Use of tryptic peptide MALDI mass spectrometry imaging to identify the spatial proteomic landscape of colorectal cancer liver metastases.

*Celine Man Ying Li[1,2], *Matthew T Briggs[3], Yea-Rin Lee[3], Teresa Tin[2], Clifford Young[3], John Pierides[4], Gurjeet Kaur[5], Paul Drew[1,2], Guy J Maddern[1,2], Peter Hoffmann[3], Manuela Klingler-Hoffmann[3#] and Kevin Fenix[1,2#]

1. Discipline of Surgery, Adelaide Medical School, The University of Adelaide, Adelaide, SA 5005, Australia.
2. The Basil Hetzel Institute for Translational Health Research, The Queen Elizabeth Hospital, Adelaide, SA 5011, Australia.
3. Clinical and Health Sciences, University of South Australia, Adelaide, South Australia, 5000, Australia.
4. SA Pathology, Royal Adelaide Hospital, Adelaide, SA, 5000, Australia
5. Institute for Research in Molecular Medicine, University Sains Malaysia, Pulau Pinang, 11800, Malaysia

(*) equal contribution.

#Correspondence: manuela.klingler-hoffmann@unisa.edu.au and kevin.fenix@adelaide.edu.au

Keywords: Colorectal cancer liver metastasis, MALDI-MSI, LC-MS/MS, biomarkers, drug targets, colorectal cancer

Abstract

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths worldwide. CRC liver metastases (CRLM) are often resistant to conventional treatments, with high rates of recurrence. Therefore, it is crucial to identify biomarkers for CRLM patients that predict cancer progression. This study utilised matrix-assisted laser desorption/ionisation mass spectrometry imaging (MALDI-MSI) in combination with liquid chromatography-tandem mass spectrometry (LC-MS/MS) to spatially map the CRLM tumour proteome. CRLM tissue microarrays (TMAs) of 84 patients were analysed using tryptic peptide MALDI-MSI to spatially monitor peptide abundances across CRLM tissues. Abundance of peptides was compared between tumour vs stroma, male vs female and across three groups of patients based on overall survival (0-3 years, 4-6 years, and 7+ years). Peptides were then characterised and matched using LC-MS/MS. A total of 471 potential peptides were identified by MALDI-MSI. Our results show that two unidentified *m/z* values (1589.876 and 1092.727) had significantly higher intensities in tumours compared to stroma. Ten *m/z* values were identified to have correlation with biological sex. Survival analysis identified three peptides (Histone H4, Hemoglobin subunit alpha, and Inosine-5'-monophosphate dehydrogenase 2) and two unidentified *m/z* values (1305.840 and 1661.060) that were significantly higher in patients with shorter survival (0-3 years relative to 4-6 years and 7+ years). This is the first study using MALDI-MSI, combined with LC-MS/MS, on a large cohort of CRLM patients to identify the spatial proteome in this malignancy. Further, we identify several protein candidates that may be suitable for drug targeting or for future prognostic biomarker development.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related deaths worldwide [1]. Approximately 50% of CRC cases metastasise to the liver [2]. CRC liver metastasis (CRLM) is normally treated with surgical resection and chemotherapy. However, most CRLM cases present with multiple liver metastases or have patients with poor health conditions that prevents surgery [3]. Most CRLMs ultimately develop chemo-resistance [4]. Even with curative intent, CRLM has a relapse rate of 50% [5]. Together these factors contribute to CRLM having a 5-year survival rate of 30% [2]. Currently, there are no predictive biomarkers for CRLM recurrence or survival. CRLM surveillance currently involves frequent computed tomographic (CT) scans and monitoring for blood carcinoembryonic antigen (CEA) levels [6, 7]. However, these surveillance strategies have failed to detect early relapse, contributing to disease progression. Thus, there is an urgent need for predictive markers for CRLM survival [5].

Quantitative image analysis for tissue histological biomarkers based on routine histochemistry or immunohistochemistry has been used for clinical diagnostics and biomarker research [8]. In cancer research, recent advances in mass spectrometry have made proteomics a useful approach for biomarker discovery [9, 10]. In particular, matrix-assisted laser desorption/ionisation (MALDI) mass spectrometry imaging (MSI) generates unbiased spatial intensity maps of molecules such as intact proteins, tryptic peptides, *N*-glycans, lipids or metabolites, which allows for the spatial mapping of tumour tissues. MALDI-MSI has been demonstrated to predict metastatic potential, disease recurrence, and survival in variety of primary cancers [11-13].

In the context of CRC, MALDI-MSI has been extensively applied to primary tumours and has identified prognostic indicators for overall survival [14-16]. For example, multiple analytes have been shown to be associated with clinical outcomes such as tumour stage, grade, metastasis, and cancer cell proliferation in primary CRC. However, there are a limited number of MALDI-MSI studies on CRLM. To the best of our knowledge, only two reports have applied MALDI-MSI to detect lipid [17] and phospholipid [18] signatures on CRLM tissue. These studies identified tumour localised lipid signatures that could have prognostic potential or can be therapeutic targets. Whereas two additional MALDI-MSI studies identified the spatial proteome with limited sample sizes preventing any correlation with survival [19, 20].

In this study, we applied tryptic peptide MALDI-MSI on a cohort of Australian CRLM patients and identified peptides that discriminate between tumour features, patient characteristics, and survival outcomes. Subsequently, liquid chromatography-tandem mass spectrometry (LC-MS/MS) was performed on tryptic peptides obtained from consecutive tissue sections to identify the associated proteins [21]. We identified discriminative proteins related to tumour features, and potential prognostic biomarkers that were associated with worse clinical outcomes in CRLM patients.

Material and Methods

Sample Collection and Tissue Specimens

A retrospective cohort of CRC patients with liver metastases were identified using the South Australian Metastatic Colorectal Cancer Registry (SAmCRC) [22]. From that cohort, patients who had undergone liver resection were selected for the study. Their corresponding FFPE tissue blocks were retrieved from SA Pathology sites based at The Queen Elizabeth Hospital, Royal Adelaide Hospital and Flinders Medical Centre, Adelaide, South Australia. Retrieved cases were reviewed by a clinical pathologist (J.P) for study suitability. Demographics and clinicopathological characteristics of patients were shown in Table 1. This study was approved by the Human Research Ethics Committee of the Central Adelaide Local Health Network under protocol number 12237.

Tissue Microarray Construction

Previously selected tissue blocks were annotated by a pathologist (J.P), clearly distinguishing regions with viable tumours from normal liver. Two tumour cores and one normal liver core per patient were inserted into the TMA recipient block. Four TMA blocks were analysed, totalling 168 CRLM tumour cores from 84 patients. TMAs were constructed using a TMA Master II (3DHISTECH, Budapest, Hungary). The completed block was sealed with paraffin and stored at 4˚C.

Hematoxylin and Eosin (H&E) Staining

FFPE tissue blocks were cut by microtome into 6 µm sections. They were first deparaffinized and rehydrated with xylene for 3 x 5 mins, dipped in ethanol (100%) for 3 x 4 mins, ethanol (90%, *v/v*) for 2 mins, ethanol (70%, *v/v*) for 2 mins, and washed with distilled water for 1 min. Subsequently, slides were stained with hematoxylin for 5 mins, then washed with running tap water for 5 mins. Slides were then dipped into 0.3% hydrochloric acid with ethanol (75%, *v/v*) for 10 seconds, followed by washing with tap water for 10 mins. Tissues were stained with eosin for 1 min and washed with ethanol (100%) for 1 min. The H&E-stained TMA slides were scanned by a NanoZoomer 1 Digital Slide Scanner (Hamamatsu Photonics K.K, Japan). Images were acquired and viewed using NDP.view2 software (Hamamatsu).

Tryptic Peptide MALDI-MSI Sample Preparation

Tryptic peptide MALDI-MSI was performed on the sectioned TMAs following the protocol as previously described, with minor modifications [23]. Briefly, 6 µm sectioned TMAs on indium tin oxide (ITO) slides were washed with RCL premium grade xylene (100%) for 2 x 5 min, HPLC grade ethanol (100%) for 2 x 2 min, and 10 mM ammonium bicarbonate for 2 x 5 min. Slides were then boiled with 10 mM citric acid (pH 6) in a conventional microwave (1250 W, Model MS2540SR, LG, China), followed by pulse-heating for 10 min and heating at 98 °C on a heating block for 30 min. Slides were cooled down at room temperature prior to washing with 10 mM ammonium bicarbonate for 2 x 1 min. Subsequently, slides were dried at ambient conditions,

marked with a water-based white-out, and scanned using a flattop scanner (Epson Perfection V600 Photo, Model J25A, Epson, Indonesia) to teach the instrument during MALDI-MSI data acquisition.

Next, 20 µg of trypsin gold (Promega) in 200 µL of 25 mM ammonium bicarbonate and acetonitrile (10%, *v/v*) was deposited onto slides using an ImagePrep instrument (Bruker Daltonic, Germany) in 30 cycles with a fix nebulization time of 1.2 sec and drying time of 15 sec followed by incubation at 37 °C for 2 h in a humidified chamber containing potassium sulphate (53.3 g) and MilliQ water (18.2 g). Next, 7 mg/mL of alpha-cyano-4-hydroxycinnamic acid (HCCA) in acetonitrile (50%, *v/v*) and trifluoroacetic acid (TFA, 0.2%, *v/v*) was deposited onto slides using an iMatrix Spray instrument (Tardo, Switzerland). Instrument-specific settings were as follows: 60 mm height, 1 mm line distance, 180 mm/s speed, 1 µL/cm^2 density, 30 cycles and 15 sec delays.

