Predicting colon cancer metastasis through spatial molecular characterization of the tumor immune microenvironment

Project Relevance

Every year, over 150,000 Americans are diagnosed with Colorectal Cancer (CRC), and annually over 50,000 individuals will die from CRC, necessitating improvements in screening, prognostication, disease management, and therapeutic options. Spatial patterns of tumor infiltrating lymphocytes (TIL) – T cell, NK cells and B cells near the primary tumor site – and their concomitant molecular alterations can determine whether there is potential for concurrent nodal and/or distant metastasis, which can indicate the risk of recurrence and mortality. In this proposal, we aim to develop low-cost assessments for TIL-specific whole transcriptomic molecular alterations that can predict whether a tumor has or will metastasize, and the risk of tumor recurrence to complement existing disease management strategies.

Project Summary
Colorectal Cancer (CRC) is both the third most common form of cancer and cause of cancer-related deaths in the United States. Examination of axillary lymph nodes at the time of surgical resection is essential for prognostication and 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 alternative assessment methods which assess lymph node involvement through indirect mechanisms would be illuminating in cases where resection is inadequate. Tumor-infiltrating lymphocytes (TIL) and other immune cell types are important prognostic indicators in CRC. The type, density, and location of TILs with respect to the tumor, in addition to tumor-specific somatic alteration profiles, can determine TIL's effect on prognosis. Furthermore, spatially dependent, immune cell specific, proteomic and transcriptomic expression patterns inside and around tumor – the Tumor Immune Microenvironment (TIME) – can discern the coordinated immune response to tumor metastasis. The comprehensive characterization of TILs is possible using highly multiplexed spatial omics technologies, but high cost and low throughput prevent their clinical deployment. *Virtual staining* can infer molecular information at low cost from tissue histology where the morphology allows. We aim to design a low-cost *Virtual Staining* test, distilled from highly multiplexed spatial molecular information, that could complement surgical lymph node dissection for recurrence risk assessments and compete with other emerging predictors (e.g., circulating tumor DNA). In a set of stage pT3 tumors with or without nodal and/or distant metastases, we will identify spatial proteomic and whole transcriptomic markers of metastasis with digital spatial profiling and Visium spatial transcriptomics of immune cells. Single-cell transcriptomic data will complement these assessments by improving the resolution and cell-type specificity of the spatial transcriptomics data, further elucidating intercellular tissue dynamics. Identified markers will be validated through lower-cost multiplexed immunofluorescence staining. Finally, we will establish histological correspondence to identified spatial metastasis markers and develop virtual staining algorithms to convert H&E-stained tissue into validated multiplexed immunofluorescent and whole transcriptomic markers. Spatial and cell-type specific patterns of molecular markers that indicate whether a patient has or is likely to develop metastasis will be identified under this framework. Inferring such information from tissue morphology can provide a low-cost and highly interpretable adjunct molecular assessment for lymph node resection, to predict recurrence risk and response to adjuvant chemotherapy. We expect that our findings will provide preliminary data for an R01 clinical trial to compare identified markers prospectively to independent metastasis predictors (e.g., liquid biopsy).

Specific Aims

Colorectal 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, laminar, submucosal and other layers of colon is prognostic, where higher tumor stage reflects invasion depth. Although assessing regional lymph node involvement is also important for determining prognosis (e.g., risk of recurrence), variable resection quality can preclude adequate assessment for disease staging, necessitating the exploration of indirect molecular profiling methods to infer the presence of nodal metastasis[1,2].

Tumor-infiltrating lymphocytes (TIL) are an important prognostic indicator as they mediate direct antitumor cell immune responses and contribute to the recruitment [3], activation, and maturation of other immune cells [4,5]. Previous studies have shown that the type, functionality, density, and location of TILs within the tumor microenvironment (TME) inform the immune response and its anti-tumoral effectiveness. Various tumor-specific characteristics, including mismatch repair alterations, determine TIL's effects on TME and prognosis.

This proposal aims to understand how spatial patterns of highly multiplexed molecular markers (proteomic and whole transcriptomic) can serve as indicators of concurrent nodal and distant metastasis and how highly multiplexed findings can be distilled into a low-cost adjunct test to improve recurrence risk assessment. For inadequately dissected lymph nodes, this adjunct test can communicate the confidence in the examination from spatial molecular information at the primary site. Morphological information from whole slide images combined with spatial transcriptomics will inform prognostication through low-cost and rapid histological inference of prognostic molecular information[6].

***Aim 1* will identify spatial proteomic markers** associated with tumor metastasis. For stage pT3 CRC tumors, spatial profiling of immune cells within regions of interest will be conducted across age, sex, MMR alteration, and grade matched patients. Spatial profiling will consider macroarchitectural components labeled by the pathologist (intratumoral, tumor immune interface, and away from tumor). Mixed Effects Machine Learning methodologies will identify potential effect modifiers to follow up on for clinical findings.

***Aim 2* will identify spatial whole transcriptomic markers** associated with tumor metastasis. Spatial whole transcriptomic analyses will complement our spatial protein expression findings and can reveal upstream recruitment factors (e.g., cytokines). To establish and improve the reproducibility of identified markers, we will leverage cutting-edge technologies such as the Visium spatial transcriptomics and 10x single-cell FFPE assays, CytAssist technology, and Vizgen Merscope device.

***Aim 3* will examine the reproducibility and scalability of identified markers** through concordance assessments with alternative multiplexed staining (e.g., multiplexed immunofluorescence). In addition, we will establish histological correspondence to identified spatial metastasis markers through joint modeling of the histomorphology, spatial transcriptomics, and protein staining. The ST assay will leverage the newly developed CytAssist technology, which allows our team to utilize autostainers and 40X image scanning technologies in Pathology for a standardized and high-resolution assessment. An ML model will be developed to predict identified markers from matched immunofluorescence stains through *Virtual Staining* of H&E slides (*VirtualProtein*) and will feature application of graph neural networks on cellular histomorphology to localize whole transcriptomic signatures at cellular resolution (*VirtualRNA*) [7–9].

This proposal will dovetail with efforts in the Center for Quantitative Biology (CQB) COBRE Single Cell Genomics Core (SCGC) to expand spatial molecular assessments and will facilitate collaboration among the CQB COBRE, the Pathology Shared Resource (PSR), Center for Molecular Epidemiology (Epi) COBRE Biorepository Core, and the bioMT COBRE Molecular Interactions & Imaging Core (MIIT) Microscopy Shared Resource (MSR). The PI has assembled a team of clinician scientist collaborators in the departments of Pathology, Epidemiology and Medical Oncology to increase the translational impact of study findings through a team science effort. These findings will be back translated to inform basic science research and the creation of computational biology tools that facilitate advanced spatial omics analyses. Access to developed software will be democratized through incorporation into the CQB COBRE Data Analytics Core's (DAC) analysis suite.

