Filter-aided expansion proteomics

Zhen Dong[1,2,3#], Yi Zhu[1,2,3#]*, Fangfei Zhang[1,2,3#], Chunlong Wu[4], Ting Chen[5], Jiayi Chen[1,2,3], Wenhao Jiang[1,2,3], Qi Xiao[1,2,3], Shu Zheng[5], Kiryl D. Piatkevich[2,3]*, Tiannan Guo[1,2,3]*

[1]Westlake Center for Intelligent Proteomics, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, Zhejiang Province, China.

[2]School of Life Sciences, Westlake University, Hangzhou, Zhejiang Province, China.

[3]Research Center for Industries of the Future, Westlake University, Hangzhou, Zhejiang Province, China.

[4]Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou, Zhejiang Province, China.

[5]Cancer Institute, Key Laboratory of Cancer Prevention and Intervention, Ministry of Education, Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China.

[#] These authors contributed equally.

[*]Corresponding authors: zhuyi@westlake.edu.cn; kiryl.piatkevich@westlake.edu.cn; guotiannan@westlake.edu.cn

Abstract

Integration of hydrogel-based tissue expansion and mass spectrometry-based proteomics enables high spatial resolution analysis of global protein expression in tissues. Here, we introduce a filter-aided tissue expansion proteomics (FAXP) method with optimized hydrogel preparation and treatment steps for spatial proteomics analysis of formalin-fixed paraffin-embedded (FFPE) specimens. Compared with our original ProteomEx method, FAXP, partially automated with robotics, exhibited 14.5-fold higher volumetric resolution, increased peptide yield leading to 250% more protein identifications, and 50% time consumption for sample preparation. We demonstrated its application in clinical colorectal tissue samples. We further integrated it with laser capture microdissection for proteomic analysis of subcellular organelles.

Keywords: Spatial proteomics, filter-aided sample preparation, FFPE tissue, tissue expansion, colorectal cancer, sub-cellular proteomics, laser capture microdissection

Introduction

To study protein expression in space at the proteome level, two major strategies have been developed. The first one is mass spectrometry imaging (MSI)[1], which usually use Matrix-Assisted Laser Desorption-Ionization (MALDI) to convert proteins or peptides in a particular region into gas-phase ions before MS analysis[2,3]. The other emerging spatial proteomics technologies are developed as the sensitivity of MS-based proteomics increases to single cell level. For example, selected regions of tissue could be isolated using laser-capture microdissection (LCM)[4-7], which might be coupled with microfluidics[8,9] or a 3D-printed device[10] or tissue clearing[11], followed by highly sensitive MS-based proteomics analysis.

Proteins on the surface of selected tissue region could also be extracted using a fine liquid junction in line with LC-MS/MS hardware[12,13].

Tissue expansion technology has been recently introduced to MS-based proteomics[14-17]. The concept of expansion proteomics (ProteomEx) is based on the expansion microscopy approaches[18-20], like protein-retention Expansion Microscopy[21], appended with MS-based proteomics analysis. First, paraformaldehyde (PFA)-fixed tissue section is treated with N-succinimidyl acrylate (NSA) to modify primary amino groups of proteins with acryl group, capable of co-polymerizing with hydrogel monomer. Then, treated tissue is infused with hydrogel monomers, and *in situ* polymerization reaction is initiated to form a tissue-hydrogel composite. To enable isotropic expansion, the formed tissue-hydrogel composite is mechanically homogenized by SDS-containing buffer at elevated temperatures (95ºC). Expanded sample is manually microdissected, and excised pieces of hydrogel are subjected to in-gel reduction/alkylation and tryptic digestion to extract peptides for LC-MS/MS analysis. Importantly, chemical treatment of biological tissue does not introduce any modifications to the extracted peptides, yielding proteomics data indistinguishable from conventional sample preparation methods. Furthermore, unlike the LCM-based techniques, ProteomEx does not require any sophisticated or custom equipment and utilizes only commercially available and highly accessible reagents and accessories for sample preparation. However, the lateral and volumetric resolution of the original ProteomEx protocol remains limited to ~160 µm and ~0.61 nL, respectively[14]. Due to the extended sample treatment timeline (~54 hours, not including MS analysis) and multiple manual steps, processing multiple samples in parallel is challenging, precluding large-scale analysis. Furthermore, the compatibility of ProteomEx with archived clinical samples was not previously demonstrated. Addressing these challenges and limitations of ProteomEx can significantly broaden its utility and promote large-scale spatial proteomic studies.

Here, we introduced a method termed filter-aided expansion proteomics (FAXP), compatible with multiple types of formalin-fixed paraffin-embedded (FFPE) tissues, including archived clinical samples. Compared to the original protocol, FAXP enables 2.2-/14.5-fold higher lateral/volumetric resolution with twice faster timeline. When integrated with laser capture microdissection, FAXP enables proteomics analysis of subcellular organelles. The FAXP method offers better reproducibility for subnanoliter sample volume and higher throughput, capable of processing 96 samples per batch across multiple batches in parallel. These enhancements stem from optimized procedures used in the sample expansion procedures and the development of filter-aided in-gel digestion for peptide retrieval. We further demonstrated its application in human colorectal cancer FFPE samples utilizing a hybrid library search strategy[22] to identify 6790 proteins associated with the advancement of colorectal tumors through four stages alongside a machine learning classifier for feature selection and spatial visualizations of the specific features.

Results

Acceleration of the ProteomEx workflow

Since the ProteomEx method[14] does not allow effective analysis of FFPE tissues, which are the most common specimens in clinic, we sought to optimize and enhance the original ProteomEx protocol to achieve several key features, including 1) compatibility with archived clinical samples; 2) faster timeline; 3) higher sample preparation throughput; 4) higher lateral and volumetric resolution. To mimic the popular format of archived clinical samples, we used 10 µm FFPE mouse liver tissue (unless otherwise stated), since it is one of the most homogenous tissues in terms of protein distribution, providing less biologically biased

comparison between small samples across tissue sections[5,23].

First, we asked whether the timeline of sample preparation and peptide recovery can be accelerated without sacrificing proteome coverage. We found that protein anchoring, hydrogel embedding, homogenization, and Coomassie staining, the most time-consuming steps in the original protocol, could be significantly accelerated (in total to 11.2 hours vs. 40.5 hours in ProteomEx) without compromising peptide and protein identifications compared to ProteomEx (Figure S1; see Methods section for the protocol details). However, we noticed that autoclavation of tissue-hydrogel composite at 105-120°C, which we employed for rapid homogenization, resulted in softer hydrogels after full expansion compared to the original homogenization treatment conditions. As softer hydrogels were harder to handle during dissection step and could potentially introduce larger distortion to the expanded tissue, we increased the hydrogel cross-linker ratio by 17.5 folds (Table S1), which enhanced hydrogel sturdiness similarly to that observed for other hydrogels reported before[24]. To evaluate tissue expansion using the altered hydrogel composition, we quantified the linear expansion factor and isotropy of expansion on cellular and whole tissue slice level using protein anchoring time as a variable (Figure 1A-G). At the cellular level, under 1-hour anchoring condition the linear expansion factors of cell nuclei visualized via DAPI staining were comparable to those measured at the whole expanded tissue (linear expansion factor (LEF) 4.42±0.31 vs. 4.69±0.17 for individual nuclei vs. whole expanded tissue, respectively; Figure 1E), indicating a consistent and uniform expansion pattern that extends from the cell nuclei to the whole tissue structure. However, extended anchoring time (2 hours) resulted in reduced expansion factors (LEF=3.10±0.58, p=4.50e-6) at the cellular level compared to those measured at the whole expanded tissue level (Figure 1E). Using 1-hour anchoring step for all further experiments, we computed the root-mean-square (RMS) measurement error in feature-length measurements after non-rigid registration of pre- and post-expansion tissue across length scales of up to 1500 µm. The analysis revealed that the RMS errors were approximately 5.5% of the measurement distance, which was slightly better than that for the original protocol (Figure 1F, G). The comparable RMS errors were also measured for tissues with rich extracellular matrix such as kidney (Figure S2), indicating a high degree of accuracy suitable for precise mapping of spatial proteome distribution onto pre-expanded tissue morphology.

We then accelerated the ProteomEx timeline and streamlined peptide recovery from microdissected tissue-hydrogel samples, addressing the most laborious and unreliable steps due to multiple manual procedures (see Note S1, Figure S3, Figure S4, and Table S2 for details). 20 mM DTT/55 mM IAA can now be performed on the entire tissue section after homogenization, reducing the need for extra manual treatment steps during in-gel digestion and streamlining the protocol.

Comparison of conventional in-gel digestion with filter-aided in-gel digestion

The original ProteomEx protocol had a limited lateral/volumetric resolution, achieving only ~160 µm/~0.61 nL, primarily due to difficulties in manually handling hydrogel samples smaller than 1 mm. Specifically, during peptide recovery small hydrogel pieces (<1 mm in dimension) underwent shrinkage caused by multiple buffer exchange steps resulting in attendant sample lost during pipetting. To overcome this limitation, we suggested immobilizing the hydrogel sample on a membrane placed into a pipette tip. We first assessed the feasibility of the proposed filter-aided in-gel digestion method (also referred to as in-tip digestion) compared to the original in-gel digestion (Figure 2A). Analysis of 1.55 nL of the tissue using timsTOF Pro 1-hour DDA-MS revealed that the in-tip digestion provided significantly higher number of peptide and protein identifications compared to in-gel

digestion (increased by 97.7% and 30.6%, respectively; Figure 2B). Correspondingly, 75.0% of peptides and 97.5% of proteins identified in in-gel digestion were also identified in in-tip digestion, and it offered an extra identification of 56.7% more peptides and 20.5% more proteins, indicating a high consistency and performance improvement in proteome coverage (Figure 2C). In-tip digestion also displayed superior capability in identification of longer peptides (Figure 2D) without preference for peptide hydrophobicity (Figure 2E), indicating an improved peptide coverage for in-tip digestion compared with in-gel digestion. Furthermore, the in-tip digestion led to a significant improvement in protein digestion efficiency measured by missed cleavage rate, with an average of around 17.2%, compared to the in-gel digestion rate of 23.8% (Figure 2F). As an additional advantage, in-tip digestion successfully identified more mouse liver-specific markers[25], including all discovered through in-gel digestion, and additionally six markers (ELOVL6, SLC35B1, PVRL1, TM6SF2, ABCA8A, and ADIPOR2) with lower abundance (Figure 2G). In-tip digestion generated more reproducible quantification outcomes with a significant improvement in coefficient of variance (less than 0.2 at peptide and protein levels; Figure 2H) and Pearson correlation (close to 1.0 at peptide and protein levels; Figure 2I). Both protocols exhibited analogous distributions of protein types and subcellular localizations as revealed by the pathway enrichment analysis using the identified proteins, indicating there was no significant biological variations from in-tip digestion (Figure 2J). Altogether, these results confirmed that filter-aided in-gel digestion method improved sample manipulation while enhancing coverage and reproducibility of proteome analysis.

