

<https://doi.org/10.1038/s41698-026-01517-8>

Mass spectrometry-based proteomics delivers in-depth proteome profiling of FFPE lung cancer biopsies from single glass slides



Olena Berkovska¹, Igor Schliemann^{1,2}, Georgios Mermelekas¹, Nazlı Ezgi Özkan¹, Mahnaz Nikpour¹, Vilde Drageset Haakensen³, Åslaug Helland³, Janne Lehtiö¹ ✉ & Lukas M. Orre¹ ✉

Clinical proteomics has the potential to add a valuable data layer to genomic and histopathological analyses in precision oncology, but its application to limited clinical material remains challenging. Here, we demonstrate that state-of-the-art mass spectrometry-based proteomics enables in-depth proteomic profiling of formalin-fixed paraffin-embedded (FFPE) clinical samples, including small diagnostic biopsies and single tissue sections mounted on glass slides. Despite minimal input material, single-slide analyses of clinical lung tumor specimens yielded biologically and clinically informative data, supporting detection of actionable proteins, immune-related signatures, and multivariate biomarkers. These results establish the feasibility of proteomics for retrospective FFPE studies and routine clinical practice, expanding opportunities for biomarker discovery and precision medicine from scarce tissue material.

Proteomics has emerged as a promising complement to immunohistochemistry (IHC) and genomic profiling in cancer research and diagnostics, offering deeper insight into tumor biology and therapeutic response^{1–6}. To date, most proteomics studies have relied on fresh-frozen tissue; however, formalin-fixed paraffin-embedded (FFPE) samples are the most widely used material in clinical practice, with millions of specimens stored in biobanks. FFPE material therefore holds great potential for the large-scale studies needed to decipher cancer complexity and advance precision medicine. Substantial progress has been made in mass spectrometry (MS)-based proteomics of FFPE samples, driven by advances in instrument sensitivity and improved workflows^{7–10}. Nonetheless, performance in clinical settings with limited material, such as biopsies and/or tissue sections mounted on glass slides, remains unclear.

In this study, we analyzed a cohort of FFPE tumor biopsy specimens from unresectable lung cancer using a single 4- μ m section on a glass slide per patient (“single-slide cohort”, $n = 68$; Fig. 1A and Supplementary Table 1). An additional, smaller cohort comprising multiple FFPE sections from biopsies or surgically resected tumor specimens served as a reference (“multi-section cohort”, $n = 15$). The samples were analyzed using label-free data-independent acquisition (DIA) proteomics on two state-of-the-art MS

instruments (Orbitrap Astral and timsTOF HT). In addition, multi-section cohort samples were analyzed on an older-generation MS instrument (Orbitrap Exploris).

Sample preparation for both cohorts followed the same general workflow, with two key differences: inclusion of a wash step and the use of an MS-based peptide amount estimation instead of a colorimetric assay for the single-slide cohort (Fig. 1A). In a small pilot analysis ($n = 6$) of the single-slide cohort, we observed substantially higher MS signal from skin keratins compared with the multi-section cohort (Supplementary Fig. 1A, B). High-abundance proteins present a challenge in MS proteomics due to ion suppression of lower-abundance peptide signals. We hypothesized that the elevated skin keratin signal resulted from contamination accumulated during handling and long-term storage of glass slides. Accordingly, we introduced a glass-washing step using a mild non-ionic detergent for preparation of the full cohort. Although this step modestly reduced skin keratin levels, the difference was not significant, and the proportion of MS signal originating from a small number of high-abundance proteins remained significantly higher than in the multi-section cohort (Supplementary Fig. 1B). Further protocol optimization may improve these results, but was beyond the scope of this study due to limited material availability. Therefore,

¹Department of Oncology and Pathology, Karolinska Institutet, SciLifeLab, Solna, Sweden. ²Department of Pathology and Cancer Diagnostics, Karolinska University Hospital, Stockholm, Sweden. ³Department of Oncology, Oslo University Hospital, Oslo, Norway. ✉e-mail: janne.lehtio@ki.se; lukas.orre@ki.se

