bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Integration of Multiple Spatial Omics Modalities Reveals Unique Insights into Molecular Heterogeneity of Prostate Cancer

Wanqiu Zhang[1,2‡], Xander Spotbeen[3,4‡], Sebastiaan Vanuytven[5,6#], Sam Kint[4,5#], Tassiani Sarretto[7], Fabio Socciarelli[8], Katy Vandereyken[4,5], Jonas Dehairs[3], Jakub Idkowiak[3,9], David Wouters[4,10], Jose Ignacio Alvira Larizgoitia[4,10], Gabriele Partel[4,10], Alice Ly[2], Vincent de Laat[3], Maria José Q Mantas[2], Thomas Gevaert[11], Wout Devlies[11,12], Chui Yan Mah[13,14,15], Lisa M Butler[13,14,15], Massimo Loda[8], Steven Joniau[11,12], Bart De Moor[1], Alejandro Sifrim[4,10], Shane R. Ellis[7], Thierry Voet[4,5], Marc Claesen[2], Nico Verbeeck[2], Johannes V. Swinnen[3,4*]

‡These authors contributed equally
#These authors contributed equally
*Corresponding author (j.swinnen@kuleuven.be)

Author affiliations:

1 STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium
2 Aspect Analytics NV, C-mine 12, 3600 Genk, Belgium
3 Laboratory of Lipid Metabolism and Cancer, KU Leuven and Leuven Cancer Institute (LKI), Leuven, Belgium
4 KU Leuven Institute for Single Cell Omics (LISCO), KU Leuven, Leuven, Belgium
5 Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, Belgium
6 Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles (ULB), Brussels, Belgium
7 Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Northfields Ave, Wollongong, NSW 2522 Australia
8 Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY, USA
9 Department of Analytical Chemistry, Faculty of Chemical Technology, University of Pardubice, Pardubice, Czech Republic
10 Laboratory of Multi-omic Integrative Bioinformatics, Department of Human Genetics, KU Leuven, Leuven, Belgium
11 Department of Urology, University Hospitals Leuven, Leuven, Belgium
12 Department of Development and Regeneration, KU Leuven, Leuven, Belgium
13 South Australian Health and Medical Research Institute (SAHMRI), Adelaide, SA 5000, Australia.
14 South Australian Immunogenomics Cancer Institute, University of Adelaide, Adelaide, SA 5005, Australia.
15 Freemasons Centre for Male Health and Wellbeing, University of Adelaide, Adelaide, SA 5005, Australia.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Competing Financial Interests:
A.L., and M.J.Q.M. are employed by Aspect Analytics NV.
M.C. and N.V. are shareholders of Aspect Analytics NV.
T.V. is co-inventor on licensed patent WO/2014/053664 (High-throughput genotyping by sequencing low amounts of genetic material).
The remaining authors declare no competing interests.

Abstract

Recent advances in spatial omics methods are revolutionising biomedical research by enabling detailed molecular analyses of cells and their interactions in their native state. As most technologies capture only a specific type of molecules, there is an unmet need to enable integration of multiple spatial-omics datasets. This, however, presents several challenges as these analyses typically operate on separate tissue sections at disparate spatial resolutions. Here, we established a spatial multi-omics integration pipeline enabling co-registration and granularity matching, and applied it to integrate spatial transcriptomics, mass spectrometry-based lipidomics, single nucleus RNA-seq and histomorphological information from human prostate cancer patient samples. This approach revealed unique correlations between lipids and gene expression profiles that are linked to distinct cell populations and histopathological disease states and uncovered molecularly different subregions not discernible by morphology alone. By its ability to correlate datasets that span across the biomolecular and spatial scale, the application of this novel spatial multi-omics integration pipeline provides unprecedented insight into the intricate interplay between different classes of molecules in a tissue context. In addition, it has unique hypothesis-generating potential, and holds promise for applications in molecular pathology, biomarker and target discovery and other tissue-based research fields.

Introduction

Recent developments in spatial-omics technologies have revolutionised tissue-based research, enabling a deeper understanding of the cellular architecture of tissues, cell-cell interactions as well as the intricate relationship between molecular identity and tissue structure. These technologies find broad applications in studying normal tissue development[1], as well as in deciphering complex diseases like neurodegenerative disorders[2,3] and different types of cancer[4–9]. To investigate the various classes of molecules (RNA, proteins, metabolites, lipids) in a spatial context, different technologies are applied, including *in situ* hybridization, *in situ* sequencing and *in situ* RNA capturing followed by *ex situ* sequencing for spatial transcriptomics[10–19] (ST), histochemistry using fluorescently labelled antibodies[19–21] or genetically encoded fluorescent protein tags[22,23] for spatial proteomics, and mass spectrometry imaging (MSI) to study the distribution of peptides, metabolites, and lipids[24–29].

The biological questions addressed by such platforms are often complex and involve molecular alterations at multiple levels of the gene expression cascade. As each of the spatial-omics technologies often captures only a partial aspect of a single molecular layer in a tissue,

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

integration of spatial multi-omics (SMO) datasets that span across the biomolecular scale is essential to provide insight in the complex interplay between the different molecular layers and to better understand the intricacy of these biological processes, and their molecular drivers. Pipelines aiming to align and correlate these datasets encounter several obstacles, with the limited coverage of the molecular landscape being a prominent example. Moreover, distinct spatial technologies may not be compatible on a single tissue section, making the use of neighbouring tissue sections necessary[30]. As the structural organisation of tissues can differ between consecutive sections, and the experimental handling for the various platforms may result in tissue deformation, precise co-registration of adjacent datasets has proved difficult but is essential in order to align corresponding regions and prevent misinterpretation. This operation may be additionally complicated by the fact that many spatial technologies acquire data at different spatial resolutions, ranging from sub-cellular for microscopy-based technologies[31,32] to 55 μm for the Visium ST platform[33] and 1-100 μm for commercial MSI platforms[34]. Therefore, a granularity matching step is required to ensure that the spatial positioning of data points from the different modalities is reasonably equivalent. Finally, in spatial-omics studies, sensitivity, especially for detecting low-abundance molecules and achieving quantitative accuracy, poses further challenges. To address these limitations, spatial omics approaches can be complemented with bulk or single-cell omics technologies. Bulk-omics methods offer high coverage and sensitivity, enabling a more comprehensive and quantitative analysis of molecular profiles[35], while single-cell omics provides insights into cellular heterogeneity within tissues, enhancing the resolution and sensitivity of spatial data[30,36].

Recent studies have used a variety of different omics integration strategies, including co-registration and/or the use of a weighted nearest neighbour strategy to partially address the issues related to the combination of multiple spatial omics technologies[8,9,37,38]. Here, we have applied a novel spatial multi-omics pipeline (SMOx), which involves multi-point co-registration and Gaussian weighted granularity matching steps for more accurate data point matching, resulting in improved integration of various spatial omics modalities. Using this pipeline, we combined Visium ST (further referred to as ST) with MALDI-MSI coupled with post-ionisation (MALDI-2), which has increased sensitivity for detecting lipids compared to conventional MALDI[39,40], and single-nucleus RNA sequencing (snRNA-seq) to perform cell-type deconvolution of ST data[41] and high resolution pathology annotations. To explore the potential of this pipeline, we applied it to study human prostate cancer (PCa), which in view of its highly inter- and intra-tumoural heterogeneous nature[42] is an ideal case study for the SMOx integration pipeline. In fact, PCa often exhibits multiple morphological disease states within a single tissue section, ranging from prostatic intraepithelial neoplasia (PIN) to various histological malignant states (ISUP grades) and cribriform, each representing distinct disease subtypes with unique biological characteristics and clinical outcomes[43]. Moreover, PCa development involves molecular changes across multiple omics classes. Besides alterations in genes and transcripts, prominent changes also occur in lipids, offering unique potential for novel therapeutic interventions[44,45]. Using PCa as a paradigm, we showcase the potential of our novel spatial multi-omics pipeline to uncover lipid-gene expression pairs that are linked to different cell populations and histopathological disease states not discernible by morphology alone and can aid in the annotation of histologically ambiguous regions. These findings illustrate the potential application of our data integration pipeline to diverse fields of tissue-

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

based research, including molecular pathology and for the generation of novel research hypotheses.

Results

Establishment of a novel spatial multi-omics data integration pipeline (SMOx) and its application in prostate cancer

With the aim to more comprehensively map the molecular heterogeneity within tissues and to better understand the interplay between different classes of molecules in a tissue context, we established a spatial multi-omics pipeline (SMOx) and applied it to the field of PCa (**Fig. 1A)**. After collecting tumour and matching non-malignant biopsies from men that underwent a radical prostatectomy (n=8, **Supplementary Table 1**), 10 µm thick frozen tissue sections were cut. Neighbouring tissue sections were subjected to ST analysis using the Visium Spatial Gene Expression platform (n=16) and MALDI-2-MSI based lipidomics (n=16). From the ST analysis, we obtained curated expression data of 18,950 genes across 42,475 individual capture spots on the ST array slides (summarised across all samples). Only spots under the tissue sections were considered for downstream computational analysis. From the MSI data, peak picking identified 12,510 peaks across the samples. The average overall spectra of cancer and non-cancer samples are presented in **Supplementary Fig. 1**. High-resolution digital H&E images of the MSI sections were annotated by two independent expert uro-pathologists. Additionally, we collected three 100µm thick cryo-sections and isolated nuclei for snRNA-seq (n=13) from the remaining tissue directly adjacent to the cryo-sections used for ST and MSI analysis, to enable the identification and mapping of distinct cell populations across the tissue by applying the 'cell2location' deconvolution algorithm[46].

To integrate the data from the ST and MSI spatial-omics modalities, which are generated on separate (neighbouring) tissue sections and are acquired at different spatial resolutions, the SMOx pipeline was designed to allow co-registration and granularity matching (**Fig. 1B-C**). As both ST and MSI routinely include H&E staining of tissue sections in their workflows[47–49], we used the H&E-stained images to align the ST and MSI data. To facilitate cross-modality comparison and uncover inherent molecular patterns, the data from the different modalities were independently subjected to dimensionality reduction using uniform manifold approximation and projection (UMAP)[50]. The UMAP representation (i.e., mapped 3 dimensional embeddings) of the MSI data[51] was used to guide the co-registration with the respective post-MSI H&E image. Subsequently, the MSI H&E image was co-registered to the H&E image of the ST Visium slide. Alignment of ST data with the respective H&E histological images was done automatically using the Space Ranger pipeline (10X Genomics), eliminating the need for manual co-registration[52]. This strategy of using H&E-stained images to align ST and MSI data reduces the complex multimodal registration to a mono-modal one between their respective H&E-stained images, to which each modality has previously been co-registered, and removes the selection effect of registering to a template image. Subsequently, a shared spatial coordinate system was created by linking the MSI data with the ST data. This approach produces a readout that can be considered a "virtual section", between the sections used for ST and MSI analysis (**Fig. 1B**). Following co-registering of MSI and ST data, we addressed the disparity in spatial scales of data acquisition points through granularity matching. In this

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

study, MSI was conducted at a spatial resolution (pixel size) of 30 μm, whereas the Visium assay has spot diameters of 55 μm with a 100 μm centre-to-centre distance between spots[33]. Consequently, each Visium spot corresponds to multiple MSI pixels with uneven amounts of coverage **(Fig. 1C)**. To match the granularity of the spatial readouts, we employed Gaussian weighting to the MSI pixels close to each ST spot after registration[53]. This procedure resulted in the generation of a comprehensive aggregated MSI feature spectrum linked to each ST spot, ensuring synchronised spatial readouts between MSI and ST data (**Fig. 1C**). The resulting co-registered and matched MSI-ST data structure can be interpreted as a data frame consisting of matched readouts for each Visium spot. Within this data frame, each row represents a single matched spot, while the columns encompass distinct dataset information at that specific spot, such as gene expression, lipid content, pathology annotations, deconvoluted cell types, etc. Thus, this integrated spatial multi-omics pipeline could link transcript-lipid pairs with tissue features.