Improving volume-dependent limit of tissue micro-sampling

The combination of filter-aided in-tip digestion with accelerated ProteomEx workflow enabled improved proteome coverage while increasing the throughput of sample preparation. To validate whether in-tip digestion improves lateral/volumetric resolution of ProteomEx, we investigated the volume-dependent limit of tissue microsampling using the optimized workflow with peptide recovery step performed following either in-gel or in-tip protocol (Figure 2K). By manually dissecting tissue microsamples with actual volumes ranging from 0.042 nL to 3.054 nL (corresponds to lateral resolution from 73 µm to 624 µm), we observed that in-tip digestion under DDA-MS and DIA-MS identified a range of 5519 to 28,931 peptides and 19,499 to 48,330 peptides, respectively (Figure 2L). In-tip digestion consistently exhibited superior peptide and protein identification performance compared to in-gel digestion in both DDA and DIA modes, except for when processing tissue samples with a volume of 3.054 nL, indicating in-tip digestion was more prone for minute sample preparation. There was a saturating plateau observed for tissue volumes between 0.763 nL and 1.357 nL using in-tip digestion with a leveling off at higher volumes. In turn, due to sample loss during peptide extraction in-gel digestion encountered a technical constraint of processing tissue volumes below 0.339 nL, which was 8 times larger than the smallest volume we were able to process with in-tip digestion. For tissue volume of 0.339 nL (the smallest volume we could process with in-gel digestion), in-tip digestion outperformed in-gel digestion by around 250% in protein identification resulting in identification of 2229 and 4844 proteins in DDA and DIA mode, respectively. Remarkably, in-tip digestion displayed compatibility with tissue samples as small as 0.042 nL in volume, effectively achieving a lateral resolution of 73 µm on 10 µm thick sections, which was equivalent to 8 murine hepatocytes[26]. These results demonstrated that in-tip digestion improved lateral and volumetric resolution by 2.85 and 8.07-fold, respectively, while providing 8.07/2.55-fold higher peptide/protein identification compared to the ProteomEx protocol with in-gel digestion under its lower limit tissue volume using DIA mode.

In addition, the approach effectively targets single nucleus by combining LCM (Figure 3 and Supplementary Video 1), resulting in detection of an average of 550 proteins with an Astral mass spectrometer. When compared to the 0.042 nL In-tip DIA-MS (Figure S5), this method identified 258 specific proteins (Figure S5A) and demonstrated significantly higher ratios for nucleus and transcription regulator-related proteins, exhibiting a 4.26- and 3.59-fold increase, respectively (Figure S5B). Furthermore, the nuclei-specific pathway enrichment analysis revealed elevated enrichment in nucleic activities (Figure S5C), such as mRNA processing and splicing pathways.

Application of FAXP on colorectal cancer patient samples

To further explore the utility of FAXP on clinical samples, we examined the proteome profiles of various subtype regions in colorectal cancer (CRC), namely N (normal), L (low-grade dysplasia), H (high-grade dysplasia), PC (peripheral carcinoma), CC (central carcinoma), and C (carcinoma), delineated using H&E staining by expert pathologist, aiming to examine spatial discrepancies at an individual patient level and evaluate variations within the same and between different subtypes across patients. We first confirmed the FAXP protocol enabled efficient Coomassie staining and isotropic expansion of the archive clinical samples (Figure 4A). Notably, the RMS error for CRC tissue was approximately 6% of the measurement distance, which was within the expected range based on mouse tissue expansion (Figure 4B). Subsequently, we performed visual-guided manual microdissection of the expanded CRC samples from three distinct patients (referred to as P1, P2, and P3) and collected samples as outlined in Figure 4C.

In total 131 microdissected tissue samples were processed using the FAXP protocol and analyzed with optimized diaPASEF-based MS settings (Table S3). Considering the protein identification, we opted for the hybrid-pan human hybrid library for subsequent quantitative analysis due to its capability to generate the highest number of identifications, resulting in the recognition of 58,038 precursors and 6790 proteins (Figure S6E). To probe differences among various groups, we categorized the identified proteins based on MS batch IDs, patient IDs, slice IDs, and disease subgroups (Figure 4D). Our results revealed a consistent number of protein identifications across batches, patients, and slices (Figure 4E; Figure S6F-G), demonstrating the reproducibility and robustness of the FAXP workflow. The systematic data analysis revealed that the number of identified proteins progressively increased with the severity of cancer stage. Specifically, number of protein identifications in L, H, C, PC, CC were 2.50%, 14.15%, 12.23%, 14.62%, and 16.61% higher, respectively, than that in normal tissue ($p = 0.36$, 8.76e-08, 1.82e-05, 7.79e-09 and 1.67e-07), with a trend for increasing number of proteins with cancer progression (Figure 4E).

In the context of quantifying differences between various stages in CRC, the study identified significantly differentially expressed proteins (DEPs) when comparing both progression stages to normal stage and adjacent stages in the progression process (Figure 4F-G; Figure S6H). Notably, there were more down-regulated DEPs observed in more advanced stages compared to less severe stages, although this trend did not hold for comparisons such as C vs. H and CC vs. PC (Figure 4F). The number of DEPs found aligned with the increasing identification difference observed during the progression stages (Figure 4E). Specifically, when examining more severe stages such as C and H compared to the normal stage, there was a similarity in the most regulated DEPs. These included upregulated DEPs such as CEAM5, IFM3, and S100P, as well as downregulated DEPs such as CLCA1 and FCGBP. In the context of tumor development regulation, CEAM5 was noted as an oncogene that promotes tumor progression and induces resistance to anoikis in colorectal carcinoma cells[27]. The expression of CEAM5 exhibited an increasing trend in pair-wise comparisons such as L vs. N

(FC = 2.43, p = 1.84e-03), H vs. N (FC = 10.17, p = 3.13e-09), and C vs. N (FC = 12.26, p = 3.13e-09). Besides, CLCA1, recognized as a tumor suppressor in CRC[28], showed a decreasing trend in comparisons including L vs. N (FC = 0.38, p = 3.08e-02), H vs. N (FC = 0.055, p = 1.14e-04), and C vs. N (FC = 0.072, p = 1.14e-04). These findings indicated the capability of the workflow used in the study to dissect the heterogeneity of different malignant changes within the same tissue slice.

To study the variations across tissues and patients, we calculated the coefficient variations of protein abundance between three consecutive slides of the same subtype (intra-patient variation) and among all punches within the same subtype (intra-subtype variation). The results showed that, across all subtypes, the CV values for intra-patient variation were higher than those for intra-subtype variation (Figure 5A). To explore the spatial relevant molecular features, we employed a supervised random forest-based machine learning model to investigate the dysregulated proteins that may define various disease subgroups. These subgroups were further classified into four clusters (Figure 5B). UMAP visualizations of all subtype samples from all patients, based on features obtained from the random forest machine-learning model, were displayed with arrows connecting the centroid of each subtype (Figure 5C). In the Sankey plot, our results showed the significance of mitochondrial dysfunctions related to metabolism (Figure 5D), which could have crucial implications for the discovery of therapeutic targets.

Furthermore, we examined the spatial expressions of AHCYL1, ARHGAP8, NAMPT, and YWHAB as the top feature in each cluster shown in Figure 5B, observing their heterogeneous distribution across different spatial regions and slices (Figure 5E), as well as the key cancer regulator KRAS[29]. Moreover, pathway-centric analysis unveiled differential spatial expression of multiple proteins within CRC-relevant Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (pathway entry: hsa05210) (Figure 5F). The distribution map provides insights into trends in protein expression across different stages. In the carcinoma stage, we observed a noticeable increase in the number of upregulated proteins involved in the p53 signaling pathway within the carcinoma area. Conversely, the ErbB signaling pathway exhibited a higher count of downregulated proteins in the same carcinoma area. These findings underscore the dynamic changes occurring in CRC samples during disease progression.

Discussion

Establishing the FAXP workflow represents a significant advancement in improving lateral/volumetric resolution and enhancing the throughput for spatially resolved proteomics analysis of biological tissues. We have also optimized the hydrogel recipe to enable effective expansion proteomics analysis of FFPE samples. Key novel or optimized steps were introduced to significantly expedite the processing time and depth of the spatial proteome (Figure 6A). These include: i) slicing FFPE samples onto customized glass slides with chamber or commercial glass slides for all the scenarios of FFPE sample placement; ii) performing dewaxing and rehydration; iii) introducing autoclave-based tissue homogenization to utilize high pressure and heat to accelerate the tissue homogenization; iv) conducting reduction/alkylation on the entire tissue slice embedded in hydrogel to reduce the hands-on steps during proteomics sample preparation and increase the reproducibility; v) performing fluorescent staining or optimizing Coomassie Brilliant Blue R-250 staining after tissue expansion for quicker staining and easier de-staining; vi) facilitating manual microdissection down to 73 μm in diameter or laser capture microdissection down to subcellular organelle for higher resolution and employing proteomics sample preparation through the developed filter-aided sample preparation to reduce sample loss and increase

sample processing throughput; vii) utilizing DIA-MS with timsTOF Pro and performing protein identification using a hybrid search library strategy to ensure proteome coverage. In addition, the workflow is partially automated with robotics.

FAXP has achieved two remarkable features in spatial resolution and sample processing throughput. This lateral resolution is 2.2 times higher than that achieved with ProteomEx. FAXP facilitates the processing of micro-dissected samples in batches of up to 96, with the simultaneous processing of multiple batches. Collectively, FAXP has reduced the total processing time by up to 53.5% compared to ProteomEx (Figure 6B). In ProteomEx, the sample loss rate during the peptide extraction step is around 6%[14]. The FAXP method minimizes both sample and peptide loss due to fewer hands-on steps and the peptide-trapping capability of the hydrogel during sample preparation.

Another key distinction of FAXP is its adaptability to varying sample types. For instance, samples rich in extracellular matrix, such as CRC tissues[30,31], require more rigorous homogenization conditions involving higher SDS concentration, elevated homogenization temperature, and longer processing time. This dynamic optimization, as demonstrated by different conditions in Figure S7A (Table S1), ensures that FAXP remains versatile and effective across diverse samples, highlighting its flexibility for clinical applications.

When considering CRC applications, disparities emerge among N, L, H, C, CC vs. PC subtypes in contrast to the global proteome, underlining the significance of analyzing distinct CRC subtypes to gain deeper insights into the disease's molecular mechanisms[32,33]. Especially, FAXP exhibits the capability to discriminate heterogeneity within inter-patient, inter-slide, and intra-tissue clinical FFPE samples. This attribute bears particular importance in clinical research, enabling precise analysis and characterization of samples, thereby advancing comprehension of diseases and potential therapeutic avenues.

There are also some limitations for FAXP workflow. Optimization of the gel expansion technique for tissues rich in collagen and extracellular matrix components represents one of the challenges. Another limitation is the staining method, which currently only uses Coomassie Brilliant Blue. This falls short compared to more advanced staining methods like immunofluorescence, which is important for precise clinical observations. To enhance its usefulness in addressing heterogeneity in clinical samples, the workflow should also be able to handle samples with irregular shapes. While FAXP does improve spatial resolution compared to ProteomEx, it still cannot achieve single-cell resolution using manual operation same as LCM technique due to limitations in sample collection method, while alternatively LCM based technique can be incorporated with FAXP to further extend to the single-cell and subcellular organelle resolution. In this paper, we have demonstrated the application of FAXP in analyzing the proteome of a single nuclear. More types of subcellular organelles could be potentially analyzed in future experiments.

