# Accelerating biomarker discovery in clinical cancer cohorts using volumetric absorptive microsampling (VAMS) devices and high-throughput mass spectrometry

N. Lucas[1,2,3]; C. Hill[1,3]; D. Pascovici[4]; R. McMahon[1,3]; B. Herbert[1,3]; E. Karsten[1,3]

1. Sangui Bio Pty Ltd, Sydney, NSW Australia; 2. The University of Sydney, Sydney NSW Australia; 3. Kolling Institute, Royal North Shore Hospital, St Leonards, NSW Australia; 4. CSIRO, Sydney, NSW, Australia

## Introduction

Volumetric absorptive microsampling devices (VAMS) allow patients to collect blood via fingerprick sampling at home. This approach enables collection of frequent longitudinal samples that can be used for health surveillance of disease, including cancer, for early diagnostics, ongoing monitoring, and recurrence. We have demonstrated that VAMS devices aid sample preparation for proteomics, facilitating the depletion of high abundance proteins via washing, prior to digestion. Reducing the dynamic range enabled the detection of more than 3000 proteins using a mid-throughput mass spectrometry (MS) method running 18 samples per day (SPD). Using a clinical cohort of non-small cell lung cancer (NSCLC), we compared the use of both mid- and high-throughput methods to identify relevant biomarkers and differentiate healthy from disease patients.

## Methods

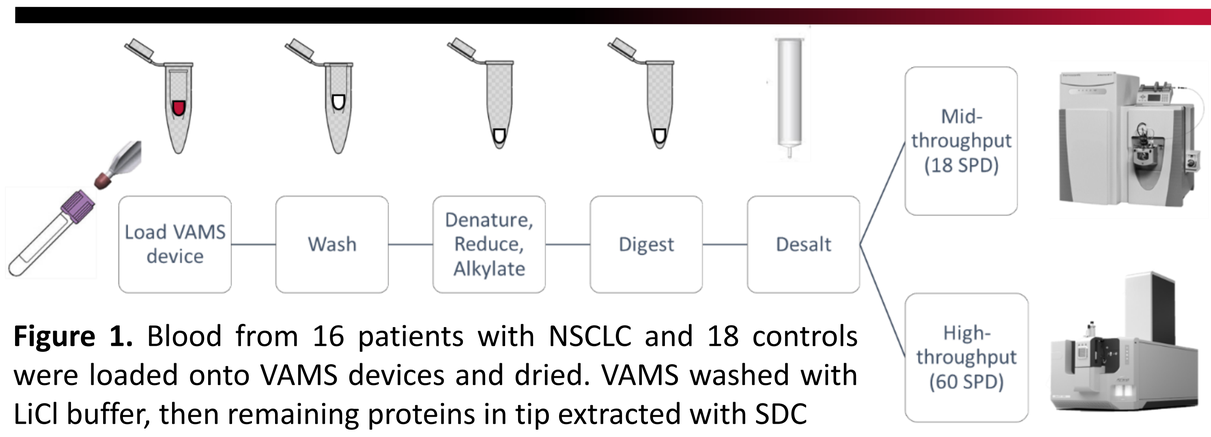



**Figure 1.** Blood from 16 patients with NSCLC and 18 controls were loaded onto VAMS devices and dried. VAMS washed with LiCl buffer, then remaining proteins in tip extracted with SDC lysis buffer, reduced, alkylated and digested with trypsin. Two MS methods used: mid-throughput 18 SPD (QEHFX, Thermo) and high-throughput 60 SPD (7600 Zeno-TOF, SCIEX). Data were analysed using a combination of DIA-NN, Mass Dynamics and ProMor.

## Results and Discussion

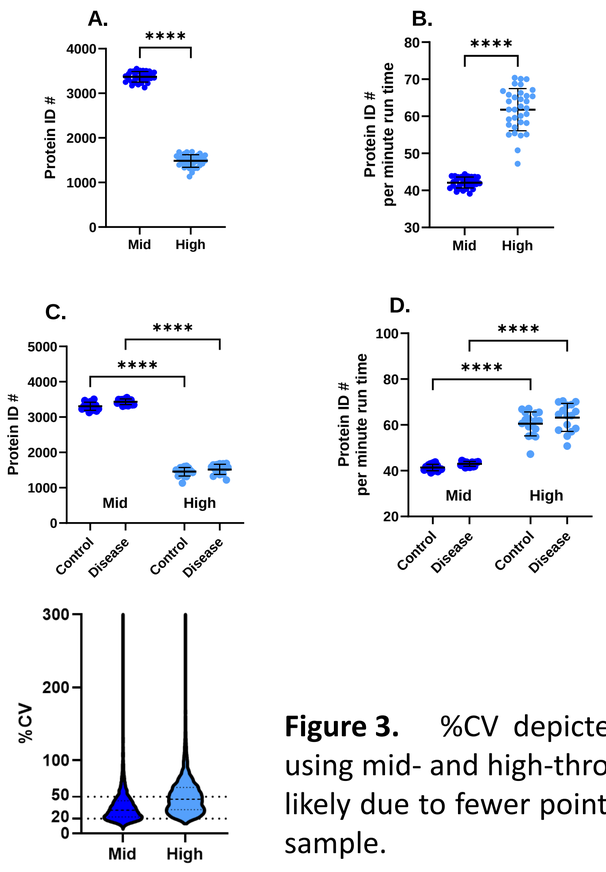



**Figure 2.** Number of protein IDs for samples analysed with a mid-throughput method (compared to a high-throughput method for (A) total acquisition time, (B) per minute of acquisition time, (C) total acquisition time stratified by disease group, or (D) per minute of acquisition time stratified by disease group. Data are mean ± SD, data is significantly different (*) if $p < 0.05$.

