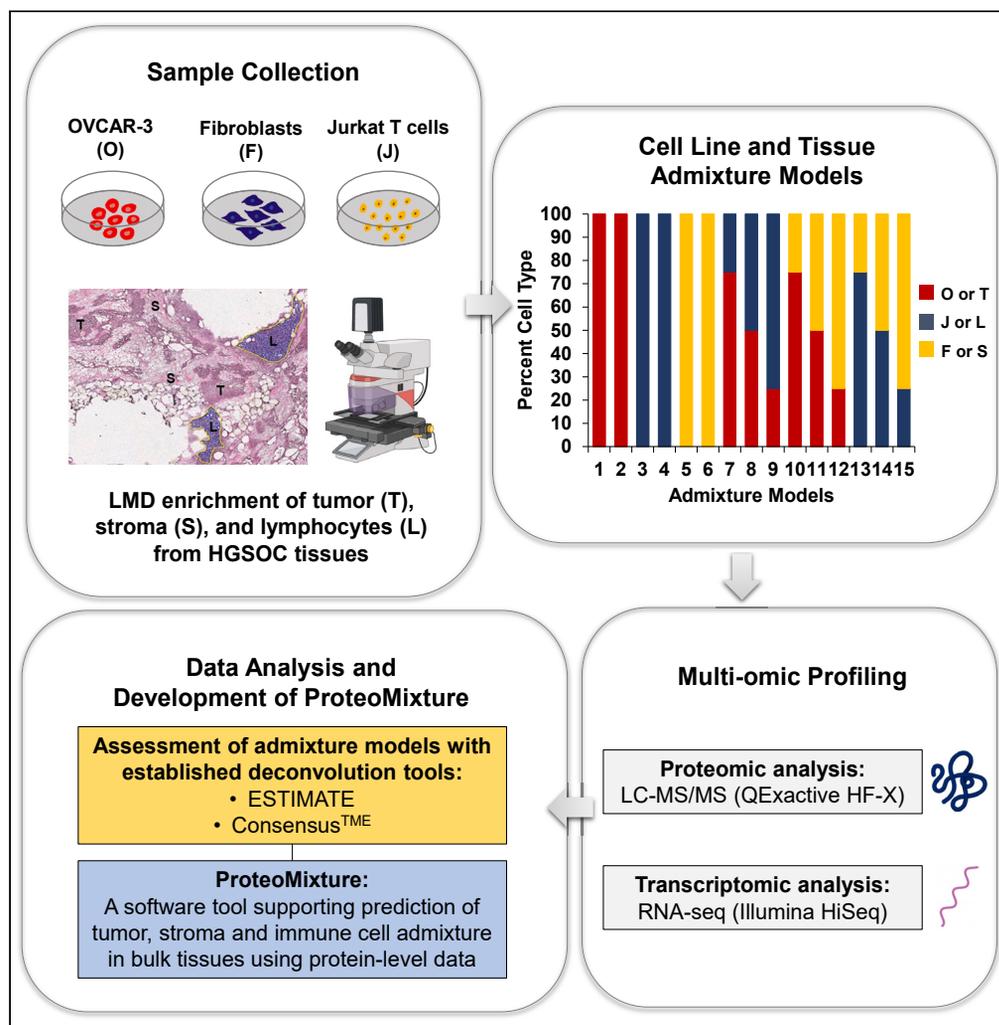


Article

ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data



Pang-ning Teng,
Joshua P. Schaaf,
Tamara Abulez, ...,
G. Larry Maxwell,
Thomas P.
Conrads, Nicholas
W. Bateman

conrads@whirc.org (T.P.C.)
batemann@whirc.org (N.W.B.)

Highlights

ProteoMixture predicts cellular admixture in proteomic data from bulk tissues

Optimized protein signatures of cellular admixture were validated in independent cohorts

Cellular admixtures model heterogeneity within the tumor microenvironment

Article

ProteoMixture: A cell type deconvolution tool for bulk tissue proteomic data

Pang-ning Teng,^{1,2,7} Joshua P. Schaaf,^{1,2,7} Tamara Abulez,^{1,2} Brian L. Hood,^{1,2} Katlin N. Wilson,^{1,2} Tracy J. Litzi,^{1,2} David Mitchell,^{1,2} Kelly A. Conrads,^{1,2} Allison L. Hunt,^{1,3} Victoria Olowu,^{1,2} Julie Oliver,^{1,2} Fred S. Park,^{1,2} Marshé Edwards,^{1,2} AiChun Chiang,^{1,2} Matthew D. Wilkerson,⁴ Praveen-Kumar Raj-Kumar,⁵ Christopher M. Tarney,^{1,6} Kathleen M. Darcy,^{1,2,6} Neil T. Phippen,^{1,6} G. Larry Maxwell,^{1,3,6} Thomas P. Conrads,^{1,3,6,*} and Nicholas W. Bateman^{1,2,6,8,*}

SUMMARY

Numerous multi-omic investigations of cancer tissue have documented varying and poor pairwise transcript:protein quantitative correlations, and most deconvolution tools aiming to predict cell type proportions (cell admixture) have been developed and credentialed using transcript-level data alone. To estimate cell admixture using protein abundance data, we analyzed proteome and transcriptome data generated from contrived admixtures of tumor, stroma, and immune cell models or those selectively harvested from the tissue microenvironment by laser microdissection from high grade serous ovarian cancer (HGSOC) tumors. Co-quantified transcripts and proteins performed similarly to estimate stroma and immune cell admixture ($r \geq 0.63$) in two commonly used deconvolution algorithms, ESTIMATE or Consensus^{TME}. We further developed and optimized protein-based signatures estimating cell admixture proportions and benchmarked these using bulk tumor proteomic data from over 150 patients with HGSOC. The optimized protein signatures supporting cell type proportion estimates from bulk tissue proteomic data are available at <https://lmdomics.org/ProteoMixture/>.

INTRODUCTION

The ovarian cancer tumor microenvironment (TME) includes various cell types such as tumor, stroma, and immune cells that can regulate tumor development and progression.^{1,2} Immune cell populations, including tumor-associated/infiltrated lymphocytes (TILs), in the TME have been shown to impact cancer prognosis and response to neoadjuvant chemotherapy (NACT).^{3,4} Proteogenomic analyses of high grade serous ovarian cancer (HGSOC) to date have largely utilized bulk tumor collections that contain widely varying admixtures of diverse cell types.^{5,6} Our group⁷ and others⁸ have shown that variations in the proportions of different cellular populations within the TME can impact correlation with different HGSOC prognostic molecular subtypes.^{9–13} Improved the characterization of cell admixture contributions to the bulk tissue proteome will support the refinement of proteogenomic signatures from bulk and enriched cell type collections.

Deconvolution of cell type proportions (cell admixture) from bulk expression data has previously been achieved by quantifying the enrichment of cell type-associated gene expression signatures. Current deconvolution tools include the Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE),¹⁴ xCell,¹⁵ Microenvironment Cell Populations-counter (MCP-counter)¹⁶ and CIBERSORTx.^{14–18} Some of these tools have also recently been merged into an integrated tool, Consensus^{TME}¹⁸, enabling the prediction of cell admixture for 18 different cell types, including fibroblasts, endothelial cells, and 16 immune-related cell types. These tools have been developed using transcript-level data, which exhibits a limited correlation to proteome abundances in cancer cells and tissues, including HGSOC,^{6,19} largely due to translational regulation.²⁰ Thus, there remains a paucity of data investigating the applicability of these signatures for characterizing cellular admixture within proteome data in HGSOC tissues. Very recent efforts by Feng et al.²¹ described the Decomprolute tool, which enables the prediction of immune cell signatures using proteomic data, and established deconvolution tools across various organ site malignancies, including ovarian cancer. Motivated by this work as well as recent efforts by our group correlating stromal cell admixture

¹Gynecologic Cancer Center of Excellence and the Women's Health Integrated Research Center, Annandale, VA 22003, USA

²The Henry M. Jackson Foundation for the Advancement of Military Medicine Inc., Bethesda, MD 20817, USA

³Women's Health Integrated Research Center, Women's Service Line, Inova Health System, Falls Church, VA 22042, USA

⁴Center for Military Precision Health, Department of Anatomy, Physiology and Genetics, Uniformed Services University of the Health Sciences, Bethesda, MD 20814, USA

⁵Chan Soon-Shiong Institute of Molecular Medicine at Windber, Windber, PA 15963, USA

⁶The John P. Murtha Cancer Center, Department of Surgery, Uniformed Services University and Walter Reed National Military Medical Center, Bethesda, MD 20889, USA

⁷These authors contributed equally

⁸Lead contact

*Correspondence: conrads@whirc.org (T.P.C.), batemann@whirc.org (N.W.B.)