MALDI-MSI Data Acquisition

MALDI-MSI data was acquired on a timsTOF fleX MS instrument (Bruker Daltonic, Germany), controlled by timsControl (v3.0, Bruker Daltonics) and flexImaging (v7.0, Bruker Daltonics) in positive mode in the mass range of *m/z* 700-3200. The instrument was calibrated using an internal calibrant [24]. Instrument-specific settings were as follows: a transfer time of 180 µs, a pre-pulse storage of 25 µs, a collision RF of 4000 Vpp, a collision energy of 10 eV, a funnel 1 (accumulation) RF of 500 Vpp, a funnel 2 (analysis) RF of 500 Vpp, a multipole RF of 500 Vpp and a quadrupole ion energy of 5 eV. Laser power was set at 60% and pixel resolution at 50µm.

MALDI-MSI Data Analysis

MALDI-MSI data was imported into SCiLS Lab v2023a (Bruker Daltonics, Germany) and pre-processed by TopHat baseline subtraction and root mean square (RMS) normalisation. Individual TMA cores were annotated based on pathological annotations by a pathologist (G.K.) from consecutive H&E-stained images scanned with NanoZoomer (Hamamatsu). The feature finding tools were then used to create a list of putative peptides. The feature finding uses the T-Rex^2 algorithm, and very strong filtering with 5% coverage and 0.5% relative intensity threshold were applied. For comparisons of peak intensities between the groups, the RMS normalised average spectra of each mass was exported as Excel files and statistical analysis performed using GraphPad Prism (v 9.0, San Diego, CA).

LC-MS/MS Sample Preparation

CRLM TMAs were sectioned using a microtome (Microm HM 325) at 6 µm thickness and placed directly into 1.5 mL Eppendorf Protein LoBind Tubes. Paraffin was removed by the addition of 100 µL of xylene (100%) for 2 min and 100 µL of ethanol (100%) for 2 min. Sections were then rehydrated by the addition of 100 µL of Milli-Q water (100%) for 2 x 2 min and 100 µL of 50 mM ammonium bicarbonate (100%) for 2 x 2 min. Then the tissue was boiled in 30 µL of 10 mM citric acid at pH 6 at 98 °C for 20 min in a Thermomixer (1500 rpm), followed by 80 °C for 2 h. Tubes were cooled to room temperature and then centrifuged at 20,000 x g and 4 °C

for 30 min. Proteins were denatured from the tissue by adding 30 µL of 8 M urea, followed by incubation at room temperature for 30 min. Subsequently, samples were analysed using a modified tryptophan assay to determine protein concentrations [25].

Protein reduction was by incubation with 2.5 µL of 0.2 M dithiothreitol (DTT) at room temperature for 1 hr on a Thermomixer (500 rpm). Proteins were then alkylated by the addition of 2.8 µL of 0.275 M 2-chloroacetamide (CA) and incubation in the dark for 30 min on a shaking Thermomixer. Protein digestion was performed with sequencing grade trypsin (Promega, USA) (1:50) at 37 °C overnight on a shaking Thermomixer, with digestion stopped by the addition of 10% formic acid (5 µL). Finally, tryptic peptides were desalted using C18 ZipTips (Millipore, Ireland) according to the manufacturer's instructions.

LC-MS/MS Data Acquisition

LC-MS/MS analysis was conducted on an EASY-nLC 1200 system coupled to an Orbitrap Exploris 480 mass spectrometer (Thermo Scientific, Bremen, Germany). Peptides (approximately 500 ng) were resuspended in 0.1% formic acid and loaded onto a 25 cm fused silica column (75 µm inner diameter, 360 µm outer diameter) heated to 50 °C. The column was packed with 1.9 µm ReproSil-Pur 120 C18-AQ particles (Dr. Maisch, Ammerbuch, Germany). Peptides were separated over a 30 min linear gradient (3 to 24% acetonitrile in 0.1% formic acid) at a flow rate of 300 nL/min. A compensation voltage of -45 V was applied from a FAIMS Pro interface (Thermo Scientific) to minimise the impact of interfering ions before implementing a data-independent acquisition (DIA) MS method. Briefly, an MS scan (*m/z* 395 to 905) was acquired at resolution 30000 (*m/z* 200) in positive ion mode before isolated precursors (50 x 10 *m/z* windows, 1 *m/z* overlap) were fragmented with higher energy collision dissociation (27.5% normalised collision energy). The resulting MS/MS spectra (*m/z* 145 to 1450) were measured at resolution 22500.

LC-MS/MS Data Analysis

Raw data files were processed with Spectronaut v17.4.230317 (Biognosys AG, Schlieren, Switzerland). A direct DIA+ analysis was conducted by searching a UniProt human proteome database (Release 2022_04, 20607 canonical entries) with BGS factory settings. We manually matched single charged *m/z* values obtained from the MALDI-MSI data with singly charged *m/z* values calculated from doubly or triply charged peptides identified using Spectronaut analysis peptides with a peptide tolerance of +/- 16 ppm.

Results

Identification of tryptic peptide signatures in CRLM samples using MALDI-MSI.

To identify tryptic peptides associated with CRLM, we first constructed four TMAs from 84 CRLM patients consisting of 168 CRLM tumour tissue cores (two per patient, Table 1). The TMAs were sectioned onto conductive ITO slides and prepared for MALDI-MSI analysis. A comprehensive and detailed MALDI-MSI

workflow is presented in Fig. 1A. For the analysis of the processed MALDI-MSI data, tumour cores were annotated as tumour (cancer cell rich) and intratumoral stroma (cancer cell poor/free) regions in SCiLS Lab by matching pathologist annotated H&E regions overlaid with the MALDI-MSI optical image (Fig. 2). Subsequently, the feature finding tool in SCiLS Lab was used to generate a list of 471 putative peptides by applying very strong filtering with 5% coverage and 0.5% relative intensity threshold. Next, the receiver operating curve (ROC) discriminating feature tool was used to identify *m/z* values with the highest area under the curve (AUC) when comparing between tumour and stroma, male and female patients and across three groups of patients' overall survival (0-3 years, 4-6 years, and 7+ years). Furthermore, denatured proteins from consecutive tissue sections were then digested in-solution into tryptic peptides for LC-MS/MS analysis to identify proteins by matching peptide sequences with measured *m/z* values by MALDI-MSI (Fig. 1B). To achieve an extensive list of peptides identified by LC-MS/MS, the raw data was analysed in Spectronaut and resulted in the identification of 3332 proteins from CRLM samples.

Identification of tryptic peptides associated with CRLM tumour regions.

First, to identify tryptic peptides that significantly distinguished between tumour and stroma, we selected MALDI-MSI *m/z* values with the highest AUC values to discriminate the two regions. From this analysis, 20 *m/z* values had AUC values of >0.5 which were then investigated further if they have significant abundances in the CRLM tissues (Table 2). The 20 *m/z* values were observed between a *m/z* range of 1000 to 2300 (Fig. 3A-B). Subsequently, a ROC plot with associated AUC values for each peptide was generated to plot the specificity and sensitivity in relation to tumour and stroma. Of the 20 *m/z* values, *m/z* 1589.876 with 0.618 AUC (95% CI: 0.563-0.674), and *m/z* 1092.727 with 0.562 AUC (95% CI: 0.504-0.620) were found to be high in intensity in tumour regions relative to stroma regions. Unfortunately, *m/z* 1589.876 and *m/z* 1092.727 could not be identified from the LC-MS/MS data (Fig. 3C-F). The remaining 18 *m/z* values with AUC values >0.5 were not statistically significant (Table 2 and Supp. Fig 1).

Identification of tryptic peptides associated with biological sex.

Next, we determined if there were differences in the spatial proteome between male and female CRLM tumour tissues. Selecting, MALDI-MSI peaks with an AUC of >0.5 resulted in the identification of 14 *m/z* values. From this analysis, 8 *m/z* values (*m/z* 2211.113, 1296.682, 1570.676, 1669.829, 1461.702, 1465.697, 1235.618, and 1366.628) were found to be significantly different in the tumour region between males and females with AUC values between 0.600-0.695 (Table 3 and Supp. Fig. 2). Interestingly, we identified 4 (*m/z* 2211.113, 1296.682, 1570.676, 1366.628) to be higher in intensity in females (Fig 4A-D), whereas another 4 (*m/z* 1669.829, 1461.702, 1465.697, 1235.618) were higher in intensity in males (Fig. 4E-H). In the stroma (Supp. Fig. 3), 8 (*m/z* 2211.113, 1296.682, 1570.676, 1669.829, 1465.697, 1366.628, 1661.060 and 1305.840) were identified to be statistically different between male and females, with 6 (*m/z* 2211.113, 1296.682, 1570.676, 1366.628, 1661.060 and 1305.840) higher in females (Fig 5A-F) compared to 2 (*m/z* 1669.829 and 1465.697) in males (Fig 5G-H). The AUC values ranged from 0.633-0.677.

In total 10 unique *m/z* values were identified to discriminate between male and female tissues (Table 3). Six of the 10 *m/z* values were associated with biological sex regardless of the region type (tumour or stroma). Using LC-MS/MS, we identified 3 peptides, [FTTDAIALAMSR] (*m/z* 1296.682) from Proteasome subunit beta type-9 (PSMB9), [SDLVNEEATGQFR] (*m/z* 1465.697) from Carcinoembryonic antigen-related cell adhesion molecule 5 (CEACAM5), and [YHLGAYTGDDVR] (*m/z* 1366.628) from Tryptase beta-2 (TPSB2) or Tryptase alpha/beta-1 (TPSAB1), while the corresponding peptides for *m/z* 2211.113, *m/z* 1570.676 and *m/z* 1669.829 could not be identified. Further, an unidentified *m/z* value (*m/z* 1461.702) and [MVNHFIAEFK] (*m/z* 1235.618) from Heat shock cognate 71 kDa protein (HSPA8) were higher in male tumour regions, while two unidentified *m/z* values (1661.060 and 1305.840) were higher in female stroma regions. Together, these results indicated that biological sex could contribute to changes to both the overall and region-specific tumour proteome.

Identification tryptic peptides linked to survival.

