Technical University of Munich

Spatial Characterization of Extracellular Matrix Peptides using Mass Spectrometry Imaging

Scientific work for obtaining the academic degree
Bachelor of Science (B.Sc.) Life Sciences Ernährungswissenschaft
at the TUM School of Life Sciences of the Technical University of Munich.

Supervisor	PD Dr. med. Dr. nat. med. habil. Kristina Schwamborn Institute of Pathology, TUM
Thesis advisor	Dr. rer. nat. Juliana Gonçalves Institute of Pathology, TUM
Submitted by	Theresa Mayer

Submitted on	August 26, 2025 in Munich

Contents

Abstract

The extracellular matrix (ECM) is a central component of tissue that plays both structural and functional roles and is significantly involved in tumor progression and therapy response. Collagen, the most abundant ECM protein, is particularly difficult to analyze by conventional methods due to its dense and cross-linked structure. The aim of this work was to spatially characterize collagen type I peptides using matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI). Specifically, we compared two proteolytic enzymes, trypsin and collagenase type III (COLase 3) in terms of their efficiency to release collagen type I peptides for detection across distinct tissue regions of formalin-fixed, paraffin-embedded archival samples from different tissue origins.

For this purpose, sections of a tissue microarray (TMA) containing 102 patient samples representing nine different tissue types, as well as whole-mount kidney sections were analyzed. Each section was digested with trypsin and/or COLase 3. The results showed clear differences in the enzyme-specific detection patterns. While COLase 3 proved to be particularly effective in detecting collagen I peptides in dense, collagen-rich regions, trypsin showed limited efficiency in digesting collagen in the same regions. There was little to no spatial correlation between the two enzymes, which underlines their complementary properties. In addition, digestion with PNGase F significantly increased subsequent digestion efficiency of trypsin, whereas COLase 3 pretreatment reduced tryptic peptide detection. The results show that the choice of enzyme should be determined specifically based on the study's aim and the tissue architecture. At the same time, the findings indicate that fundamental statements about an entire tissue sample using only one proteolytic enzyme should only be made with caution, due to differences in the enzymatic effectiveness observed in various tissue regions. Therefore, combined or serial digestion strategies are necessary for comprehensive ECM characterization. In addition, definition of different tissue regions, by co-registration of the data with histological annotations, can help better compare between similar regions from different samples. This work provides methodological foundations for optimized protocols for spatial analysis of ECM and highlights the importance of enzyme selection and sequence for reliable imaging of collagen in healthy and diseased tissues.

Zusammenfassung

Die extrazelluläre Matrix (ECM) ist ein zentraler Bestandteil des Gewebes, der sowohl strukturelle als auch funktionelle Aufgaben erfüllt und maßgeblich an der Tumorprogression sowie an der Wirksamkeit therapeutischer Ansätze beteiligt ist. Kollagen, das am häufigsten vorkommende ECM-Protein, ist aufgrund seiner dichten und vernetzten Struktur mit herkömmlichen Methoden nur schwer zu analysieren. Ziel dieser Arbeit war es, die räumliche Verteilung von Kollagen Typ I mittels bildgebender Massenspektrometrie (Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry Imaging, MALDI-MSI) zu untersuchen. Dabei wurden zwei proteolytische Enzyme, Trypsin und Kollagenase Typ III, hinsichtlich ihrer Fähigkeit verglichen, die Detektion von Kollagen Typ I in verschiedenen Regionen von formalinfixierten, in Paraffin eingebetteten (FFPE) Gewebeproben zu ermöglichen.

Zu diesem Zweck wurden Schnitte eines Gewebe-Microarrays (TMA) mit 102 Patientenproben aus neun verschiedenen Gewebetypen sowie Nierenschnitte analysiert, wobei jede Probe mit Trypsin und/oder Kollagenase Typ III behandelt wurde. Die Ergebnisse zeigten deutliche enzymspezifische Unterschiede in den Detektionsmustern. Während Kollagenase Typ III sich als besonders effektiv beim Nachweis von Kollagen Typ I in dichten, kollagenreichen Regionen erwies, zeigte Trypsin eine begrenzte Effizienz beim Abbau von Kollagen in genau diesen Regionen. Es bestand keine oder nur eine geringe räumliche Korrelation zwischen den beiden Enzymen, was die komplementären Eigenschaften aufzeigt. Darüber hinaus erhöhte der vorherige Verdau mit PNGase F die nachfolgende Effizienz von Trypsin, wohingegen die Vorbehandlung mit Kollagenase Typ III den Nachweis tryptischer Peptide reduzierte. Die Ergebnisse zeigten, dass die Wahl des Enzyms spezifisch an das Studienziel und die entsprechende Gewebestruktur angepasst werden sollte. Gleichzeitig deuten die Ergebnisse darauf hin, dass grundlegende Aussagen über eine gesamte Gewebeprobe hinweg unter der Verwendung von nur einem proteolytischen Enzyms aufgrund der unterschiedlichen enzymatischen Effektivität in den verschiedenen Geweberegionen nur mit Vorsicht zu treffen sind. Daher sind kombinierte bzw. serielle Enzymstrategien für eine umfassende ECM-Charakterisierung notwendig. Darüber hinaus kann das Verknüpfen der MALDI-MSI Daten mit histologischen Annotationen einen besseren Vergleich zwischen vergleichbaren Regionen aus verschiedenen Proben ermöglichen. Diese Arbeit liefert eine methodische Grundlage für die Entwicklung optimierter Protokolle zur räumlichen Analyse der ECM und unterstreicht die Bedeutung der Enzymwahl und -reihenfolge für die zuverlässige Darstellung von Kollagen in gesundem und pathologisch verändertem Gewebe.

List of Figures

List of Abbreviations

ambic	Ammonium Bicarbonate
BLAST	Basic Local Alignment Search Tool
COL1A1/ COL1A2	Collagen type I α1/α2 chain
COLase 3	Collagenase Type III
ECM	Extracellular Matrix
FFPE	Formalin-Fixed, Paraffin-Embedded
H&E	Hematoxylin and Eosin
IHC	Immunohistochemical
LC	Liquid Chromatography
m/z	Mass-to-Charge Ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization
MSI	Mass Spectrometry Imaging
PNGase F	Peptide-N-Glycosidase F
pRCC	Papillary Renal Cell Carcinoma
psi	pounds per square inch
TFA	Trifluoroacetic Acid
TIC	Total Ion Count
TMA	Tissue Microarray
TOF	Time-of-Flight
α-CHCA	α-Cyano-4-hydroxycinnamic acid

Bibliography

[1] C. Bonnans, J. Chou, and Z. Werb, 'Remodelling the extracellular matrix in development and disease', *Nat Rev Mol Cell Biol*, vol. 15, no. 12, pp. 786–801, Dec. 2014, doi: 10.1038/nrm3904.

[2] A. D. Theocharis, S. S. Skandalis, C. Gialeli, and N. K. Karamanos, 'Extracellular matrix structure', *Advanced Drug Delivery Reviews*, vol. 97, pp. 4–27, Feb. 2016, doi: 10.1016/j.addr.2015.11.001.

[3] P. Lu, K. Takai, V. M. Weaver, and Z. Werb, 'Extracellular matrix degradation and remodeling in development and disease', *Cold Spring Harb Perspect Biol*, vol. 3, no. 12, p. a005058, Dec. 2011, doi: 10.1101/cshperspect.a005058.

[4] T. R. Cox and J. T. Erler, 'Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer', *Dis Model Mech*, vol. 4, no. 2, pp. 165–178, Mar. 2011, doi: 10.1242/dmm.004077.

[5] Y. Jiang, H. Zhang, J. Wang, Y. Liu, T. Luo, and H. Hua, 'Targeting extracellular matrix stiffness and mechanotransducers to improve cancer therapy', *Journal of Hematology & Oncology*, vol. 15, no. 1, p. 34, Mar. 2022, doi: 10.1186/s13045-022-01252-0.

[6] C. Frantz, K. M. Stewart, and V. M. Weaver, 'The extracellular matrix at a glance', *J Cell Sci*, vol. 123, no. 24, pp. 4195–4200, Dec. 2010, doi: 10.1242/jcs.023820.

[7] S. Xu *et al.*, 'The role of collagen in cancer: from bench to bedside', *J Transl Med*, vol. 17, no. 1, p. 309, Sep. 2019, doi: 10.1186/s12967-019-2058-1.

[8] S. C. Zambrzycki *et al.*, 'Profiling of collagen and extracellular matrix deposition from cell culture using in vitro ExtraCellular matrix mass spectrometry imaging (ivECM-MSI)', *Matrix Biology Plus*, vol. 24, p. 100161, Dec. 2024, doi: 10.1016/j.mbplus.2024.100161.

[9] J. K. Macdonald, A. S. Mehta, R. R. Drake, and P. M. Angel, 'Molecular analysis of the extracellular microenvironment: from form to function', *FEBS Letters*, vol. 598, no. 6, pp. 602–620, 2024, doi: 10.1002/1873-3468.14852.

[10] A. Naba, 'Ten Years of Extracellular Matrix Proteomics: Accomplishments, Challenges, and Future Perspectives', *Mol Cell Proteomics*, vol. 22, no. 4, p. 100528, Mar. 2023, doi: 10.1016/j.mcpro.2023.100528.

[11] D. Li, J. Yi, G. Han, and L. Qiao, 'MALDI-TOF Mass Spectrometry in Clinical Analysis and Research', *ACS Meas. Sci. Au*, vol. 2, no. 5, pp. 385–404, Oct. 2022, doi: 10.1021/acsmeasuresciau.2c00019.

[12] P. M. Angel and R. M. Caprioli, 'Matrix-Assisted Laser Desorption Ionization Imaging Mass Spectrometry: In Situ Molecular Mapping', *Biochemistry*, vol. 52, no. 22, p. 10.1021/bi301519p, Jun. 2013, doi: 10.1021/bi301519p.

[13] J. G. Swales, G. Hamm, M. R. Clench, and R. J. A. Goodwin, 'Mass spectrometry imaging and its application in pharmaceutical research and development: A concise review', *International Journal of Mass Spectrometry*, vol. 437, pp. 99–112, Mar. 2019, doi: 10.1016/j.ijms.2018.02.007.

[14] J. K. Macdonald *et al.*, 'Optimization of Collagenase Proteomics for Improved Mass Spectrometry Imaging Peptide Identification', *Anal. Chem.*, vol. 97, no. 14, pp. 7672–7681, Apr. 2025, doi: 10.1021/acs.analchem.4c04818.

[15] C. L. Clift, R. R. Drake, A. Mehta, and P. M. Angel, 'Multiplexed imaging mass spectrometry of the extracellular matrix using serial enzyme digests from formalin-fixed paraffin-embedded tissue sections', *Anal Bioanal Chem*, vol. 413, no. 10, pp. 2709–2719, Apr. 2021, doi: 10.1007/s00216-020-03047-z.

[16] P. M. Angel *et al.*, 'Mapping Extracellular Matrix Proteins in Formalin-Fixed, Paraffin-embedded Tissues by MALDI Imaging Mass Spectrometry', *J Proteome Res*, vol. 17, no. 1, pp. 635–646, Jan. 2018, doi: 10.1021/acs.jproteome.7b00713.

[17] M.-F. Hsueh, A. Khabut, S. Kjellström, P. Önnerfjord, and V. B. Kraus, 'Elucidating the Molecular Composition of Cartilage by Proteomics', *J. Proteome Res.*, vol. 15, no. 2, pp. 374–388, Feb. 2016, doi: 10.1021/acs.jproteome.5b00946.

[18] R. R. Drake, T. W. Powers, K. Norris-Caneda, A. S. Mehta, and P. M. Angel, 'In Situ Imaging of N-Glycans by MALDI Imaging Mass Spectrometry of Fresh or Formalin-Fixed Paraffin-Embedded Tissue', *Current Protocols in Protein Science*, vol. 94, no. 1, p. e68, 2018, doi: 10.1002/cpps.68.

[19] A. Ly *et al.*, 'Site-to-Site Reproducibility and Spatial Resolution in MALDI–MSI of Peptides from Formalin-Fixed Paraffin-Embedded Samples', *PROTEOMICS – Clinical Applications*, vol. 13, no. 1, p. 1800029, 2019, doi: 10.1002/prca.201800029.

[20] C. L. Clift, A. Mehta, R. R. Drake, and P. M. Angel, 'Multiplexed Imaging Mass Spectrometry of Histological Staining, N-Glycan and Extracellular Matrix from One Tissue Section: A Tool for Fibrosis Research', in *Multiplexed Imaging: Methods and Protocols*, E. Zamir, Ed., New York, NY: Springer US, 2021, pp. 313–329. doi: 10.1007/978-1-0716-1593-5_20.

[21] C. T. McDowell, X. Lu, A. S. Mehta, P. M. Angel, and R. R. Drake, 'Applications and continued evolution of glycan imaging mass spectrometry', *Mass Spectrometry Reviews*, vol. 42, no. 2, pp. 674–705, 2023, doi: 10.1002/mas.21725.

[22] L. A. McDonnell, A. Walch, M. Stoeckli, and G. L. Corthals, 'MSiMass List: A Public Database of Identifications for Protein MALDI MS Imaging', *J. Proteome Res.*, vol. 13, no. 2, pp. 1138–1142, Feb. 2014, doi: 10.1021/pr400620y.

[23] 'BLAST: Basic Local Alignment Search Tool'. Accessed: Aug. 22, 2025. [Online]. Available: https://blast.ncbi.nlm.nih.gov/Blast.cgi

[24] R. Bischoff, 'Enzymatic liberation of myogenic cells from adult rat muscle', *Anat Rec*, vol. 180, no. 4, pp. 645–661, Dec. 1974, doi: 10.1002/ar.1091800410.
[25] T. S. Høiem *et al.*, 'An optimized MALDI MSI protocol for spatial detection of tryptic peptides in fresh frozen prostate tissue', *PROTEOMICS*, vol. 22, no. 10, p. 2100223, 2022, doi: 10.1002/pmic.202100223.
[26] M. Dilillo *et al.*, 'Ultra-High Mass Resolution MALDI Imaging Mass Spectrometry of Proteins and Metabolites in a Mouse Model of Glioblastoma', *Sci Rep*, vol. 7, no. 1, p. 603, Apr. 2017, doi: 10.1038/s41598-017-00703-w.
[27] P. Goettig, 'Effects of Glycosylation on the Enzymatic Activity and Mechanisms of Proteases', *International Journal of Molecular Sciences*, vol. 17, no. 12, p. 1969, Dec. 2016, doi: 10.3390/ijms17121969.
[28] J. Y. Lee *et al.*, 'Targeted Mass Spectrometric Approach for Biomarker Discovery and Validation with Nonglycosylated Tryptic Peptides from N-linked Glycoproteins in Human Plasma', *Mol Cell Proteomics*, vol. 10, no. 12, p. M111.009290, Dec. 2011, doi: 10.1074/mcp.M111.009290.
[29] C. M. DeRosa, S. D. Weaver, C.-W. Wang, N. Schuster-Little, and R. J. Whelan, 'Simultaneous N-Deglycosylation and Digestion of Complex Samples on S-Traps Enables Efficient Glycosite Hypothesis Generation', *ACS Omega*, vol. 8, no. 4, pp. 4410–4418, Jan. 2023, doi: 10.1021/acsomega.2c08071.
[30] C. Miersch, K. Stange, and M. Röntgen, 'Effects of trypsinization and of a combined trypsin, collagenase, and DNase digestion on liberation and in vitro function of satellite cells isolated from juvenile porcine muscles', *In Vitro Cell.Dev.Biol.-Animal*, vol. 54, no. 6, pp. 406–412, Jun. 2018, doi: 10.1007/s11626-018-0263-5.
[31] K. Linka *et al.*, 'Unraveling the Local Relation Between Tissue Composition and Human Brain Mechanics Through Machine Learning', *Front. Bioeng. Biotechnol.*, vol. 9, Aug. 2021, doi: 10.3389/fbioe.2021.704738.
[32] J. M. Barnes, L. Przybyla, and V. M. Weaver, 'Tissue mechanics regulate brain development, homeostasis and disease', *Journal of Cell Science*, vol. 130, no. 1, pp. 71–82, Jan. 2017, doi: 10.1242/jcs.191742.
[33] K. Song, Z. Yu, X. Zu, G. Li, Z. Hu, and Y. Xue, 'Collagen Remodeling along Cancer Progression Providing a Novel Opportunity for Cancer Diagnosis and Treatment', *International Journal of Molecular Sciences*, vol. 23, no. 18, p. 10509, Jan. 2022, doi: 10.3390/ijms231810509.
[34] E. Henke, R. Nandigama, and S. Ergün, 'Extracellular Matrix in the Tumor Microenvironment and Its Impact on Cancer Therapy', *Front. Mol. Biosci.*, vol. 6, Jan. 2020, doi: 10.3389/fmolb.2019.00160.