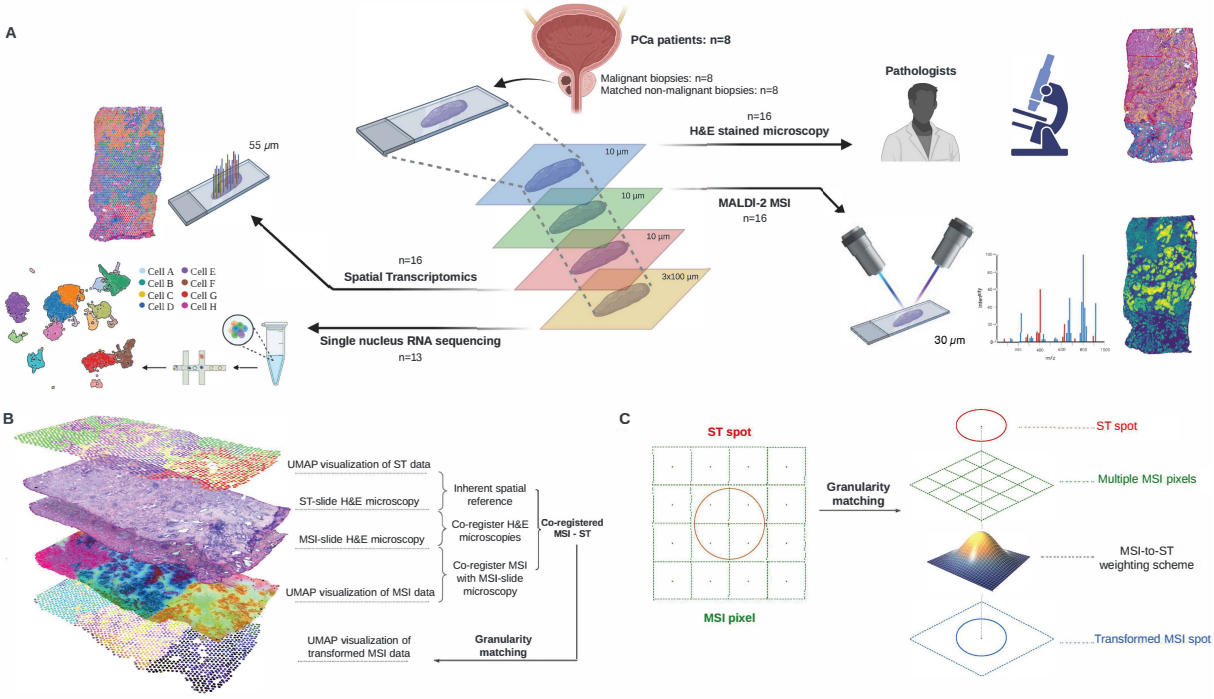


Figure 1: **Study design and overview of the spatial multi-omics integration pipeline (SMOx). A.** Schematic representation of the various datasets generated on prostate cancer samples and the applied integrated spatial multi-omics workflow. **B.** Strategy for co-registration and integration of different spatial multi-omics data layers generated on consecutive tissue sections. To link ST data with MSI data, UMAP visualizations of different modalities were co-registered in two steps based on the haematoxylin and eosin (H&E) stainings of the respective slides and MSI data, and a UMAP visualisation of transformed MSI data was generated after granularity matching. **C.** Overview of the granularity matching step to bridge the scale gap between MSI and ST data. Spatial readouts were aligned, and Gaussian weighting was applied to generate a transformed MSI spot.

Application of the SMOx pipeline to identify correlated transcript-lipid pairs linked with histological annotations

Using our spatial multi-omics integration strategy, we explored the spatial correlation of transcripts and lipids in a PCa tissue section of a single patient (patient 929) as a proof of concept. The level of detail in the annotations created by the histopathologist on the digital H&E image, along with the transformation of annotations in the shared coordinate system, is

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

depicted in **Fig. 2A**. In this tissue section, five histologically distinct regions could be discerned: benign epithelium, stroma, PIN, tumour ISUP 3 and immune cells. When UMAP embeddings are projected onto tissue locations, similar patterns emerge among the distinct spatial modalities, indicating that despite the diverse nature of spatial data readouts (genes, lipids, etc.) (**Fig. 2B**), they all contain relevant molecular information reflecting the distinct histopathological features in the sample.

To link the different levels of molecular information, we conducted correlation analysis assessing the relationship between individual transcripts and MSI features (most of which are lipids), along with their spatial localization in the tissue (**Fig. 2C-D, Supplementary Fig. 2**). Among the positively correlated lipid-transcript pairs (correlation coefficient over 0.65), we observed PC 34:2 $[M+H]^+$ and *TMEFF2* (a transmembrane proteoglycan with multiple roles in cancer[54]) as one of the most highly correlated pairs in regions annotated as tumour ISUP3. PIN included PE 40:6 $[M+H]^+$ and *AMACR* (an established PCa marker[55]) as a highly correlated pair. LPE 16:0 $[M+H]^+$ and *NEFH* (a neurofilament-encoding gene maintaining cell integrity) were identified in the benign epithelium region. In contrast, a notable spatial negative correlation (-0.631) was observed between the lipid SM 34:1;O2 $[M+H]^+$ and the *AGR2* gene (a disulphide isomerase gene family implicated in tumour metastasis[56]), with their expression detected in tumour ISUP 3 and PIN regions, respectively. Interestingly, several of the identified transcripts listed in **Fig. 2C** encode enzymes that are directly involved in lipid metabolism, including DEGS1 (a sphingolipid desaturase[57]), HPGD (a 15-hydroxyprostaglandin dehydrogenase associated with PCa risk[58]), AMACR (an enzyme involved in peroxisomal beta-oxidation of branched-chain fatty acids and one of the most widely used histological markers upregulated in PCa[59]), and ACSM1 (a medium chain Acyl-CoA Synthetase that is overexpressed in PCa[60,61]). Importantly, these genes are up-regulated in areas annotated by pathologists as PIN and tumour ISUP3, in comparison to the regions annotated as benign epithelium. Many other transcripts are not directly implicated in lipid metabolism but are known to have a cell type-specific expression pattern and are differentially expressed upon cancer development. These include transcripts of genes *TMEFF2*, *KLK3* (encoding the prostate-specific antigen biomarker), *PCA3*, *AGR2*, and *NEFH*. These observations suggest that transcript-lipid correlations are associated with disease state and specific cell populations, and may reveal functional interactions.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

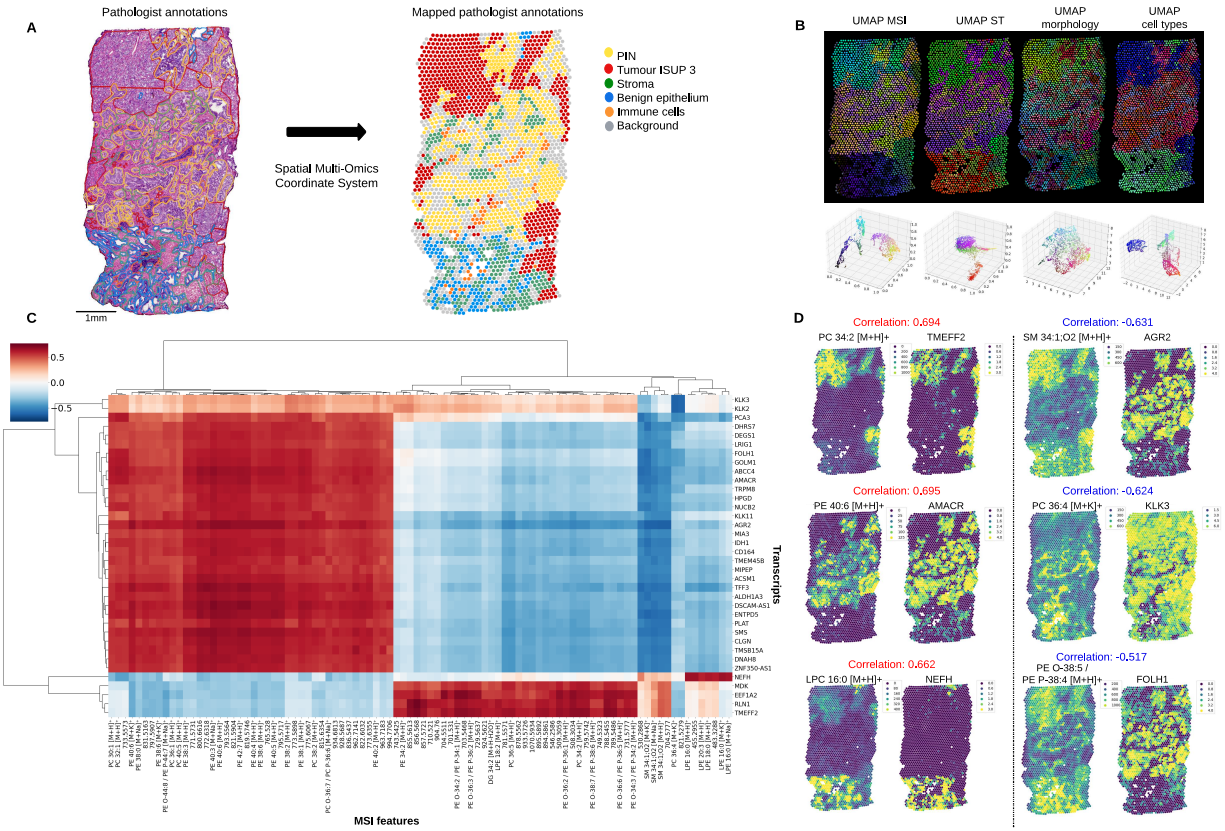


Figure 2. Integration of spatially resolved multi-omics reveals spatially correlated lipid-gene pairs in a prostate cancer sample. A. Highly detailed pathologist annotations of a representative prostate cancer sample (929_cancer) on the post-MSI H&E-stained tissue section and their spatial distribution projected on the shared coordinate system. **B.** UMAP representations of individual modalities including MS imaging data, ST data, morphological features extracted from pathology annotations, and cell types derived from snRNA-seq data after spot deconvolution. The top row represents a hyperspectral visualisation of the UMAP data on the shared coordinate system using the RGB colour scheme. The bottom row shows the scatter plots of the resulting 3-dimensional UMAP embeddings for each modality. **C.** Hierarchically clustered heatmap of the correlation matrix between MSI features and ST data. **D.** Spatial visualisation of representative positively and negatively correlated lipid-transcript pairs in the 929-cancer sample.

