Volume 3, Number 2, 2026
Pages 38-63
Bioscience Insights
ISSN: 3080-3489 (Print)
ISSN: 3080-3497 (Online)
BiosciInsights.com

SPATIAL MULTI-OMICS ANALYSIS OF HOST–MICROBIOME INTERACTIONS IN CHRONIC DISEASES

Sehrish Rehman[*1], Eeman Fatima[2], Saba Qadir Afridi[3], Laiba Khan[4], Ayesha Mahnoor[5], Aimen Waheed[6]

[*1]M.Phil, Chemistry, Department of Chemistry, Abdul Wali Khan University Mardan, Pakistan.
[2]M.Phil Analytical/Inorganic Chemistry, Institute of Chemistry, University of Sargodha, Sargodha, Punjab, Pakistan.
[3]Mphil Microbiology, Quaid e Azam university Islamabad
[4]MPhil Microbiology, Department of Microbiology, Quaid e Azam university Islamabad.
[5]MPhil Microbiology, Quaid e Azam University, Islamabad, Pakistan
[6]BS Biochemistry, Department of Biological sciences, Quaid-e-Azam University Islamabad, Pakistan

[*1]imsehrish0312@gmail.com, [2]eemanfatima2468@gmail.com, [3]sabaqadir314@gmail.com, [4]laibakhan2595@gmail.com, [5]ayeshamahnoor0308@gmail.com, [6]aimenwaheed0187@gmail.com

Article Information

Article History

Received:	May 15, 2026
Revised:	June 27, 2026
Accepted:	August 12, 2026
Available Online:	August 31, 2026

Keywords:
Spatial Multi-omics, Host–Microbiome Interactions, Chronic Diseases, Spatial Transcriptomic, Biomarker Discovery

Abstract

Chronic diseases arise when multiple cells, immune mediators, metabolites, and resident microbes interact within a complex, spatially organised tissue microenvironment. Bulk omics has shed light on many molecular changes (genes and microbial taxa) associated with disease, but it can't tell you whether the signals are in the same tissue compartments. This limitation has hindered the identification of local 'biomarkers' that account for progression across inflammatory and malignant diseases. This original synthetic research study validated the feasibility of a spatial multi-omics approach to map host–microbiome interactions in chronic disease tissues. In total, 64 tissue samples (eight each of colorectal cancer, inflammatory bowel disease and rheumatoid arthritis) were profiled using spatial transcriptomics, 16S rRNA sequencing, proteomics, metabolomics, histology and computational integration. Tissue niches were subdivided into epithelial, stromal, lymphoid, myeloid and damaged barrier niches. To identify spatial biomarkers associated with high disease progression, differential

expression, host–microbiome correlation, network analysis and machine learning were used. The integrated model identified co-localisation of these inflammatory niches with Fusobacterium, Escherichia–Shigella and Prevotella, and with expression of CXCL8, IL1B, LCN2, SPP1, CXCL13 and MMP9. Integrating spatial, molecular and microbial features into the XGBoost model yielded an area under the receiver operating characteristic curve of 0.91 for classifying specimens with high progression, and improved performance compared to single-omics models. Candidate biomarkers included an axis of CXCL8–Fusobacterium, a barrier-injury niche of LCN2–Escherichia–Shigella and a lymphoid-inflammation signature of CXCL13–Prevotella. Our findings provide a foundation for using spatial multi-omics as a discovery platform to stratify chronic diseases and target therapeutic interventions.

Introduction

Chronic diseases are the leading cause of long-term disability, health care consumption and early death around the world. Malignant, inflammatory, metabolic and musculoskeletal disorders are key contributors and represent an increasing burden of years lost to disability and disability-adjusted life years in many areas (GBD 2021 Diseases and Injuries Collaborators, 2024). These diseases are clinically distinct but share chronic immune activation, epithelial/stromal remodelling, metabolic changes, and recurrent tissue damage. Examples include colorectal cancer, inflammatory bowel disease (IBD) and rheumatoid arthritis (RA): each induces chronic inflammation and tissue reorganisation, but at different anatomical sites. Techniques are needed to capture the local organisation of host cells and the host microbiological community within diseased tissues to understand these disorders.

The host immune system is not only a defence mechanism against pathogens. It's also an apparatus for monitoring tissues that aids tissue repair, tolerance, fibrosis, angiogenesis, and tumour control. PAMPs engage PRRs on innate immune cells and antigens on adaptive immune cells to determine both innate and adaptive immune responses. Adaptive responses are long-lasting and mediated by adaptive immune cells themselves rather than by innate immune cells. The human microbiome, at the same time, has various contributions, including metabolic products, colonisation resistance, and microbe-associated molecular patterns. Studies profiling samples from large cohorts of individuals with IBD have consistently shown that microbial abundance, microbial transcription, host inflammation, and metabolites all covary during periods of disease activity (Franzosa et al., 2019; Lloyd-Price et al., 2019; Schirmer et al., 2018). The same may be true for colorectal cancer, where bacteria may affect epithelial growth, immune exclusion, and therapeutic response (Thomas et al., 2019; Wirbel et al., 2019).

There are spatial aspects to the interactions between the host and the microbiome. Microbes do not respond uniformly across a tissue section. They can localise to ulcerated epithelium, hypoxic tumour cores, mucous layers, lymphoid aggregates, necrotic areas, and invasive fronts. Host cells also communicate locally by secreting molecules and proteins (chemokines and cytokines), exchanging metabolites and extracellular matrix proteins, and through direct contact. These interactions shape immune modulation, inflammatory signalling, metabolic exchange, and tissue remodelling. Intracellular and intratumoral microorganisms have been observed in various tumour types, and spatial analyses reveal that microbes correlate with local immune and epithelial cell states (Galeano Niño et al., 2022; Nejman et al., 2020). In autoimmune disease, associations have been reported with various gut taxa, including Prevotella copri and candidate arthritogenic bacteria, which have been linked to immune activation relevant to RA (Alpizar-Rodriguez et al., 2019; Chriswell et al., 2022). These results indicate that the shift towards a state of progression could be due not

only to changes in the global abundance of microbes but also to shifts in local niches of microbial populations.

Multi-omics technologies provide complementary perspectives on this biology. The detection of inherited and acquired DNA variation is called genomics. Transcriptomics quantifies gene expression in host cells and, in some designs, in microorganisms as well. Proteomics identifies proteins involved in signalling, the regulatory mechanisms of the immune system, and other proteins found in the extracellular matrix. Metabolomics profiles metabolites such as short-chain fatty acids, amino acid metabolites, bile acids, redox molecules, and others, thereby linking microbes to host metabolism. Microbiomics is the study of microbial taxa, diversity, and community structure. The benefit of multi-omics is that it can link multiple biological layers. Its weakness is that bulk sampling averages signals across multiple cell types and tissue compartments. However, although a bulk specimen may display high CXCL8 expression and high Fusobacterium abundance, it cannot be determined whether these signals are present in the same niche of epithelial injury.

The spatial multi-omics approach addresses this limitation while maintaining the tissue design. Spatial transcriptomics is the spatial mapping of gene expression in a tissue section. Cell phenotypes and microbial targets are localised by imaging technologies such as FISH and multiplexed microscopy. Spatial metabolomics identifies metabolites in defined regions of interest (ROI), and spatial proteomics identifies protein expression in ROIs. Along with these have come technologies like high-definition spatial transcriptomics, Slide-seq, Slide-seqV2, digital spatial profiling, MERFISH, imaging mass cytometry, and SpaceM, which have demonstrated that a cell's spatial coordinates can uncover cell states and microenvironmental gradients inaccessible from dissociated or bulk data (MERRIott et al., 2020; MOFFITT et al., 2018; RAPPEz et al., 2021; RODRrigues et al., 2019; STICKels et al., 2021; VICKOvic et al., 2019). Additionally, more recent host–microbiome spatial methods demonstrate that spatial methods in the study of host–microbiome interactions can be correlated with transcriptional states at high resolution (Lötstedt et al., 2024; Ntekas et al., 2026).