List of Tables

1. Introduction

The extracellular matrix (ECM) is a complex, three-dimensional network that not only provides a physical scaffold but also regulates a wide range of cellular processes, including growth, migration, differentiation, survival, homeostasis, and morphogenesis [1]. Primarily composed of fibrillar structural proteins such as collagens, elastin, fibronectin, laminins, various glycoproteins, proteoglycans, and glycosaminoglycans, the ECM undergoes permanent remodeling through tightly regulated processes of synthesis, degradation, and post-translational modifications to maintain tissue homeostasis [2], [3]. However, in conditions such as fibrosis, tumor growth, and metastasis, this finely balanced system becomes disrupted, resulting in pathological ECM remodeling with extensive alterations in its composition and structure [4]. Collagen, the most abundant ECM protein, plays a central role in this process, as its increased deposition and modification are not passive byproducts of tumor growth but actively contribute to altered cell behavior, metastasis, and therapy response [4], [5], [6]. These effects are driven by multiple mechanisms controlled by tumor cells: Mutated genes like p53 and altered signaling pathways such as transforming growth factor β (TGF-β)/Smad control the expression of collagen-forming enzymes and directly affect the amount and composition of collagen in the ECM [7]. In addition, post-translational modifications such as hydroxylation, glycosylation, and cross-linking by enzymes such as prolyl hydroxylases and lysyl oxidases stabilize and densify the collagen network [7]. As a result, ECM density and stiffness increase, which activates mechanosensitive pathways, leading to invasive cell behavior, epithelial-mesenchymal transition, and enhanced cell motility [5], [6]. In addition, tumor cells interact with collagen via receptors such as integrins and discoidin domain receptors 1 and 2 (DDR1/2) and secrete matrix metalloproteinases to degrade and remodel the matrix, which further promotes migration and progression [7].

Despite the fundamental role of collagen in tumor progression, its analysis within intact tissue remains a major challenge. A key reason for this is the constant remodeling of the ECM described above, which requires analytical methods to not only detect the spatial localization, but also be able to capture its dynamic biochemical changes over time [8]. Methods such as transcriptome analyses are only of limited use for this purpose, as they provide information about gene expression but do not reflect the actual molecular state, structure, or degree of modification of ECM proteins [9]. A further obstacle lies in the special biochemical nature of central ECM components such as fibrillar collagens and elastin. They are highly cross-linked, extensively modified, and form higher-order molecular structures, such as collagen triple helices, which makes them poorly soluble and

limits their accessibility for standard biochemical methods like SDS-PAGE, immunoprecipitation, pull-down assays, or mass spectrometry [10].

To overcome these limitations and preserve spatial context while analyzing molecular composition, matrix-assisted laser desorption/ionization mass spectrometry imaging (MALDI-MSI) offers a promising approach. In this soft ionization technique, the analyte is applied together with an organic matrix to a conductive target plate and irradiated with a pulsed laser. The matrix absorbs the laser energy and facilitates the analytes desorption and ionization through energy transfer and secondary reactions. The ions generated in this way, predominantly with a single charge, are then accelerated by an electric field into a time-of-flight (TOF) mass spectrometer. Separation is based on the mass-to-charge ratio (*m/z*), with lighter ions reaching the detector faster than heavier ones [11]. This technique provides several advantages: it allows the spatial mapping of molecular analytes including proteins/peptides, lipids, and metabolites directly within intact biological tissues without the need for target-specific reagents such as antibodies, as required in conventional methods like immunohistochemistry [12], [13]. In addition, its simple operation, high analytical throughput, and high tolerance against contaminants have established the method as an important tool in modern clinical diagnostics and research, where MALDI-TOF is widely used for a broad range of applications [11]. These include the identification of pathogens, the diagnosis of various diseases through biomarker analysis, nucleic acids analysis such as SNP genotyping and DNA methylation, and the analysis of small molecules including metabolites, amino acids, and lipids [11].

Several approaches have already been developed to improve access to ECM components and thereby enhance their detection in mass spectrometry-based analyses [14], [15], [16], [17]. Early strategies focused on extraction methods. For example Hsueh et al. showed that guanidinium hydrochloride extraction yielded mainly non-collagenous proteins, while large parts of the collagen network remained in the residue [17]. Improved approaches, including surfactant-assisted extraction, decellularization, removal of interfering polysaccharides, and in situ enzymatic digestion significantly increase the yield of collagen-associated peptides and enable a more comprehensive characterization of ECM composition and organization [17]. However, both standard and optimized extraction methods destroy the tissue architecture, resulting in the loss of spatial information on ECM components. To preserve the spatial context, Clift et al. introduced a multiplexed MALDI-MSI approach in which formalin-fixed, paraffin-embedded (FFPE) sections underwent sequential enzymatic digestions with chondroitinase ABC, peptide-N-glycosidase F (PNGase F), elastase, and COLase 3 [15]. This stepwise enzyme application enabled the in situ detection of chondroitin sulfate glycosaminoglycans, *N*-glycans, elas-

tin, collagens, and other ECM peptides within the same tissue section. The authors proposed that by gradually removing specific ECM components, the subsequent enzymes gained better access to previously inaccessible structures, which increased the peptide yield while preserving the spatial distribution of the molecules for MALDI imaging [15]. Although this strategy significantly improved ECM coverage, its multi-step and time-consuming implementation limits its practicability for routine or large-scale studies.

Nevertheless, the study clearly demonstrates that enzyme choice and sequence are critical determinants of detection efficiency, as enzymatic digestion represents a central step in standard MALDI-MSI workflows. Trypsin has long been the enzyme of choice for proteomic applications due to its wide applicability, as it preferentially cleaves at the C-terminal side of arginine and lysine [14]. However, it shows reduced activity when these are followed by proline, which is a common sequence constellation in G-P-X-rich triple-helical regions of collagen [2], [14]. In contrast, collagenase type lll (COLase 3), a bacterial matrix metalloproteinase, allows more targeted access to ECM components, as it cleaves these native structures and effectively digests both soluble and insoluble collagens [14]. Building on this concept, the present work systematically compares trypsin and COLase 3 to investigate how both enzymes influence the detection of the spatial distribution of ECM peptides within the same tissue. The analysis focuses on type I collagen, a highly abundant and widespread ECM component, consisting of α1(I) and α2(I) chains encoded by the COL1A1 and COL1A2 genes [2], [9]. By mapping these enzyme-specific patterns, the aim of this work is to determine the strengths of each enzyme in accessing different morphological regions in patient derived FFPE tissues. The results provide insights into in situ protease activity, serving as the basis for the development of progressively optimized MALDI MSI workflows that enable a comprehensive yet practical characterization of the ECM.

For this purpose, a tissue microarray (TMA) with 102 patient samples from nine different tissue types (pancreas, liver, breast, kidney, tonsil, brain, stomach, lung, leiomyoma) and whole mount kidney sections were analyzed.

2. Materials and Methods

2.1. Materials

Chemicals and Reagents

Indium-tin-oxide (ITO) slides, 0.1 % Poly-L-lysine (Sigma), xylene (Sigma), isopropanol (Fluka), ethanol (Sigma), methanol (Roth), citraconic acid anhydride (Sigma), hydrochloric acid (Thermo Fisher Scientific), Trizma base (Carl Roth), calcium chloride (Sigma), ammonium bicarbonate (Sigma), glycerol (Sigma), PNGase F Prime (N-Zyme Scientifics, Lot No. NZ-2024-0233), collagenase type III (Worthington Biochemical), trypsin (Promega, Product No. V511A, Lot No. 0000560791), potassium sulfate (Carl Roth), α-cyano-hydroxycinnamic acid (Sigma), acetonitrile (Sigma), trifluoracetic acid (Merck), ammonium phosphate monobasic (Sigma), peptide calibration standard (Bruker), red phosphorus (Sigma), COL1A1 antibody (E8F4L clone; rabbit anti-human/mouse; #72026, Cell Signaling Technology)

Equipment

Oven (VWR, DRY-Line 53), desiccator (Duran Group), water bath (GFL), decloaking chamber (NXGen by BioCare), TM Sprayer (HTX Technologies), syringe pump (New Era Pump Systems), liquid chromatography (LC) pump (KNAUER), digestion chamber (Clip & Close, Emsa 550mL), benchtop sonicator (Bandelin), slide scanner (Reflecta), rapifleX MALDI Tissue-typer mass spectrometer (Bruker Daltonik)

2.2. Methods

The experimental procedures followed established protocols for the analysis of *N*-glycans described by Drake et al. [18], trypsin-based peptide imaging according to Ly et al. [19] and collagenase-based peptide imaging by Clift et al [20].

Sample cohort

Two sections of FFPE TMAs and three sections of wholemount FFPE tissue of human papillary renal cell carcinoma (pRCC) were used. The sections underwent different enzymatic treatments, summarized in **Figure 1**.

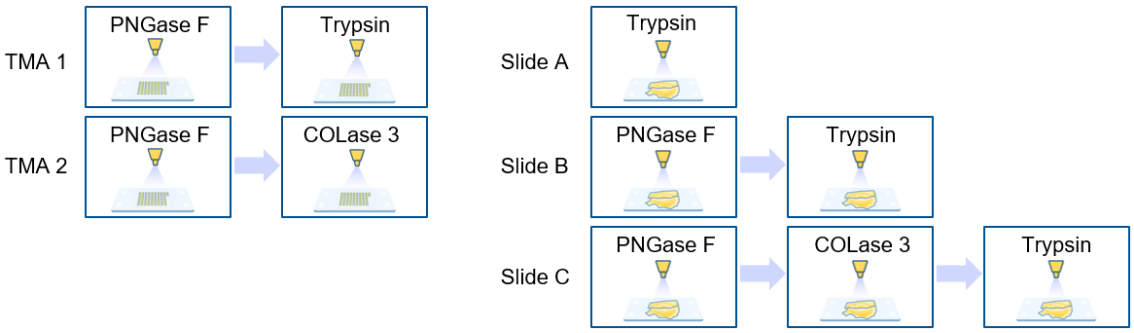


Figure 1. Overview of enzymatic pretreatments applied to two sections of the TMA (1-2) and three consecutive cuts of kidney sections (A-C). The TMAs underwent PNGase F digestion followed by either trypsin (TMA 1) or COLase 3 digestion (TMA 2). As for the kidney sections, slide A was treated with trypsin only, slide B with PNGase F followed by trypsin, and slide C with PNGase F followed by COLase 3 and subsequently trypsin. The image was created with BioRender.

The TMA consisted of 102 cores with 1 mm diameter, including both tumor and healthy tissue samples from nine different tissue types: pancreas (n = 10), liver (n = 12), breast (n = 10), kidney (n = 9), tonsil (n = 11), brain (n = 10), stomach (n = 17), lung (n = 10), and leiomyoma (n = 13). To minimize potential measurement bias, the tissue samples were randomly distributed across the block, as illustrated in **Figure 2**.

The tissue samples were obtained from the archive of the Institute of Pathology of the Technical University of Munich. The TMA was constructed by the Comparative Experimental Pathology core facility following a predefined layout.

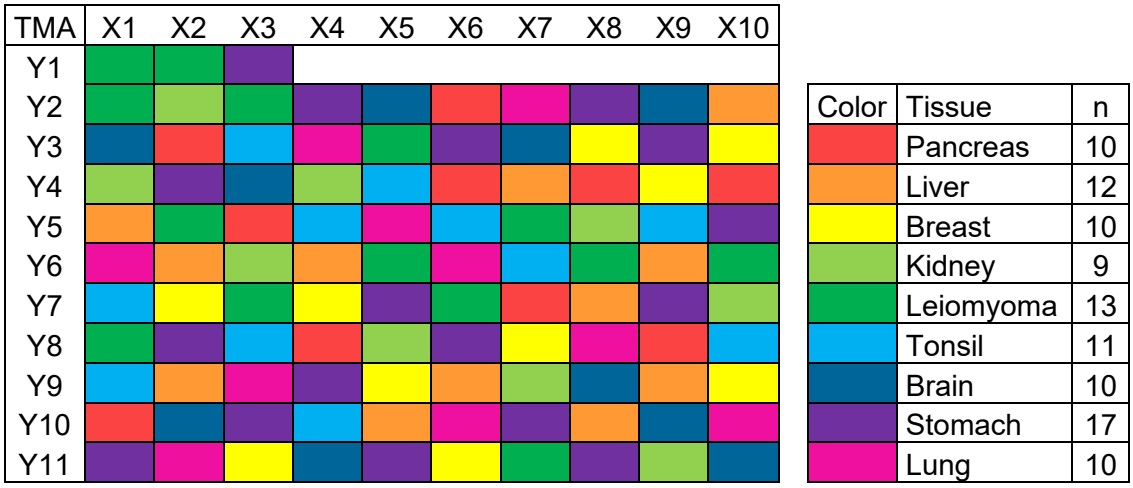


Figure 2. Layout of the TMA with color-coded tissue types

Sample preparation

4 µm sections, cut from the TMA or the kidney tissue blocks, were mounted onto indium tin oxide-coated slides that had been treated with a 0.1% Poly-L-lysine aqueous solution. To enable proper adhesion of the tissue sections, the slides were incubated at 37°C for at least three hours or overnight. Afterward, they were kept at room temperature in a dry environment until further processing.

PNGase F treatment

For deparaffinization, the slides were placed in a slide rack and heated in an oven at 80°C for 15 minutes to melt the paraffin and to further improve tissue adhesion to the glass. This was followed by two xylene baths (3 minutes each) to remove the liquefied paraffin. The slides were then rehydrated through a graded ethanol series: 100%, 95%, and 70% (1 minute each) and two ultra-pure water baths (3 minutes each). Finally, the slides were dried in a desiccator for 10 minutes. To revert cross-linking effects induced by the fixation protocol and expose specific epitopes, antigen retrieval was performed [21]. For this purpose, a citraconic buffer was prepared by pouring 50 mL ultra-pure water into a 100 mL bottle, adding 50 µL of citraconic acid anhydride and 4 µL of 6 M hydrochloric acid (HCl). The bottle was filled up with ultra-pure water to a final volume of 100 mL. With a pH strip the pH was checked to be approximately 3.0 ± 0.5. The buffer solution was transferred into a plastic coplin jar, so that the slides were fully covered by the fluid. The coplin jar was sealed and placed in the preheated water bath at 85°C for 30 minutes. The slides were cooled down to room temperature by exchanging the water with room temperature ultra-pure water three times and placed in a desiccator for approximately 10 minutes until completely dry. To prepare the enzyme, 1000 µL ultra-pure water was added into an empty Eppendorf tube. Half of the solution was pipetted into a tube with 100 µg of PNGase F Prime to dissolve the enzyme. After solubilizing the enzyme, the solution was resuspended in the remaining water in the Eppendorf tube and vortexed for a few seconds. The enzyme solution was sprayed onto the slides with a TM-Sprayer, with the assistance of a syringe pump, employing the following settings: 15 layers, 45°C, 25 µL/min flow rate, 1200 mm/min velocity, 3 mm track spacing, crisscross pattern, and 10 pounds per square inch (psi) nitrogen flow. The slides were incubated in a digestion chamber containing saturated potassium sulfate (K_2SO_4) solution, prepared by dissolving 75 g of K_2SO_4 in 500 mL of de-ionized water, in a preheated oven for 2 h at 38.5 ± 1.5°C.

Following digestion, the slides were labeled with four reference marks using a white marker and were scanned at a resolution of 1200 dots per inch (dpi). This step is necessary for positional alignment between fleXControl (4.2) and fleXImaging (6.0) when defining the regions for data acquisition. The matrix solution was prepared by adding 200 mg α-Cyano-4-hydroxycinnamic acid (α-CHCA) to a falcon tube and adding 14 mL ultra-pure water, 14 mL acetonitrile, and 28 µL trifluoroacetic acid (TFA). The matrix was vortexed and sonicated using a benchtop ultrasound bath until it was completely dissolved. Every matrix used in this work was prepared similarly to a final concentration of 7 mg/mL of α-CHCA in 50% acetonitrile/ H_2O/ 0.1% TFA. The matrix was applied with

the TM-Sprayer, with assistance of the LC pump, using the following parameters: 10 layers, 79°C, 100 µL/min flow rate, 1300 mm/min nozzle velocity, 2.5 mm track spacing, crisscross pattern, and 10 psi nitrogen flow. Before the sample measurement, 1 µL calibration solution (peptide calibration standard and/ or red phosphorous suspension) was applied to a tissue-free region where matrix had been applied on the slide.