With the use of a longer gradient and LC column for the mid-throughput method, significantly more proteins were identified in comparison to the high-throughput method (mean and SD of 3370 ± 117 and 1483 ± 137, respectively).




**Figure 3.** %CV depicted as (A) violin plots of all samples analysed using mid- and high-throughput methods. The higher variability is most likely due to fewer points across a peak and high dynamic range of the sample.

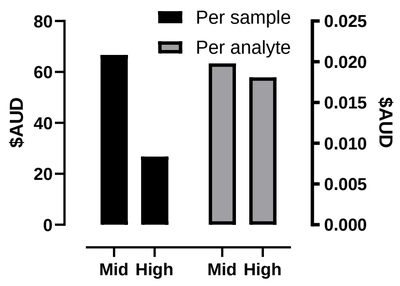
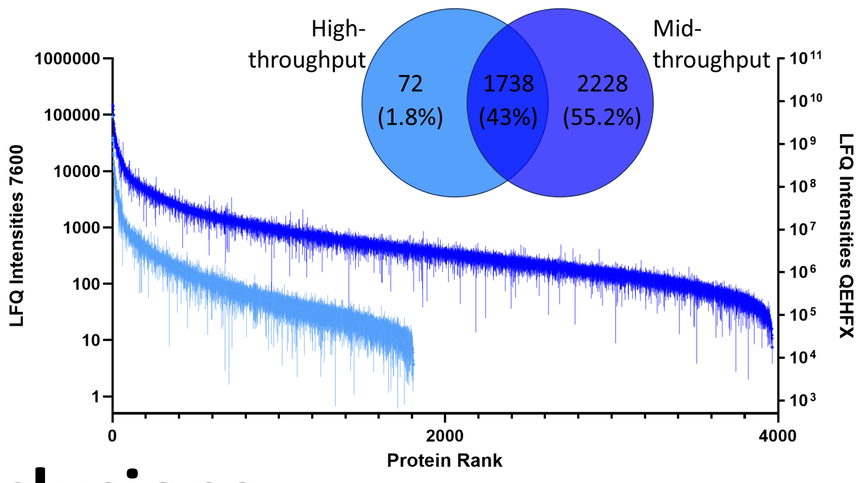



**Figure 4.** Cost comparison in $AUD between mid- and high-throughput proteomic analysis per sample (left y-axis) and per analyte per sample (right y-axis).

**Figure 5.** Dynamic range of LFQ intensities of ranked proteins analysed using both high-throughput methods (left y-axis) and mid-throughput methods (right y-axis) blood samples (n = 34). Data presented as mean ± SD. Overlap in detected proteins illustrated by Venn Diagrams. Washing of VAMS tips significantly improves dynamic range over conventional blood preparation methods as seen with mid-throughput method.



