Proteome Science

REVIEW

Open Access

Spatial proteomics: unveiling the multidimensional landscape of protein localization in human diseases

Mengyao Wu[1], Huihui Tao[1,3,4*†], Tiantian Xu[1], Xuejia Zheng[2], Chunmei Wen[1], Guoying Wang[1], Yali Peng[1] and Yong Dai[1,2,5†]

Abstract

Spatial proteomics is a multidimensional technique that studies the spatial distribution and function of proteins within cells or tissues across both spatial and temporal dimensions. This field multidimensionally reveals the complex structure of the human proteome, including the characteristics of protein spatial distribution, dynamic protein translocation, and protein interaction networks. Recently, as a crucial method for studying protein spatial localization, spatial proteomics has been applied in the clinical investigation of various diseases. This review summarizes the fundamental concepts and characteristics of tissue-level spatial proteomics, its research progress in common human diseases such as cancer, neurological disorders, cardiovascular diseases, autoimmune diseases, and anticipates its future development trends. The aim is to highlight the significant impact of spatial proteomics on understanding disease pathogenesis, advancing diagnostic methods, and developing potential therapeutic targets in clinical research.

Keywords Spatial proteomics, Mass spectrometry, Fluorescent antibody, Disease, DSP

Introduction

Proteins are the functional molecules of all cells and the effectors of all biological processes. Spatial proteomics is a multidimensional advanced technique focused on exploring the spatial distribution and interactions of proteins within cells and tissues. This field aims to provide a profound understanding of the subcellular and tissue-level localization, distribution, and interactions of proteins, offering unprecedented insights into organelle functions and cellular biological processes. This review focuses on tissue-level spatial proteomics, which extends beyond subcellular structures to analyze protein distribution and expression patterns on a broader scale.

The fundamental principle involves studying proteins through techniques such as ifluorescence labeling, mass spectrometry, and imaging. The integrated analysis of these diverse techniques reflects the multidimensionality of spatial proteomics, providing various layers of information such as protein quantification, visual distribution, and spatial imaging, thereby offering a comprehensive view of protein localization and interactions. Given that proteins directly reflect the functional state of cells and vary in expression due to cell type, life cycle, disease states, and treatment methods [1, 2], spatial proteomics

†Huihui Tao and Yong Dai two authors contributed equally to this work.

*Correspondence:
Huihui Tao
475433093@qq.com
[1] School of Medicine, Anhui University of Science & Technology, Huainan, China
[2] The First Hospital of Anhui University of Science and Technology, Huainan, China
[3] Key Laboratory of Industrial Dust Deep Reduction and Occupational Health and Safety of Anhui Higher Education Institutes, Huainan, China
[4] Anhui Province Engineering Laboratory of Occupational Health and Safety, Huainan, China
[5] Present Address: Joint Research Center for Occupational Medicine and Health of IHM, Anhui University of Science and Technology, Huainan, China

is critically important for the study of disease mechanisms, analysis of molecular mechanisms, and identification of biomarkers. In recent years, rapid technological advancements have allowed spatial proteomics to achieve significant milestones in academic research globally, particularly in constructing organ tissue spatial atlases, studying microenvironments and diseases, exploring cell interactions, and identifying biomarkers and drug targets [3–8] (Fig. 1). Despite challenges in sample preparation, data acquisition, and analysis [9], the potential and value of spatial proteomics are increasingly recognized, making it a key tool in driving medical and disease research. This review aims to outline the characteristics of spatial proteomics and its application in fields such as neurodegenerative diseases, cancer, and cardiovascular diseases, elucidating changes in disease-related protein spatial expression profiles to foster the discovery of new diagnostic markers and the development of new treatment methods.

Overview of spatial proteomics

Characteristics of spatial proteomics

Spatial proteomics is an innovative approach that combines mass spectrometry with advanced imaging

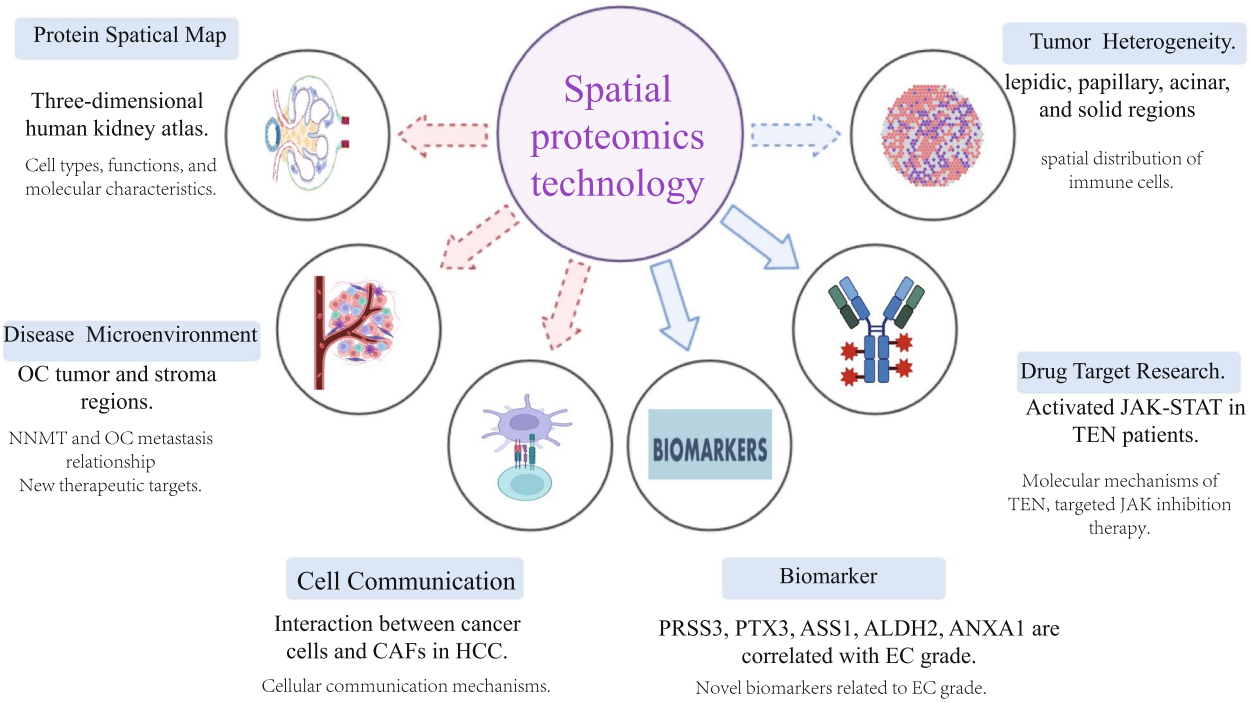

Fig. 1 Examples of related findings and troubleshooting in spatial proteomics. From left to right, the figure illustrates: spatial proteome mapping, microenvironment and disease, cell–cell interaction, biomarker discovery, drug target development, and tumor heterogeneity research. Spatial Proteome Mapping: A high-resolution 3D reference atlas of the human kidney has been constructed, encompassing cellular, pathway, and genetic elements. This atlas reveals the cell types, functions, and molecular characteristics of a healthy kidney, facilitating the understanding of kidney disease classification and pathogenesis (refer to reference [3]). Microenvironment and Disease: In ovarian cancer (OC), non-enzymatic methyltransferase (NNMT) is upregulated in the tumor microenvironment and has been identified as a regulator promoting metastasis. Experimental validation shows that NNMT expression correlates with shorter survival periods, and inhibiting its activity can effectively suppress cancer metastasis. This research elucidates the relationship between NNMT and OC metastasis, enhancing the understanding of OC metastasis mechanisms (reference [4]). Cell–Cell Interaction: In hepatocellular carcinoma (HCC), cancer-associated fibroblasts (CAFs) and their protein expression differences were studied. It was found that proteins released by CAFs interact with the TLR2 receptor on HCC cancer cells, influencing HCC progression. This study addresses the challenge of isolating proteins from specific cells and reveals cell communication mechanisms in liver cancer progression (reference [5]).Biomarker Discovery: In endometrial cancer (EC), five proteins—PRSS3, PTX3, ASS1, ALDH2, and ANXA1—were identified in sentinel lymph nodes (SLNs) of EC patients. Their expression is significantly correlated with tumor grading, providing insights for early diagnosis and grading of EC (reference [6]). Drug Target Development: In toxic epidermal necrolysis (TEN), research has shown that interferon signaling pathways are active and pSTAT1 is activated in immune cells and keratin-forming cells of TEN patients. This supports the safety of using JAK inhibitors (JAKi) to treat TEN, highlighting their potential as drug targets (reference [7]).Tumor Heterogeneity: In melanoma heterogeneity research, specific proteins such as DMBT1 and MARCKS exhibit varied expression in squamous, papillary, acinar, and solid regions of tumors and are associated with recurrence risk. Immunosuppressive cells increase in the core region of solid tumors. This study addresses the challenge of understanding melanoma tumor heterogeneity, clarifying its internal molecular characteristics and immune cell distribution.(reference [8])

techniques, allowing researchers to accurately obtain high-resolution localization information of proteins within cells. Subcellular spatial proteomics has previously been summarized. For example, Lundberg et al. extensively covered methods and applications, encompassing mass spectrometry-based organelle analysis, protein–protein interaction networks, and whole-cell imaging of protein localization, highlighting the link between abnormal subcellular localization and diseases. Other researchers have reviewed various methods for subcellular protein localization, such as fluorescence imaging, protein proximity labeling, organelle purification, and mass spectrometry, emphasizing the importance of protein localization in cellular signaling, growth, proliferation, movement, and programmed cell death (see detailed techniques in references [10, 11]). Subcellular-level proteomics research has been utilized in cystic fibrosis (CF) studies. For instance, John J et al. analyzed the transport pathways of CFTR protein in CF, revealing that golgin proteins are involved in the aberrant transport of CFTR, which contributes to the disease's molecular pathology [12]. Meanwhile, Hirst et al. focused on the critical role of AP-5 in the transport process from late endosomes to the Golgi apparatus. Additionally, studies have shown that VX-809 (lumacaftor) induces extensive mitochondrial remodeling, facilitating CFTR functional restoration [13, 14]. These studies enhance the understanding of CF's molecular pathology and facilitate the development of new diagnostic and therapeutic targets. Given this, the review will focus on the introduction of spatial proteomics at the tissue level, building on the solid foundation provided by subcellular spatial protein analysis.