<https://doi.org/10.1016/j.isci.2024.109198>



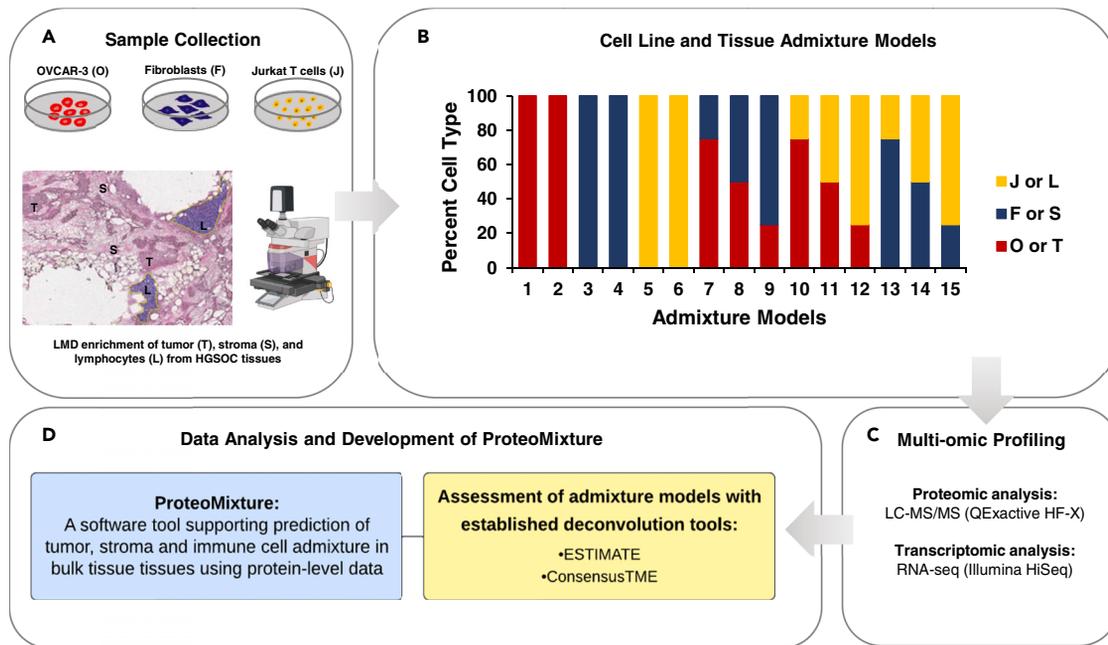


Figure 1. Analytical Workflow Supporting Proteogenomic Analysis of Cellular Admixture in High Grade Serous Ovarian Cancer (HGSOC)

(A and B) Cell admixtures were generated by mixing varying amounts of cell type-associated peptide digests or extracted RNA collected from cell line models or subpopulations of cell of interest from HGSOC tissues. 1–11: Cell Line and Tissue Admixtures for LC-MS/MS and RNA-seq. 12–15: Cell Line Admixture for LC-MS/MS and RNA-seq; Tissue Admixture for RNA-seq.

(C and D) Cell admixtures were analyzed by LC-MS/MS and RNA-seq followed by assessment in established deconvolution and prognostic molecular subtype prediction tools. Protein signatures predicting tumor, stroma, and immune cell admixture were optimized and validated in independent HGSOC cohorts. Protein-level signatures were distilled into a publicly available tool supporting cell type deconvolution from bulk tissue proteomic data (<https://lmdomics.org/ProteoMixture/>).

with the prediction of the mesenchymal (MES) subtype,⁷ a molecular subtype correlating with poor disease prognosis in HGSOC, we examined proteomic signatures of tumor, stroma, and immune cell admixture in HGSOC.

Our study describes an evaluation of the performance of matched transcriptome and proteome data generated from a contrived admixture series of HGSOC tumor, stroma/fibroblasts, and immune cells using existing deconvolution and prognostic molecular subtype prediction tools. We have further investigated the impact of cell type admixture on the correlation of protein and transcript abundances. We describe optimized protein signatures for tumor, stroma, and immune cell admixtures and their performance in classifying proteome data from enriched and bulk tissue collections for multiple, independent HGSOC patient cohorts. We provide these signatures as part of a publicly available tool, ProteoMixture, supporting cell type deconvolution from bulk tissue proteomic data (<https://lmdomics.org/ProteoMixture/>).

RESULTS

Proteogenomic analysis of HGSOC cell admixture models

We generated cell admixtures consisting of defined percentages of tumor, stroma, and immune cell populations from either cultured cell line models or laser microdissection (LMD) enriched tissues. Specifically, *in vitro* cell line admixtures were generated using HGSOC tumor cells (OVCA3),²² fibroblast cells (to mimic stromal cells) established from an *in situ* ovarian cancer,^{23,24} and a model of T-cells (Jurkat).²⁵ LMD harvested tissue mixtures were generated using enriched populations of tumor, stroma, and immune-infiltrated stroma cells pooled from five women diagnosed with HGSOC (Figure 1; Tables S1, S2, S3, S4, and S5). Global proteome and transcriptome analyses of cell admixtures were performed using a quantitative, multiplexed proteomic approach employing tandem mass tags (TMTs) and liquid chromatography, high-resolution tandem mass spectrometry (LC-MS/MS), and RNA sequencing (RNA-seq), respectively. Global proteome and transcriptome analyses quantified $6,683 \pm 783$ proteins and $>20,000$ transcripts across all admixtures (Tables S6, S7, S8, and S9).

Proteome and transcriptome data from cell admixtures reflecting quartile percentages of tumor, stroma, and immune cell types were evaluated by principal component analysis (PCA) (Figure 2). PCA of the top 100 variably abundant proteins and transcripts by median absolute deviation (MAD) showed that admixtures that comprise one predominant cell type (tumor, stroma, or immune cells) form largely distinct clusters that transition to clusters of related sample compositions across cell type dilution series (Figure 2). PCA analysis of the top 100 variably abundant proteins explained $56.6 \pm 4.67\%$ and $41.45 \pm 6.15\%$ ($\leq 14.84\%$ CV, Figures 2A and 2B) of the variance between cell admixture conditions generated using cell lines or tissue samples, respectively. Additionally, our proteomic analysis included biological replicate conditions

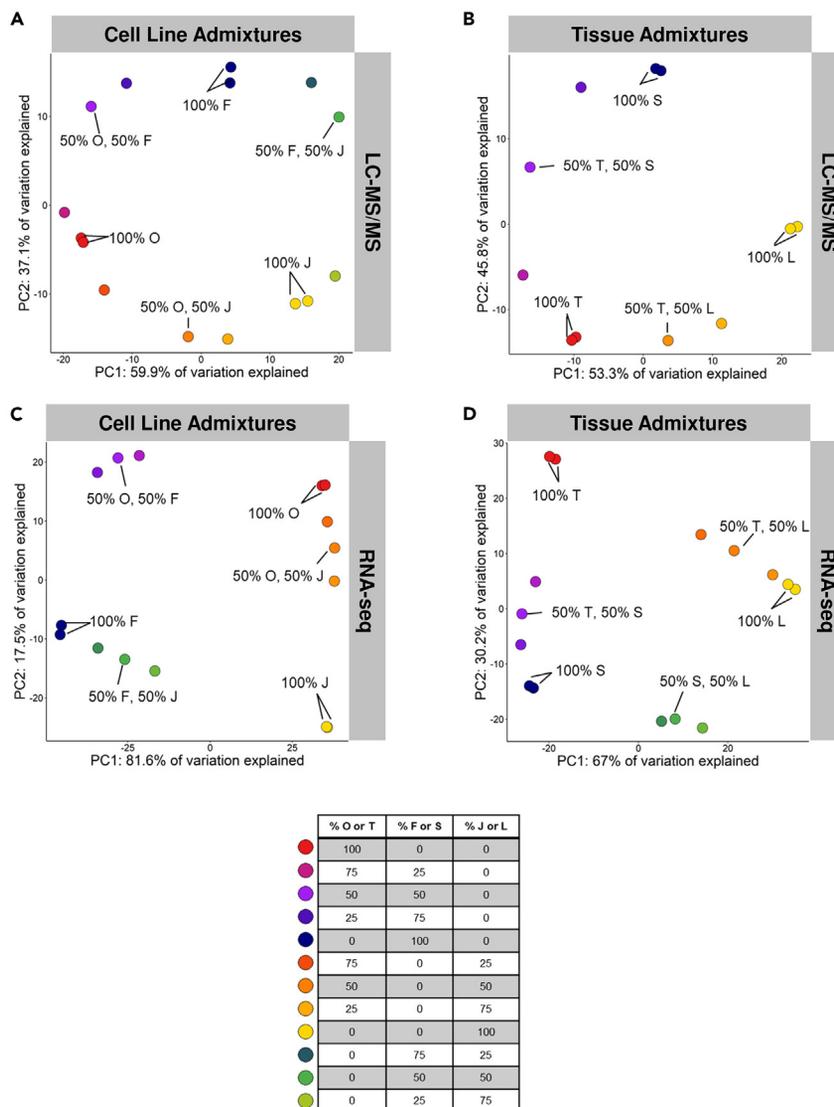


Figure 2. Principal component analysis (PCA) of cell admixture models

PCA analyses using the top 100 most variably abundant proteins (median absolute deviation) in (A) cell line admixtures and (B) LMD enriched cell populations from HGSOC tissues. PCA analyses using the top 100 most variably abundant transcripts in (C) cell line admixtures and (D) LMD enriched cell populations from HGSOC tissues. Abbreviations: OVCAR-3 (O), Fibroblast (F), Jurkat T cells (J), Tumor (T), Stroma (S), Lymphocytes (L).

of 100% tumor (OVCAR-3), stroma (fibroblast), and immune enriched (Jurkat) cell samples which are closely clustered in PCA analyses (Figures 2A and 2B), suggesting high reproducibility of our analytical workflows. Similar analyses of the top 100 variably abundant transcript explained $74.30 \pm 10.32\%$ and $23.85 \pm 8.98\%$ ($\leq 37.65\%$ CV, Figures 2C and 2D) of the variance between cell admixtures generated using cell lines or tissue samples. Correlation analysis of 6,097 proteins and transcripts co-quantified across non-admixed cell populations of interest identified that purified tumor (Spearman's $\rho = 0.48 \pm 0.005$) and immune enriched samples (Spearman's $\rho = 0.483 \pm 0.002$) exhibited the highest mean correlation of transcript and protein abundances, while enriched stroma samples were significantly lower (Spearman's $\rho = 0.32 \pm 0.04$, Figure S1A; Table S10). When we examined tumors admixed with other cell types, the correlation between transcript and protein abundances decreased as the percentage of stroma or lymphocytes increased (Figures S1B and S1C; Table S10).

