Analytical validation of the PROphet computational model for clinical benefit prediction and decision-making tool in metastatic NSCLC

Itamar Sela[1], Coren Lahav[1], Ben Yellin[1], Yehonatan Elon[1], Michal Harel[1]

[1]OncoHost LTD

Abstract

The PROphet test provides a decision-making tool for first-line treatment of non-small cell lung cancer (NSCLC) patients without driver mutations. The test is based on plasma proteomics profiling using the SomaScan platform, followed by a machine learning-based analysis. During the PROphet predictor development, a set of predictive proteins was identified, termed resistance-associated proteins (RAPs). Here we set to examine the analytical validity of the SomaScan assay as the biomarker measurement method in the PROphet test, focusing either on all proteins measured using this platform, or specifically on the RAPs. Experimental precision analysis displayed a median coefficient of variation (CV) of 3.9% and 4.7% for intra-plate and inter-plate examination, respectively, when studying all proteins, with no difference between technicians. Notably, the RAPs displayed lower median CV values and lower CV range. The median accuracy rate of measurements obtained in sites was 88% and 94% for all proteins and for the RAPs, respectively, with no significant difference between technicians. Computational precision examination displayed high precision for 13 out of 14 examined samples, with median standard error of 0.1689. Cross-platform comparison between SomaScan platform and other proteomics platforms, including immunoassays, proximity extension assay (PEA) and mass spectrometry, displayed a median Spearman coefficient of 0.51 and 0.53 for all examined proteins and RAPs, with no significant difference between the two protein sets. Last, the effect of the cross-platform correlations on the prediction capabilities was examined, displaying only a minor effect of the correlation on the prediction capabilities of the PROphet predictor, with a median of 80% agreement between predictions obtained between the two examined proteomic platforms. Taken together, these results demonstrate strong analytical performance for the SomaScan technology, while the RAPs displayed improved capabilities in most of the analytical performance analyses.

Introduction

Standard of care for driver mutation-negative, metastatic non-small cell lung cancer (NSCLC) comprises immune checkpoint inhibitors (ICIs) that target PD-1, PD-L1 or CTLA-4 proteins expressed on tumor and immune cells[1]. The integration of ICI-based therapies into first- and second-line treatments for NSCLC has led to significant improvements in survival and quality of life in comparison to chemotherapy[2].

Factors that guide treatment choices in driver mutation-negative NSCLC include tumor PD-L1 expression level, tumor mutational burden, the extent of disease and histology. In the first-line setting, patients with PD-L1 ≥50% are typically treated with PD-(L)1 inhibitors as a monotherapy, whereas PD-(L)1 inhibitors combined with chemotherapy is preferred in cases of extensive disease and for patients with

PD-L1 <50%[1]. While PD-L1 testing is performed routinely for all newly diagnosed NSCLC patients, clinical evidence shows that the PD-L1 biomarker neither guarantees response to first-line ICI monotherapy in patients with high PD-L1 expression, nor rules out the possibility of response in patients with low or negative PD-L1 expression [3-7]. It is currently acknowledged that additional predictive biomarkers are needed to accurately classify patients and support treatment decisions[8].

We recently developed and validated a computational predictive model that analyzes proteomic profiles in pre-treatment blood plasma to predict clinical benefit from anti-PD-(L)1-based therapy with or without chemotherapy in NSCLC patients[9]. The model, which serves as the computational component of the commercially available PROphet® test, computes a PROphet score by integrating predictions from a large collection of proteomic biomarkers. The PROphet score (a value from 0 to 10) is a scaled metric related to the probability of achieving progression-free survival for at least 12 months. Patients with a PROphet-POSITIVE result (score ≥5) survive significantly longer than patients with a PROphet-NEGATIVE result (score <5), demonstrating that the model effectively stratifies patients according to survival outcomes. Notably, when integrated with PD-L1 expression levels, the PROphet result provides additional guidance for treatment selection. Specifically, it identifies PD-L1 ≥50% patients who are more likely to benefit from combination ICI-chemotherapy in comparison to ICI monotherapy despite high PD-L1 levels, and differentiates between PD-L1 <50% patients who are likely to benefit from the recommended combination ICI-chemotherapy and those who are not. Accordingly, the PROphet® test is intended as an aid for the clinical management of NSCLC patients.

PROphet score computation requires the expression levels of 388 plasma protein biomarkers collectively termed Resistance-Associated Proteins (RAPs). These RAPs are measured in a sample of the patient's blood plasma using the highly multiplexed, aptamer-based SomaScan assay, developed by SomaLogic, Inc. The assay relies on Slow Offrate Modified Aptamers (SOMAmers) – chemically-modified, single stranded oligonucleotides that bind to native proteins with high affinity and specificity – and is capable of measuring approximately 7000 human proteins over a broad range of concentrations[10-12]. Protein targets detected by the assay include receptors, growth factors, hormones and signaling proteins covering a wide range of biological functions (Figure S1). As such, the SomaScan assay has been used for biomarker discovery in multiple fields including cancer[9,13,14], infectious disease[15], autoimmune disease[16], cardiovascular disease[17], neurological disorders[18] and more, as well as for different applications, such as monitoring[19] or pharmacodynamic analysis[20].

Here, we investigate the analytical validity of the PROphet test in OncoHost's CLIA-certified laboratory where testing is performed. We focus on assay accuracy and reproducibility with respect to the RAP set as well as the entire set of measured proteins. We also assess computational precision of the test and compare performance across different platforms.

Methods

Sample collection

Plasma samples were collected from non-small cell lung cancer (NSCLC) patients either at baseline (before treatment commencement; termed here as T0) or before the second treatment (termed here as T1). Samples were obtained as part of a clinical study (PROPHETIC; NCT04056247). Clinical parameters were

collected, including Progression Free Survival (PFS) data. All clinical sites received IRB approval for the study protocol and all patients provided appropriate written informed consent. Blood samples were collected from each patient into EDTA-anticoagulant tubes, and the plasma was isolated from whole blood by centrifugation at 1200 x g at room temperature for 10-20 minutes within 4 hours of venipuncture. Plasma supernatant was collected and stored frozen at -80°C. Samples were shipped frozen to the analysis laboratory. The collection protocol is consistent with the Somalogic plasma collection guidelines [REF].

Proteomic profiling

Plasma samples from 90 patients were profiled using the SomaScan V4.1 platform (Boulder, CO)[11,12,21]. The assay simultaneously measures a total of 7596 protein targets, out of which 7289 targets are human proteins. SomaScan technology uses fluorescence to detect protein levels and the assay results are provided in relative fluorescence units (RFU). Runs were performed at two sites (OncoHost's laboratory in North Carolina; SomaLogic laboratory).

Plasma samples from 76 patients were profiled using the Olink proximity extension assay (PEA; Olink Proteomics, Uppsala, Sweden) on 13 human Olink Target 96 panels (Cardiometabolic, Cardiovascular II, Cardiovascular III, Cell Regulation, Development, Immune Response, Inflammation, Metabolism, Neuro Exploratory, Neurology, Oncology II, Oncology III, Organ damage) as previously described [22]. Two samples per patient were profiled (i.e., T0 and T1 samples). The assay results are provided in relative units called Normalized Protein eXpression (NPX) calculated from Ct values (as reported using real-time qPCR) following data normalization.