In gene expression research, although mRNA levels are typically considered direct indicators of gene activity, the correlation between mRNA and protein abundance is often low, failing to accurately reflect actual protein abundance. This lack of correlation arises from various mechanisms: mRNA molecules are prone to degradation due to regulation by RNA-binding proteins and miRNAs, and translation efficiency can limit protein synthesis [15, 16]. Moreover, post-translational modifications and protein transport barriers or mislocalization, as well as protein degradation pathways like the ubiquitin–proteasome system, further affect protein function and abundance [17, 18]. Thus, measuring mRNA levels alone cannot fully reflect the actual distribution and functional state of proteins. Spatial proteomics not only allows for the direct monitoring and quantification of protein changes and modifications but also provides new perspectives and methods for discovering disease-related protein biomarkers and understanding the complex structure of the human proteome in disease contexts.

Traditional proteomics primarily focuses on the compositional structure and function of proteins, reflecting the overall protein expression changes in an organism through identification and quantitative analysis. In contrast, spatial proteomics offers higher spatial resolution and comprehensiveness, allowing direct analysis of protein localization and interactions [19]. This aids in comprehensively understanding disease mechanisms and discovering new protein biomarkers and drug targets. Spatial proteomics enables high-throughput quantitative and localization analysis of proteins in histological samples. Depending on the quantification method, current spatial detection techniques for proteins are mainly divided into three categories: fluorescence-based antibody, mass spectrometry-based, and sequencing-based spatial proteomics methods.

Fluorescence-based antibody spatial proteomics methods

Fluorescence-based antibody spatial proteomics methods employ specific antibodies or fluorescent probes to label target proteins, ensuring precise protein localization. Fluorescence microscopy is then used to observe and record the spatial distribution of proteins at the tissue and cellular levels, offering high resolution. Immunohistochemistry was the earliest method in this category, detecting protein expression in tissue sections using specific antibodies. Its advantage lies in preserving the original tissue structure of proteins. However, a limitation of this traditional approach is that it requires precise antibody matching and can only localize one protein per experiment. To increase the diversity of protein detection and overcome the limitations of immunohistochemistry, some research groups have introduced multiplex immunofluorescence techniques. This method extends protein detection diversity by using multiple fluorescent probes, allowing simultaneous detection of up to 100 protein biomarkers in a single tissue section, significantly enhancing high-throughput analysis capabilities [20]. Nevertheless, it encounters challenges of spectral overlap with different fluorescence signals, limiting its applicability for quantitative analysis of multiple proteins in complex samples. Tyramide Signal Amplification (TSA), a popular multiplex immunofluorescence technique, uses tyramide enzyme to enhance the fluorescent signal of target proteins, improving protein sensitivity [21]. Increasing signal intensity facilitates the detection of low-abundance proteins but introduces potential issues of cross-reactivity or non-specific binding, requiring longer multiple staining cycles. This method is suitable for diverse labeling or localization of low-abundance proteins. The development of sequential immunofluorescence (seqIF) has addressed

these challenges. This technique automates the application of antibodies and their elution in a continuous process, combining integrated microscopy with microfluidic chips for in situ imaging, optimizing sample processing workflows and providing a reliable method for detecting multiple biomarkers [22]. It features efficient antibody incubation and washing capabilities, yet its experimental procedures are complex, suitable for high-resolution localization and interaction studies of proteins in tissue or cell culture samples. One of the most representative imaging-based spatial proteomics methods is CODEX, which primarily uses antibody-conjugated DNA barcodes, multiple rounds of staining and elution, and imaging systems to identify barcode sequences, achieving high-throughput detection of multiple proteins in samples. The method allows simultaneous detection of multiple high-resolution markers, with a detection limit of 50 proteins [23, 24]. It is applicable to various sample types such as fresh and fixed tissues (e.g., paraffin sections), but its dependence on antibodies limits its application to a broader range of markers (Table 1). It is well-suited for highly multiplexed biomarker detection and spatial distribution analysis at the single-cell level, particularly in exploring tumor microenvironments.

Mass spectrometry-based spatial proteomics methods

As mass spectrometry technology has rapidly evolved, it has become a mainstream technique for protein analysis. It identifies proteins based on the sequence of their unique peptides. The principle involves extracting proteins from tissue samples and subjecting them to proteolytic digestion with trypsin, followed by separation of the protein peptides using a mass spectrometer. The analysis of mass spectrometric signals by the spectrometer and image analysis software reveals changes in protein expression. Spatial proteomics methods based on mass spectrometry are divided into spatially untargeted and targeted proteomics.

Non-targeted mass spectrometry methods maximize proteome coverage, commonly used for biomarker development [31]. Liquid chromatography-mass spectrometry (LC–MS) was one of the earliest spatially untargeted methods, combining the advantages of liquid chromatography and mass spectrometry for the separation and purification of proteins in mixtures [26]. LC–MS can analyze complex biological samples such as blood, urine, and cell lysates with high sensitivity in studying protein quantification, post-translational modifications, and protein interactions. It is suitable for research on disease biomarkers and drug development. However, limitations in automation and complex sample preprocessing restrict its widespread application in clinical disease research. In contrast, Mass Spectrometry Imaging (MSI) is a relatively fast and cost-effective method that, as a key tool combining mass spectrometry and imaging technologies, allows for label-free acquisition of data on metabolites, lipids, peptides, and proteins from tissue slices. MSI is suitable for fresh or frozen tissue sections, requiring minimal sample preparation and using various ionization techniques to measure hundreds of potential analytes' relative spatial abundance in almost any tissue section [32, 33]. Sample preparation for MSI is relatively simple, with high performance, but limited quantitative capabilities. It is ideal for unbiased spatial distribution analysis at the tissue level, such as protein expression patterns in cancer tissues or drug distribution in tissues. Secondary ion mass spectrometry (SIMS), the first MS imaging technology applied in the 1960s, is suitable for solid

Table 1 Summary of key characteristics of major spatial proteomics techniques

Categories	Methods	Samples	Advantages	Disadvantages	Reference
FA-based	MIF	FFPE/FF	High specificity Multiplex biomarkers	Fluorescence overlap interference	[20]
	CODEX	FFPE/FF	Highly multiplexed detection; fewer fluorescent	Antibody dependency	[23, 24]
MS-based	MALDI-MSI	FFPE/FF	No fluorescence interference High throughput	susceptible to matrix effects. Difficulties in detecting low abundance proteins.proteins	[25]
	DESI-MSI	FFPE/FF	Simple sample preparation. fast imaging	Low resolution and sensitivity; ionization affected by spraying	[26]
	LCM-MS	FFPE/FF	High precision, biomarker screening	Limited resolution	[27]
	IMC	FFPE/FF	No fluorescence interference.higher resolution	Complex equipment; high cost	[28]
	MIBI	FFPE/TMA	Ion beam detection; higher resolution	Complex equipment; slow throughput, long cycles	[29]
Seq-based	DSP	FFP/FF/TMA	Suitable for diverse samples, high sensitivity	antibody dependency	[30]

