

Multi-Center Validation of a Low-Cost, High-Throughput Recurrence Risk Assessment for Colon Cancer through Spatial Transcriptomic Inference from Routine Staining

Project Relevance

Every year, over 150,000 Americans are diagnosed with Colon Cancer (CRC), and annually over 50,000 individuals will die from CRC, necessitating improvements in screening, prognostication, disease management, and therapeutic options. Current multiplexed spatial profiling assays, used to evaluate the spatial distribution of cell types like tumor-infiltrating lymphocytes (TILs) and their molecular changes, show promise in identifying biomarkers for recurrence risk, aiding targeted therapy decisions— yet their clinical use is hindered by high costs, limited throughput, and inconsistency. Our project proposes to refine and validate an affordable, high-throughput technique (**Virtual Molecular Inference— VMI**) that can decipher intricate gene activity patterns within the tumor immune microenvironment (TIME) utilizing routinely collected tissue stains across multiple medical centers, holding the potential to significantly enhance CRC recurrence prediction, tumor staging, and aid in the identification of individuals who would benefit from adjuvant treatments.

Project Summary

Colon Cancer (CRC) is the third most common form of cancer and cancer-related deaths in the United States. Examination of regional lymph nodes at the time of surgical resection is essential for prognostication. While it is important to maximize the number of lymph nodes assessed, recent population-based studies have shown that evaluation of lymph node involvement is usually incomplete or inadequate. This can impact the accuracy of tumor staging and downstream disease management options, such as whether the patient should receive adjuvant chemotherapy. Developing alternatives to assess lymph node involvement through indirect mechanisms could be illuminating in cases where lymph node sampling is inadequate. Spatial expression patterns of various cell types, including tumor-infiltrating lymphocytes (TIL), both within and around the primary tumor, the Tumor Immune Microenvironment (TIME), have been identified as important CRC prognostic indicators, illuminating a coordinated immune response. The comprehensive, spatial characterization of these cellular lineages is possible using highly multiplexed spatial omics technologies, but high cost and low throughput prevent their clinical deployment. *Virtual Molecular Inference (VMI)* can infer spatial molecular information (e.g., spatial transcriptomics– ST) at low cost from tissue histology where the morphology allows. In prior COBRE research, we created a cost-effective VMI test demonstrating highly precise gene expression inference to study factors relating to CRC metastasis. However, its role as a complementary tool for conventional recurrence assessments needs validation through performance evaluation across diverse patient and tumor contexts in various geographic locations. In the largest set of CRC tumors to undergo ST profiling in a multi-institutional setting to-date, we will identify spatial whole transcriptomic markers of recurrence with Visium ST (**IDENTIFY**). Single-cell transcriptomic data collected on serial tissue sections will complement these assessments by improving the resolution and cell-type specificity of ST, further elucidating intercellular tissue dynamics with appropriate validation. We will then establish histological correspondence to identified spatial metastasis markers by refining **VMI** algorithms to convert H&E-stained tissue into ST markers in a multi-center setting (**REFINE**). Spatial and cell-type specific patterns of molecular markers that indicate whether a patient is likely to develop recurrence/metastasis will be identified under this framework– forming the *TNM-Recurrence* (TNM-R) score. Inferring such information from tissue morphology can provide a low-cost and highly interpretable recurrence risk assessment to determine response to adjuvant chemotherapy. Finally, a multi-center comparative assessment of this digital assay to other prevailing recurrence markers (e.g., Immunoscore, ctDNA) (**VALIDATE**) and iterative human factors development of the digital display (**INTERFACE**) will set the stage for a type I effectiveness-implementation clinical trial which will assess the impact of this tool on clinical decision making, downstream health outcomes, and cost-effectiveness to inform implementation strategies.

Specific Aims

Colon cancer (CRC) has an annual incidence in the United States of approximately 150,000 new cases and a 63% 5-year survival rate. Successive invasion into epithelial, lamina propria, submucosal and other layers of colon is prognostic, where higher tumor stage reflects invasion depth. Evaluating the involvement of lymph nodes is crucial for determining a patient's prognosis, such as the likelihood of cancer recurrence. However, inconsistencies in lymph node dissection can hinder a thorough evaluation for staging the disease. This challenge necessitates the use of molecular profiling as an indirect method to detect nodal metastasis, or the adoption of other strategies to improve the assessment of recurrence risk.

Numerous cell types within the primary tumor environment can signal the risk of cancer recurrence. Among these, tumor-infiltrating lymphocytes (TILs) are gaining prominence as they not only orchestrate direct anti-cancer immune responses but also play a pivotal role in the recruitment, activation, and maturation of additional immune cells. TILs' influence on the tumor immune microenvironment (TIME) and patient prognosis is increasingly recognized, with factors such as mismatch repair alterations modulating their impact. However, comprehensive analysis of these cellular interactions through spatial multiplexing, essential for biomarker discovery and clinical application, remains elusive due to financial, throughput, and reliability constraints. Therefore, there is a pressing need for alternative methods that can accurately deduce this molecular information, thereby facilitating widespread evaluation of spatial molecular patterns for research and clinical use.

This proposal is focused on deciphering how spatial distributions of extensively multiplexed molecular markers (encompassing the whole transcriptome) can predict cancer recurrence or metastasis. We aim to streamline these insights into an affordable, supplementary test that enhances the accuracy of recurrence risk evaluation. By integrating morphological data from whole slide images with spatial transcriptomics, we plan to develop a cost-effective and rapid method for computationally deriving prognostic molecular information from these images—**Virtual Molecular Inference (VMI)**. Building on our previous COBRE project, which made significant strides in understanding the spatial dynamics of CRC metastasis and developing/applying VMI to various tissue types through several feasibility studies. We now seek to further this work at scale in a multi-centric setting and compare its effectiveness to prevailing assays. Our goal is to create and validate a clinically applicable tool for assessing recurrence risk in CRC patients, the TNM-Recurrence score, which is based on data derived from VMI and tailored to diverse clinical situations.

Aim 1: (IDENTIFY) Multicenter Biomarker Identification Using Visium ST Data – Collaboratively analyze Visium Spatial Transcriptomics data from both Dartmouth and Cedars Sinai. Focus on identifying molecular markers associated with CRC metastasis and recurrence across diverse patient populations and datasets.

Aim 2: (REFINE) VMI Algorithm Calibration and Refinement in a Multicenter Context – Utilize identified biomarkers to calibrate and refine the **VMI** algorithm. Emphasize the adaptation and optimization of the algorithm for accuracy and reliability, ensuring applicability and validation across different institutional settings.

Aim 3: (VALIDATE) Multicenter Validation of VMI and Comparative Analysis with Existing Assays. Conduct a multicenter retrospective study to validate **VMI**'s efficacy using unseen whole slide images through calculation of the *TNM-Recurrence Score (TNM-R)*. Ensure the algorithm's robustness and generalizability across different datasets. Compare the **VMI**'s performance for recurrence prediction head-to-head with existing assays— 1) CDX2 expression, 2) circulating tumor DNA (ctDNA) markers, and 3) Immunoscore— to assess its relative efficacy, specificity, and sensitivity.

Aim 4: (INTERFACE) Development and Usability Testing of Clinical Decision Support Interface. Focus on the development of a clinician-oriented interface for **VMI**, emphasizing user-friendliness and clarity in prognostic information presentation. We will conduct thorough usability testing. This process includes iterative refinements based on clinician feedback to ensure the interface is practical, intuitive, and effective.

This proposal aims to validate the effectiveness of our approach on a broader scale by combining the unique infrastructures and patient populations of Dartmouth College and Cedars Sinai. Previously, Dr. Levy's work at Dartmouth successfully bridged the CQB COBRE Single Cell Genomics Core and the Pathology Shared Resource, significantly improving specimen quality and cost-effectiveness. Building on this, we will leverage the molecular research capabilities at both institutions, namely Dartmouth's *Augmet* and Cedars' *Molecular Twins*, to facilitate comprehensive comparative analyses. Drs. Levy and Vaickus, as Co-PIs, have curated a collaborative team of clinician-scientists to enhance the translational relevance of our study, ensuring that findings will not only contribute to advancing basic science research but also aid in developing computational biology tools for sophisticated spatial omics analyses. This study will reveal how spatial expression patterns associate with recurrence, laying the groundwork for a multicenter trial to assess the impact of these findings on clinical decisions and patient outcomes. The outcomes will be crucial in planning and executing a hybrid type I trial, focusing on both the clinical effectiveness and implementation of discoveries into routine clinical practice.

Research Strategy

A. SIGNIFICANCE

A.1. Burden of Colon Cancer and Assessment Challenges. Colon cancer (CRC) is the third leading cause of cancer both globally and within the United States (~8% of cancer deaths)¹. The incidence of CRC is shifting towards younger demographics who are not included in established screening programs^{2,3}. While modifying specific risk factors such as epigenetics, environment, and diet can potentially reduce CRC incidence, there exists an immediate and pressing need to develop more accurate, faster, and cost-effective solutions for CRC screening and prognostication⁴⁻⁶. This vital necessity forms the core objective of the work detailed in this proposal. Disease management of resected CRC includes a key step: lymph node dissection to determine the initial N-stage at treatment, which serves as a proxy for recurrence risk. This is typically followed by adjuvant therapy for patients with involved lymph nodes. However, assessing tumor metastasis remains a formidable challenge, as the quality of lymph node dissection can vary widely, potentially leading to inaccurate assessments of recurrence risk. For instance, a study found that only 37% of colon cancer cases received proper evaluation of regional lymph nodes^{7,8}. The number of lymph nodes retrieved during dissection is a significant prognostic factor; retrieving a high number (>20) is associated with a better prognosis⁹⁻¹². Inadequate dissection and analysis not only impact prognostication but also influence selection of appropriate treatment strategies. This reality underscores the imperative for studying molecular alterations at the primary tumor site for indirect evidence of metastasis / recurrence risk, supplementing traditional pathological TNM staging.

A.2. The Crucial Role of Tumor Infiltrating Lymphocytes in CRC Prognostication. Our proposal focuses on understanding the Tumor Immune Microenvironment (TIME), a complex network of cancerous and non-cancerous cells, signaling molecules, and structural components like blood vessels and extracellular matrix¹³. The spatial patterns and densities of different cellular lineages, notably Tumor Infiltrating Lymphocytes (TILs) – including T cells, B cells, and NK cells – play a crucial role in the anti-tumoral response. Influenced by factors like Microsatellite Instability (MMR) status¹³⁻¹⁵, this immune contexture and other TIME components offer valuable prognostic information that our study aims to explore and leverage.

A.3. Emerging Assays for Assessing Recurrence Risk. There are several pathology assays that complement traditional pTNM (Tumor-Node-Metastasis) staging and offer independent risk factors for recurrence. Key among these are: **1) Immunoscore:** This assay quantifies the density of cytotoxic T cells both within and surrounding the tumor, providing a robust measure of the immune system's engagement with the tumor¹⁶. **2) CDX2 Assay:** A signature of colon epithelial differentiation – CDX2-negative tumors exhibit a highly aggressive, immature progenitor-cell phenotype^{17,18}. **3) Circulating Tumor DNA (ctDNA):** This approach involves assessing mutations across several molecular pathways (e.g., vascular endothelial growth factor pathway)^{19,20}.

However, these assays have limitations that constrain their clinical application. Immunoscore, for instance, only assess two lymphocyte lineages (CD3/CD8, a small subset of the immune cells present). The variation in immune responses among individuals further complicates analysis. Additionally, Immunoscore requires manual/digital pathology interpretation, which disrupts the clinical workflow and adds cost^{16,21-24}.

A.4. Transformative Potential of Spatial Transcriptomics to Expand the Scope and Resolution of Spatial Biomarkers. The spatial organization of immune cells and other important lineages within TIME is complex and challenging to decode and utilize clinically. Recent technological advances in spatial omics, including 10x Genomics Spatial Transcriptomics (ST) and Nanostring Digital Spatial Profiling (DSP)²⁵⁻²⁷, are revolutionizing our ability to dissect this complexity at unprecedented spatial resolutions. These platforms allow for the high-resolution mapping of whole transcriptomes (e.g., WTA) within tissue sections, enabling a detailed comparison of TILs and other important cellular subpopulations across the TIME. This advanced technique offers several advantages over assays like Immunoscore and CDX2: **1) Comprehensive TIME Analysis:** ST provides a holistic understanding of the spatial organization of **all cell types** within TIME, transcending the limitations of IHC, typically used to analyze only a limited number of molecular markers on serial sections. This includes detailed mapping of specific cell types, their interactions, and the spatial distribution of gene expression patterns. Such comprehensive profiling offers a broader view of tumor biology, potentially revealing novel disease mechanisms and therapeutic targets²⁵. **2) Identification of Spatially Distinct Patterns:** Traditional assays that typically profile bulk specimens might miss spatially distinct gene expression patterns critical for understanding disease progression, overlooking cell type and location. ST excels in identifying these patterns. **3) Enhanced Predictive Accuracy and Novel Insights:** By analyzing the entirety of the TIME spatial transcriptomics enhances the accuracy of recurrence risk prediction and facilitates the discovery of new, targetable biomarkers and pathways.