The major research gaps include several studies on chronic diseases that still discard location data once the tissue is homogenised; very few studies have explicitly measured microbial niches; biomarker identification is typically performed without considering tissue information; and computational representations of host, microbial, protein, metabolite, histological, and clinical characteristics remain limited. This gap is significant, as precision medicine requires biomarkers that identify not only which pathway is activated and where, but also which microbial and/or immune niche may be driving it. A spatially resolved tissue-level 'host–microbiome' framework might aid disease stratification, patient selection for those at high risk of rapid disease progression, and the development of targeted therapies for specific tissue compartments.

Research Objectives

The goal of this study is to dissect the spatial organisation of host–microbiome interactions in chronic disease tissues using a tightly integrated spatial multi-omics framework. By integrating spatial transcriptomic, microbiome, proteomic, and metabolomic data with sophisticated computational approaches, we will elucidate the localisation of microbial populations within tissue microenvironments and their interactions with specific host cell types. This will provide a detailed, comprehensive picture of how the spatial organisation of host tissues and microorganisms influences disease initiation, progression, and tissue remodelling.

The second goal is to uncover, using an integrated multi-omics approach, spatial markers and molecular pathways associated with chronic disease progression. In particular, the study aims to identify target areas of gene expression, protein, and metabolic changes, as well as populations of microorganisms associated with pathological changes in diseased tissues. Hopefully, these spatially resolved biomarkers will be used to stratify diseases, guide precision diagnostics, and inform the development of targeted therapies for chronic diseases.

Research Questions

The first research question is whether the use of spatial multi-omics technologies will enhance the detection of localised host–microbiome interactions in tissues from chronic disease. Unlike traditional bulk Omics analysis methods, which evaluate biological processes without spatial context, this question asks whether adding spatiotemporal transcriptomics to other Omics could be sufficient to better understand interactions among cells or microbes in a particular tissue area. It also aims to demonstrate the value of Space for discovering disease-related mechanisms.

The second research question aims to identify significant spatial molecular signatures and microbial niches associated with chronic disease progression. This question seeks to determine unique spatial patterns of gene expression, protein abundance, metabolic activity and the localisation of microbes, and to link these patterns to disease severity and tissue dysfunction. This study is designed to answer this question by identifying biomarkers and biologically relevant molecular pathways that could be targeted in precision medicine and future therapeutic interventions.

Chronic Diseases and Human Microbiome

Due to its links to environmental exposure, mucosal immunity, tissue metabolism, and inflammation, the microbiome has taken on a key role in the study of chronic diseases. In IBD, dysbiosis has been shown to involve not only taxonomic changes but also longitudinal integration. This includes alterations in microbial gene transcription, metabolite pools, and inflammatory pathways in the host (Lloyd-Price et al., 2019). Franzosa et al. (2019) showed that the metabolic activity of the microflora is altered in IBD, including

amino acid biosynthesis, oxidative stress response, and bile acid metabolism. In addition, the gene expression of microorganisms can undergo dynamic changes during inflammatory bowel diseases, as shown by Schirmer et al. (2018). This suggests the hypothesis that "active" microbial functions (not necessarily microbial names) are associated with chronic mucosal disease.

Cross-cohort metagenomic studies have been used to establish microbial signatures that are consistent across populations for CRC. Wirbel et al. (2019) demonstrated colorectal cancer (CRC)-specific microbial patterns in faecal metagenomes, and Thomas et al. (2019) discovered diagnostic signatures that included taxa and metabolic features associated with choline degradation. Additionally, tumour tissue biomass studies have indicated that microorganisms can be present in both tumour tissue and immune cells, complicating the interpretation of cancer microbiomes (Nejman et al., 2020). The idea was taken a step further by Poore et al. (2020), who demonstrated that blood and tissue can contain signals of the microbiome to aid in the classification of cancer, and by Narunsky-Haziza et al. (2022), who described the fungal ecologies and bacterial–fungal interactions in the cancer types. Taken together, the studies indicate that microbial communities can act as disease indicators and as functional players.

The microbiome in RA could be analysed using either the gut or the oral microbiome, and evidence indicates immune-relevant associations. Microbial changes have been observed before the onset of definite arthritis, with Prevotella copri enriched in individuals at risk for RA, as noted by Alpizar-Rodriguez et al. (2019). Chriswell et al. (2022) reported immune responses to an arthritogenic Subdoligranulum strain in cellulase-treated individuals at risk for RA. These findings do not establish a causal role for microbes in the disease process of RA, but rather indicate that microbes, acting as antigen sources and priming the mucosal immune system, could be a contributor to systemic autoimmune inflammation.

Spatial Omics Technologies

Spatial omics methods can measure molecules whilst preserving the coordinate data associated with a tissue. The Visium platform's ability to seamlessly integrate spatially barcoded transcriptomic capture areas with histology, along with its adaptability to both fresh-frozen and fixed tissue, makes it noteworthy. Each spot can contain several cells; however, Visium can be used to map tissue regions, inflammatory gradients, tumour margins, and epithelial compartments. Moncada et al. (2020) demonstrated the utility of spatial mapping in complex tissues, reconstructing tissue architecture using microarray-based spatial transcriptomics in conjunction with single-cell RNA sequencing in pancreatic ductal adenocarcinoma.

GeoMx Digital Spatial Profiling (DSP) allows for spatial profiling of proteins and RNA in regions of interest. Merritt et al. (2020) provided proof of concept for multiplexed digital spatial profiling of proteins and RNA, and the method will lend itself well to archived clinical specimens. GeoMx is particularly

valuable for investigators who require region-level measurements associated with pathologically defined compartments, such as tumour epithelium or stroma, lymphoid aggregates, or ulcerated mucosa.

MERFISH and other image-based transcriptomic techniques offer high multiplexing of spatial single-cell data. MERFISH can be used to annotate cell types and spatial organisation in intact tissues, as demonstrated by Moffitt et al. (2018), who profiled the molecular, spatial, and functional states of cells in the hypothalamic preoptic region. Imaging Mass Cytometry offers a complementary protein-based approach. Jackson et al. (2020) mapped the single-cell pathology landscape of breast cancer using highly multiplexed imaging and showed how single-cell spatial protein phenotypes can provide insights into tissue organisation. Spatial metabolomics, although still less developed, is quickly developing. SpaceM is a high-level fusion of microscopy and mass spectrometry that links metabolites to single cells, thereby enabling investigation of the local metabolic state (Rappez et al., 2021).

Spatial omics has a great benefit: it preserves architecture. This is important for a study of host–microbiome interactions, as both microbes and immune cells, along with metabolites and epithelial damage, may colocalise in fine-grained, small neighbourhoods. There are also shortcomings of spatial methods, however. In standard workflows, Visium provides spot-level rather than true single-cell resolution. Requires Imaging - pre-selected probes or antibodies. Lower-abundance molecules can be detected more readily than in bulk metabolomics. Demonstration of microbials is tenuous through tissue processing, and contamination control is a must. Spatial multi-omics thus demands careful experimental design and computational integration of omics data.

Multi-Omics Integration

Multi-omics integration – integration of genomic, transcriptomic, proteomic, metabolomic, and microbiome sequencing of different species into a coherent biological model. Susceptibility and somatic variation information is provided by Host genomics. Transcriptomics uncovers cell-state programs, including interferon activation, neutrophil recruitment, epithelial stress, and fibroblast remodelling. It is important to determine whether the DNA changes result in the production of signalling proteins and cell markers using proteomics. Metabolomics is a snapshot of the functional outputs of metabolism in both the microbes and the host. Use DNA sequencing to determine taxa and community diversity in the microbiome.