Legend: *FA* Fluorescence Antibody, *MS* Mass Spectrometry, *Seq* Sequencing, *MIF* Multiplex immunofluorescence, *CODEX* co-detection by indexing, *MALDI-MSI* Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging, *DESI-MSI* Desorption Electrospray Ionization Mass Spectrometry Imaging, *LCM-MS* Laser Capture Microdissection Mass Spectrometry, *IMC* Imaging Mass Cytometry, *MIBI* Multiplexed Ion Beam Imaging, *DSP* Digital Spatial Profiling *FFPE* Formalin-Fixed Paraffin-Embedded, *FF* Frozen Section, *TAM* Tissue Microarray

samples like tissue sections or material surfaces. SIMS achieves extremely high spatial resolution by bombarding the sample surface with an ion beam and analyzing the generated mass spectrometry signals. However, its ionization capability for proteins limits its application in complex protein samples [34]. SIMS is suitable for high-resolution imaging analysis of protein expression in specific regions or cell types but requires highly pure samples and clean surfaces. The most typical imaging mass spectrometry technique introduced thereafter was Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging (MALDI-MSI). MALDI-MSI typically combines with time-of-flight (TOF) mass spectrometry. It suits various sample types. requiring uniform matrix coating on tissue, followed by laser ablation of the solid matrix and analytes from tissue sections, with TOF analyzers deducing the mass-to-charge ratio of desorbed ions to determine ion mass [25]. Although the spatial resolution of MALDI-MSI, ranging from 5–50 μm [25, 35], is lower compared to SIMS technology, and the choice of matrix significantly influences ionization results, presenting challenges in detecting low-abundance proteins, it is characterized by high throughput, capable of detecting 50–100 peptides without destroying tissue. Therefore, this method is suitable for quantifying the spatial distribution of multiple molecules and biomarkers in tissue sections.It has been widely applied in quantifying and detecting various biomarkers in tissues, serum, plasma, cerebrospinal fluid, and urine [36, 37]. Desorption Electrospray Ionization Mass Spectrometry Imaging (DESI-MSI) technology, which developed subsequently, uses an electrospray ion source to generate ions on the surface of liquid samples [38], allowing direct molecular information retrieval from tissue surfaces without the need for matrix selection, thus avoiding the matrix effects observed in MALDI-MSI. DESI-MSI is suitable for various sample types with simple preparation. However, it has lower spatial resolution and is used in studies that require rapid screening of large-scale samples or do not necessitate high-resolution information. Although MALDI-MSI and DESI-MSI can unbiasedly analyze the spatial distribution of molecules within tissues, they also have limitations. The lack of predefined targets increases the difficulty of identifying peptides in the vast amount of data. High-efficiency chromatographic separation techniques can effectively separate complex samples and reduce matrix effects, thereby enhancing resolution. The absence of chromatographic pre-separation in the aforementioned methods results in complex and overlapping signals in mass spectrometry data, where low-abundance peptides are often overshadowed by high-abundance signals, limiting the capacity for in-depth analysis. Additionally, sample heterogeneity and variations in ionization efficiency pose challenges for accurate quantitative analysis. To overcome these challenges, laser capture microdissection (LCM) technology has emerged. LCM is an advanced tissue separation technology that combines laser cutting and image analysis, allowing for the rapid, precise selection and localization of regions of interest (ROI) from tissue samples. This one-step collection avoids sample contamination from manual handling and achieves separation at the single-cell level without losing spatial information [39, 40]. Laser Capture Microdissection-Mass Spectrometry (LCM-MS) is the most widely applied method. It involves processing and mass spectrometry analysis of fixed, stained tissue sections, correlating spatial information with data through image analysis software to identify protein expression and interactions within tissues [27] (Fig. 2).

Accurate analysis of protein expression in tissue samples is a key challenge in proteomics research, where LCM-MS has made significant advances, making it suitable for studying specific tissue separations. With the continuous development and expansion of LCM-MS's analysis capabilities, researchers can now achieve higher throughput in protein analysis. For example, a novel high-throughput protein mapping method combines voxelization techniques with high-throughput LC-FTICR mass spectrometry to analyze smaller tissue units in the mouse brain, successfully generating expression patterns for 1,028 proteins. This technique utilizes stable isotope labeling and AMT tag strategies to enhance the reliability and reproducibility of the analysis. However, further improvements in resolution are needed [41]. Min Ma et al. proposed the Micro-scaffold Assisted Spatial Proteomics (MASP) strategy, utilizing 3D-printed micro-scaffolds to achieve higher spatial resolution. They mapped over 5,000 proteins in the mouse brain, achieving uniform micro-compartmentalization and precise spatial information retention across the entire tissue. Although this requires more complex analysis procedures, the technique significantly enhances sample processing efficiency and proteomics analysis capability for small samples [42]. The development of NanoPOTS technology has enabled researchers to perform grid-based sampling at a 100 μm resolution on mouse uterine tissue sections, achieving automated mass spectrometry imaging of 2,000 proteins. This technique significantly improves upon LCM-MS, particularly through the use of microfluidic sample processing and automation, which reduces sample loss and enhances sensitivity and throughput [43]. Integrating LCM, MSI, and spatial perception algorithms, the research team revealed the spatial heterogeneity of protein expression and tumor-associated markers in human atypical teratoid/rhabdoid tumors. This approach further enhanced the identification of protein spatial expression

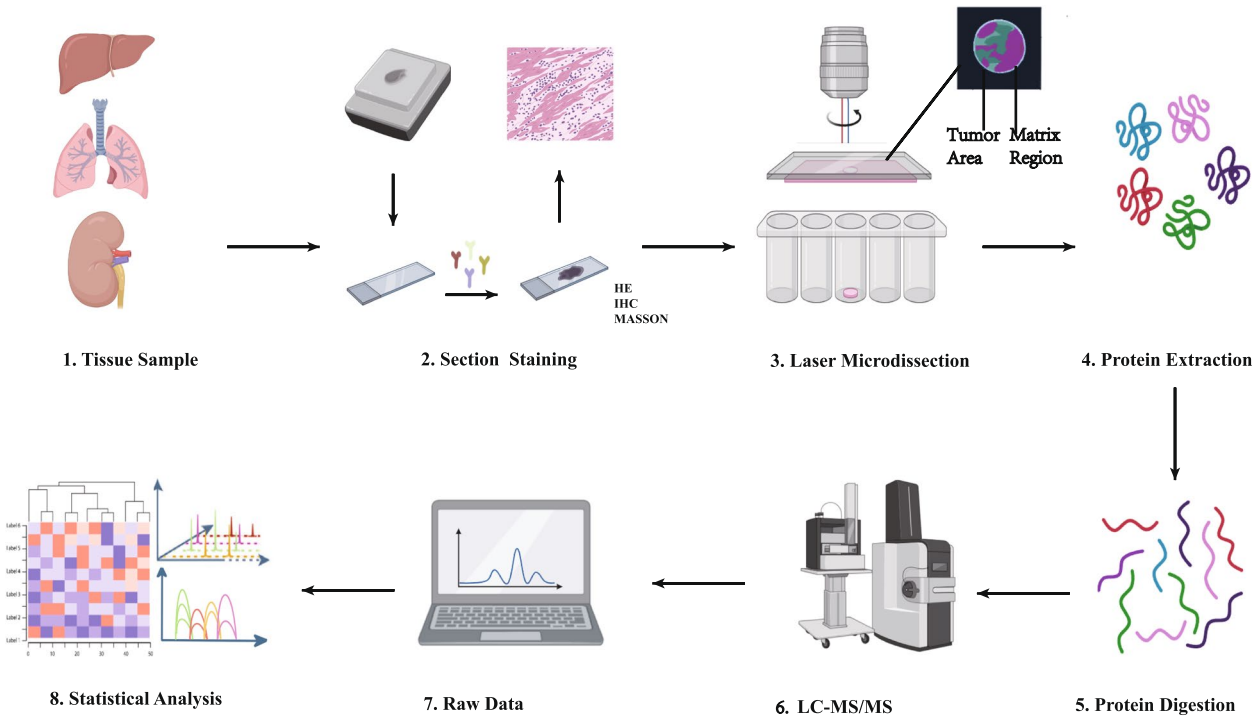

Fig. 2 Operational procedure of spatial proteomics mass spectrometry methods. 1. Sample Collection: Different tissue samples are collected. 2. Slice Staining: Tissue samples were sectioned and mounted onto glass slides. Following natural air-drying, appropriate staining techniques, such as Hematoxylin and Eosin (H&E) staining, immunohistochemical staining, Masson's trichrome staining, etc., were chosen based on experimental design. 3. Laser Fiber cutting: The laser microdissection system is used for precise localization and microdissectionof the regions of interest under a microscope. 4. Protein Extraction: were extracted using protease or protein extraction buffer.. 5. Protein Digestion: The extracted proteins were subjected to digestion to release peptide fragments. 6. LC–MS/MS: The peptide solution is injected into the liquid chromatography system, where separation of different peptide segments is achieved using a chromatographic column. Subsequently, the separated peptide segments are introduced into the mass spectrometer, where they are ionized. The mass spectrometer evaluates the protein mass and relative abundance by analyzing the mass-to-charge ratio of ions in the peptide segments. 7. Raw Data: The data can be preprocessed through baseline correction, peak fitting, and other methods. 8. Statistical Analysis: The utilization of data visualization tools for differential analysis on raw data allows for the revelation of information concerning the distribution and abundance of specific proteins across different spatial context

and provided a more in-depth characterization of tissue heterogeneity [44].

Spatial targeted mass spectrometry methods play crucial roles in current research for verifying pre-determined proteins. Key techniques include Imaging Mass Cytometry (IMC) and Multiplexed Ion Beam Imaging (MIBI). IMC involves staining tissue with metal-conjugated antibodies, followed by scanning the tissue with a laser to excite the metal ions and generate mass-to-charge signals, which are analyzed to localize and quantify proteins. It is suitable for high-dimensional spatial resolution studies like tumor microenvironment analysis, offering high multiplexing capability and robustness but with complex sample preparation and high costs [28]. MIBI, a variant of IMC, is based on secondary ion mass spectrometry analysis, distinguishing different antibody markers through ion beam-excited secondary ion signals [29]. It excels in high-resolution spatial analysis and can simultaneously analyze up to 100 targets, overcoming the resolution limitations of MALDI-MSI. It is ideal for highly multiplexed and high-resolution analysis of complex biological samples. However, the high cost of metal tags and long acquisition periods limit its widespread application [45, 46]. These methods still require improvements in tissue sample preparation and accuracy. Sample preparation issues like tissue washing and protease digestion may cause protein delocalization, affecting spatial resolution and accuracy. Recent research has proposed several improvements in sample preparation and data processing, such as Filter-Aided Sample Preparation (FASP), Single-Pot, Solid-Phase-enhanced Sample Preparation (SP3), and optimized MS data acquisition strategies [47–50]. These methods significantly improve sample preparation efficiency and data reliability, broadening mass spectrometry's application scope.