Performance of cell admixture models using established deconvolution and molecular subtype classification tools

We analyzed proteome and transcriptome data with previously published cell deconvolution tools established using gene expression or transcript-level data, i.e., ESTIMATE¹⁴ and Consensus^{TME}¹⁸ (Tables S11, S12, S13, S14, S15, S16, S17, and S18). We first assessed the overlap between proteins and transcripts comparing 1) stroma and immune score signature gene candidates in ESTIMATE, and 2) fibroblast and immune score gene signature candidates in Consensus^{TME}. An average of 46% of gene signature constituents across these tools were

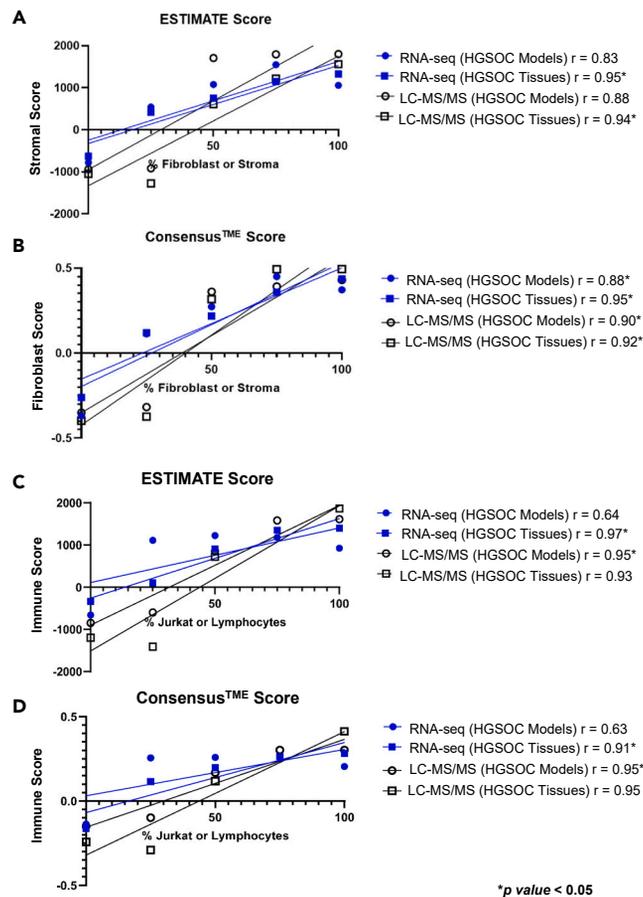


Figure 3. Comparison of Protein and Transcript-Level ESTIMATE and Consensus^{TME} Deconvolution Scores across Cell Admixture Conditions

Correlation analysis of transcriptome (RNA-seq) and proteome (LC-MS/MS) scores for quartile dilutions from cell lines and tissue admixtures corresponding to ESTIMATE stroma scores (A), Consensus^{TME} fibroblast scores (B), ESTIMATE immune scores (C), or Consensus^{TME} immune scores (D), (Tables S11, S12, S13, S14, S15, S16, S17, S18, and S19). (r) corresponds to Pearson correlation coefficients.

co-quantified in both global proteome and transcriptome datasets (Figure S2A). We then correlated the resulting ESTIMATE and Consensus^{TME} scores with proportional cell population admixtures for cellular subpopulations of interest (Figures 3A–3D). The ESTIMATE stromal score (Pearson’s $r > 0.9$, p value < 0.05) and the Consensus^{TME} fibroblast score (Pearson’s $r > 0.9$, p value < 0.05) were positively correlated with the percent stroma cell admixture from HGSOC tissues using proteomic or transcriptomic data (Figures 3A and 3B; Table S19). In cell line admixtures, higher correlations between the Consensus^{TME} fibroblast score and percent fibroblast (Pearson’s $r > 0.88$, p value < 0.05) were observed as compared to the correlation between ESTIMATE stromal score and percent fibroblast (Pearson’s $r > 0.82$, p value > 0.05) (Figure 3; Table S19). Immune scores from ESTIMATE or Consensus^{TME} scaled proportionally with increasing immune cell populations using co-quantified proteins (Pearson’s $r > 0.9$, p value < 0.05) from cell line admixtures or transcripts (Pearson’s $r > 0.9$, p value < 0.05) from HGSOC tissue admixtures as input (Figures 3C and 3D; Table S19). ESTIMATE tumor purity scores positively correlated with percent OVCAR-3 and tumor cell collections between proteomic and transcriptomic data (Pearson’s $r > 0.9$, p value < 0.05) (Figure S2B; Table S19). Global proteome data collected from LMD enriched HGSOC tissue admixtures was also analyzed using a proteome-based immune cell deconvolution tool, Decomprolute.²¹ Decomprolute scores for CD8⁺ T cells and B cells were highly correlated with the percent immune cell HGSOC tissue admixture conditions (Pearson’s $r > 0.94$, p value < 0.01) (Figure S3). Decomprolute also characterized LMD enriched lymphocyte cell populations as being comprised of predominantly CD8⁺ T cells (Decomprolute score = 0.42 ± 0.01) and B cells (Decomprolute score = 0.28 ± 0.02) (Table S20).

Prognostic molecular subtypes derived from gene signature analyses have been described in HGSOC patient tumors,¹⁰ from which the immunoreactive (IMR) and MES correlate with known immune or stroma cell populations, respectively.^{8,12} We explored the impact of cell admixture using transcriptome data from LMD enriched tissue samples on prognostic molecular subtype classifications using the consensusOV tool.¹⁰ Cell admixtures predominated by tumor cells were largely classified as differentiated (DIF) subtype. Admixtures with high proportions of stroma or immune cells were classified as MES or IMR subtypes, respectively (Figure S4). Tumor and lymphocyte cell admixture classifications transitioned from IMR to DIF subtype between 20 and 50% lymphocytes. Tumor and stroma admixture classifications

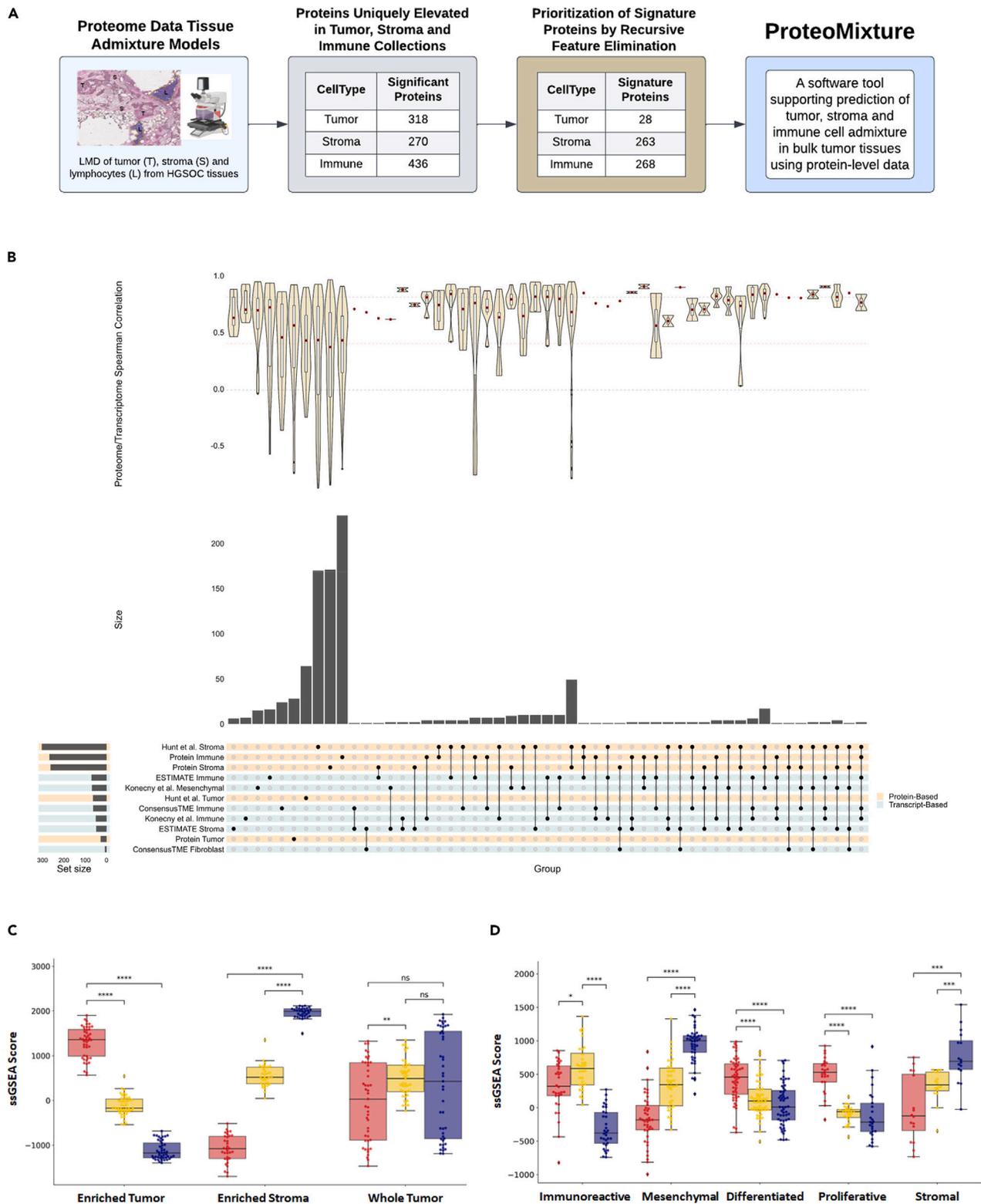


Figure 4. Optimization of protein signatures enabling the prediction of tumor, stroma, and immune cell admixture in proteomic data from enriched and bulk HGSOE tissues

(A) Strategy for the selection and prioritization of protein-level cell type signatures for the development of ProteoMixture.