Identification of molecular expression patterns associated with histological annotations

To further explore the alignment of MSI data and ST expression patterns with pathology annotations, we applied non-negative matrix factorisation (NMF) to extract correlated features from the individual modalities of the integrated spatial multi-omics dataset of sample 929_cancer. The resulting correlation heatmap between MSI NMF components and pathology annotations demonstrated significant positive correlations for each pathological annotation (PIN, benign epithelium, tumour IUSP3 and stroma) except for immune cells (likely due to their greater cell type variability) (**Fig. 3A top panel**). The spatial distribution of these NMF components, along with pathology annotations within the common coordinate system are illustrated in **Fig. 3B**, accentuating the congruence between lipidome and histology. Similarly, significant positive correlations were observed between ST NMF components and the different annotated regions (including immune cells) (**Fig. 3A bottom panel** and **Fig. 3B**). For MSI NMF component 9, which highly correlates with PIN, the pseudo-spectrum of the biomolecular ion peaks that are involved in the expression of the MSI NMF component's spatial patterns is

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shown, along with the gene expression profile of the PIN-associated ST NMF component 13 (**Fig 3A**). Integration of both spatial-omics modalities within the shared coordinate system and subsequent hierarchical clustering revealed strong correlations between specific NMF components of both modalities (**Fig. 3C**). Differentially expressed lipids and genes that contribute significantly to the molecular characteristics of the different annotated areas are listed in **Fig. 3D**. The high level of correlation provides further credence to a spatial relationship between readouts of the ST and MSI data, and overall agreement with the pathology annotations.

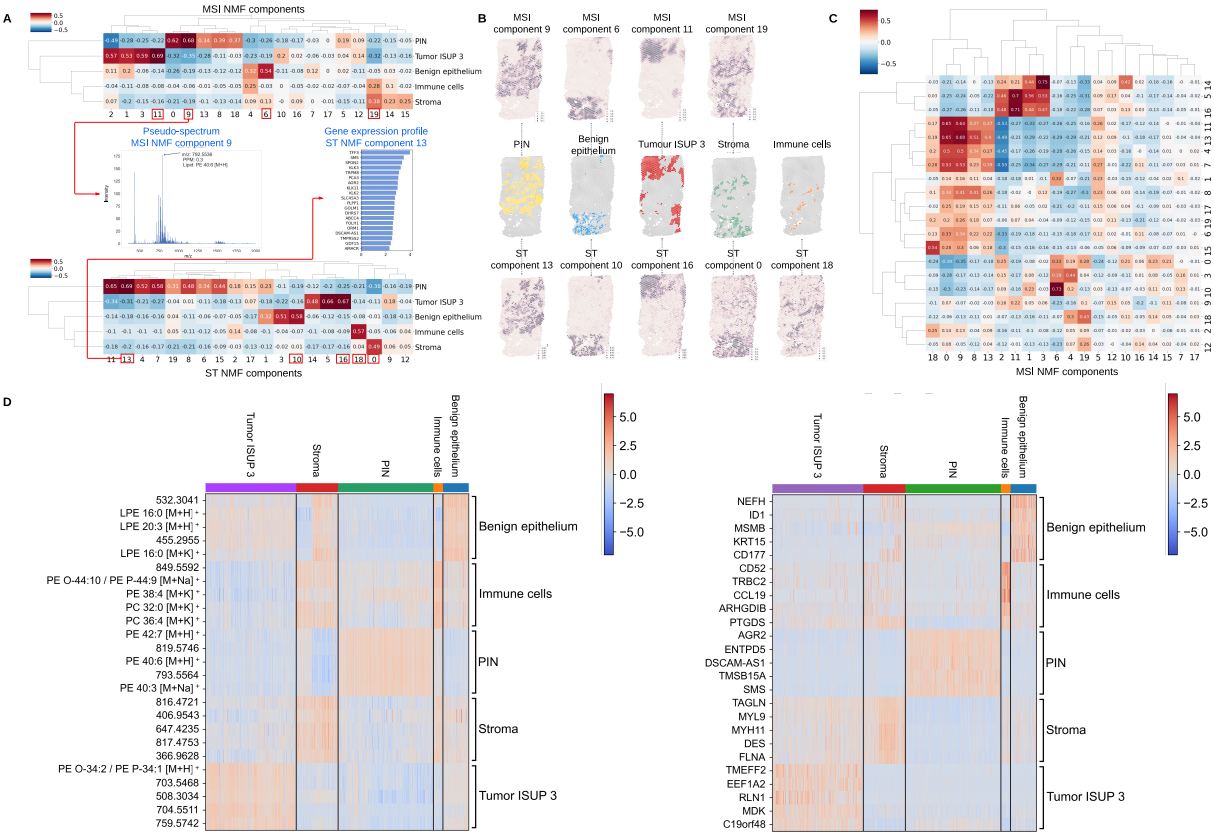


Figure 3. Correlation between histological and molecular profiling. A. Heatmap showing correlations between MSI NMF components and the five histologically distinct regions from sample 929_cancer. A representative pseudospectrum for MSI NMF component 9 and the most enriched genes of the correlating ST component 13 are shown. **B.** Spatial visualisation of representative MSI NMF components (top panels) correlating with ST NMF components (bottom panels) for each histological feature (middle panel). **C.** Hierarchically clustered heatmap of the overall correlations between MSI NMF and ST NMF components. Numerical values correspond to the respective correlation coefficient of each MSI NMF – ST NMF pair. **D.** Heatmaps illustrating MSI features (mostly lipids) (left panel) and genes (right panel) exhibiting the most significant differential expression within each of the five histological regions in the 929_cancer sample.

Linking lipids and transcripts to specific cell types in the spatial coordinate system

Due to the granularity of the matched MSI-ST data set with a spatial resolution defined by the spatial-omics analysis techniques used (ST: 55 μm resolution; MSI 30 μm resolution), each spot in the spatial coordinate system corresponds to multiple cells, hampering the assignment of lipids and transcripts to specific cell types. To address this issue, we applied snRNA-seq-based spot deconvolution using the 'cell2location' deconvolution algorithm[46], based on snRNA-seq data from a set of 100 μm thick serial tissue sections neighbouring those utilised

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for ST. The resulting UMAP clusters were annotated based on curated marker genes[62–64], leading to the identification of 14 distinct cell types **(Fig. 4A)**. Of note, the TC1 (Tumour Cell 1) cell type was mainly enriched in both the tumour ISUP 3 and PIN region, whereas TC2 (Tumour Cell 2) was linked mainly to the PIN region only (**Fig. 4B-C**). Interestingly, cells with a club cell-specific gene signature, which have recently been associated with PCa carcinogenesis[63], localised to sub-regions within the PIN-annotated areas. Correlation analysis between MSI and ST data from spots enriched in specific cell types revealed unique lipid profiles associated with a cell type along with marker genes and enzymes involved in lipid metabolism, as illustrated or TC1 and TC2 cells, club cells and macrophages (**Fig. 4D)**. This analysis revealed that ISUP 3-associated TC1 cells and PIN-associated TC2 cells correlated with long chain polyunsaturated phospholipids (mainly PE), with an additional enrichment for hexosylceramides (HexCer) in the TC2 cell component. Cholesteryl esters (CE), mainly identified as [2M+H]+ ions[65] and known to accumulate in lipid droplets, were associated with club cells. Macrophage marker genes, like *HLA-D* genes[66], *LYZ*[67] and *CXCL5*[68], correlated best with ether lipids, CE and long polyunsaturated phosphatidylcholines (PC). Complementary gene signature analysis of lipid metabolism pathways revealed an overall high expression of genes linked to lipid metabolism pathways in luminal and TC2 cells, along with cell type-specific pathway activations (**Fig. 4E**). Examples include the upregulation of fatty acid biosynthesis and elongation in cancer cells (TC1 and/or TC2), steroid biosynthesis in luminal cells, and arachidonic acid metabolism in macrophages and mast cells, in line with expectations and data available in the literature[44,69,70]. Projection of these pathways in the UMAP representations confirmed these links and provided an additional layer of granularity, for instance revealing a specific subset of cells within the TC1 and TC2 clusters displaying elevated expression of genes involved in linoleic acid metabolism (**Fig. 4F**). These findings illustrate the benefit of linking lipids and transcripts to specific cell types in the spatial coordinate system.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

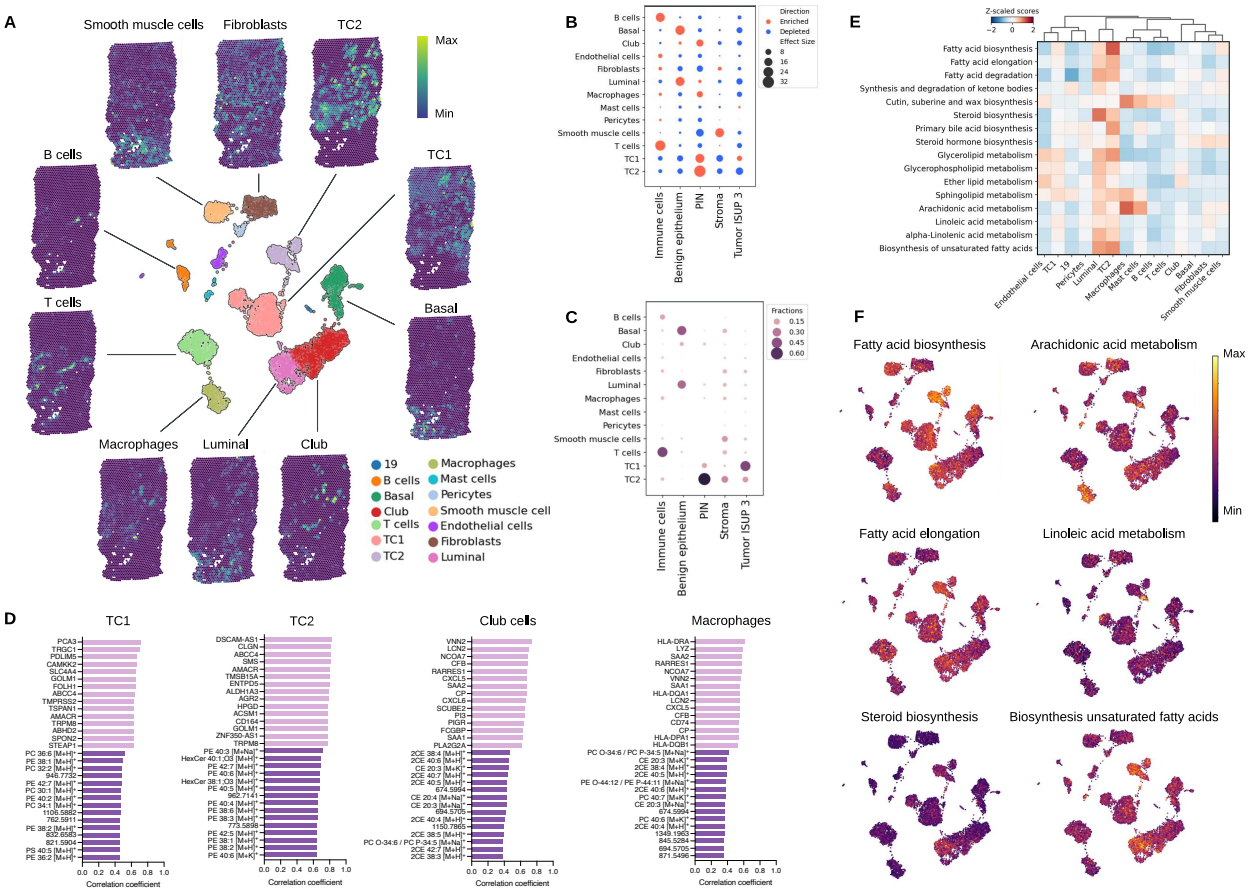


Figure 4. Spot deconvolution reveals cell type specific transcriptional and lipidomic profiles. A. UMAP representation of annotated snRNA-seq clusters and spatial localisation of the main cell types in the spatial transcriptomics data after spot deconvolution using cell2location[42]. TC1: tumour cluster 1, TC2: tumour cluster 2. **B.** Dot plot displaying the enrichment (orange) and depletion (blue) of cell types within the regions demarcated by the pathologist. **C.** Dot plot illustrating the relative abundance of a specific cell type within each pathologically annotated region. **D.** MSI features (pink) and transcripts (purple) with the highest positive correlations with TC1, TC2, club cells and macrophages. **E.** KEGG enrichment of differentially expressed genes involved in lipid metabolism per annotated cell type based on the snRNA-seq data. Colours indicate mean up- (red) or down- (blue) regulation of the pathway, while colour intensity indicates significance level. **F.** UMAP representation of the enrichment of different lipid metabolism pathways per cell type.

