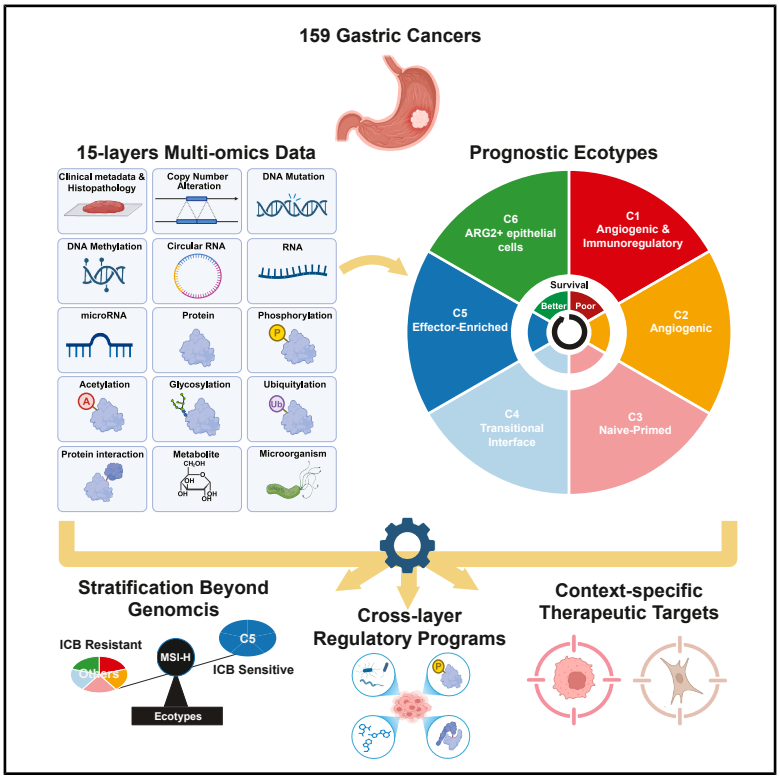
Cell Reports
Medicine

Article

A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights

Graphical abstract



Highlights

- Fifteen-omics atlas provides a foundational resource for gastric cancer research
- Tumor ecotypes capture TME diversity, prognosis, and immunotherapy response
- Ecotype functional programs defined by proteogenome, metabolome, and microbiome
- Multi-omics reveal therapeutic targets specific to ecotypes and cell types

Authors

Yuefan Wang, Lindsey K. Olsen, Fenglong Jiao, ..., Hui Zhang, Bing Zhang, Clinical Proteomic Tumor Analysis Consortium

Correspondence

huizhang@jhu.edu (H.Z.),
bing.zhang@bcm.edu (B.Z.)

In brief

Wang et al. present a 15-layer multi-omics resource that integrates tumor ecotype classification with deep molecular profiling to define functional gastric cancer ecosystems. By linking microenvironment architecture to regulatory programs and context-specific therapeutic vulnerabilities, the study refines tumor stratification and advances precision oncology.

Wang et al., 2026, Cell Reports Medicine 7, 102756
May 19, 2026
https://doi.org/10.1016/j.xcrm.2026.102756

CellPress

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights

Yuefan Wang,[1,29] Lindsey K. Olsen,[2,3,29] Fenglong Jiao,[4,30] Chenwei Wang,[2,3,30] Kevin X. Jiang,[5,6,30] Yongchao Dou,[2,3,30] Yingwei Hu,[1,30] Liyuan Jiao,[1,30] Wenrong Chen,[2,3,30] John M. Elizarraras,[2,3,30] Pratik Khare,[7] Nengneng Yu,[8] Huili Zhu,[9] Lijun Chen,[1] Tung-Shing M. Lih,[1] Pınar O. Eser,[6] Fernanda Martins Rodrigues,[10,11,12] Zhiao Shi,[2,3] Cissy Zhang,[7] Clinton Yu,[4] David I. Heiman,[6] Yuxing Liao,[2,3] Paul W. Shafer,[2,3] Seunghyuk Choi,[2,3] Jong Min Choi,[2,3] Sara R. Savage,[2,3] Eric J. Jaehnig,[2,3] Jonathan T. Lei,[2,3] Yanling Sun,[2,3] Chien-Wei Peng,[10,11,12] Zhenyu Sun,[1] Paul Morenkov,[4] Kevin Zhang,[13] Yifat Geffen,[6] Julian Hess,[6] Chandan Kumar-Sinha,[14,15] D.R. Mani,[16] Li Ding,[10,11,12,17] Gad Getz,[6,18,19] Qing Kay Li,[1] Gilbert S. Omenn,[20] Anne Le,[7] Galen Hostetter,[21] Chelsea J. Newton,[21] Shuang Cai,[22] Karen A. Ketchum,[22] Ana I. Robles,[23] Mehdi Mesri,[23] Parham Minoo,[24] M. Constanza Camargo,[25] Eunkyung An,[23] Ralph H. Hruban,[1] Zaoxing Liu,[8] Mathangi Thiagarajan,[26] Anders B. Dohlman,[5,6,28] Ramon U. Jin,[10,17,27,28] Lan Huang,[4,28] Daniel W. Chan,[1,28] Hui Zhang,[1,28,*] Bing Zhang,[2,3,28,31,*] and Clinical Proteomic Tumor Analysis Consortium