Identifying metastasis related molecular alterations at the primary site will inform recurrence risk assessments and CRC disease management options (i.e., adjuvant chemotherapy). Findings will elucidate how spatially dependent lymphocyte expression patterns inform a coordinated immune response to nodal metastasis. We expect research findings to fuel a clinical trial, funded through an R01 grant mechanism (e.g., PAR-22-131), which will compare the adjunct test prospectively to independent indicators of metastasis and recurrence risk: 1) CDX2 expression, 2) circulating tumor DNA (ctDNA) markers, and 3) immunoscore[10–12]. As compared to these technologies, we expect the adjunct test to balance tradeoffs in efficiency, cost, access, and comprehensiveness.

Research Strategy

Significance

Burden of Colon Cancer and Assessment Challenges. Colorectal cancer (CRC) is the third leading cause of cancer both worldwide and in the United States and accounts for approximately 8 percent of cancer-related deaths[13]. CRC incidence is shifting towards younger demographics who are not included in established screening programs[14,15]. While modifying specific risk factors (e.g., epigenetics, environment, diet)[16–18] can be effective in informing and/or curbing CRC incidence, there is a concurrent and vital need to develop more accurate, faster, and lower cost solutions for CRC screening and prognostication– and this is the goal of the work detailed in this proposal. Disease management of CRC often includes lymph node resection to determine N-stage as a proxy for recurrence risk, after resection at the primary site at the time of diagnosis. This is typically followed by adjuvant therapy for patients with positive lymph nodes. Assessing tumor metastasis is difficult because not all metastatic lymph nodes can be surgically removed, potentially leading to inaccurate assessments of recurrence risk. Lymph node evaluation outside of specialized centers can be inadequate, as shown by a study where only 37% of colon cancer cases had proper evaluation of regional lymph nodes [19–22].. Retrieving a high number of lymph nodes (>20) has been linked to better prognosis but also higher morbidity. Inadequate resection and analysis can affect prognostication and treatment selection, highlighting the need for studying molecular alterations at the primary site to indirectly infer metastasis and recurrence risk beyond pathological staging.

Role of Tumor Infiltrating Lymphocytes on Prognostication in the Context of Existing Digital Pathology Assays. Tumor Infiltrating Lymphocytes (TIL) are crucial in characterizing and modulating the Tumor Microenvironment (TME) and Tumor Immune Microenvironment (TIME) for prognosis and immunotherapy[23]. The TME is a complex system of malignant and benign cells, blood vessels, and extracellular matrix. The spatial distribution, density, and relationships of T cell, B cell, NK cell, and other immune infiltrates play a significant role in the antitumoral response, modified by Microsatellite Instability (MSI) status[23–25]. Several digital pathology assays complement pTNM staging and provide independent risk factors for recurrence. Assays for assessing tumor recurrence risk include: 1) Immunoscore, which measures cytotoxic T-cell density within and around the tumor, 2) CDX2, a pluripotent signature indicating tumor's ability to bypass immune response, and 3) circulating tumor DNA (ctDNA) for mutations across several molecular pathways (e.g., vascular endothelial growth factor pathway). Highly multiplexed assays capturing multiple aspects of metastasis simultaneously could facilitate the formulation of additional tests.

Spatial Localization of Immune Signatures. The development of spatial omics technologies such as 10x Genomics Spatial Transcriptomics (ST) or GeoMX Digital Spatial Profiling (DSP) has enabled multiplexing findings (e.g., whole transcriptome, WTA) at incredible spatial resolution[26]. Existing applications of spatial profiling include comparisons of TIL subpopulations across the TME and automated scoring systems that have been developed which infer TIL information from standard morphology stains as a digital biomarker (*virtual stain*)[27–30], with limited translational application[31]. Despite the breadth of literature, few studies thus far have attempted to connect comprehensive, high-multiplexing of TILs offered by spatial omics technologies with the capacity to predict lymph node involvement and recurrence risk. Current assays are high cost, low throughput and have limited reproducibility, which limits their application as a clinical test. However, inferring spatial molecular information from tissue morphology can inform assessments at low-cost and high-throughput [32,33].

Innovation. This proposal improves disease prognostication with a low-cost and high throughput virtual staining spatial molecular assessment of tumor metastasis, complementing inadequate lymph node resection evaluations. Innovations include: 1) investigating spatial cell-type specific proteomic, transcriptomic changes associated with metastasis, 2) a low-cost validation framework for molecular associations, and 3) innovative statistical and machine learning methods. These improvements provide: 1) higher multiplexing, 2) accurate cell-typing, and 3) address batch-level variation. The impact is far-reaching: 1) serving as an adjunct assessment to lymph node resection, 2) providing additional information for optimal therapeutics, and 3) reducing cost barriers to local and distant metastasis assessment for global health applications. The goal is to infer the degree of lymph node involvement or recurrence risk in situations where lymph node assessment is biased/inadequate, with similar power and lower cost than highly multiplexed molecular profiling [34] and other prognostic indicators[10–12,35].

Impact of the COBRE Center on Long-Term Career Ambitions. With my prior training and experiences, I am well-positioned to develop cutting-edge technologies that are user-centered, trusted by clinicians, and have a high potential for adoption. My design philosophy is centered around fully incorporating the feelings and ideas of the end user, specifically the practicing clinician. However, I noticed that many AI technologies developed for

pathology are disconnected from their core users. I am determined to bridge this gap and create technologies that meet the needs of those who use them most. In my meetings with senior leadership, we quickly discovered a shared interest in building research infrastructure to facilitate AI research within the department. This collaboration ultimately led to my faculty hiring at Dartmouth Health (DH). As an Assistant Professor in Pathology and Dermatology, I am eager to further explore the potential of digital pathology technologies with the help of essential mentorship and resources provided by COBRE project leader support (**see Plan to Leverage CQB Resources**). Since starting my research lab, I have mentored dozens of students from the undergraduate to MD-PhD level, structured into 5 project teams, and dozens more through a national high school internship program I started four years ago. Many of these students are passionate about developing research in the context of this proposal and have submitted works ultimately accepted for publication across top-refereed journals and conference venues. COBRE funding would allow me to attract top student talent and provide the necessary mentorship to structure my lab, establish national collaborations, and receive much needed administrative support. Mutually synergistic collaborations with COBRE cores highlighted throughout the **Approach** will facilitate the collection of preliminary data necessary to carry-out an R01-funded clinical trial which envisions how this adjunct test can impact clinical decision making through a healthcare delivery science framework. With COBRE support, I am excited to accelerate the development of my research program and make meaningful contributions to the field of digital pathology and fully realize the aspirations of CQB COBRE center to develop and apply innovative single-cell and spatial technologies.

Approach. Overview. The aims of this proposal represent distinct assessment methods: Aims 1 and 2 will **identify** metastasis-related biomarkers, while Aim 3 will **validate** and **infer** these biomarkers; Aim 1) spatial molecular profiling via lymphocyte-specific proteomic profiling, Aim 2) spatial whole transcriptomic profiling and single cell RNAseq analysis, and Aim 3) establishment of a low-cost adjuvant assessment method through a) independent validation with immunofluorescence (IF) and b) virtual staining with tissue morphology (H&E). Aims are broken down into subsections which detail: 1) tissue molecular profiling and imaging (**Profile**), 2) data analysis (**Assess**), 3) expected key findings (**Deliverable**), and 4) preliminary data / progress from COBRE Phase I funding, statistical power, limitations and minimum success criteria (**Feasibility**). Subsections will reference data collected from specific tissue sections (**Data Collection**).