PROphet model

The PROphet model provides clinically-relevant information that can potentially aid in patient management and treatment decisions. For a thorough description of the algorithm and its clinical validity, please refer to Christopoulos et al.[9]. In brief, the model is based on proteomic profiling of a plasma sample obtained at T0 from an advanced stage non-small cell lung cancer (NSCLC) patient undergoing anti PD-1 / PD-L1 treatment, either as a monotherapy or in combination with chemotherapy. The model takes as an input the expression level of 388 proteins (termed as resistance-associated proteins, or RAPs) and the patient's sex. The model output is clinical benefit probability prediction for each patient (clinical benefit is defined as progression-free survival of at least 12 months). Based on the clinical benefit probability, the patients are assigned a PROphet-positive or -negative result using the median clinical benefit probability as a threshold.

Experimental precision

Intra-plate experimental precision was examined based on four technical replicates of four different patient samples, all assayed in the same running batch. Two sample preparation procedures were performed on two different days by two different technicians. The intra-plate experimental precision was calculated for each technician and was quantified using Coefficient of Variation (CV). In the case of sample disqualification, a minimum of three replicates was required. The CV was calculated for each protein and was assessed for each technician by the ratio $\sigma/\mu \times 100$, where σ denotes the standard deviation of all four replicates, and μ is the mean protein level across all four replicates.

Inter-plate experimental precision was determined on four patient samples assayed in four separate runs on four different days, by two technicians. Sample preparation was performed by two

different technicians, and the precision was calculated per technician. In the case of sample disqualification, a minimum of three replicas was required. For each sample, all the RFU values were measured and CV was calculated for each protein.

Computational precision

Computational precision is defined as the extent to which the algorithm output for a given sample is consistent between multiple runs. Computational precision was determined by the percentage of replicates that were assigned the same result by the algorithm (e.g., for a PROphet-negative sample the consistency is given by the percentage of replicates that were labeled by the algorithm as negative). Computational precision was calculated based on four analytical runs. Sample preparation was performed by two technicians on four days. Either six or ten replicates of 14 different samples were assayed in four different runs. The computational precision was calculated for all replicates of each sample.

Experimental accuracy

Accuracy was defined as the extent to which the protein levels measured at OncoHost's laboratory coincide with the measurements performed by SomaLogic for the same samples. The RFU measurements acquired by SomaLogic were taken as the reference value. Accuracy rate was quantified by the percentage of accurate protein measurements. A protein measurement is accurate if the ratio $RFU_{OH\ lab}$ / $RFU_{SomaLogic\ lab}$ is between 0.8 and 1.2 (QC ratio acceptance range, as defined by SomaLogic), where $RFU_{OH\ lab}$ is the protein level measured at OncoHost's laboratory in North Carolina and $RFU_{SomaLogic\ lab}$ is the reference value measured by SomaLogic for the same sample. Fourteen samples were examined.

Cross-platform comparisons of protein measurements

Comparison between the measurements obtained in the SomaScan platform and other protein measurement platforms was performed using Spearman rank correlation. The analysis was performed based on two sources: (i) 152 samples (samples from 76 patients collected at two time points) for which protein expression level measurements were performed using both SomaScan and PEA platforms. (ii) To expand to additional platforms measuring protein expression levels, as well as to broaden the coverage of the cross-platform analysis, external datasets with correlation comparisons were added to the analysis. Overall, 10 external datasets were added for the correlation comparison (supplementary table 3), covering comparison between SomaScan versus immunoassays, PEA and mass spectrometry (MS). Spearman correlation coefficients were obtained; since the correlations reported in Dammer et al.[23], were Pearson correlations, the Spearman correlation coefficient was calculated to maintain the same coefficient calculation method for all studies. Raffield and colleagues reported correlations measured in different platforms; the average data that they reported were used here (Table S2-9 of reference [24]). For cases reporting more than one value for a given protein (e.g., in the study of Kukova et al. [25], where samples were obtained before and after operation), the median level was used.

In addition to the comparison of protein level measurements across different platforms detailed above, cross-platform consistency was further evaluated for the algorithm outcome based on measured protein levels. Specifically, PROphet results (positive or negative) were compared for protein levels as measured using either SomaScan or PEA. The same samples were assayed using both platforms, and the PROphet model was applied on both. The RAPs that were used for the prediction model were selected based on the Spearman correlation between the two platforms; RAPs with correlation coefficient above

a given threshold were used for each prediction assessment. To avoid random selection of a threshold, different thresholds were examined between 0.05 and 0.95.

Data analysis

Data analysis was performed using python, R, GraphPad Prism (San Diego, California, USA, http://www.graphpad.com) and Perseus computational platform[26,27]. Log2 transformation was applied unless stated otherwise. For CV calculation, non-transformed RFU levels were used. Enrichment analysis was done using 1D enrichment test[28]. Proteomaps was done using Liebermeister et al. tool[29].

Results

The commercially available PROphet test is based on proteomic measurements followed by computational analysis. As such, analytical validation of the test was investigated on two levels: (i) experimental validation, in which reproducibility and sensitivity of measured proteins were assessed; (ii) model validation, in which reproducibility of PROphet scoring was assessed.

Experimental precision

Intra-plate experimental precision, which describes the variability of sample measurements in the same running batch, was determined by calculating the coefficient of variation (CV) of a set of technical replicates from a single SomaScan assay run. Samples from four patients were assayed. Sample preparation was performed by two technicians on two different days, and intra-plate experimental precision was determined per sample, per technician. For the entire set of human proteins measured in the SomaScan assay V4.1, the median CV was 3.284% - 5.812% (overall median of 3.9165%), with a median Q1 and Q3 of 2.601% and 6.053%, respectively (Figure 1A and Table S1A). The CV range differed between the four samples, but did not show variability between technicians (paired t-test p-value = 0.8170; Figure S2A). Functional analysis showed that membranal proteins and proteins involved in DNA-related processes (e.g., DNA repair, DNA metabolic process) were enriched with lower CV values, while mitochondrial-related proteins, as well as metabolism-related processes were enriched with higher CV values (1D-enrichment analysis; Table S2).

The PROphet model relies on 388 RAPs that were selected from a total of 7288 measured proteins based on their measurement stability and the difference in their expression level between patients with or without clinical benefit[9]. When comparing the CV of the RAPs to that of the entire set of proteins, there was no significant difference in the median CV, though there was a trend towards lower CV values for the RAP set (p-value = 0.0619; Figure 1B-C and S2B; Table S1B), where the RAP set had a median CV of 3.474% compared to 3.917% in the entire protein set. The CV values of the entire protein set spanned over a larger range than that of the RAP set, while the RAPs were significantly enriched with lower CV values (1D enrichment analysis, p-value = 6.18258E-05; Figure 1D). The intra-plate CV range of the RAP set was significantly lower than that of the entire protein set (paired p-value = 0.0003; Figure 1E). Furthermore, the median CV of the RAP set was similar between technicians (paired t-test p-value = 0.0687; Figure S2C).

The inter-plate experimental precision describes the assay's variability between runs, days and technicians. As in the intra-plate analysis, we examined inter-plate experimental precision by calculating the CV for four patient samples prepared by two technicians and assayed in four different running batches.

The median CV for the entire protein set was between 3.826 and 5.731, with a median Q1 and Q3 of 2.876% and 7.825% respectively (Figure 2A and Table S3A), similar to the median intra-plate CV. The inter-plate CV was comparably lower for the RAP set (Figure 2B-D, paired t-test p-value = 0.0212). The histogram of the inter-plate CV for the entire protein set spanned over a larger range than that of the RAP set, while the RAP set was significantly enriched with lower CV values (1D enrichment analysis, p-value = 1.4496E-09; Figure 2E). Last, for the RAP set, there was a trend towards a smaller inter-plate CV range, although it was not significantly different from that of the entire protein set (Figure 2F; paired t-test p-value = 0.0526; Table S3B). Functional analysis showed that mitochondrial and metabolism-related processes were significantly enriched with higher CV values, while membranal proteins, as well as extracellular proteins were enriched with lower CV values, similar to the trend that was observed for intra-plate CV (1D-enrichment analysis; Table S4). Altogether, these results demonstrate that the SomaScan assay serves as a precise proteomics platform in terms of technical replicates either within or between running batches. Superior analytical precision was achieved with RAPs in comparison to all proteins.