[1]Department of Pathology, the Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[2]Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, USA
[3]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA
[4]Department of Physiology & Biophysics, University of California, Irvine, Irvine, CA, USA
[5]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[6]Cancer Program, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA
[7]Gigantest, Baltimore, MD, USA
[8]Department of Computer Science, University of Maryland, College Park, MD, USA
[9]Department of Gastrointestinal Medical Oncology, Division of Cancer Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[10]Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA
[11]McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO, USA
[12]Department of Genetics, Washington University in St. Louis, St. Louis, MO, USA
[13]Department of Neuroscience, Krieger School of Arts and Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[14]Department of Pathology, University of Michigan, Ann Arbor, MI, USA
[15]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA
[16]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[17]Siteman Cancer Center, Washington University in St. Louis, St. Louis, MO, USA
[18]Department of Pathology, Harvard Medical School, Boston, MA, USA
[19]Krantz Family Center for Cancer Research and Department of Pathology, Massachusetts General Hospital, Boston, MA, USA
[20]Departments of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics, and Environmental Health, University of Michigan, Ann Arbor, MI, USA
[21]Van Andel Research Institute, Grand Rapids, MI, USA
[22]ICF, 530 Gaither Road Suite 500, Rockville, MD, USA
[23]Office of Cancer Clinical Proteomics Research, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Rockville, MD, USA
[24]Department of Pathology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada
[25]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA
[26]Frederick National Laboratory for Cancer Research, Frederick, MD, USA
[27]Division of Oncology and Gastroenterology, Washington University School of Medicine, St. Louis, MO, USA
[28]Senior author
[29]These authors contributed equally
[30]These authors contributed equally
[31]Lead contact
*Correspondence: huizhang@jhu.edu (H.Z.), bing.zhang@bcm.edu (B.Z.)
https://doi.org/10.1016/j.xcrm.2026.102756

SUMMARY

Gastric cancer is marked by profound molecular and microenvironmental heterogeneity that limits therapeutic progress. Here, we present a 15-layer multi-omics atlas that integrates genomics, epigenomics, transcriptomics, proteomics, multiple post-translational modifications (PTMs), protein-protein interactions, metabolomics, and microbiome profiles from 159 primary gastric adenocarcinomas and 30 matched normal adjacent

tissues. Using cell-state deconvolution, we define tumor ecotypes that refine genomic and histological subtypes by capturing distinct tumor microenvironment architectures linked to clinical outcomes and potential associations with immunotherapy response. Multi-omics integration prioritizes genomic and epigenomic aberrations and their associated vulnerabilities; defines ecotype-specific transcriptional programs, signaling pathways, PTMs, protein interaction networks, and metabolic regulation; and identifies microbiome features linked to ecotypes and resistance pathways. We further prioritize ecotype-, genomic subtype-, and cell type-specific targetable proteins using proteomic and PTM analyses within a tumor microenvironment context. This comprehensive atlas provides a systems-level blueprint for decoding gastric cancer heterogeneity and advancing precision oncology.

INTRODUCTION

Gastric cancer (GC) remains a leading cause of cancer-related mortality worldwide, with poor outcomes driven by marked biological heterogeneity and limited therapeutic efficacy.[1–3] Large-scale genomic and transcriptomic studies, including those from The Cancer Genome Atlas (TCGA)[4] and the Asian Cancer Research Group (ACRG),[5] have identified key driver mutations and defined molecular subtypes based on features such as Epstein-Barr virus (EBV) infection, microsatellite instability (MSI), chromosomal instability (CIN), and transcriptional signatures of *TP53* activity and epithelial-to-mesenchymal transition (EMT). However, these studies incompletely resolve downstream functional consequences and rely on bulk measurements that obscure cellular heterogeneity and tumor microenvironment (TME) architecture.