Data Collection. IRB approval has been obtained for this study. In collaboration with the Pathology Shared Resource (PSR), we will access 150 tissues from stage pT3 Colon adenocarcinoma patients biobanked in Pathology. Cases were identified through a retrospective search of pathology reports and slides at Dartmouth Hitchcock Medical Center (DHMC), a *Center of Excellence* with high quality nodal data, from 2016 to 2022. Stage was determined using the pTNM staging system, which balances local invasion versus metastasis as an overall prognostic marker. The cohort will be restricted by pathological T-stage III (i.e., pT3) using the pTNM staging system to control for local invasiveness. This will identify biomarkers that provide prognostic value beyond the current prognostic staging system. Demographic information and tissue characteristics will be collected for each specimen as detailed in **Table 1**. Tissue resections from the patient will be partitioned into tissue blocks that will be sectioned into layers with 5-20 micron depth each (**Figure 1;** sections described in **Table 1**). DSP slides will be IF stained for the following markers: SYTO (nuclei), CD45 (immune cell), and PanCK (epithelial/tumor) for targeted profiling of lymphocytes. A GI pathologist will view H&E and IF whole slide images (WSI) simultaneously to annotate sections and label any spatially resolved information by three distinct macroarchitectural regions: intratumoral (*intra*), invasive

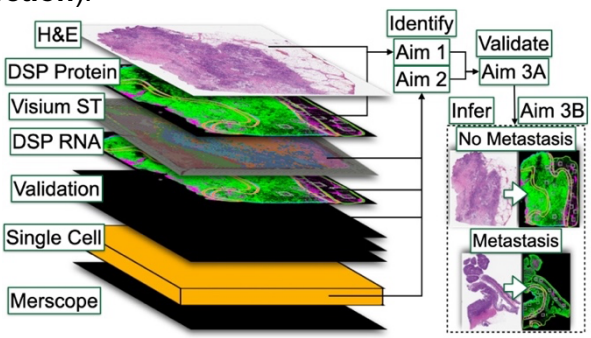


Figure 1: Overview of specimen sectioning

Table 1: Tabular description of demographics, tissue characteristics and assayed tissue sections

Demographics & Tissue Characteristics		Block Sections (5 um thick)	Assay	Associated Aim
Age	Invasion (T-Stage)	Layer 1 (n=150)	H&E	Aim 1
Sex	Nodal Metastasis (N-Stage)	Layer 2 (n=64)	Proteomics DSP, IF, Co-registered H&E	Aim 1
Tumor Grade	Distant Metastasis (M-Stage)	Layer 3 (n=32)	Visium ST, Co-registered H&E	Aim 2
Tumor Site (e.g., cecum, rectum)	Tissue Slides (Lymph Nodes)	Layer 4 (n=16)	RNA DSP, IF, Co-registered H&E	Aim 2
Gross Tissue Dimensions	MMR Gene Loss Expression (MLH1, PMS2, MSH2, MSH6), using IHC	Layer 5 (n=100)	Validation IF Stains, Co-registered H&E	Aim 3
		Layer 6 (n=16)	Single Cell (50-um thick)	Aim 2
Macroarchictectural Annotations	Time of Recurrence or Right Censoring	Layer 7 (n ≤ 16)	Vizgen Merscope	Aim 2, if needed/costs

margin (*inter*) and away from tumor (*away*). For all aims, patient age/sex will be modeled as a potential confounder / effect modifier through multivariable regression.

Aim 1: Establish lymphocyte-specific spatial immuno-oncology protein markers of metastasis.

1-Profile: Proteomic Digital Spatial Profiling (DSP) of lymphocytes within distinct macroarchitectural contexts. We will profile 32 patients without metastasis and 32 patients with nodal metastasis (half of these patients with both nodal and distant metastasis) (n=64 patients). For each patient, 24 regions of interest (ROI) will be placed among the 3 macroarchitectural regions (1,536 total ROIs/nested observations)[36]. Immune cells are isolated within each ROI via image segmentation of the CD45 stain and profiled through targeted ultraviolet (UV) cleavage of attached oligo tags. The Nanostring nCounter will quantify immune cell protein expression across 40 immuno-oncology markers. Returned data will include protein expression measurements for each ROI, tagged with x,y coordinate, an ROI-specific nuclei count, and co-registered H&E and IF slides from same section. ROIs will be filtered based expression relative to negative control, normalized, and log2 transformed.

1-Assess. Develop methods to identify differentially expressed clinical markers of nodal/distant metastasis in the TME, accounting for batch and spatial autocorrelation. Identify statistical interactions between markers/patient characteristics using machine learning, verified by fitting machine learning-derived interactions in traditional statistical models. Identify groups of metastasis markers with gene modules. Bayesian hierarchical linear regression models will establish associations with metastasis (nodal and/or distant metastasis). This will identify lymph node and distant metastasis related markers based on their differential expression within specific macroarchitecture (*intra*, *inter*, *away*), adjusting for potential confounding (deficient mismatch repair, assessed through IHC– dMMR, age, sex) (**Figure 2A,B,D,E**)[37,38]. Batch and case-level variation are captured with random intercepts and an exponential decay kernel. Effect estimates will be derived from the posterior distribution with *post hoc* comparisons via *emmeans*[39,40].

When applied to spatial omics, the assumption of independence of nested observations in most ML methods for omics data can lead to distorted study findings. To address this issue, we plan to develop a classifier that can estimate the probability of tumor metastasis based on all markers (x_i). This will be accomplished by fitting tree boosting models, $f_\phi(\vec{x_i})$, in a Mixed Effects Machine Learning modeling framework (MEML), such as Gaussian Process Boosting and hierarchical Bayesian Additive Regression Tree (BART)[41,42]. These methods will account for patient-specific/spatial autocorrelation by incorporating random intercepts and a Gaussian process. Significant interactions identified from the MEML method will be included as interaction terms in a Bayesian hierarchical logistic regression model to report relevant effect modifiers (e.g., the effect of CD20 conditional on age). (**Figure 2C**)[36,43]. Performance will be evaluated using a C-statistic, the Bayes Factor and WAIC. Feature selection will incorporate the Horseshoe LASSO and projection predictive selection methods[44,45]. A weighted gene co-expression network analysis (WGCNA) and CDS (conserved, differential, specific) differential co-expression analysis will cluster genes into modules for association with disease metastasis[46].

1-Deliverable: Identify stains for independent validation. As this is a proteomic assessment, **1-Assess** findings will motivate lower-cost staining (*Aim 3A*) to recapitulate findings. Effect modifiers revealed in **1-Assess** will suggest stains (or set of stains) specific to age, sex or pathway (e.g., MLH1-loss).