Integrated data analysis reveals molecular heterogeneity in neoplastic disease areas

Whereas the spatial distributions of MSI NMF components and ST NMF components in sample 929_cancer overall matched well with the histological annotations, the region annotated as PIN by the pathologists, was captured by two NMF components in both modalities – MSI NMF component 9 and 18 and ST NMF components 13 and 15 (**Fig. 5A-B**). Additionally, snRNA-seq-based spot deconvolution revealed that one of the components of each modality (MSI NMF 9 and ST NMF 13) corresponded well with the spatial distribution of TC2 cells (**Fig. 5C**). The other component (MSI NMF component 18 and ST NMF component 15) aligned with a specific subregion of PIN corresponding to the spatial distribution of epithelial club cells and macrophages. Examination of the H&E stainings of these latter regions confirmed the presence of a neoplastic gland-like structure. These observations underscore the potential of the integrated data analysis approach to reveal molecular heterogeneity in

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

neoplastic disease areas that might not be discernible through morphological examination alone.

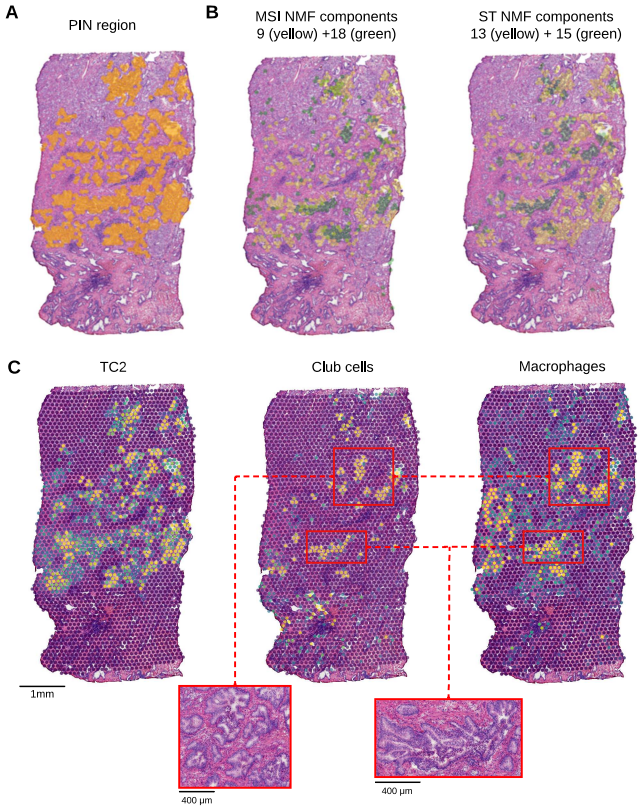


Figure 5. Integration of spatially resolved omics unveils molecular heterogeneity in PIN. A. Pathologist annotation of PIN in sample 929_cancer (highlighted in yellow). **B** MSI and ST NMF components correlating with the annotated PIN region in panel A. **C.** Localisation of TC2, club cells and macrophages based on snRNA-seq-deconvoluted ST data. Inserts show H&E stainings of the indicated areas.

Integrated spatial multi-omics aids in the histological annotation of morphologically ambiguous tissue areas

Integrated SMOx analysis of the entire set of 8 PCa samples overall revealed a strong concordance with the pathologist's annotations. Nevertheless, in several instances our integrated data-analysis pipeline (based on ST and MSI NMF components) suggested alternative histopathological assignments for areas in the tissues. This was for instance the case in sample 941_normal in which a region initially labelled as PIN exhibited closer molecular concordance with tumour ISUP4 (**Fig. 6A-B**). This observation was further supported by the snRNA-seq data, which indicated the presence of a population of TC1 cells. Furthermore, we observed a match between the top correlated MSI features, predominantly lipids, and genes of the ISUP 4-associated component with the top correlated MSI features and genes of the TC1 cell population (**Fig. 6C**). Re-examination of this region by the pathologist confirmed the presence of a micro-area of infiltrating prostate adenocarcinoma cells (lower inset box, **Fig. 6A-B**). Another ambiguous tissue area was identified in sample 931_cancer, which, based on the H&E staining of the frozen tissue section was annotated by the pathologist as "other" **(Fig. 6D).** Based on the ST and MSI NMF components (the latter one showing the strongest spatial correlations), our integrated approach identified this region as a subcluster within the ISUP5 tumour region. snRNA-seq data deconvolution further

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

revealed the presence of two distinct tumour clusters, TC1 highly correlated with ISUP5 and TC2 correlated with the ambiguous region (**Fig. 6E-F**). Histopathological re-evaluation of this region assigned it as cribriform. These results highlight the potential of spatial multi-omics to aid in histopathological annotation of ambiguous tissue areas.

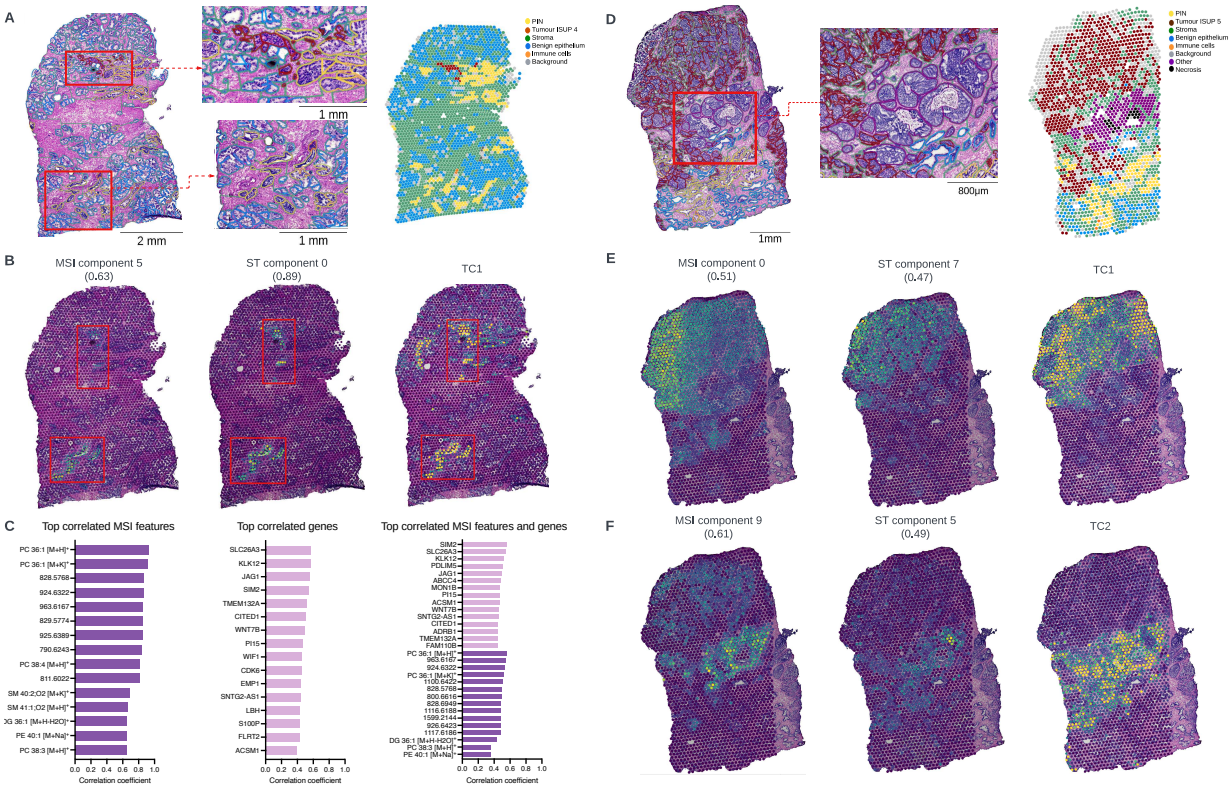


Figure 6. Integrated spatial multi-omics supporting histological annotation of ambiguous tissue areas. A. Highly detailed pathologist annotations of a non-malignant prostate sample (941_normal) on the post-MSI H&E-stained tissue section and their spatial distribution projected on the shared coordinate system. Inserts show H&E stainings of tumour ISUP 4 and a morphologically ambiguous region. **B.** Distributions of MSI NMF component 5, ST NMF component 0 and TC1 cell population based on snRNA-seq-deconvoluted ST data. Numerical values (0.63 and 0.89) correspond to the respective correlation coefficient of the MSI and ST NMF with the region annotated by the pathologists as tumour ISUP 4. **C.** Top correlated MSI features and genes of MSI NMF component 5 and ST NMF component 0, respectively. **D.** Highly detailed pathologist annotations of a high-grade prostate cancer sample (931_cancer) on the post-MSI H&E-stained tissue section and their spatial distribution projected on the shared coordinate system. Inserts show H&E stainings of an area annotated by the pathologist as “other”. **E.** Distributions of MSI NMF component 0, ST NMF component 7 and TC1 cell population based on snRNA-seq-deconvoluted ST data. Numerical values (0.51 and 0.47) correspond to the respective correlation coefficient of the MSI and ST NMF with the region annotated by the pathologists as tumour ISUP 5. **F.** Distributions of MSI NMF component 9, ST NMF component 5 and TC2 cell population based on snRNA-seq-deconvoluted ST data. Numerical values (0.61 and 0.49) correspond to the respective correlation coefficient of the MSI and ST NMF with the region annotated by the pathologists as “other”.