Single-cell and spatial transcriptomic analyses have identified recurring multicellular ecosystems, or ecotypes, reflecting coordinated tumor-stroma-immune interactions.[6–11] Although these studies highlight the importance of microenvironmental organization in shaping tumor behavior, they remain largely transcriptome-centric and do not capture regulatory activity at the protein, post-translational modification (PTM), protein-protein interaction (PPI), or metabolic levels.

Proteomic and proteogenomic studies have begun to bridge this gap by measuring protein abundance and signaling activity,[12–15] yet most focus primarily on phosphorylation and rarely integrate additional regulatory layers such as glycosylation, acetylation, ubiquitination, PPIs, metabolomics, or tumor-associated microbiota. Moreover, limited integration with cell state and ecosystem architecture constrains functional interpretation.

To address these limitations, we integrate tumor ecotype classification with deep multi-omics characterization within the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC). We profiled primary adenocarcinomas, tumor-matched normal adjacent tissues (NATs), and blood samples from 159 GC patients. We expanded CPTAC standardized pipelines[16] to incorporate intact glycoproteomics, cross-linking mass spectrometry (XL-MS) for PPI mapping, metabolomics, and tumor-resident microbiome profiling, generating 15 data layers. Using cell-state deconvolution,[17] we define ecotypes that stratify tumors beyond genomic and histological subtypes and link TME architecture to survival and immunotherapy response. Through integrative analyses of newly generated and published bulk and single-cell datasets, we map functional consequences of genomic and epigenomic alterations, delineate ecotype-specific proteogenomic and metabolic programs, characterize microbiome associations, and prioritize therapeutic targets across tumor subtypes and cell types. This framework provides a systems-level approach for decoding tumor ecosystems and informing therapeutic development.

RESULTS

Comprehensive molecular characterization

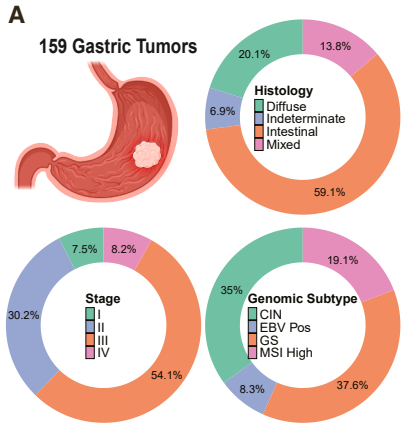
To comprehensively characterize GC heterogeneity, we assembled a cohort of 165 treatment-naive primary non-cardia gastric adenocarcinomas and 41 paired, histologically confirmed NATs, excluding patients with known hereditary cancer predisposition. After quality control, 159 tumors and 30 NATs were retained (Figure S1A; STAR Methods). This CPTAC cohort reflects GC diversity in age, clinically recorded gender, histology, and stage (Figure 1A; Table S1) and is predominantly derived from Eastern European patients, complementing prior Asian-focused proteogenomic studies.[12–14]

Homogenized flash-frozen tissues underwent multi-omics profiling. Whole-exome sequencing and whole-genome sequencing (WGS) identified somatic mutations and copy-number alterations (CNAs); methylation arrays captured DNA methylation; RNA sequencing (RNA-seq) and microRNA (miRNA) sequencing quantified mRNA, circular RNA, and miRNA abundances; tandem mass tag proteomics measured global protein abundance, phosphorylation, glycosylation, and acetylation; data independent acquisition proteomics quantified ubiquitination; XL-MS mapped PPIs; metabolomics profiled tumor metabolites; tumor-associated microbiota were inferred from WGS and RNA-seq; and clinical annotations were extracted from metadata. Together, these analyses generated 15 data layers comprising over 385,000 high-confidence features (Figures 1B, S1B, and S1C; Table S1). We further derived higher-order features including genomic aberration metrics, tumor composition metrics, regulatory and pathway activity scores, TCGA[4] and ACRG[5] subtype classifications (Figures S1D–S1G), and microbial community profiles (Figure 1C).