However, there is notable underutilization of spatial transcriptomics in clinical translational research. The uptake of these advanced technologies is hindered by several factors: the high costs associated with spatial omics assays, their relatively low throughput, and challenges in achieving reproducibility^{28,29}. This restricts the

scope of large-scale association studies that are essential to determine the true clinical value of emerging diagnostic tests. Furthermore, understanding risk factors for recurrence across diverse conditions and demographic groups is critical to support the widespread clinical implementation of these technologies.

A.5. Introducing VMI Technologies for Low Cost, High Throughput Analysis. Pathological examination of tissue samples, traditionally conducted through a variety of staining techniques, remains a cornerstone in the diagnosis and prognosis of colon cancer (CRC). However, these procedures are time-consuming, labor-intensive, and prone to variability, depending on the skill and experience of the pathologist and dissector. To address these challenges, we have developed innovative automated scoring systems which utilize spatial profiling technologies. These systems transform morphological features into digital biomarkers, offering a new dimension in pathology analysis. A key advancement in this area is the development of Virtual Molecular Inference (VMI) technologies. Leveraging sophisticated machine learning algorithms, these technologies simulate the appearance and diagnostic information of specialized molecular stains/markers using only whole slide images (WSIs) of routine stains (H&E) as input³⁰⁻³⁴. By training models to recognize and infer complex patterns of cellular organization and gene expression from WSI, our approach seeks to 'virtually' reproduce the key diagnostic spatial molecular information (**VMI**) that spatial transcriptomics provides, without the associated time, cost, manual interpretation and variability^{35,36}. The clinical implications of this work are significant: 1) Accelerated Diagnosis: **VMI** can significantly reduce the time required to prepare and analyze tissue samples, enabling faster, more timely interventions. 2) Enhanced Precision: By reducing the subjectivity and variability associated with traditional staining and interpretation, **VMI** offers the promise of more consistent and precise diagnostic assessments. 3) Cost-Effective Analysis: **VMI** eliminates the need for costly and potentially hazardous reagents, and reduces the labor required for sample preparation and analysis. 4) Expanded Accessibility: As a digital technique, **VMI** can be easily deployed in settings where access to advanced pathology services is limited, expanding access to high-quality cancer diagnostics. In our proposal, while **VMI** spans a range of technologies such as *virtual mIF/mIHC staining*, our primary emphasis and subsequent references will focus on ST inference.

B. INNOVATION. TNM-R Scoring System: A New Paradigm in CRC Assessment, Derived from VMI. We propose that leveraging advanced **VMI** technology can significantly enhance our ability to evaluate recurrence risk in colon cancer. By inferring spatial molecular information from routine tissue sections, **VMI** can help identify recurrence risk factors that extend beyond current biomarkers and staging techniques. Our innovation significantly enhances the existing TNM staging system by incorporating a **VMI-derived** recurrence risk score (TNM-R), leading to more precise risk stratification, improved treatment decision-making, and enhanced disease management. By examining gene expression across different tissue architectures, the TNM-R score will incorporate novel spatial features which correlate with CRC recurrence and metastasis in a framework familiar to oncologists and surgeons. Our hypothesis centers on the premise that highly multiplexed spatial transcriptomics, inferred from standard tissue morphology (**VMI**), can offer a cost-effective and efficient method for detailed molecular assessment. While recurrence risk can be predicted from H&E-stained slides alone, these methods fall short in leveraging the comprehensive information offered by ST, particularly in revealing underlying mechanisms. This project aims to tap into the full potential of ST through **VMI**, not only enhancing recurrence risk assessment but also deepening the precision of histological analysis.

Advancing interpretable neural network architectures for spatial inference. High-resolution deep learning (DL) methods will reliably infer recurrence-related biomarkers from standard tissue stains (**VMI**). We will evaluate the performance of various model architectures to predict diverse biomarker panels, examining tradeoffs between predictiveness, biological relevance and integrating existing biological knowledge to boost accuracy. Our work is distinct from previous work that applied deep learning to arbitrarily defined gene sets with unclear relevance.

Furthermore, improved tissue/imaging quality has enhanced our ability to accurately detect these markers³⁷.

Emphasis on clinical utility of TNM-R through head-to-head comparison. A comparative analysis of our **VMI** algorithms against established assays like Immunoscore, CDX2, and ctDNA, will assess their efficacy, specificity, and sensitivity, in enhancing patient stratification and treatment planning in colon cancer management.

Rigorous validation of VMI results using single cell techniques. Utilizing Xenium, ancillary staining, and single-cell transcriptomics, we will validate the patterns inferred from H&E slides to ensure accuracy and reliability.

While our ultimate goal is to gauge the tool's effectiveness in a multicenter clinical trial, we acknowledge several imminent challenges: 1) Consistency and reliability across varied institutional settings. This involves ensuring accuracy for diverse patient demographics and adapting to distinct specimen preparation methods and clinical practices. 2) Influence on medical diagnostics and expert confidence. Assessing how our tools impact colon cancer decision making and benchmarking them against expert judgment. This involves presenting a range of diagnostic display options to gauge confidence in the assessments. 3) User interface refinement: Tailoring the user interface of our diagnostic tools to meet the specific needs of pathologists and oncologists.

Implications. TNM-R aims to enhance traditional recurrence risk assessments by leveraging advanced **VMI** techniques to infer risk with accuracy comparable to, or potentially surpassing, existing methods, while also being more cost-effective and interpretable, making spatial assessment more feasible and beneficial on a wider scale.

Why not predict recurrence from H&E slides alone? 1) Our prior research indicates that incorporating spatial transcriptomics data into initial model training significantly enhances performance in tasks such as metastasis prediction (among others) compared to using H&E slides alone³⁸, 2) **VMI** significantly expands the potential for spatial molecular assessments in large-scale studies, enabling more extensive biomarker discovery and development by reducing the cost, time, personnel and instrumentation to acquire data, 3) TNM-R offers an interpretable recurrence risk score to aid clinical decision-making which could lead to changes in tissue stain ordering practices and open up new avenues for personalized treatment strategies.

C. STUDY TEAM. Our interdisciplinary, translational research team is uniquely positioned to develop and implement sophisticated **VMI** technologies for recurrence assessment, with strengths spanning AI, oncology, single-cell/spatial genomics, and digital pathology, and a desire to translate innovations into clinical practice.

Core MPI team. Led by Drs. Joshua Levy (Cedars Sinai) and Louis Vaickus (Dartmouth Health) (**see MPI Leadership Plan**), both recognized for their significant contributions to digital pathology and AI. Together, they have collaborated on over 40 peer-reviewed papers in the last four years, increasingly focusing on spatial transcriptomic and proteomic profiling relevant to this grant. Their joint efforts have led to the establishment of mature digital pathology infrastructure across two research institutions, including development of four research/translational computing clusters and a national pathology AI internship program. Dr. Levy, as former Director of Machine Learning Research at Dartmouth Health and in his current role at Cedars Sinai as Director of Digital Pathology Research, brings a wealth of experience in developing and utilizing research infrastructure at both institutions. His work enabled through Pilot and Project COBRE funding has been pivotal in gathering the preliminary data for this R01 project, developing relationships across shared resources and cores such as Genomics and Pathology which led to significant advancements in data quality and acquisition. Dr. Vaickus, as a GI pathologist and an expert in pathology informatics and deep learning, has been integral to advancing AI applications in pathology. For instance, his work in developing graph neural network methodologies for whole slide images is of significant relevance to this proposal. R01 funding would mark a significant transition from their respective individual career development awards (COBRE, K08) to broader research and translational impact.

Pathological Expertise. GI pathologists, including Drs. Louis Vaickus and Xiaoying Liu from Dartmouth, as well as Dr. Maha Guindi from Cedars, will provide clinical expertise on spatial assessments and advise on how TNM-R can augment staging practices. Dr. Guindi serves as Director of GI Pathology and Medical Director of Digital Pathology— she has experience in NIH funded comparative assessments to ctDNA for CRC recurrence.

GI Oncology. Alexandra Gangi is Director of the Gastrointestinal Tumor Program in the Department of Surgery and is a board-certified surgical oncologist and surgeon whose primary research interests include metastatic colon cancer. She has received NIH funding for studying ctDNA and liver metastasis in context of colon cancer and will consult on how the **VMI** and TNM-R score technologies can enhance disease management practices.

Spatial Genomics Infrastructure and Analysis. Fred Kolling leads the genomics core at Dartmouth Cancer Center and serves as associate director of shared resources and adds essential expertise in spatial transcriptomics and computational biology. His knowledge of Visium technology is critical for guiding our experimental design and ensuring high-quality DNA/RNA sample integrity. Dr. Kolling's collaborative history with Pathology Shared Resources and Dr. Levy, particularly in spatial proteomics and transcriptomics, adds vital expertise. Complementing the team, from Cedars Sinai is Alexander Xu, the Associate Director of the Spatial Profiling Core. He brings his analytical expertise in spatial biomarker development. Working alongside him is Simon Knott, the Director of the Genomics Shared Resource at Cedars, who contributes his knowledge and experience in spatial assays. Kolling, Xu, Knott and Levy have working relationships with 10x Genomics at both institutions, providing cost benefits and enhancing collaborative opportunities (see 10X letter of support).

Statistical Oversight. Complementing Dr. Levy's biostatistics background, Dr. Jiang Gui will offer additional statistical support, having mentored Dr. Levy in statistical methods during his COBRE project. Dr. Gui's expertise is especially notable in high-dimensional omics risk prediction models and in managing complex dependencies and risk assessments across various cancer types. Together, Drs. Levy and Gui are Co-PIs on a Department of Defense project focused on biomedical DL methods for natural language processing.

Pathology Informatics, Cloud Architecture, and Institutional Infrastructure. Keluo Yao and Louis Vaickus, as Pathology Informatics Directors at Cedars and Dartmouth respectively, will lead efforts to streamline data collection, from specimen handling to WSI aggregation, clinical data retrieval, and synchronize records with Pathology LIS systems and clinical EHRs. Further, our team includes DH members skilled in both human-centered design (HCD) and human-computer interaction (HCI) design, as well as cloud engineering. Dr.

Elizabeth Murnane will provide expertise in HCD/HCI for enhancing acceptability and usability of our tools, while Dr. Parth Shah will guide cloud infrastructure across institutions via *Augmet* platform while facilitating access to complex comparative assay data (ctDNA) at Dartmouth. Dr. Karine Sargsyan who directs the OncoBiobank at Cedars Cancer will provide similar ctDNA access at Cedars through *Molecular Twins* initiative.

D. PRELIMINARY RESULTS. Data Demonstrating

Feasibility of Proposed Research. Our team has been instrumental in delineating multiplexed spatial proteomic and transcriptomic signatures associated with CRC metastasis and recurrence. This, coupled with our work in spatial RNA inference from histology, positions us uniquely to advance the field, highlighted through the following recent research (**Figure 1**):

1) Mapping Transcriptomic and Proteomic Signatures. Under the previous COBRE project, we published spatial transcriptomic/proteomic signatures related to colon tumor metastasis/recurrence using ST and DSP, stratified by tissue architecture. For example, we identified biomarkers such as granzyme B and forkhead box P3, within and around tumor for CRC metastasis risk, especially in scenarios where lymph node examination does not provide clarity^{39,40}.

2) Large Scale Validation of VMI Technologies. We have published several large-scale validation studies have affirmed the credibility and utility of **VMI** technologies across various pathological conditions (Liver Fibrosis Staging, SOX10 IHC, etc.)^{31,32,41}. In a separate study, we also developed inference algorithms for spatial proteomic data for lymphocyte prediction from co-registered same-section IF and H&E in CRC, using detection (YOLOv8) and cell graph neural networks³³. Our project aims to explore the clinical utility of these tools, specifically by using spatial features derived from **VMI** to develop a TNM-R score.

3) Pilot Feasibility Studies Marrying Histological Imaging with Spatial Transcriptomics. This integration showcased excellent capability in inferring spatial RNA patterns of 1000 genes through various algorithms in several pilot colon cohorts^{37,38,42,43}. We also observed that certain algorithms performed better based on the biological pathways under examination. For instance, Graph Neural Networks (GNNs) were more effective in predicting genes associated with large-scale architectural changes (e.g., E2M transition), while Convolutional Neural Networks (CNNs) identified local features related to proliferative activity and metabolism. In a separate study, contrastive pretraining methods comparing ST and underlying histology improved the representative capacity of inferred histomolecular data for downstream tasks, such as inference of metastasis from the primary site^{37,38,42,43}. A study we conducted on skin tumors yielded similar findings⁴⁴.