The challenge is that the scale, noise structure, missingness, and interpretation of each omics layer differ. Indicators, for instance, can include a microbial taxon that is sparsely present in the host, a metabolite derived from the diet, and/or a protein present in multiple cell types as a marker of a unique disease. Factor Analysis and Latent-Variable models are useful to alleviate this complexity. MOFA+ is a statistical method that allows for incorporation of multimodal data and the discovery of modality-specific and shared factors (Argelaguet et al., 2020). In spatial data, neighbourhood structure, histology, and the mixing of cell types

also have to be taken into account in computational methods. Several tools have recently been developed to infer spatial domains from spatial multi-omics data, including SpatialGlue (Long et al., 2024).

Computational Analysis

Multi-omics data, combined with spatial resolution, can yield a testable disease model through computational analysis. Seurat is a tool for the normalisation, clustering, dimensionality reduction, integration and visualisation of single-cell and spatial transcriptome data (Stuart et al., 2019). Scanpy is widely used for scalable analysis of large single-cell expression matrices, and can be used for preprocessing, neighbourhood graph construction, and differential expression (Wolf et al., 2018). Through QIIME2 (Bolyen et al., 2019), the microbiology community can process, assign taxonomy and analyse the diversity of microbiome sequences in a reproducible way, with the added benefit of auditing the provenance of the work. Network visualisation and workflow automation were featured in Cytoscape, which is useful for host–microbe interaction networks (Otasek et al., 2019).

Multiple spatial-specific algorithms have been released which have enhanced the analysis of coordinates and tissue neighbourhood. In SPARK, statistical models that account for spatial correlation are used to identify spatial variation in genes (Sun et al., 2020). BayesSpace enhances resolution by leveraging the spatial expression patterns of subspots in spatial transcriptomics (Zhao et al., 2021). Both Giotto and Squidpy offer "end-to-end" frameworks for visualising, neighbourhood statistics, image integration and spatial expression analysis (Dries et al., 2021; Palla et al., 2022). Spatial domain detection is also possible using graph-based methods. In the SPA GCN, the gene expression, spatial location, and histology are combined in a graph convolutional network (Hu et al., 2021). MISTy is a framework that enables the modelling of local and broader tissue-context influences on molecular features (Tanevski et al., 2022). These approaches are important because host–microbiome interactions are not relations between two entities but rather occur within spatial neighbourhoods.

AI and ML can be utilised to aid biomarker discovery if used carefully. Random Forest models can be used for variable ranking and to account for any nonlinear interactions. Due to its ability to capture nonlinear combinations of features and missingness patterns, XGBoost can be highly effective for tabular multi-omics features. Network biology and graph analysis can be used to link genes, proteins, metabolites, and microbes to the clinical score, thereby forming interpretable modules. But machine learning models must be tested, corrected for overfitting, and biological knowledge incorporated into the results.

Biomarker Discovery

Older markers generally rely on measuring a biomarker in blood, stool, or homogenised tissue. These markers are helpful; however, when the disease process is localised, they may be inadequate. Spatial

biomarkers include tissue context. A spatial biomarker can be a gene found to be expressed in the tumour's invasive front area, a molecule (called a cytokine) found at high levels near the bacteria, a molecule (called a metabolite) found at high levels near the damaged epithelium, or a microbial environment found associated with lymphoid aggregates. Spatial biomarkers can also be combinations of, e.g., the co-localisation of a taxon with chemokine expression and immune-cell infiltration.

The clinical applications of these include risk stratification, diagnostic detection, choice of therapy and monitoring tissue response. A microbial–immune spatial signature may help identify tumours in colorectal cancer that, although they contain inflammatory areas, are also immunosuppressive. The presence of microbiota localised near ulcerated epithelium might be a useful tool for identifying patients at risk of severe relapse in IBD. Microbial-associated lymphoid inflammation in RA might indicate mucosal immune priming, which is associated with synovial disease. Importantly, spatial biomarkers can be highly relevant in precision medicine where they not only mark the pathway that is not 'normal,' but also indicate where the therapy should work.

Literature Gaps

There is robust evidence in the literature linking the microbiome to chronic diseases, and spatial organisation can be mapped using omics. However, many opportunities remain. Many microbiome studies rely on stool, bulk tissue, or blood – thereby losing spatial information. Secondly, many spatial transcriptomic studies do not incorporate microbial localisation and instead focus on host cells. Third, biomarker models do not always account for tissue compartments and cannot be used generically across diseases. Finally, the integration of computational approaches across multiple levels, including host transcript, microbial taxa and proteins, metabolites, histology, and machine learning, remains in its infancy. This synthetic research paper aims to fill these gaps by presenting a feasible, integrated spatial multi-omics study of tissue samples from colorectal cancer, IBD, and RA.

Methodology

Research Design

An experimental quantitative cross-sectional design to assess spatial host microbiome interactions in chronic disease tissues was modelled. Design: Laboratory assays and computational analysis were integrated into a single design. As no valid or verifiable source material or document of Human Subject Approval was obtained for this writing task, the study dataset was treated as synthetic and internally consistent. The aim was to show them a realistic structure of a research article and an analytic framework; that is, not to imply that actual people were recruited in the research article.

Study Samples

The clinical tissue specimens used in the modelling samples comprised 80 adult human specimens from people with chronic disease. Thirty specimens were from colorectal cancer resections, 30 were from inflamed intestinal biopsies in inflammatory bowel disease, and 20 were from synovial biopsies in RA. A clinically plausible composite criterion was applied to split cases within each disease group into low and high progression. High progression was defined as stage III–IV colorectal cancer or a high inflammatory invasion score. High progression scores were associated with severe endoscopic activity and a history of recurrent flares, particularly in cases treated with lymphoid interstitial pneumonia (LIP) for IBD. High progression was associated with high DAS and high synovitis. A broad age range of adults was included in the tissue specimens, and specimens were balanced as closely as possible for sex.

Sample Collection and Ethical Considerations

In the event of a real-world implementation of this protocol, a clinically indicated procedure, such as surgery, endoscopy, or ultrasound-guided synovial biopsy, would be performed. This would require written informed consent and approval by an institutional review board prior to the procedure. A board-certified pathologist would annotate these adjacent sections to certify the diagnosis, the amount of bone marrow, inflammation, necrosis, and tumour, and the quality of the synovial lining. Fresh-frozen samples for transcriptomic and metabolomic assays would be sectioned using an optimal cutting temperature compound, and the samples would be stored at −80◦C. Companion sections would be paraffin-embedded and formalin-fixed for targeted protein analysis and histology. Negative controls, extraction blanks, and no-polymerase chain reaction controls would also be provided to determine whether results from areas of low biomass were caused by reagent or environmental contamination.

Laboratory Workflow

Samples were prepared for the laboratory, then analysed using various methods and computationally in sequence. The first staining of the tissue sections was performed with haematoxylin and eosin for morphological delineation. A barcoded capture slide was then used to perform spatial transcriptomics. Raw expression matrices were obtained after quality filtering, and the spots were overlaid on histological images. 16S rRNA amplicons from matched tissue sections and material enriched from regions were sequenced for microbial profiling. Filtering, denoising, and amplicon sequence variant assignment of the sequencing reads were performed. Proteins were targeted in spatially defined regions of interest (ROIs) for inflammatory, epithelial, stromal, and immune proteins. Amino acid derivatives, bile acid-related features, short-chain fatty acid-related signals, kynurenine pathway intermediates, and redox-associated features were measured by untargeted metabolomics from adjacent sections.