To identify potential prognostic markers for CRLM, we classified our patient cohort into 3 groups based on survival (0-3 years, 4-6 years, and 7+ years). Specific to the tumour region (Supp. Fig 4), we identified 6 *m/z* values that had an AUC >0.5 between 0-3 and 7+ years survival. From this list, 5 *m/z* values (*m/z* 1325.754, 1529.738, 1481.857, 1305.840 and 1661.060) achieved statistical significance (Table 4). [DNIQGITKPAIR] (*m/z* 1325.754) from histone H4 had the highest AUC value of 0.720 (95% CI: 0.6320-0.8086) (Fig. 6A). Additionally, [VGAHAGEYGAEALER] (*m/z* 1529.738) from Haemoglobin subunit alpha (HBA1), [REDLVVAPAGITLK] (*m/z* 1481.857) from Inosine-5'-monophosphate dehydrogenase 2 (IMPDH2), and 2 unidentified *m/z* values (*m/z* 1305.840 and 1661.060) were all relatively high abundant in patients who had poor survival outcomes (0-3 years post-surgery) compared to those who have survived for 7 or more years (Fig. 6). Analysing the stromal regions (Supp. Fig. 5), only histone H4 replicated the tumour analysis showing higher abundance in patients with poor (0-3 years) survival outcomes (Supp. Fig. 6). Together, our MALDI-MSI analysis identified peptides that associated with duration of survival in CRLM patients.

Discussion

CRLM is a highly lethal and heterogeneous disease with poor prognostic tests available [26, 27]. While previous studies have reported on the use of MALDI-MSI in CRLM tissue, most have used small sample sizes to only describe the proteomic landscape of CRLM [19, 20]. To the best of our knowledge, this study is the first study to utilise MALDI-MSI combined with LC-MS/MS to identify potential protein biomarkers in CRLM patients. Furthermore, we applied this technology on the largest cohort of CRLM patients collected in Australia.

The tumour-stroma ratio is a well-known independent prognostic indicator for CRC, with high levels of stroma associating with poor survival [28, 29]. Thus, we determined if there are consensus proteins that could discriminate between the two regions in our heterogenous CRLM sample cohort by MALDI-MSI. MALDI-MSI analysis between tumour and intratumoral stroma regions, identified peptides that associate with tumour regions. Among the 20 *m/z* values identified using MALDI-MSI, 2 *(m/z* 1589.876 and 1092.727) reached statistical

significance, however it was not possible to establish matches with our LC-MS/MS data. Nevertheless, our MALDI-MSI data was able to find features that distinguish tumours from stroma.

While it is widely appreciated that biological sex can greatly affect cancer biology and treatment outcomes [30, 31], proteomic analyses that describe such contributions, especially in CRLM, are rarely reported. In this study, we investigated if biological sex was associated with differences in the spatial proteome of CRLM. We identified PSMB9, CEACAM5 and TPSB2/TPSAB1 and 3 unidentified *m/z* values (*m/z* 2211.113, 1570.676 and 1669.829) that had significant differences in abundance between male and female CRLM tissues. These were sex-specific for both tumour and stroma regions. Interestingly, HSPA8 and an unidentified *m/z* value (*m/z* 1461.702) localised to tumours specific to males, and 2 unidentified *m/z* values (*m/z* 1305.840 and 1661.060) localised to stroma, that were specific to females. Together these results show for the first time that biological sex may have an underappreciated role in the spatial proteome of CRLM that may have currently unknown consequences to its pathophysiology. Further research is required to determine if expression of these proteins affect sex specific CRLM progression and treatment outcomes.

We report that histone H4, HBA1, IMPDH2, and two unidentified *m/z* values (*m/z* 1305.840 and 1661.060) are more abundant in tumour tissues of CRLM patients with poor survival. In this study, only histone H4 expression in tumour and/or stroma was enriched in patients that survived for >3 years. Histones are key epigenetic regulators. Their post-translational modifications (acetylation, methylation and phosphorylation) can be linked with gene expression changes that lead to cancer progression including CRC and hence are being investigated for their therapeutic and prognostic potential [32]. Histone H4 upregulation has been linked to platinum-based chemotherapy resistance in malignancies [33] and may explain its association with poor prognosis in our patient cohort. Interestingly, histone H4 was identified to be elevated in CRC patient plasma by mass spectrometry (detecting the [DNIQGITKPAI] peptide sequence) [34], and ELISA [35]. Thus, histone H4 could potentially be developed as a blood biomarker to identify CRLM patients with worse prognostic outcomes. The relationship between histone H4 tumour tissue expression, plasma levels, and survival outcomes is unknown and warrants further investigation. HBA1 is the alpha subunit of hemoglobulin. Anaemia, measured by blood hemoglobulin levels, has been linked with tumour hypoxia and poor outcomes in solid tumours [36]. However, the specific contribution of HBA1 in cancer progression is unknown. Recent single cell RNA-sequencing analysis of gastric cancers shows that HBA1 is overexpressed in gastric cancer cells. Importantly, patients with high tumour HBA1 expression led to poor overall survival supporting our findings in CRLM [37]. The contributions of IMPDH2 in CRC progression is well established. IMPDH2 is an isoform of IMPDH, an enzyme critical for biosynthesis of purine nucleotides and is essential for DNA synthesis [38]. IMPDH2 is commonly upregulated in malignancies [39] including primary CRC [40]. IMPDH2 has been shown to promote CRC tumorigenesis [41], metastatic potential [42] and in methotrexate resistance [43]. Thus, IMPDH2 has been previously implied as a potential prognostic marker for CRC outcomes and may be useful as a drug target. Collectively, our study identified previously reported and novel tryptic peptides candidates that may be able to predict poor outcomes in CRLM patients. Considering the reported function of these proteins in cancer, these may also be used as novel targets in CRLM therapy development.

In this study we used a manual indirect identification approach, where we matched the MALDI-MSI data with data from an independent LC-MS/MS experiment. As part of this workflow, we have implemented internal calibrants to allow high accuracy peptide matching [24]. However, there are limitations in this study to consider, including the identification of peptides with the similar *m/z* values which fit within the same MALDI-MSI mass tolerance window, therefore potentially resulting in mis-assignments. Additional validation by *in situ* tandem MS fragmentation to confirm the identity of the peptides may improve peptide identification [44]. However, this approach is only feasible for peptides with high signal intensity. Recently, several automated annotation tools have been developed, which even include post-translational modifications [24, 45]. These tools might be used to identify more of the 471 peptides from the LC-MS/MS data. Moreover, additional methods such as immunohistochemistry could be employed to validate the abundance of the identified proteins in the metastatic tumour regions. A second independent cohort of patients is required to validate our findings and confirm the prognostic potential of the identified proteins that associated with poor CRLM survival outcomes. Nonetheless, this study provides novel and valuable insights into CRLM biology and identifies potential prognostic biomarkers and targets that are needed in CRLM.

Conclusions

In summary, this is the first study to utilise MALDI-MSI with LC-MS/MS to identify potential spatial proteomic tissue biomarkers in the largest cohort of Australian CRLM patients to date. Using these workflows, spatial proteomic features present in CRLM were revealed and changes within the tumour proteome identified in relation to clinical features, including sex and overall survival. Importantly, we reveal that biological sex could impact the spatial proteome profile of CRLM tumours and identified several proteins that associate with poor survival in CRLM patients. These findings should be further explored to identify new prognostic markers and therapeutic targets that are highly needed in CRLM.

References

[1] E. Morgan *et al.*, "Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN," *Gut*, vol. 72, no. 2, pp. 338-344, 2023, doi: 10.1136/gutjnl-2022-327736.

[2] H. Zhou *et al.*, "Colorectal liver metastasis: molecular mechanism and interventional therapy," *Signal Transduction and Targeted Therapy*, vol. 7, no. 1, p. 70, 2022/03/04 2022, doi: 10.1038/s41392-022-00922-2.

[3] Y. Dong and T. Gruenberger, "Surgical management of colorectal liver metastases—a practical clinical approach," *European Surgery*, vol. 55, no. 3, pp. 94-99, 2023/06/01 2023, doi: 10.1007/s10353-023-00796-w.

[4] H. Zhou *et al.*, "Colorectal liver metastasis: molecular mechanism and interventional therapy," (in eng), *Signal Transduct Target Ther*, vol. 7, no. 1, p. 70, Mar 4 2022, doi: 10.1038/s41392-022-00922-2.

[5] T. Reinert *et al.*, "Circulating tumor DNA for prognosis assessment and postoperative management after curative-intent resection of colorectal liver metastases," (in eng), *International journal of cancer*, vol. 150, no. 9, pp. 1537-1548, May 1 2022, doi: 10.1002/ijc.33924.

[6] J. H. Kang and S. H. Choi, "Imaging study for colorectal liver metastasis: beyond the diagnosis and to the prognosis," (in eng), *Hepatobiliary Surg Nutr*, vol. 8, no. 6, pp. 666-668, Dec 2019, doi: 10.21037/hbsn.2019.10.06.

[7] K. Sasaki *et al.*, "Pre-hepatectomy carcinoembryonic antigen (CEA) levels among patients undergoing resection of colorectal liver metastases: do CEA levels still have prognostic implications?," (in eng), *HPB : the official journal of the International Hepato Pancreato Biliary Association*, vol. 18, no. 12, pp. 1000-1009, Dec 2016, doi: 10.1016/j.hpb.2016.09.004.

[8] H. Lee *et al.*, "Quantitative Proteomic Analysis Identifies AHNAK (Neuroblast Differentiation-associated Protein AHNAK) as a Novel Candidate Biomarker for Bladder Urothelial Carcinoma Diagnosis by Liquid-based Cytology," (in eng), *Mol Cell Proteomics*, vol. 17, no. 9, pp. 1788-1802, Sep 2018, doi: 10.1074/mcp.RA118.000562.