4) Enhancing scope, scale and precision of ST data inferred from histology alone. We pioneered the development of cell graph neural networks (CellGNN) to interpret spatial mRNA expression from WSIs based on the alignment of imaging features from detected cells across a slide to co-registered scRNASeq data from serial sections using ST to facilitate the mapping⁴². In a separate study, enhancements in tissue quality (enabled through automated tissue staining and 40X magnification—commonly referred to as 40X and referred henceforth—image scanning via Leica Aperio GT450 with spatial integration via the 10X CytAssist assay) led to substantial improvements in predictive performance³⁷. This contrasts with nearly all ST **VMI** studies that utilize lower imaging quality, unsuitable for clinical use^{28,45,46}.

5) Ongoing large-scale inference efforts. We are applying **VMI** internally at DH to study large-scale molecular changes and metastasis/recurrence risk across 1000 CRC WSI and 300 skin WSI.

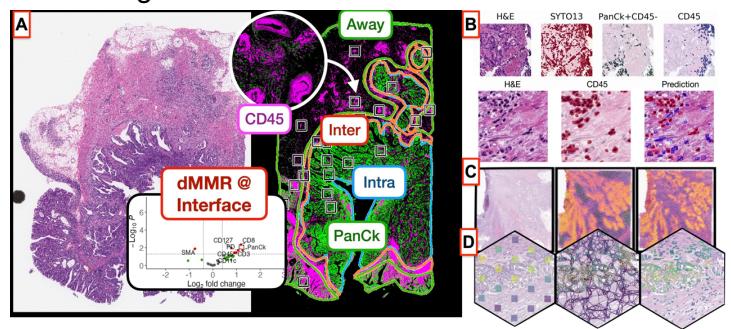


Figure 1: Select Preliminary Data Aims 1&2: **A)** Identification of spatial markers of metastasis at distinct tissue architectures, **B)** Cell graph neural networks facilitate VMI various immunostains in Colon; **C)** ST inference through CNN patch-based VMI; **D)** Cell GNN VMI for single cell ST inference

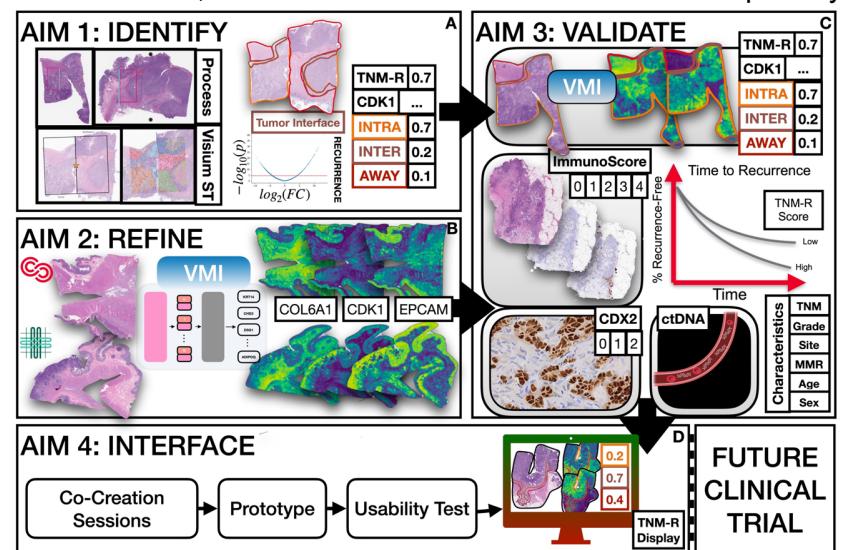


Figure 2: Approach Overview: **A)** Identify spatial markers of recurrence from Visium ST through differential expression/time-to-event analysis, **B)** Refine: VMI models on Visium ST from Cedars and Dartmouth, inference on held out WSI; heatmaps on right from unpublished VMI trained on COBRE data; **C)** Validate: Comparative assessment of TNM-R score, derived from spatial VMI features, to other prevailing assays, **D)** Interface: Iterative design process to develop and test interface for future clinical trial.

6) Collection efforts. Our team has successfully collected nearly 100 Visium ST profiles from a diverse range of tissues, including 50 colon, 16 skin, 24 breast, and 8 placenta— manuscripts are in preparation affirming initial algorithmic findings. Initial data collection was performed at Dartmouth, with unique patient demographics, specimen preparation and imaging infrastructure, warranting multicenter collection.

E. APPROACH. Overview. This R01 proposal, building on yet distinct from prior COBRE insights, uniquely integrates biomarker development with clinical application, showcasing the enhancement of prognostication through the fusion of modern computational tools and traditional pathology. In Aim 1 (**IDENTIFY**), we will identify recurrence biomarkers through Visium ST across two institutions. Aim 2 (**REFINE**) advances **VMI** techniques within a multicenter framework, with the primary goal of accurately inferring ST from routine H&E histology slides. The subsequent Aim 3 (**VALIDATE**) involves using the developed algorithms to predict on held out slides and a comparative analysis, where **VMI** biomarkers are evaluated against prevailing recurrence risk predictors. Finally, Aim 4 (**INTERFACE**) is dedicated to the iterative development of a user-friendly tool for clinical use, aimed at enhancing decision-making and collaboration between pathologists and oncologists. Mitigating dependence of aims. Aims 1 (biomarker identification) and 2 (**VMI** algorithms) operate independently, though provide foundational insights for Aim 3, which applies **VMI** on unseen WSI to identify recurrence markers (TNM-R) with comparative validation (**Figure 2**). Strong preliminary data underpins success in these areas, reducing risks associated with interdependence of aims. Aim 4, dedicated to tool configuration, operates independently from earlier aims.

E.1. Aim 1 (IDENTIFY): Multicenter Biomarker Identification Using Visium ST Data. The goal of this aim is to identify spatial biomarkers for CRC recurrence/metastasis using Visium ST, harnessing Cedars and Dartmouth's resources, profiling across various tissue types and patient characteristics, for **VMI** biomarker validation in Aim 3.

E.1.1. Multi-Center Data Collection. 1.1.1. Cohort Characteristics. We have secured IRB approval at Dartmouth (pending at Cedars) for a study involving 250 specimens for WSI and ST profiling. Selected patients at Dartmouth and Cedars Sinai from 2016 to 2023 (**Table 1**), cover ascending, transverse and descending colon adenocarcinoma cases with 3–5-year recurrence outcomes, excluding patients who received prior chemotherapy or immunotherapy. We will collect comprehensive demographic and tissue specimen information. Matching based on age, sex, tumor grade, tissue size, MMR status, tumor staging assessed using pTNM system. MMR status is determined by immunohistochemical analysis of MLH1/PMS2/MSH2/MSH6 expression. Tissue resections will be processed into paraffin blocks, sectioned into layers ranging from 5-20 microns in depth as required for each assay.

1.1.2. Whole Slide Imaging. To be collected at 40X resolution via Leica Aperio GT450 Scanners.

1.1.3. Visium ST Profiling. Profile up to 18,000 genes per sample, capturing spatial variations at near single-cell resolution. The workflow integrates CytAssist technology for efficient sample preparation, including deparaffinization, rehydration, and H&E staining of FFPE tissue sections, followed by high-resolution imaging and precise spatial detection of RNA probes. Pathologists place capture areas aligned with diverse tissue macroarchitectures. Histotechnicians will manually dissect these areas for genomics analysis. Sequencing is conducted on an Illumina NovaSeq, leveraging barcoded spots for detailed mRNA capture. The Spaceranger software aligns CytAssist sections with corresponding H&E stains, ensuring accurate co-registration, enabling development of algorithms for predicting RNA transcript counts at 8 to 55 µm resolution from H&E.

1.1.4. Integration of Single Cell Data with ST for comprehensive understanding of spatial tissue composition.

1) Profile: Employing the Chromium Flex assay, we will analyze disaggregated FFPE tissue sections through single cell RNA sequencing (scRNASeq) with same transcriptomic probes as Visium ST, enabling near single-cell spatial resolution and identification of cell type proportions. 2) Preprocess: Data processed with CellRanger v7.1.0 provides quality control metrics and expression matrices. 3) Label: We use label transfer from a public scRNA-seq dataset of colon cells to our in-house CRC scRNA-seq dataset, employing a semi-supervised variational autoencoder (scANVI), scBERT (Large Language Model-based), and marker-based methods for cell labeling^{47–51}. 4) Map and Deconvolve: SpatialScope⁵² package for deconvolution (e.g., endothelial, CD4 T-cells), aligning spatial transcriptomics data with serial section scRNA-seq to map cell-type distributions across CRC sections, providing detailed cellular abundance— prior preliminary analysis across ten CRC samples.

Table 1: Cohort Characteristics: Initial data collected from prior COBRE work and estimated characteristics for Aims 1&2 (Visium ST data) and Aim 3 (comparative assays); LoF=Loss of Function

Characteristic	COBRE (N=57)	Aims 1&2 (N=250)	Aim 3 (N=2000)
Sex (Male)	33 (57.9%)	143 (57.2%)	1163 (58.1%)
Age (<50)	9 (15.8%)	39 (15.6%)	322 (16.1%)
Age (50+)	48 (84.2%)	232 (92.8%)	1685 (84.2%)
Ascending/Transverse	29 (50.9%)	110 (44.0%)	1025 (51.2%)
Descending/Other	27 (47.4%)	124 (49.6%)	954 (47.7%)
T-stage=3	57 (100.0%)	100 (40.0%)	800 (40.0%)
N-Stage=0	29 (50.9%)	131 (52.4%)	1029 (51.4%)
N-Stage=1	20 (35.1%)	83 (33.2%)	696 (34.8%)
N-Stage=2	8 (14.0%)	29 (11.6%)	279 (14.0%)
Distant Metastasis	19 (33.3%)	75 (30.0%)	580 (29.0%)
Local Recurrence	17 (29.8%)	65 (26.0%)	524 (26.2%)
MLH1/PMS2 LoF	16 (28.1%)	65 (26.0%)	517 (25.9%)
MSH2/MSH6 LoF	0 (0.0%)	10 (4.0%)	80 (4.0%)
Visium ST	57 (100.0%)	250 (100.0%)	0 (0.0%)
CDX2	0 (0.0%)	0 (0.0%)	2000 (100.0%)
Immunoscore	0 (0.0%)	0 (0.0%)	2000 (100.0%)
ctDNA	0 (0.0%)	0 (0.0%)	700 (35.0%)
Dartmouth	57 (100.0%)	125 (50.0%)	1000 (50.0%)
Cedars	0 (0.0%)	125 (50.0%)	1000 (50.0%)

1.1.5. Cohort Expansion. Initially, our study at Dartmouth focused on ~50 Visium-assayed specimens, mainly from pathological T-stage III (pT3). We are now broadening this to include stages pT2 to pT4, aiming for a total of 250 specimens (125 each from Dartmouth and Cedars), to control for tumor invasiveness with enhanced statistical power. To optimize resources, we will place four specimens onto each capture area, similar to a Tissue Microarray (TMA). This method, which supports profiling of eight patients per slide, requires only 32 tissue slides for CytAssist, a cost-effective strategy successfully implemented in our prior COBRE project³⁷.

1.1.6. Tissue Annotation. GI pathologists will annotate H&E WSI sections, identifying three tissue regions: intratumoral, invasive margin (around tumor), and away from tumor. They will also label CRC architectures like epithelium, lamina propria, and muscularis propria. We have annotated more than 700 WSIs⁵³.

1.1.7. Public Data Release. We are creating the largest publicly accessible Colon ST dataset, distinguished by its pairing with 40X tissue scans and **VMI**. The dataset, including patient demographics, WSI, ST and scRNASeq data, will be made available in public repositories. We will adhere to NGFF and MITI standards, and store data in OME-TIFF/ZARR/JSON format, ensuring broad bioimaging access. This release includes tissue masks, segmentation masks, polygonal contours, and instance segmentation masks. Our team's expertise in image co-registration, annotation, and machine learning analysis will ensure effective data integration³³.