Collagenase type III **(COLase 3) treatment**

After PNGase F treatment, the matrix was removed from TMA 2 and slide C with short washes of 5 seconds in methanol and ethanol. The antigen retrieval was performed in a 10 mM Tris buffer solution which was prepared by filling 900 mL of deionized water into a 1 L bottle, adding 1.21 g Trizma base, and bringing the volume of water up to 975 mL. The pH was adjusted to 9 with either 1 M NaOH or 1 M HCl. The slides were placed in a slide carrier and submerged in Tris buffer solution. The tissue samples were heated in a water bath at 95°C for 30 minutes. To cool down the slides to room temperature the buffer was slowly replaced by room temperature ultra-pure water. The slides were dried in a desiccator for 10 minutes until completely dry. To prepare the buffer for the COLase 3, 197.6 mg of ammonium bicarbonate and 11.10 mg of calcium chloride were dissolved in 100 mL of ultra-pure water and again adjusted to a pH of 7.25. COLase 3 solution was prepared by adding 1 mL of the ammonium bicarbonate (ambic) buffer to a 0.2 mg vial of collagenase to a final concentration of 0.2 µg/µL. The COLase 3 was applied with the TM-Sprayer and a syringe pump with the following parameters: 15 layers, 45°C, 25 µL/min flow rate, 1200 mm/min velocity, crisscross pattern, and 10 psi nitrogen flow. The slides were incubated in a digestion chamber with saturated K_2SO_4 solution for 5 h at 37.5°C. The α-CHCA matrix was sprayed using the TM-Sprayer and the LC pump with following parameters: 14 layers, 79°C, 70 µL/min flow rate, 1300 mm/min nozzle velocity, 2.5 mm track spacing, crisscross pattern, and 10 psi nitrogen flow. After the matrix application, the slides were briefly dipped in cold 5 mM ammonium phosphate monobasic solution, which was prepared by first producing a 500 mM stock solution by dissolving 2.88 g ammonium phosphate monobasic in 50 mL of ultra-pure water. The 5 mM working solution was obtained from dilution of 500 µL of the stock solution into 49.5 mL of ultra-pure water in a conical tube and mixed well prior to use. After drying the slides vertically in the drying chamber, the calibration spot was applied, and the slides were measured in the mass spectrometer.

Trypsin treatment

A representative workflow for the enzyme combination PNGase F and trypsin, as used in TMA 1 and slide B, is shown in **Figure 3**.

FFPE tissue slides were deparaffinized by heating them in an oven at 80°C for 15 minutes, followed by immersion in two consecutive xylene baths (5 minutes each). The slides were then rehydrated through a graded series of alcohol baths: 100% isopropanol, followed by ethanol at 100%, 96%, 70%, and 50% (5 minutes each). Finally, they were rinsed in ultra-pure water for 5 seconds and dried in the desiccator for approximately 5 minutes. Dewaxing and rehydration steps were skipped for the samples TMA 1, slide B and C, which have been previously dewaxed, rehydrated, and digested with other enzymes. For these samples, the matrix was removed by short consecutive washing steps (approximately 5 seconds) in methanol and ethanol. All samples were subjected to antigen retrieval performed in a decloaking chamber using ultra-pure water at 110°C for 20 minutes. The slides were then cooled down to room temperature by exchanging the water three times and placed in a desiccator for approximately 10 minutes until completely dry. To prepare the trypsin solution, 800 µL of a 0.0125% glycerol solution was added to 200 µL of a frozen 100 mM ammonium bicarbonate in deionized water stock solution stored in a 1.5 mL Eppendorf tube. This resulted in a buffer solution with a final concentration of 20 mM ambic and 0.01% glycerol. 800 µL of this buffer solution was added to one vial of sequencing grade modified trypsin (20 µg), mixed well on a vortexer, and stored on ice until it was used. The enzyme was sprayed onto the tissue using a TM-Sprayer and a syringe pump with following settings: 16 layers, 30°C, 15 µL/min flow rate, 750 mm/min velocity, 2 mm track spacing, crisscross pattern, and 10 psi nitrogen flow. The slides were placed in a digestion chamber with saturated K_2SO_4 solution to guarantee stable humidity during the digestion process for two hours at 50°C. If trypsin was the first enzyme applied, four reference marks were added on off tissue regions for positional teaching and the slide was scanned. The matrix was sprayed using the TM-Sprayer and a LC pump with the following settings: 4 layers, 75°C, 120 µL/min flow rate, 1200 mm/min nozzle velocity, 3 mm track spacing, heel-to-heel pattern, and 10 psi nitrogen flow. After the matrix application, 1 µL of the calibration solution was applied, and the slides were measured in the mass spectrometer.

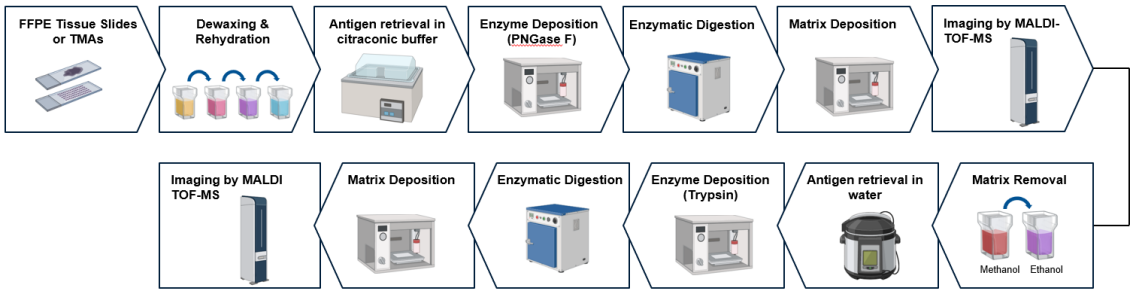


Figure 3. Representative workflow for MALDI-MSI of FFPE tissue using sequential PNGase F and trypsin digestion. This image was created with BioRender.

Measurement setup

The samples were analyzed by the mass spectrometer in positive-ion reflector mode, collecting 500 laser shots per pixel at a frequency of 10 000 Hz. The laser was defined to raster in a 50 x 50 µm range at each position and set to M5 small conformation. The sample rate was set to 1.25 Giga samples per second (GS/s). The mass spectra acquisition were conducted with a laser power between 85-100%, adjusted so that the detection signal of the test measurement reached intensity values in the range of 10^4. The mass range was set to either 600-3200 for peptide or 900-3200 for *N*-Glycan analysis.

The detector check and calibration were performed before every measurement.

Histological validation and tissue annotations

Histological staining was performed for histomorphological evaluation of the tissues. Hematoxylin and eosin (H&E), Van Gieson, and immunohistochemical (IHC) staining with an antibody against COL1A1 were performed for the kidney sections, and H&E and IHC staining for the TMA sections. In addition, an experienced pathologist made histological annotations of the relevant tissue regions using QuPath software (version 0.6.0).

Data Analysis

The MALDI MSI data were analyzed by SCiLS Lab software (version 2025b Pro). Prior to data analysis, the data were prepared by applying baseline correction and total ion count (TIC) normalization.

For peak detection, feature lists were generated using the *Feature Finding* tool, selecting all regions except the matrix, running the *work on individual spectra* mode, limiting the maximum peaks per spectrum to 100 and using every 16th spectrum. With the *sliding window* function, the threshold was set to 10% of the maximum intensity to exclude the background. The resulting feature list was then filtered to remove non-monoisotopic and matrix-related peaks. These lists were used both for segmentation and matching the peaks with the reference lists. Segmentation was performed to demonstrate reproducibility, again selecting all regions except the matrix and using *work on individual spectra* mode. It was based on five classes and was generated using k-means clustering, with correlation distance as the metric and no denoising applied. In addition, the *find values co-localized to feature* function was used to compare and semi-quantitatively represent the spatial distribution both within individual slides and between different slides, always considering the histological annotations. This analysis was also conducted in *work on*

individual spectra mode. For comparative analysis across different datasets, the SCiLS Ion Image Mapper was employed to identify spatial colocalizations.

To visualize signal intensities, SCiLS Lab was also used to create intensity boxplots for individual peaks in different regions. Since the software does not allow to show signal intensities of several *m/z* values in different regions at the same time, the average intensities for the respective tissue regions were calculated with SCiLS Lab software in the feature table using *add intensity column(s)* in the spot intensity statistics with *peak maximum* interval processing mode. Subsequently, the maximum intensity of the highest collagen-associated peak in the spectrum was determined across all histological annotations, also using the *peak maximum* interval processing mode. The intensities were then baseline-corrected by subtracting the value representing the background noise, followed by adjustment to the maximum signal in the dataset by dividing through the maximum intensity value in Excel (**Table 1**). This approach allowed the visualization of intensities from different datasets within the same bar chart.

Table 1. Adjustments applied to visualize signal intensities from trypsin and COLase 3 digests in the same diagram

	Kidney		TMA	
	trypsin	COLase 3	trypsin	COLase 3
m/z	1105.677	843.898	1105.570	843.553
maximum intensity	31.21	16.89	19.28	35.27
baseline subtraction value	-0.23	-0.16	-0.13	-0.20

Two reference sources were used to identify collagen-associated peptides: the MSI mass list from the International Mass Spectrometry Imaging Society (IMSIS) [22] for the tryptic collagen peptides, and a comprehensive list of COLase 3 derived peptides by Macdonald, J. K. et al [14].

For inclusion of individual *m/z* values in the comparative analysis between the tryptic peptides and COLase 3 peptides, two selection criteria were defined:

1. Sequence overlap between enzyme digests: only *m/z* values whose corresponding peptide sequences had an overlap of at least 5 amino acids between tryptic- and COLase 3 digested peptides were considered. This criterion served as an indication that the peptides originated most likely from the same protein of origin and the same region in the protein structure, implying that their spatial distribution was comparable.

2. Spatial correlation: to ensure data comparability, it was assessed whether different collagen-derived peptides resulting from the same digestion showed similar spatial distribution within the tissue. Pearson correlation coefficients were calculated between each pair of selected *m/z* values. Signals with a consistently low correlation ($r < 0.5$) to all others were excluded, as structural proteins such as COL1A1 are expected to have a consistent distribution. If a signal does not match this pattern, it likely indicates a mismatch.

The correlation scores that informed the selection process are provided in **Appendix A1**.

It should be noted that individual peptide sequences, and thus also the corresponding *m/z* values, cannot be exclusively assigned to a single protein. The presence of a *m/z* peak alone therefore does not constitute definitive evidence for the occurrence of a specific protein. To illustrate this uncertainty, example assignments are provided in **Appendix A2**: for the value 1105.58 *m/z* Basic Local Alignment Search Tool (BLAST) searches were performed (database: UniProtKB/Swiss-Prot; organism: Homo sapiens [taxid:9606]; algorithm: blastp) to list possible proteins containing the same peptide sequence [23]. For 1139 *m/z* (1139.5948 *m/z* and 1139.6518 *m/z*), the COLase 3 peptides list created by Macdonald et al. [14] was used to identify possible proteins with different amino acid sequences that could result in the same *m/z* value.

3. Results

3.1. Quality Control

K-means clustering was applied, and the resulting cluster distributions were evaluated to assess comparability between datasets and to explore biological correlations within each dataset and between datasets.

Whole mount kidney sections

For the kidney tissue, segmentation analysis based on *N*-glycan signals (57 *m/z* features) was performed with five clusters. In slide B (**Figure 4**) the clusters were clearly separated and closely reflected the underlying tissue biology. In slide C, clustering also distinguished the main tissue regions; however, the cluster boundaries were less well defined compared to slide B, especially cluster 4 (light blue), and the upper part of cluster 3 in slide B (pink) and cluster 4 (orange) in slide C.

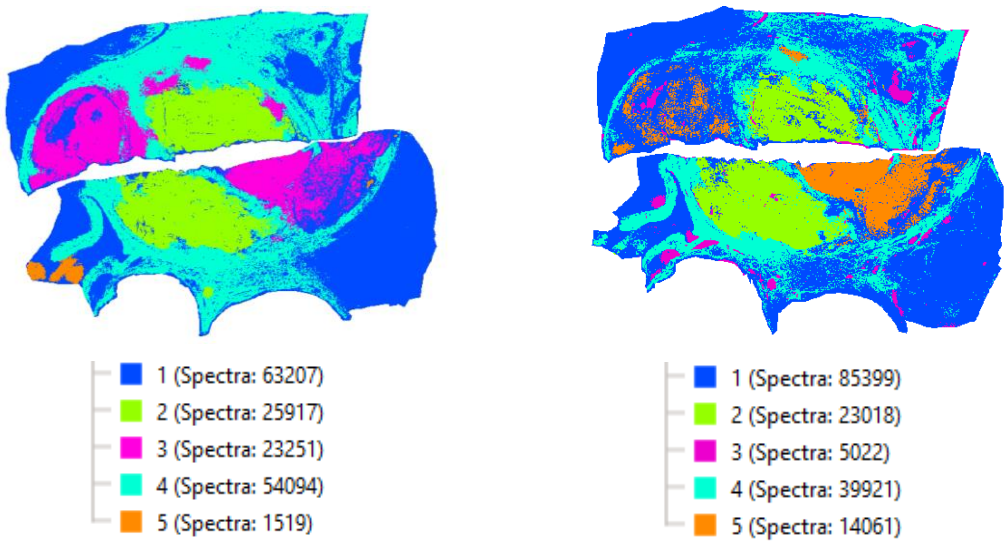


Figure 4. Slide B (left) and slide C (right) segmented into 5 classes based on identical 57 *m/z* peak list after PNGase F treatment, with the corresponding segmentation trees shown below each image.

Tissue Microarray

Segmentation analysis of the TMA, based on the *N*-glycan signals (101 *m/z* features), showed a largely consistent spatial distribution of the clusters for both datasets. Since a cluster containing just a single spectrum (cluster 3, TMA 1 – **Figure 5**) was generated, a sixth class was introduced to improve comparability between the datasets. The segment colors appeared at similar frequencies and in comparable groups of tissue cores. Local

differences in the homogeneity of individual clusters were observed, but no systematic deviations between the slides were detected.

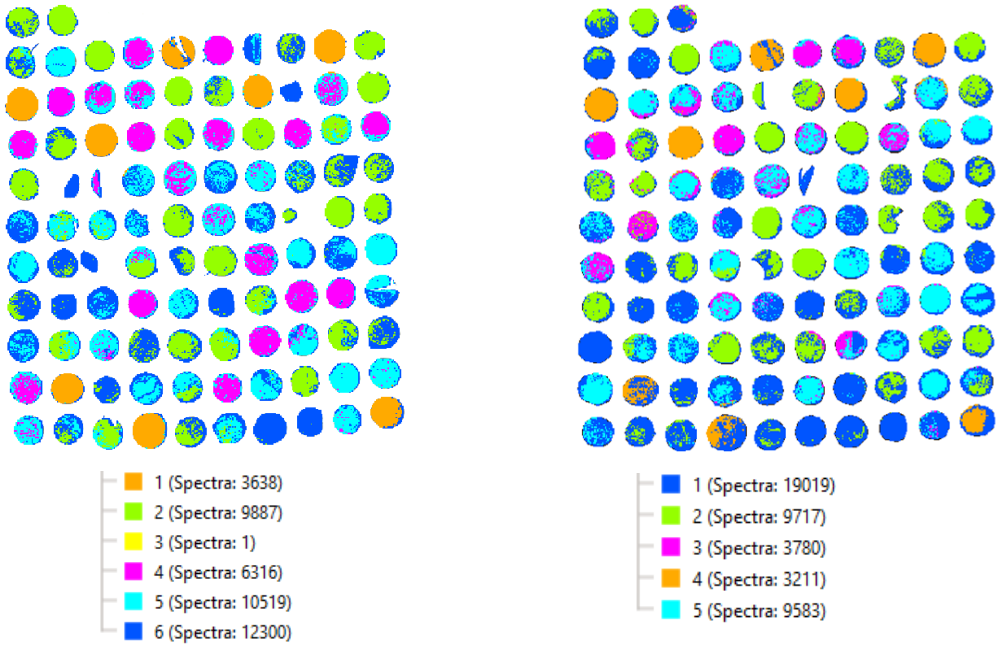


Figure 5. Segmentation of TMA 1 (left, 6 classes) and TMA 2 (right, 5 classes) based on identical 101 *m/z* peak list after PNGase F treatment, with the corresponding segmentation trees shown below.

To ensure that segmentation differences in the TMA and kidney sections did not influence the subsequent enzymatic treatments, ion images of several *N*-glycan signals were visually compared (**Appendix A3**). The high concordance of the spatial distribution of these signals across slides indicates a comparable molecular baseline.

3.2. Histological annotations

Following mass spectrometry acquisition, tissue sections were stained with H&E, and histomorphological regions were annotated by a pathologist in **Figure 6** to improve data interpretability and ensure a more refined analysis and outcome.