Sequencing-based spatial proteomics methods

Digital Spatial Profiling (DSP) is an emerging sequencing-based spatial proteomics and transcriptomics technology, providing a means for highly multiplexed analysis at the tissue section level. Suitable for tissue sections or cultured samples, DSP involves fixation and sectioning to ensure sample quality and efficiency. DSP's principle involves in situ binding of probe-tagged antibodies to tissue samples, selecting regions of interest (ROIs) through UV light exposure to separate probe tags from bound proteins, and collecting tags in a 96-well plate, each well corresponding to an ROI. Ncounter system quantifies probe tags in wells to determine protein expression abundance. DSP's advantages include high spatial targeting, allowing for multiple immunofluorescence pretreatments of tissue sections, selecting ROIs, and strong multiplexing capabilities and high sensitivity, which make it adaptable to diverse sample types. Nevertheless, the technique's dependency on antibodies and the risk of signal overlap could potentially affect the reliability of the results. DSP is widely used in various tissue sample studies, particularly in cancer, immune, and neurological disease research. For example, Cabrita et al. used DSP to study melanoma patient samples, analyzing molecular characteristics of tertiary lymphoid structures formed by tumor-associated CD8 + T cells and CD20 + B cells, identifying TLS-related molecules predictive of survival outcomes [51]. Another example involved DSP analysis of tumor and stroma regions in pancreatic cancer patient tissues, identifying key markers related to treatment, providing important clues for biomarker screening [30].

Data analysis

Analyzing spatial proteomics data is a complex process encompassing multiple steps and involving various specialized software and platforms for data processing and interpretation. While sample preparation is the starting point of the workflow, this review focuses on the critical tools needed from data acquisition to visualization. Below is a brief overview of the data analysis steps and the tools used.

Data acquisition and preprocessing

Data acquisition is divided into image data acquisition and protein data acquisition. For image data acquisition, the CM1950 high-precision cryostat from Leica Microsystems provides uniform, undamaged sections [52]. The 3D HISTECH Pannoramic SCAN is a whole-slide scanner for tissue section digitization, offering high-resolution imaging through built-in optical lenses and image sensors [53]. Its advantages include compatibility with various tissue stains, automated operation, image stitching capabilities, and a user-friendly interface. Additionally, the PALM MicroBeam laser microdissection system and Zeiss Zen software from ZEISS are used, with Zeiss Zen offering intelligent automation tools for image acquisition. Leica LAS X supports high-speed image acquisition and precise positioning and, when paired with an automated slide scanner, efficiently and automatically extracts data from microscope images, featuring an intuitive interface and powerful image processing capabilities [8].

For protein data acquisition, the timsTOF Pro mass spectrometer from Bruker is commonly used [44]. It integrates trapped ion mobility spectrometry (TIMS) with quadrupole time-of-flight (QTOF) technology, achieving a scanning speed of over 120 Hz without compromising resolution, thereby improving acquisition efficiency and sensitivity for comprehensive proteome coverage. The Orbitrap Astral mass spectrometer is also frequently used, known for its ability to integrate with various sample preparation and imaging techniques for precise peptide measurement, offering multifunctionality and data accessibility [54]. The Orbitrap Astral mass spectrometer is also frequently used, known for its ability to integrate with various sample preparation and imaging techniques for precise peptide measurement, offering multifunctionality and data accessibility [55]. These instruments require control software such as Xcalibur from Thermo Fisher Scientific [4], RawTools [56], and the rawrr R package [57] to read raw data files generated by the mass spectrometers.

Initial data processing is fundamental to data analysis, involving preliminary processing of image and mass spectrometry data to ensure data quality and format consistency. Data preprocessing includes noise reduction, background correction, and normalization. Image data preprocessing focuses on image segmentation, while mass spectrometry data focuses on peak extraction. Post-acquisition, image data is denoised and background corrected using tools like ImageJ, CellProfiler, and Fiji to eliminate non-specific fluorescence. Threshold or region segmentation algorithms are then used to identify different tissue regions, followed by multi-channel alignment and data standardization to mitigate technical variability. For mass spectrometry data, software such as MaxQuant [58] and Proteome Discoverer is used for peak extraction.

Database engines

Database search and parsing are central stages in mass spectrometry data analysis. This process involves constructing a theoretical spectral library based on protein sequences from the database and matching it with the preprocessed spectra to identify peptide sequences and protein information. Database searching is complex, with search engines playing a critical

role. During this process, steps such as aligning mass spectrometry data and correcting for shifts are completed to ensure data reliability, providing high-quality input for the search engines. Different mass spectrometry data acquisition methods correspond to different search engines, making it crucial to choose the appropriate database engine for data parsing. For instance, Sequest, developed by Thermo Fisher Scientific, is a classic search engine used for DDA data acquisition analysis, identifying peptides and proteins from tandem mass spectrometry (MS/MS) data. MaxQuant is a widely used database search engine for DDA acquisition, incorporating the Andromeda search engine, a probability-based engine using a target-decoy strategy to identify peptides and proteins and control the false discovery rate (FDR) [59]. Proteome Discoverer is another search engine offering extensive plugins and work support [60]. Initially designed for DDA data, its integrated features also support DIA acquisition data through extensions and integration with other software. The MSFragger search engine developed by the Nesvizhskii lab employs a unique fragment ion indexing algorithm to speed up searches, supporting various MS data formats for DDA acquisition peptide identification. When integrated with FragPipe, it also supports DIA acquisition database searches [61], providing a comprehensive pipeline combining multiple tools and algorithms for large-scale proteomic data processing and accurate peptide identification. X!Tandem is another search engine for identifying proteins in tandem mass spectrometry data, offering flexibility and scalability for different experimental conditions [62].

The ordered data acquisition characteristics of DDA (Data-Dependent Acquisition) make it suitable for direct use with database search engines. However, DIA (Data-Independent Acquisition) data collection is unordered, and its multiplexing capability allows for the simultaneous collection of MS2 data from multiple peptides, increasing data richness but also complexity. Therefore, DIA data cannot be directly used with traditional database search engines and requires specialized tools for processing.

DIA-NN [63] is a common tool for handling DIA data, utilizing deep neural networks to identify and quantify mass spectrometry signals, capturing a more comprehensive range of proteins and facilitating proteome coverage and in-depth analysis. MSPLIT-DIA's advantage lies in its unbiased data collection of all ions, enhancing protein identification capabilities [64]. These tools significantly improve the depth and accuracy of DIA data analysis, enriching the methodologies available for mass spectrometry data analysis.

Statistical analysis and visualization

GO analysis, KEGG analysis, and PPI analysis are commonly used functional annotation tools. EggNOG-mapper is recommended for GO annotation [65]. For KEGG analysis, tools like KEGG Mapper and KAAS are often used. KEGG Mapper is an online tool that allows users to map proteins to KEGG pathways [66]. STRING, a comprehensive PPI database, plays a crucial role in constructing and visualizing PPI networks. Diamond is another frequently used software, known for its fast matching of protein sequences to reference genomes, suitable for large-scale data comparisons. These analyses provide a foundation for statistical inference and visualization. Perseus is a multifunctional software tool widely used for the statistical analysis and visualization of proteomics data, supporting various experimental data processing and multi-language programming extensions [67]. Excel and R (such as LIMMA [68], clusterProfiler [69] and SPOT [70] are also options, though Excel is limited in handling large-scale datasets. SPOT provides a more advanced, flexible, and precise statistical method by evaluating the spatial distribution of cells across multiple candidate radii.Various visualization tools vividly display patterns and structures in the data. ggplot2 (R) and Plotly (Python) are two primary choices. ggplot2 offers highly customizable data visualization options, such as heatmaps and scatter plots, suitable for showcasing spatial protein distribution in tissues. Plotly is an open-source visualization tool supporting the creation of interactive charts and data analysis visualizations.Additionally, Cytoscape, InteractiVenn software [69], are used for visual data analysis, with InteractiVenn being particularly useful for Venn diagram analysis.