[9] M. A. Reymond and W. Schlegel, "Proteomics in cancer," (in eng), *Adv Clin Chem*, vol. 44, pp. 103-42, 2007, doi: 10.1016/s0065-2423(07)44004-5.

[10] Y. W. Kwon *et al.*, "Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery," (in English), *Frontiers in Medicine*, Review vol. 8, 2021-September-22 2021, doi: 10.3389/fmed.2021.747333.

[11] K. L. Sheng, L. Kang, K. J. Pridham, L. E. Dunkenberger, Z. Sheng, and R. T. Varghese, "An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression," *Scientific reports*, vol. 10, no. 1, p. 21246, 2020/12/04 2020, doi: 10.1038/s41598-020-78126-3.

[12] S. Meding *et al.*, "Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer," (in eng), *J Pathol*, vol. 228, no. 4, pp. 459-70, Dec 2012, doi: 10.1002/path.4021.

[13] A. C. Hristov *et al.*, "HMGA2 protein expression correlates with lymph node metastasis and increased tumor grade in pancreatic ductal adenocarcinoma," *Modern Pathology*, vol. 22, no. 1, pp. 43-49, 2009/01/01 2009, doi: 10.1038/modpathol.2008.140.

[14] T. Gemoll, S. Strohkamp, K. Schillo, C. Thorns, and J. K. Habermann, "MALDI-imaging reveals thymosin beta-4 as an independent prognostic marker for colorectal cancer," *Oncotarget*, vol. 6, no. 41, 2015. [Online]. Available: https://www.oncotarget.com/article/6103/text/.

[15] A. Hinsch *et al.*, "MALDI imaging mass spectrometry reveals multiple clinically relevant masses in colorectal cancer using large-scale tissue microarrays," *Journal of Mass Spectrometry*, vol. 52, no. 3, pp. 165-173, 2017, doi: https://doi.org/10.1002/jms.3916.

[16] B. Martin *et al.*, "A Mass Spectrometry Imaging Based Approach for Prognosis Prediction in UICC Stage I/II Colon Cancer," (in eng), *Cancers*, vol. 13, no. 21, Oct 26 2021, doi: 10.3390/cancers13215371.

[17] N. H. Patterson *et al.*, "Assessment of pathological response to therapy using lipid mass spectrometry imaging," *Scientific reports*, vol. 6, no. 1, p. 36814, 2016/11/14 2016, doi: 10.1038/srep36814.

[18] S. Shimma, Y. Sugiura, T. Hayasaka, Y. Hoshikawa, T. Noda, and M. Setou, "MALDI-based imaging mass spectrometry revealed abnormal distribution of phospholipids in colon cancer liver metastasis," *Journal of Chromatography B*, vol. 855, no. 1, pp. 98-103, 2007/08/01/ 2007, doi: https://doi.org/10.1016/j.jchromb.2007.02.037.

[19] L. Moritz *et al.*, "Characterization of Spatial Heterogeneity in Metastasized Colorectal Cancer by MALDI Imaging," in *Preprints*, ed: Preprints, 2023.

[20] A. Turtoi *et al.*, "Organized proteomic heterogeneity in colorectal cancer liver metastases and implications for therapies," *Hepatology*, vol. 59, no. 3, pp. 924-934, 2014, doi: https://doi.org/10.1002/hep.26608.

[21] R. Casadonte *et al.*, "Imaging mass spectrometry to discriminate breast from pancreatic cancer metastasis in formalin-fixed paraffin-embedded tissues," (in eng), *Proteomics*, vol. 14, no. 7-8, pp. 956-64, Apr 2014, doi: 10.1002/pmic.201300430.

[22] Y. Tomita *et al.*, "Survival improvements associated with access to biological agents: Results from the South Australian (SA) metastatic colorectal cancer (mCRC) registry," *Acta Oncologica*, vol. 55, no. 4, pp. 480-485, 2016/04/02 2016, doi: 10.3109/0284186X.2015.1117135.

[23] T. W. Powers *et al.*, "MALDI imaging mass spectrometry profiling of N-glycans in formalin-fixed paraffin embedded clinical tissue blocks and tissue microarrays," (in eng), *PLoS One*, vol. 9, no. 9, p. e106255, 2014, doi: 10.1371/journal.pone.0106255.

[24] J. O. R. Gustafsson *et al.*, "Internal calibrants allow high accuracy peptide matching between MALDI imaging MS and LC-MS/MS," (in eng), *J Proteomics*, vol. 75, no. 16, pp. 5093-5105, Aug 30 2012, doi: 10.1016/j.jprot.2012.04.054.

[25] J. R. Wiśniewski and F. Z. Gaugaz, "Fast and Sensitive Total Protein and Peptide Assays for Proteomic Analysis," *Analytical Chemistry*, vol. 87, no. 8, pp. 4110-4116, 2015/04/21 2015, doi: 10.1021/ac504689z.

[26] J. Guinney *et al.*, "The consensus molecular subtypes of colorectal cancer," *Nature Medicine*, vol. 21, no. 11, pp. 1350-1356, 2015/11/01 2015, doi: 10.1038/nm.3967.

[27] B. A. Alves Martins, G. F. de Bulhões, I. N. Cavalcanti, M. M. Martins, P. G. de Oliveira, and A. M. A. Martins, "Biomarkers in Colorectal Cancer: The Role of Translational Proteomics Research," (in eng), *Front Oncol*, vol. 9, p. 1284, 2019, doi: 10.3389/fonc.2019.01284.

[28] G. W. van Pelt *et al.*, "The tumour-stroma ratio in colon cancer: the biological role and its prognostic impact," (in eng), *Histopathology*, vol. 73, no. 2, pp. 197-206, Aug 2018, doi: 10.1111/his.13489.

[29] K. Zhao *et al.*, "Artificial intelligence quantified tumour-stroma ratio is an independent predictor for overall survival in resectable colorectal cancer," (in eng), *EBioMedicine*, vol. 61, p. 103054, Nov 2020, doi: 10.1016/j.ebiom.2020.103054.

[30] I. Baraibar *et al.*, "Sex and gender perspectives in colorectal cancer," (in eng), *ESMO Open*, vol. 8, no. 2, p. 101204, Apr 2023, doi: 10.1016/j.esmoop.2023.101204.

[31] C. H. Li *et al.*, "Sex differences in oncogenic mutational processes," *Nature Communications*, vol. 11, no. 1, p. 4330, 2020/08/28 2020, doi: 10.1038/s41467-020-17359-2.

[32] X. An, X. Lan, Z. Feng, X. Li, and Q. Su, "Histone modification: Biomarkers and potential therapies in colorectal cancer," *Annals of Human Genetics*, vol. 87, no. 6, pp. 274-284, 2023, doi: https://doi.org/10.1111/ahg.12528.

[33] R. Wang *et al.*, "Histone H4 expression is cooperatively maintained by IKKβ and Akt1 which attenuates cisplatin-induced apoptosis through the DNA-PK/RIP1/IAPs signaling cascade," *Scientific reports*, vol. 7, no. 1, p. 41715, 2017/01/31 2017, doi: 10.1038/srep41715.

[34] P. Van den Ackerveken *et al.*, "A novel proteomics approach to epigenetic profiling of circulating nucleosomes," *Scientific reports*, vol. 11, no. 1, p. 7256, 2021/03/31 2021, doi: 10.1038/s41598-021-86630-3.

[35] E. ÖZGÜR, M. KESKIN, E. E. YÖRÜKER, S. HOLDENRIEDER, and U. GEZER, "Plasma Histone H4 and H4K20 Trimethylation Levels Differ Between Colon Cancer and

Precancerous Polyps," *In Vivo*, vol. 33, no. 5, pp. 1653-1658, 2019, doi: 10.21873/invivo.11651.

[36] S. Walrand, R. Lhommel, P. Goffette, M. Van den Eynde, S. Pauwels, and F. Jamar, "Hemoglobin level significantly impacts the tumor cell survival fraction in humans after internal radiotherapy," *EJNMMI Research*, vol. 2, no. 1, p. 20, 2012/05/19 2012, doi: 10.1186/2191-219X-2-20.

[37] W. Yu, G. Chen, J. Yan, X. Wang, Y. Zhu, and L. Zhu, "Single-cell sequencing analysis reveals gastric cancer microenvironment cells respond vastly different to oxidative stress," *Journal of translational medicine*, vol. 20, no. 1, p. 250, 2022/06/03 2022, doi: 10.1186/s12967-022-03411-w.

[38] E. C. Thomas *et al.*, "Different Characteristics and Nucleotide Binding Properties of Inosine Monophosphate Dehydrogenase (IMPDH) Isoforms," *PLOS ONE*, vol. 7, no. 12, p. e51096, 2012, doi: 10.1371/journal.pone.0051096.

[39] S. Kofuji and A. T. Sasaki, "GTP metabolic reprogramming by IMPDH2: unlocking cancer cells' fuelling mechanism," *The Journal of Biochemistry*, vol. 168, no. 4, pp. 319-328, 2020, doi: 10.1093/jb/mvaa085.

[40] Y. He *et al.*, "Identification of IMPDH2 as a tumor-associated antigen in colorectal cancer using immunoproteomics analysis," (in eng), *Int J Colorectal Dis*, vol. 24, no. 11, pp. 1271-9, Nov 2009, doi: 10.1007/s00384-009-0759-2.

[41] Q. Zhang *et al.*, "c-Myc–IMPDH1/2 axis promotes tumourigenesis by regulating GTP metabolic reprogramming," *Clinical and Translational Medicine*, vol. 13, no. 1, p. e1164, 2023, doi: https://doi.org/10.1002/ctm2.1164.

[42] S. Duan *et al.*, "IMPDH2 promotes colorectal cancer progression through activation of the PI3K/AKT/mTOR and PI3K/AKT/FOXO1 signaling pathways," *Journal of Experimental & Clinical Cancer Research*, vol. 37, no. 1, p. 304, 2018/12/05 2018, doi: 10.1186/s13046-018-0980-3.

