Spatial proteomics with hyperplexed immunofluorescence imaging predicts risk of colorectal cancer recurrence and infers recurrence-specific network biology

1642

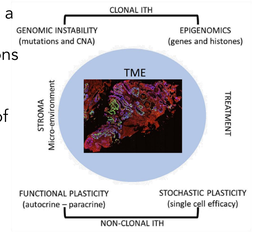
Shikhar Uttam[1], Andrew M. Stern[1,2], Samantha A. Furman[1], Filippo Pullara[1], Fiona Ginty[3], D. Lansing Taylor[1,2,4,5], S. Chakra Chennubhotla[1,5]

[1]University of Pittsburgh, Pittsburgh, PA, [2]Drug Discovery Institute, Pittsburgh, PA, [3]General Electric Global Research, Niskayuna, NY, [4]Hillman Cancer Center, Pittsburgh, PA, [5]Spintellx, Inc., Pittsburgh, PA

Emails: shf28 @ pitt.edu, {chakra, lans} @ spintellx.com

ABSTRACT

Motivation: Spatial context of the tumor microenvironment plays a critical role in disease progression, recurrence and response to therapy. In this work, our goal is to harness the spatial interactions for cancer prognosis.



Introduction: Colorectal cancer (CRC) is the third leading cause of cancer related deaths in United States with recurrence after resection-with-a-curative-intent being frequently implicated in these deaths. The basis for CRC recurrence is multifactorial including MSI status, lymph nodes etc., and may involve dysregulation of heterocellular signaling among tumor cells and their microenvironment.

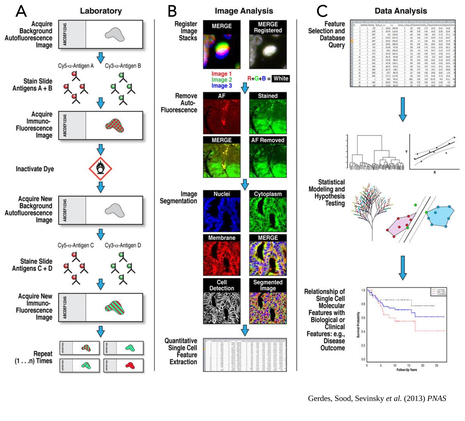
Based on hyperplexed immunofluorescence imaging with **Cell DIVE™** (GE Healthcare, Issaquah, WA) (**Fig. 1**) [1], single cell analysis and novel computational analyses, we have developed a recurrence-risk prediction method that samples these signaling networks within the epithelial and stromal domains of the tumor microenvironment and provides improved performance over current state-of-the-art recurrence-risk prediction assays.

Data: In the retrospective study presented here, we used 56 hyperplexed immunofluorescence biomarkers associated with canonical oncogenic pathways, metabolism, immune response and other colon cancer hallmarks to spatially profile resected tissue samples from 432 chemo-naïve CRC patients.

Results: Using epithelial- and stromal-domain expression and cellular co-expression diversity of the biomarkers, we predicted the risk of CRC recurrence with a concordance index of 0.91. We also generated training and validation sets from the CRC patient cohort and demonstrated that the area under the curve (AUC) of the prediction receiver operating characteristic (ROC) was 0.885. We utilized stratified bootstrapping to show that the AUC was stable with a standard deviation of 0.01. Significantly, the penalized model selection used within our method allowed us to infer epithelial and stromal-domain protein networks that are specific to the risk-of-recurrence. Despite the limited sampling that is intrinsic to tissue microarrays, we were able to capture immune cell infiltration and the differential modulation of these outcome specific networks.

Conclusions: Our CRC recurrence-risk prediction method exploits the spatial proteomics computational and systems pathology platform using hyperplexed immunofluorescence imaging and single cell analysis. This study demonstrates the potential of spatial proteomics to not only reveal the underlying systems pathophysiology, but also predict risk of CRC recurrence. Inferring outcome- and domain-specific CRC networks will enable biomarkers mechanistically linked to disease progression to be determined and their causality corroborated. In turn, this knowledge can potentially be used to inform optimal therapeutic strategies for individual patients.

METHODS AND RESULTS



Gerdes, Sood, Sevinsky *et al.* (2013) *PNAS*

Figure 1: Cell DIVE[1] sample analysis workflow. (**A**) In the laboratory, background autofluorescence (AF) tissue images are acquired before subsequent application of fluorescent dye-conjugated primary antibodies. Stained images are then acquired, followed by dye inactivation and restaining with new directly conjugated antibodies. New images are acquired, and the cycle is repeated until all target antigens are exhausted. (**B**) Stained images are registered, background AF is removed from each stained image, and images are segmented into epithelial and stromal regions. (**C**) Data analysis can consist of a variety of statistical and visual explorations.