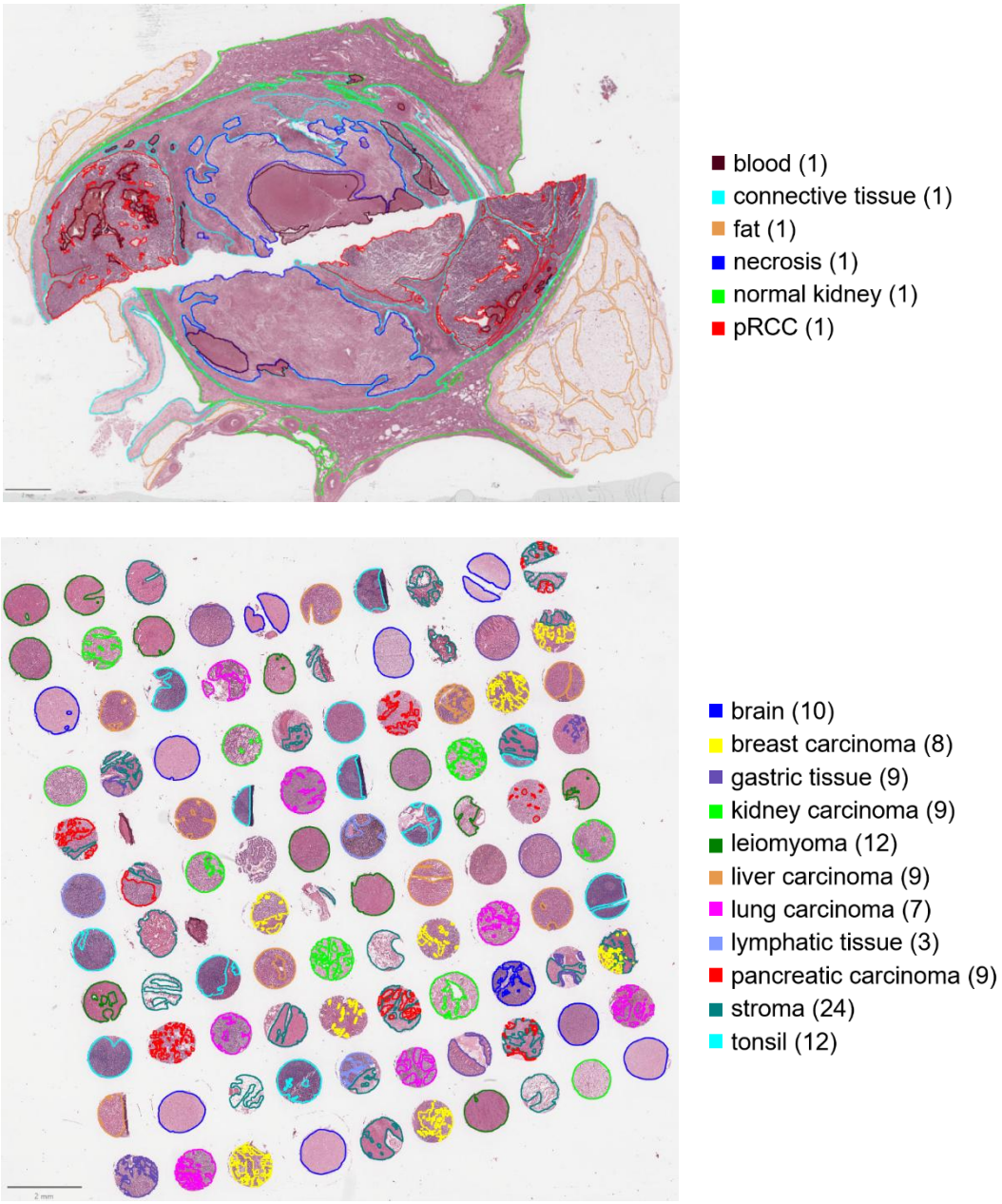


Figure 6. Histological annotations of kidney (top) and TMA (bottom) sections performed by an experienced pathologist using QuPath software. The number of annotated regions is indicated in parentheses.

3.3. Whole mount kidney sections

Peptide expression patterns

After applying the two filter criteria described in the method section, a total of three *m/z* values from trypsin digestion on slide A, B and C and six *m/z* values from COLase 3 digestion on slide C were selected for the comparative analysis in the kidney tissue. As no prior mass calibration was performed before the mass spectrometry measurement of the slide C digested with COLase 3, the detected *m/z* values deviate more noticeably

from the theoretical values in the reference list [14], which is documented in **Appendix A4**. However, the assignment of the signals is additionally based on the second selection criterion described in the Methods section, which requires that collagen-derived peptides from the same digestion show a similar spatial distribution within the tissue. Since the selected peaks showed a consistent correlation, their identification can be considered reliable despite the mass deviation.

Table 2. *m/z* values and correlation scores (Pearson r) for COL1A1-derived peptides detected in slide B (trypsin) and slide C (COLase 3). Peptide sequences are shown with the overlapping regions between the enzymatic digests highlighted in color for clarity. Corresponding correlation data for slide A and C (both trypsin digested) with COLase 3 treated slide C are provided in **Appendix A5**.

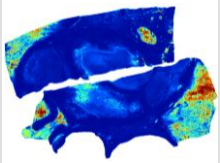
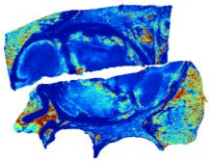
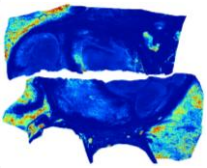
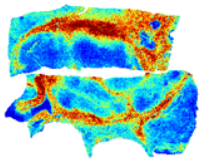
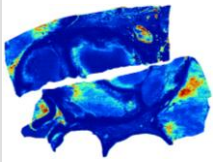
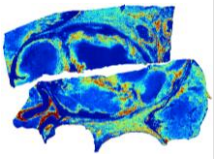
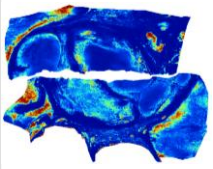
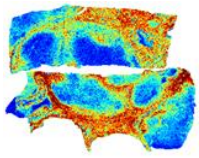
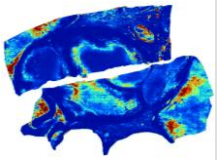
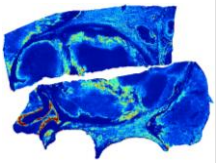
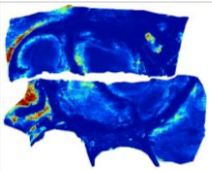
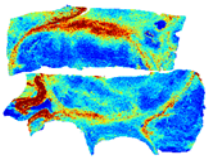
Enzyme	*m/z*	Gene	Peptide sequence	Pearson r
TRY	898.595	COL1A1	GVVGLPGQR	-0.27
COL	1140.277	COL1A1	GIAGQRGVVGLP	
TRY	1105.677	COL1A1	GVQGPPGPAGPR	0.17
COL	823.818	COL1A1	GPVGVQGPP	
TRY	1459.772	COL1A1	GSAGPPGATGFPGAAGR	0.08
COL	797.848	COL1A1	GAAGRVGPP	0.00
COL	1041.070	COL1A1	GAKGARGSAGPP	
TRY	1459.772	COL1A1	GSAGPPGATGFPGAAGR	-0.08
COL	1344.459	COL1A1	GPPGATGFPGAAGRV	
TRY	1459.772	COL1A1	GSAGPPGATGFPGAAGR	-0.06
COL	1360.450	COL1A1	GATGFPGAAGRVGPP	

Ion images were generated to investigate the spatial distribution of these peptides, as shown in **Table 3**.

In general, the three tryptic peptides showed a largely similar distribution pattern with only a very low signal in the connective tissue and similarly weak in parts of the normal kidney tissue, as indicated by the dark blue color in the ion images. This region with low collagen detection was significantly broader and more pronounced in slides A and C than in slide B. However, stronger signals were also observed in certain areas, marked by green to red colors, for example in specific regions of the adipose tissue and at the borders surrounding the connective tissue, as well as in parts of the necrotic areas and normal kidney tissue. In contrast, the ion images after COLase 3 digestion showed a pronounced expression of COL1A1 in the regions where low expression of COL1A1 was detected after tryptic digestion, the connective tissue and parts of the normal tissue. This explains the low or even negative correlation values registered in **Table 2**. Other regions

such as adipose tissue, necrotic areas, and tumor showed, in relative terms, lower signal intensities compared to connective tissue or normal kidney tissue.

Table 3. Ion images of COL1A1-related *m/z* values after tryptic digestion (slides A-C, left columns), shown alongside the corresponding COLase 3 peptide signals in slide C (right column). Each row displays spatial distributions of matched peptide pairs with shared sequence motifs.

Trypsin				COLase 3	
m/z	Slide A	Slide B	Slide C	*m/z*	Slide C
898.513				1140.277	
1105.582				823.818	
1459.701				1344.459	

Peptide expression across histological regions

The distribution of collagen within the tissue will be shown in **Figure 7** using three peptide pairs with sequence overlap between slide B (trypsin digestion) and slide C (COLase 3 digestion).

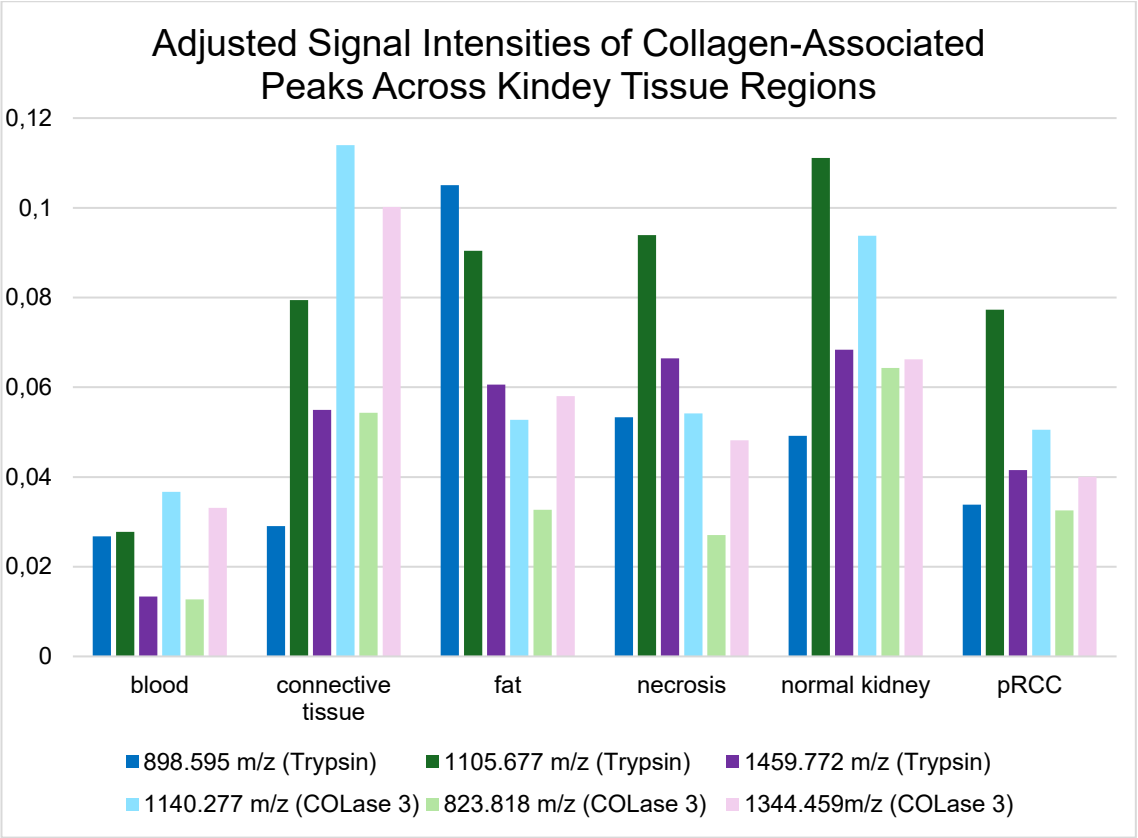


Figure 7. Average signal intensities of selected collagen-associated peptide pairs in kidney tissue regions after digestion with trypsin (slide B) or COLase 3 (slide C). Intensities were TIC-normalized and adjusted to the absolute maximum intensity observed for a collagen-associated peak in the dataset, facilitating the comparison of relative peptide intensity of each digestion method across different datasets. The selected *m/z* pairs originate from the same collagen peptides and were chosen exemplarily to provide a broad presentation in enzymatic performance.

After tryptic digestion, the regions normal kidney and necrosis showed the highest signal intensities for the selected peptides, followed by fat. In contrast, connective tissue, tumor, and blood showed lower intensities. For the COLase 3-derived peptides, the highest signal intensities were measured in connective tissue and normal kidney followed by fat and necrosis. Signal intensities were relatively low in blood and tumor. Among the trypsin-derived peptides, the signal of 1105.677 *m/z* was particularly prominent, consistently reaching the highest relative intensities across several tissue regions.

Some differences in signal intensity were also observed between the trypsin slides with different pretreatments. For all three *m/z* values, the median in the signal intensity boxplots (**Figure 8**) was consistently the highest in slide B, which underwent PNGase F treatment prior to trypsin digestion, while no clear difference was observed between slide A without pretreatment and slide C with PNGase F + COLase 3 pretreatment. The only notable difference was at 1459.701 *m/z*, where slide A had a lower median intensity than slide B, but reached higher values in the upper quartile (Q3). This pattern shows overall lower intensities, but occasionally very high signal intensities, which can also be seen in the corresponding ion images (**Table 3**).

To verify whether the differences observed in the trypsin-derived signals reflected actual effects or were merely due to general differences in overall intensity, the three strongest matrix signals across all three slides were examined (**Figure 9**). These peaks served as an indicator of the general intensity level. Overall, higher values were observed in slide A, while slide B and C differed only slightly from each other, with slide C showing a tendency toward higher intensities.

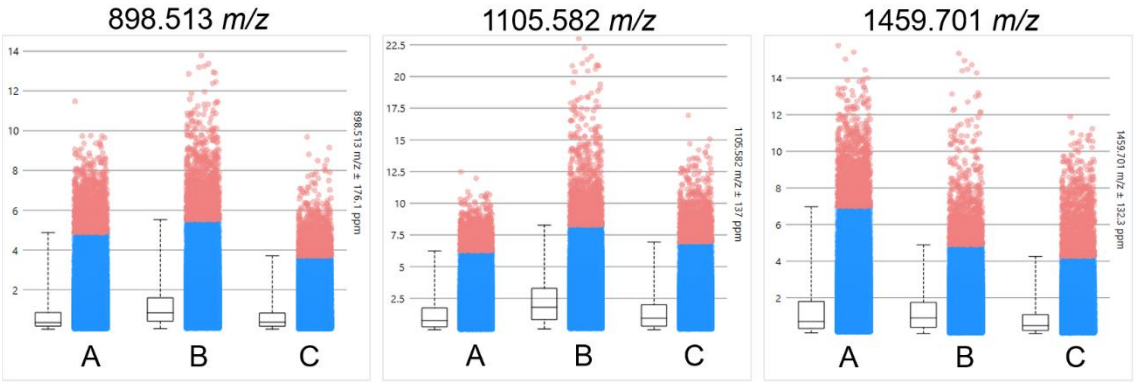


Figure 8. Intensity boxplots of jointly TIC-normalized signals from three selected tryptic peptides of COL1A1, across the kidney section based on histological annotations: 898.513 *m/z* (left), 1105.582 *m/z* (middle), 1459.701 *m/z* (right). For each plot, the first bar represents slide A (no pretreatment), the second slide B (PNGase F pretreatment) and the third slide C (PNGase F + COLase 3 pretreatment).

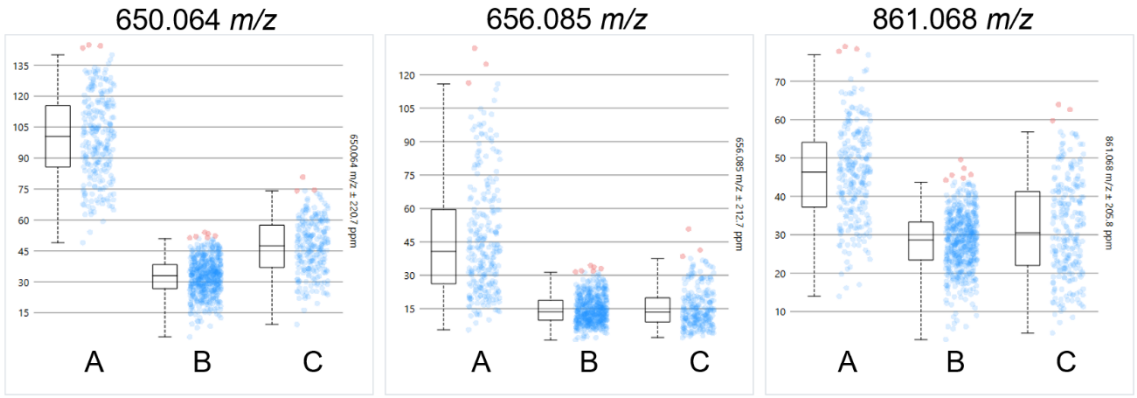


Figure 9. Intensity boxplots of jointly TIC-normalized matrix-associated peaks across slide A (no pretreatment), B (PNGase F pretreatment), and C (PNGase F + COLase 3 pretreatment) corresponding to a pure matrix region without tissue. These three peaks represent the highest intensity matrix signals in the slides. They serve as reference signals to indicate the overall intensity level of the spectra, providing a basis to assess whether differences observed in the trypsin-derived peptide signals are due to true effects or general variations in overall signal intensity.