1-Feasibility: Preliminary data collected on 32 patients informs study power, reveals metastasis-related markers and effect modifiers, but more data is required to achieve significant findings. Potential limitations include batch effects, biased ROI selection, and potential for Type I error. To date, 840 ROI have been collected, leading to an initial publication of MEML effect modifiers (**Figure 2**) and a follow up peer-reviewed publication identifying clinical markers inside the tumor, at the invasive margin, and away, associated with extracellular matrix remodeling (e.g., GZMB), immune suppression (e.g., FOXP3), T-cell exhaustion (e.g., CD8), etc.; initial results can be explored at https://levylab.shinyapps.io/ViewColonDSPResults [36,47]. An empirical power analysis, motivated by initial published findings, was conducted based on 100 simulated datasets (binary outcomes drawn via $y_i \sim Binomial(1, p_i)$) using the statistical model discussed in **1-Assess** at a sample

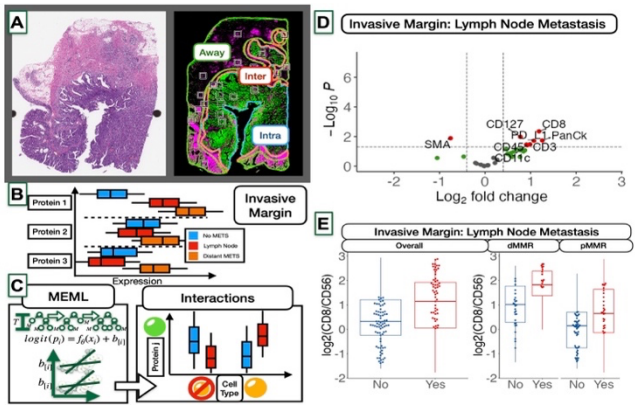


Figure 2: Aim 1: A) H&E/IF slide and macroarchitectures of interest; B) Differential expression in architectures for proteins; C) Identify effect modifiers via MEML (e.g., CD20 and age); D) Example volcano plot, preliminary findings; E) Box plots, preliminary findings

size of 64, given batch effects and estimated effect size of 0.2, yielded a power of 0.81 after multiplicity adjustment ($\alpha = 0.05$, Bonferroni adjustment). Sensible weakly informative priors centered around 0 will guard against Type I error in addition to multiplicity adjustments[48]. **Minimum success criteria–** Identify at least three proteins or effect modifiers for validation.

Aim 2: Spatial cell-type transcriptomic signatures.

2-Profile: Comprehensive transcriptomic profiling using the DSP RNA assay and Visium Spatial Transcriptomics (ST) platform. Up to 18,000 genes can be profiled using ST to complement proteomics. ST offers a greater chance to elucidate spatial variation due to the comprehensive coverage of barcoded spots. Thirty-two Visium ST slides (16 patients with nodal/distant metastasis), subset from the proteomics cohort, will be assayed. Capture areas (**Figure 3A**) will be placed by a pathologist to ensure adequate overlap with previous tissue macroarchitectural annotations (*intra/inter/away*). Capture areas will be manually cut by a histotechnician from the PSR and sent to the SCGC for profiling. ST uses spatially barcoded spots to capture polyadenylated mRNA molecules and register their coordinates within the capture area. The CQB COBRE has adapted several innovative protocols to streamline and standardize sample preparation, including housing of the innovative CytAssist technology[49,50], which facilitates the transfer of spatial transcriptomic probes directly from tissue slides, allowing for auto-staining and image scanning in Pathology, improved assay chemistry, and the ability to operate on archival FFPE blocks.

We plan to collect 16 10x single cell RNASeq samples to improve the resolution and deconvolve Visium spot data and at most 12 Vizgen Merscope samples to validate ST findings on the single cell level.

2-Assess: Apply methods from 1-Assess. In addition– 1) identify ST genes that exhibit spatial variation related to metastasis, 2) leverage 10x single cell data to infer spatial signaling patterns between cells, adjust for and estimation effect modification by cell-types through deconvolution, 3) identify metastasis-related pathways. Concordance between DSP (in situ hybridization probes and NGS readout) and ST at representative spots for 16 patients (8 metastasis patients) will establish internal validity of research findings[51] using a negative binomial likelihood, and compare statistical precision of effect estimates.

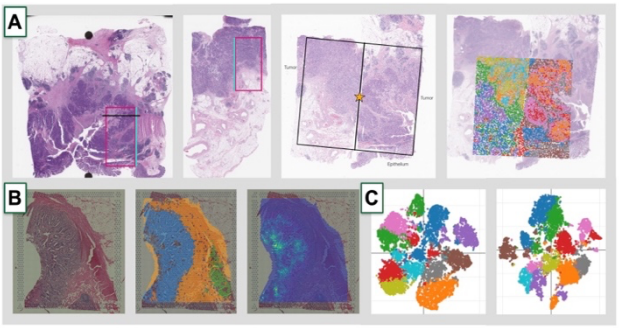


Figure 3: Aim 2: A) CytAssist workflow between PSR and SCGC to profile two samples simultaneously; B) Example spatial transcriptomics analysis; C) Collected two FFPE single cell

Spatial variation. Statistical and ML methods (e.g., SPARK, Moran's I, cluster analysis, gaussian process) will identify spatially variable genes within and across slides by incorporating information from adjacent spots (i.e., spatial autocorrelation)[52–56]. Indices of spatial clustering will be compared to identify which spatially variable genes exhibit different clustering patterns between patients by metastasis status and/or macroarchitecture.

Single cell integration. In collaboration with the CQB COBRE Data Analytics Core (DAC), expression profiles for spots within sections will also be deconvolved into their constituent cell type proportions (e.g., endothelial cells, CD4T, etc.; $\vec{p}$) through established deconvolution approaches, some of which leverage single cell information mapped to spatial locations[53,57–59]. Methods which integrate information on ligand-receptor and cytokine expression with co-localized expression patterns will inform which cell-types are likely to interact [60,61]. Effect estimates from **2-Assess** will be compared to adjustment for and effect modification by cell-type. *Pathway analysis.* Significant genes from **2-Assess** will undergo gene set enrichment (GSEA) and pathway analyses on well-known pathway databases (e.g., KEGG)[62].

2-Deliverable: Similar to 1-Deliverable, including spatially variable/interacting genes/cells, and pathways will inform knowledge on TIL cell-type specific molecular changes.