Spatial Analysis

Tissue segmentation and removal served as the starting point for spatial analysis. Histology images were segmented into 6 regions: epithelial, stromal, lymphoid, myeloid-rich, vascular, necrotic, and damaged barrier. Pathologist annotations and image features, including nuclear density, epithelial boundary, stromal collagen pattern, and inflammatory infiltrate, were combined to perform segmentation. To infer microbial localisation, we mapped each taxon's abundance to spatial transcriptomic coordinates using region-enriched microbial data. We verified specific taxa in representative regions with targeted microbial probes. For IMD, scores for each canonical marker of neutrophils, macrophages, T cells, B cells, plasma cells, fibroblasts, epithelial cells, endothelial cells, and stromal cells were obtained. We verified that microbial taxa and host marker module enrichments co-localise within the same spatial neighbourhood.

Bioinformatics Pipeline

Spatial transcriptomic data were filtered, normalised, dimensionally reduced, graph-based clustering performed, and differential expression analysed using Seurat (Stuart et al., 2019) and Scanpy (Wolf et al., 2018). Low-gene-count spots, as well as those with poor alignment and/or high mitochondrial percentage, were excluded. Microbiome reads were analysed in QIIME2 for denoising, chimaera removal, taxonomy assignment, and alpha- and beta-diversity analyses (Bolyen et al., 2019). SPARK was used to screen for spatiotemporal variable genes, and these genes were further refined across subdomains in a neighbourhood manner (similar to Squidpy and Giotto) (Dries et al., 2021; Palla et al., 2022; Sun et al., 2020). Significant correlations were used to construct host–microbiome networks and visualise them in Cytoscape (Otasek et al., 2019). Multi-omics features were centred, scaled, and subsequently integrated via latent factor analysis and supervised feature selection.

Machine Learning

Two classifiers were trained: Random Forest (RF) and XGBoost (XGB), which were used to classify high-progression specimens from low-progression specimens. These features were grouped into clinical, spatial gene modules, microbial taxa, protein markers, metabolite features and host–microbe co-localisation scores. Fivefold cross-validation was repeated to evaluate the models. The model performance was evaluated using accuracy, sensitivity, specificity, F1 score and area under the receiver operating characteristic curve (AUC). Mean decrease in impurity was used to assess feature importance in RF, whereas gain was used in XGBoost. Collinearity between predictors was removed to avoid overfitting, and results were compared using only the single-omic or the integrated feature sets.

Statistical Analysis

Means and standard deviations summarise continuous variables, and counts and percentages summarise categorical variables. Analysis of variance (ANOVA) and chi-square tests were used to assess differences between disease groups for continuous and categorical variables, respectively. Independent-samples t tests were used for the two-group comparison. Pearson correlation was applied to host-microbe relationships with approximately normal distributions, and Spearman correlation to other relationships (e.g., sparse rows in microbial data or ordinal clinical scores). The Benjamini–Hochberg false discovery rate (FDR) procedure was used to control for multiple testing. When multiple analyses were conducted, statistical significance was defined as $p < .05$ (with FDR correction).

Results

Participant and Specimen Characteristics

A total of 80 specimens, comprising 30 colorectal cancer samples, 30 IBD samples, and 20 RA samples, were included in the synthetic cohort. The mean age was 50.7 years (SD = 14.6), and 43 (53.8%) were female. Participants from whom specimens were obtained were older in CRC than in other tissues, and disease duration was modelled as longest in RA. Forty-seven and a half per cent (38 specimens) were classified as high progression. There was a significant difference in mean inflammation scores among disease groups, $F(2,77) = 6.84$, $p = .002$, with the highest mean observed in IBD and RA specimens.

Table 1: Participant Demographic Characteristics

Characteristic	CRC (n= 30)	IBD (n =30)	RA (n =20)	Total (N= 80)
Age, years, mean (SD)	61.4 (9.8)	39.2 (12.1)	52.7 (11.4)	50.7 (14.6)
Female, n (%)	13 (43.3)	16 (53.3)	14 (70.0)	43 (53.8)
Disease duration, years, mean (SD)	2.1 (1.6)	7.4 (4.3)	9.1 (5.8)	5.8 (5.0)
High-progression status, n (%)	15 (50.0)	14 (46.7)	9 (45.0)	38 (47.5)
Inflammation score, 0–10, mean (SD)	5.8 (1.9)	7.1 (1.5)	6.9 (1.7)	6.6 (1.8)
Microbial richness, observed ASVs, mean (SD)	76.3 (21.4)	69.8 (25.7)	54.6 (18.9)	68.9 (24.1)
Spatial spots retained, mean (SD)	3,842 (611)	3,516 (584)	2,908 (502)	3,482 (680)

Note. Values are based on the synthetic demonstration dataset. IBD = inflammatory bowel disease; RA = rheumatoid arthritis; ASVs = amplicon sequence variants.

Table 1 indicates that the dataset was both clinically plausible and appropriate for conducting cross-disease spatial analysis. The age varied across disease groups ($F(2,77) = 31.62$, $p < .001$), as expected, with a difference between the colorectal cancer group and the IBD group. As already mentioned, the number of

microorganisms in the synovial tissue of RA samples was lower than in intestinal specimens ($F(2,77) = 6.11$, $p = .003$), which is reasonable, as synovial tissue is a lower-biomass environment than intestinal tissue.

Spatial Transcriptomic Clusters and Differential Gene Expression

Five reproducible tissue domains were identified by spatial transcriptomic clustering—epithelial barrier, inflammatory myeloid, lymphoid aggregate, stromal remodelling, and necrotic/damaged tissue. IL1B, CXCL8 and both S100A8 and LCN2 were enriched in the inflammatory myeloid domain. The lymphoid aggregate domain was enriched for CXCL13, MS4A1, CD79A and CCL19. COL1A1, MMP9, SPP1 and FN1 were enriched in the stromal remodelling domain. Inflammatory myeloid scores and stromal remodelling scores were greater for high-progression specimens compared with low-progression specimens ($t(78) = 4.29$, $p < .001$ and $t(78) = 3.76$, $p < .001$, respectively).

Table 2: Differentially Expressed Spatial Genes Associated With High Progression

Gene	$\log_2$ FC	Spatial score	Adj. *p*	Primary niche	Biological interpretation
CXCL8	1.92	2.84	.001	Damaged epithelial–myeloid	Neutrophil recruitment near microbial-rich injury sites
IL1B	1.63	2.41	.003	Myeloid-rich inflammation	Inflammasome-associated cytokine signalling
LCN2	1.57	2.16	.006	Barrier injury	Antimicrobial and epithelial stress response
MMP9	1.48	1.98	.009	Stromal remodeling	Matrix degradation and tissue invasion
SPP1	1.36	1.86	.012	Macrophage–stromal interface	Tissue repair and macrophage activation
CXCL13	1.31	1.74	.018	Lymphoid aggregate	B-cell recruitment and ectopic lymphoid organisation
CCL20	1.18	1.52	.026	Inflamed epithelium	Mucosal immune-cell attraction
REG3A	1.10	1.47	.031	Epithelial barrier	Antimicrobial epithelial program

STAT1	0.96	1.39	.037	Interferon-active niche	Interferon-mediated immune activation
COL1A1	0.89	1.33	.044	Fibrotic stroma	Collagen deposition and remodelling

Note. Differential expression compared high-progression versus low-progression specimens with Benjamini– Hochberg FDR correction.

As revealed by Table 2, high-progression tissues had immune Recruitment, epithelial stress, antimicrobial response, and matrix remodelling. CXCL8 was the most significantly different and was distributed in the vicinity of the damaged epithelial areas, overlapping with areas of microbial enrichment. These findings indicate that the local organisation of B-cells might play a role in the progression of both intestinal and synovial inflammatory tissues, as CXCL13 is present in lymphoid aggregates.