In addition to the overall higher signal intensity of slide B for the trypsin-derived peptides, a higher number of peaks was also detected. To count the peaks, a threshold value of 10% of the respective maximum intensity was defined for each slide and all peaks above this value were counted. Under these conditions, 240 peaks were identified in slide A, 251 in slide B and 207 in slide C. Especially in the spectra of the slides B and C (**Figure**

10, bottom), peaks in the 1100-1300 *m/z* range are noticeably reduced in slide C compared to slide B, being either absent or with very low intensity.

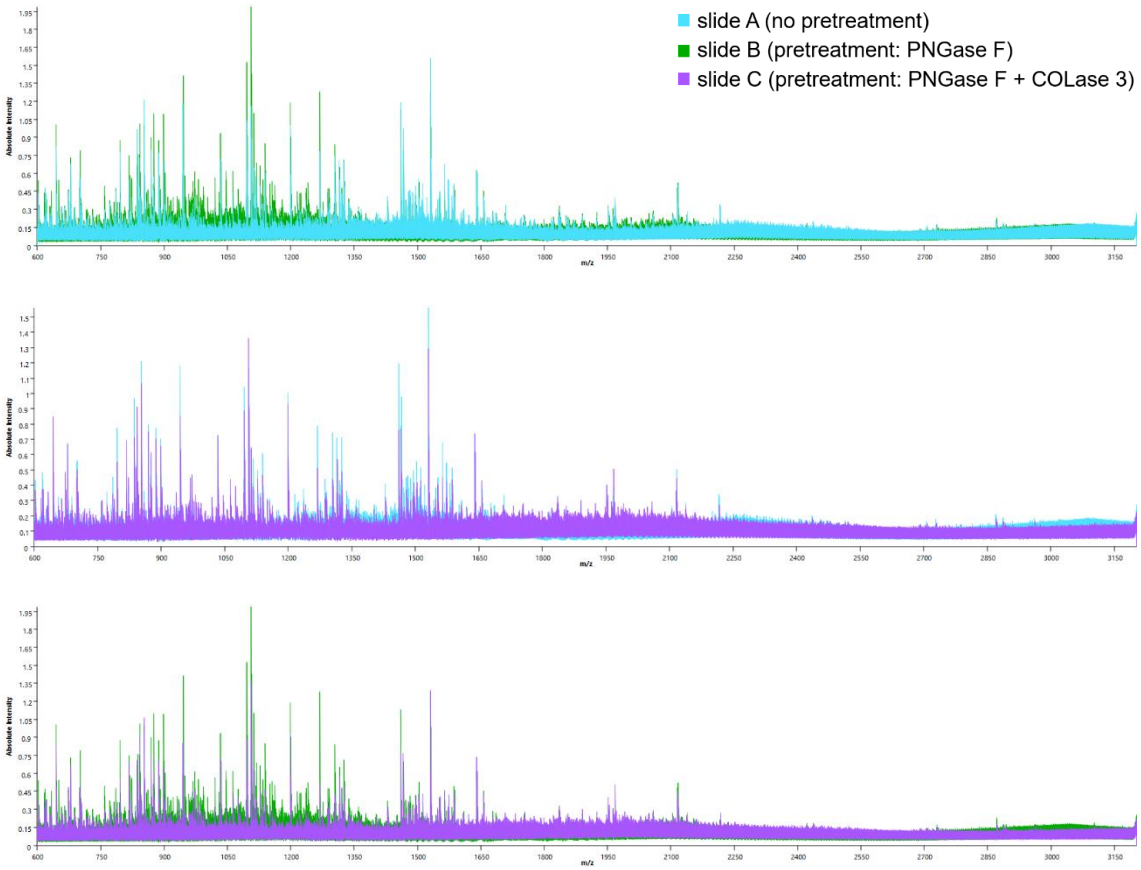


Figure 10. Mass spectra obtained after trypsin digestion of kidney sections subjected to different pretreatments. Top: slide A (blue) compared with slide B (green). Middle: slide A (blue) compared with slide C (purple). Bottom: slide B (green) compared with slide C (purple).

A differential examination of individual tissue regions (**Figure 11**) shows that the observed difference with the higher median values in slide B (**Figure 8**) is primarily due to higher signal intensities in the connective and normal kidney tissues. In all the other regions: blood, fat, necrosis, and pRCC, especially at 898.513 *m/z* and 1105.582 *m/z*, no clear differences were found. The elevated values across all histological annotations in the upper quartile of slide A at 1459.701 *m/z* can be attributed primarily to the fat and necrosis regions, which also observe high local signal intensities in the corresponding ion images, as visible by the intense red areas in **Table 3**. To exclude methodological differences, again these findings were validated with intensity boxplots of the matrix (**Figure 9)**.

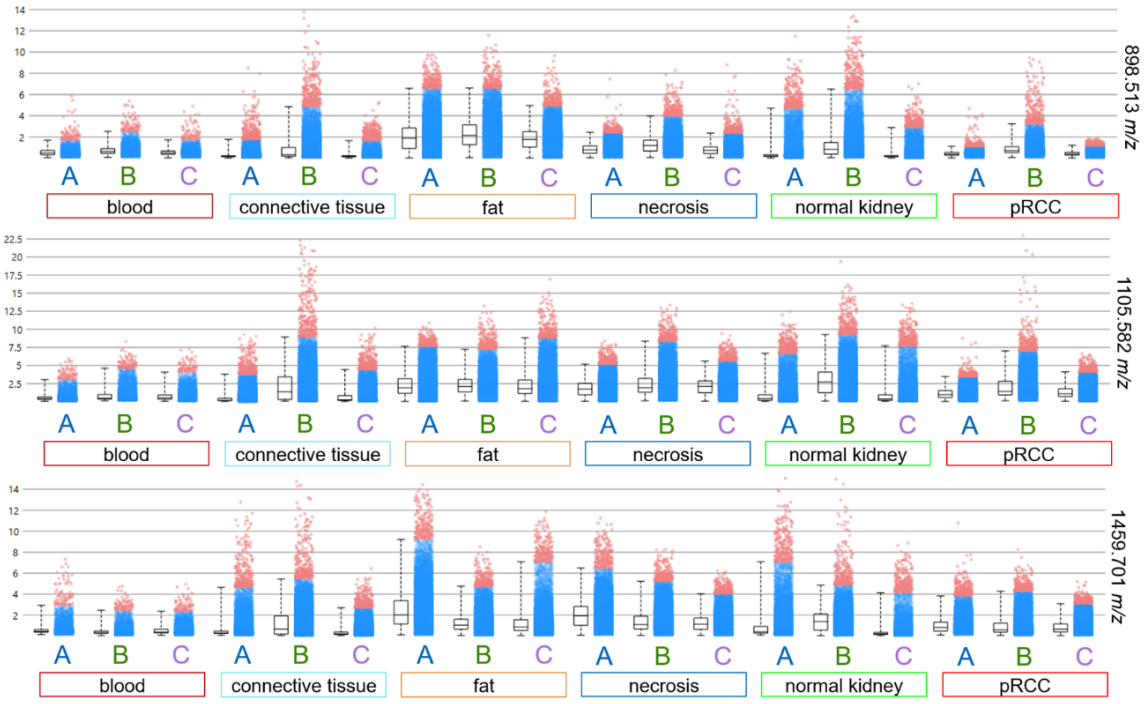


Figure 11. Region-specific signal intensity boxplots of three tryptic peptides (898.513 *m/z* (top), 1105.582 *m/z* (middle) and 1459.701 *m/z* (bottom)) across histologically annotated kidney tissue regions. The three boxplots of each group represent slide A, slide B and slide C, from left to right respectively.

3.4. Mixed tissue microarray

Peptide expression patterns

Following the selection criteria mentioned in the methods section, the same three tryptic peptides derived from collagen were also detected and evaluated on TMA 1 and six COLase 3 derived peptides on TMA 2. The peptide sequences and correlation scores between the different datasets are summarized in **Table 4**.

Table 4. COL1A1-derived peptides, including their *m/z* values and correlation scores (Pearson r) after trypsin (TMA 1) or COLase 3 (TMA 2) digestion. In the peptide sequences, the overlapping regions between the enzymatic digests are highlighted in color.

Enzyme	*m/z*	Gene	Peptide sequence	Pearson r
TRY	898.528	COL1A1	GVVGLPGQR	-0.32
COL	1082.790	COL1A1	IAGQRGVVGLP	-0.14
COL	1421.897	COL1A1	GPQGIAGQRGVVGLP	
TRY	1105.570	COL1A1	GVQGPPGPAGPR	-0.16
COL	823.549	COL1A1	GPVGVQGPP	
TRY	1459.695	COL1A1	GSAGPPGATGFPGAAGR	-0.19
COL	797.531	COL1A1	GAAGRVGPP	

TRY	1459.695	COL1A1	GSAGPPGATGFPGAAGR	-0.13
COL	1343.833	COL1A1	GPPGATGFPGAAGRV	
TRY	1459.695	COL1A1	GSAGPPGATGFPGAAGR	-0.22
COL	1359.801	COL1A1	GATGFPGAAGRVGPP	

To assess the expression patterns of the selected peptides, the corresponding ion images from the two enzymatic treatments were compared. **Figure 12** shows the spatial distribution of the *m/z* values after trypsin or COLase 3 digestion across the tissue cores. The expression patterns differed significantly between the two treatments. Many cores show high COL1A1 expression after COLase 3 digestion, while very low expression was detected in the same location after tryptic digestion. Similarly, the reversed trend was also observed, with some cores showing high expression of COL1A1 after tryptic digestion, but low expression after COLase 3 digestion. A few examples that well illustrate this observation are shown in **Figure 13**.

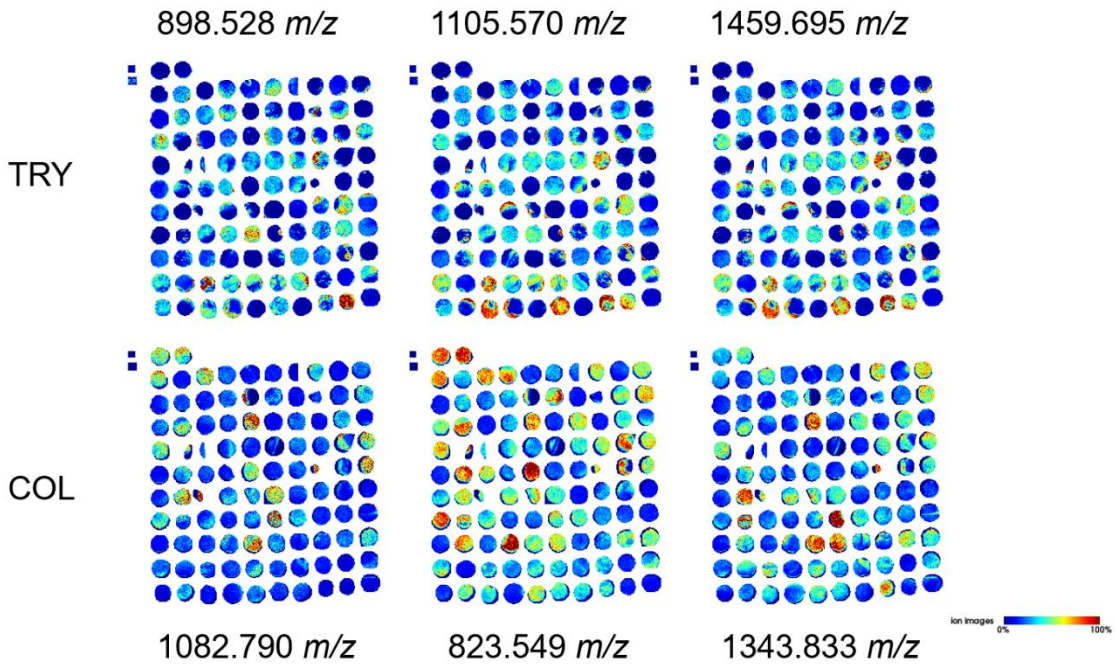


Figure 12. Ion images of TMA 1 (top row, trypsin digestion) and TMA 2 (bottom row, COLase 3 digestion), jointly TIC-normalized, showing spatial distribution of collagen-associated *m/z* values with matching peptide sequences across both enzymatic conditions.

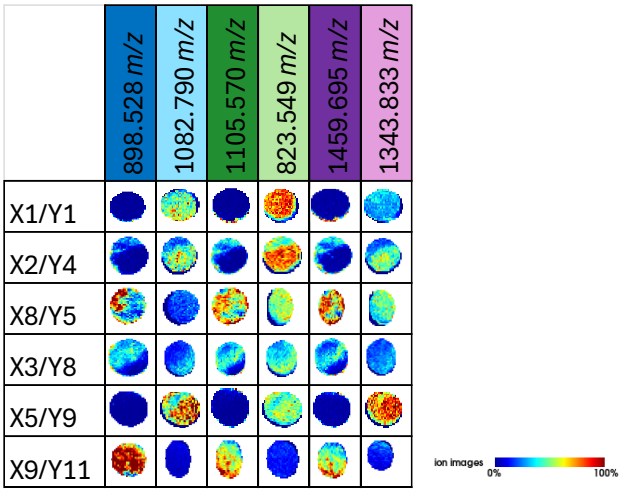


Figure 13. Comparison of COL1A1 peptide fragments derived from trypsin and COLase 3 digestion in some selected tissue cores. *m/z* values to matching peptide sequences are color-coded. Core labels can be referenced in **Figure 2**.

These visual differences were substantiated by the poor peptide colocalization between the two datasets, as determined by Pearson's correlation coefficients, which were consistently very low ($r \leq 0.1$) or negative (**Table 4**).

Peptide expression across tissue types

To discover tissue-specific patterns in the detection of collagen-associated peptides, the average intensities were evaluated for each tissue and enzyme treatment and adjusted to the maximum intensity (**Figure 14**). The values are comparable within an enzyme treatment, a direct comparison between trypsin and COLase 3 is not intended.

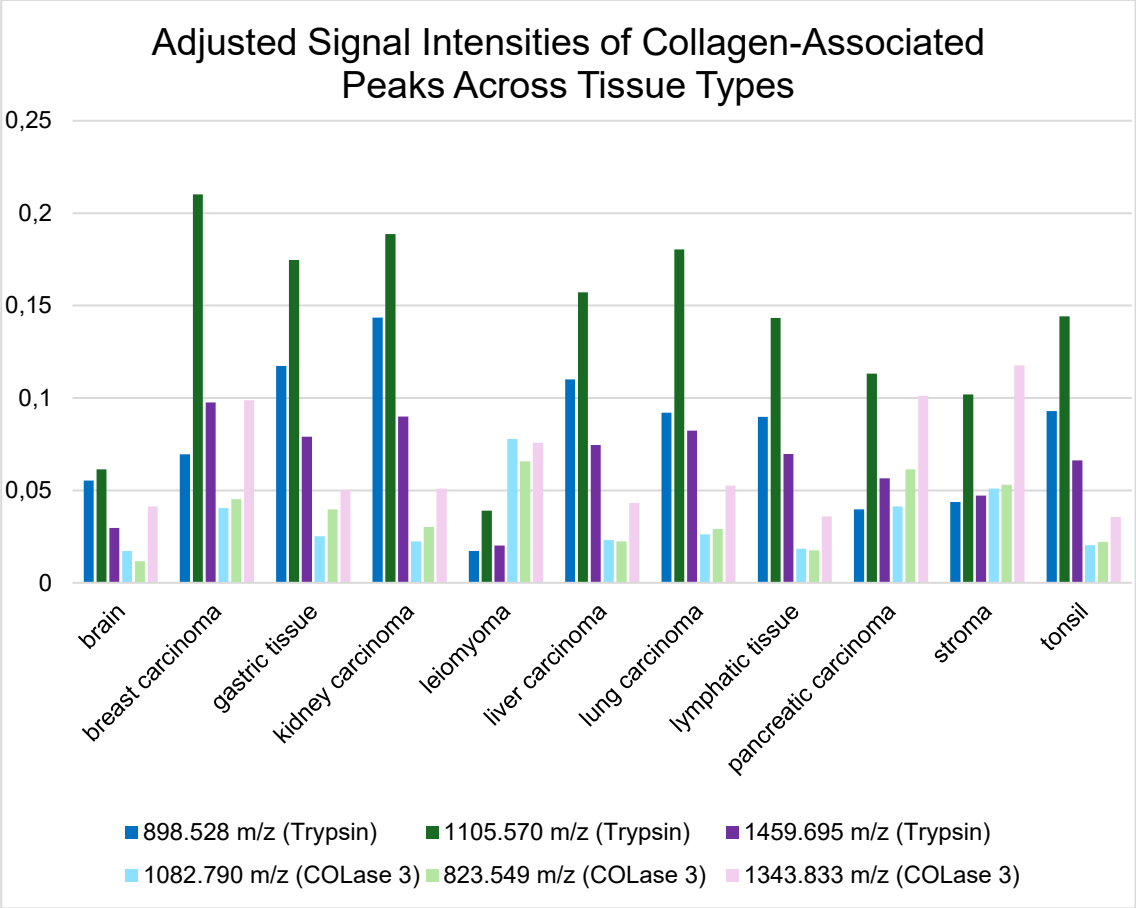


Figure 14. Adjusted average signal intensities of selected collagen-associated *m/z* peaks in different tissue types after digestion with trypsin (TMA 1) or COLase 3 (TMA 2). Intensities were TIC-normalized and adjusted to the absolute maximum intensity observed for a collagen-associated peak in the dataset.