Experimental accuracy

An important feature for a clinical test is the ability to obtain reproducible measurements between different testing sites. According to SomaLogic, when the ratio between measurements obtained in two different sites lies within a range of 0.8-1.2 (per protein), the measurement is accepted as being accurate. To determine experimental accuracy rate, 14 patient samples were assayed in our laboratory by two technicians, and the results were compared to those obtained by the SomaLogic laboratory. The accuracy rate in the entire protein set was high, spanning between 72% and 96% (Figure 3A), and there was no difference between the two technicians in any of the examined samples (Figure S3). Notably, the RAPs displayed higher accuracy in all samples (Figure 3B), in accordance with the observation that high reproducibility of a given protein improves its signal to noise ratio, thereby increasing its probability to consistently differentiate between different patient populations. Overall, we demonstrate a high accuracy when comparing between two sites, with no observable difference between technicians.

Computational precision

Computational precision describes the consistency of an algorithm result when the procedure is applied repeatedly to multiple measurements of a single sample. To evaluate the test's computational precision, we ran 14 different patient samples repeatedly in four running batches on four different days, with the addition of inter-plate experimental variability. Examination of the scaled clinical benefit probability showed that repeated measurement of 13 samples returned the same PROphet result as the reference sample (Figure 4A-B), indicating very high consistency between repeated measurements. In addition, the standard error was low for all samples, with a median of 0.1689 (Figure 4C).

Cross-platform comparisons of protein measurement

Next, we sought to study the consistency between SomaScan measurements compared to other protein detection platforms. Such comparison enables an assessment and verification of protein measurements obtained using different platforms. To this end, we ran 152 samples from 76 patients in SomaScan and proximity-extension assay (PEA)-based platforms. The PEA platform enabled the profiling of 1196 proteins (1161 non-redundant proteins), 1105 of which were common to the SomaScan platform. In addition, previous studies have compared between the SomaScan assay and other protein biomarker detection platforms, including traditional immunoassays, PEA and mass spectrometry (MS)[13,14,23-25,30-34]. We

combined 10 such studies in order to obtain a larger coverage of proteins, as well as to add additional sources for comparison (Table S3). Altogether, there was a Spearman rank correlation for 2247 proteins (Figure 5A and 5B). The median correlation was 0.5077, with a good agreement between correlations obtained from different studies (Figure S4 displays the histogram of the standard deviation for each protein, with a median standard deviation of 0.17041), though there were cases with either lower or higher correlation profiles. Notably, there were major differences in the cohort size between the different studies, which may affect the results. Immunoassays, which enable measurement of fewer proteins, varied in their correlation profile, ranging from a median of 0.1639 to 0.6875. The Spearman correlations between MS and SomaScan displayed a similar trend as was observed for the correlation profile of PEA-SomaScan, with slightly improved median correlation of 0.5366. In addition to inter-platform correlations, protein verification can be done based on cis-protein quantitative trait locus (cis-pQTL) analysis, as well as by verification of proteins using MS measurements (Figure S5). Examination of published data[24,34-36] revealed 2036 proteins that were verified by at least one approach, while approximately half of them were verified by two or more studies. When examining proteins that were verified either using cis-pQTL, MS-based validation or Spearman correlation equal to or above 0.5, in total 2491 proteins remained, covering 34% of the proteome examined by SomaScan. Correlation data was available for 191 out of 388 RAPs, showing a very similar trend to the entire protein set (Figure 5A and 5B). Comparison between RAPs and all proteins showed no significant overall difference in the correlation coefficient distribution (Mann Whitney test p-value for the median correlation = 0.9667). When adding cis-pQTL and MS-based verification together with RAPs with Spearman correlation above or equal to 0.5, 199 proteins remained accounting for more than 50% of the RAPs, 40 of which were verified by all approaches.

In order to examine whether the correlation between SomaScan and other platform measurements can affect the prediction capabilities of the PROphet model, we examined PROphet output results when running the model on either PEA or SomaScan data. Exploring different thresholds of protein measurement correlations between the two platforms showed a stable correlation between the PROphet score (values 0-10) obtained using SomaScan and that obtained using PEA, with a median Pearson correlation of 0.74 and with negligible effect on the threshold of correlation between PEA and SomaScan (Figure 6A). Furthermore, there was a high degree of agreement between the PROphet result (PROphet NEGATIVE or POSITIVE) obtained using each platform, with a median of 80% of samples with the same PROphet result (Figure 6B). Overall, these findings imply that the correlation between the platforms has only a minor effect on the prediction capabilities of the PROphet test.

Discussion

Circulating blood contains thousands of proteins involved in a vast range of biological processes. Alterations in blood protein levels can denote disease onset or progression, as well as mechanisms underlying response or resistance to therapy. As such, the blood proteome is a rich source of potential biomarkers that can provide valuable information for different medical applications, including diagnosis, normal physiological status assessment, evaluating treatment benefit or monitoring disease progression[37-39]. Despite the promise of blood-based proteomics, discovery of novel bona-fide protein biomarkers is challenging due to three main reasons: (i) blood matrix is extremely complex, displaying a large dynamic range between low- and high-abundance plasma proteins[38,39]. Advances in proteomic profiling technologies have helped to better handle this complexity, leading to a deeper examination of

the plasma proteome and enabling the discovery of new blood-based biomarkers[37,40]; (ii) proteins exist in a range of different structures due to splicing, variations in amino acid sequence, protein complex formation, or post-translational modification. This diversity in protein structure, termed proteoform[41-43], may introduce noise when measuring proteins in a given proteomic platform, as well as when comparing between platforms; (iii) human blood samples are highly heterogeneous owing to clinical, biological and environmental variability. Given these challenges, it is important to perform a thorough assessment of the biomarker discovery platform of choice, as well as to explore its analytical validity. Here we investigated the analytical validity of the SomaScan assay currently used as the biomarker measurement method in the PROphet test.

Overall, the SomaScan platform displayed strong analytical capabilities, with low CV values within the same running batch assayed on different days, between batches and between technicians. In addition, high accuracy was observed between sites, consistent with previous reports[10,12,44]. The CV values are lower than those reported for MS- and PEA-based proteomics[45-47]. In addition, repeated measurements of the same sample displayed high consistency in terms of obtaining the same PROphet result. These properties are important not only for a clinical test, but also for the biomarker discovery process. Technologies with higher CV values require larger cohorts to overcome the inherent noise in clinical data as well as to handle the complexity of the plasma proteome; therefore, one of the advantages of SomaScan's technology for biomarker discovery is that it performs well even with small cohorts. Notably, the RAPs displayed improved analytical properties in comparison to the entire set of proteins measured by the assay. Specifically, lower CV values and a higher accuracy rate were obtained with the RAP measurements. This may result from the process of RAP selection during model development, where the proteomic dataset was first narrowed down to a set of proteins with high analytical reliability. In addition, RAPs were selected iteratively using a cross-validation process, where in each iteration a different subset of patients was examined; this enabled the selection of proteins that are more confidently different between the examined groups (patients with and without clinical benefit). Clinical datasets can be complex, noisy, and contain many variables explored on a limited cohort, making it difficult to identify meaningful biomarkers that can be generalized to a new cohort of patients. Overall, the analytical superiority of the RAP set demonstrates the importance of applying good practice when identifying predictive biomarkers during model development.