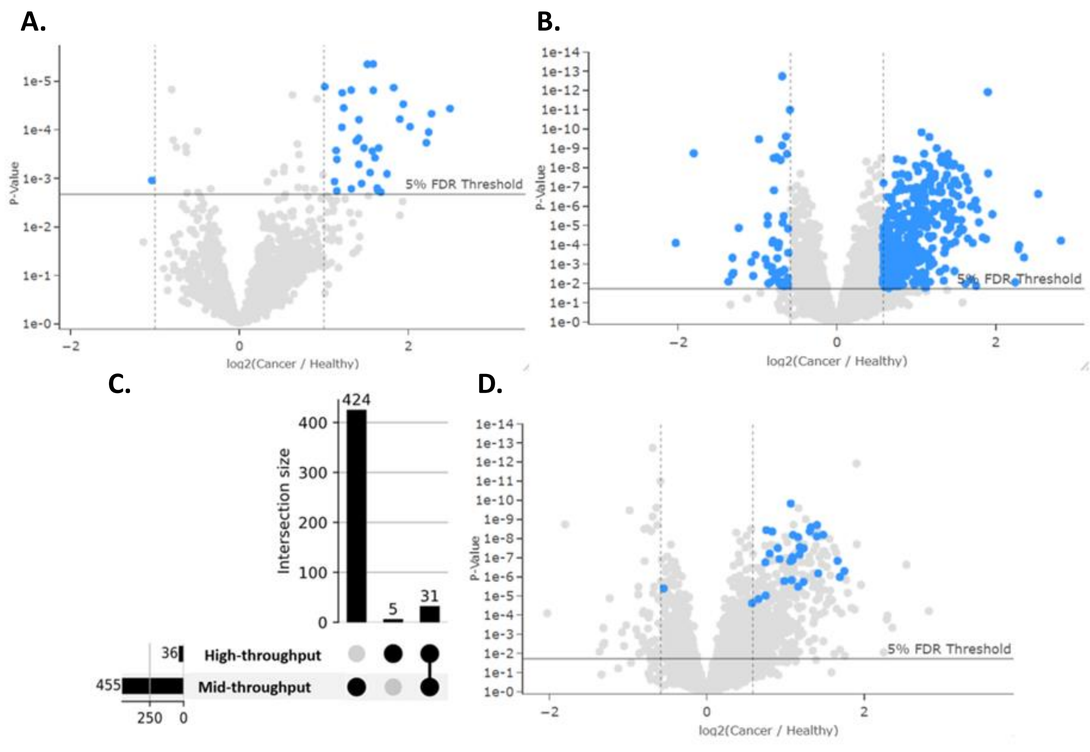




**Figure 6.** Differentially expressed proteins between controls and disease above the threshold of 5% FDR and fold change ratio of 1.5 (blue dots) as expressed by (A) a volcano plot of results using high-throughput methods, (B) a volcano plot of results using mid-throughput methods, (C) an upset plot of the overlap of proteins identified in each analysis, and (D) a volcano plot of the mid-throughput analysis with the overlapping proteins from both analyses highlighted in blue. 36 (high-throughput) and 455 (mid-throughput) differentially expressed proteins were discovered for each method. There were 31/36 of these identified proteins that overlapped.

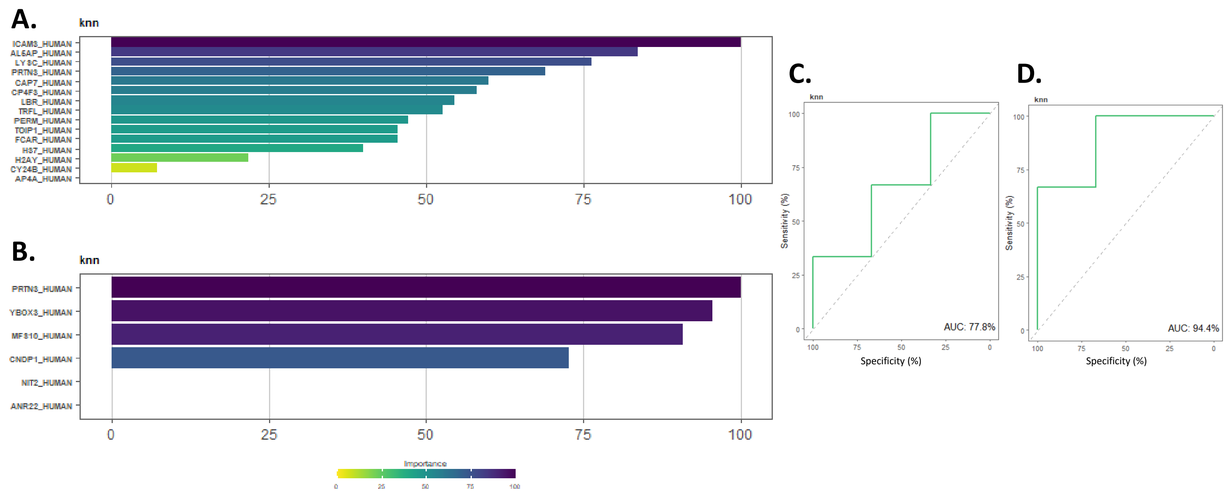



**Figure 7.** Fourteen-marker prediction model for the high-throughput method (A) Receiver operating characteristic curve (ROC) analysis, using both k-nearest neighbours (knn) algorithm (C). Six-protein prediction model for the mid-throughput method (B) and ROC analysis using knn algorithm (D). Both methods performed well at discriminating NSCLC patient from controls, with AUC of 77.8%, and 94.4% for high and mid, respectively.

## Conclusions

- Although the mid-throughput method gave more IDs overall, on a per-minute of run time basis the high-throughput method produced 1.5-fold more IDs.
- Machine learning algorithms identified several markers including PRTN3, CNDP1, NIT2, MNDA, ICAM3 which have been identified previously in lung cancer and show promise as future markers to be included in validation studies. There was one protein common between the two methods, Myeloblastin (PRTN3), a serine protease, which has been linked to KRAS mutations in lung cancer patients.
- The combination of our novel sample preparation methods with both the mid- and high-throughput protocols produced high quality data enabling differential biomarkers to be defined.
- Study demonstrates the potential use for microsampling in cancer biomarker discovery which has real-world clinical utility, providing a path to at-home sampling and patient centric monitoring.

**Acknowledgements**

The mass spectrometry was carried out using Sydney MS facilities at the University of Sydney

Sangui Bio

Natasha Lucas
natasha@sanguibio.com
www.sanguibio.com