Evaluating the Performance of ProteoMixture, a Proteomics-Based Cell Deconvolution Tool in Pan-Cancer Data for > 1000 Patient Tumors

Joshua P. Schaaf [1,2], Tamara Abulez [1,2], Pang-ning Teng [1,2], Brian L. Hood [1,2], Kelly A. Conrads [1,2], Tracy Litzi [1,2], Allison Hunt [1,3,4], Kathleen M. Darcy [1,2,3], Neil T. Phippen [1,3], George L. Maxwell [1,3,4], Thomas P. Conrads [1,3,4], Nicholas W. Bateman [1,2,3]

[1]Gynecologic Cancer Center of Excellence, Department of Gynecologic Surgery and Obstetrics, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Bethesda, MD, USA. [2]The Henry M. Jackson Foundation for the Advancement of Military Medicine Inc., Bethesda, MD, 20817, USA. [3]Murtha Cancer Center Research Program, Department of Surgery, Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Bethesda, Maryland, 20889, USA. [4]Women's Health Integrated Research Center, Women's Service Line, Inova Health System, Falls Church, VA, USA.

SCHOOL OF MEDICINE — Uniformed Services University

Background

- Cellular heterogeneity within the tumor microenvironment has been shown to impact cancer prognosis and treatment.
- xCell[4] and ESTIMATE[3] provide subtype enrichment scores from gene expression.
- ProteoMixture[1] provides subtype enrichment scores from protein abundance.

Objective
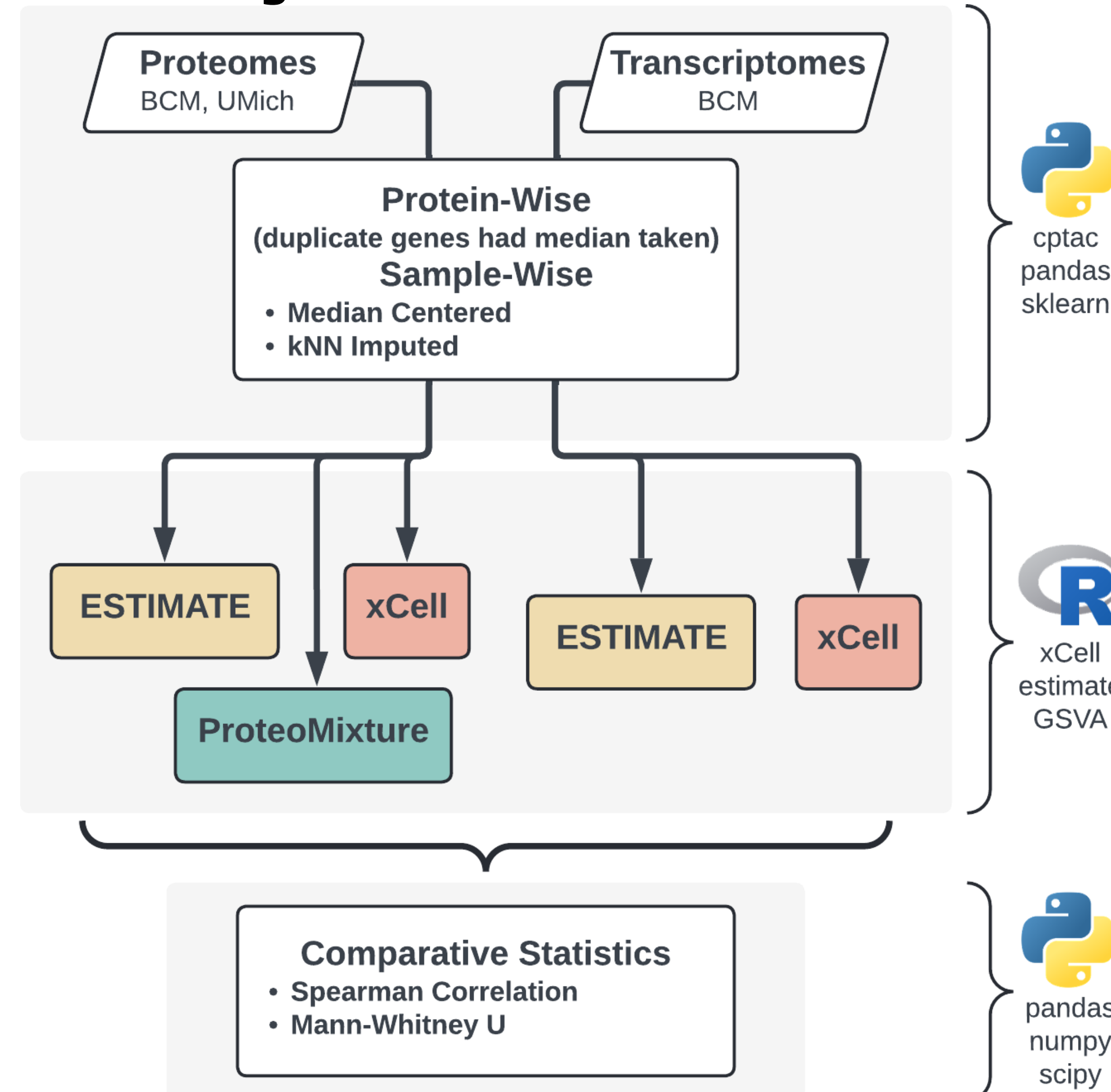
Compare performance between ProteoMixture and xCell/ESTIMATE in pan-cancer bulk tumor samples in the Clinical Proteomic Tumor Analysis Consortium[2] (CPTAC).