Multivariate statistical analysis, dimensionality reduction, and clustering analysis play crucial roles in visualizing and mining information from spatial proteomics data. Multivariate statistical analysis is used to uncover relationships among multiple variables in a dataset. Commonly used software packages include R, Python, and SAS [71]. Dimensionality reduction aims to reduce the spatial dimensions of data while retaining complex dataset features for visualization and analysis. Principal Component Analysis (PCA) is an unsupervised method for dimensionality reduction, transforming high-dimensional data into a lower-dimensional space to identify sources of data variance. Recommended software packages include R (e.g., prcomp, PCAtools packages) and Python.Partial Least Squares Discriminant Analysis (PLS-DA) is a supervised dimensionality reduction method that identifies significant patterns between independent and dependent variables, suitable for analyzing high-dimensional data like protein expression and metabolomics data. Recommended software packages

include R (using mixOmics package) [72] and Python. In addition to PCA and PLS-DA, t-SNE and UMAP are popular dimensionality reduction techniques for visualizing proteins [73], t-SNE is a nonlinear dimensionality reduction technique using Student's t-distribution to define similarity between data points, optimizing the distribution of high-dimensional data points in a lower-dimensional space. However, it has limitations such as stochasticity and potential loss of original data structure. UMAP, based on manifold learning, preserves global structure better.Clustering analysis categorizes similar samples or variables into groups to reveal data structure. Hierarchical Cluster Analysis (HCA) is a common method that builds a dendrogram to organize data points, suitable for classification and identification without predetermined cluster numbers, with software packages including R (e.g., ComplexHeatmap package) and Python (e.g., seaborn package) [74]. Traditional clustering method like K-means divides data points into K clusters iteratively, ensuring each data point is closest to the center of its cluster, making it simple and suitable for large-scale datasets. Research integrates PCA and K-means to group PCA-reduced data rapidly, identifying differential spatial protein profiles in Staphylococcus aureus infection in mouse kidneys [75]. These tools provide robust visual support for in-depth analysis of complex data.

Progress in disease research through spatial proteomics

Cancer disease

Cancer remains one of the most challenging and serious health issues globally, with continually rising incidence and mortality rates imposing significant burdens on healthcare systems worldwide. Consequently, exploring the molecular mechanisms of cancer and identifying effective treatment and prevention strategies have become critical areas of medical research. Spatial proteomics, as a cutting-edge tool, provides a more comprehensive and precise perspective for unveiling disease molecular mechanisms. It also paves new pathways for personalized and precise cancer treatment strategies (Table 2). For instance, Witzke et al. used Fourier-transform infrared imaging for label-free tissue annotation and collection in bladder cancer samples. They performed LC–MS/MS proteomic analysis on cancer, inflammation, and connective tissue regions, identifying three candidate biomarkers: AHNAK2, KRT6A, and ASPN, which were validated through immunohistochemistry. AHNAK2 showed minimal expression in low-grade bladder cancer but was significantly upregulated in high-grade invasive bladder cancer, indicating its potential as a biomarker for distinguishing between bladder inflammation and cancer [55]. In 2019, a research team used laser microdissection to isolate specific cell types from the tumor and stroma of ovarian cancer patients. Mass spectrometry analysis identified upregulated expression of nicotinamide N-methyltransferase (NNMT) in the stroma. Subsequent in vivo and in vitro experiments confirmed NNMT's role in the differentiation of cancer-associated fibroblasts (CAFs). Inhibiting NNMT activity was found to reverse the CAF phenotype, suggesting a significant role for NNMT in the development and treatment of ovarian cancer [4].

In the field of non-small cell lung cancer (NSCLC) research. Zugazagoitia et al. collected and constructed

Table 2 Advancements in the Application of Spatial Proteomics in Cancer Research

Disease Type	Protein Name and location	References
Bladder cancer	AHNAK2 (high-grade bladder cancer zone)	[56]
Ovarian cancer	NNMT (fibroblast area of ovarian cancer stroma)	[4]
Non-small cell lungcancer	VISTA(tumor area), SFPQ (nuclear area of tumor patient's mesenchymal stem cells), B7-H3、GITR、PTEN、PD- (tumor area)	[76–79]
Breast cancer	α-SMA (in vitro breast cancer stroma) PD-L1、B7H3、LAG3 (primary tumor region)	[80, 81]
Glioblastoma	ALCAM, ANXA11, AltProt, IP_652563RPS14 (tumor region B) PPP1R12A、RPS14 (tumor region A) SERPINA6 (peripheral blood)	[72, 82]
Hepatocellular carcinoma	PD-L1 (tumor margin)	[83]
Head and neck squamous cell carcinoma	B2M (tumor area)	[84]
Gastric cancer	PD-L1 (tumor area)	[85]
Esophageal cell carcinoma	IDO1 (tumor microenvironment zone)	[86]
Endometrial cancer	RSS3 (stage 3 zone of endometrial cance), ASS1 (stage2zone of endometrial cance), ALDH2, ANXA1 (stage 1.zone of endometrial cance)	[6]

tumor tissue microarrays (TMAs) from 81 NSCLC patients. Using digital spatial profiling (DSP) technology to analyze cellular and molecular characteristics within the tumor microenvironment, they found that VISTA and CD127 were associated with PD-1 checkpoint blockade therapy. High expression levels of CD56 and CD4 in immune cell compartments were significantly correlated with patients' progression-free survival (PFS) and overall survival (OS). Further validation using multiplex immunofluorescence confirmed the distribution of CD56 + immune cells in the tumor stroma, supporting the DSP findings [75]. Moutafi et al., using high-throughput DSP technology, to detect proteins in tumor, leukocyte, macrophage, and immune stroma regions. They identified 71 candidate proteins, with high CD66b expression in the immune stroma region correlating with shorter overall survival (OS). Quantitative immunofluorescence validated these findings, providing evidence for CD66b as a biomarker for ICI resistance in NSCLC [84]. Another study by Yang et al., combining spatial proteomics with mesenchymal stem cell (MSC) pathway analysis, studied nuclear proteins in mesenchymal stem cells (MSCs) from NSCLC patients, discovering that SFPQ overexpression in lung cancer MSCs promotes proliferation, chemoresistance, and invasiveness. SFPQ knockdown reduced CD44v6 expression, with further in vitro and animal model experiments confirming SFPQ's role in MSC characteristics and invasiveness through CD44v6 regulation. It suggesting SFPQ as potential lung cancer therapeutic targets [87]. Subsequently, advanced NSCLC patients treated with the bispecific antibody KN046 were included in the study. Spatial analysis of proteins and RNA in tumor and stromal regions was conducted using DSP technology. Survival analysis and predictive model construction based on the results confirmed that differential expression of proteins such as B7-H3, CD45RO, GITR, CD34, PTEN, and PD-L1 in tumor and stromal regions was significantly associated with the efficacy of KN046 treatment. These findings reveal the impact of tumor heterogeneity on immune therapy response, offering new perspectives for improving treatment outcomes in advanced NSCLC patients [76].

Additionally, Vathiotis et al. using digital spatial profiling and immunofluorescence, revealed α-SMA as a novel biomarker for resistance to trastuzumab treatment in breast cancer, which is crucial for predicting patient responses to trastuzumab therapy [77]. Vathiotis et al. utilized DSP to analyze the expression of multiple proteins in tumor, leukocyte, and stromal regions of early-stage HER2-positive breast cancer tissues. The study noted that high expression of α-smooth muscle actin (α-SMA) in leukocyte and stromal regions was associated with shorter disease-free survival (DFS). QIF validation confirmed that high stromal α-SMA expression was linked to recurrence during trastuzumab treatment. Independent cohort validation using NeoALTTO cohort identified α-SMA as a novel biomarker for trastuzumab resistance in breast cancer [78]. A study also comprehensively characterized the tumor microenvironment (TME) of primary and metastatic lesions in HER2-positive breast cancer patients, finding high expression of various immune cells and markers (such as B cells, CD3, CD4) in primary tumors and exploring the potential roles of PD-L1, B7H3, LAG3, and other markers in immune checkpoint inhibitor (ICI) therapy [79]. Recently, Duhamel et al. applied MALDI-MSI combined with nanoLC-MS/MS to conduct an in-depth molecular characterization of glioblastoma multiforme (GBM) patients, classifying tumors into ABC regions. High expressions of ALCAM, ANXA11, and AltProt IP_652563 were correlated with poor prognosis, while RPS14 and PPP1R12A were associated with better outcomes. These markers were validated in an additional cohort of 50 patients, providing valuable insights into the molecular features of the disease [71]. Another study comparing local and peripheral blood proteins in glioblastoma patients found a significant decrease in SERPINA6 in the plasma. AUROC analysis confirmed SERPINA6 as a potential biomarker for glioblastoma progenitor cells [80]. Utilizing single-cell proteomics and multiplex immunofluorescence co-expression analysis in untreated and IL-2 injected regions of metastatic melanoma. The study found that the loss of membrane MHC class I expression is associated with treatment resistance. This discovery offers new insights into melanoma's response to IL-2 therapy [81]. In hepatocellular carcinoma (HCC) research, a team utilized CODEX and multiple techniques to observe interactions between tumor-associated macrophages (TAMs) and mucosal-associated invariant T (MAIT) cells. They found that MAIT cells exhibited dysfunction in the tumor microenvironment, characterized by high PD-1 expression. Subsequently, the S3-CIMA machine learning algorithm identified interactions between these two cell types. In vitro co-culture experiments and in vivo mouse model experiments confirmed that TAMs affect MAIT cell function via the PD-1/PD-L1 pathway, providing new targets for cancer immunotherapy [82]. In pre-PD-1 antibody treatment assessment of biopsy samples from head and neck squamous cell carcinoma (HNSCC), high expression of β-2 microglobulin (B2M) was observed in the tumor region. Quantitative immunofluorescence and survival analysis validated B2M as a biomarker for assessing immune therapy response and its correlation with survival rates [88]. Koh V et al. analyze tumor and stromal areas in gastric cancer patients, identifying differential proteins related to cancer expression. They found

that high expression of PD-L1 in tumor areas contributed to understanding immune escape mechanisms, informing future gastric cancer treatments [83]. Another study collected samples from esophageal squamous cell carcinoma (ESCC) patients. They established microarrays to analyze protein differences in tumor regions and the tumor microenvironment (TME). A scoring model was constructed to predict the risk of recurrence and distant metastasis, validating the role of IDO1 in assessing the risk of recurrence or metastasis in ESCC [85]. Aboulouard et al. studied endometrial cancer (EC) of different grades (I, II, III) and corresponding sentinel lymph nodes (SLNs). They identified numerous differentially expressed proteins and those overexpressed in specific tumor grades using mass spectrometry. Immunohistochemistry further validated five key proteins (PRSS3, PTX3, ASS1, ALDH2, and ANXA1) as potential biomarkers for EC and SLN grading. This research offers improved methods for individual stratification and diagnosis of EC [6].