Among 157 genomically profiled tumors, TCGA classification identified EBV-positive (8%), MSI (19%), CIN (35%), and genomically stable ([GS]; 38%) subtypes (Figures S1D–S1F). In microsatellite stable (MSS) tumors, recurrent mutations mirrored previous reports,[4,18] including *TP53*, *ARID1A*, *PIK3CA*, *CDH1*, *MAP2K7*, and *SMAD4* (Figures 1D and S1H). MSI tumors showed a higher frequency of mutations than reported by TCGA (Figures 1E and S1H), likely reflecting increased

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS

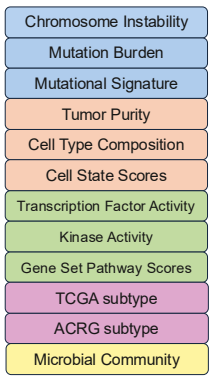


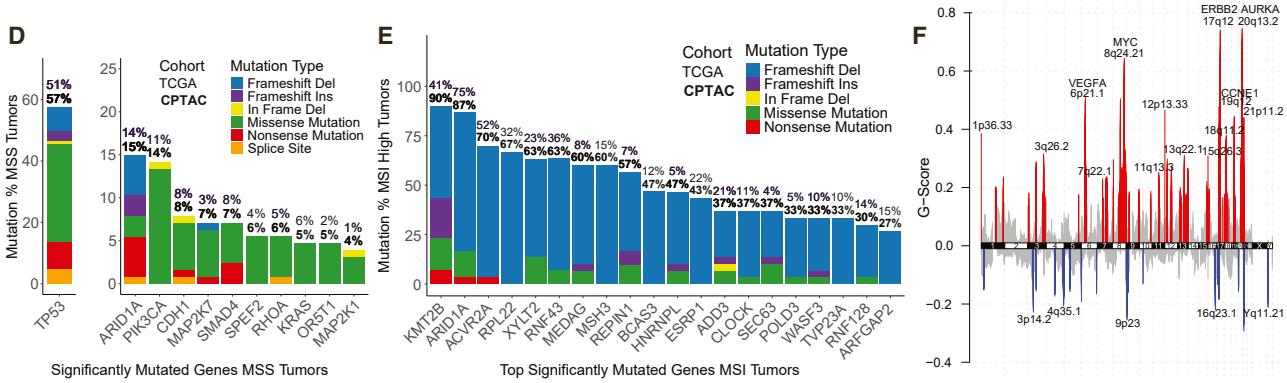
B

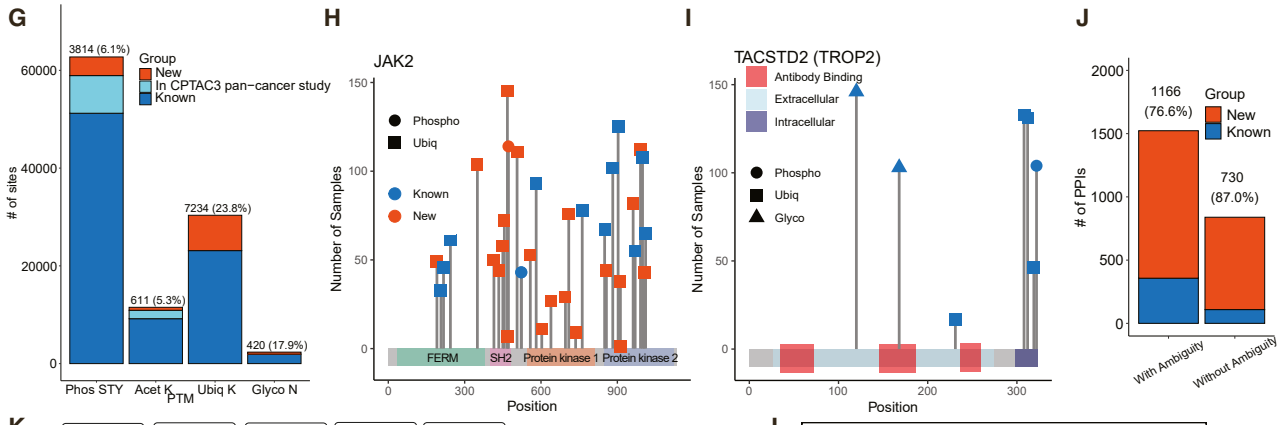
Platform	Data Layer	Feature Level	Feature Count
Sample metadata	Histopathology	Sample	35
WGS WES	CNA Somatic Mutations	Gene Coding mutation	62,659 72,919
Methylation Array	DNA Methylation	Gene	19,569
RNA-seq	RNA Circular RNA	Gene Gene	62,696 8,323
miRNA-seq	miRNA	Mature miRNA	2,582
TMT Proteomics TMT Proteomics **TMT Proteomics** **TMT Proteomics** **DIA Proteomics** **XL-MS**	Proteins Phosphorylation **Glycosylation** **Acetylation** **Ubiquitylation** **Protein interaction**	Gene Site **Site-specific glycoform** **Site** **Site** **Cross linked peptide**	11,294 62,823 **36,199** **11,549** **30,661** **3,813**
Metabolomics	**Metabolites**	**Metabolite**	**444**
WGS/RNA-seq	**Microbiome**	**Species**	**315**