Previous studies have compared between SomaScan measurements and those obtained using other proteomics platforms. Here we summarized multiple such studies and added a dataset that compares between SomaScan and PEA, encompassing approximately 30% of the proteome examined by SomaScan V4.1. When adding protein verification using pQTL or MS analysis, along with high-correlation across platforms, our analysis covered 30% of the SomaScan proteins and 50% of the RAPs, all of which can be considered as having high-confidence agreement between platforms. Importantly, low correlation between platforms does not necessarily result from specificity or cross-reactivity issues in one of the examined platforms. There may be multiple reasons for such discrepancy, including different proteoforms examined in the different platforms or affinity differences between technologies. Overall, the specificity of the SomaScan assay arises from two main factors[12]. (i) Aptamer design procedure, where aptamers are designed to specifically bind to their target proteins with high affinity and selectivity using a process that involves screening of a large library of aptamers for those that specifically bind to the target protein. During the aptamer design process, each aptamer is characterized thoroughly, leading to the selection of

very high affinity aptamers. (ii) Quality control procedures including performance testing with known protein standards and performance monitoring over time.

Last, the effect of the inter-platform correlation was tested by comparing PEA and SomaScan based predictions. Of note, the PEA technology used in this study measures only 102 out of 388 RAPs upon which the model is based. Therefore, we could not test the ability of the PEA platform to reproduce the results using all 388 proteins. Despite this limitation, the ability to achieve consistent results on two different platforms was remarkable, with high correlations between the output of the two technologies. Importantly, the correlations between the results obtained using the two platforms were stable, regardless of the inter-platform measurement correlation threshold. This indicates that cross-platform correlations did not play a role. This result is a powerful demonstration for the compatibility of the PROphet model approach to high throughput proteomic data analysis. Overall, these findings imply that the PROphet model is compatible with different proteomic platforms, highlighting its inherent robustness.

To conclude, here we explored the analytical validity of the SomaScan platform for clinical benefit prediction of NSCLC patients both experimentally and computationally. Overall, the platform displayed strong analytical capabilities, enabling robust and reproducible predictions.

Acknowledgements

We thank Dr. Kimberly McGregor for feedback and support.

Funding

This study was supported by OncoHost LTD.

Conflicts of interest

MH, CL, IS, BY and YE are employees of OncoHost.

Figure legends

Figure 1: Intra-plate experimental precision. A. Coefficient of variation (CV) calculated for each protein based on measurements of four technical replicates of four different samples by two technicians. The lines indicate the median and interquartile range of each sample. B. CV for the 388 resistance associated proteins (RAPs), as determined by the PROphet model. The line indicates the median of each sample. C. The median intra-plate CV for all proteins and for RAPs only. Paired t-test p-value is presented. D. Histogram of the median CV values of all proteins (dark blue) compared to the median CV values of the RAPs (light blue). E. The range of intra-plate CV for all proteins and for RAPs. Paired t-test p-value is presented.

Figure 2: Inter-plate experimental precision. A. Coefficient of variation (CV) calculated for each protein based on measurements of four replicates assayed in four different runs by two technicians. The lines indicate the median and interquartile range of each sample. B. CV for the 388 resistance associated proteins (RAPs), as determined by the PROphet model. The line indicates the median of each sample. C. The median inter-plate CV for all proteins and for RAPs only. Paired t-test p-value is presented. D. Histogram of the median CV values of all proteins (dark blue) compared to the median CV values of the RAPs (light blue). E. The range of inter-plate CV for all proteins and for RAPs. Paired t-test p-value is presented.

Figure 3: Accuracy rate of measurements between sites. The measurements were performed at OncoHost's laboratory and a reference laboratory (SomaLogic's laboratory). Shown are the median accuracy rates per technician for the entire protein set (A) and for the RAP set (B). The accuracy for each protein measurement was determined based on the ratio between the measurements obtained in each lab. An accurate measurement is defined when the ratio lies between 0.8 and 1.2. The plot displays the accuracy rate (the fraction of accurate protein measurements) per sample.

Figure 4: Computational precision of the PROphet model. A. The PROphet result obtained in multiple repeated measurements of the same sample. The line indicates the median of each sample. B. The consistency percentage of the algorithm to provide the same PROphet result (negative or positive). C. The standard error for the consistency.

Figure 5: Correlations between SomaScan platform and other proteomics platforms. A. A heatmap displaying the Spearman rank correlation coefficients for all proteins with available correlation data (left) and the RAPs with available correlation data (right). B. The distribution of the Spearman rank correlation coefficients for all proteins with available correlation data (top) and the RAPs with available correlation data (bottom).

Figure 6: PROphet model result evaluation using SomaScan or proximity extension assay (PEA) based platforms. A. The Pearson correlation between the PROphet result (positive / negative) as obtained using SomaScan or PEA platforms. The different protein correlation thresholds (measured using Spearman rank coefficient) are indicated on the x-axis. A secondary x-axis displays the number of proteins that were used by the PROphet model to obtain the PROphet result. B. The consistency percentage between the two platforms for each protein correlation threshold. A consistency is defined by the fraction of samples that obtained the same PROphet result in both platforms.

Supplementary figure legends

Figure S1. Voronoi plot displaying functional groups explored using SomaScan technology. Each small polygon represents a protein that is measured using SomaScan. The polygons are grouped based on KEGG functional groups. The map was generated using Proteomaps.

Figure S2. Intra-plate precision analysis. A. Intra-plate range comparison between technicians for all proteins. Paired t-test p-value is presented. B. The median intra-plate CV for all proteins compared to the median CV of the RAPs. C. Intra-plate range comparison between technicians for RAPs. Paired t-test p-value is presented.

Figure S3. Comparison between the median accuracy rate of the two technicians. The p-values were calculated using Mann-Whitney test and were adjusted using Benjamini Hochberg false discovery rate multiple hypothesis correction (indicated as a.p in the figure). The interval indicates the inter-quartile range.

Figure S4. Histogram of standard deviation of the median correlation between SomaScan and other platforms.

Figure S5. Orthogonal validation of proteins measured using SomaScan. A. A heatmap describing orthogonal verification using cis-pQTL or mass-spectrometry for all proteins (left) or the RAPs (right). B. Venn diagram showing the number of proteins verified using either one of the three methods.