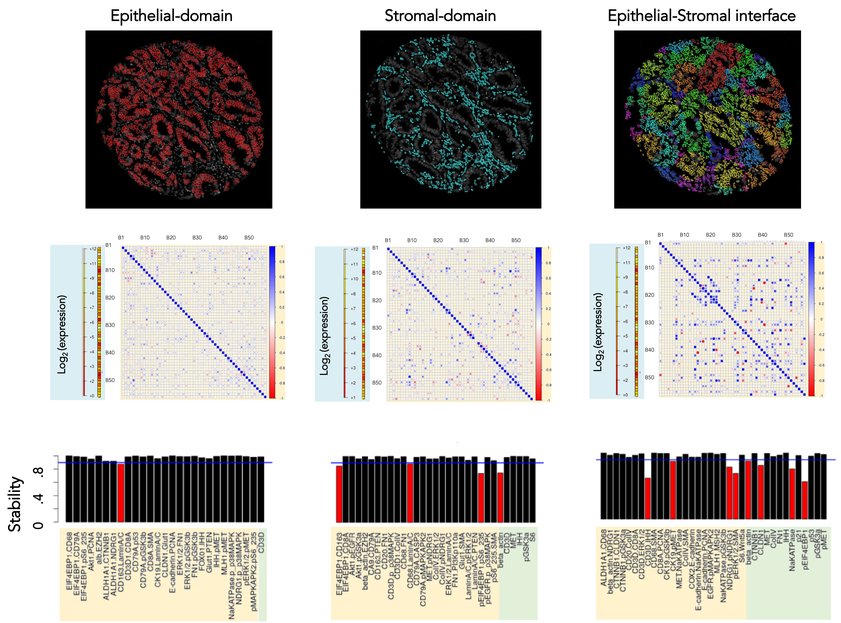


Figure 2. (**Top**) Spatial domain decomposition model with epithelial cells identified by E-cadherin staining, with individual cells delineated by Na+K+ATPase cell membrane marker and DAPI based nuclear staining. (**Middle**) Intensity expressions and Kendell rank correlations between biomarker pairs provide a very high-dimensional feature space for each tumor spot and are in turn fed into a recurrence-guided learning model for feature selection. (**Bottom**) Features are tested for stability in their contribution to recurrence prognosis at a 90% concordance threshold. The domain-specific features are finally combined into a single spatial-domain prognostic test.

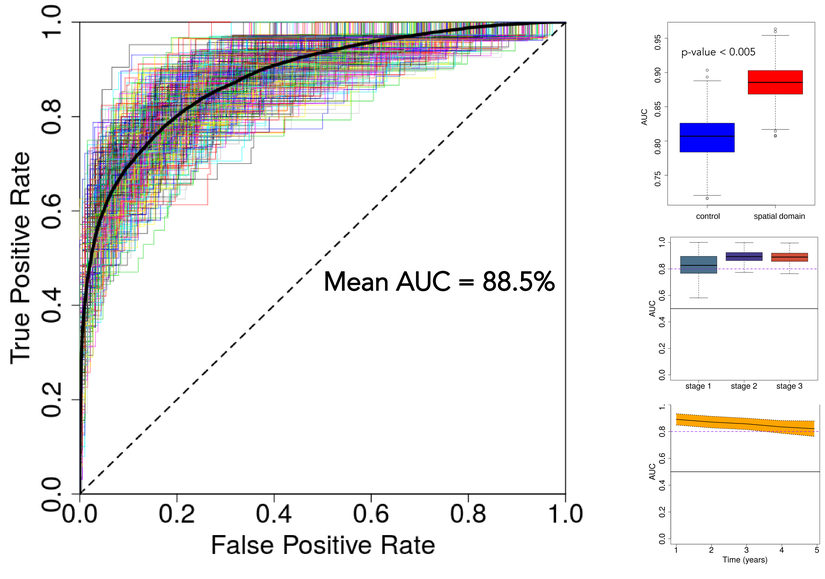


Figure 3. Learning recurrence-guided and spatially informed prognostic test for CRC recurrence. (**Left**) ROC curves were obtained by bootstrapping patient data set to generate 500 pairs of independent training and testing sets using stratified sampling. Mean area under the curve (AUC) for the ROCs is 88.5% with a standard error of 0.1%. (**Right-Top**) Spatial model of deconvolving tumor spots into epithelial, stromal and epithelial-stromal domains provides a significant performance improvement over a control model without domain decomposition. (**Right-Middle**) Predicting the risk of recurrence in individual patients from all Stages remains high with with mean AUC of bootstrapped ROC curves for the three stages respectively being 82.1%, 89.4% and 88.6%. Standard error in these mean AUC values is 0.4%, 0.2%, and 0.2% respectively. (**Right-Bottom**) Time-dependent AUC values of the bootstrapped ROC curves remains consistent and stable (95% confidence interval shown) with only a small, expected and graceful degradation moving away from resection timepoint.