In summary, the introduction of the FAXP workflow represents a significant step forward in hydrogel-based spatial proteomics, addressing challenges in tissue expansion, sample preparation and data analysis. Its notable strengths include increased throughput of sample preparation, better sample compatibility, improved spatial/volumetric resolution, and enhanced reproducibility.

Materials and Methods

Preparation of mouse organ FFPE tissue samples

All animal maintenance and experimental procedures were conducted according to the Westlake University Animal care guidelines, and all animal studies were approved by the

Institutional Animal Care and Use Committee (IACUC) of Westlake University, Hangzhou, China, under animal protocol #22-083-GTN.

Male C57BL/6J mice (four-month old) were anesthetized deeply using 1% sodium pentobarbital to perform the procedures. Subsequently, transcardial perfusion was performed using 1× phosphate-buffered saline (PBS) and 4% paraformaldehyde (PFA; Electron Microscopy Sciences, the United States). The mouse liver and kidney were dissected and postfixed in 4% PFA for 6 hours at 4°C. Fixed tissues were dehydrated using increasing ethanol concentrations: 75%, 95%, and 100% for 30 minutes each. Next, the dehydrated tissues underwent infiltration with paraffin wax at 60°C. The embedded tissues were sliced into either 5-µm or 10-µm in thickness, utilizing a rotary microtome (Leica RM2255, Germany) according to a standard protocol and mounted on the glass slides.

Collection of FFPE samples from patients

Archival FFPE colorectal tissue samples were acquired from three patients diagnosed with CRC at the Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, China. The samples were sliced to a thickness of 8 µm with a rotary microtome (Leica RM2255, Germany). The tumor grading followed the Tumor, Node, Metastasis (TNM) classification system of colorectal cancer, as illustrated in Figure 4. Informed consent was obtained from all patients prior to their surgeries. Patient-derived tissues were embedded in paraffin according to the approved institutional review board protocol at the same hospital. Human tissue samples were collected with the approval of the Institutional Ethics Committee of the Second Affiliated Hospital of Zhejiang University School of Medicine (No. 2020-322). The study was also approved by the ethics committee of Westlake University (Permission number: 20220913GTN001).

Tissue expansion and staining

The chemical composition and storage conditions for all reagents and buffers used for tissue expansion, staining, and sample preparation are described in Table S4. Suppliers and catalog numbers for chemicals, reagents, and accessories used for FAXP are listed in Table S5. The paraffin-embedded sections were mounted onto glass slides and subjected to deparaffinization. Two rounds of 10-minute heptane treatments were performed. The sections then underwent a rehydration procedure involving decreasing ethanol concentrations: 100%, 90%, and 75% for 5 minutes each followed by rehydration in ddH_2O for 5 minutes. The rehydrated FFPE tissue sections were then ready for tissue expansion. The rehydrated CRC tissue sections were subjected to H&E staining following the manufacturer's protocol (Shanghai Yuanye Bio-Technology, China).

The tissue expansion procedure followed the previously described protocol[14], incorporating certain modification. Briefly, rehydrated tissue sections were rinsed in 1× PBS and treated with protein anchoring solution for either 1 hour or 2 hours at 25°C. Protein anchoring time was reduced compared to the original protocol because the thinner tissue sections were used in this study (5-µm or 10-µm vs. 30-µm). The anchoring solution was removed by washing the samples three times with the anchoring termination buffer for 5 minutes each. The NSA-anchored tissue was incubated with activated monomer solution in a gelation chamber at 4°C to infuse monomers. Subsequently, the treated tissue was transferred to a vacuum oven for polymerization reaction at 37°C in a nitrogen gas atmosphere for 2-3 hours. The formed tissue-hydrogel composite was transferred to a dish with the protein denaturation buffer. The buffer and homogenization condition was adapted from the dExPath protocol[34]. To improve mechanical stability of the hydrogel after autoclaving, we increased cross-linker concentration 17.5-fold from original (Table S1). Homogenized samples were transferred to 9

cm petri dishes and washed with 1× PBS three times. Reduction and alkylation of proteins were performed by adding the reduction buffer (20 mM dithiothreitol (DTT) in 1× PBS) and alkylation buffer (55 mM iodoacetamide (IAA) in 1× PBS) and incubating for 30 minutes, respectively. The samples were washed with 1× PBS three times for 10 minutes each time. The hydrogel sample was incubated in ddH_2O and allowed to expand to its maximum size. We found out that using Coomassie Brilliant Blue (Sangon, China) staining allowed us to stain and expand samples in less than 1 hour compared to the almost 10 hours of staining and expansion in the original protocol. Imaging was performed using Zeiss Fluorescence Stereo Zoom Microscope (Axio Zoom.V16) in brightfield mode, controlled by ZEN 3.1 software. Furthermore, we found DAPI staining was also compatible with our workflow (Figure S8). By using hydrogel samples with a diameter of 488 μm derived from mouse liver slices, we could identify an average of 830 peptides and 350 proteins using 0.5-hour DIA-MS by timsTOF Pro.

Optimization of reduction and alkylation conditions

Three hydrogels of mouse liver FFPE slices (10 μm) were employed to determine the best conditions for protein reduction and alkylation on whole tissue slice samples. Four sets of reduction and alkylation conditions (10 mM DTT and 55 mM IAA; 20 mM DTT and 55 mM IAA; 20 mM TCEP and 55 mM IAA; 45 mM TCEP and 100 mM IAA) were then tested separately on the gels (as shown in Figure 1H). The combination of 20 mM DTT/55 mM IAA resulted in the highest carbamidomethyl modification rate close to an average 98% (Figure 1I). This condition exhibited comparable rates of cysteine-containing peptide and protein identifications when compared to the ProteomEx method (indicated No. 1 in Figure 1I), while also demonstrating superior peptide identification performance.

Proteomic sample preparation using conventional in-gel digestion

Manual microdissections were performed on the expanded and Coomassie-stained tissue-hydrogels (following reduction and alkylation). The excised tissue-hydrogel samples were first de-stained and dehydrated. Protein digestion was done using 12.5 ng/μL trypsin (Hualishi Tech. Ltd, China).

The supernatant of the digested peptide solutions was collected first. After that, 100 μL of 25 mM ABB was added to collect the resulting supernatant. Then, 100 μL of 50% ACN and 2.5% formic acid (FA) was added and vortexed for 10 minutes; this step was repeated three times with the supernatant collection each time. Next, 100 μL of 100% ACN was added until the hydrogel fragments became small and white. The peptide solutions were then concentrated using a SpeedVac to 20-30 μL. Purification was done using C18 micro spin columns (Thermo Fisher Scientific, the United States). After purification, the samples were dried with a SpeedVac, stored at -80°C, and ready for later LC-MS/MS analysis.

Proteomic sample preparation using filter-aided in-gel digestion

One plug of the C18 disk (Empore, the United States) was assembled into the 200 μL tip to form a spintip device. After transferring the excised tissue-hydrogel samples, the samples were dehydrated. Protein digestion was then performed using 12.5 ng/μL trypsin (Hualishi Tech. Ltd, China).

After collecting the digested peptide solutions, the resulting peptide solutions were concentrated to dryness using a SpeedVac and stored at -80°C, ready for later LC-MS/MS analysis.

Determination of detection range of the minimal volume

Tissue volumes were estimated using commercially available biopsy punches based on needs (0.35 mm, 0.5 mm, 0.75 mm, 1 mm, 1.5 mm, 2 mm, 3 mm; Integra Miltex, the United States). Specifically, the formula to calculate the volume is: volume (nL) = $\pi \times (D/2F)^2 \times h \times 10^{-6}$, where D is the diameter (μm) of the punch, F is the expansion factor calculated as the square root of the ratio of the tissue area after expansion to the area before expansion, and h is the thickness (μm) of the slice. The detection range calibration curves were estimated using Local Polynomial Regression Fitting estimation (loess) using the number of identified peptides and proteins.

Single nucleus collection, sample preparation and proteomics analysis

The hydrogel sample of 5 μm FFPE mouse liver was attached onto a polyester (POL) frame slide (Leica, Germany). Single nuclei were individually imaged and cut by a MMI CellCut (Molecular Machines & Industries, Germany) and collected by diffuser caps (Molecular Machines & Industries, Germany). Collected samples were then processed through 10 μL tip under filter-aided in-gel digestion. Peptides were analyzed through an Orbitrap Astral mass spectrometer (Thermo Fisher Scientific, the United States) using DIA mode under the liquid chromatography conditions and MS parameters suggested by the Thermo Fisher Scientific.

Liquid chromatography

The peptide samples were separated on a 15 cm × 75 μm silica column custom packed with 1.9 μm 100 Å C18 aqua installed into nanoElute® system (Bruker Daltonics, Germany). The mobile phase was mixed with buffer A (MS grade H_2O + 0.1% FA) and buffer B (MS grade ACN + 0.1% FA). The flow rate was set at 300 nL/min. For both 30-minute and 60-minute linear LC gradients, the buffer B (%) was linearly increased from 5% to 27% in 25 minutes and 50 minutes separately, followed by an increase to 40% within 5 minutes and 10 minutes separately, and a further boost to 80%. The results showed that 1-hour LC gradient resulted in more peptides and proteins, along with an increase in the mean peak full width at half maximum (FWHM) from an approximate average of 2.5 to 3.5 (Figure S7B).

Proteomic data acquisition using MS

Eluted peptides were analyzed in a trapped ion mobility spectrometry (TIMS) combined with a quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonics, Germany) via a CaptiveSpray nano-electrospray ion source. MS data were acquired using the Bruker Otofcontrol v6.2 and HyStar v5.1 mass spectrometer control software.

Data-dependent acquisition (DDA) was performed in the PASEF mode with 10 PASEF scans per topN acquisition cycle. The ramp time was set at 100 ms, achieving a total cycle time of 1.17 s. The ion mobility was scanned from 0.7 to 1.3 Vs/cm^2. MS1 and MS2 acquisition was performed in the m/z range from 100 to 1700 Th. Precursors that reached a target intensity of 20,000 cts/s were dynamically excluded for 0.4 minute. Singly charged precursors were excluded by their position in the m/z-ion mobility plane. For the data-independent acquisition Parallel Accumulation Serial Fragmentation (diaPASEF) mode, the majority of settings were the same as in DDA mode. We defined the isolation windows as shown in Table S3. For the determination of the detection range, both DDA and DIA methods were used for sample collection. A total of 33 DDA files and 33 DIA files were collected for evaluation purposes.