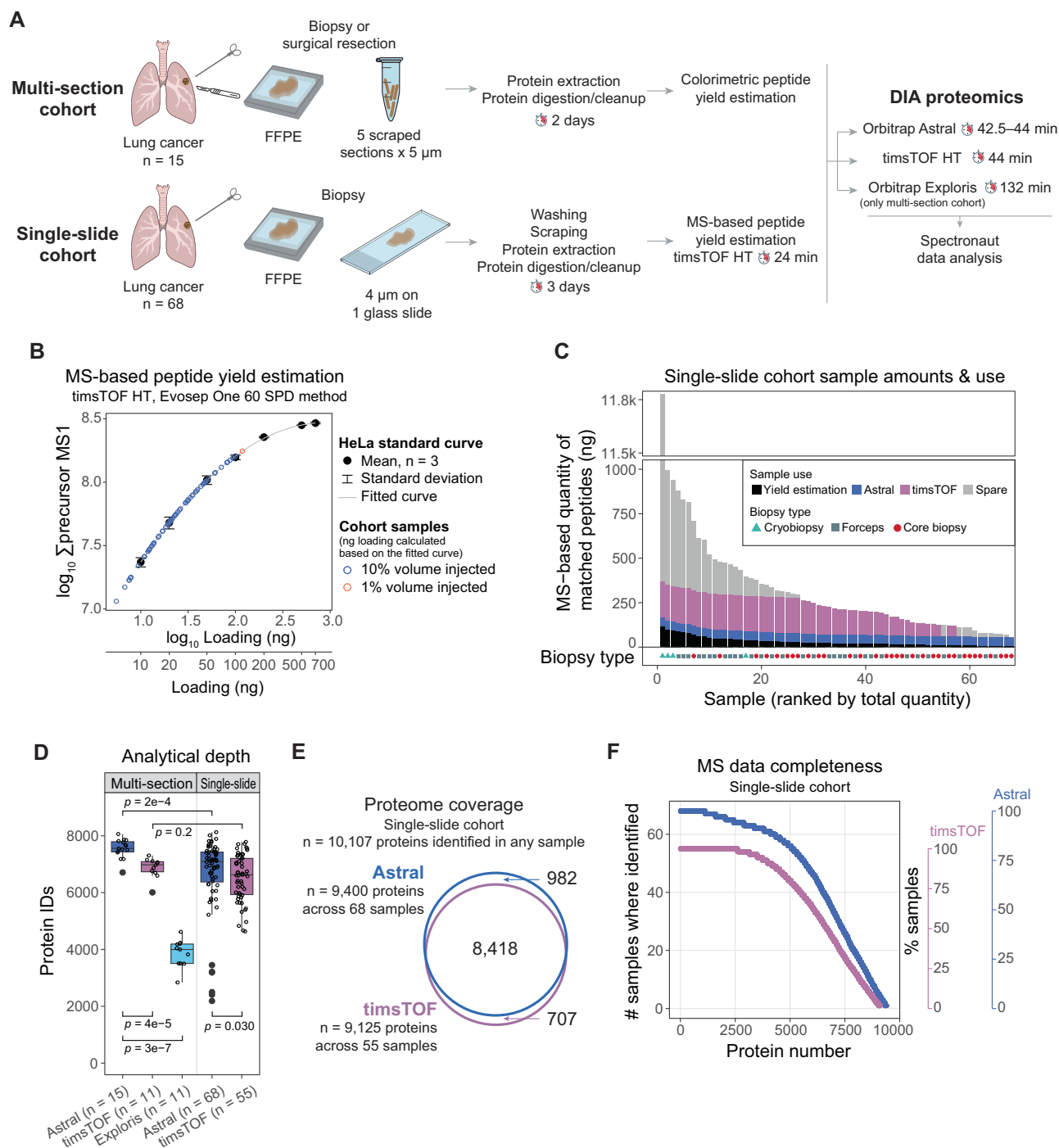


Fig. 1 | MS-based analysis of formalin-fixed, paraffin embedded (FFPE) clinical tumor samples delivers in-depth, proteome-wide profiling. **A** Schematic overview of the study design. **B** HeLa-based standard curve relating injected peptide amount to total MS1 signal, with single-slide cohort samples ($n = 68$) projected onto the curve to estimate injected peptide amounts. The curve was modeled using a quadratic polynomial regression relating $\log_{10}$ -transformed values. **C** Estimated peptide yields extracted from glass slides in the single-slide cohort (calculated as shown in **B**)

and the proportion of each sample used for the different analyses. **D** Number of proteins (gene-centric) identified per sample in MS-based analyses of the multi-section and single-slide cohorts. P values were calculated using two-sided Wilcoxon rank-sum test. The number of samples in each dataset is indicated in the figure. **E** Total number of proteins (gene-centric) identified in the single-slide cohort analyses and the overlap between the Astral and timsTOF datasets. **F** Number of samples in which each protein was identified in the single-slide cohort analyses.

we will consider skin keratin contamination as a quality control parameter in the downstream analysis.

The second protocol adaptation for the single-slide cohort involved peptide yield estimation. Colorimetric assays are commonly used in bulk proteomics to quantify total peptide yield and normalize injection amounts; however, they require relatively large amounts of material (several hundred nanograms), and their accuracy can be compromised by interference from

non-peptide components in FFPE samples. To address these limitations, we employed an MS-based estimation strategy similar to that described by Tüshaus et al.⁷. Briefly, single-slide cohort samples were analyzed using a fixed injection volume alongside HeLa standards (10–700 ng injections). A HeLa standard curve was generated using summed MS1 precursor signal, and peptide amounts in clinical samples were calculated from this curve (Fig. 1B). We refer to these values as “MS-based quantity of matched

peptides", as this estimate does not account for fragmented, cross-linked, or otherwise non-canonical peptides.

MS-based matched peptide yields in the single-slide cohort ranged from 55 ng to 1 µg per sample, except for one large cryobiopsy yielding more than 10 µg (Fig. 1C). These values were substantially lower than those obtained from the multi-section cohort, in which 5–250 µg of colorimetrically quantified peptides were extracted (Supplementary Fig. 1C). Notably, biopsy size showed a moderate-to-strong correlation with the estimated peptide yield (Supplementary Fig. 1D). Although biopsy length cannot be used to reliably normalize injection amounts for full MS analyses, these correlations can inform pre-run peptide yield estimation to ensure the samples fall within the HeLa standard curve. Importantly, using only 10% of each sample for the initial yield estimation left sufficient material for at least one full MS analysis of all single-slide cohort samples (Fig. 1C). The selection of injection amounts for full MS analyses balances a trade-off between analytical depth and long-term instrument performance, as well as the number of samples with sufficient material for equal injections: higher peptide loading can increase protein identifications but also necessitate more frequent instrument maintenance. Based on these considerations, we selected a 50-ng injection for the Astral instrument. For the timsTOF analysis, we used 200 ng when sufficient material was available; otherwise, we used the remaining sample amount, but no less than 50 ng.

Across the full single-slide cohort analyses, we obtained deep proteomic coverage, identifying up to approximately 8000 proteins (gene-centric) per sample and 10,000 proteins across the cohort (Fig. 1D–F). The median number of proteins was 7097 and 6629 in the Astral and timsTOF datasets, respectively. While the analytical depth in the single-slide cohort was significantly lower than in the multi-section cohort (median 7097 vs 7557 and 6629 vs 6980 proteins in the Astral and timsTOF datasets, respectively; Fig. 1D and Supplementary Fig. 1E), higher injection amounts, as expected, increased the number of identifications (Supplementary Fig. 1F, G). This analytical depth reflected a compromise aimed at maximizing the cohort size by minimizing the amount of injected material and using all available (unselected) samples. For comparison, another study optimized for maximal analytical depth in rapid DIA analyses of FFPE tissues reported a median of approximately 9100 identified proteins when injecting 1 µg of peptides from a small cohort of 13 lung tumors⁸.