References

1. Hanna NH, Schneider BJ, Temin S, et al. Therapy for Stage IV Non-Small-Cell Lung Cancer Without Driver Alterations: ASCO and OH (CCO) Joint Guideline Update. *J Clin Oncol*. 05 10 2020;38(14):1608-1632. doi:10.1200/JCO.19.03022
2. Pabani A, Butts CA. Current landscape of immunotherapy for the treatment of metastatic non-small-cell lung cancer. *Curr Oncol*. Jun 2018;25(Suppl 1):S94-S102. doi:10.3747/co.25.3750
3. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med*. Jul 09 2015;373(2):123-35. doi:10.1056/NEJMoa1504627
4. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med*. Oct 22 2015;373(17):1627-39. doi:10.1056/NEJMoa1507643
5. Mok TSK, Wu YL, Kudaba I, et al. Pembrolizumab versus chemotherapy for previously untreated, PD-L1-expressing, locally advanced or metastatic non-small-cell lung cancer (KEYNOTE-042): a randomised, open-label, controlled, phase 3 trial. *Lancet*. May 04 2019;393(10183):1819-1830. doi:10.1016/S0140-6736(18)32409-7
6. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Five-Year Outcomes With Pembrolizumab Versus Chemotherapy for Metastatic Non-Small-Cell Lung Cancer With PD-L1 Tumor Proportion Score ≥ 50. *J Clin Oncol*. Jul 20 2021;39(21):2339-2349. doi:10.1200/JCO.21.00174
7. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med*. 11 10 2016;375(19):1823-1833. doi:10.1056/NEJMoa1606774
8. Arora S, Velichinskii R, Lesh RW, et al. Existing and Emerging Biomarkers for Immune Checkpoint Immunotherapy in Solid Tumors. *Adv Ther*. 10 2019;36(10):2638-2678. doi:10.1007/s12325-019-01051-z
9. Petros C. A novel plasma proteomic computational model for predicting therapeutic benefit from immune checkpoint inhibitors in non-small cell lung cancer.
10. Candia J, Daya GN, Tanaka T, Ferrucci L, Walker KA. Assessment of variability in the plasma 7k SomaScan proteomics assay. *Sci Rep*. Oct 13 2022;12(1):17147. doi:10.1038/s41598-022-22116-0
11. Gold L, Ayers D, Bertino J, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PLoS One*. Dec 2010;5(12):e15004. doi:10.1371/journal.pone.0015004
12. Kim CH, Tworoger SS, Stampfer MJ, et al. Stability and reproducibility of proteomic profiles measured with an aptamer-based platform. *Sci Rep*. May 30 2018;8(1):8382. doi:10.1038/s41598-018-26640-w
13. Finkernagel F, Reinartz S, Schuldner M, et al. Dual-platform affinity proteomics identifies links between the recurrence of ovarian carcinoma and proteins released into the tumor microenvironment. *Theranostics*. 2019;9(22):6601-6617. doi:10.7150/thno.37549

14. Graumann J, Finkernagel F, Reinartz S, et al. Multi-platform Affinity Proteomics Identify Proteins Linked to Metastasis and Immune Suppression in Ovarian Cancer Plasma. *Front Oncol*. 2019;9:1150. doi:10.3389/fonc.2019.01150
15. Filbin MR, Mehta A, Schneider AM, et al. Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Rep Med*. May 18 2021;2(5):100287. doi:10.1016/j.xcrm.2021.100287
16. Stanley S, Vanarsa K, Soliman S, et al. Comprehensive aptamer-based screening identifies a spectrum of urinary biomarkers of lupus nephritis across ethnicities. *Nat Commun*. May 04 2020;11(1):2197. doi:10.1038/s41467-020-15986-3
17. Williams SA, Ostroff R, Hinterberg MA, et al. A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci Transl Med*. Apr 06 2022;14(639):eabj9625. doi:10.1126/scitranslmed.abj9625
18. Abdi IY, Bartl M, Dakna M, et al. Cross-sectional proteomic expression in Parkinson's disease-related proteins in drug-naïve patients vs healthy controls with longitudinal clinical follow-up. *Neurobiol Dis*. Feb 2023;177:105997. doi:10.1016/j.nbd.2023.105997
19. Tarca AL, Romero R, Bhatti G, et al. Human Plasma Proteome During Normal Pregnancy. *J Proteome Res*. Nov 04 2022;21(11):2687-2702. doi:10.1021/acs.jproteome.2c00391
20. Hyland PL, Chekka LMS, Samarth DP, et al. Evaluating the Utility of Proteomics for the Identification of Circulating Pharmacodynamic Biomarkers of IFNβ-1a Biologics. *Clin Pharmacol Ther*. Jan 2023;113(1):98-107. doi:10.1002/cpt.2778
21. Kraemer S, Vaught JD, Bock C, et al. From SOMAmer-based biomarker discovery to diagnostic and clinical applications: a SOMAmer-based, streamlined multiplex proteomic assay. *PLoS One*. 2011;6(10):e26332. doi:10.1371/journal.pone.0026332
22. Lundberg M, Eriksson A, Tran B, Assarsson E, Fredriksson S. Homogeneous antibody-based proximity extension assays provide sensitive and specific detection of low-abundant proteins in human blood. *Nucleic Acids Res*. Aug 2011;39(15):e102. doi:10.1093/nar/gkr424
23. Dammer EB, Ping L, Duong DM, et al. Multi-platform proteomic analysis of Alzheimer's disease cerebrospinal fluid and plasma reveals network biomarkers associated with proteostasis and the matrisome. *Alzheimers Res Ther*. Nov 17 2022;14(1):174. doi:10.1186/s13195-022-01113-5
24. Raffield LM, Dang H, Pratte KA, et al. Comparison of Proteomic Assessment Methods in Multiple Cohort Studies. *Proteomics*. Jun 2020;20(12):e1900278. doi:10.1002/pmic.201900278
25. Kukova LZ, Mansour SG, Coca SG, et al. Comparison of Urine and Plasma Biomarker Concentrations Measured by Aptamer-Based versus Immunoassay Methods in Cardiac Surgery Patients. *J Appl Lab Med*. Nov 2019;4(3):331-342. doi:10.1373/jalm.2018.028621
26. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods*. 09 2016;13(9):731-40. doi:10.1038/nmeth.3901
27. Tyanova S, Cox J. Perseus: A Bioinformatics Platform for Integrative Analysis of Proteomics Data in Cancer Research. *Methods Mol Biol*. 2018;1711:133-148. doi:10.1007/978-1-4939-7493-1_7
28. Cox J, Mann M. 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC Bioinformatics*. 2012;13 Suppl 16:S12. doi:10.1186/1471-2105-13-S16-S12
29. Liebermeister W, Noor E, Flamholz A, Davidi D, Bernhardt J, Milo R. Visual account of protein investment in cellular functions. *Proc Natl Acad Sci U S A*. Jun 2014;111(23):8488-93. doi:10.1073/pnas.1314810111
30. Liu RX, Thiessen-Philbrook HR, Vasan RS, et al. Comparison of proteomic methods in evaluating biomarker-AKI associations in cardiac surgery patients. *Transl Res*. Dec 2021;238:49-62. doi:10.1016/j.trsl.2021.07.005