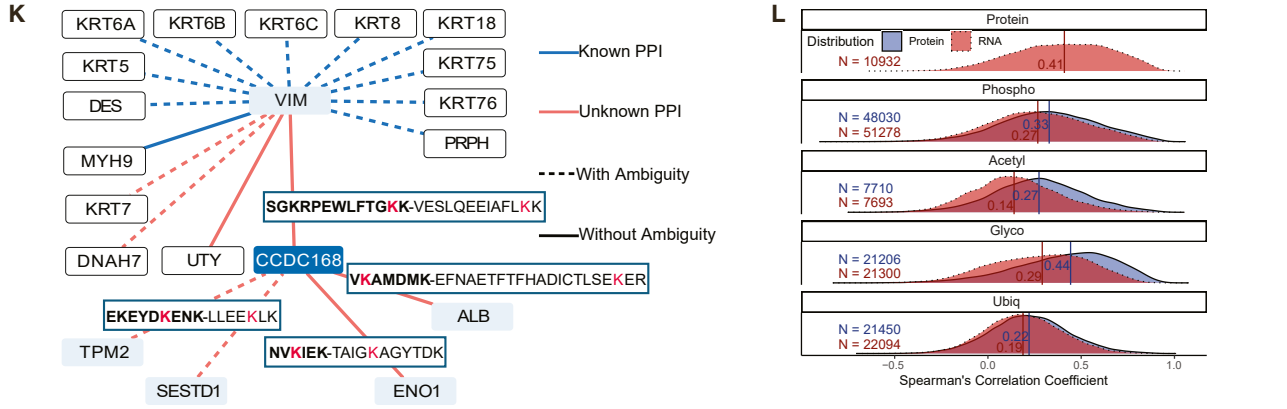
Total > 385,000 features

C









(legend on next page)

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

sequencing depth. Somatic CNA analysis identified both arm-level and focal alterations, including recurrent focal amplifications at 17q12 (*ERBB2*), 20q13.2 (*AURKA*), 8q24.21 (*MYC*), 6p21.1 (*VEGFA*), and 19q12 (*CCNE1*) (Figures 1F and S1D; Table S1).

Deep PTM profiling substantially expanded the known PTM landscape beyond existing databases[19–21] and CPTAC3 studies.[22] Among PTM sites identified in this study, 6.1% of serine/threonine/tyrosine (S/T/Y) phosphorylation sites, 5.3% lysine (K) acetylation sites, 23.8% lysine ubiquitination sites, and 17.9% asparagine (N)-linked glycosylation sites were newly reported (Figure 1G). This PTM atlas expands opportunities for mechanistic and therapeutic discovery. For example, JAK2, a key immune regulatory kinase controlled by ubiquitin-mediated turnover,[23] harbored 21 previously unannotated ubiquitination sites and one novel phosphorylation site (Figure 1H). Also, we detected glycosylation and ubiquitination sites at or near therapeutic epitopes of TACSTD2 (TROP2, Figure 1I), a surface antigen targeted by US Food and Drug Administration (FDA)-approved antibody-drug conjugates (ADCs) and under clinical evaluation in GC,[24] underscoring the importance of incorporating PTM information into therapeutic epitope selection and biomarker development.

We established large-scale XL-MS profiling in a tumor cohort, generating a GC-specific PPI network that substantially expands beyond curated PPI databases[25–27] (Figure 1J). With residue-level resolution, this network reveals novel interaction interfaces and contextualizes understudied proteins. For example, the poorly characterized cancer-testis antigen CCDC168 interacted with EMT-associated proteins VIM and TPM2, while known VIM interactions were confirmed (Figure 1K).

Quantitative integration revealed widespread post-translational regulation. The median mRNA-protein correlation was moderate (Spearman's $\rho = 0.41$), consistent with other cancers,[28] whereas PTM-protein correlations were similarly modest ($\rho = 0.22$–0.44), and PTM-mRNA correlations were weaker (Figure 1L).