We next evaluated the effect of skin keratin contamination on analytical depth and quantification validity. Skin keratin levels showed a strong negative correlation with the number of identified proteins, an expected effect for high-abundance proteins (Supplementary Fig. 2A). Consistently, this effect was even more pronounced in a broader analysis of the 20 most abundant proteins (Supplementary Fig. 2A). In our cohort, HBB (hemoglobin subunit beta) was a notable contributor to MS signal, highlighting the well-known challenge of plasma and blood contamination in MS samples (Supplementary Fig. 2B). Nonetheless, despite the lack of preselection for MS analysis, over 90% of samples yielded more than 5,000 identified proteins, with only six of 68 samples classified as outliers in analytical depth.

To exemplify quantitative validity and assess the impact of high abundance proteins on quantification, we evaluated concordance among six proteins in the MCM complex, whose abundances are expected to correlate¹. Across the cohort, quantitative agreement was strong, with low overall variability among MCM complex members (Supplementary Fig. 2C, D). However, standard deviation between MCM proteins tended to increase in samples with higher levels of skin keratins. This suggests that caution should be exercised when interpreting quantitative data from low quality samples such as those contaminated with blood, plasma, or skin keratins.

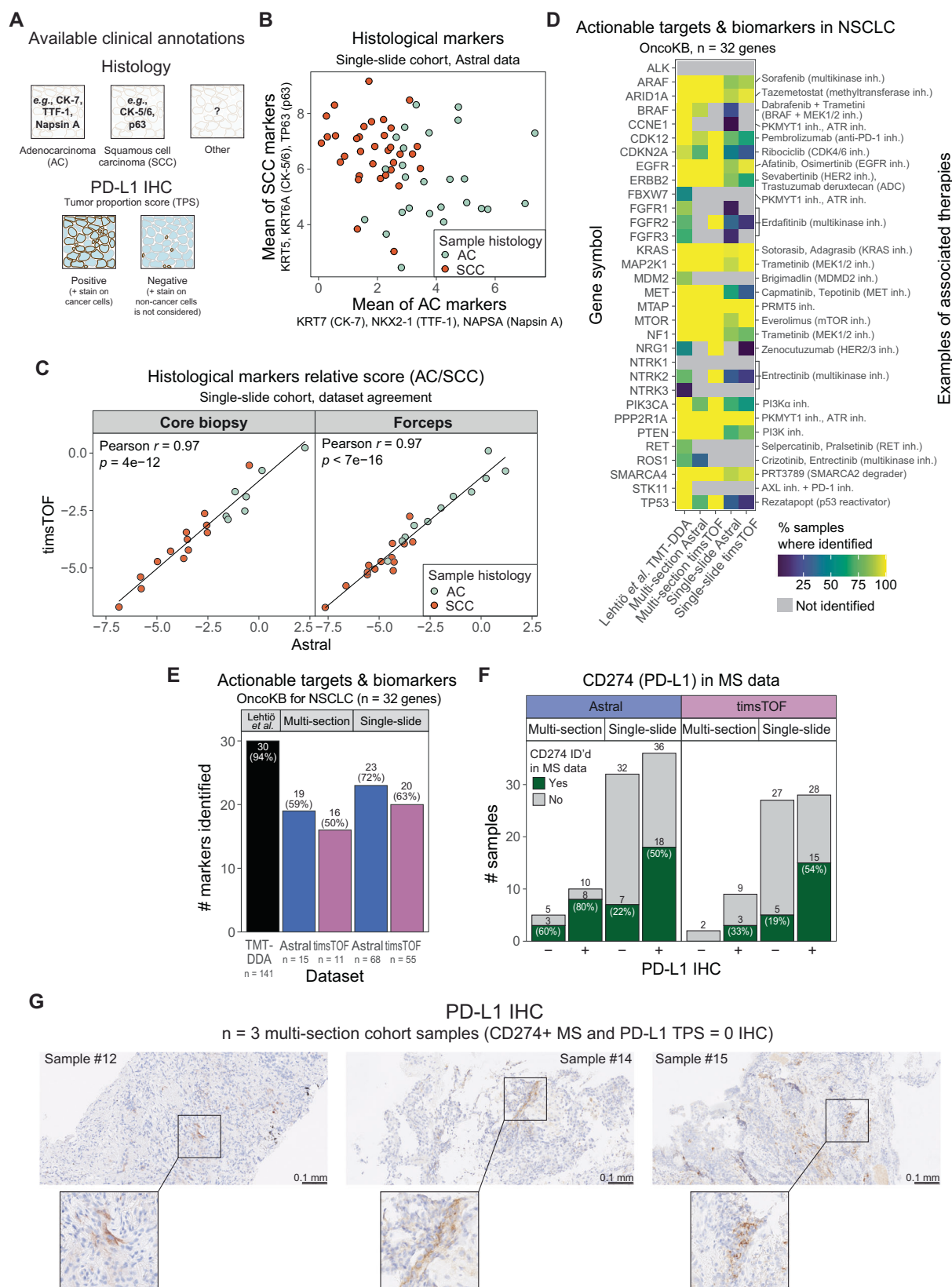
We next evaluated the biological and clinical relevance of the generated proteomic profiles. To provide biological context, we integrated publicly available protein lists of interest and clinical annotations with our MS data (Figs. 2A, E and 3D). MS-based quantification of histological markers routinely used in pathology to distinguish adenocarcinoma (AC) from

squamous cell carcinoma (SCC) enabled resolution of AC and SCC samples, with strong concordance between the Astral and timsTOF datasets (Fig. 2B, C and Supplementary Fig. 2E). While previous studies have reported disagreement between proteomics and IHC-based histology annotations for outlier cases⁵, the separation of the two histological groups in the single-slide cohort was nonetheless not as complete as expected. To investigate this further, we performed a subgroup analysis based on the biopsy type and observed a stronger distinction between AC and SCC samples in core biopsies than forceps biopsies (Fig. 2C). Forceps biopsies can pinch through the bronchial epithelium, introducing normal tissue into samples and thus dilute the tumor signal. This raises a concern regarding the informativeness of bulk proteomics of biopsy samples without corresponding H&E or IHC evaluations. Macro- or microdissection of the samples could reduce the signal contamination from adjacent normal tissue and enrich for tumor cells. Furthermore, the availability of the corresponding scanned images could help inform the proteomics data analysis regarding immune and stromal infiltration of the tumor, thus facilitating filtering or deconvoluting the data to reduce dilution of the tumor signal.