31. Tin A, Yu B, Ma J, et al. Reproducibility and Variability of Protein Analytes Measured Using a Multiplexed Modified Aptamer Assay. *J Appl Lab Med*. Jul 2019;4(1):30-39. doi:10.1373/jalm.2018.027086
32. Katz DH, Robbins JM, Deng S, et al. Proteomic profiling platforms head to head: Leveraging genetics and clinical traits to compare aptamer- and antibody-based methods. *Sci Adv*. Aug 19 2022;8(33):eabm5164. doi:10.1126/sciadv.abm5164
33. Lim SY, Lee JH, Welsh SJ, et al. Evaluation of two high-throughput proteomic technologies for plasma biomarker discovery in immunotherapy-treated melanoma patients. *Biomark Res*. 2017;5:32. doi:10.1186/s40364-017-0112-9
34. Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Synergistic insights into human health from aptamer- and antibody-based proteomic profiling. *Nat Commun*. Nov 24 2021;12(1):6822. doi:10.1038/s41467-021-27164-0
35. Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. *Nature*. 06 2018;558(7708):73-79. doi:10.1038/s41586-018-0175-2
36. Emilsson V, Ilkov M, Lamb JR, et al. Co-regulatory networks of human serum proteins link genetics to disease. *Science*. 08 24 2018;361(6404):769-773. doi:10.1126/science.aaq1327
37. Deutsch EW, Omenn GS, Sun Z, et al. Advances and Utility of the Human Plasma Proteome. *J Proteome Res*. 12 03 2021;20(12):5241-5263. doi:10.1021/acs.jproteome.1c00657
38. Geyer PE, Holdt LM, Teupser D, Mann M. Revisiting biomarker discovery by plasma proteomics. *Mol Syst Biol*. 09 26 2017;13(9):942. doi:10.15252/msb.20156297
39. Geyer PE, Voytik E, Treit PV, et al. Plasma Proteome Profiling to detect and avoid sample-related biases in biomarker studies. *EMBO Mol Med*. Nov 07 2019;11(11):e10427. doi:10.15252/emmm.201910427
40. Van Gool A, Corrales F, Čolović M, et al. Analytical techniques for multiplex analysis of protein biomarkers. *Expert Rev Proteomics*. 04 2020;17(4):257-273. doi:10.1080/14789450.2020.1763174
41. Aebersold R, Agar JN, Amster IJ, et al. How many human proteoforms are there? *Nat Chem Biol*. Feb 14 2018;14(3):206-214. doi:10.1038/nchembio.2576
42. Forgrave LM, Wang M, Yang D, DeMarco ML. Proteoforms and their expanding role in laboratory medicine. *Pract Lab Med*. Jan 2022;28:e00260. doi:10.1016/j.plabm.2021.e00260
43. Smith LM, Kelleher NL, Proteomics CfTD. Proteoform: a single term describing protein complexity. *Nat Methods*. Mar 2013;10(3):186-7. doi:10.1038/nmeth.2369
44. Candia J, Cheung F, Kotliarov Y, et al. Assessment of Variability in the SOMAscan Assay. *Sci Rep*. 10 27 2017;7(1):14248. doi:10.1038/s41598-017-14755-5
45. Wik L, Nordberg N, Broberg J, et al. Proximity Extension Assay in Combination with Next-Generation Sequencing for High-throughput Proteome-wide Analysis. *Mol Cell Proteomics*. 2021;20:100168. doi:10.1016/j.mcpro.2021.100168
46. Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst*. Mar 23 2016;2(3):185-95. doi:10.1016/j.cels.2016.02.015
47. Andersen LC, Palstrøm NB, Diederichsen A, Lindholt JS, Rasmussen LM, Beck HC. Determining Plasma Protein Variation Parameters as a Prerequisite for Biomarker Studies-A TMT-Based LC-MSMS Proteome Investigation. *Proteomes*. Dec 01 2021;9(4)doi:10.3390/proteomes9040047

Figure 1

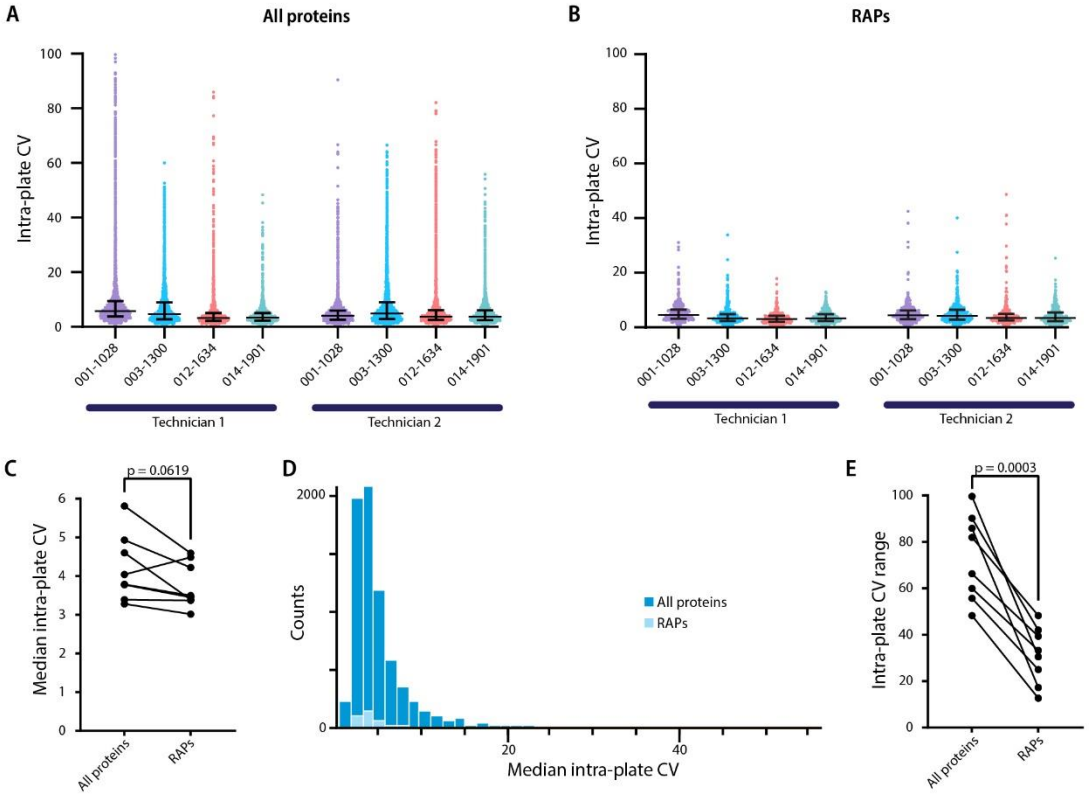

A
All proteins
Intra-plate CV
100
80
60
40
20
0
001-1028
003-1300
012-1634
014-1901
Technician 1
Technician 2
B
RAPs
C
p = 0.0619
Median intra-plate CV
All proteins
RAPs
D
Counts
2000
All proteins
RAPs
20
40
Median intra-plate CV
E
p = 0.0003
Intra-plate CV range

