Mixed Effects Machine Learning on Spatially Localized Immuno-Oncology Markers for Colon Metastasis Prediction

Joshua J. Levy PhD, Carly A. Bobak PhD, Mustafa Nasir-Moin BS, Eren M. Verizoglu BS, Scott M. Palisoul BS, Rachael E. Barney BS, Lucas A. Salas MD MPH MSc PhD, Brock C. Christensen PhD, Gregory J. Tsongalis PhD, Louis J. Vaickus MD PhD
Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

Dartmouth-Hitchcock

ABSTRACT

- **Spatially resolved** characterization of transcriptome and proteome promises to further **clarify cancer pathogenesis**
- **Batch effects** and **nested dependencies** within slide threaten to preclude classifier deployment by learning batch signal
- **Mixed effects machine learning (MEML)** methods overcome batch effects to **communicate key disease findings**

INTRODUCTION

- **Spatial Omics Analyses**: Spatial localization of highly multiplexed gene and protein markers to estimate:
  - Spatially variable genes
  - Spatial clustering patterns indicative of disease status
  - Interactions between co-localized cellular populations
  - Integrate with imaging morphology, single cell RNASeq
- **Nanostring GeoMx Digital Spatial Profiler (DSP):**
  - Immuno-fluorescent (IF) antibody stains, linked to UV cleavable oligo tags, demarcate cell lineages
  - Region of interest (ROI) selection
  - Image segmentation of cells for UV cleavage
  - Quantifies cleaved oligo tags for protein expression
- **Machine Learning:** Heuristic search $f_\theta(x_i)$ for nonlinear transformation and interactions relates input to output
  - Classification and regression trees (CART): Tackles prohibitive dimensionality/collinearity via conditional decision splits from nonparametric bootstrapping
  - Performance degradation from data clustering (e.g., patient, batch, spatial autocorrelation, reagents)
- **Mixed Effects Machine Learning (MEML):**
  - Gaussian Process: Dependency structure Σ captures batch/correlation: $y_i = \beta \cdot \overrightarrow{x_i} + b_{[i]} + \epsilon_i, b_{[i]} \sim N(0,\Sigma)$
  - MEML: Replace fixed effects component with machine learning model: $y_i = f_\theta(x_i) + b_{[i]} + \epsilon_i, b_{[i]} \sim N(0,\Sigma)$
  - Identifies interactions robust to data clustering

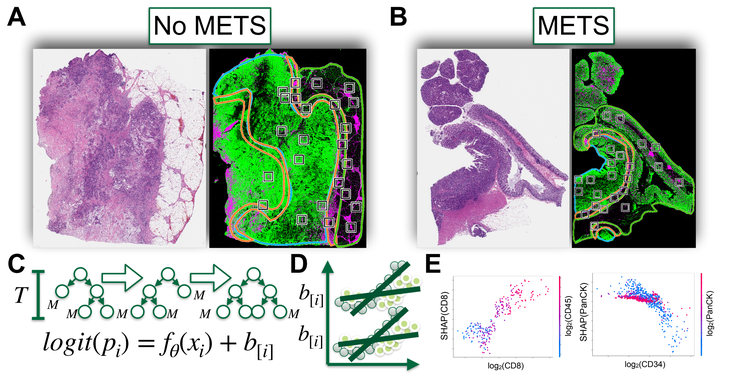

Figure 1: DSP workflow identifies batch-invariant interactions with MEML

METHODS

- **Colorectal Cancer (CRC):**
  - Annual incidence in the United States: ~150,000 new cases
  - 63% 5-year survival rate
  - Somatic alterations (e.g., MMR, APC) initiate tumorgenesis
  - Metastatic potential (METS) characterized by local invasion
  - Tumor infiltrating lymphocytes (TIL) important prognostic factor
- **Goal:** Develop machine learning classifier to relate spatial TIL protein expression from initial biopsy to nodal/distant metastasis
- **Experimental Design:**
  - 35 patients, pTNM stage 3, age/sex/MMR/site matched
  - Use H&E stain to annotate IF stain with 3 macro-architectural regions (***intra**-tumoral, tumor **inter**face, **away** from tumor*)
  - Segment immune cells within 24 ROIs/slide (840 total ROI)
  - UV cleavage/quantification of 39 immuno-oncology proteins
- **Tasks:**
  - Predict tumor interface, within patient (***WS***) and held-out (***OOS***)
  - Predict tumor METS, across all and within macro-architectures
- **Comparison Algorithms:**
  - Fixed effects: Random Forest, XGBoost
  - MEML: Gaussian Process Boosting (GPBoost) with spatial exponential kernel; Bayesian Additive Regression Trees (BART)
  - Traditional Approaches: Bayesian Generalized Linear Mixed Effects Models (BGLMM); with interactions extracted from GPBoost with SHAP (BGLMM-Int), unpenalized for odds ratios