We next assessed identification of clinically actionable drug targets and predictive biomarkers listed in the OncoKB database for lung cancer¹¹. Across datasets, 63–72% of these markers were identified in the single-slide analyses, compared with 50–59% in the multi-section datasets, a difference most likely attributable to the small cohort size in the context of cancer heterogeneity (Fig. 2D, E). We then examined MS-based identification of PD-L1 in relation to tumor PD-L1 status determined by immunohistochemistry (IHC), which is routinely used to guide immunotherapy selection in lung cancer. CD274 (PD-L1) was identified in 80% of IHC-positive cases in the multi-section Astral dataset, but in only 33–54% of positive samples in the remaining datasets (Fig. 2F), suggesting limited sensitivity for CD274 detection in FFPE samples using the current workflows. Overall, CD274 was detected in 25–75% of samples across datasets. Notably, most identifications occurred in PD-L1 IHC-positive tumors (Fig. 2F), whereas the remaining detections likely reflect PD-L1 expression by non-cancer cells in the tumor microenvironment (which is not captured by the IHC scoring, Fig. 2G), or tumor heterogeneity, as the PD-L1 IHC staining was performed on tissue sections distinct from those used for MS analysis.

Individual biomarkers such as PD-L1 can inform clinical decision making; however, multivariate biomarkers are increasingly recognized as required to capture the complexity of tumor biology, therapy response, and resistance mechanisms for more precise therapy selection¹². The ability to generate data at scale for quantitative analysis of marker sets and molecular signatures is a key strength of MS-based proteomics. Here, as a proof-of-concept, we demonstrate signature-level analyses using the tumor inflammation signature (TIS) composed of 18 markers proposed by Ayers et al.¹³, which characterizes the tumor immune microenvironment (TIME), a key determinant of response to immunotherapy (Fig. 3A, B)¹⁴. Although originally defined at the mRNA-level data and now implemented in RNA-based gene expression assays, proteomics offers several advantages over transcriptomics. Proteome-level information reflects the final molecular tumor phenotype and drug target levels more accurately due to frequently low mRNA-protein correlations^{5,15}. Moreover, proteins are more stable than RNA, enabling more robust profiling of archived material¹⁶. In the single-slide cohort, the TIS score showed strong agreement between the Astral and timsTOF datasets (Fig. 3C). Notably, although the TIS scores were significantly higher in PD-L1-positive tumors than in PD-L1-negative, the difference was small, and the large variability in both groups highlights that TIME assessment provides information orthogonal to PD-L1 status, in agreement with what has been previously shown⁵. However, the ability to interpret the derived TIS scores is limited by the absence of tumor cellularity and histopathological assessment information.

Beyond molecular signature quantification, in-depth proteomics enables exploratory analyses; therefore, we next examined immune-related signals at a broader scale. Using immune cell-type gene sets published by



Charoentong et al.¹⁷, including 782 proteins as a reference, we detected 58% of the markers in the single-slide cohort (Fig. 3D). We then stratified the single-slide cohort by abundance of the pan-immune cell marker PTPRC (CD45; Fig. 3E) and compared proteomic profiles between the top and bottom tertiles (Fig. 3F–H). In the Astral dataset, 25% of identified immune

cell markers were significantly elevated in the PTPRC-high group, whereas only 1% were significantly lower. Finally, gene-set enrichment analysis (GSEA) using immune-related MSigDB hallmark gene sets¹⁸, as expected, revealed significant enrichment of IFN γ and inflammatory response pathways in the PTPRC-high group (Fig. 2I).

Fig. 2 | MS-based analysis of single FFPE slides provides biologically sound data and identifies clinically actionable targets. **A** Schematic overview of the available clinical annotations used to provide context for the proteomic observations. **B** MS-based quantitative signature of histological markers in adenocarcinoma (AC, $n = 26$) and squamous cell carcinoma (SCC, $n = 34$) samples in the single-slide cohort Astral dataset. **C** Relative levels of histological markers (log-transformed adenocarcinoma marker score minus log-transformed squamous cell marker score, as shown in **B**) in core biopsies (left; $n = 20$) and forceps biopsies (right; $n = 25$) from the single-slide cohort in the Astral and timsTOF datasets. Identification of actionable drug targets and predictive biomarkers in non-small-cell lung cancer (NSCLC) as listed in the

OncoKB database with **D** annotation of the sample percentage where identified and **E** summary statistics. An in-depth TMT-DDA dataset of 141 NSCLC tumors covering 13,975 proteins (Lehtiö et al.⁵) was used as a reference proteomics dataset. **F** Identification of CD274 (PD-L1) in MS proteomics data stratified by PD-L1 status, as determined by tumor proportion score (TPS) assessment performed in clinical routine, where TPS > 0 is positive (+) and TPS = 0 is negative (–) PD-L1 expression. **G** Representative PD-L1 immunohistochemistry (IHC) images from three multi-section cohort samples with tumor proportion score (TPS) = 0, but CD274 identified in MS proteomics data. PD-L1 staining is absent in tumor cells; positive staining is observed in macrophages and other stromal cells.

In conclusion, our study demonstrates that state-of-the-art MS instrumentation and optimized workflows enable rapid, in-depth proteomic analysis of clinical FFPE specimens, including small biopsies and limited material on single glass slides. Although multiple sections remain advantageous for maximizing analytical depth, single-slide analyses can still provide biologically and clinically informative data suitable for both targeted and exploratory applications.