2-Feasibility: Similar to 1-Feasibility. Preliminary data collected in collaboration with PSR and SCGC. We have profiled thirteen ST slides, according to **2-Profile**, through a pilot initiative among the Levy Lab, PSR, DAC and SCGC. We successfully tested single cell and ST FFPE protocols and compared unsupervised gene expression clusters across slides for both single cell and ST (**Figure 3B,C**). Our team has collaborated with the Pathology Shared Resource (PSR) and SCGC to become one of the first groups in the country to trial the CytAssist protocols and the 10x single cell FFPE assay. This has given us a clearer understanding of how to address data quality issues and batch effects while increasing the number of samples per capture area through careful dissection and placement of tissue to maximize the study power while reducing costs. The collection of additional single cell data, both in-house and through public repositories will supplement the sample size which

is limited due to fiscal constraints. Single cell data will be supplemented using publicly available Colon data collected from *The Human Cell Atlas (HCA)* and *Single Cell Portal* projects, which provides access to scRNASeq expression for over 40 million cells across dozens of tissue types[60,63,64]. The DAC will also help improve data quality (i.e., expand the number of genes) through mapping and integration of this single cell data with the Visium spots [54,59,65–68]. Nuclei detection machine learning techniques (developed for other projects) applied to corresponding histological slides will improve cell type localization[9,69–73]. To lend additional credence to our findings, reproducibility of ST will also be assessed via Vizgen Merscope assay by selecting the *top-600* markers from the differential expression and spatial variation analyses (ranked by standardized effect estimates) and performing similar assessments on the single cell level. Minimum success criteria includes identifying at least 100 genes which are differentially expressed within one or more spatial architectures. Based on a simulation through models specified in **2-Assess**, with effect estimates derived from our thirteen pilot slides, we will be able to detect these spatial effects in the Visium data, a sample size of 32 with a power of at least 0.94 after Bonferroni adjustment. Our team is well-positioned to execute on *Aim 2* pending cohort expansion to capture patient-level variation. Multiplicity issues can be ameliorated through gene set weighting and Bayesian methods[74–76].

Aim 3: Developing and Scaling Low-Cost Adjuvant Assessments as Clinical Trial Precursor

Aim 3a: Independent Validation of Stains Predictive of Metastasis

3a-Profile: Perform fluorescence imaging of top 10 mRNA/protein markers from 1-Deliverables. IF stains will be selected based on effect size, statistical significance, and perceived pathway relevance from TME literature. Two 6-plex IF stains (one stain for nuclei) will be imaged using the PerkinElmer Vectra3 (bioMT MSR) for 100 cases (TMAs if needed). Cells will be separated through image segmentation of the nuclear stain and tagged by patient ID, macroarchitecture, slide coordinates (x,y) and a 10-dimensional vector representing protein expression represented by 16-bit precision image intensity, thresholded to indicate positive staining.

3a-Assess: Recapitulate 1-Assess findings and demonstrate prognostic potential of low-cost stains through metastasis prediction. Differential expression will be compared to ordinal stain scores assigned by collaborating pathologists. These findings will be used to inform markers for recurrence in a subsequent grant submission given appropriate follow up time[34].

3a-Deliverable: Similar to 1-Deliverable, with greater expected statistical power than *Aim 1*.

3a-Feasibility: Existing IF stains collected with DSP profiling can trial 3a-Assess. RNA may be more difficult to validate in IF. For example, presence of PanCK at the tumor invasive margin quantified using the DSP was associated (**1-Feasibility**; not pictured) with metastasis. PanCK target identified with DSP has matched IF stain for comparison. Validation of selected mRNA targets will depend on degree to which RNA translates to protein. **Statistical power** is expected to be greater than that in *Aim 1* and *2;* we used the same statistical model as aim 2 with an expected sample size of 150, yielding a power of 0.98 (**3a-Deliverable**). **Minimum Success Criteria–** Six of ten stains confirm equivalent/added prognostic value.

Aim 3b: *Virtual Staining* of Validation Stains and ST for Low-Cost Cellular Molecular Assessment

3b-Profile: Co-register same-layer morphology (H&E) and molecular (IF, ST spots; 1,3A-Profile) information. Set aside 150 serial section H&E WSI for validation (**Table 1**).

3b-Assess: Develop two Virtual Staining tools, *VirtualProtein* and *VirtualRNA*, to infer spatial molecular information from H&E, and assess inferred spatial molecular information similar to 3a-Assess. Localize nuclei for Virtual Staining at the cellular resolution. ***VirtualProtein* recapitulates spatial protein expression from H&E*. *VirtualRNA* recapitulates spatial mRNA expression from H&E at cellular resolution.** *VirtualProtein* will be trained using a generative adversarial network (GAN) which takes as input the H&E WSI and outputs the pixel-wise presence and intensity of the multiplexed stains from **3A-Profile** (**Figure 4A**). Since GANs cannot operate on the entire WSI, the GAN will be modified to incorporate macroarchitectural information with Graph Neural Networks (GNN) for visual/predictive consistency [7–9,77–79].

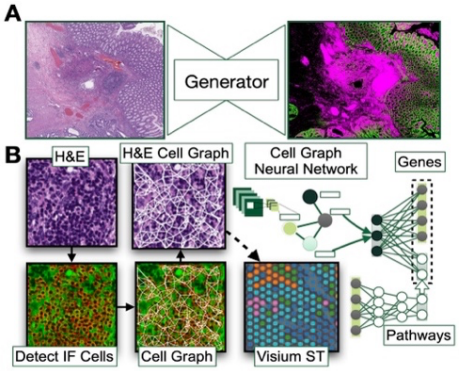


Figure 4: Aim 3– Virtual: A) Protein; B) RNA

We will use co-registered *IF* and *H&E* stains of nuclei and immune cells to build a highly accurate immune cell prediction tool by training an object detection neural network using the Detectron2 framework[72,80]. Graphs will be constructed over cells, identified by applying the detection network on **2-Profile** H&E images, to form *cell graphs* based on spatial proximity. A convolutional neural network (CNN) will extract cell-level morphological information and tissue context, stored as node attribute vectors [7,81–83]. *VirtualRNA* will train GNN to pool

information from *cell graphs* to predict the mRNA expression patterns at ST spots (**Figure 4B**). Graph/image-based interpretation methods will identify cells and morphology related to gene expression. As inference on 18,000 genes is intractable, we will predict gene expression aggregated across pathways using a *sparse pathway neural network*[84–88].

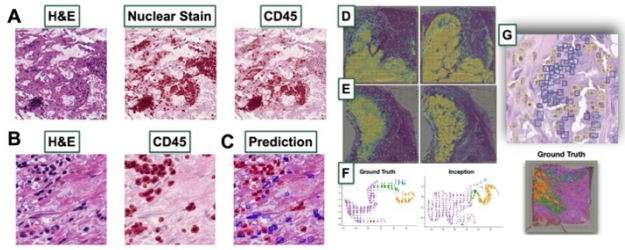


Figure 5: Aim 3 preliminary findings: A) Tissue restaining with CD45; B-C) VirtualProtein inference; D-E) VirtualRNA inference; F) Predicted expression clusters similar to ground truth; G) Unpublished RNA inference results, cell-graph neural networks

3b-Deliverable: Similar to 3a-Deliverable. In addition, develop an image viewer oncologists can use to interface with *Virtual Stain* results[89]. We anticipate *Virtual Stains* can be ordered at even lower cost compared to IF stains from *3A*. We aim to identify proteomic and mRNA markers which can be predicted from morphology and used to prognosticate at similar power as **3a-Deliverable**. We expect effect estimates to be more precise than Aims *1*-2, while circumventing costs, labor and turnaround time associated with staining.