E.1.2. Spatially-Resolved Biomarker Discovery Workflow: Focusing on the abundance, proportions, and arrangement of cells within TIME, we aim to uncover patterns linked to clinical outcomes such as recurrence and metastasis. This approach involves three key steps: **1) Extract Spatial Features:** Improve upon traditional methods like IHC or H&E staining by leveraging highly-multiplexed spatial resolution to study novel cellular, genetic, and spatial configurations predictive of patient outcomes. **2) Identifying Individual Markers:** Stratifying patients based on spatial features and employing differential expression or time-to-event analysis to evaluate the biomarkers' effectiveness in predicting metastasis or recurrence. **3) Developing Composite Markers:** Utilizing advanced machine learning techniques to identify a panel of predictive markers.

E.1.3. Spatial Features Extraction with Visium Assay. We will utilize the Visium assay to extract a comprehensive set of spatial features correlated with colon cancer recurrence. These features include: **1) Differential Expression Analysis:** Compare gene expression by recurrence status, enabling a more nuanced understanding of how gene expression varies across different tissue type, architecture, and recurrence status (see E.1.4). **2) Cell Type Proportions Incorporating Single Cell RNA-Seq:** Integrate scRNASeq to accurately determine proportions of various cell types at each ST spot⁵⁴. **3) Cell-Type Specific Differential Expression:** Investigate differences in gene expression by recurrence, stratified by TIME cell-type using the C-SIDE algorithm⁵⁵. **4) Cell-Cell Interaction Ligand-Receptor Analysis:** Explore interactions between different cell types^{56,57}. **5) Spatially Variable Genes:** Identify genes whose expression levels vary spatially across slide via SPARK, Moran's I, cluster analysis⁵⁸. **6) Neighborhood Enrichment Analysis:** Examine local spatial context of cells to understand enrichment of certain cell types or gene expressions in specific neighborhoods within tissue²⁵. **7) Pathway Activity Scores and Factor Analysis:** Use tools like decoupler to acquire pathway activity scores or factor analysis to group and aggregate expression, providing functional insights⁵⁹.

E.1.4. Statistical Analysis in Biomarker Identification. Focus on identifying biomarkers that differentiate patients with and without tumor recurrence, stratified based on dichotomized endpoints at 3 and 5 years, as well as time-to-event analyses using Cox proportional hazards (CoxPH) models to estimate instantaneous hazards⁶⁰. For instance, differential expression analysis based on dichotomized recurrence endpoints will utilize either negative binomial or lognormal regression modeling for gene expression y_i , considering confounders/effect modifiers (\vec{z}_i) like tissue architecture, metastasis status (N/M-stage), MMR, age, sex, race/ethnicity, random hospital and patient intercepts b , and spatial autocorrelation $k_{\Sigma[i]}$, where distance between two spots is $d(s_i, s_j)$:

$$(1) y_i \sim N(\log(\mu_i), \sigma^2); \mu_i = \beta_0 + \beta_1 \text{recurrence}_i + \vec{\beta}_z \cdot \vec{z}_i + b_{\text{patient}[i]} + b_{\text{hospital}[i]} + k_{\Sigma[i]}$$

Findings will be stratified by tissue architecture using interaction term $\text{architecture}_i * \text{recurrence}_i$. **We will control for age and sex as biological variables and proactively oversample minority races and ethnicities to reduce bias wherever feasible.** We will additionally consider time-to-event based associations with recurrence, applying CoxPH given aggregate spatial features \vec{x} : $h(t|\text{recurrence}_i) = h_0(t)\exp(\mu_i)$; $\mu_i = \beta_0 + \vec{\beta}_1 \vec{x} + \vec{\beta}_z \cdot \vec{z}_i + b_{\text{hospital}[i]}$. To account for the assumption of dependence of nested spatial observations and avoiding false positive findings by assuming i.i.d., we will develop a classifier to estimate tumor recurrence to derive a biomarker panel, utilizing tree boosting models (e.g., GPBoost, BART) within a Mixed Effects Machine Learning (MEML) framework^{39,61-63}, using Gaussian processes for patient/hospital effects and spatial autocorrelation: $\mu_i = f_\phi(\vec{x}_i) + b_{\text{patient}[i]} + b_{\text{hospital}[i]} + k_{\Sigma[i]}$. Note: $k_{\Sigma} \sim MVN(0, \Sigma)$; $\Sigma_{i,j} = \alpha^2 \exp(-\rho^2 d(s_i, s_j)^2)$

E.1.5. Pathway analysis for molecular mechanisms. Conduct using g:Profiler, enrichR and GSEA, focusing on statistically significant genes to identify key biological pathways from databases like GO and KEGG⁶⁴⁻⁶⁶.

E.1.6. Power Analysis. Our power analysis is based on the differential expression model from *Eq. 1* under the null hypothesis that the marker does not stratify patients by recurrence (assuming no tissue architecture, $H_0: \beta_1 = 0, H_A: \beta_1 \neq 0$). We expect at least 100 genes/spatial features which are differentially expressed within one or more spatial architectures. Simulating 1000 datasets of 1000 spatially-variable genes, 100 of them differentially expressed by recurrence with effect size $\beta_1 = 0.3$ via *Eq. 1* (*simr*)⁶⁷ yielded power 0.97 for Bonferroni-adjusted $\alpha = 0.00005$ and an obtainable sample size of 250 patients, representing approximately 1 million Visium spots.

E.1.7. Potential Pitfalls and Alternative Approaches. **1) Cost effective tissue annotations at scale:** Due to the study's scale, we are considering efficient annotation methods, such as integrating Segment Anything Model (SAM)^{68,69}. SAM uses contrastive self-supervised deep learning to create accurate tissue region contours from coarsely annotated regions. **2) Recurrence modeling through differential expression.** Modeling spot-level expression as an outcome effectively handle technical replicates to increase our effective sample size, avoiding the data aggregation issues of treating it as a predictor, justifying comparison with MEML approaches. **3) Alternative approaches to mitigate multiple comparisons:** Limit analyses to variable/spatially variable genes, Bayesian prior for Type I control, selection methods such as the Horseshoe LASSO, variance inflation factor and projection predictive selection⁷⁰⁻⁷². **4) Resolution enhancement for spatial features.** To enhance the resolution of Visium data for spatial feature extraction, we will utilize models like GraphST and SpatialScope for single-cell expression inference where ST, scRNASeq and WSI are available⁷³. **5) Validating findings.** For validating biomarker findings, we plan to assess the reproducibility of the Visium ST findings using the 10X Genomics Xenium ST assay⁷⁴. This will involve selecting and conducting similar evaluations on the top-100 markers from *E.1.4.* at the single-cell level, on select specimens within budgetary constraints. Post-Xenium, sections will undergo IF/IHC staining for further validation using routine histological stains via the Pathology and Microscopy Shared Resources. **6) Enhanced pathway analysis.** Weighted gene co-expression network analysis and Conserved, Differential, Specific differential co-expression will demonstrate how gene-gene associations, spatial features, can vary by tumor progression^{40,75}. We will consider using enrichR Knowledge Graph⁶⁶, a tool that not only condenses results from diverse pathway databases to prioritize genes but provides integrated network visualization to reveal hidden gene associations.

Table 2: Select neural network architectures compared for VMI development

Approach	Description
CNN Variants	Convolutional Neural Networks, such as Inceptionv3, to capture histomorphological features across different scales.
Patch-Level CNN-GNN	Employ CNNs for extracting fine-grained histological features from tissue image patches, coupled with Graph Neural Networks to interpret patch-level features within the broader spatial tissue context.
Vision Transformer (ViT)	Similar to GNN. Analyze discrete tissue patches by integrating contextual data from adjacent areas, thereby offering a holistic view of tissue organization.
CellGNN	Utilize CNNs for detailed morphological analysis at the cellular level after cell detection, with GNN processing to map cell-cell relationships within a larger tissue environment. Elucidate cell-tissue interactions. Trained end-to-end to enhance convolutional and graph features.
Multi-Level Fusion	Synthesize and integrate feature sets derived from the above methodologies to construct a comprehensive, multi-faceted model for spatial transcriptomics, potentially enhancing predictive accuracy.
Mixture of Experts	Implement an ensemble learning strategy, selecting the most effective algorithm for predicting gene expression in each specific gene, thereby aiming to optimize the precision of gene expression profiling.

E.2. Aim 2: (REFINE) VMI Algorithm Calibration and Refinement in a Multicenter Context. Independent from Aim 1, refine initial **VMI** algorithms developed at Dartmouth using multi-institutional ST/WSI data. Compare/combine **VMI** model architectures/targets to predict ST data from standard H&E WSI. Validate that **VMI** data inferred from H&E staining effectively represent tissue histology, confirmed through pathway analysis comparing VMI data across tissue architectures.

E.2.1. Data collection. Same as *E.1.1*.

E.2.2. WSI Preprocessing for Subsequent

Feature Extraction. To generate tissue patches corresponding to molecular data, we will employ two distinct approaches for image feature extraction from H&E WSI: **1) Patch-Based Extraction:** Segment WSI into non-overlapping patches, centered around each Visium spot at a resolution of approximately 55 microns per spot (~256 pixels at 40X resolution). Patches will serve as input for **VMI** model predictions. **2) Cell-Based Extraction:** Leverage large-scale same section co-registered IF and H&E stains collected from prior COBRE, along with large-scale Lizard (large-scale CRC nuclei instance segmentation dataset) and in-house annotated CRC nuclei instance detection datasets (~60,000 annotated cells)⁷⁶. Use YOLOv8 panoptic segmentation framework to construct cell graphs based on spatial proximity of cells in H&E WSI, with small image patches around cells as features (nucleus centric tissue patches)^{33,42,77}.

E.2.3. Overview of Neural Network Approaches for ST Inference and finetuning. We plan to compare several **VMI** algorithms to infer ST from H&E slides using paired same-section Visium spots, see **Table 2**. Spot-level gene expression will be predicted through a zero-inflated negative binomial loss function. Performance will be assessed using Area Under the Curve (AUC) for dichotomized expression; Mean Absolute Error (MAE), and Spearman correlation will be used for counts. Additionally, predictive performance will be evaluated across key subgroups, categorized by variables like age, sex, race/ethnicity and MMR status. The best performing model for each gene will be integrated into **VMI**, i.e., ensemble-based Mixture of Experts, to enhance accuracy.

We will explore methods to improve **VMI** performance including: 1) pretraining models on histology data³⁸, 2) exploring self-supervised techniques such as utilizing large scale contrastive datasets and cross-modal

pretraining^{38,78}, 3) Superresolution of gene expression using SpatialScope at the cell level⁵², 4) Cell Type Proportion Prediction and 5) Topological Alignment and Co-Mapping of Single Cell Image and scRNASeq (i.e., Wasserstein Loss)⁴². We have previously shown these methods can enhance model performance.

E.2.4. Strategy for Gene-Level Inference

Target Selection. In selecting gene panels for **VMI**, we recognize the complexities of gene expression prediction, influenced by diverse biological pathways and their relevance. Our multi-strategy approach, outlined in **Table 3**, aims to balance extensive gene prediction with the necessity for targeted, biologically plausible, and clinically pertinent inference targets. This involves the development of specialized **VMI** component models, each tailored to different, relevant subsets of genes influenced by key pathways.

E.2.5. Initial Validation of Inferred ST Markers

through Prediction of Underlying Histological Structures. In our validation process and in line with prior work, we will leverage the **VMI** predictions to segment and classify different tissue architectures annotated in E.1.1.6. This is to ensure genes associated with each type of tissue are accurately identified **AND** relevant to the expected histological structures. This will be accomplished using advanced machine learning techniques, such as BART, GNN and MEML. To confirm the effectiveness of our approach, we will employ interpretation tools like Captum, SHAP, and GNNExplainer, along with differential expression analysis and spatial generalized linear mixed effects models^{39,79-81} (**Figure 3**). Our validation criteria are twofold: firstly, the inferred **VMI** markers should predict the underlying histology with at least the same accuracy as the original ST data from Aim 1. Secondly, there should be a significant overlap in the genes and biological pathways identified between the **VMI** and original ST markers, indicating consistency and reliability in our inferential methodology.

E.2.5. Power Analysis. Aim 2 Primary

Endpoint: Average predictive accuracy for **VMI** algorithms described in E.2.2-E.2.4, **Secondary endpoints** 1) Ability to predict underlying tissue histological structures and 2) recapitulate pathways tied to tissue architecture between ST and **VMI** predicted genes. Our power analysis is based on primary endpoint of AUC and Spearman correlation between predicted/true expression (expected to exceed the literature benchmark, $H_0: \rho = 0.6, H_A: \rho > 0.6$). With 250 patients and 2,500 Visium spots per patient, we expect to have an AUC of at least 0.8 compared to nominal value of 0.62 using two-sided z-test, yielding power of 80% and Bonferroni adjusted type I error of 5%. For comparing Spearman correlation with benchmark, a sample size of 250 achieves 80% power to detect a difference of 0.18 between the null hypothesis correlation of 0.6 and the alternative hypothesis correlation of 0.78 using a one-sided hypothesis test with a significance level of 0.000003 (Bonferroni correction across 18,000 genes)⁸².