We identified contamination from adjacent normal tissue and/or high abundance proteins to be key factors that determine sample quality in FFPE cancer biopsies on glass slides. These challenges may be mitigated through macrodissection and the use of freshly cut sections, respectively. Alternatively, laser capture microdissection (LCM) allows for more focused research questions through the selective analysis of tumor cells or other specific cell types^{10,19,20}. Regions of interest can be selected either through manual annotation by a clinical pathologist, which limits throughput, or through automated image analysis, which requires expensive specialized equipment. While these techniques represent exciting developments in proteomics and enable unprecedented spatial resolution, they offer lower throughput and may be less compatible with clinical routines or large cohort studies. Nevertheless, bulk proteomics can be further strengthened by integrating complementary imaging and cell composition data. Based on our observations, we recommend careful handling of slides and FFPE sections to minimize contamination from skin and dust and, if possible, use macrodissection to enrich tumor regions. Furthermore, histological assessment of adjacent sections is crucial to support more accurate interpretation of proteomics data.

By demonstrating the quantification of clinically actionable proteins, immune-related markers, and multivariate signatures, our findings highlight the value of proteomics for capturing functional tumor phenotypes beyond single-marker assessments, consistent with growing evidence that complex biomarkers are essential for precision oncology. Importantly, we show that these capabilities extend to archived FFPE biopsies derived from unresectable cases, enabling retrospective analyses of clinically annotated cohorts despite limited tissue availability. This opens opportunities for detailed proteome analysis of previously underutilized specimen collections of biopsies from late-stage patient cohorts, a patient population where new oncology drugs are typically initially tested. Our study demonstrated that clinical proteomics is a practical and powerful modality for precision medicine and cancer research.

Methods

Clinical sample collection

The study was performed in accordance with the Declaration of Helsinki. The multi-section FFPE cohort comprised of 15 samples collected from lung cancer patients at the Karolinska University Hospital in Solna, Sweden. The surgical samples originated from resected lung tumors during routine treatment of patients. The biopsy samples (core needle or forceps) were collected during diagnostic procedures for suspected lung cancer patients. The study was approved by the Regional Ethical Review Board in Uppsala, Sweden (registration nos. 2021-01931 and 2024-04958-02). Informed consent was obtained from all the patients. The FFPE sample preservation was performed as part of the clinical routine. The samples were used for

research only after routine diagnostic procedures were performed and sufficient material remained. Five consecutive sections (5 µm each) were then taken for and shortly before proteomic analysis. FFPE blocks were sectioned, and the sections were mounted on glass slides. Macrodissection was then performed to minimize non-tumor content by scraping off areas identified as non-tumor region by a qualified pathologist. The remaining parts of the sections were then scraped into Eppendorf tubes. All samples were processed rapidly, and the glass slides were not stored prior to analysis. PD-L1 staining was performed as part of clinical routine using SP263 (Ventana) antibody.

The single-slide FFPE cohort comprised of 68 samples collected from lung cancer patients at the Oslo University Hospital in Oslo, Norway. The biopsy samples (core needle, forceps, or cryobiopsy) were collected as part of the DART trial approved by the Regional Ethical Committee for Medical Health Research Ethics, REK South-East in Oslo, Norway (reference no. 48655). Informed consent was obtained from all the patients. The FFPE sample preservation was performed as part of the clinical routine with 4-µm tissue sections mounted on glass slides. The glass slides were initially prepared for other analyses, leftover slides were stored for several years, after which they were used for proteomic analyses.

The meta data for the samples, including histological and PD-L1 status annotations, peptide yields, generated datasets, and the number of identified proteins, is provided in Supplementary Data 1.