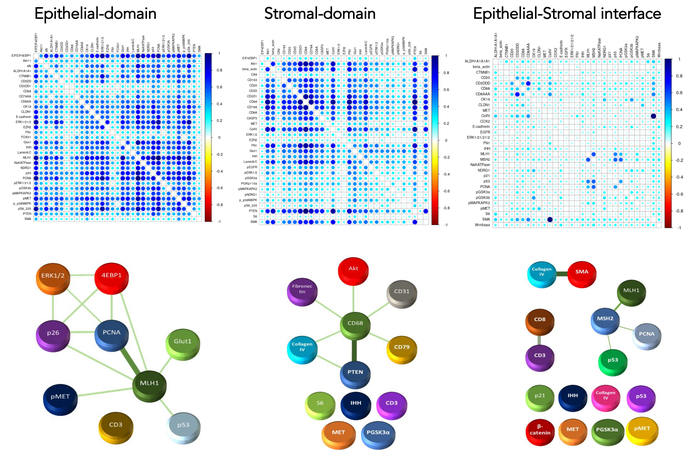


Figure 4. Identifying spatial-domain networks to discriminate patient cohorts with and without evidence of disease recurrence with (**Top**) distance graphs where nodes are biomarkers and edge weights quantify the information distance between the two cohorts. (**Bottom**) Spatial-domain networks derived by thresholding information distances reveal the heterogeneous nature of cell populations and signaling pathways. For example, the epithelial-stromal domain network reveals the prognostic role of three dominant sub-networks associated with tumor-invading T lymphocytes, disruption in DNA mismatch repair cellular process, and the role of cancer associated fibroblasts in the desmoplastic microenvironment.

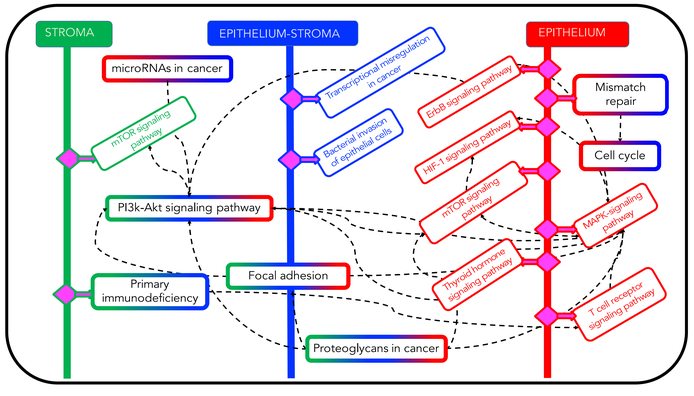


Figure 5. Validation of spatial domain networks with systems biology. We used curated databases to identify pathways enriched by biomarkers within each of the spatial domain networks. For illustration, we show pathways enriched in a majority of – at least two of the three – spatial domains. Since their identification is based on the spatial-domain networks we identified as most significant for CRC recurrence prognosis, these pathways play a differentially important role in prognosis of CRC recurrence. Interestingly, these pathways reveal a close connection to the CRC consensus molecular subtypes (CMS).

FUTURE DIRECTIONS

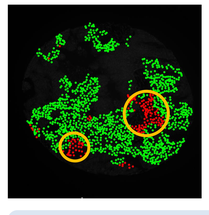


Figure 6. The computational and systems pathology platform developed here focuses on interrogating tumor microenvironments with spatial analytics. An important goal for future work is to examine how risk is spatially organized within tumor cores. The discovery of high risk microdomains within tumor cores can potentially reveal space specific risk signatures within tumor microenvironments that are associated with disease progression and CRC metastasis.

REFERENCES

1. Gerdes et al. Proc Natl Acad Sci U S A. 2013 Jul 16;110(29):11982-7
2. Spagnolo et al. Cancer Res. 2017 Nov 1;77(21):e71-e74

We thank the support of NIH-NCI ITCR grant 1U01CA204826.
D. L. Taylor and S. C. Chennubhotla are co-founders, equity stakeholders and officers of SpIntellx