Methods

Creating ProteoMixture

LMD enrichment of tumor (T), stroma (S), and lymphocytes (L) from HGSOC tissues
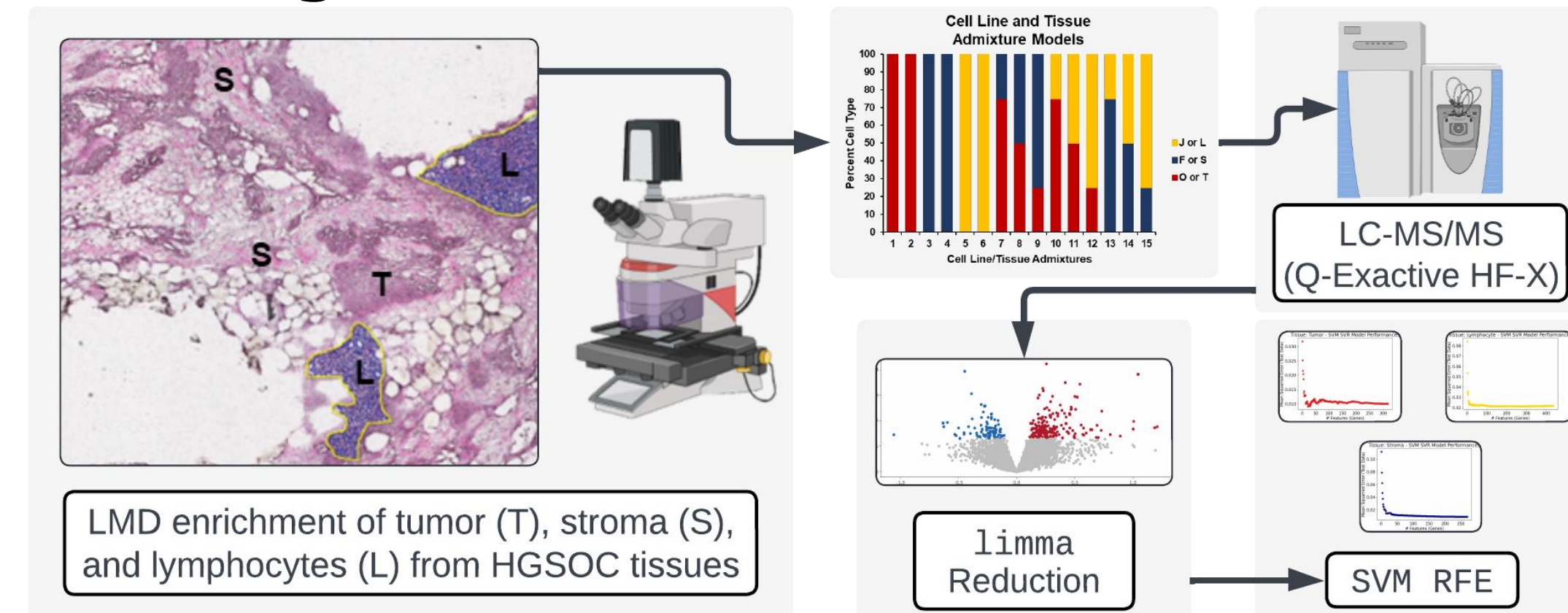