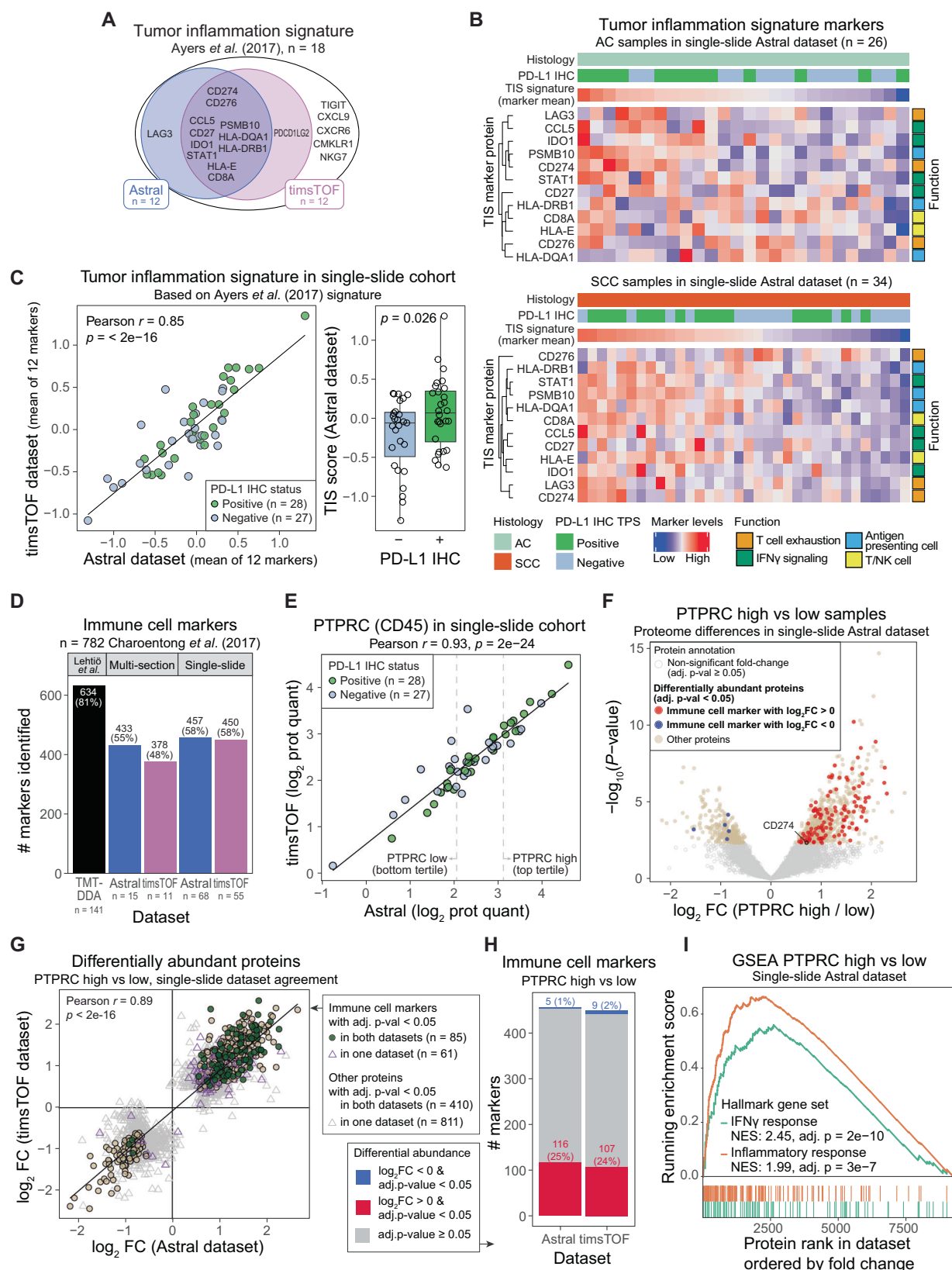
Sample preparation

For the single-slide cohort, the slides were pre-washed with 0.01% n-Dodecyl-B-D-Maltoside (DDM) by dipping and shaking the slide in the solution for approximately 10 s, rinsing in water twice, and drying. The FFPE material was then scraped off using a sterile scalpel into Eppendorf tubes.

All samples were prepared for proteomics analysis using the FenoPrep™ kit (Cat# 532001, FenoMark Diagnostics AB, Sweden) per the manufacturer's instructions. The method utilized xylene-free deparaffinization based on high-temperature, high-energy ultrasonication²¹. Briefly, the lysis buffer was added to the samples, heated under shaking (95 °C, 400 rpm, 30 min), after which the sample tubes were placed in the VialTweeter Ultrasonic processor (Hielscher Ultrasonics GmbH, Germany; settings: amplitude 100%, pulsation mode 100%, 10 cycles of 60 s on/30 s off). The samples were then heated under shaking (95 °C, 400 rpm, 30 min). The samples were then placed in the ultrasonic processor for 5 more cycles. The samples were centrifuged (13,000 × g, 10 min) and the supernatant was transferred for further processing. The proteins were alkylated. Thereafter, on-bead digestion using Lys-C and trypsin and clean-up using magnetic beads were performed. For the multi-section cohort, the peptide concentration was measured using a microBCA assay with HeLa digest as the standard. For the single-slide cohort, MS-based peptide yield estimation was performed (see next section).

MS-based peptide yield estimation using timsTOF HT

A timsTOF HT mass spectrometer (Bruker) connected to an EVOSEP LC system via the CaptiveSpray 2 source was used for DIA-PASEF LC-MS/MS analysis. Each sample was analyzed after transfer of 10%



peptide solution volume on Evotips following the manufacturer's instructions (except for one sample where 1% of the volume was used, loading amounts are provided in the [supplementary materials](#)). The peptides were separated on a PepSep C18 column (Bruker, 8 cm $\times$ 150 μ m, 1.5 μ m), using the 60SPD fixed method. For the dia-

PASEF analysis the window scheme was calculated using the py_diAID tool (<https://github.com/MannLabs/pydiAID>). The capillary voltage was set at 1500 V, stepping collision energy at 32, 40, 50 eV, dia-PASEF scan range 100–1700 m/z in positive mode, and IMS service ramp time of 100 ms.

Fig. 3 | MS-based proteomics of single FFPE slides quantifies molecular signatures and provides opportunities for in-depth exploratory analyses. **A** Markers from the tumor inflammation signature (TIS) proposed by Ayers et al.¹³ and identified in the single-slide cohort. **B** Scaled levels of TIS proteins in adenocarcinoma (AC, top) and squamous cell carcinoma (SCC, bottom) samples in the single-slide Astral dataset. Columns ordered by the signature marker mean; rows clustered by hierarchical clustering using Spearman correlation distance. PD-L1 immunohistochemistry (IHC) annotation like in Fig. 2. **C** Mean of scaled TIS protein levels in single-slide Astral and timsTOF datasets (left) and stratified by PD-L1 status (right). *P* value was calculated using two-sided Welch two-sample *t*-test. **D** Identification of immune cell markers published by Charoentong et al.¹⁷. The same Lehtiö et al.⁵ reference dataset was used as in Fig. 2. **E** Protein quantities of an immune cell marker PTPRC (CD45) in the single-slide cohort. Tertiles were defined based on the Astral

dataset and included samples not analyzed in the timsTOF dataset. **F** Differential abundance analysis using DEqMS comparing PTPRC high versus low tertiles in the single-slide cohort Astral dataset (*n* = 23 per group). Immune cell marker labels refer to the Charoentong et al.¹⁷ list. **G** Comparison of differentially abundant proteins like in **F** indicated in the Astral and/or timsTOF datasets (sample groups defined using Astral data). **H** Percentage of immune cell markers (Charoentong et al.¹⁷ list) that are differentially abundant in the PTPRC high versus low comparison in the Astral and/or timsTOF datasets. **I** Gene-set enrichment analysis (GSEA) using ranked fold changes in PTPRC high versus low tertile comparison in the single-slide cohort Astral dataset (shown in **F**), assessing enrichment of MSigDB interferon gamma (IFN γ) and inflammatory response hallmark gene sets¹⁸. FC fold change, NES normalized enrichment score.