E.2.6. Model Accessibility and Integration. We will create a user-friendly API for easy uploading and analysis of WSI which returns **VMI** data in OME-ZARR format. Additionally, a web application will be developed using the ViV software, allowing interactive **VMI** result viewing. For wider research collaboration, selected models will be shared on public platforms like HuggingFace, promoting open science, encouraging further development^{83,84}.

E.2.7. Potential Pitfalls and Alternative strategies. Improving Specimen quality. A key challenge in accurately predicting spatial expression patterns is maintaining high-quality specimen preparation. To address this challenge, our approach involves the use of the 10X CytAssist device. CytAssist streamlines the preparation and profiling of tissue samples using custom slides that are designed for imaging, eliminating the need for specialized tissue mounting mediums. Importantly, this method is compatible with automated staining processes and supports 40X image scanning for maximum spatial resolution. CytAssist not only enhances the efficiency of our workflow but also ensures adherence to the high-quality standards essential in pathology. Additionally, we've reduced assay costs by combining multiple specimens per capture area, ensuring broader patient and tumor

Table 3: Strategies to enhance prediction of genes through specialized VMI

Strategy	Description
All Genes and Biological Pathways	Implement comprehensive pretraining across all biological pathways, use transfer learning to focus on specific pathways through pathway-specific models.
Spatial Variation, Original Visium Data Insights	Apply methods like SPARK to identify high spatial autocorrelation genes across slides expected to have high performance, combine with Visium findings from Aim 1 on recurrence-related genes.
Recursive Target Selection and Genetic Algorithm	Iterative selection and optimization for top-performing genes, progressively refine gene set towards those with highest predictive accuracy for expression inference & recurrence prediction.
Random Selection and Non-redundant Features	Begin with a random selection of gene subsets, followed by a focused identification of non-redundant genes that are maximally predictive of the remaining genes, permitting additional inference.
Unsupervised Gene Groupings	Combine hierarchical clustering and factor analysis to identify key gene sets, cross-reference with established gene sets to validate.
Clinical Utility and Established Gene Set	Integrate clinically significant genes (e.g., CDX2, CD3, CD8) from assays with established gene sets, ensuring relevance in empirical data and clinical utility.

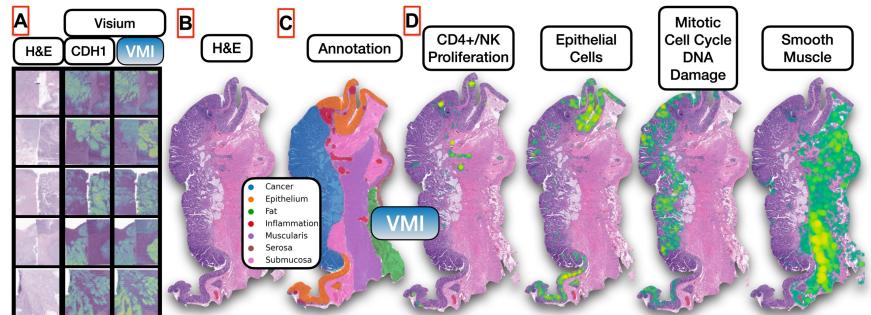


Figure 3: Preliminary data illustrating Aim 2 feasibility: A) Published work on VMI inference using Cell-GNN on held-out Visium sections, VMI models are applied to B-C) held-out WSI, annotated by pathologist, D) unpublished VMI prediction heatmaps of select genes corresponding to EnrichR pathways demonstrating agreement with histology annotations

Cancer
Epithelium
Fat
Inflammation
Muscularis
Serosa
Submucosa

representation. Pathway-level inference strategies. To tackle the issue of noise and dropouts in gene expression prediction, pathway-level inference consolidates expression into **pathway activities**, yielding more robust, interpretable, and manageable analysis. This is accomplished through: **1) Sparse Pathway Neural Network**: leverages local connectivity between genes and pathways to identify pathway signatures maximizing mutual information with histology⁸⁵, **2) Decoupler**: synthesizes complex gene expressions into pathway activity scores using a generalized linear model. Post hoc prediction leveraging gene correlational structure and potentially additional stains. To enhance the predictive accuracy of our neural network, which may inherently favor certain molecular pathways, we intend to employ post hoc models like XGBoost and GPBoost. These will utilize the correlational structure and interactions among genes to predict the expression of individual genes from the inferred data. Such models, enable focused predictions of gene expressions, considering their varying predictive abilities as identified in 1.B.4. By capitalizing on the covariance between genes, we can efficiently predict one gene's expression and assess its predictive power for others. Employ data augmentation. To enhance our dataset for our predictive models, we will employ generative and diffusion modeling for data augmentation. This method will generate synthetic histology images from RNA profiles, diversifying our training data and improving model robustness and generalizability. Our expertise in developing GANs and other adversarial architectures positions us well for the effective implementation^{31,32,38}. Single Cell Validation. Technologies like Xenium will validate findings at single-cell level (e.g., cell-cell) will complement our existing data, offering detailed insights. Validation through identification of clinically grounded features. For instance, using **VMI** to assess lymphovascular invasion (LVI) with N0 CRC cases to guide chemotherapy decisions or assessing tumor budding using VMI to identify PanCk within and away from the tumor bulk. Compare both to manual assessment (**Figure 3**).

E.3. Aim 3: (VALIDATE) Multicenter Validation of VMI and Comparative Analysis with Existing Assays. This aim expands the scope of **VMI** algorithms to large-scale cohorts, applying them to whole slide images (WSIs) of complete tissue sections. This progression moves beyond the Visium ST capture area limitations encountered in Aim 1 and extends **VMI** to WSIs lacking ST data. Such expansion is critical for enhancing spatial transcriptomics accessibility across varied CRC samples. It enables comprehensive epidemiological studies and more accurate predictive assessments of CRC histology, metastasis, and recurrence in broader real-world scenarios, where current ST methods are cost-prohibitive. Central to this aim is the development and calibration of the **TNM-R** score, tailored for predicting CRC recurrence. Our evaluation will thoroughly assess the efficacy of the TNM-R score derived from our VMI technology by comparing it against well-established prognostic benchmarks, namely the Immunoscore, CDX2 assay, and circulating tumor DNA (ctDNA) analysis. While the comparative analysis will cover a broad **patient cohort of 2000**, ctDNA profiling will be conducted on a **targeted subset of 700 patients**, selected based on availability. The aim is to meticulously determine the prognostic precision of our VMI technology relative to these conventional modalities. This structured comparison intends to validate the clinical relevance and prognostic accuracy of the VMI approach within a vast and varied clinical framework, thereby confirming its potential as a dependable instrument for CRC prognosis.

E.3.1. Multicentric data collection. Retrospective collection from 2016 to 2026 will include two phases: initially at the study's start and again in 2026, aligning with two years of Dartmouth's ctDNA collection (see 3.1.2), to facilitate analysis of three-year endpoints by the study's conclusion. **3.1.1. Cohort Expansion of WSI and Patient Metadata.** Expanding retrospective data collection to include: WSI scanning: Alongside the WSIs from Aims 1-2, we will also scan WSIs from a significantly larger cohort (**up to n=2000 patients; Table 1**) which have not been profiled on Visium. This expansion will enhance the diversity and breadth of our dataset. For slides that have been profiled for Visium, we will scan WSIs from the original, intact tissue samples as well, emphasizing the entire tissue area rather than just the capture areas. This comprehensive approach will provide a more complete view of the tissue histology. Patient Data Collection for Stratification: Similar to Aim 1, we will collect patient/tumor characteristics, enabling us to stratify patients based on metastasis and recurrence outcomes. By encompassing a wide range of tissue samples and patient characteristics, we aim to strengthen the reliability and applicability of our findings.

3.1.2. Data Collection for Comparative Assessments: The assays for comparison include: **1) Immunoscore**: Measures the density of CD3+/CD8+ T lymphocytes, around and within the tumor. **2) CDX2**: A homeobox protein assay that evaluates the expression of CDX2 in tumor cells, where loss is linked to poorer outcomes. **3) Circulating Tumor DNA (ctDNA)**: Non-invasive detection of tumor-specific variants in plasma samples, helping in assessing recurrence risk and molecular residual disease. **4) Traditional Patient Characteristics and TNM Stage**: Standard clinical metrics will serve as comparative baseline, assessing added prognostic value provided by our assay. Immunoscore will measure CD3 and CD8 lymphocyte densities within and around the tumor, using both traditional (5-point ordinal scale, 0-4) and, potentially, new digital scoring systems (stain deconvolution and cell detection) for precise quantification. For the CDX2 assay, pathologists will assess its expression qualitatively

and quantitatively (3-point ordinal scale, 0-2). Our approach to ctDNA analysis will ensure uniformity across Dartmouth and Cedars. We will consider control variables like MMR, BRAF/KRAS mutations, and ctDNA levels, amongst patient-level confounders. Considering the complementary nature of these assays, comparisons will emphasize how VMI's additional predictive value enhances and integrates with insights from existing assays.

3.1.3. Supportive Infrastructure from Dartmouth and Cedars. The Molecular Twin Data Commons initiative at Cedars-Sinai Medical Center is a pivotal component of our project, providing biospecimen data integration for precision oncology research⁸⁶. This platform will assist in the collection and analysis of ctDNA from CRC patients, offering access to a comprehensive database that includes 350 CRC patient plasma samples with ctDNA, blocks and WSI. Dr. Levy will collaborate with Cedars Cancer Biobank director Dr. Karine Sargsyan to leverage this infrastructure, facilitating the retrieval of high-quality ctDNA and tissue samples for digital pathology analysis, particularly focusing on CD3/CD8/CDX2 staining. Dr. Sargsyan's experience and role in managing large-scale biobank collections will be instrumental in expanding the depth and breadth of our dataset— for instance, she will facilitate access to WSI from WHO-IARC and ADOPT cohorts with more than 2500 CRC patients with associated metadata (e.g., recurrence).

At Dartmouth, similar support is provided by *Augmet*, led by Dr. Parth Shah. This infrastructure integrates Whole Exome Sequencing (WES) amongst other NGS assays with corresponding WSIs, enhancing accessibility and management of biospecimen data. Starting March 2024, Dartmouth will initiate a biobanking effort for all CRC plasma samples, incorporating a 100-gene panel NGS assay. This initiative will be supported in-kind by Dr. Shah, ensuring that ctDNA sequencing is integrated as a standard reimbursable procedure, reducing overall costs for our project—we expect a sample size at Dartmouth similar to Cedars Sinai. Both these infrastructures—Molecular Twins at Cedars and Augmet at Dartmouth—will significantly contribute to the collection and analysis of biospecimens, underpinning the comparative evaluation of our **VMI** algorithms against traditional assays.

E.3.2. VMI on WSIs and Spatial Feature Extraction: Utilizing the most effective algorithms from E.2. to perform **VMI** to generate spatial RNA profiles on WSIs without ST data collected, with spatial feature extraction similar to E.1.3. Utilizing algorithms from E.2.5 (Histological Validation), we will automatically segment tissue regions (as outlined in E.1.1.6 Tissue Annotation) for precise spatial feature quantification. This automation aims to consistently enhance the reproducibility of TNM-R scores across various tissue architectures.

E.3.3. TNM-R Score Development and Validation. The TNM-R score, a tool for assessing colon cancer recurrence risk, integrates biomarkers from Aim 1 and expression profiles predicted using Aim 2 algorithms (**VMI**). Ranging from 0 to 1, this score is derived from time-to-event models (CoxPH, ML models, etc.) applied to **VMI** data. The models facilitate the breakdown of the score into specific markers for different tissue regions, using regression coefficients and/or model interpretations. Calibration involves predicting patient outcomes like metastasis or recurrence, using the inferred ST data stratified by algorithmically-localized tissue architecture. Patient-level factors will also be included to refine prediction accuracy. Further, it is expected that **VMI** derived data will be able to replicate predictive genes/pathways and molecular signatures identified in Aim 1. Preliminary data underpins the feasibility of this large-scale inference. This step is vital to evaluate the practical utility of our **VMI** methods in enhancing patient prognosis and treatment strategies.