Protein identification and quantification

DDA data were analyzed using FragPipe (version 18.0) platform[35,36] with the MSFragger (version 3.5) search engine against a FASTA file from SwissProt containing 16,985 mouse protein entries (downloaded in September 2018) and 16,985 decoy sequences for mouse samples and a FASTA file from SwissProt containing 20,386 human protein entries

(downloaded in November 2022) and 20,386 decoy sequences. The "IM-MS" was selected as the MS data type under "Input LC-MS Files" section. A fragment mass tolerance was set at 0.05 Da. The enzyme used for digestion was trypsin, with cleavage occurring after "KR" residues but not preceding "P" residues. To evaluate the cysteine modification rates, carbamidomethylation was considered variable modification, while in the other DDA files, it was set as fixed.

DIA data analysis was performed by DIA-NN (version 1.8.1)[37,38]. The false discovery rate for precursors was set at 1%. Fixed modifications included "C carbamidomethylation," while variable modifications included "N-term M excision" and "Methionine oxidation." The peptide length range was set between 7 and 50. The precursor charge range was set between 2 and 4. Both m/z ranges for precursor and fragment ions were set between 100 and 1700. "Unrelated runs," "Use isotopologues," and "MBR" options were selected. Heuristic protein inference was chosen under mouse liver-based studies. Protein inference was set as "off," and the quantification strategy was set as "Robust LC (high precision)." Other settings remained as default values. The DIA data searching employed an in-house generated library for mouse liver-based studies. For the single-nucleus studies, together with the data from 0.042 nL of In-tip DIA-MS, proteomic data was searched using library-free mode by DIA-NN. Similar as the settings shown above, the modifications were mainly made for precursor (300-1800) and fragment ion (200-1800) m/z ranges. For the CRC application, proteomic data was searched using library-free mode by DIA-NN. This generated a predicted library, which was then merged with an in-house generated pan-human library to form a hybrid library. Only the precursors do not present in the pan-human library were used in combining the spectral library. The hybrid library was then used for a library-based search. The peptide and protein numbers in the DDA analysis were extracted from the FragPipe peptide.tsv and protein.tsv files, whereas in the DIA analysis, they were acquired from the stripped sequences in DIA-NN pr.matrix and unique proteins in the pg.matrix.

Batch design and data analysis for MS acquisition

The MS batches were systematically arranged, balanced, and randomized, considering patient ID, slide ID, region, and replicate number features (Figure S6A). This resulted in acquiring six batches, with 22 injections per batch. It's worth noting that one file failed due to an MS bug during acquisition, resulting in a total of 131 files. The coefficient of variation and Pearson correlation analysis for global precursors and proteins demonstrated robust reproducibility in the MS data (Figure S6B). When examining pooled samples across different batches, we observed high consistency with pool samples (Figure S6C), and the missed cleavage rates exhibited low and consistent values (Figure S6D). We evaluated five distinct spectral searching conditions aimed at achieving deepest identification performance, including a colon library developed in-house with default DIA-NN parameters and optimal DIA-NN parameters for mass range (which was also optimized for retention time window); a hybrid-library search approach using the combined library from predicted library obtained via a DIA-NN library-free search for those proteins not included in the spectral library; a pan-library consisting of multiple tissue types with a hybrid search approach.

Feature selection using the random forest model

The unsupervised clustering approach involved selecting the protein expression matrix derived from each patient sample. The k-means method was then applied for clustering spatial locations using the proteome abundance into four clusters for each slice, utilizing different shapes to represent the cluster category on a location map and colors to indicate the subtype. Finally, the central point of each cluster was calculated to generate a disease process

pathway map. We used a random forest-based model to select the top 100 features that separated four disease subgroups (normal, low-grade dysplasia, high-grade dysplasia, carcinoma) using 70% of samples as the training set and 30% of total samples as the test set. This approach yielded an accuracy of 81% and an AUC of 0.967. These selected features were then analyzed with Fuzzy c-means soft clustering (provided by mfuzz packages) into four clusters. Proteins upregulated in each disease subgroup were visualized as network modules, with proportions of each cluster indicated as pie charts.

Spatial visualization of protein expression

With the mapped locations of punches based on microscopic images, each protein was visualized as a z-score using scaled colors, where subgroup boundaries were determined manually. The pathway visualization highlighted the proteins within the KEGG colorectal cancer pathway (pathway entry: hsa05210).

Statistical analysis

Unpaired Welch's t-test was performed for all number comparisons, and statistical visualizations were drawn in PRISM (10.0.3) and R (version 4.2.2). Isotropic analysis calculations were performed using MATLAB (version R2021a), following the previously published protocol[14]. The differentially expressed protein comparison was performed at a threshold of adjusted P-value <0.05 and fold change >2 (P-values were adjusted by the method of Benjamini–Hochberg). GO term and pathway enrichment analyses were performed by ingenuity pathway analysis (IPA), in which a right-tailed Fisher's exact test estimated the P-value.

As to the single nucleus analysis, the protein matrix from DIA-NN was transformed using a logarithmic scale and then normalized across samples using quantile normalization. The Pearson correlation coefficient was calculated after the removal of missing values. Unless otherwise specified, missing values were imputed by using the minimum value. Between-group differences were evaluated using Welch's t-test, and the P values were adjusted using the Benjamini–Hochberg procedure in R (version 4.4.0). Gene ID conversion and Gene Ontology (GO) enrichment analysis were executed using the clusterProfiler (version 4.11.0), with all default options except for the maxGSSize which was set to 2500. The annotation of GO terms was performed using org.Mm.eg.db (version 3.18.0). Redundant GO terms were simplified using the default options of the clusterProfiler::simplify function. The Reactome pathway-based analysis was conducted using the ReactomePA (version 1.47.0) package.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[39] partner repository with the dataset identifier PXD041412.

Acknowledgments

We thank the NSFC grant (Grant No. 92259201 to Dr Tiannan Guo), the National Key R&D Program of China (Grant No. 2022YF0608403 to Dr Yi Zhu) and the Key Research and Development Program of Zhejiang Province (Grant No. 2022C03037 to Dr Tiannan Guo). Figure 6A was created with Biorender.com. We thank Cuiji Sun for teaching gel-making approaches based on ProteomEx, and Liang Yue for sharing home-built DIA libraries. Declaration of generative AI and AI-assisted technologies in the writing process during the preparation of this work the authors used ChatGPT to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Author contributions

T.G., Y.Z., and K.D.P conceived the project. Z.D., Y.Z., K.D.P., and C.W. developed and benchmarked the method. T.C. provided the clinical samples. Z.D., C.W., and J.C. performed the proteomic experiments. Q.X., and Z.D. participated in the robotic automation system establishment. F.Z., Z.D., and W.J. conducted proteomic data analysis and Z.D. performed the isotropic analysis. D.Z., F.Z., K.D.P. and T.G. wrote the manuscript, and the others revised the manuscript. T.G., Y.Z., and K.D.P. supervised the project.

Competing interests

T.G. and Y.Z. are shareholders and C.W. is the staff of Westlake Omics (Hangzhou) Biotechnology Co., Ltd., where ProteomEx-related technologies are commercialized. The remaining authors declare no competing interests.

References

1 Taylor, M. J., Lukowski, J. K. & Anderton, C. R. Spatially Resolved Mass Spectrometry at the Single Cell: Recent Innovations in Proteomics and Metabolomics. *Journal of the American Society for Mass Spectrometry* **32**, 872-894 (2021). https://doi.org:10.1021/jasms.0c00439

2 Moore, J. L., Patterson, N. H., Norris, J. L. & Caprioli, R. M. Prospective on Imaging Mass Spectrometry in Clinical Diagnostics. *Molecular & Cellular Proteomics* **22**, 100576 (2023). https://doi.org:10.1016/j.mcpro.2023.100576

3 Claes, B. S. R. *et al.* MALDI-IHC-Guided In-Depth Spatial Proteomics: Targeted and Untargeted MSI Combined. *Analytical Chemistry* **95**, 2329-2338 (2023). https://doi.org:10.1021/acs.analchem.2c04220

4 Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nature Biotechnology* **40**, 1231-+ (2022). https://doi.org:10.1038/s41587-022-01302-5

5 Makhmut, A. *et al.* A framework for ultra-low-input spatial tissue proteomics. *Cell Systems* **14**, 1002-1014.e1005 (2023). https://doi.org:https://doi.org/10.1016/j.cels.2023.10.003

6 Rosenberger, F. A. *et al.* Spatial single-cell mass spectrometry defines zonation of the hepatocyte proteome. *Nature Methods* **20**, 1530-1536 (2023). https://doi.org:10.1038/s41592-023-02007-6

7 Xu, R. *et al.* Spatial-Resolution Cell Type Proteome Profiling of Cancer Tissue by Fully Integrated Proteomics Technology. *Analytical Chemistry* **90**, 5879-5886 (2018). https://doi.org:10.1021/acs.analchem.8b00596

8 Kwon, Y. *et al.* Hanging drop sample preparation improves sensitivity of spatial proteomics. *Lab on a Chip* **22**, 2869-2877 (2022). https://doi.org:10.1039/d2lc00384h

9 Zhu, Y. *et al.* Nanodroplet processing platform for deep and quantitative proteome profiling of 10-100 mammalian cells. *Nature Communications* **9** (2018).

https://doi.org:10.1038/s41467-018-03367-w

10 Ma, M. *et al.* In-depth mapping of protein localizations in whole tissue by micro-scaffold assisted spatial proteomics (MASP). *Nature Communications* **13** (2022). https://doi.org:10.1038/s41467-022-35367-2

11 Bhatia, H. S. *et al.* Spatial proteomics in three-dimensional intact specimens. *Cell* **185**, 5040-+ (2022). https://doi.org:10.1016/j.cell.2022.11.021

12 Wu, Q. & Lu, H. in *Mass Spectrometry Imaging of Small Molecules Methods in Molecular Biology* Ch. Chapter 13, 181-194 (Springer US, 2022).

13 Ryan, D. J. *et al.* MicroLESA: Integrating Autofluorescence Microscopy, In Situ Micro-Digestions, and Liquid Extraction Surface Analysis for High Spatial Resolution Targeted Proteomic Studies. *Analytical Chemistry* **91**, 7578-7585 (2019). https://doi.org:10.1021/acs.analchem.8b05889

14 Li, L. *et al.* Spatially resolved proteomics via tissue expansion. *Nature Communications* **13**, 7242 (2022). https://doi.org:10.1038/s41467-022-34824-2

15 Drelich, L. *et al.* Toward High Spatially Resolved Proteomics Using Expansion Microscopy. *Analytical Chemistry* **93**, 12195-12203 (2021). https://doi.org:10.1021/acs.analchem.0c05372

16 Bai, Y. H. *et al.* Expanded vacuum-stable gels for multiplexed high-resolution spatial histopathology. *Nature Communications* **14** (2023). https://doi.org:10.1038/s41467-023-39616-w

17 Yat Ho, C., Koralege, C. P., Dominick, P.-J., Stephanie, M. C. & Ruixuan, G. Gel-assisted mass spectrometry imaging. *bioRxiv*, 2023.2006.2002.543480 (2023). https://doi.org:10.1101/2023.06.02.543480

18 M'Saad, O. & Bewersdorf, J. Light microscopy of proteins in their ultrastructural context. *Nature Communications* **11**, 3850 (2020). https://doi.org:10.1038/s41467-020-17523-8

19 Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543-548 (2015). https://doi.org:doi:10.1126/science.1260088

20 Ku, T. *et al.* Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nature Biotechnology* **34**, 973-981 (2016). https://doi.org:10.1038/nbt.3641

21 Tillberg, P. W. *et al.* Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nature Biotechnology* **34**, 987-+ (2016). https://doi.org:10.1038/nbt.3625

22 Zhang, F. *et al.* A Comparative Analysis of Data Analysis Tools for Data-Independent Acquisition Mass Spectrometry. *Molecular & Cellular Proteomics* **22** (2023).