DIA analysis using Orbitrap Astral

For the multi-section cohort, an Orbitrap Astral coupled to a Vanquish NEO System (Thermo Fisher Scientific) was used. The injection volume was 2 μ l for each sample containing 500 ng of peptides. The samples were trapped on a C18 guard-desalting column (Acclaim PepMap NEO, 300 μ m $\times$ 5 mm, nanoViper, C18, 5 μ m, 100 Å) and separated on a 25 cm long C18 column (25 cm Aurora Ultimate XT column, C18, 1.7 μ m, 75 μ m $\times$ 25 cm). The nanocapillary solvent A was 0.1% formic acid in water; and solvent B was 0.1% formic acid in acetonitrile. At a constant flow of 0.25 μ l/min, the curved gradient went from 6% B up to 40% B in 32 min, followed by a steep increase to 100% B in 5 min.

For the single-slide cohort, an Orbitrap Astral System (Thermo Fisher Scientific) coupled to an EVOSEP LC system was used. Each sample was analyzed after the transfer of 50 ng of peptides on Evotips following the manufacturer's instructions. The peptides were separated on a PepSep C18 column (Bruker, 15 cm $\times$ 75 μ m, 1.5 μ m), using the 30SPD fixed method with a total method duration of 44 min.

The standard MS parameters for DIA acquisition were set as follows: a spray voltage of 2000 V, no sheath or auxiliary gas flow, and a heated capillary maintained at 280 °C. Full-scan mass spectra were acquired with a scan range of 380–980 *m/z*, an AGC target value of 300%, maximum injection time (IT) of 10 ms, and a resolution of 240,000. The DIA windows were set to 2 *m/z* in a precursor mass range of 380–980 *m/z* and subjected to fragmentation (nCE: 25%) with a scan range of 145–1450 *m/z*, maximum IT of 3 ms, AGC target value of 500%.

DIA analysis using timsTOF HT

A timsTOF HT mass spectrometer (Bruker) connected to an EVOSEP LC system via the CaptiveSpray 2 source was used for DIA-PASEF LC-MS/MS analysis.

For the multi-section cohort, 11 samples were analyzed after transfer of 500 ng of peptides on Evotips following the manufacturer's instructions. The peptides were separated on a C18 IonOpticks column (particle size 1.6 μ m, 75 μ m ID, 15 cm length) at a flow rate of 0.25 μ l/min; solvent A (0.1% formic acid in water), solvent B (0.1% formic acid in acetonitrile), using the 30SPD fixed method with a total method duration of 44 min.

For the single-slide cohort, 56 samples sample was analyzed after transfer of 50–200 ng of peptides on Evotips following the manufacturer's instructions (loading amounts are provided in the Supplementary Data 1). The peptides were separated on a PepSep C18 column (Bruker, 15 cm $\times$ 75 μ m, 1.5 μ m), using the 30SPD fixed method with a total method duration of 44 min.

For the dia-PASEF analysis the window scheme was calculated using the py_diAID tool (<https://github.com/MannLabs/pydiaid>). The capillary voltage was set at 1500 V, stepping collision energy at 32, 40, 50 eV, dia-PASEF scan range 100–1700 *m/z* in positive mode, and IMS service ramp time of 100 ms.

One injection in the single-slide cohort failed, resulting in the final dataset of 55 samples.

DIA analysis using Orbitrap Exploris 480

An Orbitrap Exploris 480 coupled to a Dionex UltiMate™ 3000 RSLCnano System was used System (Thermo Fisher Scientific). The injection volume was 2 μ l for each sample containing 2 μ l. Samples were trapped on a C18 guard-desalting column (Acclaim PepMap 100, 75 μ m $\times$ 2 cm, nanoViper, C18, 5 μ m, 100 Å) and separated on a 25-cm long C18 column (25 cm Aurora Ultimate XT column, C18, 1.7 μ m, 75 μ m $\times$ 25 cm). The nanocapillary solvent A was 0.1% formic acid in water; and solvent B was 0.1% formic acid in acetonitrile. At a constant flow of 0.25 μ l/min, the curved gradient went from 6% B up to 40% B in 120 min followed by a steep increase to 100% B in 5 min. The standard MS parameters for DIA acquisition were set as follows: a spray voltage of 1900 V, no sheath or auxiliary gas flow, and a heated capillary maintained at 275 °C. Full-scan mass spectra were acquired with a scan range of 375–1175 *m/z*, an AGC target value of 100%, maximum injection time (IT) of 100 ms, and a resolution of 120,000. The DIA windows were set to 25 *m/z* in a precursor mass range of 375–1175 *m/z* and subjected to fragmentation (nCE: 28%) by HCD with a scan range of 145–1450 *m/z*, maximum IT of 100 ms, AGC target value of 100%, and a resolution of 15,000.

DIA data search and preprocessing

DIA data were analyzed in Spectronaut (Biognosys, v.20) using directDIA mode and searched against ENSEMBL protein database (GRCh38.pep.all). The peptide quantities were summarized into proteins (gene-centric) using the MaxLFQ algorithm and only gene-specific peptides. The data was normalized using local normalization. Both imputed (global imputing) and non-imputed data were exported. Additionally, precursor-level non-imputed and non-normalized data, as well as the total ion chromatograms, were exported. Non-imputed data were used for identification analyses, while imputed exports were used as quantitative data. Total ion chromatograms and precursor-level exports were used for MS signal calculations; additionally, precursor-level data were used to assess the contribution of high-abundance proteins.

Downstream bioinformatics analysis

All downstream bioinformatics analyses were performed in R (v. 4.5.2). To calculate the sample-wise normalized protein quantities, the MS2 area values exported from Spectronaut were divided by the median of all quantification in the respective sample; thereafter, the quantities were log₂ transformed. For high abundance protein analysis, precursor-level MS1 data were exported. For the total ion chromatogram area under the curve (TIC AUC) analysis, the TIC chart data were exported from Spectronaut. Thereafter, the AUC was calculated using the AUC() function (trapezoid method) from the DescTools R package (v. 0.99.60). The differential abundance analysis was performed using the DEqMS package^{22,23} (v. 1.26.0) using imputed data and median precursor counts per protein. *P* values were adjusted using the BH method. The gene set enrichment analysis (GSEA) was performed using the GSEA() function from the clusterProfiler R package²⁴ (v. 4.16.0) based on the ranked fold-change values derived from the DEqMS analysis. The hallmark gene sets originated from the MSigDB

database¹⁸ (msigdb R package v. 25.1.1). All the used statistical tests and the corresponding sample sizes are specified in the figures and figure legends.

