Virtual Picrosirius Red Generation Enables Spatial Analysis of Collagen Biology in Colorectal Cancer

Jennifer Choi, Joshua Levy
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth Health

ABSTRACT

• **Virtual collagen staining from routine H&E**
One step latent diffusion generates Picrosirius Red like images from colorectal cancer H&E.

• **Integration with spatial transcriptomics**
Quantified collagen architecture was linked to 18 molecular programs across 41 CRC specimens.

• **Spatial context reveals collagen biology**
Collagen molecular relationships were heterogeneous across tumors, but spatial models repeatedly identified ECM and fibroblast associated structure.

INTRODUCTION

Collagen Remodeling in Colorectal Cancer

• Collagen is a major structural component of the CRC extracellular matrix and is extensively remodeled during tumor progression.

• Changes in fiber length, width, density, branching, and organization may reflect stromal and tumor microenvironment states.

• Spatial transcriptomics captures molecular activity but does not directly measure physical collagen architecture.

Limitations of Picrosirius Red

• PSR enables direct visualization and quantification of collagen fibers.

• PSR is not routinely collected in clinical pathology, limiting its use in large retrospective cohorts.

• H&E is routinely available and may provide a scalable source for computational collagen reconstruction.

Study Goal

• Generate virtual PSR from H&E using one step latent diffusion.

• Quantify collagen architecture from generated images.

• Determine how collagen organization relates to spatial molecular programs and tumor proximity.

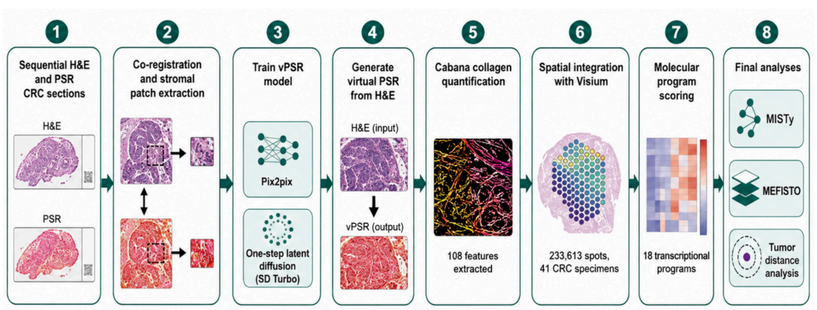

Figure 1: Workflow for virtual PSR generation and spatial transcriptomics analysis

METHODS

Virtual PSR Generation

• Four sequentially stained CRC H&E and PSR sections.

• 40,740 paired patches for training and validation.

• 4,527 paired patches for testing.

• Compared Pix2pix with a one step latent diffusion model based on SD Turbo.

Collagen Quantification

• Generated vPSR images analyzed using Cabana.

• 108 collagen features extracted.

• Eight primary features used for downstream spatial modeling, including fiber length, fiber area, lacunarity, fractal dimension, and branchpoint density.

Spatial Transcriptomic Integration

• Approximately 233,613 Visium spots across 41 CRC specimens with usable collagen measurements.

• Collagen features integrated with 18 transcriptional programs representing epithelial, fibroblast, ECM, immune, endothelial, and proliferative states.

Spatial Modeling

• **MISTy:** compared collagen measured at each spot with collagen from its six nearest spatial neighbors.

• **MEFISTO:** identified shared spatial latent factors between collagen architecture and molecular programs.

• **Tumor distance analysis:** evaluated collagen molecular relationships across increasing distance from pathologist annotated tumor boundaries.

RESULTS

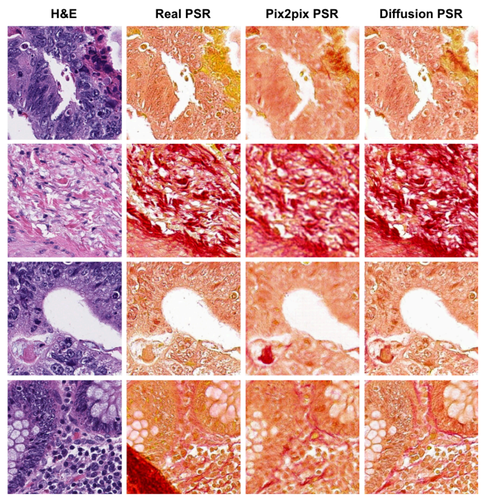

Figure 2. Virtual PSR model comparison. Representative H&E, real PSR, Pix2pix vPSR, and one step diffusion vPSR patches. Diffusion showed greater agreement with real PSR across collagen features, with a median Spearman correlation of **ρ = 0.667 versus 0.559** for Pix2pix.

RESULTS

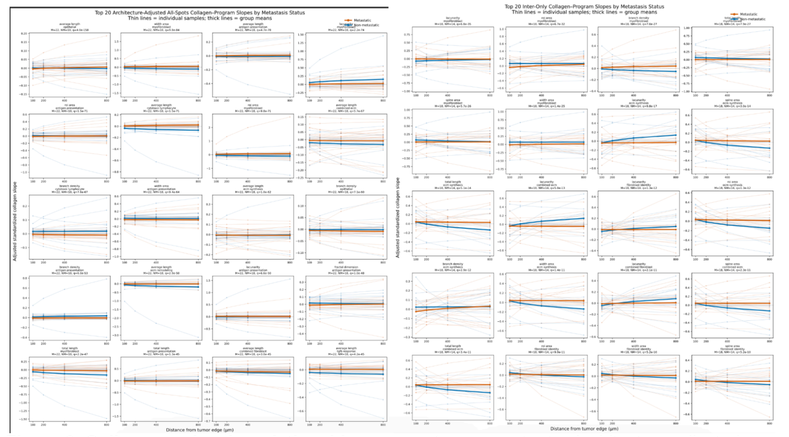

Figure 3. Collagen molecular relationships across tumor contexts. Architecture adjusted collagen program associations across CRC specimens, illustrating substantial variation in the magnitude and direction of associations between tumors.

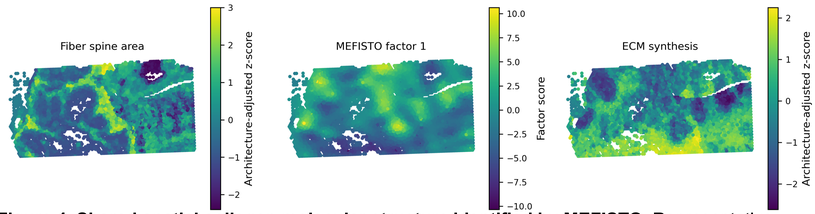

Figure 4. Shared spatial collagen molecular structure identified by MEFISTO. Representative specimen showing spatial correspondence among **fiber spine area, MEFISTO factor 1, and ECM synthesis**, demonstrating a latent spatial pattern jointly represented in collagen architecture and transcriptional activity.

CONCLUSION

Main Findings

• One step latent diffusion reconstructed quantitative collagen architecture from routine H&E more accurately than Pix2pix.

• Collagen architecture did not define one universal transcriptional state across CRC specimens.

• Spatial models revealed relationships that were obscured in conventional cohort wide association analyses.

• Neighboring collagen context and shared ECM and fibroblast programs emerged as recurring spatial features.

Limitations

• Strong between tumor heterogeneity limits universal collagen molecular associations.

• Virtual PSR was trained using a relatively small number of sequentially stained tissue sections.

• Spatial relationships require validation in independent CRC cohorts.

edit.