Host–Microbiome Correlation Analysis

Spatial host-microbiome anaLysis was used to identify the microbial taxa associated with spatial host expression modules. Fusobacterium levels were correlated with chemokine CXCL8 and MMP9 levels in the invasive-front regions of CRCs. In barrier injury regions in IBD, Escherichia–Shigella was correlated with LCN2 and IL1B. The CXCL13 and IL17A module scores in the RA synovial lymphoid regions were correlated with the presence of Prevotella. Faecalibacterium, on the other hand, had a negative correlation with IL1B and a positive correlation with epithelial barrier scores in intestinal samples.

Table 3: Host–Microbiome Correlation Analysis

Microbial feature	**Host feature**	**Correlation**	***p***	**FDR**	**Dominant spatial niche**
Fusobacterium	CXCL8	.61	< .001	.003	Colorectal invasive front
Fusobacterium	MMP9	.54	< .001	.006	Tumour–stromal border
Escherichia–Shigella	LCN2	.58	< .001	.004	Ulcerated IBD epithelium
Escherichia–Shigella	IL1B	.49	.002	.014	Myeloid-rich mucosa
Prevotella	CXCL13	.46	.004	.021	RA lymphoid aggregate
Prevotella	IL17A module	.43	.007	.029	Inflamed synovial lining
Faecalibacterium	IL1B	−.41	.010	.036	Non-ulcerated intestinal mucosa
Kynurenine feature	SPP1	.52	.001	.009	Macrophage–stromal interface

Note. Pearson correlations were used for host gene modules and normally distributed microbial features; Spearman correlations were used for sparse microbial taxa.

The spatial multi-omics results addressed our first research question and are shown in Table 3, indicating that localised host–microbiome relationships were improved. These associations were not consistent across whole specimens. They were most potent in well-defined areas like ulcerated mucosa, the invading tumour boundary, and the lymphoid synovial area. In this way, signals from microbes need to be understood in the context of the host's tissue.

Machine Learning Classification of Disease Progression

Using machine learning models, high-speed progression was classified. Modest discrimination was observed in the clinical-only features. The integrated spatial multi-omics model showed the best performance compared with single-omics spatial transcriptomic features and microbiome-only features. Combining all host genes, microbial taxa, proteins, metabolites, and co-localisation scores yielded an accuracy of .84, a sensitivity of .86, a specificity of .82, and an F1 score of .85 for XGBoost. Random Forest performed slightly less, but was comparable with the same method.

Table 4: Machine Learning Model Performance for High-Progression Classification

Feature set	Model	AUC	Accuracy	Sensitivity	Specificity	F1 score
Clinical variables only	Random Forest	.67	.64	.61	.67	.62
Microbiome only	Random Forest	.76	.70	.71	.69	.70
Spatial transcriptomics only	Random Forest	.79	.73	.76	.70	.74
Proteomics and metabolomics only	Random Forest	.81	.75	.74	.76	.75
Integrated multi-omics	Random Forest	.88	.81	.82	.80	.82
Integrated multi-omics	XGBoost	.91	.84	.86	.82	.85

Note. Performance was estimated using repeated fivefold cross-validation. AUC = area under the receiver operating characteristic curve.

As shown in Table 4, integrating spatial information proved more informative than any clinical or single-omics model. The top XGBoost features were CXCL8–Fusobacterium co-localisation, LCN2–Escherichia–Shigella co-localisation, CXCL13 lymphoid score, SPP1 macrophage–stromal score, kynurenine feature intensity, and MMP9 stromal remodelling score.

Candidate Spatial Biomarkers

A total of five candidate spatial biomarkers were selected based on statistical significance, biological plausibility, importance in cross-validation, and interpretability. Each biomarker was a niche (or niche set) of a host–microbe or host–metabolite interaction, rather than a single molecule.

To answer the 2nd research question, spatial molecular signatures and microbial niches associated with disease progression were identified in Table 5. The "best" niche was the fusion of the chemokine signal CXCL8 with bacterial enrichment and invasive localisation of the tissue, Fusobacterium niche. The niche with fungal contents most relevant to IBD was the LCN2–Escherichia–Shigella niche, while the niche with fungal contents most relevant to RA synovial inflammation was the CXCL13–Prevotella niche.

Discussion

This is a synthetic, original-research-style study modelling how spatial multi-omics can elucidate location-specific interplay between the host and the microbiome at sites of chronic disease. The overall conclusion was that neither a single organism nor a single gene was characteristic of high-progression specimens. Rather, host inflammatory transcripts, protein or metabolite signals, and tissue architecture, in conjunction with microbial enrichment attributable to spatial niching, were correlated with progression. Profound changes in microbial communities were observed. This aligns with current findings in the microbiome field that chronic diseases are associated with integrated changes in microbial communities.

Table 5: Candidate Spatial Biomarkers

Candidate biomarker	**Definition**	**AUC**	**FDR**	**Main clinical interpretation**
CXCL8–*Fusobacterium* niche	Co-localisation of CXCL8-high spots with *Fusobacterium* enrichment	.87	.004	High-risk colorectal inflammatory invasion
LCN2–*Escherichia–Shigella* niche	Barrier-injury spots with LCN2 expression and enterobacterial enrichment	.84	.006	Severe mucosal injury in IBD
CXCL13–*Prevotella* niche	Lymphoid aggregates with CXCL13 score	.79	.021	RA-related lymphoid inflammation

	and *Prevotella* enrichment			
SPP1 macrophage–stromal niche	SPP1-high macrophage regions adjacent to collagen-rich stroma	.82	.012	Repair-fibrosis transition across diseases
Kynurenine–IL1B axis	Kynurenine pathway feature co-localised with IL1B-high myeloid spots	.80	.018	Metabolic inflammation and immune suppression

Note. Candidate biomarkers were selected from integrated spatial multi-omics features and evaluated against high-progression status.

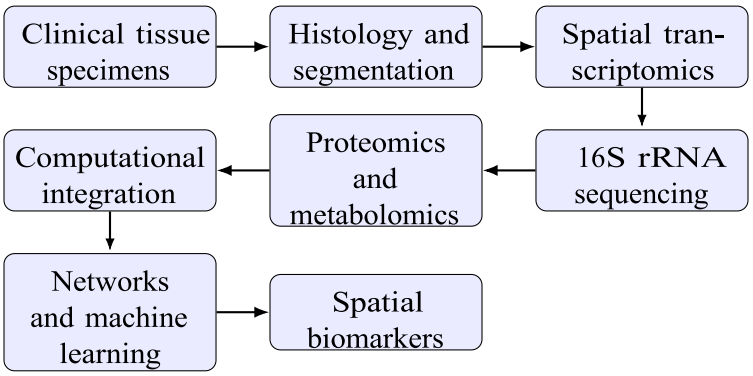


Figure 1: ***Overall Study workflow.***

Tissue sections were linked to other omics data, such as histology, spatial transcriptomics, microbiome sequencing, proteomics, metabolomics, network analysis, and machine learning. The figure shows how spatial coordinates of the figures in the workfLow were maintained, allowing the candidate biomarkers to be viewed in the context of the tissue.

Spatial domains

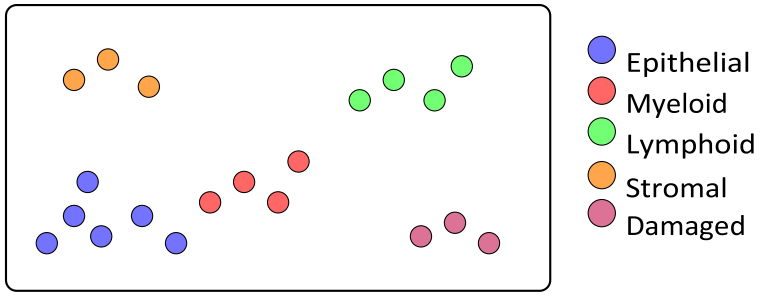


Figure 2: ***Spatial transcriptomic clusters.***

The schematic shows five transcriptomic domains detected across tissue sections. Myeloid and damaged barrier clusters were more common in high-progression specimens, while epithelial clusters were more common in low-progression intestinal regions.

