Untargeted LC-MS Metabolomics Screening of 500,000 Serum Samples Identifies Novel Biomarkers for Cancer, Cardiovascular, & Liver Diseases

Boris Sarvin[1]; Mark Vernik[1]; Avi Shoshan[1]; Eldad Kepten[1]; Carmel Shor[1]; Ori Kronfeld[1]; Avishai Gavish[1]; Shira Shaham-Niv[1]; Tomer Shlomi[1,2]
[1]MetaSight Diagnostics LTD, Rehovot, Israel \ [2]Faculty of Biology and Computer Science, Technion—Israel Institute of Technology, Haifa, Israel
boris@metasightdx.com / office@metasightdx.com

Introduction

- Untargeted metabolomics is widely used for developing disease diagnostics. It is commonly applied to identify novel blood biomarkers via small-scale retrospective case-control studies, which typically suffer from technical confounders that hinder reproducibility across studies. Hence, growing efforts are made to shift biomarker research to large-scale prospective-cohort trials. Accordingly, Buergel et al. have recently analyzed ~120,000 samples from the UK-Biobank with NMR-based metabolomics (~150 metabolites/sample).
- Here, we report the first population-scale untargeted metabolomics screening of 500,000 patients using LC-MS - the Israeli multi-OMICS Serum Screening (IMOSS-500K) study. We describe developing a cost-effective and robust high-throughput LC-MS metabolomics method, quality control techniques, and discovery of metabolic signatures predictive of diverse clinical conditions.

Methods

- Serum samples (~500,000) collected from hundreds of clinics in Israel were obtained from an Israeli HMO between July 2021 and April 2022.
- We established a semi-automated system for high-throughput liquid-liquid metabolite extraction from serum via a robotic liquid handling system; utilizing organic solvents to precipitate protein prior to LC-MS analysis. Rapid separation was performed using a dual-channel LC system equipped with amide columns by ballistic gradient with a high flow rate.
- A proprietary computational pipeline was developed to analyze the unparalleled amount of raw LC-MS data.

High Throughput Analysis

LC separation in one column within sub-minute of MS acquisition time, in parallel to washing and equilibration of the second column, provide an overall sample running time of 55 seconds and supporting the analysis of up to ~1,200 samples per system per day.

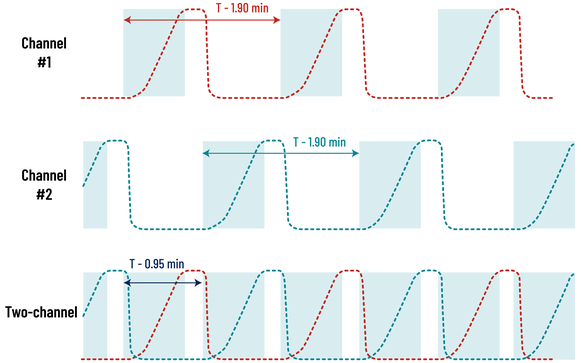

Figure 1. Scheme of high throughput analysis with synchronization of MS acquisition window (highlighted in light blue) and active separation/ system preparation for two LC channels.

Analytical Performance

Enhanced Sensitivity

Ultranarrow LC peaks enable not only to maintain comparable chromatographic peak capacity to traditional methods but also increase metabolite SNR (~20% compared to a traditional LC method of 25 mins).

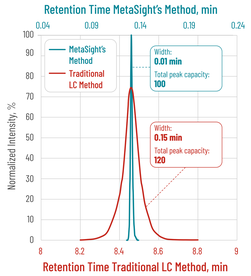

Figure 2. Enhanced sensitivity of MetaSight's method due to narrow chromatographic peaks and similar peak capacity to traditional LC method

Reliable reproducibility

High scan rate enables accurate quantification with an average ion FWHM of ~1.2 s, with ~18 scans per peak. A pooled serum QC sample was run every 10 samples across the entire study to support retention time and intensity normalization.

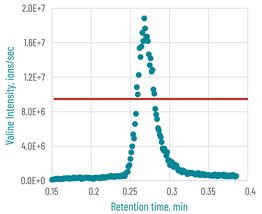

Figure 3. Accurate quantification achieved by a high number of scans per chromatographic peak.