E.3.4. Comparative Assessment to Leading Assays & Statistical Analysis: The culmination of this aim features the comparative evaluation of different assays' ability to stratify recurrence risk using a Cox proportional hazards model. Each assay, like TNM-R, Immunoscore, CDX2, and ctDNA, will have its own CoxPH model with relevant covariates, assessed through hazard ratios, confidence intervals, and concordance index (C-index). Model comparisons will use likelihood ratio tests and Akaike Information Criterion (AIC), alongside metrics like NRI and IDI. Kaplan-Meier curves will visualize risk stratification and predictive accuracy. A combined multivariable Cox model will analyze all assays' collective contributions, with Harrell's C-index evaluating performance. The Fine and Gray subdistribution hazard model will handle competing risks^{87,88}, like death before recurrence censoring. Statistical significance of the TNM-R score in the presence of these other assays and covariates $\vec{z}_i, \beta_0 + \beta_1 TNM-R_i + \beta_2 CDX2_i + \beta_3 Immunoscore_i + \beta_4 \vec{b}_{\text{hospital}[i]} + \dots$, where $\beta_1 > 0$ and hospital-level random intercepts $b_{\text{hospital}[i]}$, will affirm its capacity as a clinically viable test in addition to the C-statistic comparisons. Subgroup analyses will be conducted by various patient characteristics (e.g., age, sex, MMR) using interaction terms.

E.3.5. Power Analysis. The primary objective of this aim is for the TNM-R score to demonstrate additional predictive value ($\beta_1 > 0, AUC_{TNMR} > AUC_{\text{other}}$) over the other assays. For simplicity, we used the *lme4* and *simr* packages in R to simulate 1000 datasets representing a binary 3-year recurrence outcome, based on the model described in E.3.4, for a cohort of 700 patients (see E.3.6) with $\beta_1 = 0.3$, yielding a high power of 0.96.

E.3.6. Potential Pitfalls and Alternative strategies. Specimen Preparation and Validation: Additional validation will be performed when direct ST measurements are unavailable. This will involve using immunostains and Xenium assays for confirmation, especially if discrepancies arise in TNM-R predictions of metastasis or

recurrence. **Performance Assessment in Diverse Tissue Contexts:** Expanding our analysis to whole slide images (WSIs) introduces new challenges, such as analyzing regions distant from tumors, like fat. We expect TNM-R performance to differ when stratifying by different tissue regions. **Confounding and Effect Modifiers:** Additionally, we are prepared to investigate patient-level factors beyond our initial dataset (age, sex, MMR status, and cancer stage) to understand their impact on our results. **Enhancing Predictiveness with Complementary Stains:** To address potential challenges in marker validation and enhance the predictive accuracy of the TNM-R, we may integrate confirmatory immunostains and conduct statistical analyses alongside inferred ST data. This strategy will leverage the algorithm's capacity to identify key markers for colon cancer prognosis, guiding the selection of immunostains for more precise traditional profiling. **Limited ctDNA Data Availability:** Due to the more recent collection efforts for ctDNA, its availability is limited compared to the larger, retrospective datasets of immunoscore and CDX2. This necessitates conducting two types of subgroup analyses: a smaller-scale analysis including all assays (**n=700 patients**) and a more extensive one excluding ctDNA (**n=2000 patients**). The limited ctDNA dataset may lead to lower statistical power for ctDNA-related analyses, in contrast to the higher power achievable in comparisons involving immunoscore/CDX2 and the TNM-R score.

E.4. Aim 4: (INTERFACE) Development and Usability Testing of Clinical Decision Support Interface. This aim involves developing and testing a Clinical Decision Support Interface with the TNM-R score for CRC recurrence risk. It utilizes WSI/VMI/TNM-R data from Aims 1-3 and emphasizes user-centered design, involving pathologists in the process to ensure needs, constraints, and other key requirements are addressed. The goal is to make the tool practical for clinical use and ready for future clinical effectiveness-implementation trials⁸⁹⁻⁹¹.

E.4.1. Formulating the TNM-R Score into a User-Friendly Format. The TNM-R score, as introduced in E.3.3, is estimated using a Cox-PH approach based on aggregate **VMI** spatial features. The focus of this aim is on optimizing the display and human interpretation of this score, rather than perfecting the underlying model. We will categorize recurrence risk into three levels—low, medium, and high—each distinguished by specific spatial features identified in WSI using **VMI**. These features might include, for example, the density of CD8+ cells in tumor regions. The presentation of the TNM-R to pathologists is key, emphasizing an intuitive display that is sufficiently informative without overloading the user. Maximizing usability will involve iterative design and pathologist evaluation to ensure the tool meets diagnostic needs and effectively communicates risk assessment.

E.4.2. Iterative Design Activities. Dr. Elizabeth Murnane will leverage her expertise in human factors, usability, and developing informatics tools, including decision support systems, to lead interface design and testing. As a faculty affiliate of the Center for Implementation Science, Murnane's role is crucial to create a tool that effectively communicates complex medical information, is responsive to the intricacies of clinical decision making, and supports clinical setting integration. **1) Formative Co-Creation Activities:** Our initial stage will involve collaborative design sessions with N=8-15 pathologists, oncologists, and other stakeholders. Additional participants will be recruited as necessary to capture a diverse range of experiences and to reach saturation in perspectives. These co-creation activities allow for exchange of ideas, consensus-building, and feature prioritization, ensuring the development process aligns with actual needs and constraints. **2) Prototyping:** Based on insights, we will develop prototypes of the decision support interface. Prototypes will iteratively progress in fidelity, from rough mockups that invite honest feedback and can be easily adapted, to high fidelity versions that are increasingly refined and functional. **3) Lab-Based Usability Testing:** Studies will evaluate understandability and usability, beginning with the think-aloud method, which invites users to verbally express thoughts while interacting with an interface⁹²⁻⁹⁴. We will also collect usability metrics using standard instruments (see E.4.4). If prototypes do not receive high scores, improved versions will be made to address issues.

Table 4: Candidate Designs for the TNM-R Display Output

Display Option	Description
Quantitative Format	Present number between 0-1, delineating low to high-risk categories.
Architecture Boxplots	Boxplot scoring each gene within tissue architectures, detailed breakdown similar to Immunoscore.
WSI Viewer	Develop viewer like VIV to illustrate multiplex VMI predictions, compatible with digital pathology platforms and export to OME-NGFF.
Combination: Reflexive Stain Ordering	1) Simplified viewer with result readouts and uncertainty indicators. 2) Quantitative format and architecture boxplots. 3) Interactive highlighting of important slide areas. 3) Hotspot digital displays to identify additional genes for ordering respective stains.

E.4.3. Display Considerations. The display for the TNM-R score will integrate feedback with current best practices in digital pathology. Co-design will explore a range of display options; **Table 4** presents candidate designs to seed ideas.

E.4.4 Pilot Testing. We will take a systematic approach to evaluating the tool's usability in clinical settings, following established conventions for sample recruitment, testing procedures, data collection, and analysis to ensure comprehensive and reliable results. **1) Participant Recruitment:** We will initially recruit N=10 users, a number consistent with usability testing practices and reflective of the expected number of GI collaborators at each participating site⁹⁵. If new usability issues continue to emerge in this sample, we will recruit additional participants until reaching consistency in identified problems (i.e., no new issues encountered). **2) Testing Procedure:** The usability test will encompass a tool training session, followed by a simulated assessment using twenty case scenarios per GI pathologist. This

References

1. Wong MC, Huang J, Lok V, Wang J, Fung F, Ding H, Zheng ZJ. Differences in incidence and mortality trends of colorectal cancer worldwide based on sex, age, and anatomic location. *Clinical Gastroenterology and Hepatology*. 2021;19(5):955–966. PMID: 32088300
2. Kasi PM, Shahjehan F, Cochuyt JJ, Li Z, Colibaseanu DT, Merchea A. Rising proportion of young individuals with rectal and colon cancer. *Clinical Colorectal Cancer*. 2019;18(1):e87–e95. PMID: 30420120
3. Patel SG, Ahnen DJ. Colorectal cancer in the young. *Current gastroenterology reports*. 2018;20(4):1–12. PMID: 34414781
4. Lao VV, Grady WM. Epigenetics and colorectal cancer. *Nature reviews Gastroenterology & hepatology*. 2011;8(12):686–700. PMID: 35945569
5. Cheng E, Ou FS, Ma C, Spiegelman D, Zhang S, Zhou X, Bainter TM, Saltz LB, Niedzwiecki D, Mayer RJ. Diet-and Lifestyle-Based Prediction Models to Estimate Cancer Recurrence and Death in Patients With Stage III Colon Cancer (CALGB 89803/Alliance). *Journal of Clinical Oncology*. 2022;JCO-21. PMID: 34995084
6. Slattery ML. Diet, lifestyle, and colon cancer. *Semin Gastrointest Dis*. 2000 Jul 1;11(3):142–146. PMID: 35643370
7. Senthil M, Trisal V, Paz IB, Lai LL. Prediction of the Adequacy of Lymph Node Retrieval in Colon Cancer by Hospital Type. *Archives of Surgery*. 2010 Sep 1;145(9):840–843. PMID: 20855753
8. Kamal Y, Schmit SL, Hoehn HJ, Amos CI, Frost HR. Transcriptomic Differences between Primary Colorectal Adenocarcinomas and Distant Metastases Reveal Metastatic Colorectal Cancer Subtypes. *Cancer Research*. 2019 Aug 15;79(16):4227–4241. PMID: 31239274
9. Baxter NN, Virnig DJ, Rothenberger DA, Morris AM, Jessurun J, Virnig BA. Lymph node evaluation in colorectal cancer patients: a population-based study. *J Natl Cancer Inst*. 2005 Feb 2;97(3):219–225. PMID: 32206551
10. Hartgrink HH, Velde CJ van de, Putter H, Bonenkamp JJ, Meershoek-Klein Kranenbarg E, Songun I, Welvaart K, Krieken JHJM van, Meijer S, Plukker JT, Elk PJ van, Obertop H, Gouma DJ, Lanschot JJ van, Taat CW, Graaf PW de, Meyenfeldt MF von, Tilanus HW, Sasako M. Extended lymph node dissection for gastric cancer: who may benefit? Final results of the randomized Dutch gastric cancer group trial. 77 [Internet]. 2004 [cited 2022 May 4]; Available from: <https://repository.ubn.ru.nl/handle/2066/58505> PMID: 16091352
11. Schofield JB, Mounter NA, Mallett R, Haboubi NY. The importance of accurate pathological assessment of lymph node involvement in colorectal cancer. *Colorectal Disease*. 2006;8(6):460–470. PMID: 16784464
12. Ong ML, Schofield JB. Assessment of lymph node involvement in colorectal cancer. *World journal of gastrointestinal surgery*. 2016;8(3):179. PMID: 36678187
13. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, Coussens LM, Gabrilovich DI, Ostrand-Rosenberg S, Hedrick CC, Vonderheide RH, Pittet MJ, Jain RK, Zou W, Howcroft TK, Woodhouse EC, Weinberg RA, Krummel MF. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med*. 2018 May;24(5):541–550. PMCID: PMC5998822
14. Jakubowska K, Koda M, Kisielewski W, Kańczuga-Koda L, Famulski W. Tumor-infiltrating lymphocytes in primary tumors of colorectal cancer and their metastases. *Exp Ther Med*. 2019 Dec;18(6):4904–4912. PMCID: PMC6878893
15. Zhang Y, Zhang Z. The history and advances in cancer immunotherapy: understanding the characteristics of tumor-infiltrating immune cells and their therapeutic implications. *Cellular & molecular immunology*. Nature Publishing Group; 2020;17(8):807–821.
16. Angell HK, Bruni D, Barrett JC, Herbst R, Galon J. The Immunoscore: Colon Cancer and Beyond. *Clinical Cancer Research*. 2020 Jan 15;26(2):332–339.
17. Dalerba P, Sahoo D, Paik S, Guo X, Yothers G, Song N, Wilcox-Fogel N, Forgó E, Rajendran PS, Miranda SP, Hisamori S, Hutchison J, Kalisky T, Qian D, Wolmark N, Fisher GA, van de Rijn M, Clarke MF. CDX2 as a Prognostic Biomarker in Stage II and Stage III Colon Cancer. *N Engl J Med*. 2016 Jan 21;374(3):211–222. PMCID: PMC4784450
18. Saad RS, Ghorab Z, Khalifa MA, Xu M. CDX2 as a marker for intestinal differentiation: Its utility and limitations. *World J Gastrointest Surg*. 2011 Nov 27;3(11):159–166. PMCID: PMC3240675