Neurological diseases

In the field of neurological disease research, spatial proteomics has shown unique value in revealing the spatial distribution of pathogenic proteins, offering crucial insights into the molecular mechanisms of complex diseases. A study utilizing mass spectrometry-based spatial proteomics analyzed protein expression in six different brain regions of Alzheimer's disease (AD) patients, identifying candidate genes such as STXBP1, CRMP1, ACTR10, and AMPH closely associated with AD pathogenesis. These genes are involved in critical processes like neurotransmitter metabolism, neural circuit formation, viral susceptibility, and APP metabolism [86]. Further, Huang Z performed a proteomic analysis on AD patients (including the hippocampus, white matter cortex, superior temporal gyrus, middle temporal gyrus, and caudate nucleus). By comparing healthy controls, AD-resistant individuals, AD dementia groups and using external datasets. They identified five differential proteins (PA1B3, TICN3, ICAM1, IRGQ, and ALIL1) related to the pathogenesis of AD, providing new insights into the pathological mechanisms of AD [89]. Walker et al. utilized hippocampal samples from pathologically confirmed Alzheimer's disease (AD) patients. Including those with dementia symptoms (DEM) and cognitively normal but AD pathology-resistant individuals (RES), defined by Mini-Mental State Examination (MMSE) scores. Using DSP to study tissue neurofibrillary tangle (NFT) regions. They observed reduced levels of soluble Aβ and lower expression of tau protein and oxidative stress-related proteins IDH1 and PINK1 in RES individuals. This suggests that RES individuals maintain health by reducing oxidative and energy stress, offering new strategies for disease intervention [90]. Another experiment compared Alzheimer's disease neuropathologic changes (ADNC) and primary age-related tauopathy (PART) by quantifying hippocampal subregion p-tau burden using software, followed by DSP for neuronal protein analysis. In both AD and PART, synaptic health was negatively correlated with local p-tau burden. This study provides further insights into the pathological mechanisms of these two conditions [68]. In research on multiple sclerosis (MS). Spatial transcriptomics was used to analyze gene expression in the gray matter regions of fresh-frozen cortical tissue samples from multiple MS patients. Spatial proteomics further validated the transcriptomic results. Identifed proteins related to neurodegeneration such as GPR37L1, TYRO3, SIRPA, and FGFR3. Single-cell RNA sequencing data and in vivo disease models revealed the multicellular mechanisms of neurodegenerative pathways in progressive MS (PMS), providing insights for MS diagnosis and potential treatments [91].

Cardiovascular diseases

Cardiovascular diseases, encompassing heart diseases, strokes, and hypertension, are major global health issues. The study of their mechanisms and potential biomarkers has been a hot topic. The application of spatial proteomics methods has opened new pathways for cardiovascular disease research. These methods allow for high-resolution and comprehensive analysis of the proteome in cardiovascular tissues and blood, deepening our understanding of protein expressions in cardiovascular tissues and the mechanisms of cardiovascular diseases. Leng L et al. applied quantitative spatial proteomics to analyze heart tissue samples from COVID-19 patients in four regions (left atrium, left ventricle, right atrium, and right ventricle). Using laser capture microdissection, they separated myocardium and microvasculature, followed by mass spectrometry to quantify protein expression. Bioinformatics analysis revealed ECM proteins such as collagen types COL1A1, COL1A2, COL5A1, glycoproteins, and the ECM-regulating factor SERPINA1 as potential markers of cardiac injury [92]. Bhatia HS developed the DISCO-MS method for use in human cardiac tissue research, followed by regional segmentation and mass spectrometry via the DISCO-bot system. Their analysis indicated elevated levels of carbonic anhydrase 6 (CA6) and glyoxalase 1 (GLO1) in atherosclerotic plaque regions. These proteins, confirmed as biomarkers through immunohistochemistry, facilitate early detection of coronary artery diseases [93]. Recently, Mai H's team employed a laser microdissection strategy to study hard and soft tissue in arterial thrombi from patients with myocardial and cerebral infarction, confirming the role of TGF-β1 in promoting thrombosis formation.

Validation studies in human and mouse carotid thrombi demonstrated that inhibiting TGF-β1 delayed thrombus formation and enhanced blood flow recovery in a mouse model. They confirmed the role of TGF-β1 in promoting thrombosis formation and providing a potential target for thrombolytic therapy [94].

Autoimmune diseases

As understanding of autoimmune disease pathogenesis deepens, spatial proteomics offers new perspectives for comprehending the complexity of these diseases at the microscopic level. For instance, the study collected skin biopsy samples from systemic sclerosis (SSc) patients and used Imaging Mass Cytometry (IMC) to differentiate cells into two lymphatic endothelial cell (LEC) types and various pericyte subgroups. Using CODEX and Tyramide Signal Amplification (TSA), the identified cell types were further characterized and subjected to clinical correlation analysis. The presence of CD34+; αSMA+; CD31+cells co-existing with immune cells and myofibroblasts was found to be positively associated with fibrotic clinical progression, underscoring their critical roles in the vascular pathology and fibrotic remodeling in SSc [95]. Zheng F and their team conducted study on the functional regions of the kidney in lupus nephritis (LN). They finding significant immune cell infiltration in glomeruli and interstitial regions, particularly NK cells, immature B cells, and neutrophils. PPI network analysis identified proteins related to immune cell infiltration. Notably, LCN2 expression increased significantly and correlated positively with NK cell immune scores. They indicated that LCN2 plays a crucial role in LN development and has potential as a diagnostic and prognostic biomarker [96]. Subsequently, Dong J analyzed the proteomic features of four types of autoimmune glomerulonephritis, including IgA nephropathy, lupus nephritis, membranous nephropathy, and minimal change disease. Bioinformatics analysis revealed significant downregulation of FLIIs in two types of glomerulonephritis, while A2M was upregulated in MN, IgAN, and LN. Abnormal activation of the complement and coagulation cascades showed differential expression across these diseases. These findings provide new directions for the diagnosis and treatment of autoimmune glomerulonephritis [97]. Additionally, a recent study using DSP technology compared protein and transcript abundance changes about ANCA-associated glomerulonephritis (ANCA-GN) patients. Researchers first used immunofluorescence staining to identify glomerular regions with varying degrees of Bowman's capsule (BC) rupture, then analyzed the regions of interest using DSP. They found that upregulation of phosphoprotein-1 and CD44 in glomeruli with more severe Bowman's capsule rupture and positively correlated with fibrosis markers, emphasizing the SPP1-CD44 axis's role in ANCA-GN progression.This finding offers new insights into the study of ANCA-GN [98].

Other diseases

Spatial proteomics has demonstrated significant research potential in a variety of other diseases, providing a unique and profound perspective for uncovering molecular mechanisms in disease progression. In respiratory diseases, FFPE lung tissue samples from deceased COVID-19 patients were stained and sectioned, focusing on key regions such as alveolar epithelial tissue (AES), bronchial epithelial tissue (BES), and vascular tissue (VES). Differential protein analysis was conducted using LC–MS. The results revealed an upregulation of proteins closely related to the virus, host restriction, and inflammatory response, offering potential avenues for the development of diagnostic markers and therapeutic targets for COVID-19 [69]. Similarly, in skin diseases, Li used spatial proteomics to successfully create a protein map of human skin stratification. They analyzed six distinct layers (stratum corneum, granular-spinous layer, basal layer, basement membrane, superficial dermis, and deep dermis). Further cellular experiments examined the impact of the extracellular matrix (ECM) on the functionality of epidermal stem cells (EpSCs). They found that downregulation of the ECM protein TGFBI in the basement membrane could affect EpSC function via the Wnt/β-catenin signaling pathway. This suggests that TGFBI could serve as a crucial marker for regulating EpSC proliferation and potentially provide a target for the clinical treatment of skin ulcers or chronic skin diseases [99].