[43] S. Peñuelas, V. Noé, and C. J. Ciudad, "Modulation of IMPDH2, survivin, topoisomerase I and vimentin increases sensitivity to methotrexate in HT29 human colon cancer cells," *The FEBS Journal*, vol. 272, no. 3, pp. 696-710, 2005, doi: https://doi.org/10.1111/j.1742-4658.2004.04504.x.

[44] L. Théron *et al.*, "A Proof of Concept to Bridge the Gap between Mass Spectrometry Imaging, Protein Identification and Relative Quantitation: MSI~LC-MS/MS-LF," (in eng), *Proteomes*, vol. 4, no. 4, Oct 26 2016, doi: 10.3390/proteomes4040032.

[45] G. Guo *et al.*, "Automated annotation and visualisation of high-resolution spatial proteomic mass spectrometry imaging data using HIT-MAP," (in eng), *Nat Commun*, vol. 12, no. 1, p. 3241, May 28 2021, doi: 10.1038/s41467-021-23461-w.

Statements and Declarations

Funding

This work was supported by a Cancer Council SA Beat Cancer Infrastructure grant with matched funding support from The Hospital Research Foundation Group (GM. and K.F.). K.F. was supported by a The Hospital Research Foundation Group Early Career Fellowship. C.L. was supported by a University of Adelaide Postgraduate Research Scholarship. The authors acknowledge Bioplatforms Australia, the University of South Australia, and the State and Federal Governments, which co-fund the NCRIS-enabled Mass Spectrometry and Proteomics facility at the University of South Australia.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

Conceptualisation C.L., M.B., M.K.H and K.F.; Collection and/or assembly of data: C.L., M.B., Y.L., T.T., C.Y., J.P., and G.K. Data Analysis and Interpretation: C.L., M.B., Y.L., C.Y., M.K.H, and K.F., resources: J.P., G.M., P.H, and K.F.; writing—original draft preparation, C.L., M.B., P.D., M.K.H and K.F.; writing—review and editing, C.L., M.B., C.Y., P.D., M.K.H. and K.F.; supervision, P.H., G.M., M.K.H, and K.F., funding acquisition, C.L., G.M., M.K.H and K.F.; All authors have read and agreed to the published version of the manuscript.

Data Availability

The data underlying this article are available in this published article and its supplementary information files. Additional data are available upon request to the corresponding author.

Ethics Approval

This study was approved by the Human Research Ethics Committee of the Central Adelaide Local Health Network under protocol number 12237.

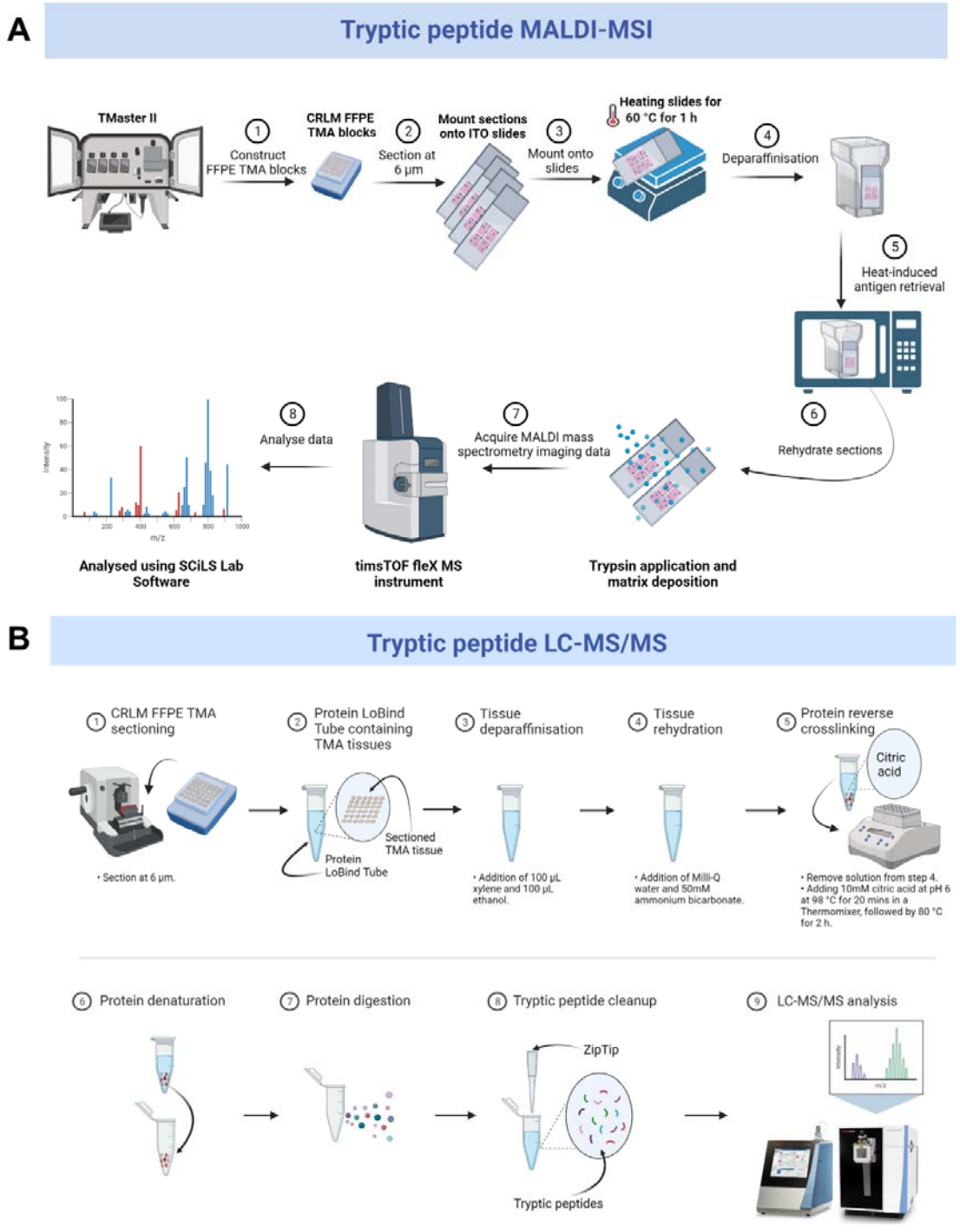
Consent to Participate

Samples collected were retrospective patient tissues and was not consented as advised by our Human Research Ethics Committee.

548

549

550 **Figures**



551

552 **Figure 1.** Overview of (A) MALDI-MSI and (B) LC-MS/MS workflows for spatial mapping and
553 characterisation of tryptic peptides from CRLM FFPE tissues (n=84). Created with BioRender.com.

554

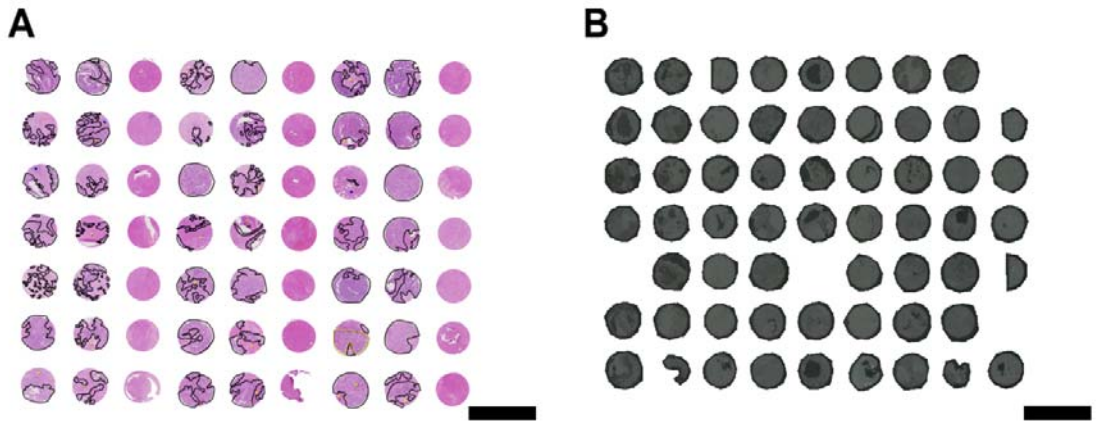


Figure 2. Representative scanned images of (A) H&E CRLM TMA cores, and (B) optical CLRM TMA cores. CRLM tumour regions and CRC tumour regions (control) in (A) were annotated by a pathologist, as shown in black and green, respectively. Scale bars (black lines) represent 3mm.