19. Chen G, Peng J, Xiao Q, Wu HX, Wu X, Wang F, Li L, Ding P, Zhao Q, Li Y, Wang D, Shao Y, Bao H, Pan Z, Ding KF, Cai S, Wang F, Xu RH. Postoperative circulating tumor DNA as markers of recurrence risk in stages II to III colorectal cancer. *Journal of Hematology & Oncology*. 2021 May 17;14(1):80. PMID: 34001194
20. Li H, Jing C, Wu J, Ni J, Sha H, Xu X, Du Y, Lou R, Dong S, Feng J. Circulating tumor DNA detection: A potential tool for colorectal cancer management (Review). *Oncology Letters*. Spandidos Publications; 2019 Feb 1;17(2):1409–1416.
21. Bruni D, Angell HK, Galon J. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nat Rev Cancer*. 2020 Nov;20(11):662–680. PMID: 32753728
22. Galon J, Mlecnik B, Bindea G, Angell HK, Berger A, Lagorce C, Lugli A, Zlobec I, Hartmann A, Bifulco C, Nagtegaal ID, Palmqvist R, Masucci GV, Botti G, Tatangelo F, Delrio P, Maio M, Laghi L, Grizzi F, Asslamer M, D'Arrigo C, Vidal-Vanaclocha F, Zavadova E, Chouchane L, Ohashi PS, Hafezi-Bakhtiari S, Wouters BG, Roehrl M, Nguyen L, Kawakami Y, Hazama S, Okuno K, Ogino S, Gibbs P, Waring P, Sato N, Torigoe T, Itoh K, Patel PS, Shukla SN, Wang Y, Kopetz S, Sinicrope FA, Scripcariu V, Ascierto PA, Marincola FM, Fox BA, Pagès F. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *The Journal of Pathology*. 2014;232(2):199–209.
23. Jorissen RN, Sakthianandeswaren A, Sieber OM. Immunoscore—has it scored for colon cancer precision medicine? *Ann Transl Med*. 2018 Nov;6(Suppl 1):S23. PMCID: PMC6291583
24. Mlecnik B, Torigoe T, Bindea G, Popivanova B, Xu M, Fujita T, Hazama S, Suzuki N, Nagano H, Okuno K, Hirohashi Y, Furuhata T, Takemasa I, Patel P, Vora H, Shah B, Patel JB, Rajvik KN, Pandya SJ, Shukla SN, Wang Y, Zhang G, Yoshino T, Taniguchi H, Bifulco C, Lugli A, Lee JKJ, Zlobec I, Rau TT, Berger MD, Nagtegaal ID, Vink-Börger E, Hartmann A, Geppert CI, Kolwelter J, Merkel S, Grützmann R, Van den Eynde M, Jouret-Mourin A, Kartheuser A, Léonard D, Remue C, Wang J, Bavi P, Roehrl MHA, Ohashi PS, Nguyen LT, Han S, MacGregor HL, Hafezi-Bakhtiari S, Wouters BG, Masucci GV, Andersson E, Zavadova E, Vocka M, Spacek J, Petruzelka L, Konopasek B, Dundr P, Skalova H, Nemejcova K, Botti G, Tatangelo F, Delrio P, Ciliberto G, Maio M, Laghi L, Grizzi F, Marliot F, Fredriksen T, Buttard B, Lafontaine L, Maby P, Majdi A, Hijazi A, El Sissy C, Kirillovsky A, Berger A, Lagorce C, Paustian C, Ballesteros-Merino C, Dijkstra J, Van de Water C, van Lent-van Vliet S, Knijn N, Mušinä AM, Scripcariu DV, Marincola FM, Ascierto PA, Fox BA, Pagès F, Kawakami Y, Galon J. Clinical Performance of the Consensus Immunoscore in Colon Cancer in the Asian Population from the Multicenter International SITC Study. *Cancers (Basel)*. 2022 Sep 6;14(18):4346. PMCID: PMC9497086
25. Moses L, Pachter L. Museum of spatial transcriptomics. *Nat Methods*. 2022 May;19(5):534–546. PMID: 35273392
26. Zeng Z, Li Y, Li Y, Luo Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biology*. 2022 Mar 25;23(1):83. PMID: 35337374
27. Burgess DJ. Spatial transcriptomics coming of age. *Nature Reviews Genetics*. 2019 Jun;20(6):317–317. PMID: 30980030
28. Monjo T, Koido M, Nagasawa S, Suzuki Y, Kamatani Y. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Scientific reports*. 2022;12(1):1–12. PMID: 35260632
29. Li Y, Stanojevic S, Garmire LX. Emerging Artificial Intelligence Applications in Spatial Transcriptomics Analysis. *arXiv:220309664 [cs, q-bio] [Internet]*. 2022 Mar 17 [cited 2022 May 5]; Available from: <http://arxiv.org/abs/2203.09664>
30. Jackson CR, Sriharan A, Vaickus LJ. A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC and evaluation on primarily melanocytic neoplasms. *Modern Pathology*. 2020 Apr 1;1–11. PMID: 32238879
31. Levy JJ, Azizzolshani N, Andersen MJ, Suriawinata A, Liu X, Lisovsky M, Ren B, Bobak CA, Christensen BC, Vaickus LJ. A large-scale internal validation study of unsupervised virtual trichrome staining technologies on nonalcoholic steatohepatitis liver biopsies. *Modern Pathology*. 2021 Apr;34(4):808–822. PMID: 33299110
32. Levy J, Jackson C, Sriharan A, Christensen B, Vaickus L. Preliminary Evaluation of the Utility of Deep Generative Histopathology Image Translation at a Mid-sized NCI Cancer Center. *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2020) - Volume 3: BIOINFORMATICS*. 2020;3:302–311.
33. Reddy R, Reddy R, Sharma C, Jackson C, Palisoul S, Barney R, Kolling F, Salas L, Christensen B, Brooks G, Tsongalis G, Vaickus L, Levy J. Graph Neural Networks Ameliorate Potential Impacts of

Imprecise Large-Scale Autonomous Immunofluorescence Labeling of Immune Cells on Whole Slide Images. Proceedings of the First International Workshop on Geometric Deep Learning in Medical Image Analysis [Internet]. PMLR; 2022 [cited 2023 Apr 1]. p. 15–33. Available from: <https://proceedings.mlr.press/v194/reddy22a.html>

34. Rivenson Y, Wang H, Wei Z, Haan K, Zhang Y, Wu Y, Günaydin H, Zuckerman JE, Chong T, Sisk AE, Westbrook LM, Wallace WD, Ozcan A. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng.* 2019 Jun;3(6):466–477. PMID: 31142829
35. Bai B, Yang X, Li Y, Zhang Y, Pillar N, Ozcan A. Deep learning-enabled virtual histological staining of biological samples. *Light: Science & Applications.* Nature Publishing Group UK London; 2023;12(1):57.
36. Rivenson Y, de Haan K, Wallace WD, Ozcan A. Emerging Advances to Transform Histopathology Using Virtual Staining. *BME Frontiers* [Internet]. Science Partner Journal; 2020 Aug 25 [cited 2022 May 5];2020. Available from: <https://spj.sciencemag.org/journals/bmef/2020/9647163/>
37. Fatemi MY, Lu Y, Diallo AB, Srinivasan G, Azher ZL, Christensen BC, Salas LA, Tsongalis GJ, Palisoul SM, Perreard L, Kolling FW, Vaickus LJ, Levy JJ. The Overlooked Role of Specimen Preparation in Bolstering Deep Learning-Enhanced Spatial Transcriptomics Workflows [Internet]. medRxiv; 2023 [cited 2023 Oct 15]. p. 2023.10.09.23296700. Available from: <https://www.medrxiv.org/content/10.1101/2023.10.09.23296700v1>
38. Azher ZL, Fatemi M, Lu Y, Srinivasan G, Diallo AB, Christensen BC, Salas LA, Kolling FW, Perreard L, Palisoul SM, Vaickus LJ, Levy JJ. Spatial Omics Driven Crossmodal Pretraining Applied to Graph-based Deep Learning for Cancer Pathology Analysis. *Pac Symp Biocomput.* 2024;29:464–476. PMCID: PMC10783797
39. Levy JJ, Bobak CA, Nasir-Moin M, Veziroglu EM, Palisoul SM, Barney RE, Salas LA, Christensen BC, Tsongalis GJ, Vaickus LJ. Mixed Effects Machine Learning Models for Colon Cancer Metastasis Prediction using Spatially Localized Immuno-Oncology Markers. *Pac Symp Biocomput.* 2022;27:175–186. PMCID: PMC8669762
40. Levy JJ, Zavras JP, Veziroglu EM, Nasir-Moin M, Kolling FW, Christensen BC, Salas LA, Barney RE, Palisoul SM, Ren B, Liu X, Kerr DA, Pointer KB, Tsongalis GJ, Vaickus LJ. Identification of Spatial Proteomic Signatures of Colon Tumor Metastasis: A Digital Spatial Profiling Approach. *Am J Pathol.* 2023 Apr 8;S0002-9440(23)00123-2. PMID: 37037284
41. Levy J, Bobak C, Azizgolshani N, Liu X, Ren B, Lisovsky M, Suriawinata A, Christensen B, O'Malley J, Vaickus L. Improving the Virtual Trichrome Assessment through Bridge Category Models [Internet]. 2021 Nov p. 2021.10.30.466613. Available from: <https://www.biorxiv.org/content/10.1101/2021.10.30.466613v1>
42. Fatemi MY, Lu Y, Sharma C, Feng E, Azher ZL, Diallo AB, Srinivasan G, Rosner GM, Pointer KB, Christensen BC, Salas LA, Tsongalis GJ, Palisoul SM, Perreard L, Kolling FW, Vaickus LJ, Levy JJ. Feasibility of Inferring Spatial Transcriptomics from Single-Cell Histological Patterns for Studying Colon Cancer Tumor Heterogeneity [Internet]. medRxiv; 2023 [cited 2023 Oct 15]. p. 2023.10.09.23296701. Available from: <https://www.medrxiv.org/content/10.1101/2023.10.09.23296701v1>
43. Fatemi M, Feng E, Sharma C, Azher Z, Goel T, Ramwala O, Palisoul S, Barney R, Perreard L, Kolling F, Salas LA, Christensen BC, Tsongalis G, Vaickus L, Levy JJ. Inferring spatial transcriptomics markers from whole slide images to characterize metastasis-related spatial heterogeneity of colorectal tumors: A pilot study. *Journal of Pathology Informatics.* 2023 Mar 29;100308.
44. Srinivasan G, Davis MJ, LeBoeuf MR, Fatemi M, Azher ZL, Lu Y, Diallo AB, Saldias Montivero MK, Kolling FW, Perrard L, Salas LA, Christensen BC, Palys TJ, Karagas MR, Palisoul SM, Tsongalis GJ, Vaickus LJ, Preum SM, Levy JJ. Potential to Enhance Large Scale Molecular Assessments of Skin Photoaging through Virtual Inference of Spatial Transcriptomics from Routine Staining. *Biocomputing 2024* [Internet]. WORLD SCIENTIFIC; 2023 [cited 2024 Jan 25]. p. 477–491. Available from: https://www.worldscientific.com/doi/10.1142/9789811286421_0037
45. Jia Y, Liu J, Chen L, Zhao T, Wang Y. THitoGene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics.* Oxford University Press; 2024;25(1):bbad464.
46. Zeng Y, Wei Z, Yu W, Yin R, Yuan Y, Li B, Tang Z, Lu Y, Yang Y. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics.* Oxford University Press; 2022;23(5):bbac297.
47. Yang F, Wang W, Wang F, Fang Y, Tang D, Huang J, Lu H, Yao J. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. *Nat Mach Intell.* Nature Publishing Group; 2022 Oct;4(10):852–866.

48. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Mol Syst Biol.* 2021 Jan;17(1):e9620. PMID: PMC7829634
49. Zhang AW, O'Flanagan C, Chavez EA, Lim JLP, Ceglia N, McPherson A, Wiens M, Walters P, Chan T, Hewitson B, Lai D, Mottok A, Sarkozy C, Chong L, Aoki T, Wang X, Weng AP, McAlpine JN, Aparicio S, Steidl C, Campbell KR, Shah SP. Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat Methods.* Nature Publishing Group; 2019 Oct;16(10):1007–1015.
50. Joanito I, Wirapati P, Zhao N, Nawaz Z, Yeo G, Lee F, Eng CLP, Macalinao DC, Kahraman M, Srinivasan H, Lakshmanan V, Verbandt S, Tsantoulis P, Gunn N, Venkatesh PN, Poh ZW, Nahar R, Oh HLJ, Loo JM, Chia S, Cheow LF, Cheruba E, Wong MT, Kua L, Chua C, Nguyen A, Golovan J, Gan A, Lim WJ, Guo YA, Yap CK, Tay B, Hong Y, Chong DQ, Chok AY, Park WY, Han S, Chang MH, Seow-En I, Fu C, Mathew R, Toh EL, Hong LZ, Skanderup AJ, DasGupta R, Ong CAJ, Lim KH, Tan EKW, Koo SL, Leow WQ, Tejpar S, Prabhakar S, Tan IB. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat Genet.* Nature Publishing Group; 2022 Jul;54(7):963–975.
51. Khaliq AM, Erdogan C, Kurt Z, Turgut SS, Grunvald MW, Rand T, Khare S, Borgia JA, Hayden DM, Pappas SG, Govekar HR, Kam AE, Reiser J, Turaga K, Radovich M, Zang Y, Qiu Y, Liu Y, Fishel ML, Turk A, Gupta V, Al-Sabti R, Subramanian J, Kuzel TM, Sadanandam A, Waldron L, Hussain A, Saleem M, El-Rayes B, Salahudeen AA, Masood A. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. *Genome Biology.* 2022 May 11;23(1):113.
52. Wan X, Xiao J, Tam SST, Cai M, Sugimura R, Wang Y, Wan X, Lin Z, Wu AR, Yang C. Integrating spatial and single-cell transcriptomics data using deep generative models with SpatialScope. *Nat Commun.* Nature Publishing Group; 2023 Nov 29;14(1):7848.
53. Levy J, Haudenschild C, Barwick C, Christensen B, Vaickus L. Topological Feature Extraction and Visualization of Whole Slide Images using Graph Neural Networks. *Pac Symp Biocomput.* 2021;26:285–296. PMID: PMC7959046
54. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, Tokcan N, Vanderburg CR, Segerstolpe Å, Zhang M, Avraham-David I, Vickovic S, Nitzan M, Ma S, Subramanian A, Lipinski M, Buenrostro J, Brown NB, Fanelli D, Zhuang X, Macosko EZ, Regev A. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat Methods.* Nature Publishing Group; 2021 Nov;18(11):1352–1362.
55. Cable DM, Murray E, Shanmugam V, Zhang S, Zou LS, Diao M, Chen H, Macosko EZ, Irizarry RA, Chen F. Cell type-specific inference of differential expression in spatial transcriptomics. *Nature methods.* 2022;19(9):1076–1087. PMID: 36050488
56. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat Methods.* 2020 Feb;17(2):159–162. PMID: 31819264
57. Mason K, Sathe A, Hess PR, Rong J, Wu CY, Furth E, Susztak K, Levinsohn J, Ji HP, Zhang N. Niche-DE: niche-differential gene expression analysis in spatial transcriptomics data identifies context-dependent cell-cell interactions. *Genome Biology.* 2024 Jan 12;25(1):14.
58. Zhu J, Sun S, Zhou X. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biology.* 2021 Jun 21;22(1):184. PMID: 34154649
59. Badia-i-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, Taus P, Dugourd A, Holland CH, Ramirez Flores RO, Saez-Rodriguez J. decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances.* 2022 Jan 1;2(1):vbac016.
60. Lin DY, Wei LJ. The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association.* Taylor & Francis; 1989 Dec 1;84(408):1074–1078.
61. Sigrist F. Latent Gaussian Model Boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2022;1–1. PMID: 35439126
62. Sigrist F. Gaussian Process Boosting. *Journal of Machine Learning Research.* 2022;23(232):1–46. PMID: 35439126
63. Tan YV, Roy J. Bayesian additive regression trees and the General BART model. *Statistics in Medicine.* 2019;38(25):5048–5069. PMID: 31460678
64. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, Merico D, Bader GD. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc.* Nature Publishing Group; 2019 Feb;14(2):482–517.

65. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 2016 Jul 8;44(W1):W90-97. PMCID: PMC4987924
66. Evangelista JE, Xie Z, Marino GB, Nguyen N, Clarke DJ, Ma'ayan A. Enrichr-KG: bridging enrichment analysis across multiple libraries. *Nucleic Acids Research*. Oxford University Press; 2023;gkad393.
67. Green P, MacLeod CJ. SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*. 2016;7(4):493–498.
68. Huang Y, Yang X, Liu L, Zhou H, Chang A, Zhou X, Chen R, Yu J, Chen J, Chen C, Liu S, Chi H, Hu X, Yue K, Li L, Grau V, Fan DP, Dong F, Ni D. Segment anything model for medical images? *Medical Image Analysis*. 2024 Feb 1;92:103061.
69. Lutnick B, Ramon AJ, Ginley B, Csiszer C, Kim A, Flament I, Damasceno PF, Cornibe J, Parmar C, Standish K, Carrasco-Zevallos O, Yip SSF. Accelerating pharmaceutical R&D with a user-friendly AI system for histopathology image analysis. *Journal of Pathology Informatics*. 2023 Jan 1;14:100337.
70. Bartonicek A, Wickham SR, Pat N, Conner TS. The value of Bayesian predictive projection for variable selection: an example of selecting lifestyle predictors of young adult well-being. *BMC Public Health*. 2021 Apr 9;21(1):695.
71. Bürkner PC. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*. 2018;10(1):395–411.
72. Bhadra A, Datta J, Polson NG, Willard B. Lasso Meets Horseshoe. *Statistical Science*. JSTOR; 2019;34(3):405–427.
73. Long Y, Ang KS, Li M, Chong KLK, Sethi R, Zhong C, Xu H, Ong Z, Sachaphibulkij K, Chen A, Zeng L, Fu H, Wu M, Lim LHK, Liu L, Chen J. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat Commun*. Nature Publishing Group; 2023 Mar 1;14(1):1155.
74. Janesick A, Shelansky R, Gottscho AD, Wagner F, Williams SR, Rouault M, Beliakoff G, Morrison CA, Oliveira MF, Sicherman JT, Kohlway A, Abousoud J, Drennon TY, Mohabbat SH, Taylor SEB. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nat Commun*. Nature Publishing Group; 2023 Dec 19;14(1):8353.
75. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008 Dec 29;9(1):559.
76. Graham S, Jahanifar M, Azam A, Nimir M, Tsang YW, Dodd K, Hero E, Sahota H, Tank A, Benes K. Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. Proceedings of the IEEE/CVF International Conference on Computer Vision [Internet]. 2021 [cited 2023 Oct 7]. p. 684–693. Available from: https://openaccess.thecvf.com/content/ICCV2021W/CDPath/html/Graham_Lizard_A_Large-Scale_Dataset_for_Colonic_Nuclear_Instance_Segmentation_and_ICCVW_2021_paper.html
77. Sohan M, Sai Ram T, Rami Reddy ChV. A Review on YOLOv8 and Its Advancements. In: Jacob IJ, Piramuthu S, Falkowski-Gilski P, editors. *Data Intelligence and Cognitive Informatics* [Internet]. Singapore: Springer Nature Singapore; 2024 [cited 2024 Jan 26]. p. 529–545. Available from: https://link.springer.com/10.1007/978-981-99-7962-2_39
78. Azher ZL, Suvarna A, Chen JQ, Zhang Z, Christensen BC, Salas LA, Vaickus LJ, Levy JJ. Assessment of emerging pretraining strategies in interpretable multimodal deep learning for cancer prognostication. *BioData Mining*. 2023 Jul 22;16(1):23.
79. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*. 2020 Jan;2(1):56–67. PMID: 32607472
80. Ying R, Bourgeois D, You J, Zitnik M, Leskovec J. GNNEExplainer: Generating Explanations for Graph Neural Networks. arXiv:190303894 [cs, stat] [Internet]. 2019 Nov 13 [cited 2020 Oct 25]; Available from: <http://arxiv.org/abs/1903.03894> PMID: 32265580
81. Jaume G, Pati P, Anklin V, Foncubierta A, Gabrani M. Histocartography: A toolkit for graph analytics in digital pathology. *MICCAI Workshop on Computational Pathology*. PMLR; 2021. p. 117–128.
82. Winterbottom A. A Note on the Derivation of Fisher's Transformation of the Correlation Coefficient. *The American Statistician*. 1979 Aug;33(3):142–143.
83. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, von Platen P, Ma C, Jernite Y, Plu J, Xu C, Scao TL, Gugger S, Drame M, Lhoest

- Q, Rush AM. HuggingFace's Transformers: State-of-the-art Natural Language Processing [Internet]. arXiv; 2020 [cited 2024 Jan 26]. Available from: <http://arxiv.org/abs/1910.03771>
84. Manz T, Gold I, Patterson NH, McCallum C, Keller MS, Herr BW, Börner K, Spraggins JM, Gehlenborg N. Viv: multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nat Methods*. Nature Publishing Group; 2022 May;19(5):515–516.
85. Levy JJ, Chen Y, Azizgolshani N, Petersen CL, Titus AJ, Moen EL, Vaickus LJ, Salas LA, Christensen BC. MethylSPWNet and MethylCapsNet: Biologically Motivated Organization of DNAm Neural Networks, Inspired by Capsule Networks. *npj Syst Biol Appl*. 2021 Aug 20;7(1):1–16. PMID: 34417465
86. Osipov A, Nikolic O, Gertych A, Parker S, Hendifar A, Singh P, Filippova D, Dagliyan G, Ferrone CR, Zheng L, Moore JH, Tourtellotte W, Van Eyk JE, Theodorescu D. The Molecular Twin artificial-intelligence platform integrates multi-omic data to predict outcomes for pancreatic adenocarcinoma patients. *Nat Cancer*. Nature Publishing Group; 2024 Jan 22;1–16.
87. Austin PC, Putter H, Lee DS, Steyerberg EW. Estimation of the Absolute Risk of Cardiovascular Disease and Other Events: Issues With the Use of Multiple Fine-Gray Subdistribution Hazard Models. *Circulation: Cardiovascular Quality and Outcomes*. American Heart Association; 2022 Feb;15(2):e008368.
88. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*. [American Statistical Association, Taylor & Francis, Ltd.]; 1999;94(446):496–509.
89. Abras C, Maloney-krichmar D, Preece J. User-Centered Design. In Bainbridge, W Encyclopedia of Human-Computer Interaction Thousand Oaks: Sage Publications. Publications; 2004. PMID: 36630999
90. van der Vegt AH, Scott IA, Dermawan K, Schnettler RJ, Kalke VR, Lane PJ. Deployment of machine learning algorithms to predict sepsis: systematic review and application of the SALIENT clinical AI implementation framework. *Journal of the American Medical Informatics Association*. Oxford University Press; 2023;30(7):1349–1361.
91. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, Denniston AK, Faes L, Geerts B, Ibrahim M. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature medicine*. Nature Publishing Group US New York; 2022;28(5):924–933.
92. Boren T, Ramey J. Thinking aloud: Reconciling theory and practice. *IEEE transactions on professional communication*. IEEE; 2000;43(3):261–278.
93. Nielsen J. Usability Engineering. Morgan Kaufmann; 1994. PMID: 29295265
94. Bul K, Holliday N, Bhuiyan MRA, Clark CCT, Allen J, Wark PA. Usability and Preliminary Efficacy of an Artificial Intelligence–Driven Platform Supporting Dietary Management in Diabetes: Mixed Methods Study. *JMIR Human Factors*. 2023 Aug 9;10(1):e43959.
95. Turner CW, Lewis JR, Nielsen J. Determining usability test sample size. *International encyclopedia of ergonomics and human factors*. CRC Press Boca Raton, FL; 2006;3(2):3084–3088.
96. Wang B, Rau PLP, Yuan T. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*. Taylor & Francis; 2023 Jul 4;42(9):1324–1337.
97. King H, Wright J, Treanor D, Williams B, Randell R. What Works Where and How for Uptake and Impact of Artificial Intelligence in Pathology: Review of Theories for a Realist Evaluation. *Journal of Medical Internet Research*. 2023 Apr 24;25(1):e38039.
98. Lewis JR. The System Usability Scale: Past, Present, and Future. *International Journal of Human–Computer Interaction*. 2018 Jul 3;34(7):577–590.
99. O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*. Elsevier; 2018;112:28–39.
100. Perrig SA. Measuring User Experience-Overview and Comparison of two Commonly Used Questionnaires. Thesis Commons; 2021 [cited 2024 Jan 29]; Available from: <https://thesiscommons.org/bkx5y/download?format=pdf>
101. Walsh T, Varsaluoma J, Kujala S, Nurkka P, Petrie H, Power C. Axe UX: Exploring long-term user experience with iScale and AttrakDiff. Proceedings of the 18th International Academic MindTrek Conference: Media Business, Management, Content & Services [Internet]. Tampere Finland: ACM; 2014 [cited 2024 Jan 29]. p. 32–39. Available from: <https://dl.acm.org/doi/10.1145/2676467.2676480>

102. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical care*. NIH Public Access; 2012;50(3):217.
103. Bürkner PC, Vuorre M. Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science*. SAGE Publications Inc; 2019 Mar 1;2(1):77–101.
104. Bürkner PC. Bayesian Item Response Modeling in R with brms and Stan [Internet]. arXiv; 2020 [cited 2024 Jan 26]. Available from: <http://arxiv.org/abs/1905.09501>