Figure 2

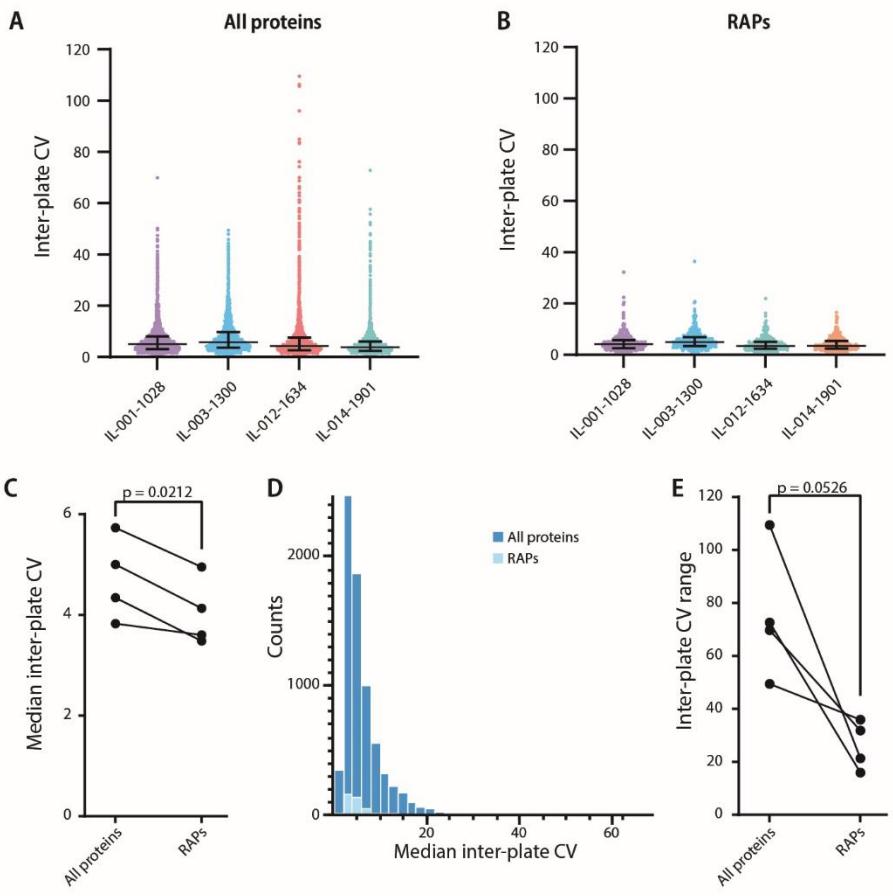

A
All proteins
Inter-plate CV
120
100
80
60
40
20
0
IL-001-1028
IL-003-1300
IL-012-1634
IL-014-1901
B
RAPs
Inter-plate CV
120
100
80
60
40
20
0
IL-001-1028
IL-003-1300
IL-012-1634
IL-014-1901
C
p = 0.0212
Median inter-plate CV
6
4
2
0
All proteins
RAPs
D
All proteins
RAPs
Counts
2000
1000
0
20
40
60
Median inter-plate CV
E
p = 0.0526
Inter-plate CV range
120
100
80
60
40
20
0
All proteins
RAPs

Figure 3

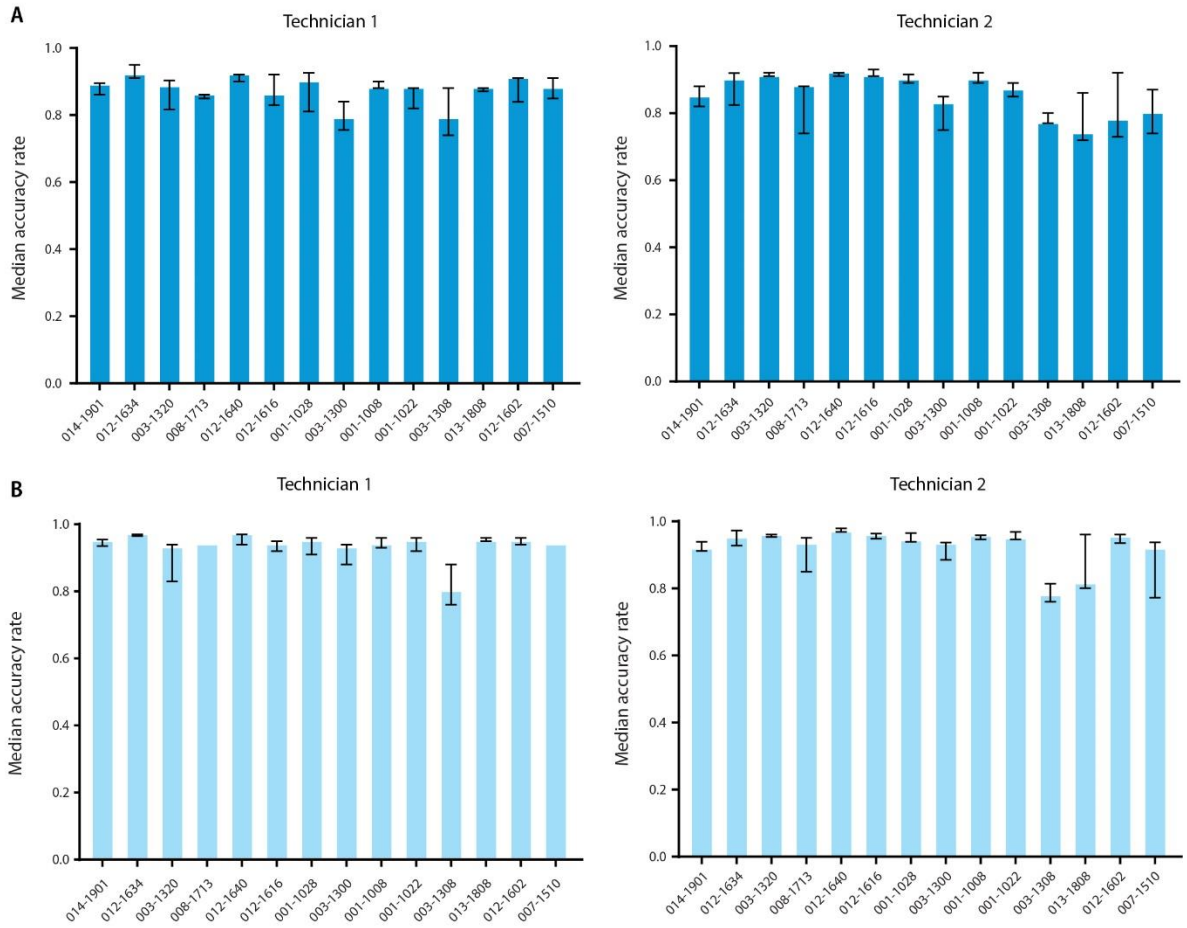

A
Technician 1
Technician 2
B
Technician 1
Technician 2
Median accuracy rate
1.0
0.8
0.6
0.4
0.2
0.0
014-1901
012-1634
003-1320
008-1713
012-1640
012-1616
001-1028
003-1300
001-1008
001-1022
003-1308
013-1808
012-1602
007-1510

Figure 4

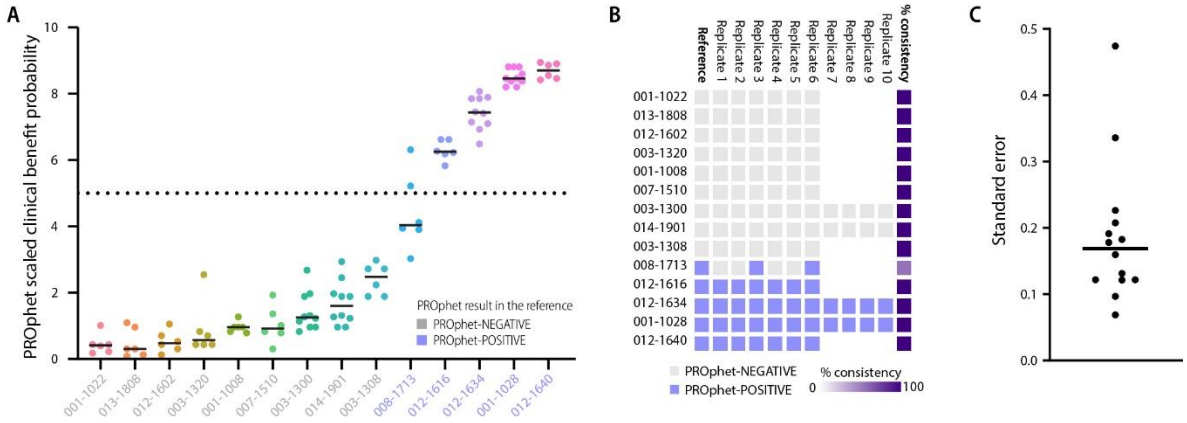

A
PROphet scaled clinical benefit probability
PROphet result in the reference
PROphet-NEGATIVE
PROphet-POSITIVE
001-1022
013-1808
012-1602
003-1320
001-1008
007-1510
003-1300
014-1901
003-1308
008-1713
012-1616
012-1634
001-1028
012-1640
B
Reference
Replicate 1
Replicate 2
Replicate 3
Replicate 4
Replicate 5
Replicate 6
Replicate 7
Replicate 8
Replicate 9
Replicate 10
% consistency
PROphet-NEGATIVE
PROphet-POSITIVE
% consistency
0
100
C
Standard error