632 https://doi.org:10.1016/j.mcpro.2023.100623
633 23 Nwosu, A. J. *et al.* In-Depth Mass Spectrometry-Based Proteomics of Formalin-Fixed,
634 Paraffin-Embedded Tissues with a Spatial Resolution of 50-200 µm. *Journal of*
635 *Proteome Research* **21**, 2237-2245 (2022).
636 https://doi.org:10.1021/acs.jproteome.2c00409
637 24 Damstra, H. G. J. *et al.* Correction: Visualizing cellular and tissue ultrastructure using
638 Ten-fold Robust Expansion Microscopy (TREx). *eLife* **11**, e85169 (2022).
639 https://doi.org:10.7554/eLife.85169
640 25 Azimifar, S. B., Nagaraj, N., Cox, J. & Mann, M. Cell-Type-Resolved Quantitative
641 Proteomics of Murine Liver. *Cell Metabolism* **20**, 1076-1087 (2014).
642 https://doi.org:10.1016/j.cmet.2014.11.002
643 26 Hammad, S. *et al.* Protocols for staining of bile canalicular and sinusoidal networks of
644 human, mouse and pig livers, three-dimensional reconstruction and quantification of
645 tissue microarchitecture by image processing and analysis. *Archives of Toxicology*
646 **88**, 1161-1183 (2014). https://doi.org:10.1007/s00204-014-1243-5
647 27 Ordoñez, C., Screaton, R. A., Ilantzis, C. & Stanners, C. P. Human
648 Carcinoembryonic Antigen Functions as a General Inhibitor of Anoikis1. *Cancer*
649 *Research* **60**, 3419-3424 (2000).
650 28 Bustin, S. A., Li, S. R. & Dorudi, S. Expression of the Ca-Activated chloride channel
651 genes and is downregulated in human colorectal cancer. *DNA and Cell Biology* **20**,
652 331-338 (2001). https://doi.org:Doi 10.1089/10445490152122442
653 29 Huang, L. M., Guo, Z. X., Wang, F. & Fu, L. W. KRAS mutation: from undruggable to
654 druggable in cancer. *Signal Transduction and Targeted Therapy* **6** (2021).
655 https://doi.org:10.1038/s41392-021-00780-4
656 30 Yuan, Z. N. *et al.* Extracellular matrix remodeling in tumor progression and immune
657 escape: from mechanisms to treatments. *Molecular Cancer* **22** (2023).
658 https://doi.org:10.1186/s12943-023-01744-8
659 31 Karlsson, S. & Nystrom, H. The extracellular matrix in colorectal cancer and its
660 metastatic settling-Alterations and biological implications. *Critical Reviews in*
661 *Oncology Hematology* **175**, 103712 (2022).
662 https://doi.org:10.1016/j.critrevonc.2022.103712
663 32 Baker, A. M. *et al.* Evolutionary history of human colitis-associated colorectal cancer.
664 *Gut* **68**, 985-995 (2019). https://doi.org:10.1136/gutjnl-2018-316191
665 33 Wang, W. *et al.* Molecular subtyping of colorectal cancer: Recent progress, new
666 challenges and emerging opportunities. *Seminars in Cancer Biology* **55**, 37-52 (2019).
667 https://doi.org:https://doi.org/10.1016/j.semcancer.2018.05.002

34 Valdes, P. A. *et al.* Decrowding Expansion Pathology: Unmasking Previously Invisible Nanostructures and Cells in Intact Human Brain Pathology Specimens. *bioRxiv* (2021). https://doi.org:10.1101/2021.12.05.471271

35 Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D. & Nesvizhskii, A. I. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nature Methods* **14**, 513-+ (2017). https://doi.org:10.1038/Nmeth.4256

36 Yu, F. *et al.* Fast Quantitative Analysis of timsTOF PASEF Data with MSFragger and IonQuant. *Molecular & Cellular Proteomics* **19**, 1575-1585 (2020). https://doi.org:https://doi.org/10.1074/mcp.TIR120.002048

37 Demichev, V. *et al.* dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nature Communications* **13** (2022). https://doi.org:10.1038/s41467-022-31492-0

38 Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature Methods* **17**, 41-+ (2020). https://doi.org:10.1038/s41592-019-0638-x

39 Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Research* **50**, D543-D552 (2022). https://doi.org:10.1093/nar/gkab1038

FIGURES

Figure 1

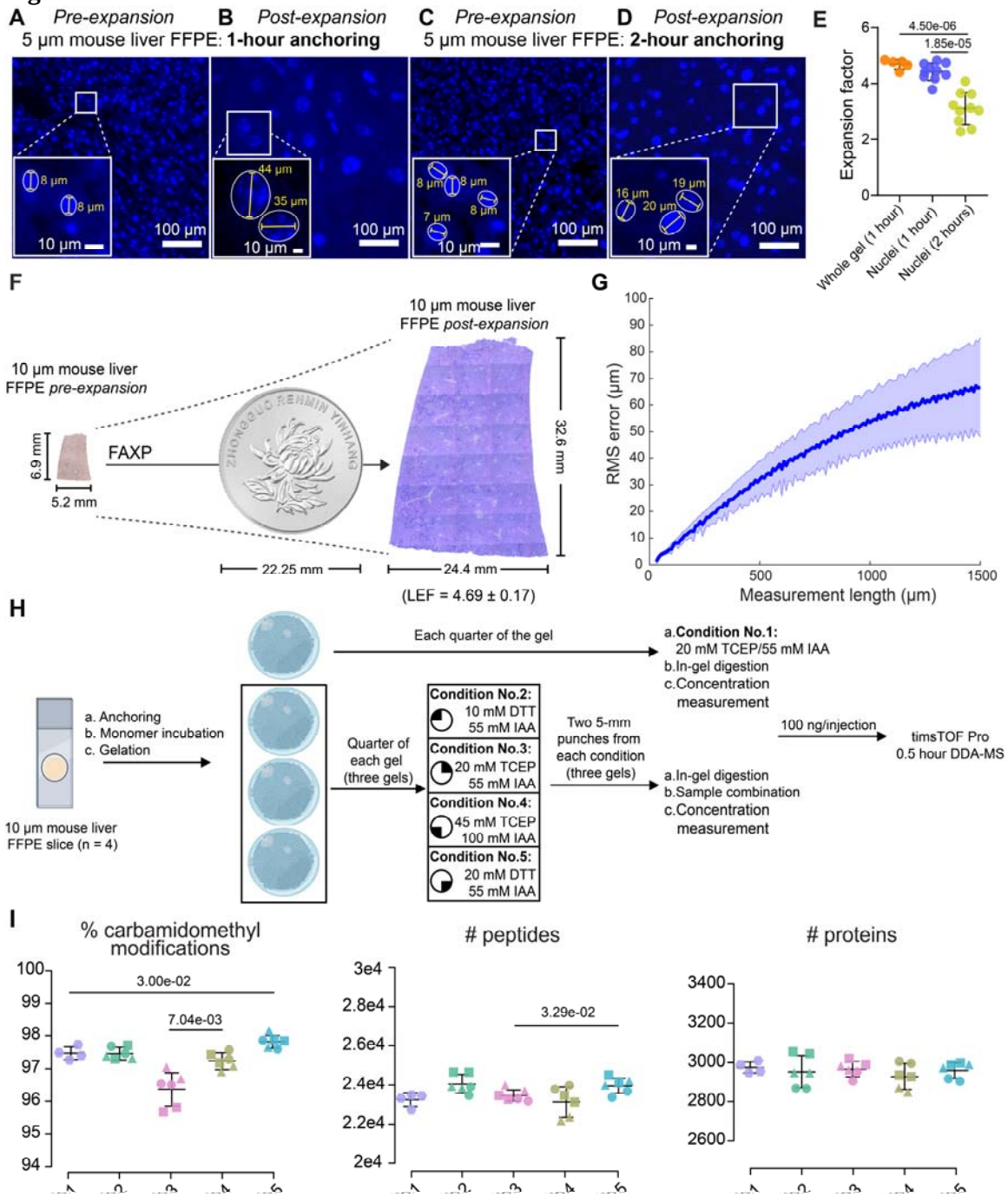

Figure 1. Optimization of gel-making procedures. (A) DAPI image of a mouse liver FFPE slice after 1-hour anchoring before expansion. (B) Post-expansion DAPI image of the same sample incubated under 1-hour anchoring. (C) DAPI image of a mouse liver FFPE slice after 2-hour anchoring before expansion. (D) Post-expansion DAPI image of the same sample incubated under 2-hour anchoring. Magnified views with white-circled nuclei and measured diameters are shown within each figure (A-D). Scale bars are indicated within each figure. (E) Comparison of expansion factors between whole gel macro-level colorimetric staining (n = 5) and micro-level DAPI nuclear staining under 1- and 2-hour anchoring (n = 10 each). All data points are presented as mean values ± standard deviation (SD). (F) Representative bright-field images of an FFPE mouse liver slice captured before expansion (left) and after expansion with Coomassie brilliant blue staining (right), in comparison to a one-dollar RMB coin, with a diameter of 22.25 mm. (G) Comparison of root-mean-square (RMS) measurement length error between pre-expansion and post-expansion liver slice images (n = 4); one example is shown in (F). (H) Study design for comparative proteomic analysis of four reduction and alkylation conditions. After gel-making for three mouse liver FFPE slices, four reduction/alkylation conditions including 10 mM DTT/55 mM IAA, 20 mM DTT/55 mM IAA, 20 mM TCEP/55 mM IAA, and 45 mM TCEP/100 mM IAA are applied to one-fourth of each gel (n = 3). Following in-gel digestion, samples from two 5-mm punches are mixed, and MS data is collected using 100 ng per injection. (I) Comparative proteomic results of the four reduction and alkylation conditions, including percentages of carbamidomethyl modifications, as well as the number of identified peptides and proteins. All data points are presented as mean values ± SD.