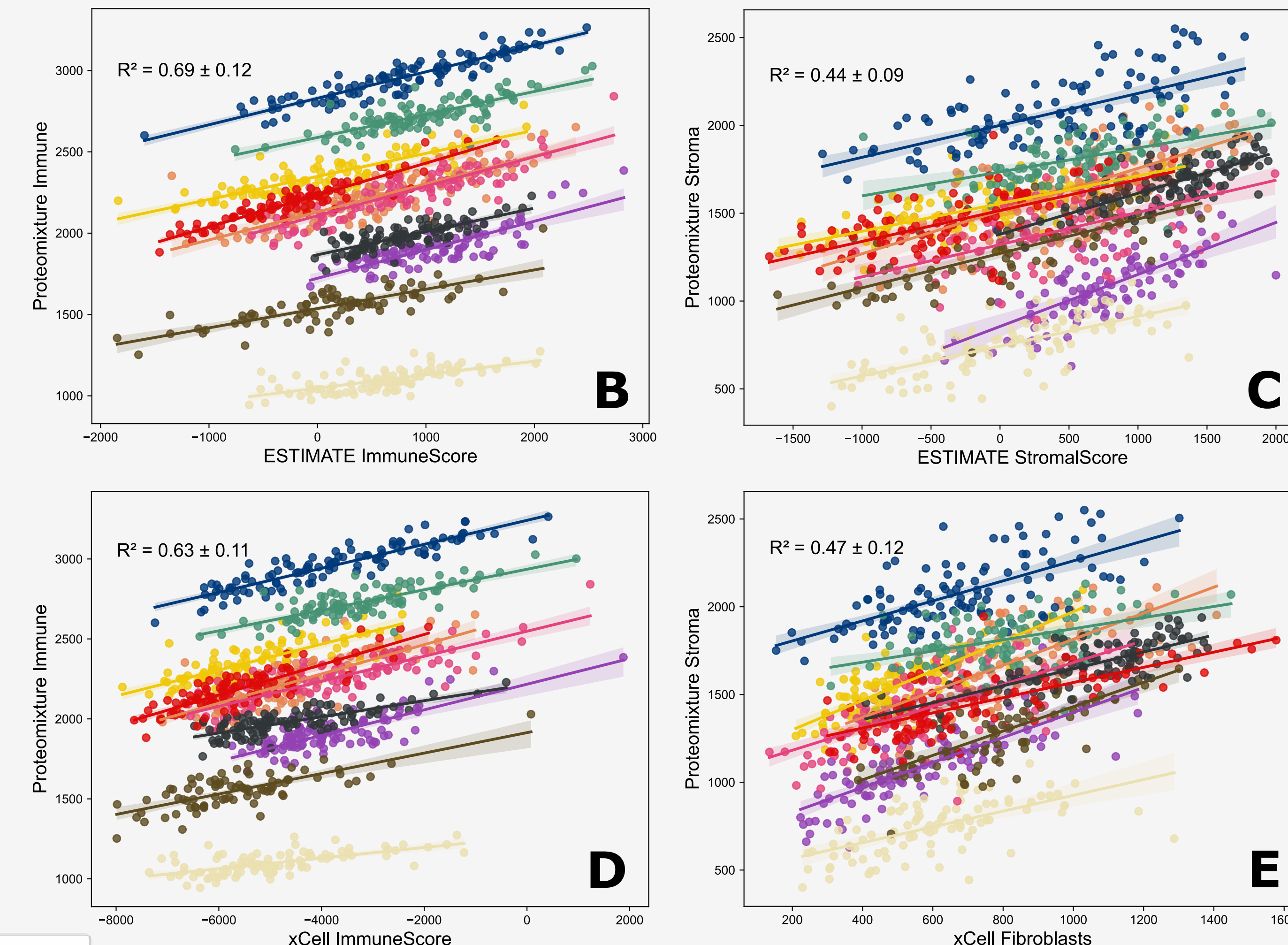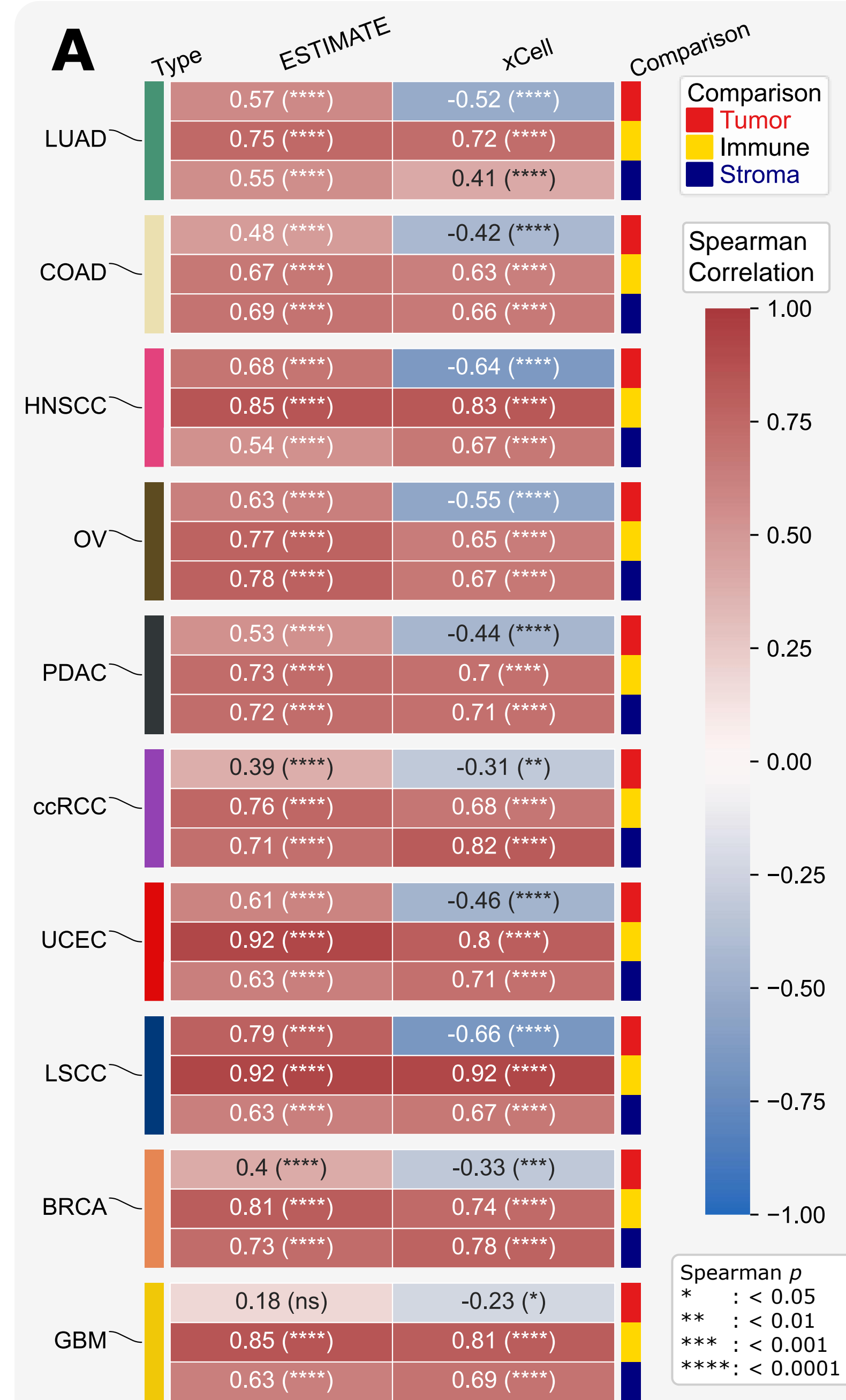
Evaluating Performance