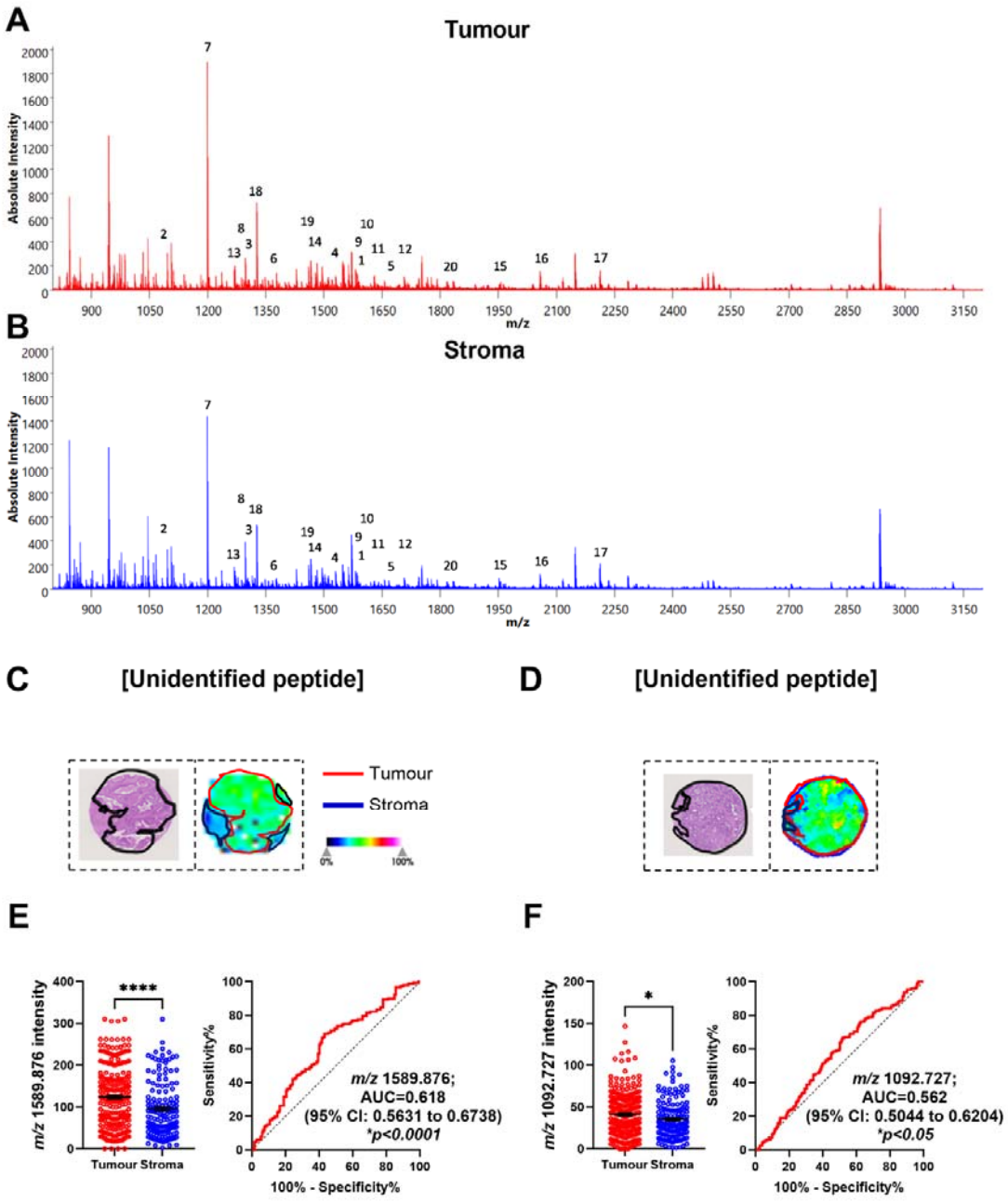


Figure 3. Identification of tryptic peptides associated with CRLM tumour regions using MALDI-MSI and LC-MS/MS. (A) Sum spectra of CRLM tumour (red) and (B) stroma (blue) in CRLM tumour cores. (C, D) H&E images (left) of a representative tumour core with tumour regions annotated in black and ion intensity maps (right) of most discriminative *m/z* values between tumour (red) and stroma (red). Tumour cores are 1.5mm in diameter. Bar graphs showing mean ± SEM and ROC plots for (E) *m/z* 1589.876 and (F) *m/z* 1092.727. **p<0.05, ****p<0.0001.* Student's t test. Each dot represents a single tumour core.

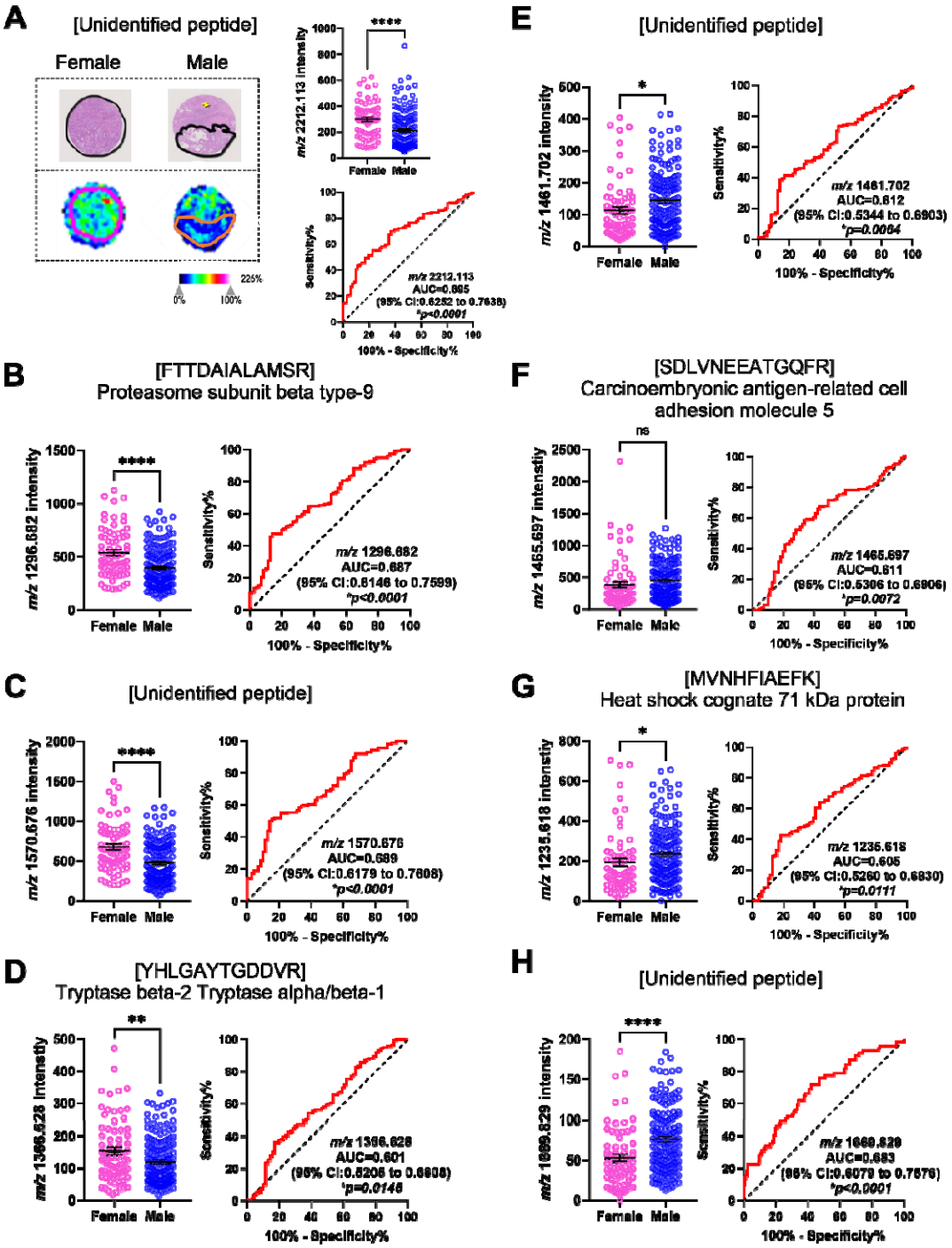


Figure 4. Significantly abundant tryptic peptides associated with male and female CRLM tumour regions, by MALDI-MSI and LC-MS/MS. (A) Representative H&E and ion intensity map of tumour region annotated CRLM tumour cores for *m/z* 2211.113. Tumour cores are 1.5mm in diameter. (A-D) Tryptic peptides (*m/z* 2211.113, 1296.682, 1570.676, and 1366.628) significantly more abundant in females relative to males. Tryptic peptides were identified as (A) unidentified, (B) Proteasome subunit beta type-9, (C) unidentified, and (D) Tryptase beta-2 Tryptase alpha/beta-1. (E-H) Tryptic peptides (*m/z* 1669.829, 1461.702, 1466.697, and 1235.618) significantly more abundant in males relative to females. These tryptic peptides were identified as (E) unidentified, (F) Carcinoembryonic antigen-related cell adhesion molecule 5, (G) Heat shock cognate 71 kDa protein, and (H) unidentified. Bar graphs showing mean ± SEM and ROC plots are presented. ****$p<0.0001$. Student's t test. Each dot represents a single tumour core.