Functional contextualization of genomic and epigenomic aberrations

We assessed associations between mutation status of 11 significantly mutated genes in MSS tumors (Figure 1D) and features across transcriptomic, proteomic, PTM, and metabolomic layers. *TP53*, *MAP2K1*, and *SMAD4* showed the broadest molecular footprints (Figure 2A; Table S2). *TP53* mutations were associated with widespread transcriptional and phosphosignaling changes, consistent with the role of *TP53* as a master regulator of gene expression and stress signaling. Canonical *TP53*-inducible genes *MDM2*, *DDB2*, and *RPS27L* were the strongest mRNA associations[29,30] (Figures 2A and S2A). *MAP2K1* mutations were linked to broad transcriptomic changes despite minimal detectable phosphosite alterations, suggesting a narrow or transient set of phosphorylation events triggering downstream transcriptional responses. In contrast, *SMAD4* mutations were associated with extensive post-transcriptional remodeling, including widespread protein, PTM, and metabolite changes, as well as kinase activity alterations (Figure 2B). Kinase activity changes were supported by phosphorylation shifts at kinase substrates independent of total protein abundance. For example, phosphorylation of multiple CSNK2A1 substrates was elevated in *SMAD4*-mutant tumors without corresponding mRNA or protein changes (Figures 2A, 2B, and S2B). While *SMAD4* knockdown has been linked to increased *CSNK2A1* expression,[31] our data reveal increased CSNK2A1 kinase activity in *SMAD4*-mutated tumors, nominating it as a candidate therapeutic vulnerability.

We next prioritized CNAs showing concordant copy number, mRNA, and protein alterations, together with differential protein abundance between tumors and NATs, identifying 1,729 candidate drivers enriched for RNA processing, nuclear transport, and amino acid metabolism (Figures 2C and 2D; Table S2). Prominent among these were amplified RNA splicing factors frequently upregulated in tumors (Figure 2E). These factors showed strong mRNA correlation with *MYC* (Figure 2F), suggesting coordinated upregulation in *MYC*-amplified tumors and indicating potential oncogene-induced dependencies.[32]

Assessing DNA methylation, we identified 102 candidate methylation drivers inversely correlated with mRNA and protein expression, enriched for cell-cell junction organization and neuropeptide signaling pathway (Figures 2G and 2H; Table S2). Among junction-associated genes (Figures 2I and S2C), *CLDN18* is therapeutically relevant, as its isoform *CLDN18.2* is targeted by FDA-approved antibody therapy in GC.[33,34] The

Figure 1. Multi-omics landscape of the CPTAC GC cohort

(A) Distributions of clinical features across the cohort.
(B) Overview of 15 multi-omics data layers; bold indicates layers not previously profiled in large GC proteogenomic studies.
(C) Derived features from (B), grouped and colored by category: genomic aberration metrics, tumor composition metrics, regulatory activity, molecular classification, and microbial communities.
(D) Mutation frequencies of significantly mutated genes (MutSig2CV, $q < 0.1$) in MSS tumors.
(E) Mutation frequencies of the top 20 significantly mutated genes (MutSig2CV, $q < 0.1$) in MSI-H tumors.
(F) Genome-wide focal CNAs, with G-score reflecting alteration frequency and amplitude.
(G) PTM sites identified in this study, categorized as newly identified, reported in CPTAC3 pan-cancer study but absent from public databases, or reported in PhosphoSitePlus.
(H) Lollipop plot of PTM sites identified on JAK2.
(I) Lollipop plot showing PTM sites identified and ADC-binding epitopes on TROP2.
(J) PPIs identified by XL-MS, categorized as newly identified or known interactions in STRING, BioGRID, BioPlex, or HuRI.
(K) XL-MS identified PPIs for CCDC168 and its partner VIM. Crosslinked peptides involving CCDC168 are mapped to their respective edges, with the CCDC168 peptides bolded, the partner peptide not bolded, and crosslinked lysine residues in red. Ambiguity arises from crosslinked peptides that map to multiple genes.
(L) Distributions of Spearman correlations in tumors: RNA-protein/PTM (red) and protein-PTM (blue). Vertical bar indicates medians. Also see Figure S1 and Table S1.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS