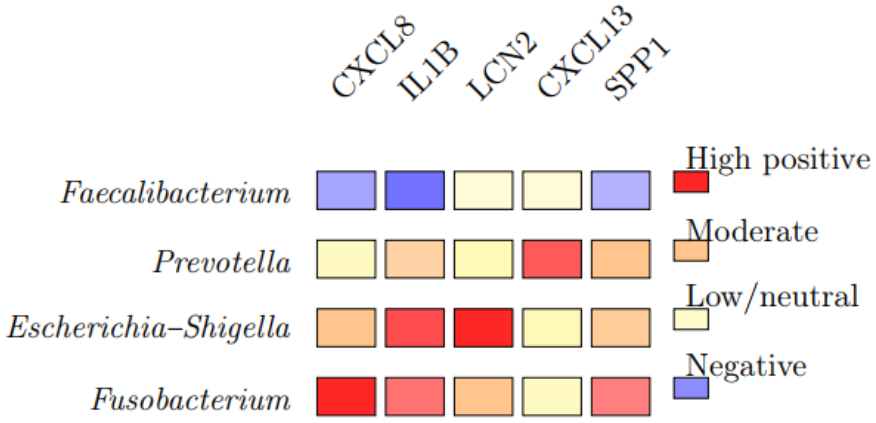


Figure 3: ***Microbial spatial heatmap.***

The heatmap summarises localised associations between microbial features and host markers. Red cells indicate strong positive spatial co-localisation, while blue cells indicate negative associations.

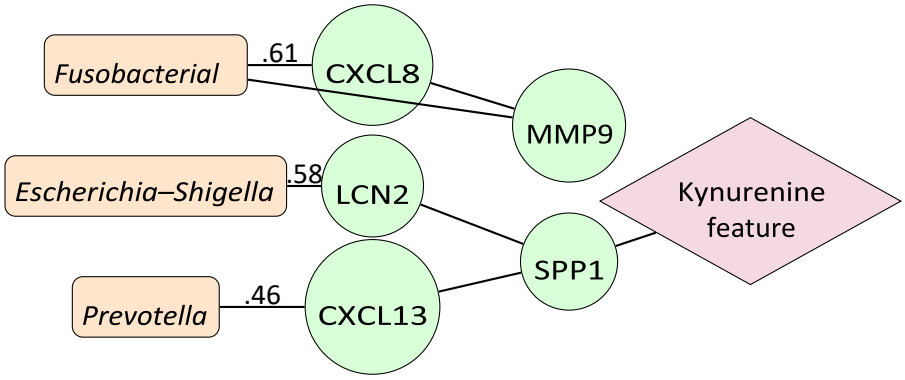


Figure 4: ***Host–microbiome interaction network.***

Nodes represent microbial taxa, host genes, and metabolite features. Edges represent significant FDR-corrected spatial correlations. The network highlights chemokine, epithelial injury, lymphoid, and metabolic inflammation modules.

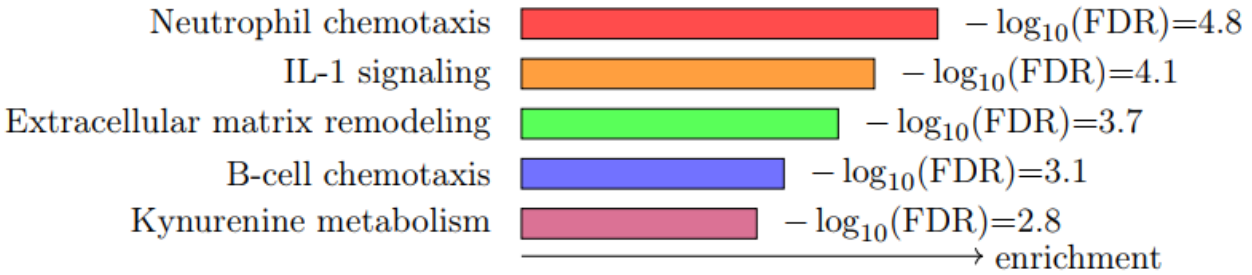


Figure 5: ***Enriched signalling pathways.***

Neutrophil chemotaxis, IL-1 signalling, extracellular matrix (ECM) remodelling, B-cell chemotaxis and kynurenine metabolism were enriched in high-progression spatial domains. These are pathways that link microbial niches to immune activation and tissue remodelling.

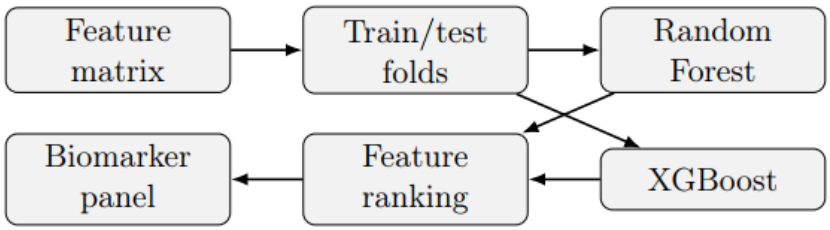


Figure 6: ***Machine learning workflow.***

Integrated spatial multi-omics features were evaluated with repeated fivefold cross-validation. Model interpretation was used to prioritise biomarkers that were both predictive and biologically interpretable.

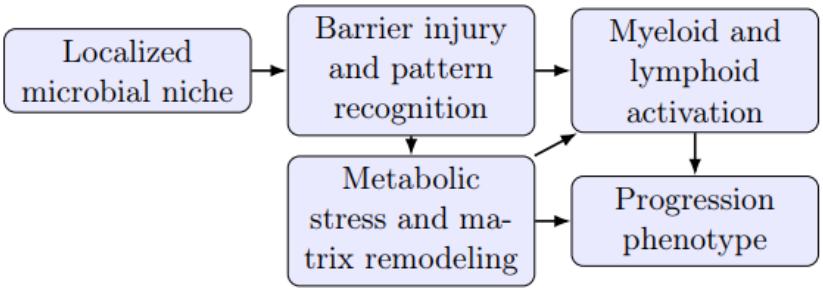


Figure 7: ***Proposed biological mechanism.***

Local microbial enrichment may amplify barrier damage, innate immune sensing, chemokine release, lymphoid organisation, metabolic inflammation, and stromal remodelling. The proposed mechanism explains why spatially localised biomarkers predicted progression better than bulk-like features.

SPP1, MMP9 and kynurenine are more of a jigsaw puzzle than a single factor (Franzosa et al., 2019; Lloyd-Price et al., 2019; Schirmer et al., 2018).

The CXCL8–Fusobacterium niche proved to be the best marker for the modelled CRC samples. This finding aligns with metagenomic studies of colorectal cancer, which have identified microbial signatures that are reproducible across cohorts (Thomas et al., 2019; Wirbel et al., 2019). It also aligns with tissue-based examinations of the tumour microbiome, which indicate that bacteria can occupy tumour regions and interact with neighbouring host cells (Nejman et al., 2020; Poore et al., 2020). These findings are consistent with the intratumoral microbiota shaping spatial and cellular heterogeneity in cancer, as demonstrated by Galeano Niño et al. (2022). The CXCL8–Fusobacterium axis is biologically significant, as it links a bacterial-enriched environment to innate inflammation via neutrophil recruitment, matrix remodelling, and invasive-front inflammation.