For trypsin, consistently high intensities were detected in kidney carcinoma and gastric tissue, with breast carcinoma also showing high expression in two of the three *m/z* values. Lung carcinoma and liver carcinoma showed medium to high intensities, tonsil and lymphatic tissue medium values, and pancreatic carcinoma, stroma, brain, and leiomyoma the lowest.

For COLase 3, the highest intensities were measured in stroma and leiomyoma, followed by pancreatic carcinoma and breast carcinoma. Medium intensities occurred in lung carcinoma and gastric tissue, followed by kidney carcinoma and liver carcinoma. Low values were observed in tonsil, lymphatic tissue, and brain.

To further explore tissue-specific patterns, stromal regions from breast, gastric tissue, liver, and tonsil were analyzed separately, as shown in **Figure 15**.

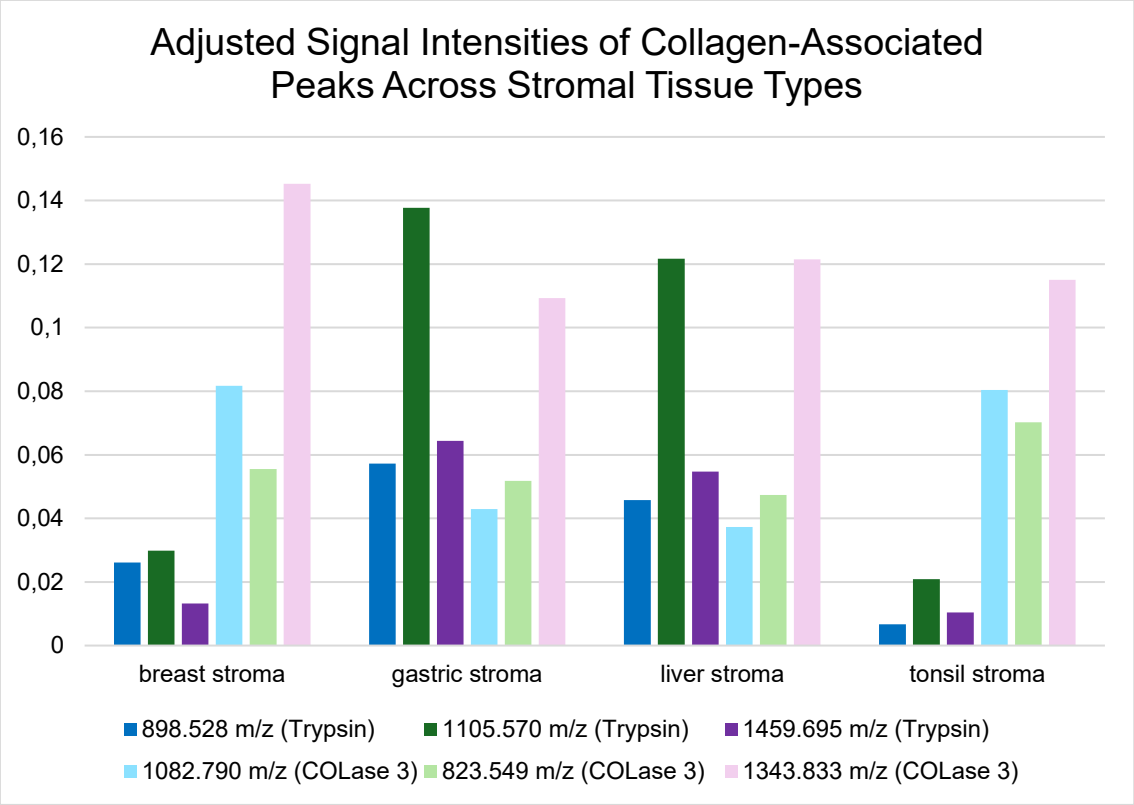


Figure 15. Adjusted average signal intensities of selected collagen-associated *m/z* peaks in stromal tissue types after digestion with trypsin (TMA 1) or COLase 3 (TMA 2). Intensities were TIC-normalized and adjusted to the absolute maximum intensity observed for a collagen-associated peak in the dataset.

A clear pattern emerged for trypsin with gastric stroma and liver stroma showing the highest intensities and tonsil stroma and breast stroma consistently showing the lowest values.

For COLase 3, the distribution was more heterogeneous. The intensities of 1082.790 *m/z* and 823.549 *m/z* largely corresponded, while 1343.833 *m/z* showed particularly in liver stroma deviating results. In general, breast and tonsil stroma tended to show high values, while gastric and liver stroma generally showed lower values.

To investigate differences between the tissue component (healthy and tumorous tissue) and the respective stroma, the signal intensities of the selected peaks were compared in pairs (**Figure 16**). The aim was to analyze how the enzymes behave both in highly collagen-rich and dense structures such as stroma, as well as in corresponding healthy or tumorous tissues.

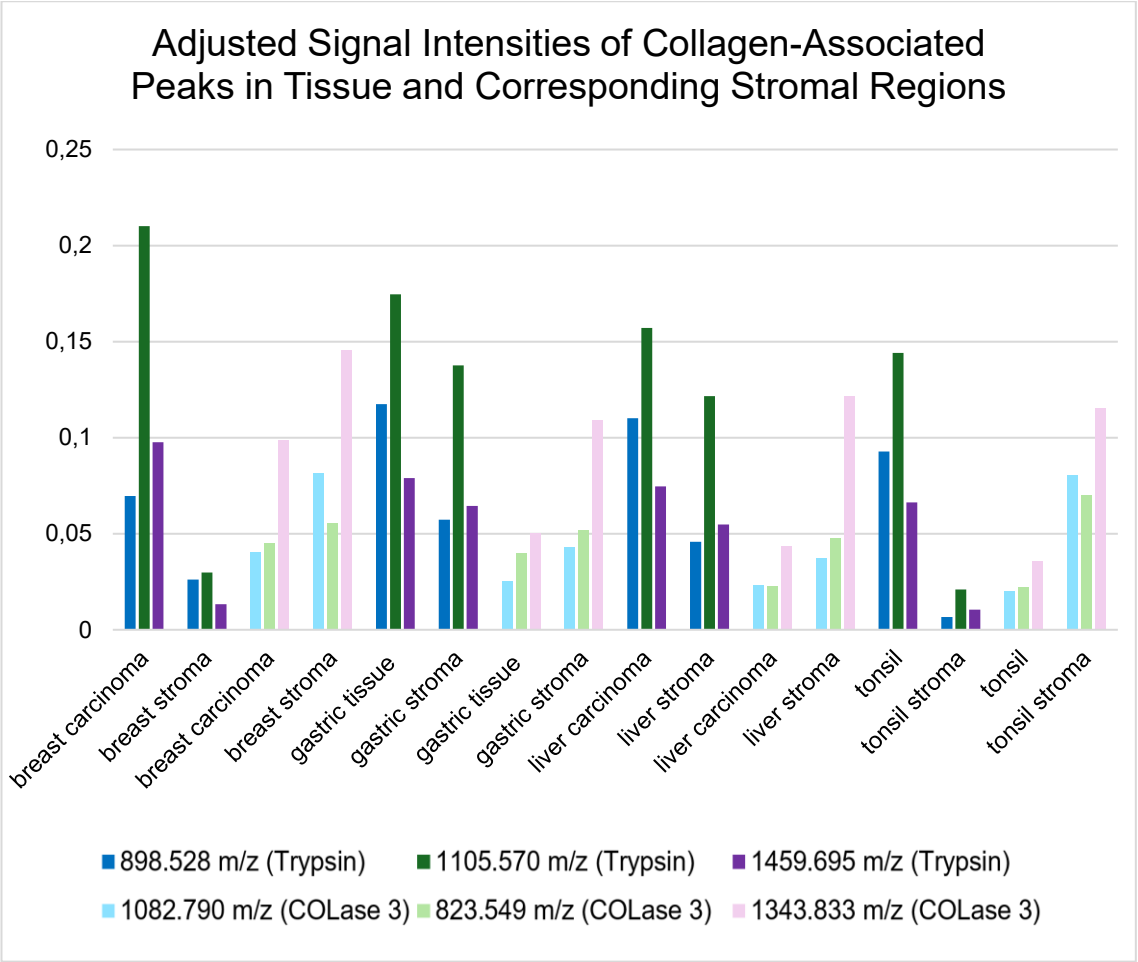


Figure 16. Adjusted average signal intensities of selected collagen-associated *m/z* peaks in tissue and corresponding stromal regions after digestion with trypsin (TMA 1) or COLase 3 (TMA 2). Intensities were TIC-normalized and adjusted to the absolute maximum intensity observed for a collagen-associated peak in the dataset.

The intensities show a consistent pattern across all three peptide pairs. After trypsin digestion, the intensities were consistently higher in the tissue component than in the corresponding stroma. After COLase 3 digestion, the opposite trend emerged with the stroma showing higher intensities than the corresponding tissue components in all pairs.

4. Discussion

Based on the segmentations and the ion images of the *N*-glycans, all samples confirmed comparability of the data sets and a common molecular baseline across all sections. The histological annotations enabled a region-specific evaluation of the peptide distributions. In the kidney sections, trypsin-derived peptides showed the highest intensities in normal kidney tissue and necrotic areas, while COLase 3-derived peptides were most prominent in connective tissue and partially in normal tissue. The correlations between the peptide signals of the two enzymes were overall low or negative. Pretreatment with PNGase F increased the tryptic signal intensity and the number of detected peaks, while pretreatment with COLase 3 decreased the subsequent trypsin signal. In the TMA, trypsin signals were strongest in several tumor tissues (e.g. kidney and breast carcinoma), whereas COLase 3 signals were predominant in stroma-rich regions and in selected tumors such as pancreatic carcinoma and leiomyoma. The stroma-specific analyses revealed tissue dependent patterns. Gastric and liver stroma showed higher tryptic signals, whereas breast and tonsil stroma showed higher COLase 3 signals. A consistent trend emerged across all paired analyses with trypsin consistently yielding higher intensities in tissue compared to stroma, while COLase 3 produced higher signals in stromal regions.

4.1. Whole mount kidney sections

COLase 3 enables better detection of collagen type I in the connective tissue

The ion images of the kidney sections showed clear differences in the spatial distribution and signal intensity of collagen-associated peptides depending on the enzyme used for digestion (**Table 3**). Particularly in the connective tissue, an almost signal-free area was detected after trypsin digestion, whereas an increased signal intensity was visible after COLase 3 digestion. This difference was not only visually apparent but could also be confirmed by region-specific analysis (**Figure 7**). While connective tissue shows the highest average signal intensity compared to all the other tissue regions after COLase 3 digestion, it was the fourth region (out of the six annotated tissue regions) with the highest expression of COL1A1 in the trypsin sample. In the other regions of the kidney section, peptide intensities did not show a strong correlation with the enzymatic treatment, indicating that both enzymes performed similarly. Signal intensities in blood and tumor remained consistently low, while fat, necrotic areas, and normal tissue showed comparable medium to high intensities, independent of the enzyme used.

These findings suggest that trypsin, although widely used in proteomics for its well-known cleavage specificity, shows limited activity in certain tissues, especially in collagen-rich regions of the ECM, such as connective tissue. This is probably due to the dense fibrillar arrangement which hinders enzymatic cleavage by trypsin, and therefore limits the detection with MALDI-TOF MSI [10]. At the edges of the connective tissue, trypsin still had comparatively good access, which is shown by areas of higher signal intensity (red color) and indicates that the ECM architecture appears to be less dense in these regions. Interestingly, Bischoff (1974) was able to show that in skeletal muscle trypsin digested the basal lamina and left only fragments of the sarcolemma, thereby enabling the release of myogenic satellite cells, whereas other enzymes, including collagenase, were unable to do so [24]. These results confirm that trypsin can effectively break down specific structural barriers such as the basement membrane, while remaining limited in compact, fibrillar, and collagen-rich tissue [24]. COLase 3, on the other hand, with high substrate specificity for native, triple-helical collagen, appears to be more effective in cleaving collagen even in compact and collagen dense tissue regions, which improves in situ detection of collagen [16].

This interpretation is further supported by the Van Gieson staining, which specifically stains collagen fibers red/pink, whereas muscle tissue and cytoplasm appear yellow, and IHC staining with an antibody against COL1A1 (**Appendix A6**). Both staining methods showed strong collagen expression in connective tissue and normal kidney tissue, while other areas showed comparatively lower expression. Notably, these staining patterns correspond more closely to the peptide distribution after COLase 3 digestion than after trypsin, suggesting that COLase 3 provides a more realistic molecular representation of collagen-rich tissue.

The improved performance of COLase 3 in detecting collagen-rich tissue regions observed in this study is consistent with previous reports. Macdonald et al., showed that COLase 3 digestion of FFPE liver tissues is particularly effective in enriching ECM proteins, including various types of collagen as well as elastin, biglycan, versican, lumican, and decorin, while trypsin digestion mainly captured proteins from intracellular compartments such as mitochondria and cytoplasm [14]. In a complementary context, Zambrzycki et al. recently introduced an in vitro ECM MSI method and demonstrated that COLase 3-based digestion allows sensitive and spatially resolved analysis of collagen deposition in cell cultures, including dynamic changes under drug treatment, genetic modification, and during time course of cancer cell growth [8]. A comparable advantage of COLase 3 was also described by Angel et al., who established a MALDI-MSI protocol for imaging ECM proteins in FFPE tissues [16]. They showed that COLase 3 digestion strongly enriched collagens and other ECM proteins, while trypsin digestion provided

only very limited collagen coverage. Importantly, they also demonstrated that COLase 3 achieved higher sequence coverage for collagens such as COL1A1 and COL3A1, with peptide distributions that appeared consistent with antibody staining patterns [16]. This is consistent with the results of the present work, in which COLase 3-based peptide maps showed a higher correlation with Van Gieson and IHC staining than the trypsin data, underscoring the more realistic molecular insight into collagen-rich tissue structures provided by COLase 3 [16].

Overall, these results highlight the added value of COLase 3 for in situ digestion of FFPE tissue for MALDI-MSI analysis of collagens and ECM components, especially in tissue contexts where trypsin reaches its limits. Beyond this general finding, the present work adds a new perspective by including systematic region-specific intensity comparisons across several annotated tissue regions. This region-specific approach not only confirms the superior performance of COLase 3 in dense connective tissue but also shows that in less collagen-rich areas, both enzymes have comparably good digestion efficiency.

Limitations of assigning individual *m/z* peaks to a single protein

In addition to the general region-specific differences between trypsin and COLase 3 digestion, the peptide at 1105.677 *m/z* stood out among the trypsin-dependent peaks (**Figure 7**). Its relative intensity was consistently higher than that of the other trypsin-derived collagen-associated peaks across several tissue regions, including blood, connective tissue, necrosis, normal kidney, and pRCC. This strong expression could indicate structural differences, such as the corresponding peptide being located in an exposed region or on the tissue surface, which makes it particularly accessible to trypsin. However, it is also possible that the signal at 1105.677 *m/z* does not originate exclusively from collagen. BLAST analyses (**Appendix A2**) showed that the underlying peptide sequence is not unique to COL1A1, but is shared by several collagens (e.g., COL5A1, COL10A1) and potentially also by non-collagen proteins (e.g., elastin microfibril interface-located protein 1, EMILIN-1) [23]. In addition, different peptide sequences from distinct proteins may generate identical or nearly identical *m/z* values, further complicating clear assignments [25]. The high intensity of this peak could therefore reflect overlapping signals from several proteins that generate the same *m/z* value. This example illustrates a general limitation of MALDI-MSI using TOF analyzers, as their mass accuracy is insufficient to reliably distinguish isobaric or near-isobaric peptides. In contrast, ultra-high-resolution instruments such as Fourier-transform ion cyclotron resonance MSI provide the resolving power required to separate these species and enable more reliable peak assignments [26].