Implementation of the SMOx workflow across the entire prostate cancer sample cohort reveals tumour-specific correlations between transcripts and lipids

Having demonstrated the ability of the integrated SMOx pipeline to extract relevant information from multi-layered molecular expression patterns within a tissue sample, we expanded our analysis to the entire PCa cohort with additional Pearson correlation analysis. Examination of pairs of transcripts and MSI features with correlation coefficients exceeding 0.4 across all 16 biopsies (comprising 8 cancer samples and 8 matching non-malignant samples), allowed us to identify pairs of transcripts and MSI features with a spatial distribution that either correlated

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

or anti-correlated throughout the cohort (**Fig. 7A)**. These included well-established links i.e., between *FASN* expression and PC34:1[44], confirming the validity of this approach, as well as several novel links illustrating the hypothesis-generating potential of the SMOx pipeline. Next, Cohen's d values were calculated between tissue types i.e., tumour versus non-tumour tissue (as annotated by pathologists on the H&E microscopy), aiming to detect transcript-MSI feature correlations that significantly changed between these tissue types. One key example is the MSMB - PE 40:2 $[M+H]^+$ pair displaying an associated Cohen's d value of 1.54, indicating a substantial effect size[71], which in this case implies a considerable difference in spatial co-expression between tumour and non-tumour samples. In fact, as illustrated by a line plot of correlation coefficients for the pair across the 16 biopsies (**Fig. 7B)** and the actual spatial distributions of the transcript and the lipid in the shared coordinate system of both tumour and non-tumour samples (**Fig. 7C)**, these markers exhibited a clear spatial co-expression in non-tumour samples, which was absent in the tumour samples (quantified with average correlation coefficients of approximately 0.5 and 0, respectively). Specifically, in the non-tumour samples a striking alignment was observed between the gene *MSMB* and lipid PE 40:2 $[M+H]^+$, whereas in tumour samples, while *MSMB* continued to exhibit a robust spatial correlation, the lipid PE 40:2 $[M+H]^+$ lost its correlation with the benign epithelial cells (highlighted in green). This spatial correlation between PE 40:2 with tumour regions was observed in the majority of tumour samples. While the precise function of this lipid remains elusive, *MSMB* has been previously shown to exhibit a higher expression in normal prostate tissue than in cancerous tissue[72,73]. These observations highlight the hypothesis-generating potential of our SMOx approach.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

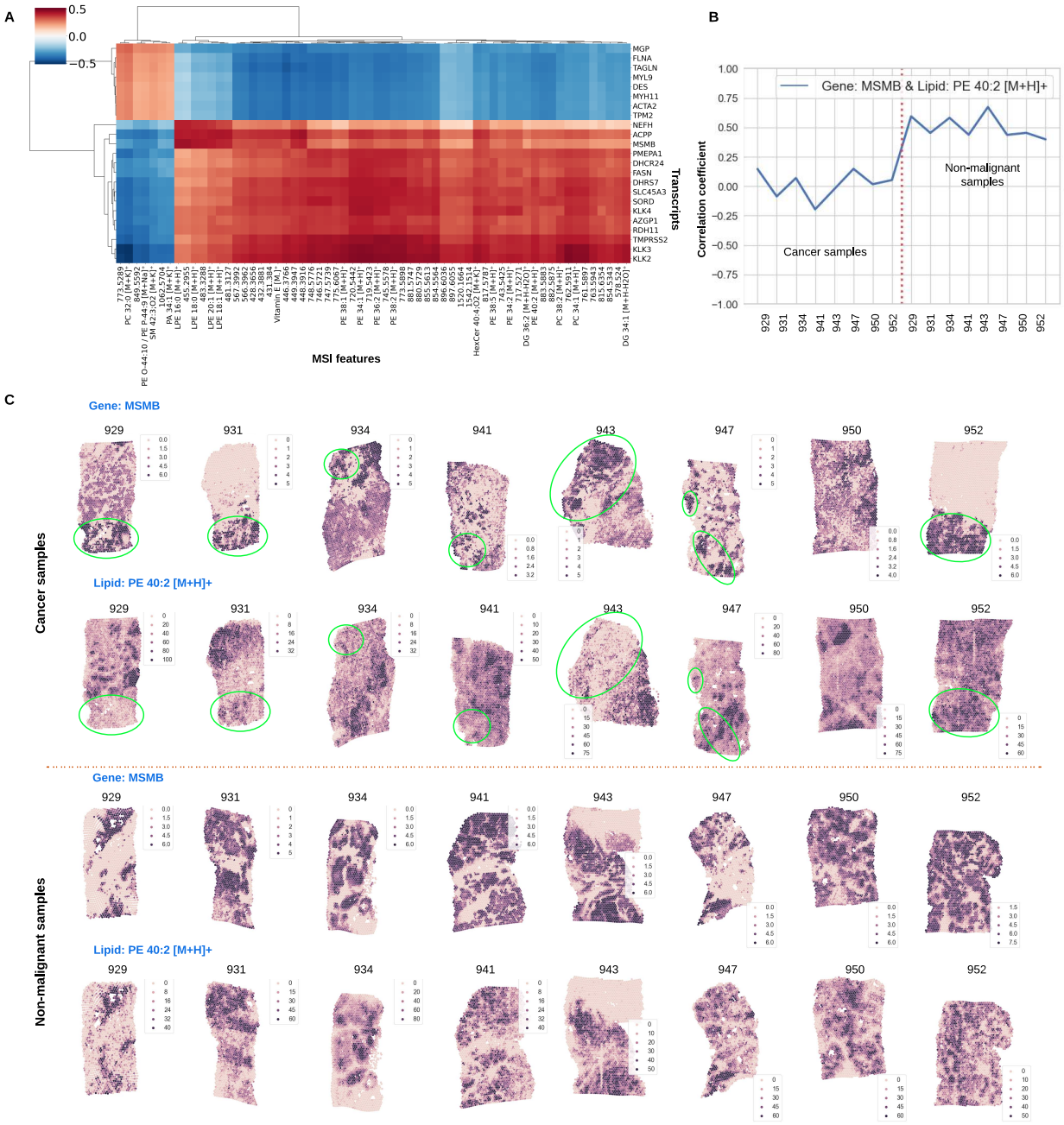


Figure 7. Implementation of the spatial multi-omic workflow across the entire prostate cancer sample cohort. A. Heatmap illustrating correlations between genes and MSI features, mostly lipids, where correlation coefficients greater than 0.4 are observed across the entirety of the sample cohort consisting of 16 biopsies (8 cancer samples and 8 matching non-malignant samples). **B.** Line plot depicting the correlation coefficients for the gene/lipid pair MSMB/PE 40:2 $[M+H]^+$ across the entirety of the sample cohort. **C.** Spatial distribution of the gene/lipid pair MSMB/PE 40:2 $[M+H]^+$ observed across the eight cancer biopsies (upper section) and the corresponding non-malignant samples (lower section). Green polygons indicate regions predominantly composed of benign epithelial cells within the cancer samples.

Discussion

Spatial multi-omics is progressively establishing itself as a powerful tool to better understand molecular events and its interactions in a native tissue context. One of the key challenges

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

related to this approach is the integration of multiple layers of molecular information that are generally acquired using a range of technologies to optimally capture the molecular information in each layer. Although the ability to conduct diverse spatial approaches, including ST and MSI analyses on a single slide has recently been reported[38], hitherto, most spatial multi-omics studies are performed on neighbouring tissue sections and generate spatial molecular maps with different spatial resolutions. These factors are further complicated by different tissue deformations due to technology-specific tissue processing, different inter-spot distances and offset of the spots. Our study, which integrates ST and MSI analysis aimed to overcome these hurdles by establishing a data-processing pipeline involving spatial co-registration and granularity matching. Co-registration refers to the mapping of pixels between modalities and is a crucial step in multimodal imaging analysis. Registration of imaging datasets is typically contingent on shared anatomical structures but can be especially challenging when modalities have distinct characteristics, or if measurements are performed at vastly different spatial scales[74]. The former challenge is typically addressed by generating a number of representative images for each modality e.g., by manually selecting from the imaging datasets or by generating images that capture spatial trends in the data by dimensionality reduction methods. The latter, multi-scale challenge refers to the fact that a spatial data point, typically referred to as a pixel or voxel, in one modality can cover a different surface area from that in another modality. These areas can range in the order of magnitude from square nanometres to hundreds of square micrometres. This problem is receiving increasing attention in the spatial biology field, due to the increasing number and complexity of spatial omics applications[75]. A granularity matching step is required to make data points reasonably equivalent between different modalities, typically via an aggregation strategy such as interpolation[8], averaging, or more advanced methods e.g., Gaussian smoothing[53,76]. In a recent spatial multi-omics study, combining MSI at 15 μm pixel size with multispectral immunofluorescence microscopy[37], a two-step co-registration was used reporting subcellular accuracy and ROIs from the microscopy projected onto the MSI data, however no granularity matching step was reported. Another study combining MSI and ST using the Visium platform used a spatial resolution of 100 μm, corresponding to the largest modality used, when integrating their data, but do not report whether aggregation was used to account for capture areas not being distributed equally across three modalities[9]. Another study used a weighted nearest neighbour strategy that mechanistically paired Visium spots in the centre of one 50 μm MSI pixel[8]. In contrast to these studies, we used a Gaussian approach that accounts for sample coverage per data acquisition area (Visium spot/MSI pixel). Together with our two-step co-registration tool, a full data analysis pipeline was established that has the ability to deal with the mentioned hurdles in spatial multi-omics data integration.

When applied to a set of clinical PCa samples, downstream integrated analysis revealed a strong spatial concordance amongst the individual modalities as seen in UMAP and NMF analysis. Importantly, agreement between the modalities in this combined approach allows more confidence in defining tissue profiles in spatial atlas approaches, assignment of molecularly heterogeneous regions not discernible by morphology, and potential uses for identifying missed cancerous regions e.g., in a diagnostic context. Using our approach to combine ST, MSI and snRNA-seq analysis on PCa samples, we were able to find lipid and transcript profiles that correlated to specific histological states and to specific cell populations. Based on these correlations between lipids and transcript profiles, we were able to assign

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

unique lipid profiles to specific cell types and disease states and to generate unprecedented insight into the heterogeneity of PCa pathology in the context of a native tissue.

One striking example of the power of our approach was the discovery of a molecularly distinct profile in a PIN region based on both ST and MSI profiling, providing more confidence in the pathological characterisation. Integrated snRNA-seq-based deconvolution revealed the presence of club cells, which recently have been linked to the pathogenesis of PCa[63]. Although the clinical significance of this finding will have to confirmed, our observation of the presence of club cells in a specific pre-malignant histological structure, may point towards that region's future potential to develop into adenocarcinoma and warrants further investigation.

When applied to the entire cohort of PCa and matched benign samples, we found that spatial transcriptomics and spatial lipidomics data largely followed histomorphological patterns. In addition, combining the two analyses allowed us to identify genetic and lipid expression profiles corresponding to different cell populations and tissue states. Our integrated spatial multi-omics approach was able to identify transcript – MSI feature pairs that correlate or anti-correlate throughout the cohort. Besides well-known associations, including the correlation between *FASN* and PC34:1[44], which provides validation for our approach, several new correlations were found that warrant future investigation. One striking example was the loss of co-expression of specific lipid-gene pairs such as PE 40:2 $[M+H]^+$ and *MSMB* in cancer samples. While the role of *MSMB* as a secreted protein with decreased expression in cancerous tissue is well established[72], the implications of a change in the abundance of correlating lipid currently remains unknown. Based on transcript – MSI feature pairs and snRNA-seq-based deconvolution of the entire cohort of samples our approach also discovered the presence of tumour cells in a histological region previously assigned as PIN and in a normal tissue section, demonstrating the ability of this approach to aid in pathological annotations. Taken together, these observations illustrate the potential of integrated spatial multi-omics approaches to reveal tissue heterogeneity and to generate new hypotheses as potential starting points for further investigation. Drawing biological conclusions from this concept study should be taken with care in view of the small number of samples that were measured and were selected based on morphological heterogeneity rather than for testing a specific hypothesis. One issue to be further resolved is the presence of batch effects[77], restricting the comparison of multiple samples to correlation analysis on each sample individually rather than collectively. While various strategies have been proposed to alleviate this issue in both ST and MSI[78], further method development is required to determine how to mitigate its effects.

As spatial technologies are rapidly evolving and are continuously gaining sensitivity, molecular coverage and spatial resolution, flexible approaches to integrate different layers of spatial and bulk molecular information will become indispensable. Integration approaches, such as the one presented here, will be crucial to better understand the complex interactions between different classes of molecules in a spatial cellular context and to apply this knowledge in diverse fields such as biomarker or target discovery, molecular pathology, and automated tissue annotation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

DATA ACQUISITION

Sample collection

Matched prostate tumour and normal samples were obtained from treatment naive high-risk PCa patients (advanced clinical stage and/or biopsy Gleason 8-10 and/or PSA ≥ 20 ng/mL) who underwent radical prostatectomy at University Hospitals Leuven, Belgium. This study was approved by the local Ethics Committee (approval number S54424) and followed the Code of Conduct for Responsible Use of Human Tissue for Research. All patients provided informed, signed consent. After surgery, 6 mm biopsy cores were extracted from the primary tumour site and a matching non-malignant biopsy was taken from the opposite side. The selection of biopsy cores was guided by a urologist and magnetic resonance imaging reports. Immediately after surgery, these biopsy cores were embedded in 3% carboxymethyl cellulose (CMC) (Sigma-Aldrich, cat.no C4888), snap-frozen in isopentane pre-chilled in liquid nitrogen, then stored at -80°C until further processing.