Results

Figure 1A. ProteoMixture Tumor, Stroma, or Immune score spearman correlations to their respective ESTIMATE and xCell scores in 10 CPTAC cancer types (BCM).

B, C, D, E. ProteoMixture immune/stroma scores correlated to the corresponding ESTIMATE (**B**, **C**) or xCell (**D**, **E**) score.
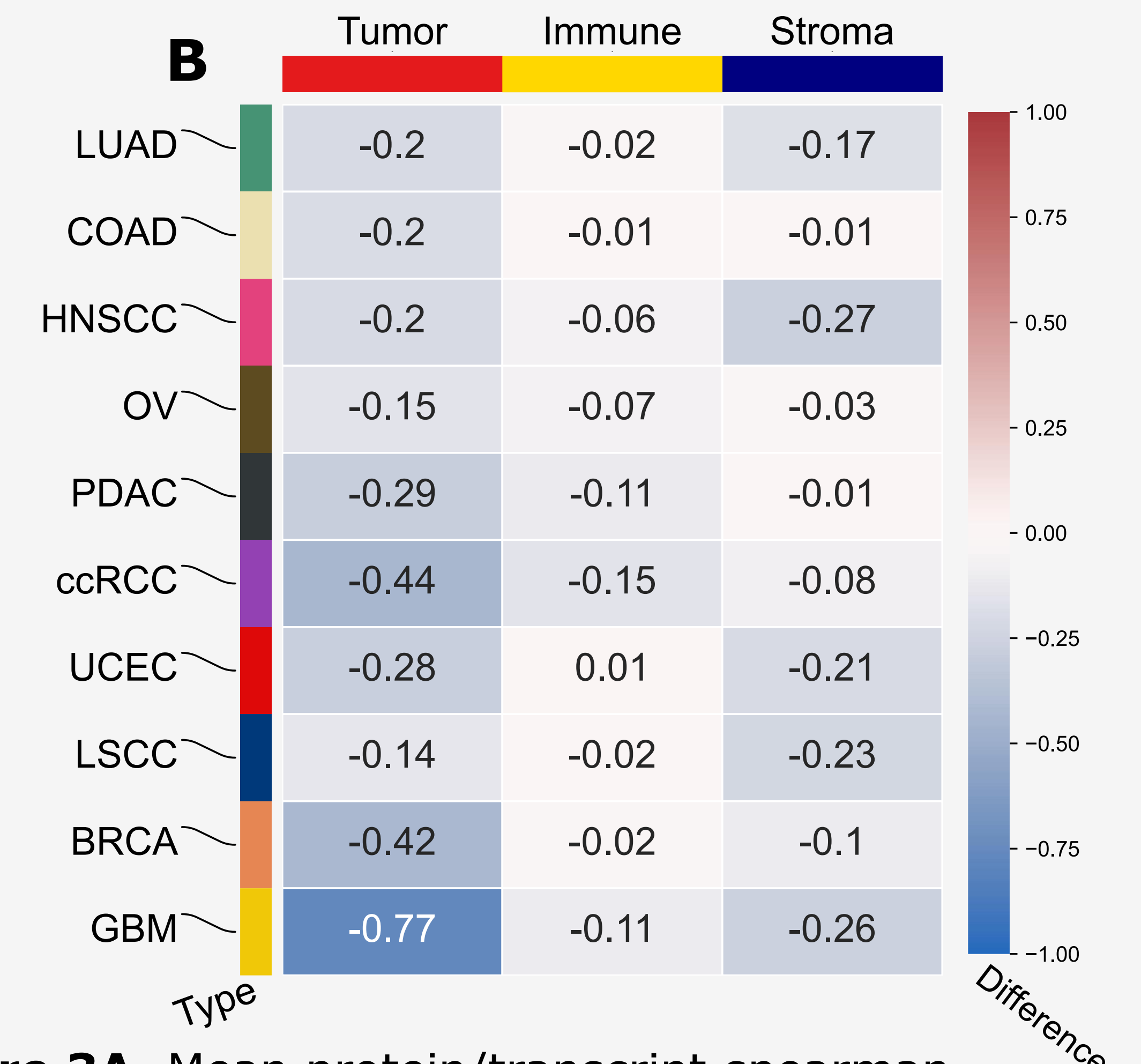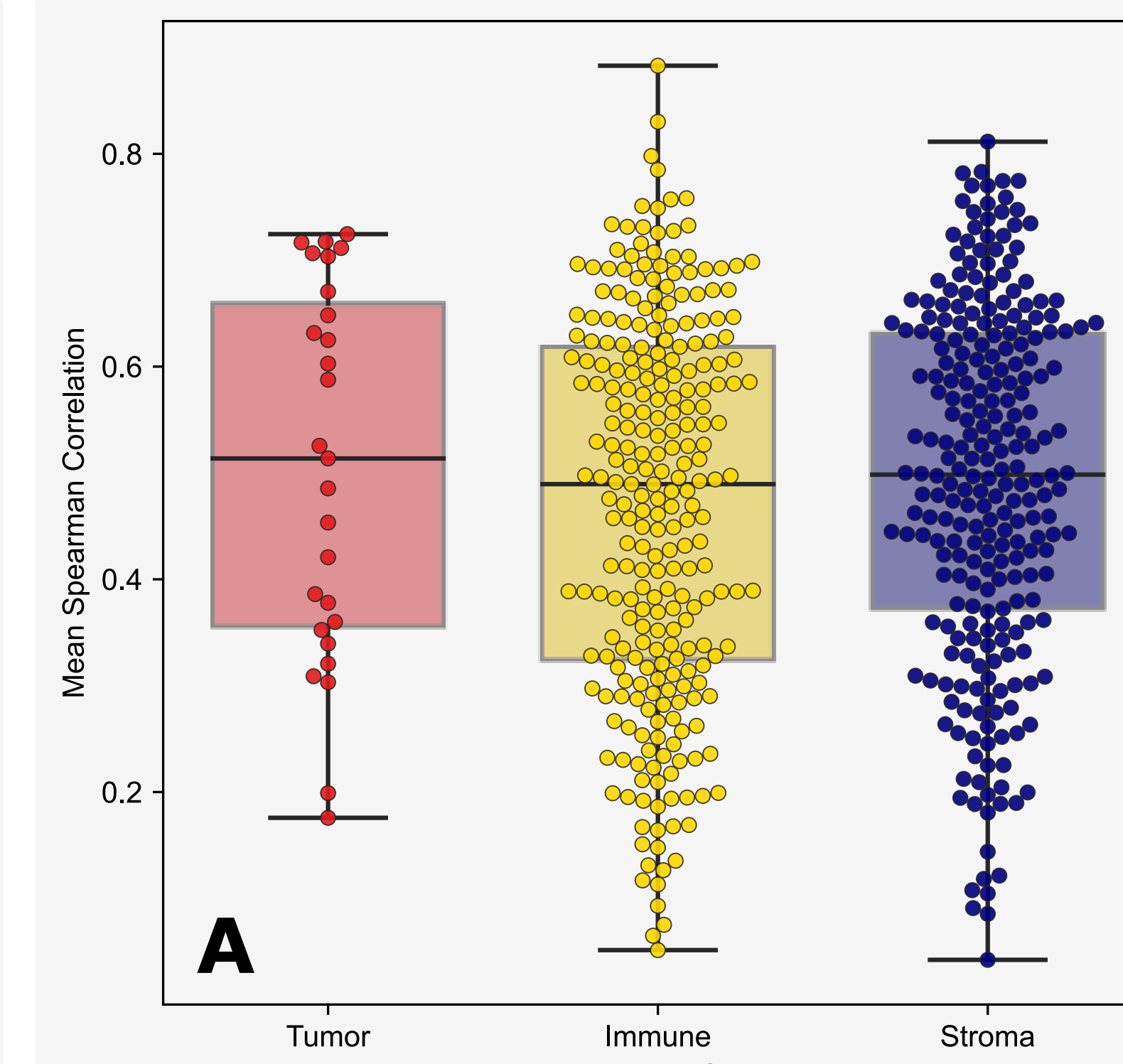