Figure 4. Continued

(B) Integrated upset plot comparing optimized protein signatures with previously published cell type or molecular subtype protein and gene signatures with companion violin plots showing protein and transcript correlation distributions for overlapping features of interest; median correlation distributions denoted with red box.

(C) Assessment of protein signatures in proteomic data from enriched and bulk HGSOE tissues (Hunt et al.,⁷ n = 9 patients with HGSOE); *0.01 < p ≤ 0.05, **p ≤ 0.001, ****p ≤ 0.0001 from Mann-Whitney U Testing.

(D) Verification of protein signatures in proteome data from bulk HGSOE tissues (Zhang et al.,⁶ n = 169 patients with HGSOE); *0.01 < p ≤ 0.05, **p ≤ 0.001, ****p ≤ 0.0001 from Mann-Whitney Testing.

transitioned from DIF to MES between 50 and 75% stroma cells. Lastly, stroma and lymphocyte admixtures transitioned from MES to IMR between 50 and 75% lymphocytes (Figure S4).

Optimization of protein signatures enabling the prediction of tumor, stroma, and immune cell admixture in proteomic data from enriched and bulk HGSOE tissues

Cell type-associated protein signatures were generated from differential analysis of LMD enriched tumor, stroma, and immune cell-infiltrated populations (not admixed) from HGSOE tissues pooled from five patient tumors (Figure 4; Table S1). More than 550 cell type-associated proteins were identified (Figure 4; Table S21). Recursive feature elimination (RFE) optimally selected 28 proteins uniquely elevated in tumor cells, 263 proteins elevated in stroma, and 268 proteins elevated in immune-enriched tissue collections (Figure 4A; Table S21). We compared these cell type-associated protein signatures with known proteomic markers of LMD enriched HGSOE tumor cell populations,⁷ the gene signature candidates used in ESTIMATE and Consensus^{TME} deconvolution tools,^{14,18} and with molecular subtype classification signatures^{10,12,13} (Figure 4B). All 28 tumor signature proteins were unique relative to previously described feature sets, while 97 (37%) and 35 (13%) of the stroma and immune signature proteins, respectively, overlapped with marker transcripts from previously described signatures (Figure 4B; Table S22). Transcript-protein correlation of proteins co-identified in cell type transcript signatures from deconvolution tools (ESTIMATE, Consensus^{TME}),^{14,18} molecular subtype classification algorithms,^{12,13} and protein signatures from LMD HGSOE⁷ was significantly greater when compared to the transcript-protein correlation of all protein signatures (p value < 0.0001) (Figures 4B and S5).

We assessed the predictive accuracy of our RFE-generated signatures for classifying tumor, stroma, and immune cells from LMD enriched tumor and stroma,⁷ and from bulk tissue collections from tumors of >150 patients with HGSOE^{6,7} (Figures 4C and 4D). As anticipated, the LMD enriched tumor exhibited significantly elevated tumor scores, while the enriched stroma exhibited significantly elevated stroma scores (Figure 4C). Bulk tissue collections exhibited a more variable correlation with protein-derived tumor, stroma, and immune enriched signatures, consistent with the ≤56% tumor purities described for these samples.⁷ We also investigated the performance of our protein-level signatures in proteome data generated for bulk tissue collections recently described⁶ by CPTAC for an additional independent cohort of >150 HGSOE tumors (Figure 4D). Tumors classified as DIF or proliferative (PRO) molecular subtypes by Zhang et al. exhibited the highest enrichment with the tumor protein signature. Further, tumors classified as MES or stromal had the highest single-sample gene set enrichment analysis (ssGSEA) scores with our stroma protein signature, while tumors classified as IMR scored highest with the immune protein signature. PCA analysis of the most variably abundant proteins (top 25% standard deviation) in the proteomic data revealed clustering of patient tumors by molecular subtype classification (Figure S6). Overlaying the ssGSEA scores calculated for each patient tumor using our stroma protein signature (Figure S6A), immune protein signature (Figure S6B), and tumor protein signature (Figure S6C) reveals that correlation with higher stroma (MES, stromal), immune (IMR), or tumor (DIF, PRO) molecular subtypes, respectively, as anticipated. In addition, although our protein-level signatures have been optimized using HGSOE tissues, we generated an analysis of global proteome data in the ProteoMixture tool from a recently described cohort of tumors collected from n = 87 patients diagnosed with lung adenocarcinoma²⁶ with the goal of assessing how well ProteoMixture scores perform in other organ sites. We correlated ProteoMixture and immune and stroma scores calculated from companion transcript-level data using ESTIMATE for this cohort and observed a high correlation between stroma (Spearman's ρ = 0.699, p < 1E-4) and immune scores (Spearman's ρ = 0.796, p < 1E-4) (Figure S7) and significant, although lower correlations, between tumor purity estimates and tumor score (Spearman's ρ = 0.47, p < 1E-4) (data not shown). We integrate these signatures into a publicly available tool supporting cell type deconvolution from bulk tissue proteomic data available here: <https://lmdomics.org/ProteoMixture/>.

DISCUSSION

Our study assessed matched proteome and transcriptome data generated from *in vitro* and *in situ*-derived admixtures of common HGSOE TME cell types using established transcript-based tools for cell type deconvolution and prognostic molecular subtype classification. Unique protein signatures were developed to classify tumor, stroma, and immune cell populations within admixture HGSOE tissue samples using protein-level data. Admixtures of tumor, stroma, and immune cell populations exhibit unique proteomic and transcriptomic profiles, which drive sample clustering in unsupervised analysis. Our group showed previously that the enrichment of tumor and stroma cells can result in markedly different proteome and transcriptome profiles for a given sample.⁷ Our findings agree with these results and further show that immune cell admixture can contribute unique proteogenomic abundance alterations, impacting the molecular profile of a given sample. We also investigated the impact of cell admixture on the abundance of co-quantified proteins and transcripts, as we previously observed that enriched stroma exhibits lower correlation trends of these features compared with enriched tumor cell populations.⁷ We identified median correlation distributions for proteins and transcripts in enriched tumor (~0.48) and stromal cell (~0.32) populations in this study, consistent

with trends previously observed in HGSOC.^{7,27} We further identified that the correlation of protein and transcripts co-quantified in immune admixed conditions exhibit comparable median correlation trends as tumor cells (~0.48), suggesting higher coordination of protein and transcript abundances is reflective of not just tumor cells but immune cell populations as well.

We also assessed the overlap of co-quantified proteins and transcripts from *in vitro* or *in situ* sample collections with gene candidates from two previously described cell deconvolution tools, ESTIMATE¹⁴ and Consensus^{TME}¹⁸. Our efforts show that protein and transcript-level data exhibit comparable performance to predict immune and stroma/fibroblast cell admixture, including the estimation of tumor purity. Although less than 50% of gene signature candidates for these tools were quantified at the protein level, these features exhibited high quantitative correlation in transcript-level data (Figure S5), likely explaining this comparable performance. A limitation of our study is that immune-enriched tissue samples were generated by collecting and pooling immune admixed cell populations within the TME of several HGSOC patient tumors, precluding our ability to characterize unique subpopulations of immune cells beyond generic assessments of immune cell admixture through the assessment of companion “immune scores.” To this end, our analysis using Decomprolute²¹ revealed a high correlation of immune admixed conditions with CD8⁺ T cells and B cells from LMD enriched lymphocyte samples (Figure S3). Future efforts will focus on generating protein-level signatures that enable the classification of distinct immune cell populations involved in immune surveillance in HGSOC, such as CD8⁺ T cells, activated CD4⁺ T cells, and plasma cells.⁴ Other cell type deconvolution tools using proteomic data that have been developed includes scpDeconv,²⁸ which was developed based on single cell proteomic data, but we did not assess the performance of this tool due to the limited feature sets quantified in single cell proteomic data in comparison with the multiplexed global proteomics workflow we applied in our study.

Using transcript-level data, we demonstrate that cell admixture directly impacts the classification of prognostic molecular subtypes¹⁰ where we identify that admixture conditions consisting of >25% of a given cell population can impact the prediction of molecular subtype classification. Our findings resonate previous results showing that cellular heterogeneity directly impacts prognostic molecular subtype classifications in HGSOC.⁷ Molecular signatures generated using transcriptomic abundances from bulk tumor samples can lead to the misinterpretation of their corresponding protein-level abundances due to stromal influences and the limited concordance between protein and transcript abundances,²⁹ a finding we also observe in this study and in our prior work.⁷ These analyses underscore the importance of assessing tumor cellularity and sample heterogeneity when interpreting prognostic molecular subtype classifications for a given sample. Further, while mRNA undergoing active translation has been shown to correlate well with protein abundance,³⁰ post-transcriptional and post-translational, regulation could contribute to a poor correlation between transcripts and protein abundance.³¹

We prioritized protein signatures based on proteins uniquely elevated in tumor, stroma, and immune-enriched cell populations using RFE. We then regressed these features relative to cell admixture dilution conditions to select an optimized set of proteins demonstrating strong performance to estimate tumor, stroma, and immune cell admixture in protein-level data alone. Our tumor, stroma, and immune protein signatures were largely unique relative to gene signatures utilized in existing cellular deconvolution or molecular subtype classification tools.^{12,14,18} Further, most of our protein signature candidates exhibited a lower correlation with cognate transcript abundances relative to features identified in previously described gene signatures. Our protein signatures successfully classified enriched tumor or stroma HGSOC samples using proteome data alone⁷ and further correlated with molecular subtype classifications estimated from bulk HGSOC tissue collections,⁶ where tumors highly admixed with stromal cell populations (i.e., MES and stromal molecular subtypes) exhibit the highest stroma protein signature scores, tumors highly admixed with immune cell populations (i.e., IMR molecular subtype) exhibit high immune protein signature scores, and tumors exhibiting higher tumor purity (i.e., DIF and PRO molecular subtypes) exhibited high tumor protein signature scores. Our ProteoMixture tool also calculates scores using ssGSEA, similarly to the ESTIMATE¹⁴ and ConsensusTME¹⁸ tools. To make these signatures readily accessible to the research community, we have developed ProteoMixture (<https://lmdomics.org/ProteoMixture/>), a tool for predicting cell types in proteomic data from HGSOC tissues. Cell type-unique protein signatures were validated and optimized for HGSOC, a disease typified by marked heterogeneity within the TME.