For each prostate biopsy, 10µm thick cryo-sections were obtained at -20ºC using a Microm HM525 NX cryostat (Thermo Scientific), followed by H&E staining. Subsequently, two independent expert pathologists performed a thorough examination to verify the presence of tumours in the tumour samples. In this study, a cohort of 8 patients was selected based on detailed pathological information including classification, ISUP grade and morphological features present (see **Supplementary Table 1** for patient-related information). Tissue blocks from the selected patients were subsequently checked for RNA quality. RNA was isolated from ten 10*µ*m sections from the selected tissue blocks using Direct-zol RNA Miniprep (Zymo Research, cat.no. R2050), and RIN score was determined using a Bioanalyzer RNA 6000 Nano kit (Agilent, cat.no 5067-1511) in combination with the Agilent Bioanalyzer software. All samples had a RIN > 7. 10µm thick sections were collected on ITO-coated glass slides (Delta Technologies, Loveland, USA) to enable matrix-assisted laser desorption/ionisation (MALDI) MSI analysis, and adjacent tissue sections of 10µm was mounted onto pre-equilibrated 10x Genomics Visium Gene Expression slides (10x Genomics, Pleasanton, CA, USA) for ST analysis. Finally, three to five tissue sections with a thickness of 100µm were collected in a pre-cooled centrifuge tube for subsequent single-nuclei RNA sequencing (snRNA-seq) analysis.

Spatial transcriptomics sample preparation and data acquisition

Tissue preparation for Visium ST analysis was performed according to the Tissue Preparation Guide (CG000240 rev C, 10x Genomics), except for embedding in 3% CMC which was conducted to stabilise the tissue for both ST and MSI analysis. Methanol fixation, H&E staining and imaging were performed according to the recommended protocol (CG000160 rev B, 10x Genomics), and subsequent cDNA and library preparation was performed according to the Visium Spatial Gene Expression Reagent Kit User guide (CG000239 rev D, 10x Genomics). The optimal tissue permeabilization time was determined upfront to be 20 minutes, following

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the Visium Spatial Gene Expression Reagent Kit - Tissue Optimization User guide (CG000238 rev A, 10x Genomics). Final libraries were sequenced to at least 50.000 reads/covered spot using a NextSeq2000 platform (Illumina, San Diego, CA, US).

Mass spectrometry imaging sample preparation and data acquisitions

Tissue sections were coated in 2,5-dihydroxybenzoic acid (DHB, AK Scientific, CA, USA) matrix using a sprayer robot (TM-Sprayer, HTX Technologies, LLC, Chapel Hill, NC, USA). DHB was dissolved in a 1:1 mixture of methanol:chloroform (v:v) at a concentration of 20 mg/mL. 12 layers of matrix solution were deposited using a flow rate of 0.12 mL/min, 10 psi of N2 pressure, nozzle temperature of 30 °C, and a 1200 mm/min velocity.
All data was acquired using an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific GmbH, Bremen, Germany) coupled to an intermediate pressure MALDI source (Spectroglyph LLC, WA, USA) as described previously[79]. A frequency tripled Nd:YLF laser (Explorer One, Spectra Physics, Mountain View, CA) emitting at 349 nm operating at 300 Hz was used for MALDI. The laser was operated using a diode current of 1.80 A and pulse energy fine-tuned using an external attenuator (PowerXP, Altechna, Vilnius, Lithuania) positioned immediately in front of the laser to yield a post-attenuation pulse energy of 1.2 μJ. A frequency quadrupled laser Nd:YAG laser emitting at 266 nm and operating at 300 Hz (NanoDPSS, Litron Lasers, Warwickshire, England) was used for laser post-ionisation (MALDI-2). This setup has been described in detail previously[80]. MALDI-2 laser pulse energy was adjusted using the internal attenuator to be 600 μJ just prior to entering the ion source. Laser pulse energy measured using a calibrated energy sensor (QE12HR-H-MB-D0, Gentec-EO, Quebec, Canada). Both lasers were externally triggered using a pulse/delay generator (QC9200, Quantum Composers, Bozeman, Montana) such that the time delay between the MALDI and MALDI-2 laser pulse was 20 μs. The mass spectrometer was operated in positive ion mode using an ion injection time of 250 ms, automatic gain control turned off, m/z range 350-2000 and a nominal mass resolution of 120,000 @ m/z 400. The pixel size was 30 μm.
All raw files were first recalibrated with Thermo Scientific Xcalibur RecalOffline using ion signals of $[\text{Vitamin E}]^+$ at *m/z* 430.38053, $[\text{PC(34:1)+K}]^+$ at *m/z* 798.54096 and a background PDMS peak at *m/z* 371.10124. Data was then converted to imzML format using Image Insight software (Spectroglyph LLC, Kennewick, WA, USA).

Tissue Staining and Histopathology Annotations

Tissue staining using H&E was conducted on tissue sections following MSI analysis. The matrix was first removed from the slide by rinsing in methanol. For the H&E staining, the glass slides were immersed in 95% ethanol, 70% ethanol and H_2O for 2 minutes each and then immersed in haematoxylin for 3 minutes. Slides were then washed with running tap water for 3 minutes and immersed in eosin for 30 seconds following another washing step with water for 3 minutes and immersion in 100% ethanol for 1 minute and then in xylene for 30 seconds. The tissues were air-dried and mounted with coverslips. Optical images were acquired using a ZEISS Axio Scan.Z1 Slide Scanner (Carl Zeiss AG, Jena, Germany) with 40x objective, pixel size 0.22 μm* 0.22 μm. The files were exported as .TIFF images.
The digitalized H&E images were annotated and verified by prostate cancer specialists from the Department of Pathology, University Hospitals Leuven and Weill Cornell Medical College,

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in a blinded fashion using the web-based annotation tool 'Annotation studio' (Aspect Analytics NV, Genk, Belgium). Pathologists annotated the whole slides in detail by placing 14 different labels 'Inflammatory infiltrate', 'Normal epithelium', 'Stroma', 'Necrosis', 'Atrophy', 'Cribriform', 'PIN', 'Nerve', 'Tumour ISUP 5', 'Tumour ISUP 4', 'Tumour ISUP 3', 'Tumour ISUP 2', 'Tumour ISUP 1', and 'Other'.

Single cell RNA sequencing sample preparation and data acquisition

Tissue sections (2-3x100 μm) adjacent to those used for spatial transcriptomics and lipidomics sections were collected in a microcentrifuge tube and subjected to single nuclei dissociation. For this, 0.3mL Tween with Salts and Tris (TST) buffer (96mM NaCl (Sigma-Aldrich, cat.no. 59222C), 10mM Tris-HCl pH 7.5 (ThermoFisher, cat.no. 15567027), 1mM CaCl2 (Merck, cat.no. 21115), 21mM MgCl2 (Sigma-Aldrich, cat.no M1028), 0.03% Tween-20 (Biorad, cat.no 1662404), 0.01% BSA (New England Biolabs, cat.no B9000S) and 0.2U/μl RNAsin (Promega, cat.no N2615)) were added to the sections and the tissue was cut into small pieces using scissors. The tissue was subsequently transferred to a KIMBLE Dounce tissue grinder (Sigma-Aldrich, cat.no D8938), together with 0.7mL TST that was used to wash the original tube. The tissue was homogenised by 10 strokes of the large clearance pestle and 10 strokes of the small clearance pestle. The homogenate was subsequently filtered using a 40μm cell strainer (pluriSelect, cat.no. 43-50040-51) and the tissue grinder and filter were washed with Salts and Tris (ST) buffer (96mM NaCl (Sigma- Aldrich, cat.no 59222C), 10mM Tris-HCl pH 7.5 (ThermoFisher, cat.no 15567027), 1mM CaCl2 (Merck, cat.no. 21115), 21mM MgCl2 (Sigma-Aldrich, cat.no M1028) and 0.2U/μl RNAsin (Promega, cat.no. N2615)). Filtered nuclei were pelleted twice for 5 minutes at 500g at 4ºC, first followed by resuspension in 500 μl ST buffer, and second followed by resuspension in 50μl of Chromium loading (CL) buffer (PBS + 0.04% BSA + 1U/μl Sigma Protector RNase Inhibitor (Millipore Sigma, cat.no. 3335399001). Nuclei were counted using a PhotonSlide™ (Logos Biosystems, South Korea) with Acridine Orange & Propidium Iodide Cell Viability assay (Westburg, cat.no. LB F23001) on a LUNA-FL™ Dual Fluorescence Cell Counter (Logos Biosystems). Nuclei were counted as the 'dead' cell fraction, and the suspension was diluted to ~1.000.000 nuclei/mL in CL buffer.
Single nuclei suspensions were used for single-nuclei gene expression analysis using the Chromium Next GEM Single Cell 3' Reagent Kits v3.1, according to the user guide (CG000315 Rev C). 7000-8000 nuclei were loaded per sample, and final libraries were sequenced to at least 25.000 reads/cell using a NextSeq2000 platform (Illumina, San Diego, CA, US).

DATA PREPROCESSING

Spatial transcriptomics data pre-processing

FASTQ-files were mapped to the GRCh38 human reference genome and spatially projected using Space Ranger v1.1.0 (10x Genomics). Before merging the datasets, ambient RNA was removed using an adjusted version of SpotClean v1.1.1 where all relevant genes are maintained[81]. Next Scanpy v.1.9.1 was used for quality control and processing[82]. Spots with less than 700 genes and 1000 reads were excluded, while genes were filtered for expression in more than 100 spots. The Scanpy toolkit was used to perform downstream processing per

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sample, including normalisation with the Scran R package (v1.26.1), log transformation and variable gene detection. The number of significant principal components for Leiden clustering and UMAP dimensionality reduction was determined using bootstrapping with a maximum of 30 components. Clustering resolutions ranging from 0.1-1.0 were assessed for stability with steps of 0.1 using clustree v0.4.2[83].

Spatial lipidomics data preprocessing

The mass spectrometry imaging data was preprocessed by performing TIC normalization, peak picking and subsequent rebinning. Peak picking was performed on the mean spectrum across all datasets. Peaks in the mean spectrum with an intensity below 0.005% of the base peak were discarded, resulting in a total of 12510 retained peaks. Intensities were binned per spectrum using a 5 ppm window around the selected peaks.

Spatial multi-omics integration pipeline

As the ST and MSI data were acquired from two serial sections, the spatial lipidomics and transcriptomics data were indirectly registered to each other by their respective H&E microscopy images. This strategy of co-registering multiple sections via one proxy section (i.e., microscopy) is inevitable in applications where multiple sections are involved. Before co-registration, representative images from the MSI data were generated for guiding the later registration step. These representative images were created via applying dimensionality reduction methods, such as NMF and UMAP. The MSI data was co-registered with its respective H&E microscopy images via a non-rigid registration workflow[84] (Aspect Analytics NV). Next, the MSI-slide H&E was co-registered to the H&E from the ST Visium slide. No additional co-registration is required to link the ST data and ST-slide microscopy as their spatial reference is inherent to the Visium ST data format. Finally, the MSI was directly linked with the ST data via a shared spatial coordinate system. MSI data was acquired at a spatial resolution (pixel size) of 30 μm, whereas the ST data was obtained with spot diameters of 55 μm. Multiple MSI pixels were aggregated into a single representative spectrum to match the ST spot via a Gaussian weighting algorithm[53].