Figure 2

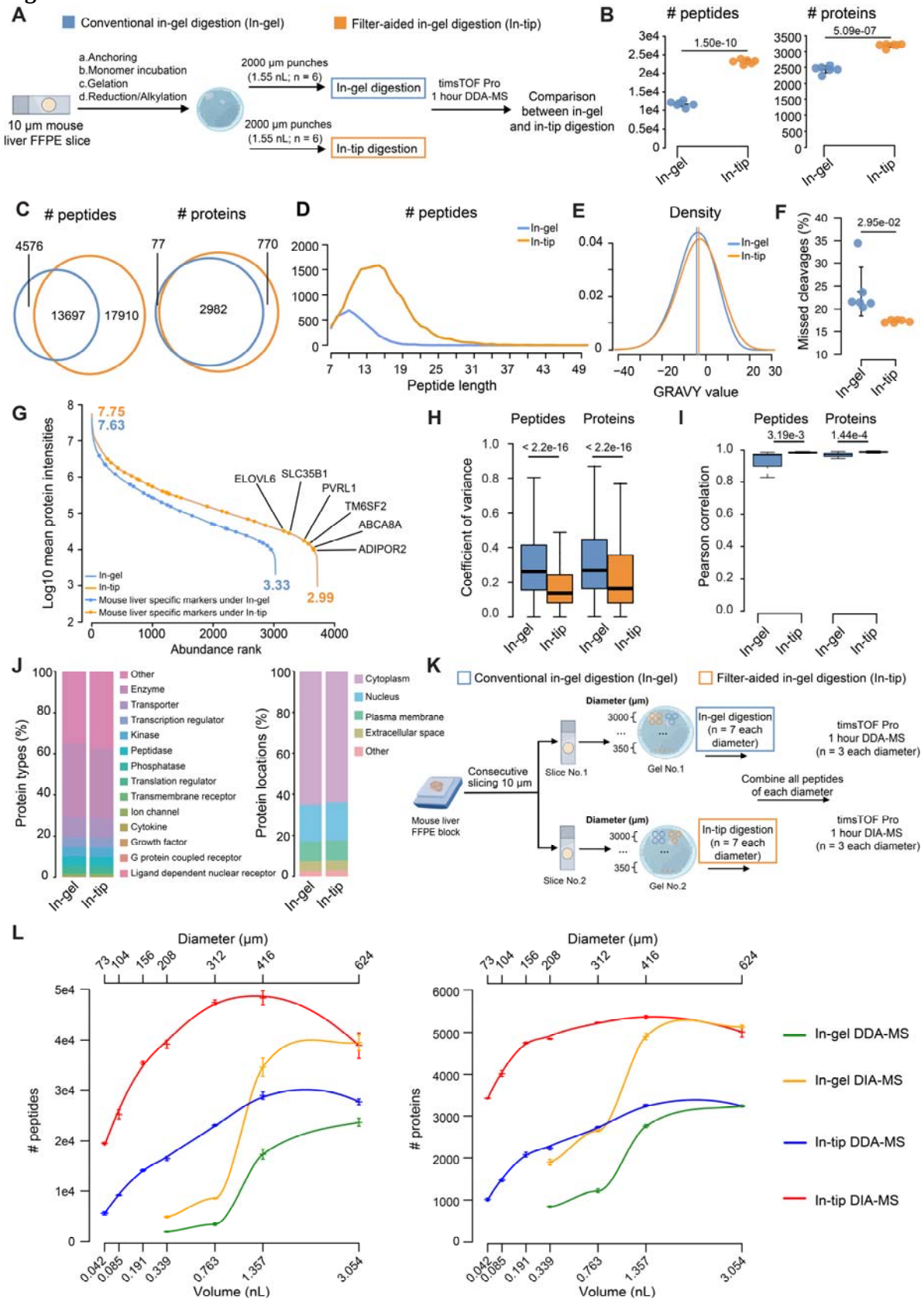
A
Conventional in-gel digestion (In-gel)
Filter-aided in-gel digestion (In-tip)
a.Anchoring
b.Monomer incubation
c.Gelation
d.Reduction/Alkylation
10 µm mouse liver FFPE slice
2000 µm punches (1.55 nL; n = 6)
In-gel digestion
2000 µm punches (1.55 nL; n = 6)
In-tip digestion
timsTOF Pro 1 hour DDA-MS
Comparison between in-gel and in-tip digestion
B
peptides
proteins
1.50e-10
5.09e-07
In-gel
In-tip
C
peptides
proteins
4576
13697
17910
77
2982
770
D
peptides
In-gel
In-tip
Peptide length
E
Density
GRAVY value
F
Missed cleavages (%)
2.95e-02
G
Log10 mean protein intensities
Abundance rank
7.75
7.63
3.33
2.99
ELOVL6
SLC35B1
PVRL1
TM6SF2
ABCA8A
ADIPOR2
Mouse liver specific markers under In-gel
Mouse liver specific markers under In-tip
H
Peptides
Proteins
< 2.2e-16
Coefficient of variance
I
Peptides
Proteins
3.19e-3
1.44e-4
Pearson correlation
J
Protein types (%)
Protein locations (%)
Other
Enzyme
Transporter
Transcription regulator
Kinase
Peptidase
Phosphatase
Translation regulator
Transmembrane receptor
Ion channel
Cytokine
Growth factor
G protein coupled receptor
Ligand dependent nuclear receptor
Cytoplasm
Nucleus
Plasma membrane
Extracellular space
Other
K
Conventional in-gel digestion (In-gel)
Filter-aided in-gel digestion (In-tip)
Mouse liver FFPE block
Consecutive slicing 10 µm
Slice No.1
Slice No.2
Diameter (µm)
Gel No.1
Gel No.2
In-gel digestion (n = 7 each diameter)
In-tip digestion (n = 7 each diameter)
timsTOF Pro 1 hour DDA-MS (n = 3 each diameter)
Combine all peptides of each diameter
timsTOF Pro 1 hour DIA-MS (n = 3 each diameter)
L
Diameter (µm)
peptides
proteins
Volume (nL)
In-gel DDA-MS
In-gel DIA-MS
In-tip DDA-MS
In-tip DIA-MS

Figure 2. Comparison between conventional in-gel digestion and filter-aided in-gel digestion applied to mouse liver FFPE slices. (A) The study design illustrates a comparative proteomic analysis between conventional in-gel digestion (in-gel digestion) and filter-aided in-gel digestion (in-tip digestion). After gel-making for a mouse liver FFPE slice, 2000-µm punches (n = 6; containing 1.55 nL of tissue volume before expansion) are subjected to in-gel and in-tip digestions separately. (B) Number comparison of the identified peptides and proteins between in-gel and in-tip digestions (n = 6 each). All data points are presented as mean values ± standard deviation (SD). (C) Venn diagram shows the overlapping of identified peptides and proteins between in-gel and in-tip digestions. (D) A curve plot shows the distribution of peptide lengths exclusively identified through in-gel and in-tip digestions, as shown in (C). (E) A curve plot displays the density distribution of hydrophobicity (GRAVY) for identified peptides between in-gel and in-tip digestions. (F) Missed cleavage rates of the identified peptides using in-gel and in-tip digestions. (G) A curve plot shows log10-transformed mean protein intensities ranked by protein abundance. Mouse hepatocyte-specific markers are shown as circles. Markers exclusively identified under in-tip digestion are specifically labeled. (H) Coefficient of variation of the quantified peptides and proteins abundance from in-gel and in-tip digestions. (I) Pearson correlation for peptide and protein quantification from in-gel and in-tip digestions. (J) The protein types and subcellular locations of the identified proteins enriched through Ingenuity Pathway Analysis (IPA) are presented. (K) Study design presents a comparative proteomic analysis among different tissue volumes processed by in-gel and in-tip digestions. Gels are prepared for two consecutive mouse liver FFPE slices, and punches (n = 7) for each of the diameters (350 µm, 500 µm, 750 µm, 1000 µm, 1500 µm, 2000 µm, and 3000 µm) are collected separately for in-gel (1000 µm - 3000 µm) and in-tip (350 µm - 3000 µm) digestions. The combined peptides prepared under each diameter using different methods are acquired with 1-hour data-dependent acquisition mass spectrometry (DDA-MS; n = 3) and data-independent acquisition mass spectrometry (DIA-MS; n = 3), respectively. (L) The number of identified peptides and proteins for different tissue volumes with corresponding diameters shown in (K). The samples are processed by in-gel and in-tip digestions, and MS data are collected by DDA-MS and DIA-MS methods, respectively. All data points are presented as mean values ± SD.

Figure 3

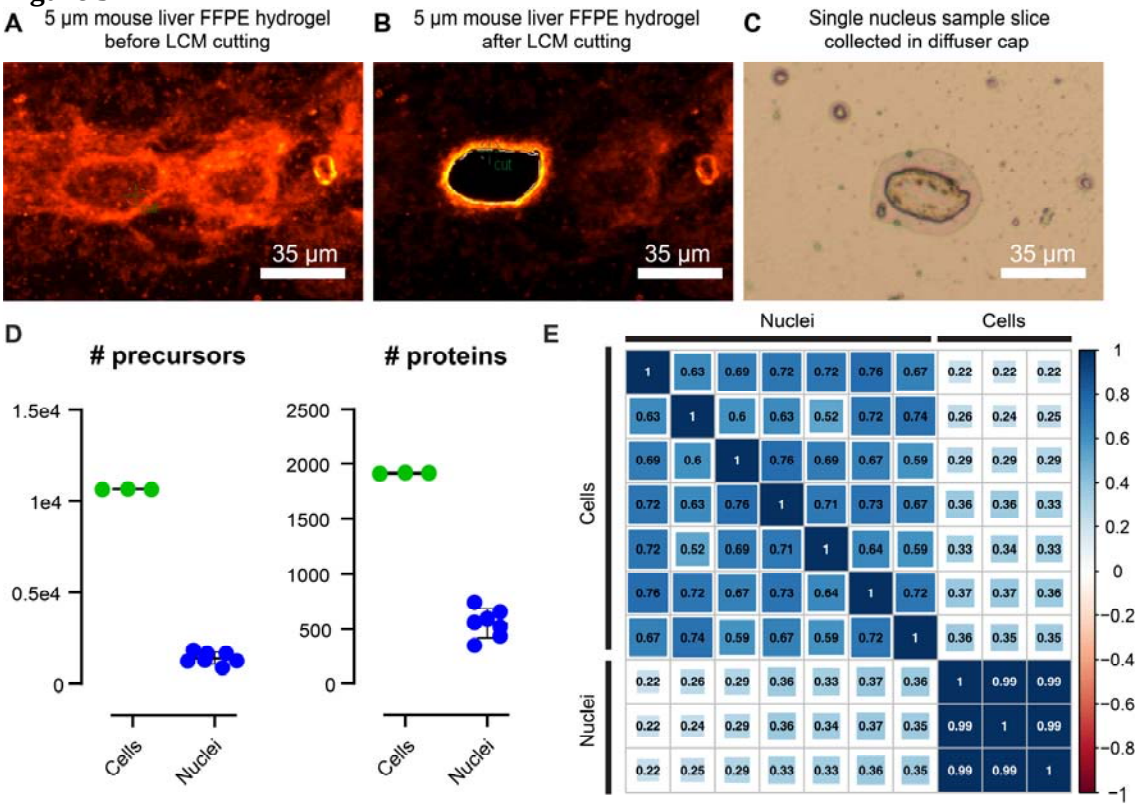

Figure 3. Study of single nucleus collection and analysis through tissue expansion, laser capture microdissection and filter-aided in-gel digestion. (A) Representative image of a mouse liver FFPE slice after expansion before laser capture microdissection (LCM) single nucleus cutting. (B) Representative image of the same sample after LCM single nucleus cutting. (C) Bright field image of the single nucleus in the diffuser cap. Scale bars are indicated within each figure. (D) Number of peptide and protein identifications between mouse hepatocytes and single nucleus in DIA mode through library-free search (data for mouse hepatocytes is from 0.042 nL of In-tip DIA-MS in Figure 2L, n=3; data for single nucleus is n=7 biological replicates from one liver slice). Data are presented as mean values ± SD. (E) Heatmap of Pearson correlations for protein quantification for each sample pair from mouse hepatocytes and nuclei. The color bar indicates the values of Pearson correlations.