3b-Feasibility: Similar to 3a-Feasibility. The PI has previously demonstrated the feasibility of *Virtual Staining*, GNN, and sparse pathway neural networks for digital pathology through published research [7–9,80]**.** An initial proteomics inference feasibility study was published demonstrating the ability to localize immune cells through prediction of a co-registered CD45 stain with an AUC of 0.94 (**Figure 5A,B**) [70]. An initial RNA inference feasibility study was published demonstrating an AUC of 0.8 for localization of 1000 RNA markers which were associated with epithelial to mesenchymal phenotype (**Figure 5C,D**) [90]. We compared CytAssist, pathology autostainers and 40x scanning versus the traditional 10x Visium workflow and found a 10-15% increase in predictive accuracy (AUC=0.85) using our specimen processing method (*in preparation*). Given the existing predictive performance, a simulation similar to Aim 2 on a cohort size of 150 yielded a power of 0.96.

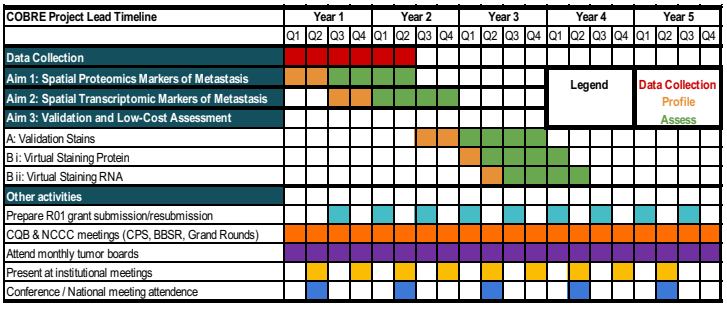
Collaborators/Mentors. Translational research team (e.g., oncology experts) described in budget justification.

Plan to Leverage CQB Resources. PI consulted with SCGC and DAC faculty to identify ways to interface with CQB cores. Grant funds will support large-scale Visium and single cell profiling, RNA quality control, and data analysis assistance. The discussions between the PI and CQB Core directors can be summarized as follows: **SCGC** (Fred Kolling, PI)– 1) incorporate newly acquired technologies (e.g., merScope, HD Visium ST) and scale FFPE ST/single-cell protocol, 2) receive up to 1/3 discount on large scale Visium ST, 3) collaborate on methods to ameliorate batch effects, 4) form spatial omics interest group, 5) specimen triage; **DAC** (James O'Malley, PI)– 1) guidance on statistical methodology for grant aims, 2) development of spatially-aware Bayesian Additive Regression Trees (BART), 3) benchmark ST against DSP, 4) software developed in COBRE project will be made publicly available through DAC HTML interface (Tim Sullivan), and 5) DAC (Owen Wilkins) will identify/leverage public single cell RNASeq data (Colon) to expand/impute ST markers[52,60,91–94]; **Mentoring/Administrative Core:** support for clinical mentor Louis Vaickus and statistical mentor Jiang Gui; R01 study planning, feedback/advice.

Publication Potential: We aim to publish 10-15 papers on project findings. PI has a strong publication record with over 50 publications in the past 4 years, majority as the first or senior author.

Presentation and Tumor Boards Enhance Collaboration, Understanding of Implementation Challenges: The PI will solicit feedback through attendance/presentations at: 1) Oncology and GI Pathology tumor boards, 2) COBRE CQB meetings, 3) Biostatistics and Bioinformatics Shared Resource (BBSR), 4) Cancer Population Sciences (CPS), 5) Cancer Center and Pathology grand rounds, and 6) national bioinformatics conferences.

R01 Clinical Trial: The data from the COBRE project will be used to formulate the adjunct test, with plans to validate the test through R33 IMAT mechanism (RFA-CA-23-003), followed by multi-center R01 validation trial will compare the adjunct test established through the COBRE project to other independent recurrence predictors (**Significance**)– R01 submissions twice a year, with the following aims: 1) formulation/user-testing of recurrence risk assessment with interpretation through a web application, 2) clinical trial assessing the 3-year risk of recurrence, and 3) study the impact and barriers to adoption of the test using an implementation science framework. The PI plans to work with the Mentoring and Administrative COBRE Core and GrantsGPS for R01 preplanning.



References

1. 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

2. 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

3. Idos GE, Kwok J, Bonthala N, Kysh L, Gruber SB, Qu C. The Prognostic Implications of Tumor Infiltrating Lymphocytes in Colorectal Cancer: A Systematic Review and Meta-Analysis. Scientific Reports. 2020 Feb 25;10(1):3360. PMID: 32099066

4. Bear HD, Chin CS. 35 - Approaches to Adoptive Immunotherapy. In: Souba WW, Wilmore DW, editors. Surgical Research [Internet]. San Diego: Academic Press; 2001 [cited 2020 Nov 21]. p. 415–434. Available from: http://www.sciencedirect.com/science/article/pii/B978012655330750037X PMID: 20223744

5. Tseng D, Schultz L, Pardoll D, Mackall C. 6 - Cancer Immunology. In: Niederhuber JE, Armitage JO, Kastan MB, Doroshow JH, Tepper JE, editors. Abeloff's Clinical Oncology (Sixth Edition) [Internet]. Philadelphia: Elsevier; 2020 [cited 2020 Nov 21]. p. 84-96.e5. Available from: http://www.sciencedirect.com/science/article/pii/B9780323476744000062 PMID: 33674323

6. Zhang M, Sheffield T, Zhan X, Li Q, Yang DM, Wang Y, Wang S, Xie Y, Wang T, Xiao G. Spatial molecular profiling: platforms, applications and analysis tools. Brief Bioinform [Internet]. [cited 2020 Nov 21]; Available from: https://academic.oup.com/bib/advance-article/doi/10.1093/bib/bbaa145/5881377

7. 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;285–296. PMID: 33691025

8. Levy JJ, Azizgolshani 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

9. 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.

10. 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

11. 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

12. Tarazona N, Gimeno-Valiente F, Gambardella V, Huerta M, Roselló S, Zuniga S, Calon A, Carbonell-Asins JA, Fontana E, Martinez-Ciarpaglini C, Eason K, Rentero-Garrido P, Fleitas T, Papaccio F, Moro-Valdezate D, Nyamundanda G, Castillo J, Espí A, Sadanandam A, Roda D, Cervantes A. Detection of

postoperative plasma circulating tumour DNA and lack of CDX2 expression as markers of recurrence in patients with localised colon cancer. ESMO Open. 2020 Sep;5(5):e000847. PMCID: PMC7513635

13. 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

14. 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

15. Patel SG, Ahnen DJ. Colorectal cancer in the young. Current gastroenterology reports. 2018;20(4):1–12. PMID: 34414781

16. Lao VV, Grady WM. Epigenetics and colorectal cancer. Nature reviews Gastroenterology & hepatology. 2011;8(12):686–700. PMID: 35945569

17. 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

18. Slattery ML. Diet, lifestyle, and colon cancer. Semin Gastrointest Dis. 2000 Jul 1;11(3):142–146. PMID: 35643370

19. 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

20. 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

21. 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

22. Ong ML, Schofield JB. Assessment of lymph node involvement in colorectal cancer. World journal of gastrointestinal surgery. 2016;8(3):179. PMID: 36678187

23. 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

24. 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

25. 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.

26. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021 Jan;18(1):9–14. PMID: 33408395

27. Nalisnik M, Amgad M, Lee S, Halani SH, Velazquez Vega JE, Brat DJ, Gutman DA, Cooper LAD. Interactive phenotyping of large-scale histology imaging data with HistomicsML. Sci Rep. 2017 Nov 6;7(1):14588. PMID: 29109450

28. SAHA M, GUO X, SHARMA A. TilGAN: GAN for Facilitating Tumor-Infiltrating Lymphocyte Pathology Image Synthesis With Improved Image Classification. IEEE Access. 2021;9:79829–79840. PMCID: PMC8224465

29. He B, Bergenstråhle L, Stenbeck L, Abid A, Andersson A, Borg Å, Maaskola J, Lundeberg J, Zou J. Integrating spatial gene expression and breast tumour morphology via deep learning. Nature Biomedical Engineering. 2020 Jun 22;1–8. PMID: 32572199

30. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596(7871):211–220. PMID: 34381231

31. Liu H, Zhao Y, Yang F, Lou X, Wu F, Li H, Xing X, Peng T, Menze B, Huang J, Zhang S, Han A, Yao J, Fan X. Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer with Deep Learning. BME Frontiers [Internet]. 2022 Mar 17 [cited 2022 May 4];2022. Available from: https://spj.sciencemag.org/journals/bmef/2022/9860179/ PMID: 34968471

32. 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

33. 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

34. Uttam S, Stern AM, Sevinsky CJ, Furman S, Pullara F, Spagnolo D, Nguyen L, Gough A, Ginty F, Lansing Taylor D, Chakra Chennubhotla S. Spatial domain analysis predicts risk of colorectal cancer recurrence and infers associated tumor microenvironment networks. Nature Communications. 2020 Jul 14;11(1):3515. PMID: 32665557

35. Reinert T, Henriksen TV, Christensen E, Sharma S, Salari R, Sethi H, Knudsen M, Nordentoft I, Wu HT, Tin AS, Heilskov Rasmussen M, Vang S, Shchegrova S, Frydendahl Boll Johansen A, Srinivasan R, Assaf Z, Balcioglu M, Olson A, Dashner S, Hafez D, Navarro S, Goel S, Rabinowitz M, Billings P, Sigurjonsson S, Dyrskjøt L, Swenerton R, Aleshin A, Laurberg S, Husted Madsen A, Kannerup AS, Stribolt K, Palmelund Krag S, Iversen LH, Gotschalck Sunesen K, Lin CHJ, Zimmermann BG, Lindbjerg Andersen C. Analysis of Plasma Cell-Free DNA by Ultradeep Sequencing in Patients With Stages I to III Colorectal Cancer. JAMA Oncology. 2019 Aug 1;5(8):1124–1131.

36. 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

37. Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software. 2017 Aug 29;80(1):1–28.

38. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A. Stan: A Probabilistic Programming Language [Internet]. Grantee Submission. 2017 [cited 2021 Apr 28]. p. 1–32. Available from: https://eric.ed.gov/?id=ED590311

39. Makowski D, Ben-Shachar MS, Chen SHA, Lüdecke D. Indices of Effect Existence and Significance in the Bayesian Framework. Front Psychol. 2019 Dec 10;10:2767. PMID: 31920819

40. Searle SR, Speed FM, Milliken GA. Population marginal means in the linear model: an alternative to least squares means. The American Statistician. Taylor & Francis; 1980;34(4):216–221.

41. Sigrist F. Latent Gaussian Model Boosting. arXiv:210508966 [cs, stat] [Internet]. 2021 May 21 [cited 2021 Jul 28]; Available from: http://arxiv.org/abs/2105.08966 PMID: 35439126

42. Tan YV, Roy J. Bayesian additive regression trees and the General BART model. Statistics in Medicine. 2019;38(25):5048–5069. PMID: 31460678

43. Levy JJ, O'Malley AJ. Don't dismiss logistic regression: the case for sensible extraction of interactions in the era of machine learning. BMC Med Res Methodol. 2020 Jun 29;20(1):171. PMCID: PMC7325087

44. Carvalho CM, Polson NG, Scott JG. Handling Sparsity via the Horseshoe. Artificial Intelligence and Statistics [Internet]. PMLR; 2009 [cited 2021 Jul 28]. p. 73–80. Available from: http://proceedings.mlr.press/v5/carvalho09a.html

45. 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.

46. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9(1):559.

47. 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 [Internet]. medRxiv; 2022 [cited 2023 Mar 8]. p. 2022.12.04.22283073. Available from: https://www.medrxiv.org/content/10.1101/2022.12.04.22283073v1

48. Gelman A, Tuerlinckx F. Type S error rates for classical and Bayesian single and multiple comparison procedures. Computational Statistics. 2000 Sep 1;15(3):373–390.

49. Singh H, Sukovich D, Mohabbat S, Li D, Kim H, Tentori A. Visium CytAssist: A Novel Platform for Spatial Transcriptomic Analysis of FFPE Sections Mounted on Standard Glass Slides. MOLECULAR THERAPY. CELL PRESS 50 HAMPSHIRE ST, FLOOR 5, CAMBRIDGE, MA 02139 USA; 2022. p. 589–590.

50. Hernandez S, Lazcano R, Serrano A, Powell S, Kostousov L, Mehta J, Khan K, Lu W, Solis LM. Challenges and Opportunities for Immunoprofiling Using a Spatial High-Plex Technology: The NanoString GeoMx® Digital Spatial Profiler. Frontiers in Oncology. 2022;2946. PMID: 35847846

51. Doreian P. Network autocorrelation models: Problems and prospects. Spatial statistics: Past, present, future. Michigan Document Services Ann Arbor; 1989;369–89.

52. Moses L, Pachter L. Museum of spatial transcriptomics. Nat Methods. 2022 Mar 10;1–13. PMID: 35273392

53. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, Li M. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021 Nov;18(11):1342–1351. PMID: 34711970

54. Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uytingco CR, Taylor SEB, Nghiem P, Bielas JH, Gottardo R. Spatial transcriptomics at subspot resolution with BayesSpace. Nat Biotechnol. 2021 Nov;39(11):1375–1384. PMID: 34083791

55. Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. Nature Methods. Nature Publishing Group; 2020 Feb;17(2):193–200.

56. 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

57. Song Q, Su J. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. Briefings in Bioinformatics [Internet]. 2021 Jan 22 [cited 2021 May 15];(bbaa414). Available from: https://doi.org/10.1093/bib/bbaa414 PMID: 33480403

58. Dong R, Yuan GC. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. Genome Biology. 2021 May 10;22(1):145.

59. Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, Beechem JM. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. Nat Commun. 2022 Jan 19;13(1):385. PMID: 35046414

60. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nature Communications. 2020 Apr 29;11(1):2084. PMID: 32350282

61. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. Nat Rev Genet. 2021 Feb;22(2):71–88. PMID: 33168968

62. 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.