Morphological feature extraction from microscopy data

Morphological features from high resolution H&E microscopy image of the MSI section were extracted using a pre-trained model that was adapted originally from the SimCLR model introduced by Chen et al[85]. The model tries to maximize the agreement between two stochastically augmented views of the same image via a contrastive loss function. Ciga et al. later applied this SimCLR model with minor modifications and pre-trained it on 57 multi-organ histopathology datasets without any labels[86]. These 57 histopathology datasets contained microscopy images with various types of staining and resolutions, which helped the model learn features with better quality. We applied this pre-trained model on high-resolution H&E-stained microscopy data, acquired at a 40x magnification, and extracted the morphological features using overlapping windows of 512 × 512-pixel patches of the image centred on each spot. Thus, each patch covers approximately 128×128 μm^2, which is similar in size to the

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

patches extracted from histopathology datasets that the model was pre-trained on by Ciga et al[86].

Dimensionality reduction via non-negative matrix factorization

We applied NMF to reduce the high dimensionality of our MSI and ST data. NMF has been widely used in both fields[87–90] as it has the non-negativity constraint and can generate interpretable parts-based representation. We used NMF to get a more in-depth understanding of each modality in a fully unsupervised manner, based on their resulting distinct spatial distributions and associated spectral or gene expressions. In short, NMF factorizes the matrix X into two non-negative matrices X=WH, where W represents the loading matrix and H is the score matrix which represents the data in lower dimensions. For example, in the case of MSI data, the lipid spectral signature of each pixel can be approximated by the additive linear combination of the column vector (spectral basis element) from W, weighted by H, which describes the contribution of each basis element to the spectrum at each pixel. We implemented NMF by the Python package sklearn.decomposition.NMF. As both MSI and ST data are known to follow Poisson distribution, we used Kullback–Leibler divergence (KL-NMF) as the cost function to incorporate Poisson noise. Multiplicative Update was used as the solver. The number of components was selected as 20.

Correlation analysis

We applied Pearson correlation by Python package numpy.corrcoef and scipy.stats.pearsonr to measure the relationship between two datasets. The output correlation coefficient ranges from -1 to 1 with 0 implying no correlation. Correlations of -1 or 1 represent an exact linear relationship (negative or positive).

Cohen's d

We calculated the Cohen's d value for gene-lipid correlations across different tissue types. Cohen's d is defined as the difference between two means divided by the pooled standard deviation for the data. The resulting Cohen's d values were ranked to find the most different gene-lipid correlations between normal and tumour tissue types. Using this approach, a set of changes in correlation were identified. We firstly calculated the Cohen's d value for gene-lipid correlations across different tissue types via computing their mean correlations in 8 tumour samples and 8 'normal' samples separately.

Processing of the single-nuclei RNA-sequencing data and cell type deconvolution

FASTQ-files were mapped to the GRCh38 human reference genome with Cell Ranger v6.1.2 (10x Genomics). Before merging the datasets, ambient RNA was removed using SoupX[91], while scDblFinder[92] was used to indicate doubles. Next Scanpy v.1.9.1 was used for quality control and processing[82]. Only cells with the following criteria were considered for further analysis: more than 500 uniquely expressed genes, less than 10% of the UMI counts mapping to mitochondrial sequences and less than 5% of the UMI counts assigned to ribosomal genes.

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

All genes that were not expressed in at least 100 cells were not considered in the downstream analysis. The Scanpy toolkit was used to perform downstream processing per sample, including normalisation, log transformation and variable gene detection. The number of significant principal components for Leiden clustering and UMAP dimensionality reduction was determined using bootstrapping with a maximum of 60 components. Clustering resolutions ranging from 0.1-1.0 were assessed for stability with steps of 0.1 using clustree v0.4.2[83]. Cell types and states were assigned to the clusters using marker genes obtained from the PanglaoDB[93] database (version of 27/03/2020), the canonical markers from Song *et al.*[63] and the NMF signatures obtained from the analysis of matching ST samples. Copy-number landscapes were inferred from the transcriptomic data with inferCNV[94] v.1.3.3 to validate clusters annotated as tumorigenic. Integration of the samples was performed in a semi-supervised manner using scANVI[95]. First the transcriptome data was integrated using scVI[96] (v.0.14.5, n_hidden=128, n_latent=50, n_layers=2, dispersion='gene-batch') with correcting batch effects and removing unwanted source of variations (total counts, percent mitochondrial genes, and percent ribosomal genes for 'continuous_covariate_keys'). After training the model for 400 epoch, scANVI using the cluster labels obtained at single-cell level, was trained for an additional 10 epochs. The integrated latent embedding generated by scANVI was used for downstream analysis (clustering and visualisation). For gene-level analyses, uncorrected counts were used. Spot deconvolution was performed using the integrated single-nuclei data per patient and Cell2Location[46]. The lipid metabolism gene sets of the KEGG pathway database[97] and the 'score genes' function of scanpy were used to asses pathway activities.

Lipid Annotation/Identification

Lipids were annotated with mass accuracy of 3 ppm based on the LipidMaps guidelines on shorthand notation for lipid structures published by Liebisch et al.[98,99] For the glycerophospholipids (GPL), the following shorthand notations of lipid classes were used: PC – phosphatidylcholines, LPC – lysophosphatidylcholines, PC O – phosphatidylcholine ethers, PC P – phosphatidylcholine plasmalogens, PE – phosphatidylethanolamines, LPE – lysophosphatidylethanolamines, PE O – phosphatidylethanolamine ethers, PE P – phosphatidylethanolamine plasmalogens, PS – phosphatidylserines, PI – phosphatidylinositols. For the sphingolipids (SPL), the following class name abbreviations were used: Cer – ceramides, SM – sphingomyelins, and Hex(n)Cer – hexosyl ceramides, where n refers to the number of hexosyl units. Cholesterol is referred to as 'Chol', cholesteryl esters as CE, acylcarnitines as CAR, monoacylglycerols as MG, and diacylglycerols as DG. Lipids are annotated at the lipid species level, namely the lipid class abbreviation is followed by a total number of carbons in fatty acyl(s) rest(s), colon, and total number of double bonds in fatty acyl rest(s). For example, PC 36:3 refers to phosphatidylcholine with a total number of carbons in both fatty acyls equal to 36 and total number of double bonds equal to 3. The same principle was used for annotation of signals corresponding to sphingolipids, but additionally the total number of oxygen was indicated after semi-colon, for example, SM 36:2;O2 refers to a sphingomyelin with sum composition of carbon atoms in N-linked fatty acyl rest and sphingoid base of 36, the number of double bonds equal to 2, and total number of O-atoms equal to 2. CE 16:0 means cholesteryl ester with palmitic acid attached to cholesterol via an ester bond. In MALDI-2 mode cholesteryl esters form dimers as showed recently by Bowman et al.[65] For m/z values annotated as dimers of cholesteryl esters, the following shorthand

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

notations were used: 2CE for cholesteryl ester dimer, followed by total number of carbons in both fatty acyls, colon, and total number of double bonds in both fatty acyls e.g., 2CE 38:4.

ACKNOWLEDGEMENT

We thank all patients that participated in this study.
WZ was supported by a Baekeland PhD grant from Flanders Innovation and Entrepreneurship (VLAIO) - HBC.20192204. XS is a recipient of a PhD fellowship (SB/1S49218N) from the Research Foundation – Flanders (FWO) and was in part supported by Opening the Future (OtF). SRE acknowledges support from the Australian Research Council (FT190100082). TS acknowledges support through the Australian Government Research Training Program Scholarship. SV and VdL were sponsored by an FWO PhD Followship strategic basic research (1S93320N and 1157918N, respectively). KV, SK and TV are supported by KU Leuven (IDN/19/039, Opening the Future (OtF)). This work was supported by a grant from the Belgian Foundation Against Cancer (STK), a Focus Group grant from the LKI Fund for Innovative Cancer Research (FIKO), the KU Leuven Opening the Future campaign, KU Leuven Incubation financing Lipometrix, a grant PRISMO from the Flemish Resilience Plan, FWO-EOS projects 30837538 (DECODE), G0F6718N (SeLMA), FWO-SBO projects S001623N (LIPOMACS) and S005319N, KU Leuven Research Fund projects C14/21/095, BOF/23/064, C16/15/059, C3/19/053, C24/18/022, C3/20/117, C3I-21-00316), US-DOD project W81XWH-19-1-0566, Industrial Research Fund (Fellowships 13-0260, IOFm/16/004) and several Leuven Research and Development bilateral industrial projects, Infrastructure projects TOP/23/014, IO12220N and I013218N, TBM Project T001919N; PhD Grant (SB/1SA1319N), the Flanders AI Research Program, VLAIO CSBO (HBC.2021.0076), Baekeland PhD (HBC.20192204), European Research Council under the European Union's Horizon 2020 research and innovation programme (ERC Adv. Grant grant agreement No 885682) and EU Interreg EMR23 EURLIPIDS, Stand up to Cancer / Kom op tegen Kanker - the Flemish cancer society, CM (Christelijke Mutualiteit), and private donations. Figure 1 was created with BioRender.com.

References

1. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647-1660.e19 (2019).
2. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976-991.e19 (2020).
3. Lee, S. *et al.* APOE modulates microglial immunometabolism in response to age, amyloid pathology, and inflammatory challenge. *Cell Rep.* **42**, 112196 (2023).
4. Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
5. Andersson, A. *et al.* Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions. *Nat. Commun.* **12**, 6012 (2021).

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

6. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma - PubMed. https://pubmed.ncbi.nlm.nih.gov/32579974/.
7. A single-cell and spatially resolved atlas of human breast cancers | Nature Genetics. https://www.nature.com/articles/s41588-021-00911-1.
8. Ravi, V. M. *et al.* Spatially resolved multi-omics deciphers bidirectional tumor-host interdependence in glioblastoma. *Cancer Cell* **40**, 639-655.e13 (2022).
9. Sun, C. *et al.* Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer. *Nat. Commun.* **14**, 2692 (2023).
10. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics - PubMed. https://pubmed.ncbi.nlm.nih.gov/27365449/.
11. Merritt, C. R. *et al.* Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
12. Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777-1792.e21 (2022).
13. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
14. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **183**, 1665-1681.e18 (2020).
15. In situ sequencing for RNA analysis in preserved tissue and cells | Nature Methods. https://www.nature.com/articles/nmeth.2563.
16. Spatially resolved, highly multiplexed RNA profiling in single cells - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4662681/.
17. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
18. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
19. Lundberg, E. & Borner, G. H. H. Spatial proteomics: a powerful discovery tool for cell biology. *Nat. Rev. Mol. Cell Biol.* **20**, 285–302 (2019).
20. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
21. Hickey, J. W. *et al.* Spatial mapping of protein composition and tissue organization: a primer for multiplexed antibody-based imaging. *Nat. Methods* **19**, 284–295 (2022).
22. Arsić, A., Hagemann, C., Stajković, N., Schubert, T. & Nikić-Spiegel, I. Minimal genetically encoded tags for fluorescent protein labeling in living neurons. *Nat. Commun.* **13**, 314 (2022).
23. Zhao, N. *et al.* A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nat. Commun.* **10**, 2947 (2019).
24. Caprioli, R. M., Farmer, T. B. & Gile, J. Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS. *Anal. Chem.* **69**, 4751–4760 (1997).
25. Wang, G. *et al.* Analyzing cell-type-specific dynamics of metabolism in kidney repair. *Nat. Metab.* **4**, 1109–1118 (2022).
26. Dewez, F. *et al.* MS Imaging-Guided Microproteomics for Spatial Omics on a Single Instrument. *Proteomics* **20**, e1900369 (2020).
27. Mund, A. *et al.* Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **40**, 1231–1240 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