Figure 4

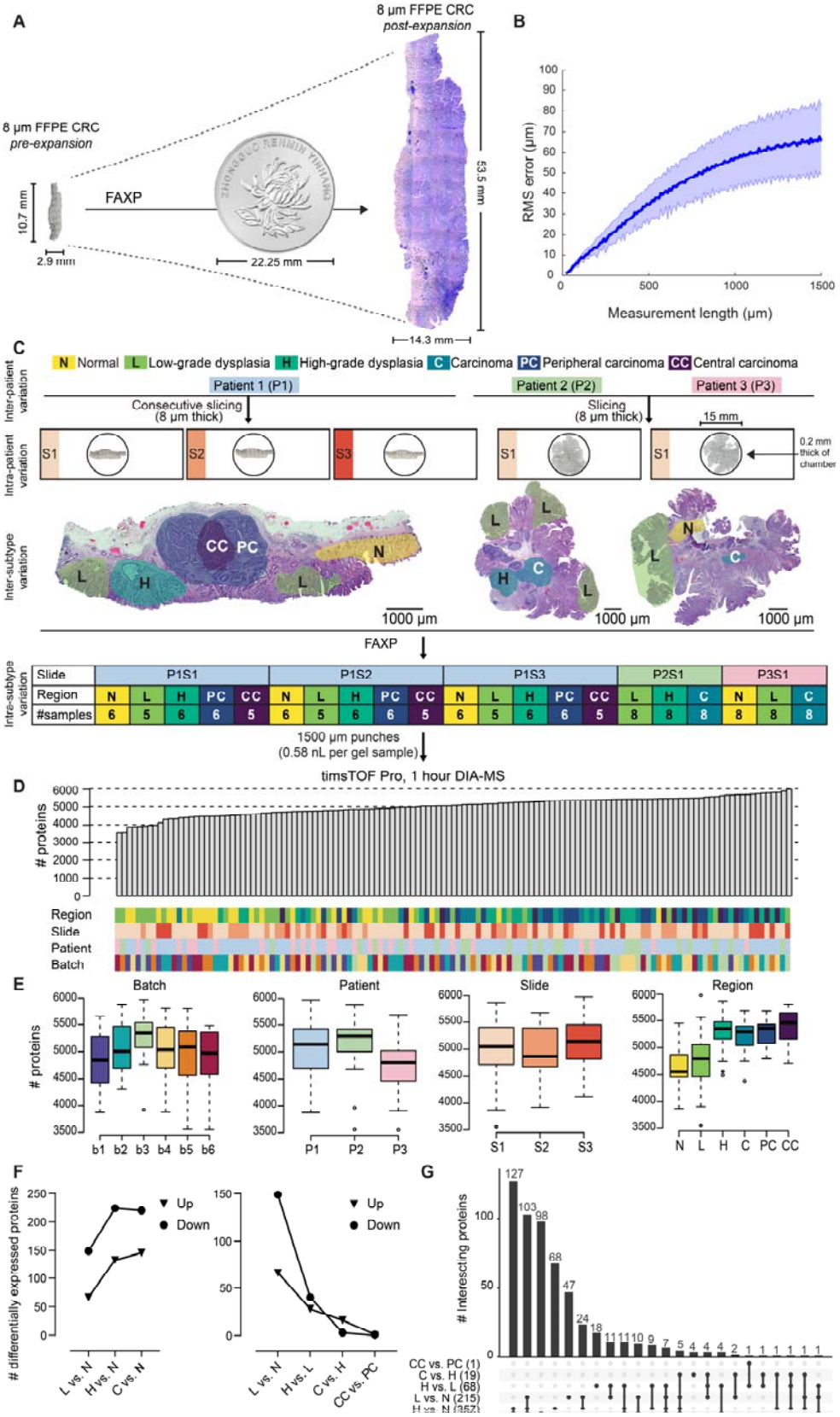
A
8 µm FFPE CRC
post-expansion
8 µm FFPE CRC
pre-expansion
FAXP
10.7 mm
2.9 mm
22.25 mm
53.5 mm
14.3 mm
B
RMS error (µm)
Measurement length (µm)
C
N Normal L Low-grade dysplasia H High-grade dysplasia C Carcinoma PC Peripheral carcinoma CC Central carcinoma
Inter-patient variation
Patient 1 (P1)
Patient 2 (P2)
Patient 3 (P3)
Consecutive slicing (8 µm thick)
Slicing (8 µm thick)
15 mm
Intra-patient variation
S1
S2
S3
0.2 mm thick of chamber
Inter-subtype variation
1000 µm
FAXP
Intre-subtype variation
Slide
Region
#samples
P1S1
P1S2
P1S3
P2S1
P3S1
1500 µm punches (0.58 nL per gel sample)
timsTOF Pro, 1 hour DIA-MS
D
proteins
Region
Slide
Patient
Batch
E
Batch
Patient
Slide
Region
proteins
b1 b2 b3 b4 b5 b6
P1 P2 P3
S1 S2 S3
N L H C PC CC
F
differentially expressed proteins
Up
Down
L vs. N
H vs. N
C vs. N
H vs. L
C vs. H
CC vs. PC
G
Interescting proteins
CC vs. PC (1)
C vs. H (19)
H vs. L (68)
L vs. N (215)

Figure 4. Application of FAXP on colorectal cancer patient samples. (A) Representative bright-field images of human colorectal cancer (CRC) slices before (left) and after (right) expansion using FAXP, in comparison to a one-dollar RMB coin, with a diameter of 22.25 mm. Expanded slices are stained with Coomassie brilliant blue for visualization. (B) Quantitative comparison of root-mean-square (RMS) measurement length error between original CRC slice images and corresponding expanded images (n = 3); one example is shown in (A). (C) Study design for the CRC application. Four different variations of patient samples are included - inter-patient variation: analysis of samples from three different patients (P1, P2, P3) to assess protein expression differences across individuals; intra-patient variation: examination of longitudinal variations using three consecutive slides from the first patient (P1S1, P1S2, P1S3); inter-subtype variation: investigation of proteomic changes across disease progression subtypes (N - normal, L - low-grade dysplasia, H - high grade dysplasia, C - carcinoma, PC - peripheral carcinoma, CC - central carcinoma); intra-subtype variation: exploration of protein expression variations within a specific disease subtype. H&E staining is performed on FFPE sample slides from three patients (P1, P2, P3). The slice IDs (P1S1, P1S2, P1S3, P2S1, P3S1) are associated with specific subtypes. The number of mass spectrometry (MS) samples for each subtype associated with each slice are indicated. 1500-µm punches (containing 0.58 nL of tissue volume before expansion) are subjected to filter-aided in-gel digestion and MS data are collected by timsTOF Pro using 1 hour DIA-MS. (D and E) Distribution of identified proteins are presented based on MS batches, patient numbers, slide numbers, and subtype regions, respectively. (F) Number of differentially expressed proteins (DEPs) in different comparisons. (G) Upset diagram showing the DEPs overlaps for selected comparisons.

Figure 5

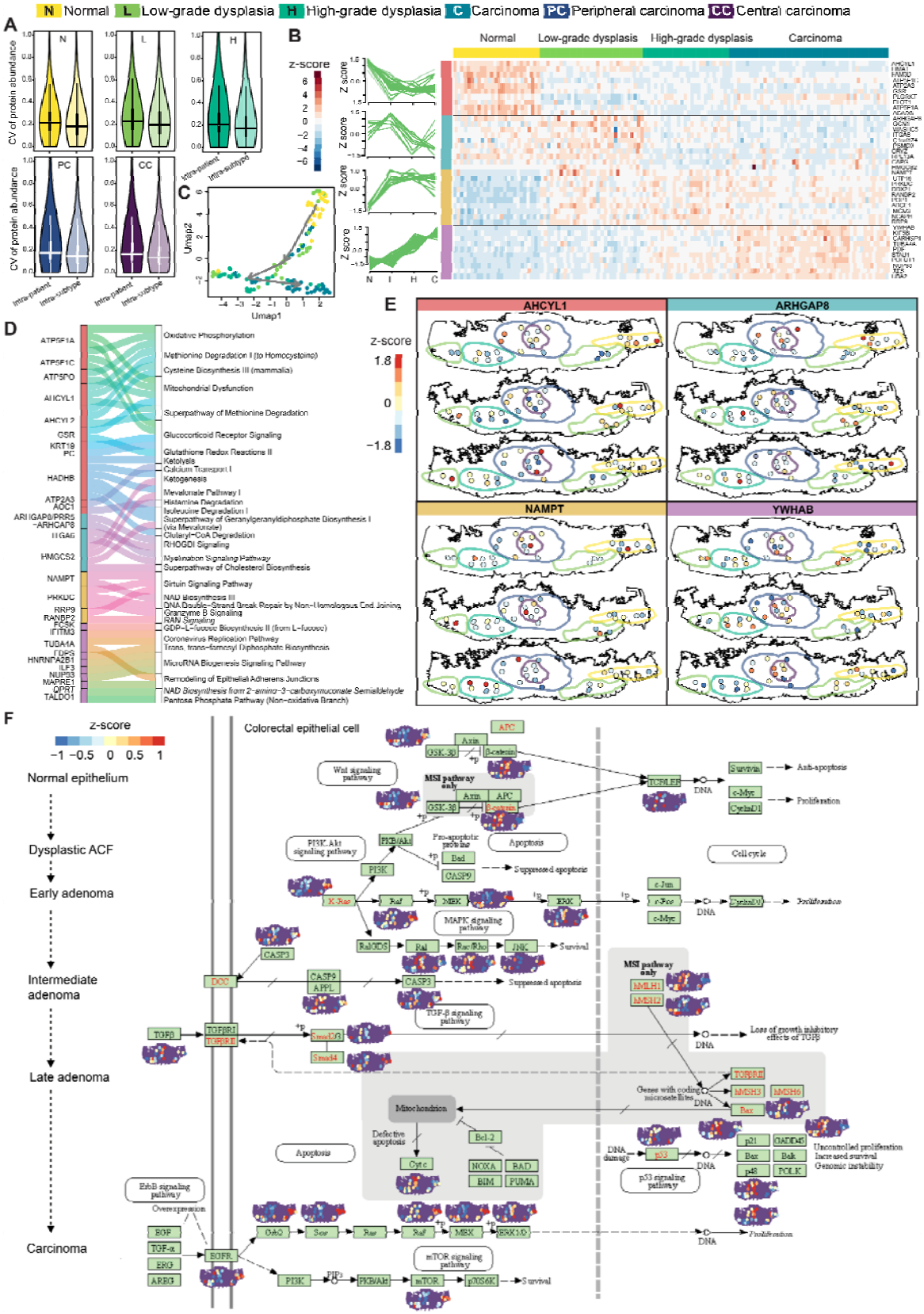
N Normal
L Low-grade dysplasia
H High-grade dysplasia
C Carcinoma
PC Peripheral carcinoma
CC Central carcinoma
A
CV of protein abundance
Intra-patient
Intra-subtype
B
z-score
Normal
Low-grade dysplasia
High-grade dysplasia
Carcinoma
C
Umap1
Umap2
D
Oxidative Phosphorylation
Methionine Degradation I (to Homocysteine)
Cysteine Biosynthesis III (mammalia)
Mitochondrial Dysfunction
Superpathway of Methionine Degradation
Glucocorticoid Receptor Signaling
Glutathione Redox Reactions II
Ketolysis
Calcium Transport I
Ketogenesis
Mevalonate Pathway I
Histamine Degradation
Isoleucine Degradation I
Superpathway of Geranylgeranyldiphosphate Biosynthesis I (via Mevalonate)
Oleoyl-CoA Degradation
RHOGDI Signaling
Myelination Signaling Pathway
Superpathway of Cholesterol Biosynthesis
Sirtuin Signaling Pathway
NAD Biosynthesis III
DNA Double-Strand Break Repair by Non-Homologous End Joining
Granzyme B Signaling
RAN Signaling
GDP-L-fucose Biosynthesis II (from L-fucose)
Coronavirus Replication Pathway
Trans, trans-Farnesyl Diphosphate Biosynthesis
MicroRNA Biogenesis Signaling Pathway
Remodeling of Epithelial Adherens Junctions
NAD Biosynthesis from 2-amino-3-carboxymuconate Semialdehyde
Pentose Phosphate Pathway (Non-oxidative Branch)
E
AHCYL1
ARHGAP8
NAMPT
YWHAB
z-score
1.8
0
−1.8
F
z-score
−1 −0.5 0 0.5 1
Normal epithelium
Dysplastic ACF
Early adenoma
Intermediate adenoma
Late adenoma
Carcinoma
Colorectal epithelial cell
Wnt signaling pathway
MSI pathway only
PI3K-Akt signaling pathway
Apoptosis
Cell cycle
MAPK signaling pathway
TGF-β signaling pathway
Mitochondrion
p53 signaling pathway
ErbB signaling pathway
mTOR signaling pathway
Survival
Anti-apoptosis
Proliferation
Suppressed apoptosis
Loss of growth inhibitory effects of TGFβ
Uncontrolled proliferation
Increased survival
Genomic instability