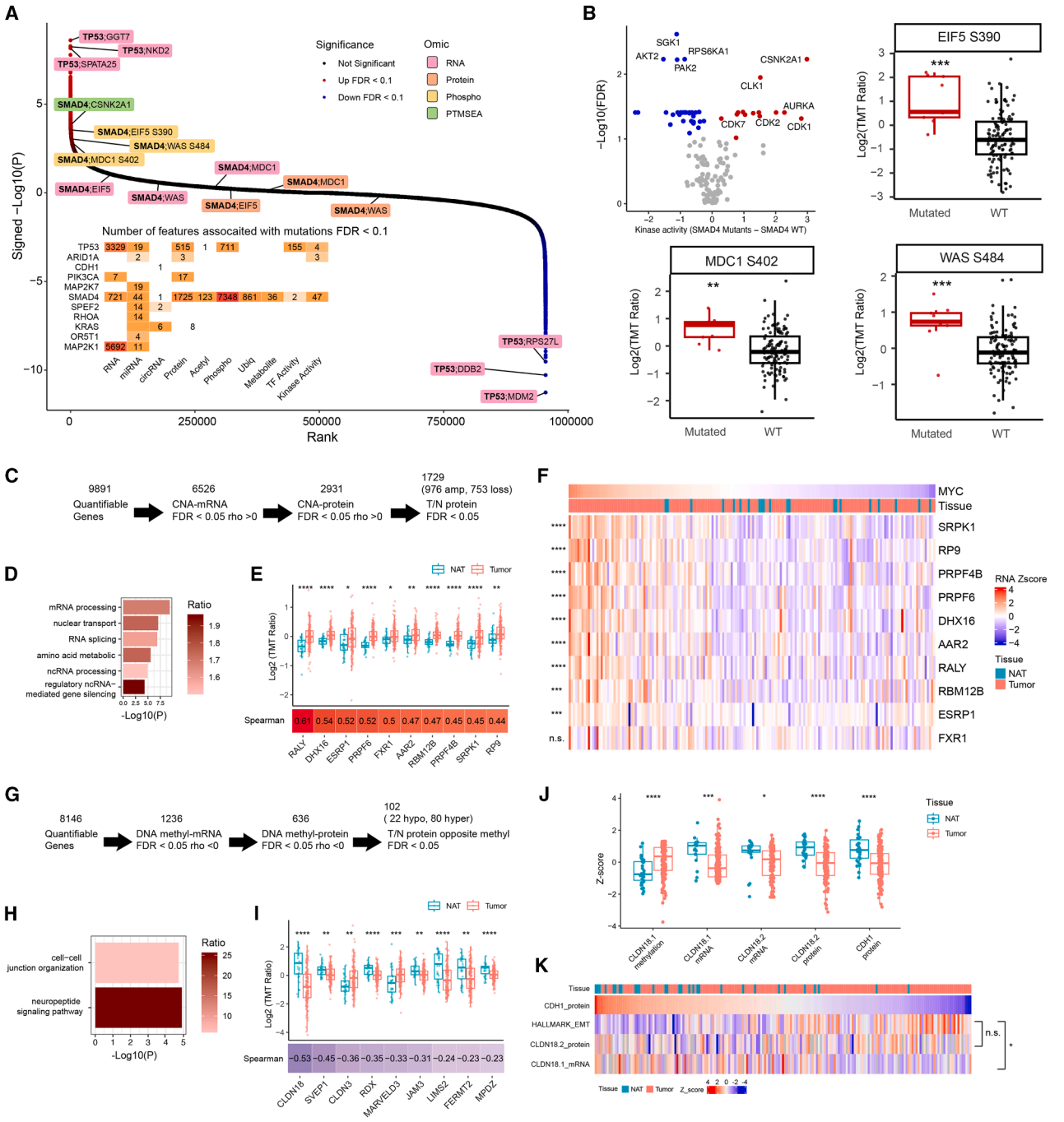


Figure 2. Functional contextualization of genomic and epigenomic aberrations

(A) Rank plot of signed $-\log_{10} p$ values for gene-feature associations in MSS tumors (Wilcoxon rank-sum test [WRS]).

(B) Volcano plot of kinase activity differences between *SMAD4*-mutant (n = 9) and *SMAD4*-WT MSS tumors (n = 118), with boxplots of phosphosite abundances for three SMAD4 substrates by mutation status (WRS).

(C) Workflow for prioritizing candidate CNA drivers by integrating CNA-mRNA and CNA-protein Spearman correlations with tumor vs. NAT differential expression (WRS).

(D) Gene Ontology Biological Process (GO BP) terms enriched among prioritized CNA drivers.

(E) Boxplots of protein abundance in tumors (n = 159) vs. NATs (n = 30) for 10 prioritized RNA splicing-related genes.

(F) Heatmap of mRNA levels of MYC and the splicing-related genes in (E) (n = 172, Spearman correlation).

(G) Workflow for prioritizing candidate methylation drivers by integrating methylation-mRNA and methylation-protein Spearman correlations with tumor vs. NAT differential expression (WRS).