Limitations of the study

Limitations of our study include the use of canonical proteoforms for the generation of protein signatures optimized for ProteoMixture. Proteoforms can reflect variations of protein modification states, such as due to post-translational modifications,³² and incorporation of proteoform-level abundances could further clarify the relationships between global protein and transcript-level abundances and will be the focus of future efforts. Additionally, our analysis of ProteoMixture performance in proteogenomic data generated from lung adenocarcinoma tissues demonstrates the proof-of-concept utility of our protein signatures to assess stroma and immune cell admixture in bulk proteomic data from other organ site malignancies. However, the lower correlation of tumor scores in lung adenocarcinoma tissues underscores a need to further refine tumor protein signatures to enable better characterization of tumor cell admixtures in organ sites beyond HGSOC, which will be the focus of future efforts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
- Lead contact

- Materials availability
- Data and code availability
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Cell lines and primary culture
 - Tissue specimens
- **METHOD DETAILS**
 - Laser microdissection (LMD)
 - Pressure cycling technology trypsin digestion of cells and laser microdissected tissues
 - RNA isolation from cells and laser microdissected tissues
 - Generation of *in vitro* cell admixtures for LC-MS/MS analysis
 - Generation of *in vitro* cell admixtures for RNA-seq analysis
 - LC-MS/MS analysis
 - Library preparation and HiSeq sequencing
 - Analysis of deconvolution tools and prognostic molecular subtypes
 - Cell type protein signature prioritization
 - Cell type protein signature analysis and validation
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2024.109198>.

ACKNOWLEDGMENTS

This study was supported in part by the U.S. Department of Defense Health Program to the Uniformed Services University for the Gynecologic Cancer Center of Excellence (HU0001-16-2-0006 and HU0001-16-2-00014). The authors would like to acknowledge Sakiyah TaQee, Persus Akowuah, Jeremy Loffredo, Glenn Gist, Salma Eltahir, Sasha Makohon-Moore, and Dr. Paulette Fauceglia for their contributions to histopathology assessment and sample preparation, and informatics data analysis. We would also like to acknowledge the patients and families who helped to make this work possible.

Disclaimer: The contents of this publication are the sole responsibility of the authors and do not reflect the views, opinions or policies of Uniformed Services University of the Health Sciences, the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc, the Department of Defense or the Departments of the Army, Navy or Airforce. Mention of trade names, commercial products or organizations do not imply endorsement by the U.S. Government.

AUTHOR CONTRIBUTIONS

N.W.B. and T.P.C. contributed to conception, experimental design, data analysis and article composition. P.N.T. performed experiments, data analysis and article composition. T.A. and J.S. contributed to bioinformatics and data analysis. J.O., V.O., F.S.P., M.E., A.C., B.L.H., K.A.C., D.M., A.L.H., T.L., P.-K.R.-K., M.D.W., N.T.P., K.M.D. and G.L.M. provided sample preparation, data generation and article review.

DECLARATION OF INTERESTS

T.P.C. is a Thermo Fisher Scientific, Inc. SAB member and receives research funding from AbbVie.

Received: August 9, 2023

Revised: November 4, 2023

Accepted: February 7, 2024

Published: February 12, 2024

REFERENCES

1. Bateman, N.W., and Conrads, T.P. (2018). Recent advances and opportunities in proteomic analyses of tumour heterogeneity. *J. Pathol.* 244, 628–637. <https://doi.org/10.1002/path.5036>.
2. Yang, Y., Yang, Y., Yang, J., Zhao, X., and Wei, X. (2020). Tumor Microenvironment in Ovarian Cancer: Function and Therapeutic Strategy. *Front. Cell Dev. Biol.* 8, 758. <https://doi.org/10.3389/fcell.2020.00758>.
3. Ovarian Tumor Tissue Analysis OTTA Consortium, Goode, E.L., Block, M.S., Kalli, K.R., Vierkant, R.A., Chen, W., Fogarty, Z.C., Gentry-Maharaj, A., Toloczko, A., Hein, A., et al. (2017). Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. *JAMA Oncol.* 3, e173290. <https://doi.org/10.1001/jamaoncol.2017.3290>.
4. Zhang, A.W., McPherson, A., Milne, K., Kroeger, D.R., Hamilton, P.T., Miranda, A., Funnell, T., Little, N., de Souza, C.P.E., Laan, S., et al. (2018). Interfaces of Malignant and Immunologic Clonal Dynamics in Ovarian Cancer. *Cell* 173, 1755–1769.e22. <https://doi.org/10.1016/j.cell.2018.03.073>.
5. Cancer Genome Atlas Research Network (2011). Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609–615. <https://doi.org/10.1038/nature10166>.
6. Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., et al. (2016). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian

- Cancer. *Cell* 166, 755–765. <https://doi.org/10.1016/j.cell.2016.05.069>.
- Hunt, A.L., Bateman, N.W., Barakat, W., Makohon-Moore, S., Hood, B.L., Conrads, K.A., Zhou, M., Calvert, V., Pierobon, M., Loffredo, J., et al. (2021). Extensive three-dimensional intratumor proteomic heterogeneity revealed by multiregion sampling in high-grade serous ovarian tumor specimens. *iScience* 24, 102757. <https://doi.org/10.1016/j.isci.2021.102757>.
 - Zhang, Q., Wang, C., and Cliby, W.A. (2019). Cancer-associated stroma significantly contributes to the mesenchymal subtype signature of serous ovarian cancer. *Gynecol. Oncol.* 152, 368–374. <https://doi.org/10.1016/j.ygyno.2018.11.014>.
 - Bentink, S., Haibe-Kains, B., Risch, T., Fan, J.B., Hirsch, M.S., Holton, K., Rubio, R., April, C., Chen, J., Wickham-Garcia, E., et al. (2012). Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PLoS One* 7, e30269. <https://doi.org/10.1371/journal.pone.0030269>.
 - Chen, G.M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D.M.A., Parmigiani, G., Birrer, M., Haibe-Kains, B., and Waldron, L. (2018). Consensus on Molecular Subtypes of High-Grade Serous Ovarian Carcinoma. *Clin. Cancer Res.* 24, 5037–5047. <https://doi.org/10.1158/1078-0432.CCR-18-0784>.
 - Helland, Å., Anglesio, M.S., George, J., Cowin, P.A., Johnstone, C.N., House, C.M., Sheppard, K.E., Etemadmoghadam, D., Melnyk, N., Rustgi, A.K., et al. (2011). Deregulation of MYCN, LIN28B and LET7 in a molecular subtype of aggressive high-grade serous ovarian cancers. *PLoS One* 6, e18064. <https://doi.org/10.1371/journal.pone.0018064>.
 - Konecny, G.E., Wang, C., Hamidi, H., Winterhoff, B., Kalli, K.R., Dering, J., Ginther, C., Chen, H.W., Dowdy, S., Cliby, W., et al. (2014). Prognostic and therapeutic relevance of molecular subtypes in high-grade serous ovarian cancer. *J. Natl. Cancer Inst.* 106, dju249. <https://doi.org/10.1093/jnci/dju249>.
 - Verhaak, R.G.W., Tamayo, P., Yang, J.Y., Hubbard, D., Zhang, H., Creighton, C.J., Fereday, S., Lawrence, M., Carter, S.L., Mermel, C.H., et al. (2013). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *J. Clin. Invest.* 123, 517–525. <https://doi.org/10.1172/JCI65833>.
 - Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612. <https://doi.org/10.1038/ncomms3612>.
 - Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220. <https://doi.org/10.1186/s13059-017-1349-1>.
 - Becht, E., Giraldo, N.A., Lacroix, L., Buttard, B., Elaroui, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W.H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 17, 218. <https://doi.org/10.1186/s13059-016-1070-5>.
 - Mavaddat, N., Michailidou, K., Dennis, J., Lush, M., Fachal, L., Lee, A., Tyrer, J.P., Chen, T.H., Wang, Q., Bolla, M.K., et al. (2019). Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am. J. Hum. Genet.* 104, 21–34. <https://doi.org/10.1016/j.ajhg.2018.11.002>.
 - Jiménez-Sánchez, A., Cast, O., and Miller, M.L. (2019). Comprehensive Benchmarking and Integration of Tumor Microenvironment Cell Estimation Methods. *Cancer Res.* 79, 6238–6246. <https://doi.org/10.1158/0008-5472.CAN-18-3560>.
 - Jarnuczak, A.F., Najgebauer, H., Barzine, M., Kundu, D.J., Ghavidel, F., Perez-Riverol, Y., Papatheodorou, I., Brazma, A., and Vizcaíno, J.A. (2021). An integrated landscape of protein expression in human cancer. *Sci. Data* 8, 115. <https://doi.org/10.1038/s41597-021-00890-2>.
 - Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. <https://doi.org/10.1038/nature10098>.
 - Feng, S., Calinawan, A., Pugliese, P., Wang, P., Ceccarelli, M., Petralia, F., Gosline, S.J., Sauro, H.M., Qian, W.J., and Wiley, H.S. (2023). Decomprolute: A benchmarking platform designed for multiomics-based tumor deconvolution. Preprint at bioRxiv. <https://doi.org/10.1101/2023.01.05.522902>.
 - Hamilton, T.C., Young, R.C., McKoy, W.M., Grotzinger, K.R., Green, J.A., Chu, E.W., Whang-Peng, J., Rogan, A.M., Green, W.R., and Ozols, R.F. (1983). Characterization of a human ovarian carcinoma cell line (NIH:OVCAR-3) with androgen and estrogen receptors. *Cancer Res.* 43, 5379–5389.
 - Siu, M.K.Y., Jiang, Y.X., Wang, J.J., Leung, T.H.Y., Han, C.Y., Tsang, B.K., Cheung, A.N.Y., Ngan, H.Y.S., and Chan, K.K.L. (2019). Hexokinase 2 Regulates Ovarian Cancer Cell Migration, Invasion and Stemness via FAK/ERK1/2/MMP9/NANOG/SOX9 Signaling Cascades. *Cancers* 11, 813. <https://doi.org/10.3390/cancers11060813>.
 - Siu, M.K.Y., Jiang, Y.X., Wang, J.J., Leung, T.H.Y., Ngu, S.F., Cheung, A.N.Y., Ngan, H.Y.S., and Chan, K.K.L. (2020). PDK1 promotes ovarian cancer metastasis by modulating tumor-mesothelial adhesion, invasion, and angiogenesis via $\alpha 5 \beta 1$ integrin and JNK/IL-8 signaling. *Oncogenesis* 9, 24. <https://doi.org/10.1038/s41389-020-0209-0>.
 - Gillis, S., and Watson, J. (1980). Biochemical and biological characterization of lymphocyte regulatory molecules. V. Identification of an interleukin 2-producing human leukemia T cell line. *J. Exp. Med.* 152, 1709–1719. <https://doi.org/10.1084/jem.152.6.1709>.
 - Soltis, A.R., Bateman, N.W., Liu, J., Nguyen, T., Franks, T.J., Zhang, X., Dalgard, C.L., Viollet, C., Somiari, S., Yan, C., et al. (2022). Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. *Cell Rep. Med.* 3, 100819. <https://doi.org/10.1016/j.xcrm.2022.100819>.
 - Hu, Y., Pan, J., Shah, P., Ao, M., Thomas, S.N., Liu, Y., Chen, L., Schnaubelt, M., Clark, D.J., Rodriguez, H., et al. (2020). Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep.* 33, 108276. <https://doi.org/10.1016/j.celrep.2020.108276>.
 - Wang, F., Yang, F., Huang, L., Song, J., Gasser, R.B., Aebersold, R., Wang, G., and yao, J. (2023). Deep Domain Adversarial Neural Network for the Deconvolution of Cell Type Mixtures in Tissue Proteome Profiling. Preprint at bioRxiv. <https://doi.org/10.1101/2022.11.25.517895>.
 - Fisher, N.C., Byrne, R.M., Leslie, H., Wood, C., Legrini, A., Cameron, A.J., Ahmaderaghi, B., Corry, S.M., Malla, S.B., Amirkhah, R., et al. (2022). Biological Misinterpretation of Transcriptional Signatures in Tumor Samples Can Unknowingly Undermine Mechanistic Understanding and Faithful Alignment with Preclinical Data. *Clin. Cancer Res.* 28, 4056–4069. <https://doi.org/10.1158/1078-0432.CCR-22-1102>.
 - Wang, T., Cui, Y., Jin, J., Guo, J., Wang, G., Yin, X., He, Q.Y., and Zhang, G. (2013). Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific. *Nucleic Acids Res.* 41, 4743–4754. <https://doi.org/10.1093/nar/gkt178>.
 - Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. <https://doi.org/10.1038/nrg3185>.
 - Smith, L.M., and Kelleher, N.L.; Consortium for Top Down Proteomics (2013). Proteoform: a single term describing protein complexity. *Nat. Methods* 10, 186–187. <https://doi.org/10.1038/nmeth.2369>.
 - Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. <https://doi.org/10.1093/nar/gkab1038>.
 - Mitchell, D., Hunt, A.L., Conrads, K.A., Hood, B.L., Makohon-Moore, S.C., Rojas, C., Maxwell, G.L., Bateman, N.W., and Conrads, T.P. (2022). Industrialized, Artificial Intelligence-guided Laser Microdissection for Microscaled Proteomic Analysis of the Tumor Microenvironment. *J. Vis. Exp.* <https://doi.org/10.3791/64171>.
 - Lee, S., Zhao, L., Rojas, C., Bateman, N.W., Yao, H., Lara, O.D., Celestino, J., Morgan, M.B., Nguyen, T.V., Conrads, K.A., et al. (2020). Molecular Analysis of Clinically Defined Subsets of High-Grade Serous Ovarian Cancer. *Cell Rep.* 31, 107502. <https://doi.org/10.1016/j.celrep.2020.03.066>.
 - Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925. <https://doi.org/10.1038/nmeth1113>.
 - Wickham, H. (2016). *Use R! 1 Online Resource (XVI, 260 Pages 232 Illustrations, 140 Illustrations in Color (Springer International Publishing))*.
 - Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32, 2847–2849. <https://doi.org/10.1093/bioinformatics/btw313>.
 - Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43, e47. <https://doi.org/10.1093/nar/gkv007>.
 - Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel,

- M., Prettenhofer, P., Weiss, R., Dubour, V., et al. (2011). Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* *12*, 2825–2830.
41. Lex, A., Gehlenborg, N., Strobel, H., Vuillemot, R., and Pfister, H. (2014). UpSet: Visualization of Intersecting Sets. *IEEE Trans. Vis. Comput. Graph.* *20*, 1983–1992. <https://doi.org/10.1109/TVCG.2014.2346248>.
42. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* *462*, 108–112. <https://doi.org/10.1038/nature08460>.
43. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* *14*, 7. <https://doi.org/10.1186/1471-2105-14-7>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
High grade serous ovarian cancer omental metastasis	IFMC/WHIRC Biobank	343WJ
High grade serous ovarian cancer omental metastasis	IFMC/WHIRC Biobank	343VT
High grade serous ovarian cancer adnexal metastasis	IFMC/WHIRC Biobank	343WN
High grade serous ovarian cancer omental metastasis	IFMC/WHIRC Biobank	343WB
High grade serous ovarian cancer omental metastasis	IFMC/WHIRC Biobank	343WH
Chemicals, peptides, and recombinant proteins		
Mayer's Hematoxylin Solution	Sigma Aldrich	Cat# MHS32
Eosin Y Solution Aqueous	Sigma Aldrich	Cat# HT110216
Buffer RLT	QIAGEN Sciences, LLC	Cat# 79216
β-mercaptoethanol	Thermo Fisher Scientific, Inc.	Cat# M3148
Triethylammonium bicarbonate (TEAB)	Sigma	Cat# T7408
Acetonitrile	Thermo Fisher Scientific, Inc.	Cat# A955-4
TRIzol	Thermo Fisher Scientific, Inc.	Cat# 15596018
NH ₄ HCO ₃	Thermo Fisher Scientific, Inc.	Cat# A643-500
Formic acid	Thermo Fisher Scientific, Inc.	Cat# 9517C
Critical commercial assays		
Pierce BCA Protein Assay Kit	Thermo Fisher Scientific, Inc.	Cat# 23225
TMTpro 16plex Label Reagent Set	Thermo Fisher Scientific, Inc.	Cat# A44520
TMTpro 18plex Label Reagent Set	Thermo Fisher Scientific, Inc.	
RNeasy Micro Kit	Qiagen	Cat# 74004
Qubit RNA HS Assay Kit	Thermo Fisher Scientific, Inc.	Cat# Q32852
High Sensitivity RNA Screentape	Agilent	Cat# 5067-5579
High Sensitivity RNA Screentape Ladder	Agilent	Cat# 5067-5581
High Sensitivity RNA Screentape Sample Buffer	Agilent	Cat# 5067-5580
Deposited data		
Raw LC-MS/MS data	This paper	ProteomeXchange Consortium, PRIDE: PXD044157
RNA-sequencing data	This paper	European Nucleotide Archive, ERP156652
Experimental models: Cell lines		
Cell Line	ATCC	NIH:OVCAR-3
Cell Line	ATCC	Jurkat, Clone E6-1
Primary Cell Line	Vitro Biopharma, Inc.	Human ovarian serous cancer associated fibroblasts
Software and algorithms		
HALO	Indica Labs	https://indicalab.com/
Mascot	Matrix Science	https://www.matrixscience.com/
Proteome Discoverer	Thermo Fisher Scientific	https://www.thermofisher.com/ us/en/home.html
Swiss-Prot	UniProt	http://www.uniprot.org/
R versions 3.6.0 and 4.2.2	CRAN	https://cran.r-project.org/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LIMMA version 3.42.2	Bioconductor	https://bioconductor.org/packages/release/bioc/html/limma.html
ggplot2 version 3.4.1	CRAN	https://cran.r-project.org/web/packages/ggplot2/index.html
consensusOV versions 1.18.0 and 1.20.0	Bioconductor	http://bioconductor.jp/packages/3.10/bioc/html/consensusOV.html
ComplexHeatmap version 2.14.0	Bioconductor	https://www.bioconductor.org/packages/release/bioc/html/ComplexHeatmap.html
GraphPad version 8.3.0	GraphPad	https://www.graphpad.com/
HiSeq Control Software (HCS)	Illumina	https://support.illumina.com/sequencing/sequencing_instruments/hiseq_2500/downloads.html
Bcl2fastq 2.17	Illumina	https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software/downloads.html
Trimomatic v.0.36	GitHub	https://github.com/timflutre/trimomatic
STAR aligner v.2.5.2b	GitHub	https://github.com/alexdobin/STAR
Subread package v.1.5.2	GitHub	https://github.com/ShiLab-Bioinformatics/subread
DESeq2 v.1.24.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ESTIMATE	R-Forge	https://r-forge.r-project.org/projects/estimate/
Consensus ^{TME}	GitHub	https://github.com/cansysbio/ConsensusTME
Python version 3.9.16	Python	https://www.python.org/
Scikit-learn package version 1.2.1	GitHub	https://github.com/scikit-learn/scikit-learn
ComplexUpset version 1.3.5	CRAN	https://cran.r-project.org/web/packages/ComplexUpset/index.html
GSVA library version 1.34.0	Bioconductor	https://bioconductor.org/packages/release/bioc/html/GSVA.html
Seaborn version 0.11.2	Seaborn	https://seaborn.pydata.org/
Matplotlib version 3.7.1	Matplotlib	https://matplotlib.org/
Scipy version 1.10.0	Scipy	https://scipy.org/