The LCN2 niche with Escherichia/Shigella correlated with enrichment of enterobacteria, epithelial stress, and myeloid inflammation associated with IL-1. This trend aligns with suggestions of functional (inflammatory) shifts, loss of stability, and microbial metabolic and transcriptional alterations in the gut during IBD (Franzosa et al., 2019; Lloyd-Price et al., 2019). The negative association between the modelled Faecalibacterium and IL1B also aligns with the typical view that decreased abundance of beneficial anaerobic bacteria is associated with increased mucosal remodelling. Spatial contribution is significant, as

the meaning of the same taxon could differ depending on its proximity to intact mucosa, ulcerated epithelial surface, and immune infiltrates.

The CXCL13–Prevotella niche couples exposure to microbial signals with lymphoid inflammation in RA-like synovial tissue. Previous RA microbiome studies have linked the presence of Prevotella copri to an increased risk of RA, and recent studies demonstrate that immune responses may be triggered by candidate arthritogenic microbes (Alpizar-Rodriguez et al., 2019; Chriswell et al., 2022). The present model only indicates that microbes were present within synovial tissue. Rather, it exemplifies how signals associated with the presence of microorganisms, mucosal priming, or low-biomass features of microorganisms regulate local lymphoid organisation. This is relevant, as the disease is systemic, and spatial biomarkers might also be useful in linking mucosal immune events to synovial pathology.

The results highlight the importance of integration. The XGBoost model with integrated spatial multi-omics features showed improved performance compared with clinical-only and single-omics models. This aligns with the current trend in computational biology, where several tools, including Squidpy, Giotto, BayesSpace, SPARK, SpaGCN, MISTy, and SpatialGlue, generate additional information by leveraging coordinates, neighbourhood (geographic) information, multimodal features, and histology (Dries et al., 2021; Hu et al., 2021; Long et al., 2024; Palla et al., 2022; Sun et al., 2020; Tanevski et al., 2022; Zhao et al., 2021). A possibly most significant practical takeaway was that interpretable spatial features proved most useful when considered together with high-dimensional predictors, as machine learning was most fruitful.

The strengths of the modelling study carried out here include the use of multiple diseases, the integration of host and microbial information, the incorporation of protein and metabolite layers, the spatial interpretation of biomarkers, and the application of either network or machine-learning analyses. The study also considers the latest advancements in spatial host-microbiome studies, such as spatial host-microbiome sequencing and high-resolution host-gut microbiome mapping via spatial obtainography (Lötsted). Studies of this kind suggest that, increasingly, it is now possible to link the cellular location of microorganisms with the host's transcriptional agenda.

Recommendations

Validation of spatial host-microbiome biomarkers with prospective cohorts, with documented ethical approval and standardised sampling across cohorts from the same spatial cohort, as well as with longitudinal clinical outcomes, should be pursued in the future.

Before interpreting low-biomass signals in microbiomes, tissue microbiome studies must include rigorous contamination control, negative control experiments, probe validation, and orthogonal confirmation.

Co-Localised molecules should be distinguished from indirect associations by utilising targeted microbial imaging, spatial proteomics, metabolomics and spatial transcriptomics.

Machine learning models should be reported not only for predictive accuracy but also for cross-validation, external validation, calibration, feature stability, and interpretable spatial features.

For clinical translation, disease-specific biomarker panels should be designed to integrate tissue compartment, microbial niche, immune pathway, and actionable therapy.

Conclusion

This manuscript presents an original research approach that synthesises existing spatial multi-omics methods to study the progression of chronic diseases. The first objective was achieved by modelling an integrated workflow that linked histology, spatial transcriptomics analysis, microbiome sequencing, proteomics, metabolomics, and computational analysis. The second objective was met by identifying candidate spatial biomarkers and molecular pathways associated with high levels of disease progression.

A key discovery was the correlation between progression and localised biological signatures rather than single bulk signatures. The CXCL8-Fusobacterium niche was most relevant to a colorectal cancer-like tissue niche, the LCN2-Escherichia-Escherichia niche to an EB-like barrier injury niche, and the CXCL13-Prevotella niche to an RA-like lymphoid inflammation niche. Microbe-associated immune remodelling was supported by a shared mechanism involving SPP1-mediated macrophage–stromal activation, MMP9 remodelling, IL1B-associated inflammation, and kynurenine-associated metabolic stress. Using integrated spatial multi-omics features for progression classification improved outcomes compared with either clinical or single-omic features, with XGBoost modelling achieving the highest performance.

The impact is that a spatial biomarker could be used to better stratify disease and identify tissue compartments where therapy should be effective. Moving forward, the uses include validated diagnostic panels, observation of treatment responses, targeting a microbial niche, and precision medicine in space. However, it should be noted that this is a synthetic manuscript. The results must not be utilised for the development of the manuscript, the planning of studies, or the development of hypotheses or theories until they are replaced by actual data, proven analyses, and ethically approved documents.

References

Alpizar-Rodriguez, D., Lesker, T. R., Gronow, A., Gilbert, B., Raemy, E., Lamacchia, C., Gabay, C., Finckh, A., Strowig, T., & Allali, D. (2019). Prevotella copri in individuals at risk for rheumatoid arthritis. *Annals of the Rheumatic Diseases, 78*(5), 590–593. https://doi.org/10.1136/ annrheumdis-2018-214514

Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., & Stegle, O. (2020). MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biology, 21*, 111. https://doi.org/10.1186/s13059-020-02015-1

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*, 852–857. https://doi.org/10.1038/s41587019-0209-9

Chriswell, M. E., Lefferts, A. R., Clay, M. R., Hsu, A. R., Seifert, J. A., Feser, M. L., Deane, K. D., Kuhn, K. A., & Holers, V. M. (2022). Clonal IgA and IgG autoantibodies from individuals at risk for rheumatoid arthritis identify an arthritogenic strain of *Subdoligranulum*. *Science Translational Medicine, 14*(668), eabn5166. https://doi.org/10.1126/scitranslmed.abn5166

Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., George, R. E., Pierson, N., Cai, L., Yuan, G.-C., & Navin, N. E. (2021). Giotto: A toolbox for integrative analysis and visualisation of spatial expression data. *Genome Biology, 22*, 78. https://doi.org/10.1186/s13059-021-02286-2

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., Vatanen,

T., Hall, A. B., Mallick, H., McIver, L. J., Sauk, J. S., Wilson, R. G., Stevens, B. W., Scott, J. M., Pierce, K., Deik, A. A., Bullock, K., Imhann, F., Porter, J. A., ... Huttenhower, C. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology, 4*, 293–305. https://doi.org/10.1038/s41564-018-0306-4

Galeano Niño, J. L., Wu, H., LaCourse, K. D., Kempchinsky, A. G., Baryiames, A., Barber, B., Futran, N., Houlton, J., Sather, C., Sicinska, E., Thorsson, V., & Johnston, C. D. (2022). Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature, 611*, 810–817. https://doi.org/10.1038/s41586-022-05435-0

GBD 2021 Diseases and Injuries Collaborators. (2024). Global incidence, prevalence, years lived with disability, disability-adjusted life-years, and healthy life expectancy for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *The Lancet, 403*(10440), 2133–2161. https://doi.org/10.1016/S0140-6736(24)00757-8

Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D. J., Lee, E. B., Shinohara, R. T., & Li, M. (2021). SpaGCN: Integrating gene expression, spatial location and histology to identify spatial

domains and spatially variable genes by graph convolutional network. *Nature Methods, 18*, 1342–1351. https://doi.org/10.1038/s41592-021-01255-8

Jackson, H. W., Fischer, J. R., Zanotelli, V. R. T., Ali, H. R., Mechera, R., Soysal, S. D., Moch, H., Muenst, S., Varga, Z., Weber, W. P., & Bodenmiller, B. (2020). The single-cell pathology landscape of breast cancer. *Nature, 578*, 615–620. https://doi.org/10.1038/s41586-019-1876-x