Figure 5. Spatial analysis of colorectal proteome expression. (A) Coefficient variations of protein abundance are depicted between three consecutive slides of the same subtype (intra-patient variation) and among all punches within the same subtype (intra-subtype variation). (B) Featured proteins, identified using a random forest-based machine learning model, are subjected to clustering using the Mfuzz algorithm. Heatmaps display the top 10 featured proteins in each cluster. (C) UMAP visualizations of all subtype samples from all patients by the features obtained from the random forest machine-learning model, with arrows connecting the centroid of each subtype. (D) A Sankey plot illustrates the correlation between enriched pathways obtained from IPA and proteins grouped based on Mfuzz clusters. (E) Spatial z-scored visualizations exhibit the expression patterns among three consecutive slides of the first patient, using the most important feature in each cluster as shown in (B). (F) Colorectal distribution maps present key proteins with average expression patterns of the three consecutive slides of the first patient involved in the KEGG pathway of colorectal cancer.

Figure 6

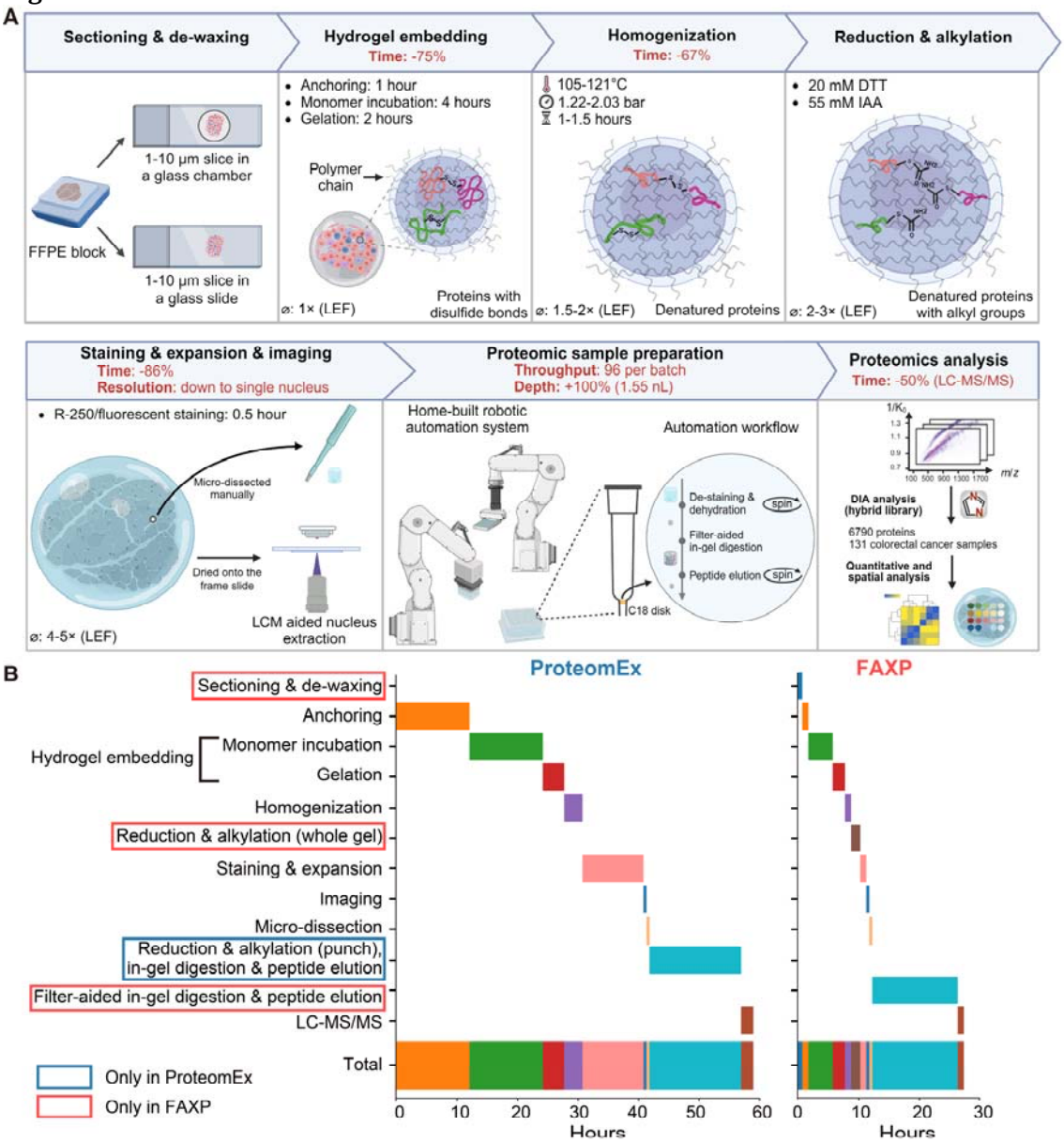
A
Sectioning & de-waxing
Hydrogel embedding
Time: -75%
Homogenization
Time: -67%
Reduction & alkylation
1-10 µm slice in a glass chamber
FFPE block
1-10 µm slice in a glass slide
• Anchoring: 1 hour
• Monomer incubation: 4 hours
• Gelation: 2 hours
Polymer chain
ø: 1× (LEF)
Proteins with disulfide bonds
105-121°C
1.22-2.03 bar
1-1.5 hours
ø: 1.5-2× (LEF)
Denatured proteins
• 20 mM DTT
• 55 mM IAA
ø: 2-3× (LEF)
Denatured proteins with alkyl groups
Staining & expansion & imaging
Time: -86%
Resolution: down to single nucleus
Proteomic sample preparation
Throughput: 96 per batch
Depth: +100% (1.55 nL)
Proteomics analysis
Time: -50% (LC-MS/MS)
• R-250/fluorescent staining: 0.5 hour
Micro-dissected manually
Dried onto the frame slide
LCM aided nucleus extraction
ø: 4-5× (LEF)
Home-built robotic automation system
Automation workflow
De-staining & dehydration
spin
Filter-aided in-gel digestion
Peptide elution
spin
C18 disk
DIA analysis (hybrid library)
6790 proteins
131 colorectal cancer samples
Quantitative and spatial analysis
B
ProteomEx
FAXP
Sectioning & de-waxing
Anchoring
Hydrogel embedding
Monomer incubation
Gelation
Homogenization
Reduction & alkylation (whole gel)
Staining & expansion
Imaging
Micro-dissection
Reduction & alkylation (punch), in-gel digestion & peptide elution
Filter-aided in-gel digestion & peptide elution
LC-MS/MS
Total
Only in ProteomEx
Only in FAXP
0
10
20
30
40
50
60
Hours
0
10
20
30
Hours

Figure 6. Comparative analysis of FAXP and ProteomEx. (A) Workflow of FAXP. The FAXP workflow begins with the sectioning and de-waxing of FFPE tissue block. Subsequently, the samples undergo treatment with a chemical anchor, followed by embedding into a hydrogel and homogenization using an autoclave. Either the updated Coomassie brilliant blue staining or fluorescent staining happen after reduction and alkylation on the whole-gel level, with expansion and imaging steps. After imaging, the expanded samples are micro-dissected manually or cut through laser capture microdissection (LCM) system, and peptides are collected through filter-aided in-gel digestion under the home-built robotic automation system for DIA-mass spectrometry (MS) data acquisition. For the human FFPE colorectal cancer samples, the MS-based proteomic data is searched in library-free mode using DIA-NN, generating a predicted library that is later combined with a pan-human library for a library-based search, which results in a protein matrix for downstream analysis. The spatial proteome expression data are further analyzed for spatial analysis. Notably, significant performance enhancements achieved in specific steps of this workflow are indicated in red below those steps (LEF - linear expansion factor). (B) Comparison of processing timeline. The bars represent the total duration time of each step under different workflows. Blue rectangles denote steps exclusively used under the ProteomEx, while red rectangles correspond to steps used under FAXP. Hands-on time is not displayed, as both the ProteomEx and FAXP workflows involve similar hands-on time when dealing with small sample amounts. Besides, ProteomEx involves reduction and alkylation on individual punches during conventional in-gel digestion, while FAXP performs these steps on the entire gel without further reduction and alkylation during filter-aided in-gel digestion. The faster processing time and improved workflow of FAXP are achieved from shorter duration time for various steps compared to ProteomEx. The data in FAXP are based on the processing of 5 μm FFPE mouse liver.