Other

PEN Membrane Glass Slides	Leica Microsystems	Cat# 11532918
96 MicroTubes in bulk (no caps)	Pressure Biosciences, Inc.	Cat# MT-96
96 MicroCaps (150uL) in bulk	Pressure Biosciences, Inc.	Cat# MC150-96
96 MicroPestles in bulk	Pressure Biosciences, Inc.	Cat# MP-96
RPMI-1640 Medium	ATCC	Cat# 30-2001
Fetal Bovine Serum (FBS)	ATCC	Cat# 30-2020
Penicillin (100 IU/mL)-Streptomycin (100 µg/mL)	ATCC	Cat# 30-2300
VitroPlus III Low Serum Complete Medium	Vitro Biopharma, Inc.	Cat# PC00B1
Non-treated T-75 flasks	Corning	Cat# 431464U
Dulbecco's Phosphate Buffered Saline (PBS)	ATCC	Cat# 30-2200
Insulin	Sigma	Cat# 19278
Acclaim™ PepMap™ 100 Å, C-18, 20 mm length, nanoViper Trap column	Thermo Fisher Scientific, Inc.	Cat# ES903
Acclaim™ PepMap™ RSLC C-18, 2 µm, 100 Å, 75 µm × 500 mm, nanoViper	Thermo Fisher Scientific, Inc.	Cat# 164536

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Dr. Nicholas W. Bateman (batemann@whirc.org).

Materials availability

This study did not generate any unique reagents.

Data and code availability

The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE³³ partner repository with the dataset identifier PXD044157.

The transcriptome data have been deposited to the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) repository with the dataset project identifier PRJEB71866 and study identifier ERP156652.

The protein signature assessment tool is available at: <https://lmdomics.org/ProteoMixture/> and supporting code is available: <https://github.com/GYNCOE/Teng.et.al.2024>.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines and primary culture

Human cell lines OVCAR-3 (NIH:OVCAR-3, female) and Jurkat T cells (Clone E6-1, male) were purchased from ATCC (Manassas, VA, USA). STR analysis were performed for OVCAR-3 and Jurkat T cells by ATCC. Human ovarian serous cancer associated fibroblasts (primary cells, female) were purchased from Vitro Biopharma, Inc. (Golden, CO, USA). OVCAR-3 was cultured in RPMI-1640 medium (ATCC, Manassas, VA, USA) supplemented with 20% fetal bovine serum (FBS) (ATCC), Penicillin (100 IU/mL)-Streptomycin (100 µg/mL) (ATCC), and 0.01 mg/mL insulin (Sigma-Aldrich, St. Louis, MO, USA). Jurkat T cells were cultured in RPMI supplemented with 20% FBS (ATCC), Penicillin (100 IU/mL)-Streptomycin (100 µg/mL) (ATCC) and maintained at $1 \times 10^5 - 1 \times 10^6$ cells/mL density. Ovarian serous cancer associated fibroblasts were cultured in VitroPlus III Low Serum Complete Medium (Vitro Biopharma, Inc.). All cells were maintained at 37°C and 5% CO₂. OVCAR-3 and ovarian serous cancer associated fibroblasts were grown in a monolayer on 10 cm² tissue culture dishes (Fisher Scientific, Inc., Hampton, NH, USA). At 70% confluency, with maintained healthy morphology, cells were washed twice with Dulbecco's Phosphate Buffered Saline (PBS) (ATCC) and collected by scraping. Jurkat T cells were grown in suspension in non-treated T-75 flasks (Corning, Glendale, AZ, USA) and harvested by centrifugation followed by two PBS washes. Replicate plates of cells were prepared for cell counting. Cell counts were obtained from TC20 Automated Cell Counter (Bio-Rad, Hercules, CA, USA).

Tissue specimens

Fresh-frozen (FF) tumor specimens (stage IIIC) were obtained from primary (adnexal mass) or metastasis (omentum) of five female patients with a primary HGSOE disease site originating at the ovary or fallopian tube (Table S1). Chemotherapy-naïve and neoadjuvant chemotherapy (NACT) individuals between the age of 40 - 76 years at the time of diagnosis were included (Table S1). The study protocol was approved under a less than minimal-risk WCG IRB-approved protocol #20122048 with waivers for consent and HIPAA authorization and evaluated under protocol #14-1679 with an exempt determination by the WCG IRB in accordance with the use of de-identified data under US Federal regulation 45 CFR 46.102(f). All experimental protocols involving human data in this study were in accordance with the Declaration of Helsinki and informed consent was obtained from all patients.

METHOD DETAILS

Laser microdissection (LMD)

FF tissue sections (10 µm thickness) were cut by cryostat and placed onto PEN membrane slides (Leica Microsystems, Deer Park, IL, USA) and processed as previously described.⁷ After staining with aqueous hematoxylin and eosin, cell type annotation and counting were performed using the HALO image analysis software (Indica Labs, Albuquerque, NM, USA). Regions of tissue annotated using HALO were exported for LMD (LMD7, Leica Microsystems) as previously described.³⁴ LMD harvested tissue for protein digestion was collected in MicroTubes (Pressure Biosciences, Inc., South Easton, MA, USA) containing 20 µL of 100 mM triethylammonium bicarbonate (TEAB, pH 8.0)/10% acetonitrile, capped and stored at -80°C until digestion. Tissue for RNA isolation was collected in 300 µL of Buffer RLT with 10% β-mercaptoethanol (QIAGEN Sciences, LLC, Germantown, MD, USA) and stored at -80°C until isolation.

Pressure cycling technology trypsin digestion of cells and laser microdissected tissues

Cells were transferred to MicroTubes for a final volume of 20 µL of 100 mM TEAB (pH 8.0)/10% acetonitrile. Cell and tissue samples underwent pressure-assisted trypsin digestion employing a barocycler (2320EXT, Pressure BioSciences, Inc.) and a heat-stable form of trypsin (SMART Trypsin, Thermo Fisher Scientific, Inc., Waltham, MA, USA). Peptide digest concentrations were determined using the bicinchoninic acid assay (BCA; Thermo Fisher Scientific, Inc.). Peptides (10 µg for cells and 4 µg for tissue samples) were labeled with isobaric tandem mass tag (TMT)

reagents according to the manufacturer's instructions (TMTpro, Thermo Fisher Scientific, Inc.). Sample multiplexes were reversed-phase fractionated (basic pH) on a 1260 Infinity II offline liquid chromatography system (Agilent Technologies, Inc., Santa Clara, CA, USA) into 96 fractions using a linear gradient of acetonitrile ($0.69\% \text{ min}^{-1}$) followed by concatenation into 36 pooled fractions. Each pooled fraction was resuspended in 25 mM NH_4HCO_3 and analyzed by LC-MS/MS.

RNA isolation from cells and laser microdissected tissues

Total RNA was isolated using TRIzol (Thermo Fisher Scientific, Inc.) according to the manufacturer's instructions and cleaned up with DNase treatment using the RNeasy Mini Kit (QIAGEN Sciences, LLC) for cell samples. RNA from tissue samples were purified using the RNeasy Micro Kit (QIAGEN Sciences, LLC.) per the Purification of Total RNA from Microdissected Cryosections Protocol including on-column DNase digestion. Initial RNA concentrations and 260/280 absorbance ratios were determined using a Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Inc.). Final RNA concentrations were determined using 1 μL of sample and the Qubit RNA HS kit (Thermo Fisher Scientific, Inc.). RNA integrity numbers (RINe) were calculated using the High Sensitivity RNA ScreenTape and Buffers on a TapeStation 4200 (Agilent Technologies, Inc.).

Generation of *in vitro* cell admixtures for LC-MS/MS analysis

Pooled admixtures ($n=13$) were generated by combining peptide digests (10 μg total) from OVCAR-3, tumor associated fibroblasts, and Jurkat T cells at pre-defined compositional ratios representing quartile dilutions of cell populations of interest (Table S2). Similarly, pooled admixtures ($n=17$; 4 μg total) were generated by combining peptide digests from LMD enriched populations of tumor, stroma, and lymphocytes (Table S3). Cell admixture conditions generated from LMD enriched tissues reflected mainly quartile dilution series, but also included select dilution combining <25% lymphocytes and tumor populations due to limited yields from lymphocyte admixed collections. Digests from multiple patients were combined to make cell type-associated pooled samples. Duplicate digest samples (from cells or tissues) were prepared for models containing a single cell type. The percentage of each cell type in a cell admixture was calculated based on μg of digest.

Generation of *in vitro* cell admixtures for RNA-seq analysis

RNA pooled admixtures ($n=16$) were generated by combining RNA samples (20 ng total) from OVCAR-3, tumor associated fibroblasts, and Jurkat T cell at pre-defined compositional ratios (Table S4). Similarly, cell type-associated RNA pooled admixtures ($n=23$) were generated by combining isolated RNA (20 ng total) from LMD enriched populations of tumor, stroma, and lymphocytes (Table S5). Cell admixture conditions generated from LMD enriched tissues reflected mainly quartile dilution series, but also included select dilution combining <25% lymphocytes and tumor populations due to limited yields from lymphocyte admixed collections. Isolated RNA from multiple patients were combined to make cell type-associated pools. Duplicate RNA samples (from cells or tissues) were prepared for models containing a single cell type. The percentage of each cell type in a cell admixture was calculated from ng of RNA.

