter-individual variability (CV>50%). Notably, the independent depletion on two depletion columns results in highly overlapping biological findings even when conditions are run separately (70% candidate conservation) and 88% of the sample pairs of the two depletions co-clustered together despite their independent processing. Overall, the introduced variance is much lower than the intra-individual variability, even in the worst-case scenario of an unbalanced study design. We envision that our semi-automated plasma depletion pipeline, in conjunction with next generation acquisition methods, will enable the unbiased and reproducible quantification of +3,000 proteins across very large cohorts.

Novel Aspect

Firstly, we demonstrate the substantial discovery benefits following depletion. Secondly, we introduce a highly reproducible and automated plasma depletion pipeline.