RESULTS

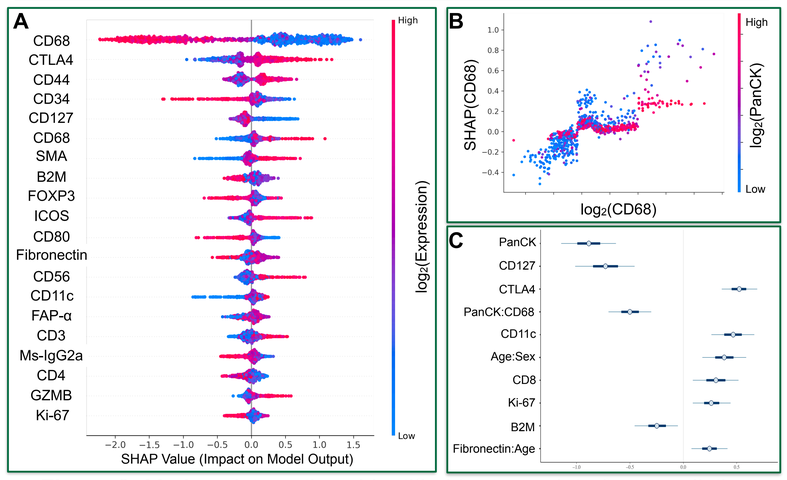

Figure 2: Markers/interactions identified for tumor interface prediction

Table 1: GPBoost and BGLMM-Int models outperform other approaches

		Fixed Effects		MEML			Bayesian Generalized Linear Mixed Models		
Task	AUC ± SE	RF	XGBoost	GPBoost	GPBoost-Coords	SP-BART	BGLMM	BGLMM-Int	BGLMM-GP
Macro	OOS	0.759±0.01	0.747±0.01	0.752±0.01	0.747±0.01	0.772±0.01	0.778±0.01	**0.785±0.009**	0.777±0.01
	WS	0.781±0.009	0.773±0.01	0.782±0.009	0.78±0.009	0.784±0.009	0.788±0.009	**0.802±0.009**	0.786±0.01
METS	Overall	0.909±0.006	0.951±0.004	**0.971±0.003**	n/a	0.897±0.006	0.852±0.008	0.896±0.006	n/a
	Intra	0.834±0.014	0.849±0.014	**0.89±0.011**	n/a	0.867±0.012	0.848±0.013	0.877±0.012	n/a
	Inter	0.881±0.013	0.866±0.013	**0.899±0.012**	n/a	0.874±0.012	0.836±0.014	0.881±0.012	n/a
	Away	0.827±0.014	0.842±0.014	**0.895±0.011**	n/a	0.887±0.011	0.885±0.012	0.885±0.011	n/a

RESULTS

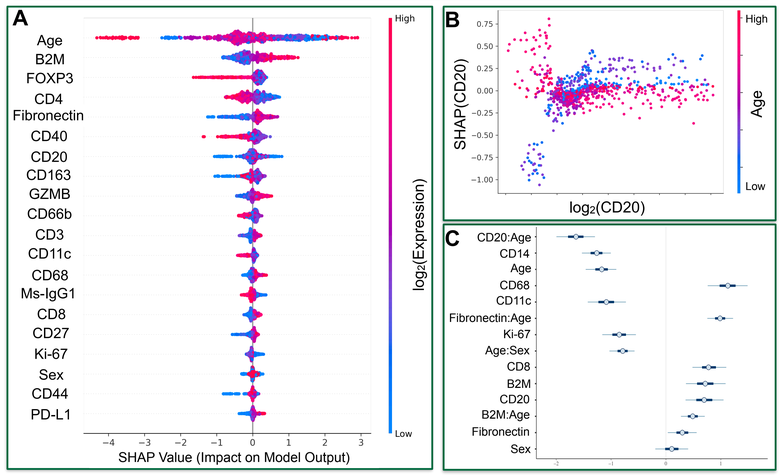

Figure 3: Markers/interactions identified for tumor METS prediction

CONCLUSION

- **Potential for Clinical Impact:**
  - MEML methods mine BGLMM statistical significant interactions
  - Elucidates disease pathology through reports of odds ratios for effect modifiers for metastasis (e.g., CD20 B-cells and aging)
  - Identified markers can be orthogonally validated/scaled with multiplexed immunohistochemistry (IHC) and IF for deployment
- **Limitations:**
  - Coarse hyperparameter search and omitted MEML methods
  - Temporal dependence for causal associations unresolved
  - Uncovers batch independent factors but requires batch exclusion during prediction on external cohorts
- **Future Directions:**
  - Clinical findings after algorithmic fine-tuning, final dataset curation
  - Explore potential for biased sampling and optimal DSP experiment planning/standardization
  - Further research on MEML methods and applying spatial kernels
  - Orthogonal validation, multiplexed modalities (e.g., multiplexed IF)
- **Data and Code Availability:**
  - GitHub: https://github.com/jlevy44/MEML_Colon_DSP_METS.
  - Data available on reasonable request, privacy/ethical restrictions.
- **Funding:**
  - NCI Cancer Center Support Grant 5P30 CA023108-37
  - Neukom Institute CompX Awards
  - NIH grant R01CA216265
  - Burroughs Wellcome Fund training grants

Acknowledgements: James O'Malley, Robert Frost, Prajan Divakar, and Christian Haudenschild for leadership, support and discussion. CGAT, EDIT, DPLM, Pathology Shared Resource, NCCC @ DHMC.

References: Available using QR code:

edit.