LC-MS/MS analysis

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analyses of TMTpro16 and TMTpro18 multiplexes were performed on a nanoflow high-performance LC system (EASY-nLC 1200, Thermo Fisher Scientific, Inc.) coupled online with an Orbitrap mass spectrometer (Q Exactive HF-X, Thermo Fisher Scientific, Inc.). Samples were loaded on a reversed-phase trap column (Acclaim™ PepMap™ 100 Å, C-18, 20 mm length, nanoViper Trap column, Thermo Fisher Scientific, Inc.) and eluted on a heated (50°C) reversed-phase analytical column (Acclaim™ PepMap™ RSLC C-18, 2 μm , 100 Å, 75 $\mu\text{m} \times 500 \text{ mm}$, nanoViper, Thermo Fisher Scientific, Inc.) by developing a linear gradient from 2% mobile phase A (2% acetonitrile, 0.1% formic acid) to 32% mobile phase B (95% acetonitrile, 0.1% formic acid) over 120 min at a constant flow rate of 250 nL/min. Full scan mass spectra (MS) were acquired using a mass range of m/z 400-1600, followed by selection of the top 12 most intense molecular ions in each MS scan for high-energy collisional dissociation (HCD). Instrument parameters were as follows: Full MS: AGC, 3×10^6 ; resolution, 60 k; S-Lens RF, 40%; max IT, 45 ms; MS2: AGC, 10^5 ; resolution, 45 k; max IT, 95 ms; quadrupole isolation, 1.0 m/z ; isolation offset, 0.2 m/z ; NCE, 30; fixed first mass, 100; intensity threshold, 2×10^5 ; charge state, 2–4; dynamic exclusion, 20 s, TMT optimization. Global protein-level abundances were generated from peptide spectral matches identified by searching .raw data files against a publicly available, non-redundant human proteome database (<http://www.uniprot.org/>, SwissProt, Homo sapiens, downloaded 12-01-2017 for cell line and 10-29-2021 for tissue) using Mascot (Matrix Science, v2.6.0), Proteome Discoverer (v2.2.0.388, Thermo Fisher Scientific, Inc.), and in-house tools using identical parameters as previously described.³⁵ Peptide spectral matches passing a false-discovery rate (FDR) expectation < 1.0% as determined by the Percolator³⁶ module of Proteome Discoverer were prioritized for downstream analysis. Quan correction was applied to all reagent ion abundances using TMTpro16 reagent lot UL296296 or TMTpro18 reagent lots WK334339 and WJ338613.

Library preparation and HiSeq sequencing

RNA library preparation and sequencing were conducted at GENEWIZ, LLC. (South Plainfield, NJ, USA). SMART-Seq v4 Ultra Low Input Kit for Sequencing was used for full-length cDNA synthesis and amplification (Clontech, Mountain View, CA, USA). Nextera XT library (Illumina, Inc., San Diego, CA, USA) was used for sequencing library preparation. Briefly, cDNA was fragmented, and adaptor was added using Transposase, followed by limited-cycle PCR to enrich and add index to the cDNA fragments. The final library was assessed with TapeStation (Agilent Technologies, Inc.). The sequencing libraries were multiplexed and clustered on a flowcell. After clustering, the flowcell was loaded on the Illumina

HiSeq instrument according to manufacturer's instructions. The samples were sequenced using a 2x150 Paired End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into fastq files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification. After investigating the quality of the raw data, sequence reads were trimmed to remove possible adapter sequences and nucleotides with poor quality using Trimmomatic v.0.36. The trimmed reads were mapped to the *Sus scrofa* reference genome available on ENSEMBL using the STAR aligner v.2.5.2b. The STAR aligner is a splice aligner that detects splice junctions and incorporates them to help align the entire read sequences. BAM files were generated as a result of this step. Unique gene hit counts were calculated by using feature counts from the Subread package v.1.5.2. Only unique reads that fell within exon regions were counted. GRCh38/hg38 was used as the human reference genome. Count level data was VST normalized by DESeq2 (version 1.24.0).

Analysis of deconvolution tools and prognostic molecular subtypes

Proteomic data from admixed samples comprised of 33.3% OVCAR-3/tumor, 33.3% fibroblast/stroma, and 33.3% Jurkat/lymphocytes were used to normalize protein abundances similarly to the pooled standard previously described.³⁵ Global proteome data was visualized by principal component analyses (PCA) using the top 100 most variable proteins or transcripts by mean absolute deviation (MAD) using ggplot2 (version 3.4.1) in R (version 4.2.2).³⁷ Gene names in the proteomic and RNA-seq datasets were harmonized with any gene synonyms used in the Consensus^{TME} and ESTIMATE tools prior to their utilization. RNA-seq data was filtered by excluding features without HUGO Gene Nomenclature Committee (HGNC) gene symbols, gene symbols that began with "LOC", or zero variant entries. RNA-seq data was gene-wise, z-score scaled and subsetted to genes co-quantified in proteomics for use with ESTIMATE and Consensus^{TME}. Pearson correlation analysis was performed between ESTIMATE or Consensus^{TME} scores and percent cell type (quartile cell percentages). Transcriptomic data from the admixed samples comprised of 33.3% OVCAR-3/tumor, 33.3% fibroblast/stroma, and 33.3% Jurkat/lymphocytes were excluded from the correlation analysis (outliers based on ESTIMATE and Consensus^{TME} scores). Spearman correlation was performed on transcriptomic data and proteomic data from tissue admixture models and heatmap of quartile percentage tissue models were generated using ComplexHeatmap (version 2.14.0).³⁸ RNA-seq data (quartile cell percentages) from HGSOc tissue cell admixture was used for subtype classification with consensusOV (version 1.20.0).¹⁰ Proteomic data from tissue admixture models were assessed by Decomprolute.²¹

Cell type protein signature prioritization

Differential analysis was performed using limma³⁹ to identify proteins uniquely elevated in LMD enriched tumor, stroma or immune tissues. Proteins passed the ≥ 1.5 fold-change (FC) and p value cutoffs (detailed in [Quantification and statistical analysis, STAR Methods](#)) were selected for the next step. Protein abundance data ([Table S6](#)) was then partitioned into three sets of training (n=10) and testing (n=6) samples stratified by tissue composition using the 'train_test_split' function in scikit-learn (version 1.2.1). Recursive feature elimination (RFE) on the respective training data was used to select each protein signature that holds the most predictive power for support vector regression models (SVR) with linear kernels. Each SVR model's target was the percent composition of the respective tissue type. Grid search was utilized to find optimal hyperparameters before performing RFE for each signature. The ranks provided by RFE for each protein signature were then used to re-train SVR models and assess their performance on their respective testing data. The SVM model with at least 15 features and had the lowest mean squared error (MSE) to the test data was chosen, and its features were determined as the protein cell type signatures. Models were trained in python (version 3.9.16) using the scikit-learn package (version 1.2.1).⁴⁰

Cell type protein signature analysis and validation

Upset plots were generated using ComplexUpset (version 1.3.5).⁴¹ Transcript-protein Spearman correlations were calculated from matched tissue admixture model data for cell type-associated protein signatures. Assignment of admixed samples with prognostic HGSOc molecular subtypes¹² was performed using consensusOV (version 1.18.0),¹⁰ where a total of 575 unique Entrez gene IDs (kindly provided by the author) were identified from 635 selected probe sets. Correlation plots were generated with seaborn version 0.11.2. Protein signatures were validated using proteomic abundance data from enriched and bulk HGSOc (n = 9)⁷ and bulk HGSOc tissues (n = 169) by performing ssGSEA with the newly derived gene sets.^{42,43} ssGSEA was performed in R (version 3.6.0) using the GSVA library (version 1.34.0).⁴³ The top 25% most variably abundant proteins by standard deviation from bulk HGSOc tissues (n = 169)⁵ were visualized by principal component analyses (PCA) and overlaid with stroma, immune, or tumor ssGSEA scores. Global proteomics data for lung adenocarcinomas and stroma and immune scores calculated from companion RNAseq data using ESTIMATE was downloaded from Soltis AR et al.²⁶ Correlation of ProteoMixture scores calculated using global proteome data with ESTIMATE scores for lung adenocarcinoma samples were generated using MedCalc (version 20.109). Package scikit-learn (version 1.2.1) was used for performing PCA, while packages seaborn (version 0.11.2), matplotlib (version 3.7.1), and scipy (version 1.10.0) were used for plotting.

QUANTIFICATION AND STATISTICAL ANALYSIS

The significance of the differential proteins from the limma analysis³⁹ was determined using the following p value cutoffs: adjusted p value < 0.05 for proteins elevated in the tumor sample when compared to stroma or immune samples; adjusted p value < 0.05 for proteins elevated in the immune sample compared to tumor or stroma samples; p value < 0.05 for proteins elevated in stroma sample compared to tumor or immune samples ([Figure 4A](#)). Pearson correlation between ESTIMATE or Consensus^{TME} scores and percent cell type was

determined using GraphPad (version 8.3.0) where the significance was determined by p value < 0.05 (Figures 3 and S2B). Pearson correlation between Decomprolute scores and percent immune cell type was performed using GraphPad (version 8.3.0) where the significance was determined by p value < 0.05 (Figure S3). Mann-Whitney Wilcoxon (M.W.W.) test was used to assess the statistical significance of ssGSEA scores between sample groups and molecular subtype groups (Figure 4, p values detailed in Figure 4 legend). Spearman's correlation between protein and transcript and p value were performed using GraphPad (Figure S1). M.W.W. test was used to assess the significance of Spearman's correlation between signatures (Figures S5A and S5B).