Public domain datasets

The Charoentong et al. immune cell markers were taken from the supplementary tables of the original publication¹⁷. The IFN gamma response and inflammatory response hallmark gene sets originated from the MSigDB¹⁸.

Data availability

MS proteomics data have been deposited on the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifiers PXD071948 (multi-section cohort) and PXD072014 (single-slide cohort).

Received: 20 January 2026; Accepted: 15 May 2026;

Published online: 29 May 2026

References

- Johansson, H. J. et al. Breast cancer quantitative proteome and proteogenomic landscape. *Nat. Commun.* **10**, 1600 (2019).
- Vasaikar, S. et al. Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**, 1035–1049.e1019 (2019).
- Gillette, M. A. et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* **182**, 200–225.e235 (2020).
- Krug, K. et al. Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* **183**, 1436–1456.e1431 (2020).
- Lehtiö, J. et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. *Nat. Cancer* **2**, 1224–1242 (2021).
- Guo, T., Steen, J. A. & Mann, M. Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature* **638**, 901–911 (2025).
- Tüshaus, J. et al. Towards routine proteome profiling of FFPE tissue: insights from a 1,220-case pan-cancer study. *EMBO J.* **44**, 304–329 (2025).
- Haines, M. et al. High-throughput proteomic and phosphoproteomic analysis of formalin-fixed paraffin-embedded tissues. *Mol. Cell. Proteom* **24**, 101044 (2025).
- Adachi, Y. et al. Time-enduring proteomic fidelity in over 30-year-old FFPE tissues: distinct proteomic signatures of hepatocellular carcinoma and adjacent non-tumor liver tissue. *Proteomics* **26**, 57–67 (2025).
- Woo, J. et al. High-sensitivity spatial proteomics for pancreatic cancer progression analysis. *Anal. Chem.* **98**, 5903–5911 (2026).
- Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* **1**, PO.17.00011 (2017).
- Passaro, A. et al. Cancer biomarkers: emerging trends and clinical implications for personalized treatment. *Cell* **187**, 1617–1635 (2024).
- Ayers, M. et al. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Investig.* **127**, 2930–2940 (2017).
- Binnewies, M. et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
- Liu, Y., Beyer, A. & Aebersold, R. On the dependency of cellular protein levels on mRNA abundance. *Cell* **165**, 535–550 (2016).
- Shao, W. et al. Comparative analysis of mRNA and protein degradation in prostate tissues indicates high stability of proteins. *Nat. Commun.* **10**, 2524 (2019).
- Charoentong, P. et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).
- Liberzon, A. et al. The molecular signatures Database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).

- Mund, A. et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).
- Rosenberger, F. A. et al. Deep Visual Proteomics maps proteotoxicity in a genetic liver disease. *Nature* **642**, 484–491 (2025).
- Pujari, G. P. et al. A high-throughput workflow for FFPE tissue proteomics. *J. Am. Soc. Mass. Spectrom* **34**, 1225–1229 (2023).
- Zhu, Y., Berkovska, O., Wang, L. et al. Differential protein expression analysis of quantitative mass spectrometry data using DEqMS. *Nat Protoc* <https://doi.org/10.1038/s41596-026-01349-7> (2026).
- Zhu, Y. et al. DEqMS: a method for accurate variance estimation in differential protein expression analysis. *Mol. Cell. Proteom.* **19**, 1047–1057 (2020).
- Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

Acknowledgements

We thank FenoMark Diagnostics for providing the FenoPrep™ kit for sample preparation. This study was funded by Cancerfonden (23 2819 Pj, 25 4349 Pj), Vetenskapsrådet (ERA PerMed 2021-00754, 2024-03294), Radiumhemmets Forskningsfonder (244203), Karolinska Institutet (2023-01372), Stockholm County council (ALF FoUI-1000396), Vinnova (ProMS 2024-01137), EU DART (965397), EU Rescuer (847912). The single-slide cohort originated from the DART trial, which was a researcher-initiated study and funded by grants from the South-Eastern Norway Regional Health Authority (grant nos: 2019119 and 26011) and AstraZeneca. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Author contributions

The study was conceptualized by O.B., L.M.O., and J.L. The study was supervised by L.M.O. Clinical samples were collected and annotated by I.S., V.D.H., and Å.H. Sample preparation was performed by N.E.Ö., G.M., and M.N. Mass spectrometry data generation was performed by G.M. and N.E.Ö. Data analysis was performed by O.B. The manuscript was written by O.B. All authors approved the final version of the manuscript.

Funding

Open access funding provided by Karolinska Institute.

Competing interests

L.M.O., J.L., and G.M. are shareholders of FenoMark Diagnostics that provides proteomics solutions to medical professionals, including a kit that was used in this study. Å.H. has participated in meetings (talks/advice/discussions) arranged by AbbVie, Takeda, AstraZeneca, Roche, Pfizer, Janssen, Eli Lilly, BMS, PierreFabre, Bayer, MSD, Novartis, Merck, Sanofi, Medcover, BeOne. All fees were paid to OUS. Å.H. has received support for research projects and studies from Roche, AstraZeneca, Novartis, Incyte, Eli Lilly, Bristol-Myers Squibb, Ultimovacs, Merck, GSK, Medcover, BeOne, and Johnson & Johnson. V.D.H. has participated in lectures organized by Bristol-Myers Squibb, MSD, AstraZeneca, Johnson&Johnson, BeOne, and Regeneron; and has participated in advisory boards of MSD, AstraZeneca, Bristol-Myers Squibb, BeOne, GSK, Immedica, Merck, and Novartis.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41698-026-01517-8>.

Correspondence and requests for materials should be addressed to Janne Lehtiö or Lukas M. Orre.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026