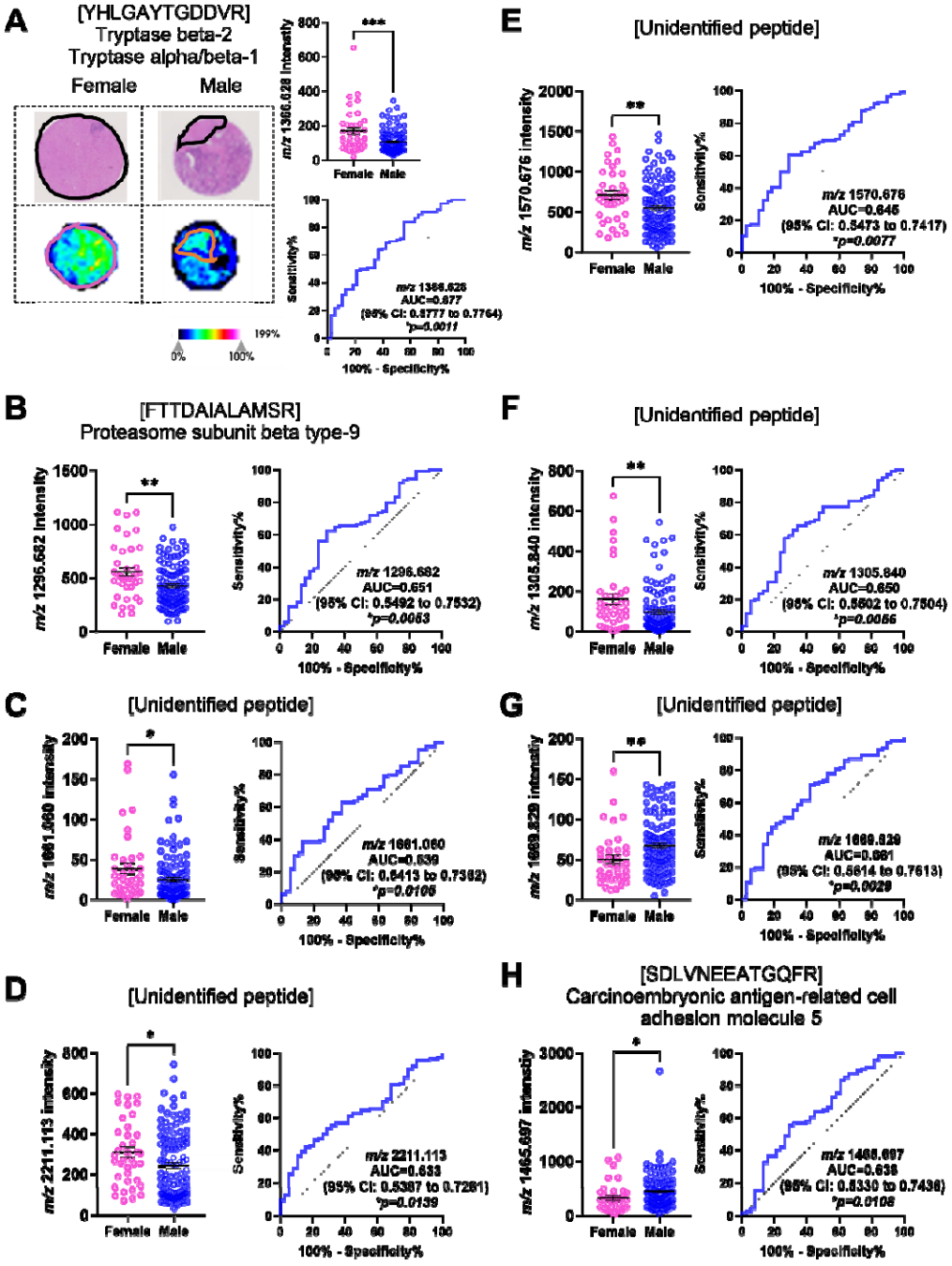
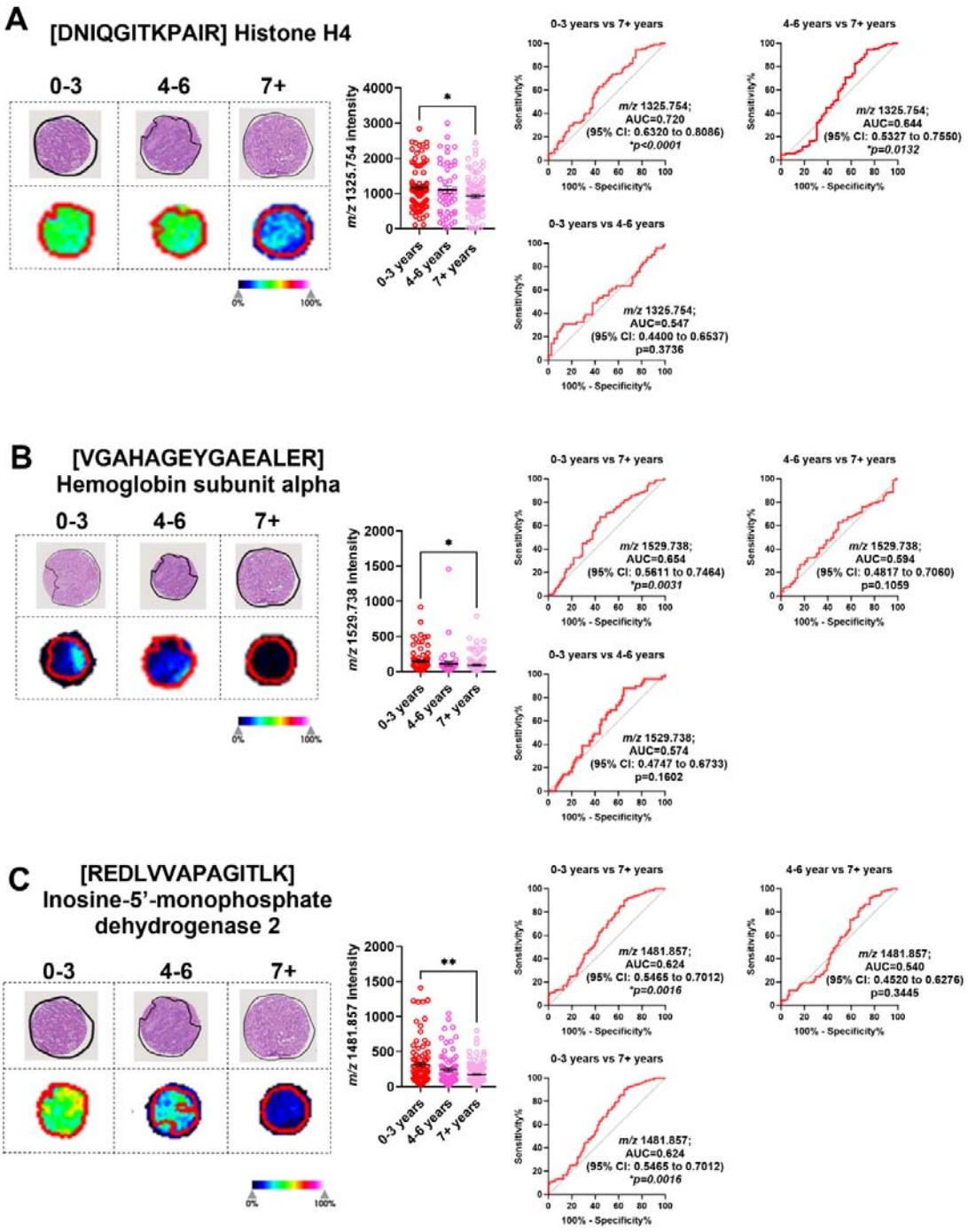


Figure 5. Significantly abundant tryptic peptides associated with male and female CRLM stroma regions, by MALDI-MSI and LC-MS/MS. (A) Representative H&E and ion intensity map of stroma region annotated CRLM tumour cores for *m/z* 1366.628. Tumour cores are 1.5mm in diameter. (A-F) Tryptic peptides (*m/z* 1366.628, 1296.682, 2211.113, 1570.676, 1661.060 and 1305.840) significantly more abundant in females relative to males. Tryptic peptides were identified as (A) Tryptase beta-2 Tryptase alpha/beta-1, (B) Proteasome subunit beta type-9, and (C-F) unidentified. (G-H) Tryptic peptides (*m/z* 1669.829 and 1461.702) significantly more abundant in males relative to females. These tryptic peptides were identified as (G) unidentified and (H) Carcinoembryonic antigen-related cell adhesion molecule 5. Bar graphs showing mean ± SEM and ROC plots are presented. ****$p<0.0001$. Student's t test. Each dot represents a single tumour core.



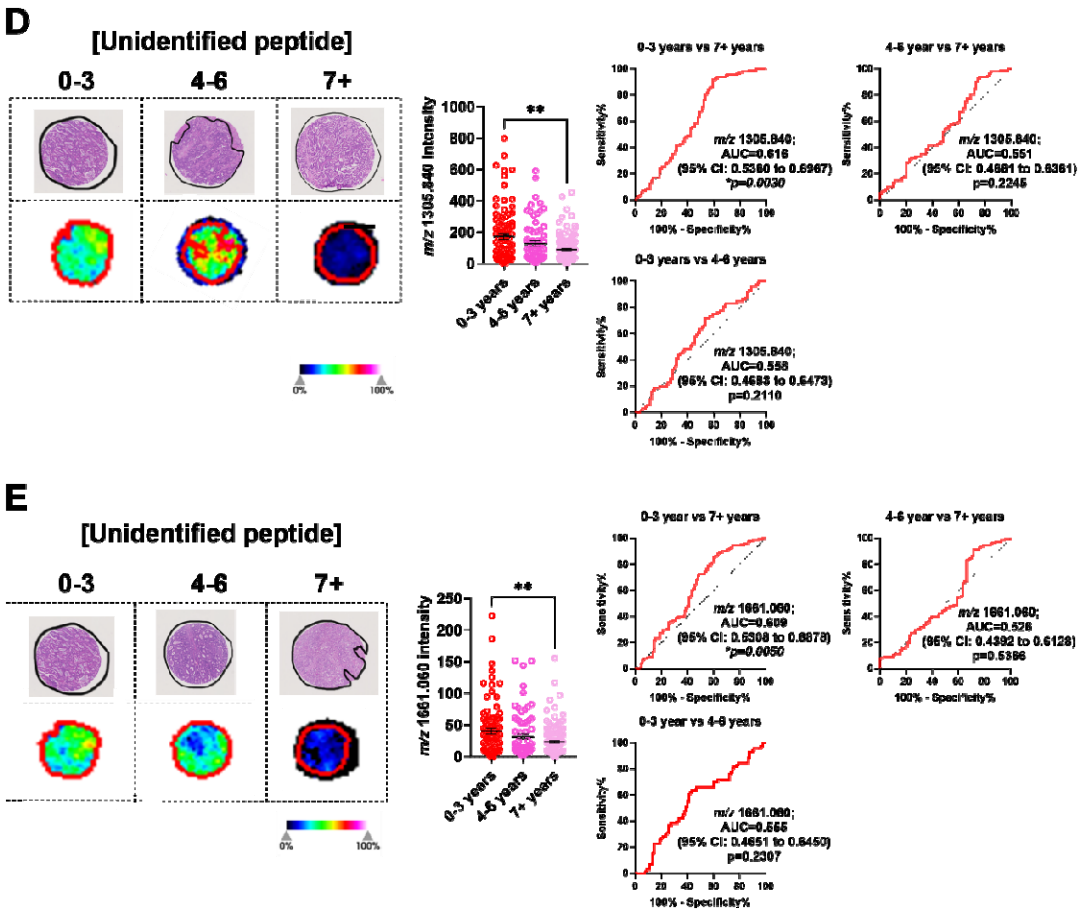


Figure 6. Significantly abundant tryptic peptides associated with poor overall survival (0-3 years) in CRLM tumour regions. CRLM cores were subdivided based on overall survival after curative-intent surgery (0-3, 4-6, and 7+ years). Representative H&E and ion intensity maps of tumour region annotated CRLM tumour cores for *m/z* 1325.754, 1529.738, 1481.857, 1305.840, and 1661.060. Tumour cores are 1.5mm in diameter. These tryptic peptides were identified as (A) Histone H4, (B) Haemoglobin subunit alpha, (C) Inosine-5'-monophosphase dehydrogenase 2 and (D-E) unidentified by LC-MS/MS. Bar graphs showing mean ± SEM and ROC plots are presented. *$p<0.05$, **$p<0.01$. Ordinary one-way ANOVA. Each dot represents a single tumour core.