Figure 3A. Mean protein/transcript spearman correlations across the 10 CPTAC cancer types, split by ProteoMixture geneset.

B. The difference between ProteoMixture scores correlated to ESTIMATE scores and ESTIMATE run on proteome correlated to ESTIMATE run on transcriptome.

C. CPTAC cancer type sample-wise spearman correlations between ProteoMixture scores on BCM and UMich proteome data.

Figure 2A. Distributions of median gene-wise protein/transcript spearman correlations across 10 CPTAC cancer types, using matrices from BCM. Annotations represent significant two-tailed Wilcoxon rank-sum comparisons between protein abundances of all proteins and deconvolution tool proteins.

B. Distributions of median gene-wise protein/transcript spearman correlations across 10 CPTAC cancer types, using matrices from BCM. Annotation represents a one-tailed Wilcoxon rank-sum result.

C. Deconvolution tool geneset completeness over 10 CPTAC cancer types in proteome data from BCM.

Conclusions

ProteoMixture performs well on pan-cancer bulk proteome data:

- ProteoMixture Immune and Stroma scores are highly correlated to their respective ESTIMATE and xCell scores in all 10 pan-cancer tumor types. (**Figure 1**)
- ProteoMixture Tumor score has lower correlation to ESTIMATE and xCell, most likely due to its originating from HGSOC tumor cells. (**Figure 1**)
- ProteoMixture genes show decreased protein/transcript correlations when compared to ESTIMATE genes. (**Figure 2B**)
- ProteoMixture genesets are better accounted for across the CPTAC cancers than transcript-based tools ESTIMATE and xCell (**Figure 2C**)

ProteoMixture Online Tool

Joshua Schaaf
(he/him)
Bioinformatics
Programmer Analyst II

My GitHub

[1]Teng et al. "ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data." *iScience*, 2024, https://doi.org/10.1016/j.isci.2024.109198
[2]Li et al. "Proteogenomic data and resources for pan-cancer analysis." *Cancer Cell*, 2023, https://doi.org/10.1016/j.ccell.2023.06.009
[3]Yoshihara et al. "Inferring tumour purity and stromal and immune cell admixture from expression data." *Nature Communications*, 2013, https://doi.org/10.1038/ncomms3612
[4]Aran et al. "xCell: digitally portraying the tissue cellular heterogeneity landscape." *Genome Biology*, 2017, https://doi.org/10.1186/s13059-017-1349-1