Pretreatment affects trypsin efficiency

Different pretreatment strategies had a noticeable impact on both signal intensity and spatial distribution of trypsin-derived peptides within the tissue. For the three *m/z* values that were examined, slide B, which was pretreated with PNGase F, showed stronger signals across the tissue (**Figure 8**) and especially in collagen-rich connective tissue and normal kidney areas (**Figure 11**). In addition, PNGase F pretreatment also resulted in a higher number of detected peaks compared to slide A without pretreatment (251 vs 240 peaks). In contrast, slide A (no pretreatment) showed lower signal intensities and a reduced number of detected peaks, indicating reduced accessibility or efficiency of trypsin. Based on the boxplots of the matrix peak (**Figure 9**), it can be observed that the overall signal intensity on slide A is slightly higher, which further amplifies the difference. This suggests that the stronger collagen-associated signals on slide B are not just due to overall intensity differences between slides but rather reflects a true effect of the PNGase F pretreatment. This positive effect in slide B can be explained by the removal of steric barriers by *N*-glycans. These sugar structures are covalently bound to proteins and can restrict the accessibility of proteolytic enzymes such as trypsin or COLase 3 to their cleaving sites, thereby reducing catalytic turnover [27]. Enzymatic deglycosylation with PNGase F removes the side chains, making the cleavage sites more accessible. This leads to more efficient proteolysis and improved detection of specific peptide fragments [15]. Several studies show that glycosylation can inhibit the effectiveness of enzymatic cleavage when the affected regions are blocked by sugar residues [15], [28], [29]. For example Clift et al. reported that deglycosylation of FFPE tissue sections with PNGase F prior to COLase 3 digestion showed the most pronounced improvement in COLase 3-derived peptide signals compared to other single-enzyme pretreatments [15]. This proves that the removal of *N*-glycans by PNGase F improves enzymatic access to the ECM and enhances the detection of collagen-associated peptides.

Due to these insights, the present study intentionally avoided using a slide with COLase 3 digestion alone. Instead, COLase 3 was used exclusively in combination with PNGase F to ensure optimal conditions for the release of collagen peptides.

However, slide C, which was pretreated with COLase 3, prior to trypsin digestion, did not lead to a further increase in signal intensity (**Figure 8**). On the contrary, in many regions, especially in connective tissue and normal kidney tissue, lower signal intensities were measured than on slide B (**Figure 11**). Even though the median of the boxplot from the matrix peak on slide C was slightly higher than on slide B (**Figure 9**), indicating an overall higher signal level, the collagen-associated signals were nevertheless reduced. In addition, a lower number of peaks was observed on slide C (207 peaks) compared to slide B

(251 peaks). This reduction is probably due to the fact that during COLase 3 pretreatment, certain peptides were already digested and subsequently washed out, so that they were no longer detectable.

Interestingly, Miersch et al. observed a synergistic effect in the isolation of satellite cells from juvenile pig muscle, in which the combined use of trypsin, collagenase type I and DNase tended to yield more cells per gram of muscle tissue compared to trypsin digestion alone. The authors attribute this increase to the more efficient dissociation of the connective tissue by the collagenase. The combination of trypsin and collagenase I is considered to digest both the basement membrane and the underlying connective tissue more effectively than trypsin alone, thereby releasing more cells. Despite the slight advantage in cell count, there were no differences between the two methods in terms of cell size, viability, or proliferation. The study thus demonstrated that an enzyme combination can increase the efficiency of tissue dissociation without altering the biological properties of the isolated cells [30].

The aforementioned study by Clift et al. not only demonstrated the specific advantage of PNGase F pretreatment, but also showed, using a mass spectrometry approach, that the serial application of multiple ECM-targeting enzymes to a single FFPE tissue section can improve access to different structural components of the ECM. Sequential treatment with chondroitinase ABC, PNGase F, elastase, and finally COLase 3 increased the average signal intensity by 3.1% and the number of identified peptides by 82% compared to COLase 3 alone. The study illustrated that pretreatments can break down structural barriers in the ECM and thereby improve enzyme accessibility [15].

Comparison with the results of the present study suggests that the effect of enzyme combination is highly dependent on the experimental context. In this specific setting, COLase 3 may have caused structural or molecular changes in the ECM, which reduced trypsin's activity. However, it should be noted that this effect was only examined at three specific *m/z* values and in a single experimental run, which limits the generalizability of the findings.

In summary, these results underscore the importance of PNGase F pretreatment for increasing trypsin efficiency, as it improves enzymatic access to glycosylated components of the ECM. At the same time, the data show that sequential pretreatment with COLase 3 can impair trypsin activity, highlighting the need for careful optimization of enzyme combinations and the order of their application for reliable peptide detection in MSI-based ECM analysis.

4.2. Mixed tissue microarray

The low to negative correlation of signal intensities in the TMA between the enzymatic treatments (**Table 4** and **Figure 12**) suggests that the results obtained in the kidney can also be transferred to other tissue types. To test this assumption, a tissue specific analysis was systematically performed considering the different tissue regions by co-registration of the mass spectrometry data with histological annotations.

Enzyme efficiency depends on tissue type and ECM architecture

The TMA results clearly demonstrate that the choice between trypsin and COLase 3 is strongly influenced by the underlying ECM architecture. Both enzyme-dependent and enzyme-independent observations were made. In this context, three ranking patterns emerged that were independent of the enzyme used, showing that brain consistently observed lower signal intensities, lung carcinoma exhibited values in the medium to high and breast carcinoma showed high intensity (**Figure 14**). This consistency suggests that these tissues contain structures that are comparably accessible to both enzymes, though not necessarily to the same extent.

This finding is expected for the brain. Fibrillar collagen is physiologically present only in small amounts, for example collagen I has a content between 0 and 0.015 ng/mg total protein [31]. The ECM consists predominantly of non-fibrillar components such as hyaluronic acid, proteoglycans and glycoproteins, which explains why collagen-associated signals are low in both digestion protocols [32].

In solid tumors profound ECM remodeling typically increases deposition of fibrillar collagens I and III, accompanied by densification, alignment and cross-linking of the fibers, which can reduce accessibility for proteolytic enzymes [33]. Nevertheless, lung and breast carcinomas show comparable intensity levels for trypsin and COLase 3, suggesting that the collagen architecture in these tumors provides a relatively good accessibility for both enzymes. In other tissues, including liver carcinoma, tonsil, lymphatic tissue, stomach tissue and kidney carcinoma, trypsin reached relatively higher intensity levels. This is particularly pronounced in kidney carcinoma, where trypsin shows the highest relative accessibility of all tissue types. Tissues showing relatively high peptide intensities after tryptic digestion are likely to be less dense, have a less cross-linked collagen architecture, contain well-exposed tryptic cleavage sites, and contain more parenchymal areas with little compact stroma, which are all factors that can favor trypsin activity.

Significant enzyme-dependent differences are particularly evident in tissues with a high stroma content and dense ECM structure, including leiomyomas, tumors such as pancreatic carcinoma and stroma. In these cases, COLase 3 constantly presents higher digestion efficiency, while trypsin consistently shows low values. The compact architecture of such tissues acts as a physical barrier to certain proteolytic enzymes [34], which explains the lower peptide release by trypsin and the clear increase in signal after COLase 3 digestion. This interpretation is supported by IHC staining using an antibody against COL1A1, whose distribution pattern is more consistent with the COLase 3 digestion results than with the trypsin data (**Appendix A6**).

This pattern corresponds to the principle observed in the kidney sections, that COLase 3 opens up the dense matrix. In highly fibrillar, densely packed and cross-linked ECM domains, the interfaces for trypsin are significantly less accessible, both sterically and structurally. COLase 3, which specifically targets native triple-helical collagen regions can release peptides more efficiently there, thus reversing the ranking in favor of stroma-rich tissues.

Stroma structure determines enzyme efficiency

Stroma-specific analysis reveals a clear, partly mirror-image pattern between the two enzymes. Stroma types with high intensities for trypsin such as gastric stroma or liver stroma achieve comparatively low values for COLase 3. Conversely, stroma types with low COL1A1 peptide intensities values after trypsin digestion, such as breast stroma or tonsil stroma, show medium to high intensities after COLase 3 digestion (**Figure 15**).

This pattern reflects the varied composition of the ECM structure in the samples examined. Gastric and liver stroma are likely to be less densely packed and less strongly cross-linked giving trypsin easier access to its cleavage sites. Tonsil or breast stroma, on the other hand, are typically denser, more branched and richer in collagen, which limits tryptic accessibility, while COLase 3 can cleave peptides more effectively there due to its specificity. The partly heterogeneous distribution of individual *m/z* signals after COLase 3 digestion further indicates that even within the same collagen type, different sequence regions are variably accessible, which underscores the importance of considering multiple *m/z* values for a valid interpretation.

These findings mirror the observations made in the kidney tissue, where trypsin and COLase 3 produced complementary results and confirm that the outcome strongly depends on the choice of enzyme. They also demonstrate that different stromal types show distinct ECM architectures varying in density, degree of cross-linking and fiber organization, which determines the relative accessibility for each enzyme.

Enzyme performance differs between stroma and parenchyma

A direct comparison of stroma and the associated reference tissue shows a consistent pattern across all tissues examined. After trypsin digestion, the signal intensities in the tissue compartment (e.g. tumor or normal tissue) are always higher than in the corresponding stroma, whereas after COLase 3 digestion, this trend is reversed with stroma showing higher values in all cases (**Figure 16**).

This consistent picture confirms the different strengths of the two enzymes. Trypsin is particularly effective in areas with less compact ECM, as often found in parenchymal tissues, while COLase 3 shows its advantages in densely packed, fibril-rich structures such as stroma. The clear separation in enzymatic preferences suggests that the ECM architecture is the decisive factor for the detection patterns.

The enzyme-dependent differences observed in the kidney section with clear advantages of COLase 3 in dense, collagen-rich tissue and better performance of trypsin in less compact structures compared to dense tissue, can also be found in the TMA data. The tissue-specific ranking patterns, the stroma-specific differences and the direct comparison between stroma and associated tissue confirm that the enzyme efficiency depends significantly on the ECM architecture. At the same time, the TMA analyses show that these principles apply not only to a single organ but across different tissue types, while organ-specific deviations caused by specific ECM compositions and architecture need to be considered when optimizing future strategies.

4.3. Limitations

Despite the clear trends that emerge in this work, the results must be interpreted with certain limitations. First, many of the observations are based on rank comparisons within the same slides, while a direct comparison of absolute signal intensities between experiments was not possible. The reason for this is that different measurements were performed with different enzymes and numerous influencing factors, such as differences in the ionizability of the peptides, matrix crystallization, variability in enzyme digestion, and technical fluctuations prevent a direct comparison of the absolute values, even after TIC normalization. Consequently, the evaluation had to be limited to relative intensity comparison within each slide.

In addition, the number of *m/z* features considered was limited, which means that individual results can only be generalized to a limited extent. Furthermore, the results were

obtained exclusively from single experiments and were therefore not confirmed by repetitions. Another point is that patient-specific variance is a potential influencing factor that could have affected the signal intensities and ranking patterns.

Another limitation relates to the calculation of Pearson correlations in Ion Image Mapper. The reliability of these correlations depends directly on the accuracy of image co-registration, which involves a certain degree of variability. This also applies to the alignment with histological annotations, where even small deviations in the co-registration can substantially affect the resulting correlation values.

Finally, it should be emphasized that individual peptide sequences and thus also the corresponding *m/z* values cannot be uniquely assigned to a single protein, as already shown in "**Limitations of assigning individual *m/z* peaks to a single protein**". This may result in overlapping contributions from different proteins, which makes it difficult to clearly assign individual peaks. Consequently, mismatches may occur, and the observed signal intensity can appear artificially increased.

4.4. Outlook

Before final conclusions can be drawn, these experiments need to be repeated using other samples to confirm the robustness of the results. Nevertheless, the present results highlight both the opportunities and challenges of MALDI-MSI in ECM proteomics. This work provides a methodological basis for optimized workflows that use complementary enzyme specificities to improve ECM coverage. Future research should therefore not only systematically investigate serial digestion strategies and the sequence of enzyme application but also expand the range of enzymes. For example, elastase or chondroitinase could provide complementary cleavage patterns and thus broaden the detectable spectrum of ECM components. In order to obtain a more complete picture, the spatial characterization should also be extended beyond collagen type I to other collagens (e.g. type III) and to non-collagen ECM proteins (e.g. elastin), while also considering post-translational modifications as an additional layer of molecular complexity [10].

Another important step is to further develop multimodal strategies. For example the combination of MALDI-MSI with immunofluorescence microscopy or with spatial transcriptomics could make it possible to link molecular composition with structural and cellular context and thus gain a more complete picture of ECM organization in health and disease [9].

5. Conclusion

The results of this study clearly show that the choice of enzymatic strategy in MALDI-MSI analysis has a decisive influence on the evaluation of ECM components, especially collagen, and that this effect is largely determined by the structural properties of the respective tissue. Both in the kidney sections and in the TMA, albeit with organ-specific deviations, a consistent overall trend could be identified. COLase 3 proved to be particularly advantageous in collagen-rich, densely packed ECM regions, where structural barriers significantly limit the cleavage activity of trypsin. Due to its substrate specificity for native, triple-helical collagen, it enables a more effective release of peptides from fibril networks that remain largely inaccessible to conventional proteases due to structural barriers. Trypsin is most effective in less densely fibrillar and weaker cross-linked ECM areas, where its well characterized cleavage specificity enables broad proteome coverage.

This observation has far-reaching consequences for proteomic workflows: analysis using trypsin alone significantly limits the results. Collagens are the most abundant protein in the human body and are central to tissue architecture. If they are inefficiently detected or not detected at all by trypsin, crucial information is lost. When using a single enzyme, general statements about the entire slide should therefore be interpreted with caution, as the efficiency of the enzyme is not homogeneous and certain tissue compartments are systematically disadvantaged. Differences in the results can have biological causes, such as stroma vs. parenchyma or disease-related changes in stroma organization during the course of a disease but can also reflect methodological effects resulting from differences in enzyme accessibility. For a valid interpretation, it is therefore essential to always take enzymatic efficiency into account. To enable a complete representation and analysis of the ECM, the use of complementary or serially applied protocols that include both enzymes is recommended. This approach allows for a more complete and balanced representation of the tissue.

Furthermore, the results highlight the central importance of appropriate pretreatments, in particular deglycosylation with PNGase F, for improving enzymatic accessibility. The removal of *N*-glycans led to significantly improved detection of trypsin fragments and underscores the need to consider posttranslational modifications as structural barriers in ECM analysis. However, the combination of COLase 3 pretreatment followed by trypsin did not lead to synergistic effects in this experimental setting, but rather to partial signal loss in some cases, which shows that enzyme combinations must be carefully optimized. Nevertheless, further and repeated experiments will be necessary to clarify whether the

observed reduction after COLase 3 pretreatment represents a consistent effect or may have occurred by chance.

In summary, it is clear that focusing solely on trypsin for ECM analysis is insufficient and potentially leads to erroneous conclusions. Collagen distribution can only be reliably shown by combining complementary enzymes and targeted pretreatments. This work not only confirms the complementary properties of trypsin and COLase 3 but also provides a methodological basis for overcoming the limitations of spatial proteomics with optimized protocols and describing collagen architecture in healthy and diseased tissue more precisely.