Overall detected ions/annotations

We quantify the concentration of > 6,000 metabolite ions (RSD < 30%)

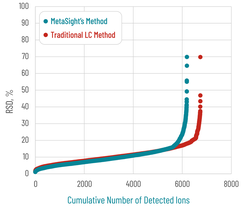

Figure 4. RSD Cumulative distribution of the detected ions for MetaSight's method versus traditional LC method.

We annotated 360 metabolites with high confidence (chemical standards or MS/MS fragmentation) and another 1150 metabolites based on accurate mass and isotopic measurements.

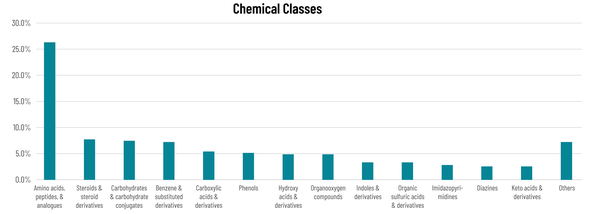

Figure 5. Distribution of chemical classes of detected metabolites.

IMOSS-500K Metabolomics Data Predicts Results of Numerous Clinical Lab Tests

- We examined the potential of the derived metabolomics data to predict the outcome of multiple blood tests conducted on the same day by the collaborating healthcare organization.
- Our measured glucose and creatinine concentrations for > 200,000 patients strongly correlated with those obtained by the clinical laboratory (Pearson R > 0.93).
- Leveraging machine learning, the metabolomics data successfully predicted out-of-range values for an additional 117 clinical tests (AUC > 0.8), including protein biomarkers, electrolytes, complete blood count, and enzyme activities.

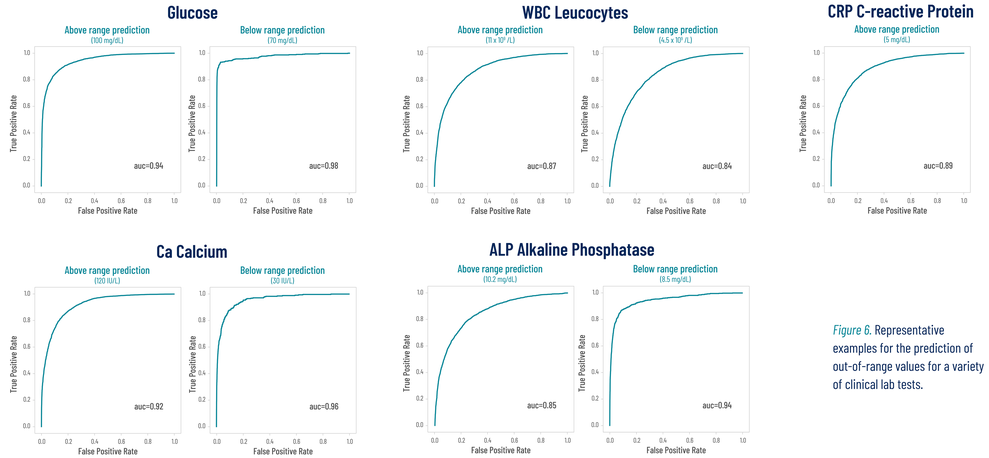

Figure 6. Representative examples for the prediction of out-of-range values for a variety of clinical lab tests.

Preliminary Metabolomics Signatures for Early Disease Diagnosis

- We identified novel robust signatures for diagnosing diverse diseases by integrating the metabolomics data with de-identified electronic health records (EHR).
- The analysis utilized samples from hundreds of patients collected prior to diagnosis, overcoming potential bias by treatment or change in lifestyle; and considering controls with diverse co-morbidities collected in a similar manner.
- Identified a novel signature for diagnosing colon cancer, congestive heart failure, and liver cirrhosis.

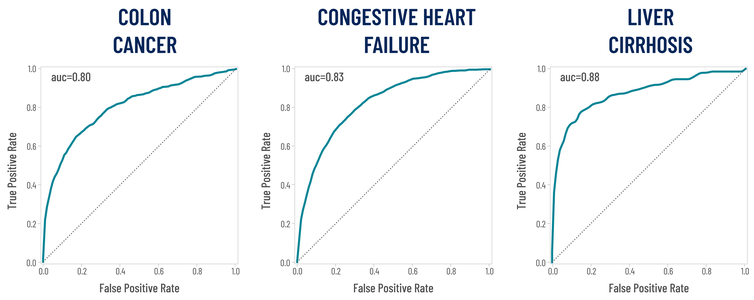

Figure 7. Diagnostic performance of identified metabolomics signatures