Conclusion and future prospects

Spatial proteomics, as a multidimensional technology, encompasses technical, spatial, and temporal dimensions. The technical dimension emphasizes the integration of various technologies, the spatial dimension provides protein information at both cellular and tissue levels, and the temporal dimension focuses on the dynamic changes of proteins. Together, these dimensions enhance our understanding of protein complexity and offer new tools for comprehensively understanding disease mechanisms and developing therapeutic strategies. This review thoroughly analyzes the progress of spatial proteomics in human disease research. Although spatial proteomics holds great potential for future research, it still faces certain challenges: (1) The sample preparation process is crucial, especially the issue of low protein extraction rates. (2) The standardization of techniques is vital as differences in sample processing methods increase experimental complexity and may lead to comparability and reproducibility issues. (3)

Effective and accurate extraction and interpretation of large-scale complex data remain key research priorities. To address these challenges, new sample preparation methods, such as hanging drop techniques, have been proposed [100]. Advanced computational tools and platforms, for example SPIAT, spaSim [101] and the Aquila data platform [102], have been developed. SPIAT, a multifunctional spatial image analysis toolkit, can process and quantify data generated by various techniques, while spaSim provides benchmarking and method validation to enhance the accuracy of data interpretation.

Future developments will focus on addressing these challenges through the integration and innovation of technologies, emphasizing the comprehensive application of spatial multi-omics. Extensive research has been conducted to combine spatial proteomics with other spatial omics data for studying various diseases. There is also increasing emphasis on developing advanced algorithms to enhance progress in data analysis, processing, and visualization. For example, the supervised machine learning algorithm MAPS has been introduced for cell type annotation in data, demonstrating superior speed and accuracy over unsupervised techniques [103]. The development of automation technologies can significantly enhance experimental efficiency and precision. These advancements not only accelerate research in the field but also deepen our understanding of disease mechanisms and facilitate the development of effective therapeutic methods. Overall, spatial proteomics is increasingly becoming an indispensable tool in modern biomedical disease research, poised to push the boundaries of disease study and treatment further, paving new paths for precision medicine and personalized therapies.

Acknowledgements
Not applicable.

Authors' contributions
CRediT author statement Yong Dai, Huihui Tao: Conceptualization. Mengyao Wu, Guoying Wang, Yali Peng: Data curation. Mengyao Wu, Huihui Tao, Tiantian Xu, Xuejia Zheng, Chunmei Wen: Formal analysis. Yong Dai, Huihui Tao: Funding acquisition. Mengyao Wu, Huihui Tao: Writing– Original Draft. Yong Dai, Writing—Review & Editing.

Funding
This work was supported by the Open Research Fund of Key Laboratory of Industrial Dust Deep Reduction and Occupational Health and Safety of Anhui Higher Education Institutes (NO. AYZJSGXLK202201002) and Research Funds of Joint Research Center for Occupational Medicine and Health of IHM (NO. OMH-2023–04).

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 29 April 2024 Accepted: 1 September 2024
Published online: 20 September 2024

References

1. Lönn P, Landegren U. Close encounters - probing proximal proteins in live or fixed cells. Trends Biochem Sci. 2017;42(7):504–15.
2. Itzhak DN, et al. Global, quantitative and dynamic mapping of protein subcellular localization. elife. 2016;5:e16950.
3. Hansen J, et al. A reference tissue atlas for the human kidney. Sci Adv. 2022;8(23):eabn4965.
4. Eckert MA, et al. Proteomics reveals NNMT as a master metabolic regulator of cancer-associated fibroblasts. Nature. 2019;569(7758):723–8.
5. Huang P, et al. Spatial proteome profiling by immunohistochemistry-based laser capture microdissection and data-independent acquisition proteomics. Anal Chim Acta. 2020;1127:140–8.
6. Aboulouard S, et al. In-depth proteomics analysis of sentinel lymph nodes from individuals with endometrial cancer. Cell Rep Med. 2021;2(6): 100318.
7. Nordmann T, et al. Single-cell spatial proteomics identifies the JAK/STAT pathway as an actionable therapeutic target in lethal cutaneous drug reactions. medRxiv. 2023.2023.11.11.23295492. https://doi.org/10.1101/2023.11.11.23295492.
8. Mund A, et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. Nat Biotechnol. 2022;40(8):1231–40.
9. Li L, et al. Spatially resolved proteomics via tissue expansion. Nat Commun. 2022;13(1):7242.
10. Lundberg E, Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol. 2019;20(5):285–302.
11. Christopher JA, et al. Subcellular proteomics Mass Spectrom Rev. 2021;1(1):32.
12. Shin JJH, et al. Spatial proteomics defines the content of trafficking vesicles captured by golgin tethers. Nat Commun. 2020;11(1):5987.
13. Hirst J, et al. Role of the AP-5 adaptor protein complex in late endosome-to-Golgi retrieval. PLoS Biol. 2018;16(1):e2004411.
14. Braccia C, et al. CFTR rescue by lumacaftor (VX-809) induces an extensive reorganization of mitochondria in the cystic fibrosis bronchial epithelium. Cells. 2022;11(12):1938.
15. Erdmann-Pham DD, Dao Duc K, Song YS. The key parameters that govern translation efficiency. Cell Syst. 2020;10(2):183–192.e6.
16. Lee JM, et al. Control of protein stability by post-translational modifications. Nat Commun. 2023;14(1):201.
17. McShane E, Selbach M. Physiological functions of intracellular protein degradation. Annu Rev Cell Dev Biol. 2022;38:241–62.
18. Unterauer EM, et al. Spatial proteomics in neurons at single-protein resolution. Cell. 2024;187(7):1785–1800.e16.
19. Mou M, et al. Application of machine learning in spatial proteomics. J Chem Inf Model. 2022;62(23):5875–95.
20. Tan WCC, et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. Cancer Commun (Lond). 2020;40(4):135–53.
21. Boguszewska K, et al. Review: immunoassays in DNA damage and instability detection. Cell Mol Life Sci. 2019;76(23):4689–704.
22. Rivest F, et al. Fully automated sequential immunofluorescence (seqIF) for hyperplex spatial proteomics. Sci Rep. 2023;13(1):16994.
23. Black S, et al. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. Nat Protoc. 2021;16(8):3802–35.
24. Goltsev Y, et al. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. Cell. 2018;174(4):968–981.e15.
25. Ogrinc Potočnik N, et al. Sequencing and identification of endogenous neuropeptides with matrix-enhanced secondary ion mass spectrometry tandem mass spectrometry. Anal Chem. 2017;89(16):8223–7.
26. Hou Y, et al. Applications of spatially resolved omics in the field of endocrine tumors. Front Endocrinol (Lausanne). 2022;13:993081.