63. Osumi-Sutherland D, Xu C, Keays M, Levine AP, Kharchenko PV, Regev A, Lein E, Teichmann SA. Cell type ontologies of the Human Cell Atlas. Nat Cell Biol. 2021 Nov;23(11):1129–1135. PMID: 34750578

64. Walker S, Abugessaisa I, Kasukawa T. Single Cell Databases: An Emerging and Essential Tool. In: Abugessaisa I, Kasukawa T, editors. Practical Guide to Life Science Databases [Internet]. Singapore: Springer Nature; 2021 [cited 2022 Oct 30]. p. 171–190. Available from: https://doi.org/10.1007/978-981-16-5812-9_9

65. 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

66. Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, Elmentaite R, Lomakin A, Kedlian V, Gayoso A, Jain MS, Park JS, Ramona L, Tuck E, Arutyunyan A, Vento-Tormo R, Gerstung M, James L,

Stegle O, Bayraktar OA. Cell2location maps fine-grained cell types in spatial transcriptomics. Nat Biotechnol. 2022 May;40(5):661–671. PMID: 35027729

67. Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat Rev Genet. 2021 Oct;22(10):627–644. PMID: 34145435

68. Littman R, Hemminger Z, Foreman R, Arneson D, Zhang G, Gómez-Pinilla F, Yang X, Wollman R. Joint cell segmentation and cell type annotation for spatial transcriptomics. Molecular systems biology. 2021;17(6):e10108. PMID: 34057817

69. Levy J, Davis M, Chacko R, Davis M, Fu L, Goel T, Pamal A, Nafi I, Angirekula A, Christensen B, Hayden M, Vaickus L, LeBoeuf M. ArcticAI: A Deep Learning Platform for Rapid and Accurate Histological Assessment of Intraoperative Tumor Margins [Internet]. medRxiv; 2022 [cited 2022 Jun 30]. p. 2022.05.06.22274781. Available from: https://www.medrxiv.org/content/10.1101/2022.05.06.22274781v1

70. Reddy R, Reddy R, Sharma C, Jackson C, Palisoul S, Barney R, Kolling F, Salas L, Christensen B, 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 [Internet]. bioRxiv; 2022 [cited 2022 Sep 23]. p. 2022.08.28.505606. Available from: https://www.biorxiv.org/content/10.1101/2022.08.28.505606v2

71. Levy JJ, Liu X, Marotti JD, Kerr DA, Gutmann EJ, Glass RE, Dodge CP, Suriawinata AA, Vaickus LJ. Uncovering additional predictors of urothelial carcinoma from voided urothelial cell clusters through a deep learning-based image preprocessing technique. Cancer Cytopathol. 2022 Aug 23; PMID: 35997513

72. Levy JJ, Vaickus LJ. Artificial Intelligence in Anatomic Pathology. Advances in Molecular Pathology. 2021 Nov;4:145–171. PMID: 33245914

73. Jackson CR, Sriharan A, Vaickus LJ. A machine learning algorithm for simulating immunohistochemistry: development of SOX10 virtual IHC and evaluation on primarily melanocytic neoplasms. Mod Pathol. 2020 Sep;33(9):1638–1648. PMID: 32238879

74. Frost HR. Tissue-adjusted pathway analysis of cancer (TPAC). bioRxiv. Cold Spring Harbor Laboratory; 2022;

75. Frost HR. Analyzing cancer gene expression data through the lens of normal tissue-specificity. PLoS Computational Biology. 2021;17(6):e1009085. PMID: 34143767

76. Frost HR. Computation and application of tissue-specific gene set weights. Bioinformatics. 2018;34(17):2957–2964. PMID: 29659714

77. Zhu J, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV). 2017. p. 2242–2251.

78. Zhang Y, de Haan K, Rivenson Y, Li J, Delis A, Ozcan A. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. Light: Science & Applications. 2020 May 6;9(1):78. PMID: 32411363

79. 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/

80. Levy J, Liu X, Marotti JD, Kerr DA, Gutmann EJ, Glass R, Dodge CP, Suriawinata AA, Vaickus LJ. Uncovering Additional Predictors of Urothelial Carcinoma from Voided Urothelial Cell Clusters Through a Deep Learning Based Image Preprocessing Technique [Internet]. bioRxiv; 2022 May p. 2022.04.30.490136. Available from: https://www.biorxiv.org/content/10.1101/2022.04.30.490136v1 PMID: 35997513

81. Jaume G, Pati P, Bozorgtabar B, Foncubierta A, Anniciello AM, Feroce F, Rau T, Thiran JP, Gabrani M, Goksel O. Quantifying explainers of graph neural networks in computational pathology. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. p. 8106–8116.

82. Ciga O, Xu T, Martel AL. Self supervised contrastive learning for digital histopathology. Machine Learning with Applications. Elsevier; 2022;7:100198.

83. Lu MY, Chen RJ, Wang J, Dillon D, Mahmood F. Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding. arXiv preprint arXiv:191010825. 2019;

84. 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

85. Zheng H, Momeni A, Cedoz PL, Vogel H, Gevaert O. Whole slide images reflect DNA methylation patterns of human tumors. npj Genomic Medicine. 2020 Mar 10;5(1):1–10. PMID: 32194984

86. Pope PE, Kolouri S, Rostami M, Martin CE, Hoffmann H. Explainability Methods for Graph Convolutional Neural Networks. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. p. 10764–10773. PMID: 34817762

87. Ahmedt-Aristizabal D, Armin MA, Denman S, Fookes C, Petersson L. A Survey on Graph-Based Deep Learning for Computational Histopathology. arXiv:210700272 [cs, q-bio] [Internet]. 2021 Jul 1 [cited 2021 Aug 27]; Available from: http://arxiv.org/abs/2107.00272 PMID: 34959100

88. 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.

89. OpenSeadragon [Internet]. [cited 2020 Jun 11]. Available from: http://openseadragon.github.io/ PMID: 34760328

90. 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 Spatially Resolved Transcriptomics Data from Whole Slide Images for the Assessment of Colorectal Tumor Metastasis: A Feasibility Study [Internet]. bioRxiv; 2022 [cited 2023 Mar 8]. p. 2022.11.24.517856. Available from: https://www.biorxiv.org/content/10.1101/2022.11.24.517856v1

91. Bergenstråhle J, Larsson L, Lundeberg J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. BMC genomics. 2020;21(1):1–7. PMID: 32664861

92. Andersson A, Bergenstråhle J, Asp M, Bergenstråhle L, Jurek A, Fernández Navarro J, Lundeberg J. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. Commun Biol. 2020 Oct 9;3(1):1–8. PMID: 33037292

93. Yuan Y, Bar-Joseph Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. Genome Biology. 2020 Dec 10;21(1):300. PMID: 33303016

94. Lopez R, Nazaret A, Langevin M, Samaran J, Regier J, Jordan MI, Yosef N. A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. arXiv preprint arXiv:190502269. 2019;