633 # Tables

634

635 **Table 1. Demographic and clinicopathological characteristics of Australian CRLM patient cohort (n=84).**

Patients	***n* = 84**
Age	
Median age	67
Age range	33-88
Early onset <50	6 (7.1%)
Late onset >50	68 (80.9%)
Missing	10 (11.9%)
Sex	
Female	25 (29%)
Male	59 (70%)
Overall survival (years)	
Median [Min, Max]	3.75 [0.36-11.94]
Mean (SD)	4.51 (2.9)
Survival groups	
0-3 years	25 (29.7%)
4-6 years	24 (28.6%)
7+ years	35 (41.7%)

636 **Table 2. Tryptic peptides and proteins that discriminate between tumour and stroma regions in CRLM by MALDI-MSI and LC-MS/MS.**

No.	*m/z* Measured by MALDI-MSI	MALDI-MSI AUC (Tumour vs Stroma)	*p*-value	Enriched in Tumour vs Stroma	*m/z* Measured by LC-MS/MS	*m/z* Charge State (+2, +3)	Tryptic Peptide Sequence Confirmed by LC-MS/MS	Theoretical Singly Charged Tryptic Peptide *m/z*	*m/z* Error (ppm)	Protein Identified	Accession No.	Protein Coverage (%)
1	1589.876	0.618	**<0.0001***	Tumour	Unidentified							
2	1092.727	0.562	**<0.050***	Tumour	Unidentified							
3	1305.840	0.558	0.053	-	Unidentified							
4	1529.738	0.545	0.135	-	510.583	+3	VGAHAGEYGAEALER	1529.734	+2.6	Haemoglobin subunit alpha	P69905	71.1
5	1661.060	0.553	0.079	-	Unidentified							
6	1368.638	0.544	0.147	-	Unidentified							
7	1198.710	0.536	0.233	-	599.856	+2	AVFPSIVGRPR	1198.706	+3.3	Actin, cytoplasmic 1 Actin, cytoplasmic 2	P60709 P63261	59.2 59.2
8	1292.665	0.529	0.337	-	646.840	+2	DLQFVEVTDVK	1292.673	-6.2	Fibronectin	P02751	9.4
9	1585.772	0.531	0.053	-	Unidentified							
10	1591.817	0.531	0.531	-	796.407	+2	EGYLQIGANTQAAQK	1591.807	+6.3	Liver carboxylesterase 1	P23141	27.0
11	1641.805	0.525	0.414	-	821.402	+2	VDTNAPDLSLEGPEGK	1641.797	+4.9	Neuroblast differentiation-associated protein AHNAK	Q09666	31.7
12	1706.776	0.527	0.371	-	853.899	+2	EGPVQFEEDPFGLDK	1706.791	-8.8	SNW domain-containing protein 1	Q13573	2.8

13	1280.719	0.521	0.485	-	640.864	+2	TEVIPPLIENR	1280.721	-1.6	Apolipoprotein B-100	P04114	0.5
14	1481.857	0.528	0.355	-	494.628	+3	REDLVVAPAGITLK	1481.869	-8.1	Inosine-5'-monophosphate dehydrogenase 2	P12268	4.5
15	1954.064	0.524	0.426	-	652.026	+3	VAPEEHPVLLTEAPLNPK	1954.064	0.0	Actin, cytoplasmic 1 Actin, cytoplasmic 2	P60709 P63261	59.2 59.2
16	2056.984	0.521	0.481	-	686.327	+3	LAEQAERYDEMVESMKK	2056.968	+7.8	14-3-3 protein epsilon	P62258	45.5
17	2211.113	0.532	0.287	-	Unidentified							
18	1325.754	0.513	0.680	-	663.380	+2	DNIQGITKPAIR	1325.754	0.0	Histone H4	P62805	54.4
19	1465.690	0.503	0.908	-	733.350	+2	SDLVNEEATGQFR	1465.692	-1.4	Carcinoembryonic antigen-related cell adhesion molecule 5	P06731	14.0
20	1833.902	0.500	0.996	-	611.979	+3	VFSNGADLSGVTEEAPLK	1833.923	-11.5	Alpha-1-antitrypsin	P01009	34.0

* Indicates statistically significant differences.

Table 3. Tryptic peptides and proteins that discriminate between biological sex in CRLM by MALDI-MSI and LC-MS/MS.

No.	*m/z* Measured by MALDI-MSI	MALDI-MSI AUC Stroma (Male vs Female)	*p*-value	Stroma Enriched (Male vs Female)	MALDI-MSI AUC Tumour (Male vs Female)	*p*-value	Tumour Enriched (Males vs Female)	*m/z* Measured by LC-MS/MS	*m/z* Charge State (+2, +3)	Tryptic Peptide Sequence Confirmed by LC-MS/MS	Theoretical Singly Charged Tryptic Peptide *m/z*	*m/z* Error (ppm)	Protein Identified	Accession No.	Protein Coverage (%)
1	2211.113	0.633	**0.0139***	Female	0.695	**<0.0001***	Female	Unidentified							
2	1296.682	0.662	**0.0053***	Female	0.687	**<0.0001***	Female	648.834	+2	FTTDAIALAMSR	1296.662	+15.4	Proteasome subunit beta type-9	P28065	18.30
3	1570.676	0.645	**0.0077***	Female	0.689	**<0.0001***	Female	Unidentified							
4	1669.829	0.661	**0.0029***	Male	0.683	**<0.0001***	Male	Unidentified							
5	1461.702	0.563	0.2489	-	0.612	**0.0064***	Male	Unidentified							
6	1465.697	0.638	**0.0108***	Male	0.611	**0.0072***	Male	733.350	+2	SDLVNEEATGQFR	1465.692	+3.4	Carcinoembryonic antigen-related cell adhesion molecule 5	P06731	14.0
7	1235.618	0.582	0.132	-	0.605	**0.0111***	Male	618.316	+2	MVNHFIAEFK	1235.624	-4.9	Heat shock cognate 71 kDa protein	P11142	31.30
8	1366.628	0.677	**0.0011***	Female	0.601	**0.0145***	Female	683.823	+2	YHLGAYTGDDVR	1366.639	-8.0	Tryptase beta-2 Tryptase alpha/beta-1	P20231 Q15661	24 24
9	1481.857	0.558	0.2874	-	0.551	0.2128	-	494.628	+3	REDLVVAPAGITLK	1481.869	-8.1	Inosine-5'-monophosphate dehydrogenase 2	P12268	4.5
10	1280.719	0.549	0.3971	-	0.550	0.2247	-	640.864	+2	TEVIPPLIENR	1280.721	-1.6	Apolipoprotein B-100	P04114	0.5
11	1441.676	0.557	0.2972	-	0.547	0.2542	-	721.343	+2	EFTPQMQAAYQK	1441.678	-1.4	Hemoglobin subunit delta	P02042	40.10
12	1661.060	0.639	**0.0105***	Female	0.520	0.6212	-	Unidentified							
13	1305.840	0.650	**0.0056***	Female	0.505	0.8958	-	Unidentified							
14	1529.738	0.557	0.2972	-	0.502	0.9697	-	510.583	+3	VGAHAGEYGAEALER	1529.734	+2.6	Hemoglobin subunit alpha	P69905	71.1

645 * Indicates statistically significant differences.

Table 4. Tryptic peptides and proteins that discriminate overall survival identified by MALDI-MSI and LC-MS/MS.

No.	*m/z* Measured by MALDI-MSI	MALDI-MSI AUC Tumour (0-3 vs 7+ Years)	*p*-value	MALDI-MSI AUC Tumour (0-3 vs 4-6 Years)	*p*-value	MALDI-MSI AUC Tumour (4-6 vs 7+ Years)	*p*-value	*m/z* Measured by LC-MS/MS	*m/z* Charge State (+2, +3)	Tryptic Peptide Sequence Confirmed by LC-MS/MS	Theoretical Tryptic Peptide *m/z*	*m/z* Error (ppm)	Protein Identified	Accession No.	Protein Coverage (%)
1	1325.754	0.720	**<0.0001***	0.547	0.3736	0.644	**0.0132***	442.589	+3	DNIQGITKPAIR	1325.754	0.0	Histone H4	P62805	54.4
2	1529.738	0.654	**0.0031***	0.574	0.1602	0.594	0.1059	765.371	+2	VGAHAGEYGAEALER	1529.734	+2.6	Hemoglobin subunit alpha	P69905	71.1
3	1481.857	0.624	**0.0016***	0.573	0.1157	0.540	0.3445	494.628	+3	REDLVVAPAGITLK	1481.869	-8.1	Inosine-5'-monophosphate dehydrogenase 2	P12268	4.5
4	1305.840	0.616	**0.003***	0.558	0.211	0.551	0.2245	Unidentified							
5	1661.060	0.609	**0.005***	0.555	0.2307	0.526	0.5366	Unidentified							
6	1589.876	0.594	0.0696	0.552	0.3281	0.665	0.0045*	Unidentified							

* Indicates statistically significant differences.