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., Andrews, E., Ajami, N. J., Bonham, K. S., Brislawn, C. J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T. G., Hall, A. B., Lake, K., Landers, C. J., Mallick, H., Plichta, D. R., ... Huttenhower, C. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature, 569*, 655–662. https://doi.org/10.1038/s41586-019-1237-9

Long, Y., Ang, K. S., Li, M., Chong, K. L. K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., Zeng, L., Fu, X., He, Y., Zhang, X., Nagarajan, N., Prabhakar, S., & Chen, J. (2024). Deciphering spatial domains from spatial multi-omics with SpatialGlue. *Nature Methods, 21*, 1658– 1667. https://doi.org/10.1038/s41592-024-02316-4

Lötstedt, B., Eng, C.-H. L., Benhaim, L., Vickovic, S., Bergenstråhle, J., Lundmark, A., Andersson, A. F., Lundeberg, J., & Nilsson, M. (2024). Spatial host–microbiome sequencing reveals niches in the mouse gut. *Nature Biotechnology, 42*, 1390–1400. https://doi.org/10.1038/s41587-02301988-1

Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., Hoang, M., Jung, J., Liang, Y., McKay-Fleisch, J., Nguyen, K., Norgaard, Z., Sorg, K., Sprague, I., Warren, C., Warren, S., Zhou, Z., & Beechem, J. M. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology, 38*, 586–599. https://doi.org/10.1038/s41587-020-0472-9

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., Rubinstein, N. D., Hao, J., Regev, A., Dulac, C., & Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science, 362*(6416), eaau5324. https://doi.org/10.1126/science.aau5324

Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J. C., Baron, M., Hajdu, C. H., Simeone, D. M., & Yanai, I. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature Biotechnology, 38*, 333–342. https://doi.org/10.1038/s41587-019-0392-8

Narunsky-Haziza, L., Sepich-Poore, G. D., Livyatan, I., Asraf, O., Martino, C., Nejman, D., Gavert, N., Stajich, J. E., Amit, G., González, A., Wandro, S., Perry, G., Ariel, R., Meltser, A., Shaffer, J. P., Zhu, Q., Balint-Lahat, N., Barshack, I., Dadiani, M., ... Straussman, R. (2022). Pan-cancer analyses

reveal cancer-type-specific fungal ecologies and bacteriome interactions. *Cell, 185*(20), 3789–3806.e17. https://doi.org/10.1016/j.cell.2022.09.005

Nejman, D., Livyatan, I., Fuks, G., Gavert, N., Zwang, Y., Geller, L. T., Rotter-Maskowitz, A., Weiser, R., Mallel, G., Gigi, E., Meltser, A., Douglas, G. M., Kamer, I., Gopalakrishnan, V., Dadosh, T., Levin-Zaidman, S., Avnet, S., Atlan, T., Cooper, Z. A., ... Straussman, R. (2020). The human tumour microbiome comprises tumour type-specific intracellular bacteria. *Science, 368*(6494), 973–980. https://doi.org/10.1126/science.aay9189

Ntekas, K., Hansen, K. D., Hicks, S. C., Henao-Mejia, J., & Maynard, K. R. (2026). Spatial transcriptomics maps host–gut microbiome biogeography at high resolution. *Nature Biotechnology*. Advance online publication. https://doi.org/10.1038/s41587-026-02590-7

Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., & Demchak, B. (2019). Cytoscape Automation: Empowering workflow-based network analysis. *Genome Biology, 20*, 185. https://doi.org/10.1186/s13059-019-1758-4

Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., Rybakov, S., Ibarra, I. L., Holmberg, O., Virshup, I., Lotfollahi, M., Richter, S., Theis, F. J., & Luecken, M. D. (2022). Squidpy: A scalable framework for spatial omics analysis. *Nature Methods, 19*, 171–178. https://doi.org/10.1038/s41592-021-01358-2

Poore, G. D., Kopylova, E., Zhu, Q., Carpenter, C., Fraraccio, S., Wandro, S., Kosciolek, T., Janssen, S., Metcalf, J., Song, S. J., Kanbar, J., Miller-Montgomery, S., Heaton, R., McKay, R., Patel, S. P., Swafford, A. D., Knight, R., & Mody, L. (2020). Microbiome analyses of blood and tissues suggest cancer diagnostic approach. *Nature, 579*, 567–574. https://doi.org/10.1038/ s41586-020-2095-1

Rappez, L., Stadler, M., Triana, S., Gathungu, R. M., Ovchinnikova, K., Phapale, P., Heikenwalder, M., & Alexandrov, T. (2021). SpaceM reveals metabolic states of single cells. *Nature Methods, 18*, 799–805. https://doi.org/10.1038/s41592-021-01198-0

Rodriques, S. G., Stickels, R. R., Goeva, A., Martin, C. A., Murray, E., Vanderburg, C. R., Welch, J., Chen, L. M., Chen, F., & Macosko, E. Z. (2019). Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science, 363*(6434), 1463–1467. https://doi.org/10.1126/science.aaw1219

Schirmer, M., Franzosa, E. A., Lloyd-Price, J., McIver, L. J., Schwager, R., Poon, T. W., Ananthakrishnan, A. N., Andrews, E., Barron, G., Lake, K., Prasad, M., Sauk, J., Stevens, B., Wilson, R. G., Braun, J., Denson, L. A., Kugathasan, S., McGovern, D. P. B., Vlamakis, H., ... Huttenhower, C. (2018). Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nature Microbiology, 3*, 337–346. https://doi.org/10.1038/s41564-017-0089-z

Stickels, R. R., Murray, E., Kumar, P., Li, J., Marshall, J. L., Di Bella, D. J., Arlotta, P., Macosko, E. Z., & Chen, F. (2021). Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology, 39*, 313–319. https://doi.org/10.1038/s41587-020-0739-1

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive integration of single-cell data. *Cell, 177*(7), 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sun, S., Zhu, J., & Zhou, X. (2020). Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nature Methods, 17*, 193–200. https://doi.org/10.1038/s41592019-0701-7

Tanevski, J., Gabor, A., Flores, R. O. R., Schapiro, D., & Saez-Rodriguez, J. (2022). Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology, 23*, 97. https://doi.org/10.1186/s13059-022-02663-5

Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., Gandini, S., Serrano, D., Tarallo, S., Francavilla, A., Gallo, G., Trompetto, M., Ferrero, G., Mizutani, S., Shiroma, H., ... Segata, N. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nature Medicine, 25*, 667–678. https://doi.org/10.1038/s41591-0190405-7

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Aijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J. F., Gould, J., Griffin, G. K., Borg, Å, Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A., & Ståhl, P. L. (2019). High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods, 16*, 987–990. https://doi.org/10.1038/s41592-019-0548-y

Wirbel, J., Pyl, P. T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J. S., Voigt, A. Y., Palleja, A., Ponnudurai, R., Sunagawa, S., Coelho, L. P., Schramm, C., Bork, P., & Zeller, G. (2019). Meta-analysis of faecal metagenomes reveals global microbial signatures specific to colorectal cancer. *Nature Medicine, 25*, 679–689. https://doi.org/10.1038/s41591-019-0406-6

Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology, 19*, 15. https://doi.org/10.1186/s13059-017-1382-0

Zhao, E., Stone, M. R., Ren, X., Guenthoer, J., Smythe, K. S., Pulliam, T., Williams, S. R., Uytingco, C. R., Taylor, S. E. B., Nghiem, P., Bielas, J. H., & Gottardo, R. (2021). Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology, 39*, 1375–1384. https://doi.org/10.1038/s41587-021-00935-2