Figure 5

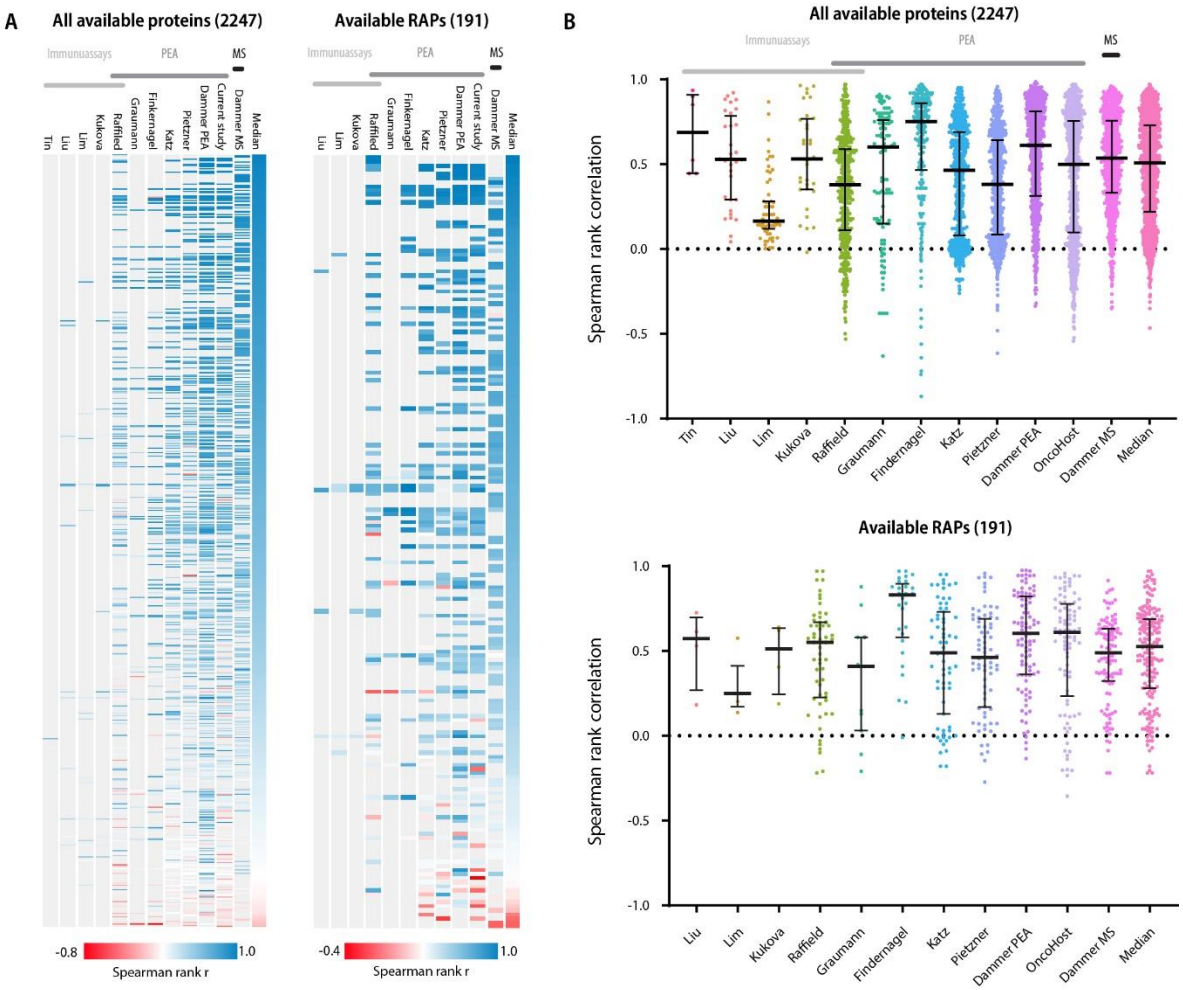
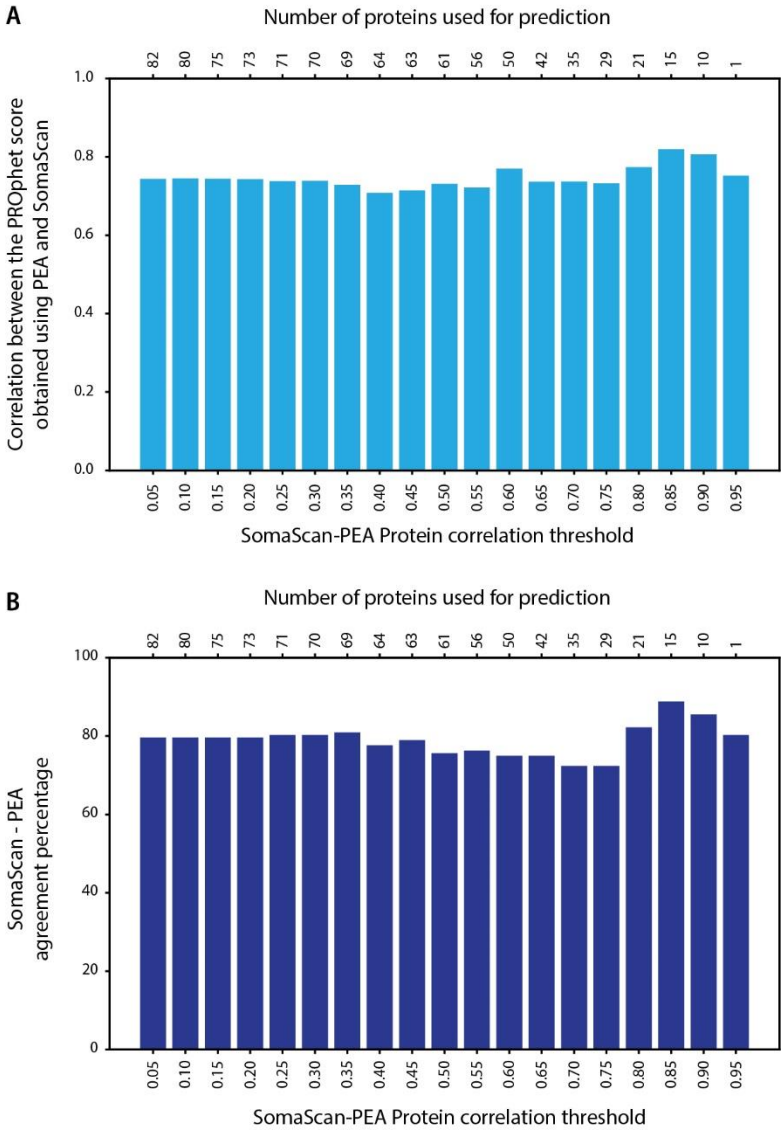

A
All available proteins (2247)
Available RAPs (191)
Immunuassays
PEA
MS
Tin
Liu
Lim
Kukova
Raffiled
Graumann
Finkernagel
Katz
Pietzner
Dammer PEA
Current study
Dammer MS
Median
-0.8
1.0
-0.4
Spearman rank r
B
Spearman rank correlation
1.0
0.5
0.0
-0.5
-1.0
OncoHost

Figure 6


A
Number of proteins used for prediction
82 80 75 73 71 70 69 64 63 61 56 50 42 35 29 21 15 10 1
Correlation between the PROphet score obtained using PEA and SomaScan
1.0
0.8
0.6
0.4
0.2
0.0
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
SomaScan-PEA Protein correlation threshold
B
Number of proteins used for prediction
82 80 75 73 71 70 69 64 63 61 56 50 42 35 29 21 15 10 1
SomaScan - PEA agreement percentage
100
80
60
40
20
0
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
SomaScan-PEA Protein correlation threshold