Appendix

A1: Pearson's correlation values used for selection criterion 2: Spatial consistency of collagen-associated *m/z* signals within one slide, based on histological annotations

Table 5. Spatial correlation scores (Pearson r) of selected *m/z* values after trypsin digestion in kidney slide A, B and C

slide A			
m/z	898.437	1105.501	1459.592
898.437	1.00	0.72	0.66
1105.501	0.72	1.00	0.94
1459.592	0.66	0.94	1.00

slide B			
m/z	898.595	1105.677	1459.772
898.595	1.00	0.58	0.49
1105.677	0.58	1.00	0.92
1459.772	0.49	0.92	1.00

slide C			
m/z	898.470	1105.508	1459.658
898.470	1.00	0.67	0.57
1105.508	0.67	1.00	0.85
1459.658	0.57	0.85	1.00

Table 6. Spatial correlation scores (Pearson r) of selected *m/z* values after COLase 3 digestion in kidney slide C

slide C						
m/z	797.848	823.818	1041.070	1140.277	1344.459	1360.450
797.848	1.00	0.68	0.80	0.79	0.73	0.65
823.818	0.68	1.00	0.57	0.53	0.35	0.48
1041.070	0.80	0.57	1.00	0.86	0.77	0.77
1140.277	0.79	0.53	0.86	1.00	0.80	0.77
1344.459	0.73	0.35	0.77	0.80	1.00	0.76
1360.450	0.65	0.48	0.77	0.77	0.76	1.00

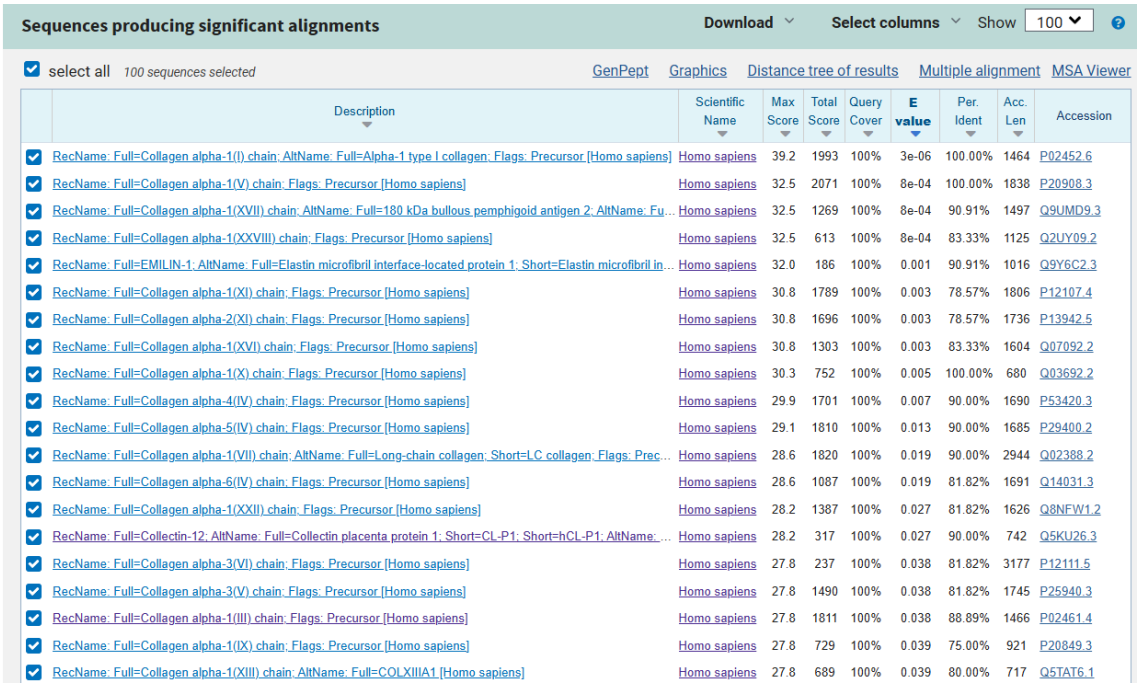
Table 7. Spatial correlation scores (Pearson r) of selected *m/z* values after trypsin digestion in TMA 1

TMA 1			
m/z	898.528	1105.570	1459.695
898.528	1.00	0.71	0.66
1105.570	0.71	1.00	0.95
1459.695	0.66	0.95	1.00

Table 8. Spatial correlation scores (Pearson r) of selected *m/z* values after COLase 3 digestion in TMA 2

TMA 2						
m/z	797.531	823.549	1082.790	1343.833	1359.801	1421.897
797.531	1.00	0.69	0.71	0.68	0.53	0.55
823.549	0.69	1.00	0.58	0.44	0.59	0.38
1082.790	0.71	0.58	1.00	0.58	0.69	0.67
1343.833	0.68	0.44	0.58	1.00	0.56	0.72
1359.801	0.53	0.59	0.69	0.56	1.00	0.65
1421.897	0.55	0.38	0.67	0.72	0.65	1.00

A2: Exemplary peptides assignments illustrating non-unique correspondence of *m/z* values to proteins



Sequences producing significant alignments Download Select columns Show 100

select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment MSA Viewer

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Collagen alpha-1(I) chain; AltName: Full=Alpha-1 type I collagen; Flags: Precursor [Homo sapiens]	Homo sapiens	39.2	1993	100%	3e-06	100.00%	1464	P02452.6
RecName: Full=Collagen alpha-1(V) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	32.5	2071	100%	8e-04	100.00%	1838	P20908.3
RecName: Full=Collagen alpha-1(XVII) chain; AltName: Full=180 kDa bullous pemphigoid antigen 2; AltName: Fu...	Homo sapiens	32.5	1269	100%	8e-04	90.91%	1497	Q9UMD9.3
RecName: Full=Collagen alpha-1(XXVIII) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	32.5	613	100%	8e-04	83.33%	1125	Q2UY09.2
RecName: Full=EMILIN-1; AltName: Full=Elastin microfibril interface-located protein 1; Short=Elastin microfibril in...	Homo sapiens	32.0	186	100%	0.001	90.91%	1016	Q9Y6C2.3
RecName: Full=Collagen alpha-1(XI) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	30.8	1789	100%	0.003	78.57%	1806	P12107.4
RecName: Full=Collagen alpha-2(XI) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	30.8	1696	100%	0.003	78.57%	1736	P13942.5
RecName: Full=Collagen alpha-1(XVI) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	30.8	1303	100%	0.003	83.33%	1604	Q07092.2
RecName: Full=Collagen alpha-1(X) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	30.3	752	100%	0.005	100.00%	680	Q03692.2
RecName: Full=Collagen alpha-4(IV) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	29.9	1701	100%	0.007	90.00%	1690	P53420.3
RecName: Full=Collagen alpha-5(IV) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	29.1	1810	100%	0.013	90.00%	1685	P29400.2
RecName: Full=Collagen alpha-1(VII) chain; AltName: Full=Long-chain collagen; Short=LC collagen; Flags: Prec...	Homo sapiens	28.6	1820	100%	0.019	90.00%	2944	Q02388.2
RecName: Full=Collagen alpha-6(IV) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	28.6	1087	100%	0.019	81.82%	1691	Q14031.3
RecName: Full=Collagen alpha-1(XXII) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	28.2	1387	100%	0.027	81.82%	1626	Q8NFW1.2
RecName: Full=Collectin-12; AltName: Full=Collectin placenta protein 1; Short=CL-P1; Short=hCL-P1; AltName: ...	Homo sapiens	28.2	317	100%	0.027	90.00%	742	Q5KU26.3
RecName: Full=Collagen alpha-3(VI) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	27.8	237	100%	0.038	81.82%	3177	P12111.5
RecName: Full=Collagen alpha-3(V) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	27.8	1490	100%	0.038	81.82%	1745	P25940.3
RecName: Full=Collagen alpha-1(III) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	27.8	1811	100%	0.038	88.89%	1466	P02461.4
RecName: Full=Collagen alpha-1(IX) chain; Flags: Precursor [Homo sapiens]	Homo sapiens	27.8	729	100%	0.039	75.00%	921	P20849.3
RecName: Full=Collagen alpha-1(XIII) chain; AltName: Full=COLXIIIA1 [Homo sapiens]	Homo sapiens	27.8	689	100%	0.039	80.00%	717	Q5TAT6.1

Figure 17. Exemplary BLAST search results for the peptide sequence GVQGPPGPAGPR (1105.58 *m/z*)

MALDI MSI M/Z Value	Mass+H (Da)	Mass Error (ppm)	MALDI Match	Experiment	Protein Description	Gene	Cellular Location	Peptide with Modifications
1139.5948	1139.592	-2.46	Both	LSE	Collagen alpha-1(I) chain	COL1A1	Extracellular	GPRGRTGDAGPV
	1139.5947	-0.09	Both	In Solution	Alpha-actinin-4	ACTN4	Cellular	IVHTIEEIEG
	1139.6058	9.65	Both	In Solution	Elastin	ELN	Extracellular	VGGVP[1]GVGGVP[1]GVG
	1139.6058	9.65	Both	LSE	Elastin	ELN	Extracellular	VGGVP[1]GVGGVP[1]GVG
1139.6518	1139.6463	-4.83	Both	In Solution	Elastin	ELN	Extracellular	FPGVGVLPGVPT
	1139.6536	1.58	Both	In Solution	Collagen alpha-1(III) chain	COL3A1	Extracellular	GPLGIAGITGARG
	1139.6536	1.58	Both	In Solution	Collagen alpha-1(I) chain	COL1A1	Extracellular	GIAGQRGVVGLP[1]
	1139.6536	1.58	Both	LSE	Collagen alpha-1(III) chain	COL3A1	Extracellular	GPLGIAGITGARG
	1139.6536	1.58	Both	LSE	Collagen alpha-1(I) chain	COL1A1	Extracellular	GIAGQRGVVGLP[1]

Figure 18. Extract from the COLase 3 peptide list for 1139 *m/z* reported by Macdonald et al. [14]

A3: Representative *m/z* distributions of *N*-glycans in TMA and kidney tissue to demonstrate reproducibility

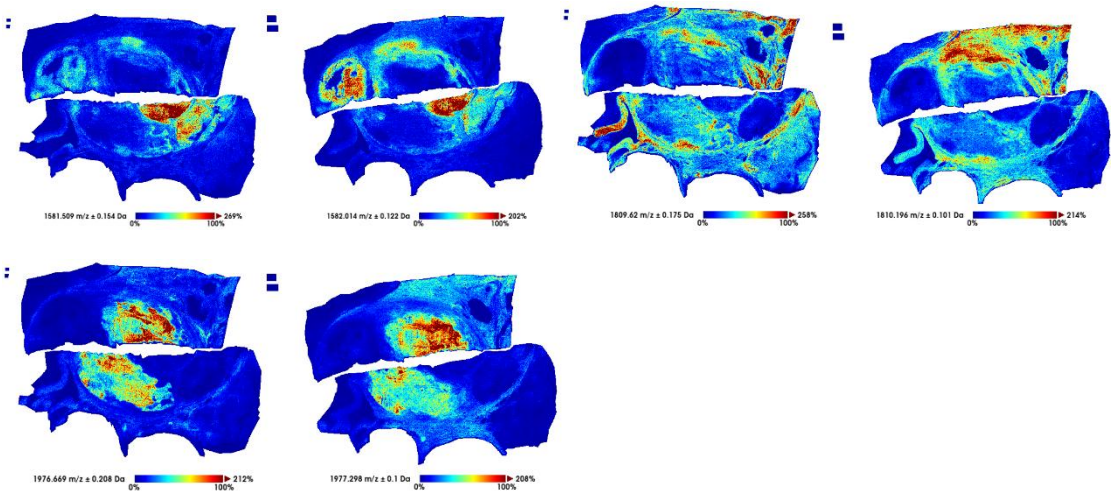


Figure 19. Ion images of *N*-glycan signals in kidney tissue. Top row (from left to right): 1851.509 *m/z*, 1582.014 *m/z*, 1809.620 *m/z*, and 1810.196 *m/z*. Bottom row (from left to right): 1976.669 *m/z* and 1977.298 *m/z*. For each *m/z* pair, the image on the left corresponds to slide B and the image on the right to slide C. The consistent spatial distributions confirm reproducibility of the experimental data.

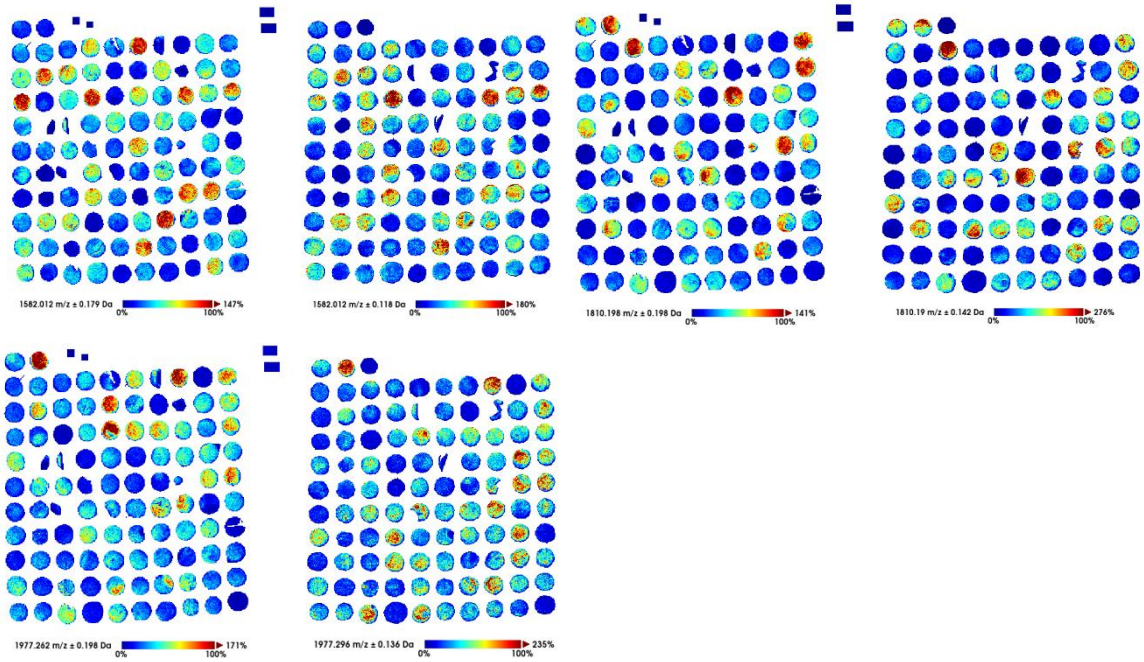


Figure 20. Ion images of *N*-glycan signals in TMA sections. Top row (from left to right): 1582.012 *m/z*, 1582.012 *m/z*, 1810.198 *m/z*, and 1810.190 *m/z*. Bottom row (from left to right): 1977.262 *m/z* and 1977.296 *m/z*. For each *m/z* pair, the image in the left column corresponds to TMA 1 and the image in the right column to TMA 2. The consistent spatial distributions across both TMAs confirm reproducibility of the experimental data.

A4: Deviations between measured and reference *m/z* values after enzymatic digestion

Table 9. Measured *m/z* values and their deviation from the theoretical reference values for selected peptides after trypsin (slide B, TMA 1) and COLase 3 (slide C, TMA 2) digestion. Data are shown for kidney tissue and TMA samples. Reference values for COLase 3-derived peptides were taken from [14], and for trypsin-derived peptides from [22].

		kidney		TMA	
	reference value *m/z*	**measured value *m/z***	**deviation**	**measured value *m/z***	**deviation**
trypsin	**898.48**	898.595	0.12	898.528	0.05
	1105.58	1105.677	0.10	1105.570	-0.01
	1459.71	1459.772	0.06	1459.695	-0.02
COLase 3	**797.4269**	797.848	0.42	797.531	0.10
	823.4313	823.818	0.39	823.549	0.12
	1041.5439	1041.070	-0.47		
	1082.6321			1082.790	0.16
	1139.6536	1140.277	0.62		
	1343.6705	1344.459	0.79	1343.833	0.16
	1359.6655	1360.450	0.78	1359.801	0.14
	1421.7864			1421.897	0.11

A5: Correlation of COL1A1 peptides across slides

Table 10. Additional correlation data (Pearson r) of COL1A1-derived peptides between trypsin-digested slides A and C and COLase 3-digested slide C. For each peptide, the respective *m/z* values are listed.

	Slide A		Slide C	
enzyme	***m/z***	**Pearson r**	***m/z***	**Pearson r**
TRY COL	898.437 1140.277	-0.34	898.470 1140.277	-0.37
TRY COL	1105.501 823.818	-0.23	1105.508 823.818	-0.21
TRY COL COL	1459.592 797.848 1041.070	-0.29 -0.32	1459.658 797.848 1041.070	-0.23 -0.30
TRY COL	1459.592 1344.459	-0.25	1459.658 1344.459	-0.20
TRY COL	1459.592 1360.450	-0.25	1459.658 1360.450	-0.26

A6: Histological and immunohistochemical staining for validation of MSI results

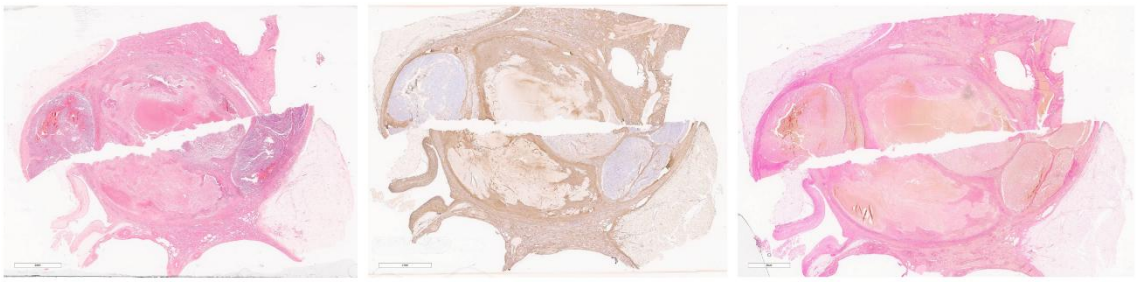


Figure 21. Histological staining of kidney tissue: H&E (left) for general morphology, IHC with antibody against COL1A1 (middle) for type I collagen and Van Gieson (right) for collagen-rich structures

Figure 22. TMA stained with H&E to assess tissue morphology (left) and IHC using COL1A1 antibody to detect type I collagen (right).