27. Herrera JA, et al. Laser capture microdissection coupled mass spectrometry (LCM-MS) for spatially resolved analysis of formalin-fixed and stained human lung tissues. Clin Proteomics. 2020;17:24.
28. Baharlou H, et al. Mass cytometry imaging for the study of human diseases-applications and data analysis strategies. Front Immunol. 2019;10:2657.
29. Patel SS, Rodig SJ. Overview of tissue imaging methods. Methods Mol Biol. 2020;2055:455–65.
30. Farren MR, et al. Immunologic alterations in the pancreatic cancer microenvironment of patients treated with neoadjuvant chemotherapy and radiotherapy. JCI Insight. 2020;5(1):e130362.
31. Sobsey CA, et al. Targeted and untargeted proteomics approaches in biomarker development. Proteomics. 2020;20(9):e1900029.
32. Ščupáková K, et al. Cellular resolution in clinical MALDI mass spectrometry imaging: the latest advancements and current challenges. Clin Chem Lab Med. 2020;58(6):914–29.
33. Ahmed M, et al. Next-generation protein analysis in the pathology department. J Clin Pathol. 2020;73(1):1–6.
34. Buchberger AR, et al. Mass spectrometry imaging: a review of emerging advancements and future insights. Anal Chem. 2018;90(1):240–65.
35. Tobias F, Olson MT, Cologna SM. Mass spectrometry imaging of lipids: untargeted consensus spectra reveal spatial distributions in Niemann-Pick disease type C1. J Lipid Res. 2018;59(12):2446–55.
36. Zubair F. MALDI mass Spectrometry based proteomics for drug discovery & development. Drug Discov Today Technol. 2021;40:29–35.
37. Maciel Li L, et al. Desorption electrospray ionization and matrix-assisted laser desorption/ionization as imaging approaches for biological samples analysis. Anal Bioanal Chem. 2023;415(18):4125–45.
38. Planque M, et al. Spatial metabolomics principles and application to cancer research. Curr Opin Chem Biol. 2023;76:102362.
39. Guo W, et al. Laser capture microdissection for biomedical research: towards high-throughput, multi-omics, and single-cell resolution. J Genet Genomics. 2023;50(9):641–51.
40. Liotta LA, et al. Laser Capture Proteomics: spatial tissue molecular profiling from the bench to personalized medicine. Expert Rev Proteomics. 2021;18(10):845–61.
41. Petyuk VA, et al. Spatial mapping of protein abundances in the mouse brain by voxelation integrated with high-throughput liquid chromatography-mass spectrometry. Genome Res. 2007;17(3):328–36.
42. Ma M, et al. In-depth mapping of protein localizations in whole tissue by micro-scaffold assisted spatial proteomics (MASP). Nat Commun. 2022;13(1):7736.
43. Piehowski PD, et al. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-µm spatial resolution. Nat Commun. 2020;11(1):8.
44. Davis S, et al. Deep topographic proteomics of a human brain tumour. Journa. 2023;14(1):7710.
45. Seal S, et al. On clustering for cell-phenotyping in multiplex immunohistochemistry (mIHC) and multiplexed ion beam imaging (MIBI) data. BMC Res Notes. 2022;15(1):215.
46. Ji AL, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. Cell. 2020;182(2):497–514.e22.
47. Leonova T, et al. Does filter-aided sample preparation provide sufficient method linearity for quantitative plant shotgun proteomics? Front Plant Sci. 2022;13:874761.
48. Moggridge S, et al. Extending the compatibility of the SP3 paramagnetic bead processing approach for proteomics. J Proteome Res. 2018;17(4):1730–40.
49. Hughes CS, et al. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat Protoc. 2019;14(1):68–85.
50. Hogrebe A, et al. Benchmarking common quantification strategies for large-scale phosphoproteomics. Nat Commun. 2018;9(1):1045. https://doi.org/10.1038/s41467-018-03309-6.
51. Cabrita R, et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature. 2020;577(7791):561–5.
52. Radtke AJ, et al. IBEX: A versatile multiplex optical imaging approach for deep phenotyping and spatial analysis of cells in complex tissues. Proc Natl Acad Sci U S A. 2020;117(52):33455–65.
53. Pedley R, et al. Novel laser capture microdissection-proteomic analysis identifies spatially distinct extracellular matrix signatures in the core and infiltrating edge of human glioblastoma. bioRxiv. 2022.2022.09.01.506199. https://doi.org/10.1101/2022.09.01.506199.
54. Stewart HI, et al. Parallelized acquisition of orbitrap and astral analyzers enables high-throughput quantitative analysis. Anal Chem. 2023;95(42):15656–64.
55. Witzke KE, et al. Integrated fourier transform infrared imaging and proteomics for identification of a candidate histochemical biomarker in bladder cancer. Am J Pathol. 2019;189(3):619–31.
56. Kovalchik KA, et al. RawTools: rapid and dynamic interrogation of orbitrap data files for mass spectrometer system management. J Proteome Res. 2019;18(2):700–8.
57. Kockmann T, Panse C. The rawrr R package: direct access to orbitrap data and beyond. J Proteome Res. 2021;20(4):2028–34.
58. Sinitcyn P, et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. Nat Biotechnol. 2021;39(12):1563–73.
59. Cox J, et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011;10(4):1794–805.
60. Zhang F, et al. Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. Proteomics. 2020;20(17–18):e1900276.
61. Yu F, et al. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. Nat Commun. 2023;14(1):4154.
62. Neagu A-N, et al. Applications of tandem mass spectrometry (MS/MS) in protein analysis for biomedical research. Molecules. 2022;27(8):2411.
63. Yang Y, Lin L, Qiao L. Deep learning approaches for data-independent acquisition proteomics. Expert Rev Proteomics. 2021;18(12):1031–43.
64. Guan S, et al. Data Dependent-Independent Acquisition (DDIA) Proteomics. J Proteome Res. 2020;19(8):3230–7.
65. Huerta-Cepas J, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol. 2017;34(8):2115–22.
66. Kanehisa M, et al. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–d592.
67. Rudolph JD, Cox J. A network module for the perseus software for computational proteomics facilitates proteome interaction graph analysis. J Proteome Res. 2019;18(5):2052–64.
68. Walker JM, et al. Spatial proteomics of hippocampal subfield-specific pathology in Alzheimer's disease and primary age-related tauopathy. Alzheimers Dement. 2024;20(2):783–97.
69. Mao Y, et al. Deep spatial proteomics reveals region-specific features of severe COVID-19-related pulmonary injury. Cell Rep. 2024;43(2):113689.
70. Samorodnitsky S, et al. A Spatial Omnibus Test (SPOT) for spatial proteomic data. bioRxiv. 2024;40:btae425.
71. Duhamel M, et al. Spatial analysis of the glioblastoma proteome reveals specific molecular signatures and markers of survival. Nat Commun. 2022;13(1):6665.
72. Krismer E, et al. AlphaPeptStats: an open-source Python package for automated and scalable statistical analysis of mass spectrometry-based proteomics. Bioinformatics. 2023;39(8):btad461.
73. Mou M, et al. Application of machine learning in spatial proteomics. J Chem Inf Model. 2022;62(23):5875–95.
74. Weiss CJ. Visualizing protein big data using Python and Jupyter notebooks. Biochem Mol Biol Educ. 2022;50(5):431–6.
75. Sharman K, et al. Rapid multivariate analysis approach to explore differential spatial protein profiles in tissue. J Proteome Res. 2023;22(5):1394–405.
76. Yang L, et al. SFPQ promotes lung cancer malignancy via regulation of CD44 v6 expression. Front Oncol. 2022;12:862250.
77. Song X, et al. Spatial multi-omics revealed the impact of tumor ecosystem heterogeneity on immunotherapy efficacy in patients with advanced non-small cell lung cancer treated with bispecific antibody. J Immunother Cancer. 2023;11(2):e006234.
78. Vathiotis IA, et al. Alpha-smooth muscle actin expression in the stroma predicts resistance to trastuzumab in patients with early-stage HER2-positive breast cancer. Clin Cancer Res. 2021;27(22):6156–63.
79. Schlam I, et al. The tumor immune microenvironment of primary and metastatic HER2- positive breast cancers utilizing gene expression and spatial proteomic profiling. J Transl Med. 2021;19(1):480.
80. Shen L, et al. Mechanistic insight into glioma through spatially multidimensional proteomics. Sci Adv. 2024;10(7):eadk1721.

81. Pourmaleki M, et al. Tumor MHC Class I expression associates with intralesional IL2 response in melanoma. Cancer Immunol Res. 2022;10(3):303–13.
82. Ruf B, et al. Tumor-associated macrophages trigger MAIT cell dysfunction at the HCC invasive margin. Cell. 2023;186(17):3686–3705.e32.
83. Koh V, et al. Hedgehog transcriptional effector GLI mediates mTOR-Induced PD-L1 expression in gastric cancer organoids. Cancer Lett. 2021;518:59–71.
84. Zugazagoitia J, et al. Biomarkers Associated with Beneficial PD-1 checkpoint blockade in Non-Small Cell Lung Cancer (NSCLC) identified using high-plex digital spatial profiling. Clin Cancer Res. 2020;26(16):4360–8.
85. Qiu C, et al. Expression of IDO1 in tumor microenvironment significantly predicts the risk of recurrence/distant metastasis for patients with esophageal squamous cell carcinoma. Lab Invest. 2023;103(12):100263.
86. Xu J, et al. Regional protein expression in human Alzheimer's brain correlates with disease severity. Commun Biol. 2019;2:43.
87. Moutafi M, et al. Discovery of biomarkers of resistance to immune checkpoint blockade in NSCLC using high-plex digital spatial profiling. J Thorac Oncol. 2022;17(8):991–1001.
88. Gavrielatou N, et al. Digital spatial profiling links Beta-2-microglobulin expression with immune checkpoint blockade outcomes in head and neck squamous cell carcinoma. Cancer Res Commun. 2023;3(4):558–63.
89. Huang Z, et al. Brain proteomic analysis implicates actin filament processes and injury response in resilience to Alzheimer's disease. Nat Commun. 2023;14(1):2747.
90. Walker JM, et al. Differential protein expression in the hippocampi of resilient individuals identified by digital spatial profiling. Acta Neuropathol Commun. 2022;10(1):23.
91. Kaufmann M, et al. Identification of early neurodegenerative pathways in progressive multiple sclerosis. Nat Neurosci. 2022;25(7):944–55.
92. Leng L, et al. Spatial region-resolved proteome map reveals mechanism of COVID-19-associated heart injury. Cell Rep. 2022;39(11):110955.
93. Bhatia HS, et al. Spatial proteomics in three-dimensional intact specimens. Cell. 2022;185(26):5040–5058.e19.
94. Mai H, et al. Spatial proteomics analysis of soft and stiff regions in human acute arterial thrombus. Stroke. 2023;54(6):1636–44.
95. Rius Rigau A, et al. Characterization of vascular Niche in systemic sclerosis by spatial proteomics. Circ Res. 2024;134(7):875–91.
96. Zheng F, et al. Spatial proteomics landscape and immune signature analysis of renal sample of lupus nephritis based on laser-captured microsection. Inflamm Res. 2023;72(8):1603–20.
97. Dong J, et al. Global-feature of autoimmune glomerulonephritis using proteomic analysisof laser capture microdissected glomeruli. Front Immunol. 2023;14:1131164.
98. Ye L, et al. Digital spatial profiling of individual glomeruli from patients with anti neutrophil cytoplasmic autoantibody-associated glomerulonephritis. Front Immunol. 2022;13:831253.
99. Li J, et al. Spatially resolved proteomic map shows that extracellular matrix regulates epidermal growth. Nat Commun. 2022;13(1):4012.
100. Kwon Y, et al. Hanging drop sample preparation improves sensitivity of spatial proteomics. Lab Chip. 2022;22(15):2869–77.
101. Feng Y, et al. Spatial analysis with SPIAT and spaSim to characterize and simulate tissue microenvironments. Nat Commun. 2023;14(1):2697.
102. Zheng Y, et al. Aquila: a spatial omics database and analysis platform. Nucleic Acids Res. 2023;51(D1):D827–34.
103. Shaban M, et al. MAPS: Pathologist-level cell type annotation from tissue images through machine learning. bioRxiv. 2024;15(1):28.

Publisher's Note