28. Ellis, S. R. *et al.* Automated, parallel mass spectrometry imaging and structural identification of lipids. *Nat. Methods* **15**, 515–518 (2018).
29. Anderson, D. M. G. *et al.* Lipid Landscape of the Human Retina and Supporting Tissues Revealed by High-Resolution Imaging Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **31**, 2426–2436 (2020).
30. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.* **24**, 494–515 (2023).
31. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
32. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
33. Spatial Gene Expression. *10x Genomics* https://www.10xgenomics.com/products/spatial-gene-expression.
34. Porta Siegel, T. *et al.* Mass Spectrometry Imaging and Integration with Other Imaging Modalities for Greater Molecular Understanding of Biological Tissues. *Mol. Imaging Biol.* **20**, 888–901 (2018).
35. Köfeler, H. C. *et al.* Recommendations for good practice in MS-based lipidomics. *J. Lipid Res.* **62**, (2021).
36. Andersson, A. *et al.* Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 1–8 (2020).
37. Goossens, P. *et al.* Integrating multiplex immunofluorescent and mass spectrometry imaging to map myeloid heterogeneity in its metabolic and cellular context. *Cell Metab.* **34**, 1214-1225.e6 (2022).
38. Vicari, M. *et al.* Spatial Multimodal Analysis of Transcriptomes and Metabolomes in Tissues. 2023.01.26.525195 Preprint at https://doi.org/10.1101/2023.01.26.525195 (2023).
39. Soltwisch, J. *et al.* Mass spectrometry imaging with laser-induced postionization. *Science* **348**, 211–215 (2015).
40. Ellis, S. R., Soltwisch, J., Paine, M. R. L., Dreisewerd, K. & Heeren, R. M. A. Laser post-ionisation combined with a high resolving power orbitrap mass spectrometer for enhanced MALDI-MS imaging of lipids. *Chem. Commun.* **53**, 7246–7249 (2017).
41. Sutton, G. J. *et al.* Comprehensive evaluation of deconvolution methods for human brain gene expression. *Nat. Commun.* **13**, 1358 (2022).
42. Carm, K. T. *et al.* Interfocal heterogeneity challenges the clinical usefulness of molecular classification of primary prostate cancer. *Sci. Rep.* **9**, 13579 (2019).
43. van Leenders, G. J. L. H. *et al.* The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. *Am. J. Surg. Pathol.* **44**, e87–e99 (2020).
44. Butler, L. M. *et al.* Lipidomic Profiling of Clinical Prostate Cancer Reveals Targetable Alterations in Membrane Lipid Composition. *Cancer Res.* **81**, 4981–4993 (2021).
45. Mutuku, S. M. *et al.* Unravelling Prostate Cancer Heterogeneity Using Spatial Approaches to Lipidomics and Transcriptomics. *Cancers* **14**, 1702 (2022).
46. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

47. Salmén, F. *et al.* Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).
48. Walch, A., Rauser, S., Deininger, S.-O. & Höfler, H. MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology. *Histochem. Cell Biol.* **130**, 421–434 (2008).
49. Kaya, I. *et al.* Histology-Compatible MALDI Mass Spectrometry Based Imaging of Neuronal Lipids for Subsequent Immunofluorescent Staining. *Anal. Chem.* **89**, 4685–4694 (2017).
50. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426 (2020).
51. Smets, T. *et al.* Evaluation of Distance Metrics and Spatial Autocorrelation in Uniform Manifold Approximation and Projection Applied to Mass Spectrometry Imaging Data. *Anal. Chem.* **91**, 5706–5714 (2019).
52. What is Space Ranger? -Software -Spatial Gene Expression -Official 10x Genomics Support. https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger.
53. Van de Plas, R., Yang, J., Spraggins, J. & Caprioli, R. M. Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. *Nat. Methods* **12**, 366–372 (2015).
54. Masood, M., Grimm, S., El-Bahrawy, M. & Yagüe, E. TMEFF2: A Transmembrane Proteoglycan with Multifaceted Actions in Cancer and Disease. *Cancers* **12**, (2020).
55. Rubin, M. A. *et al.* alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *JAMA* **287**, 1662–1670 (2002).
56. Bu, H. *et al.* The anterior gradient 2 (AGR2) gene is overexpressed in prostate cancer and may be useful as a urine sediment marker for prostate cancer detection. *The Prostate* **71**, 575–587 (2011).
57. Voelkel-Johnson, C., Norris, J. S. & White-Gilbertson, S. Interdiction of Sphingolipid Metabolism Revisited: Focus on Prostate Cancer. *Adv. Cancer Res.* **140**, 265–293 (2018).
58. Qi, X., Wang, Y., Hou, J. & Huang, Y. A Single Nucleotide Polymorphism in *HPGD* Gene Is Associated with Prostate Cancer Risk. *J. Cancer* **8**, 4083–4086 (2017).
59. alpha-Methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer - PubMed. https://pubmed.ncbi.nlm.nih.gov/11926890/.
60. Guo, Y., Ren, C., Huang, W., Yang, W. & Bao, Y. Oncogenic ACSM1 in prostate cancer is through metabolic and extracellular matrix-receptor interaction signaling pathways. *Am. J. Cancer Res.* **12**, 1824–1842 (2022).
61. Shrestha, R. *et al.* ACSM1 and ACSM3 regulate prostate cancer fatty acid metabolism to promote tumour growth and constrain ferroptosis. 2022.10.13.511039 Preprint at https://doi.org/10.1101/2022.10.13.511039 (2022).
62. Henry, G. H. *et al.* A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep.* **25**, 3530-3542.e5 (2018).
63. Song, H. *et al.* Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat. Commun.* **13**, 141 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

64. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses | Nature Communications. https://www.nature.com/articles/s41467-023-36325-2.
65. Bowman, A. P. *et al.* Evaluation of lipid coverage and high spatial resolution MALDI-imaging capabilities of oversampling combined with laser post-ionisation. *Anal. Bioanal. Chem.* **412**, 2277–2289 (2020).
66. Ma, R.-Y., Black, A. & Qian, B.-Z. Macrophage diversity in cancer revisited in the era of single-cell omics. *Trends Immunol.* **43**, 546–563 (2022).
67. Alasoo, K. *et al.* Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription. *Sci. Rep.* **5**, 12524 (2015).
68. Roca, H. *et al.* Apoptosis-induced CXCL5 accelerates inflammation and growth of prostate tumor metastases in bone. *J. Clin. Invest.* **128**, 248–266 (2018).
69. Cai, C. *et al.* Intratumoral de novo steroid synthesis activates androgen receptor in castration-resistant prostate cancer and is upregulated by treatment with CYP17A1 inhibitors. *Cancer Res.* **71**, 6503–6513 (2011).
70. Xu, M. *et al.* Arachidonic Acid Metabolism Controls Macrophage Alternative Activation Through Regulating Oxidative Phosphorylation in PPARγ Dependent Manner. *Front. Immunol.* **12**, 618501 (2021).
71. Sawilowsky, S. S. New Effect Size Rules of Thumb. *J. Mod. Appl. Stat. Methods* **8**, 597–599 (2009).
72. Dahlman, A. *et al.* Evaluation of the prognostic significance of MSMB and CRISP3 in prostate cancer using automated image analysis. *Mod. Pathol.* **24**, 708–719 (2011).
73. Bergström, S. H., Järemo, H., Nilsson, M., Adamo, H. H. & Bergh, A. Prostate tumors downregulate microseminoprotein-beta (MSMB) in the surrounding benign prostate epithelium and this response is associated with tumor aggressiveness. *The Prostate* **78**, 257–265 (2018).
74. Tuck, M. *et al.* Multimodal Imaging Based on Vibrational Spectroscopies and Mass Spectrometry Imaging Applied to Biological Tissue: A Multiscale and Multiomics Review. *Anal. Chem.* **93**, 445–477 (2021).
75. Palla, G., Fischer, D. S., Regev, A. & Theis, F. J. Spatial components of molecular tissue biology. *Nat. Biotechnol.* **40**, 308–318 (2022).
76. Jones, A., Townes, F. W., Li, D. & Engelhardt, B. E. Alignment of spatial genomics data using deep Gaussian processes. *Nat. Methods* 1–9 (2023) doi:10.1038/s41592-023-01972-2.
77. Balluff, B., Hopf, C., Porta Siegel, T., Grabsch, H. I. & Heeren, R. M. A. Batch Effects in MALDI Mass Spectrometry Imaging. *J. Am. Soc. Mass Spectrom.* **32**, 628–635 (2021).
78. Deconvolution Tactics and Normalization in Renal Spatial Transcriptomics - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8793484/.
79. Design and Performance of a Novel Interface for Combined Matrix-Assisted Laser Desorption Ionization at Elevated Pressure and Electrospray Ionization with Orbitrap Mass Spectrometry - PubMed. https://pubmed.ncbi.nlm.nih.gov/28613836/.
80. Sarretto, T. *et al.* Selective Mass Spectrometry Imaging of Aromatic Antioxidants Using Sequential Matrix-Assisted Laser Desorption and Resonant Photoionisation. *Anal. Sens.* **2**, e202100052 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.08.28.555056; this version posted August 28, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

81. Ni, Z. *et al.* SpotClean adjusts for spot swapping in spatial transcriptomics data. *Nat. Commun.* **13**, 2971 (2022).
82. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
83. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* **7**, giy083 (2018).
84. Verbeeck, N. *et al.* Automated Anatomical Interpretation of Ion Distributions in Tissue: Linking Imaging Mass Spectrometry to Curated Atlases. *Anal. Chem.* **86**, 8974–8982 (2014).
85. Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. in *Proceedings of the 37th International Conference on Machine Learning* 1597–1607 (PMLR, 2020).
86. Ciga, O., Xu, T. & Martel, A. L. Self supervised contrastive learning for digital histopathology. *Mach. Learn. Appl.* **7**, 100198 (2022).
87. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).
88. SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes - PubMed. https://pubmed.ncbi.nlm.nih.gov/33544846/.
89. Verbeeck, N., Caprioli, R. M. & Van de Plas, R. Unsupervised machine learning for exploratory data analysis in imaging mass spectrometry. *Mass Spectrom. Rev.* **39**, 245–291 (2020).
90. Nijs, M., Smets, T., Waelkens, E. & De Moor, B. A mathematical comparison of non-negative matrix factorization related methods with practical implications for the analysis of mass spectrometry imaging data. *Rapid Commun. Mass Spectrom. RCM* **35**, e9181 (2021).
91. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* **9**, giaa151 (2020).
92. Doublet identification in single-cell sequencing data using scDblFinder - PMC. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9204188/.
93. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data - PubMed. https://pubmed.ncbi.nlm.nih.gov/30951143/.
94. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
95. Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol. Syst. Biol.* **17**, e9620 (2021).
96. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
97. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
98. Liebisch, G. *et al.* Shorthand notation for lipid structures derived from mass spectrometry. *J. Lipid Res.* **54**, 1523–1530 (2013).
99. Liebisch, G. *et al.* Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures. *J. Lipid Res.* **61**, 1539–1555 (2020).