(legend continued on next page)

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

analyzed methylation probe mapped to *CLDN18.1*, which showed increased methylation and reduced mRNA expression in tumors, indicating epigenetic repression at the *CLDN18* locus. In parallel, *CLDN18.2* mRNA and protein levels were also lower in tumors compared to NATs (Figure 2J). Notably, the epithelial marker CDH1 was significantly decreased at the protein level in tumors (Figure 2J), consistent with EMT-associated disruption of epithelial junctions. Together, these findings suggest that CLDN18.2's therapeutic relevance may reflect altered membrane accessibility during EMT rather than tumor-specific overexpression, consistent with prior reports.[35,36] Because CLDN18.2-high tumors span both epithelial and mesenchymal states (Figure 2K), incorporating additional markers, such as *CLDN18.1* methylation, CLDN18 mRNA levels, or EMT indicators, may better identify tumors undergoing EMT while retaining surface-accessible CLDN18.2 for targeted therapy.

Tumor ecotypes reveal TME-driven stratification beyond genomic subtypes

To move beyond genomics-centric classification[4] and better resolve TME architecture, we applied a validated bulk RNA-seq deconvolution method[17] to estimate the relative levels of 71 epithelial, stromal, and immune cell states. Consensus clustering identified six tumor ecotypes (C1–C6, Figure 3A and S3A–S3C; Table S3). Clustering 37 significantly variable cell states defined five co-occurring TME programs (Figure 3A): a Transitional Interface program reflecting epithelial-stromal crosstalk and low-grade inflammation; an Immune Effector-Enriched program representing an inflamed and immune-activated TME; a Naive-Primed Immune program indicating immune infiltration with limited activation; an Immunoregulatory program reflecting an immune-excluded, tissue-maintaining niche potentially contributing to tumor persistence and progression, and an Angiogenic Immunosuppressive program representing a pro-tumorigenic TME. Across five external GC cohorts with available survival data (from four independent studies[4,5,37,38]), Cox regression and meta-analysis showed that Transitional Interface and Immune Effector-Enriched programs were associated with favorable prognosis, Naive-Primed Immune was neutral, and Immunoregulatory and Angiogenic Immunosuppressive programs predicted poor outcomes (Figures 3A; Table S3; STAR Methods).

Ecotypes represented distinct combinations of the TME programs (Figure 3A). C1 was enriched for Immunoregulatory and Angiogenic Immunosuppressive programs, had the highest stromal content (Figure 3B), and was enriched for ACRG EMT subtype and diffuse histology. C2 was Angiogenic Immunosuppressive dominant without accompanying Immunoregulatory features. C3 was defined by the Naive-Primed Immune program. C4 aligned with the Transitional Interface program and had the highest tumor purity and epithelial content (Figure 3B). C5 combined Immune Effector-Enriched and Naive-Primed Immune programs, showed the highest immune infiltration (Figure 3B), expressed CD274 (PD-L1) and IDO1, was enriched for MSI-related mutational signature SBS15, and under-represented *TP53* mutations. C6, the least frequent ecotype, showed selective enrichment of AGR2+ epithelial cells, suggesting a unique, epithelial-focused TME state. Among clinical features, only age was associated with ecotype, with C4 patients older than those in C1 and C3 (Figures S3D–S3E).

A multiclass classifier trained using ecotype labels and cell state profiles from our cohort reproduced ecotype patterns across five external GC datasets with diverse ancestry and geographic composition[4,5,37,38] (Figures S3F–S3G; Table S3; STAR Methods). Survival analysis in these external datasets showed that C1 had the poorest prognosis, C2 and C3 intermediate, and C4–C6 favorable outcomes ($p = 1.28e{-}06$, Figure 3C). Multivariable Cox regression, adjusting for age, clinically recorded gender, stage, and Lauren subtype confirmed ecotype as an independent predictor ($p = 1.97e{-}05$). C1 outcomes were worse than C2 despite the shared Angiogenic Immunosuppressive program, suggesting that the additional presence of the Immunoregulatory, tissue-maintaining program may lead to a more suppressive and treatment-resistant TME. Because our cohort lacks sufficient long-term survival data, we trained a diffusion-augmented AI model[39] on TCGA and ACRG datasets—without ecotype input—to predict survival (STAR Methods). When applied to our cohort, this model independently identified C1 as the poorest-prognosis ecotype (Figure S3H).

Ecotypes stratified genomically similar tumors by TME architecture in our cohort (Figure 3D) and the TCGA/ACRG cohort (Figure 3E). GS tumors were enriched in C1 and C3, and CIN tumors in C2 and C4. GS tumors classified as C1 vs. C3 differed markedly in TME programs (Figure 3A) and survival outcomes (Figure 3F). Diffuse-type tumors similarly segregated into C1 and C3 (Figures 3G and 3H) with distinct outcomes (Figure 3I). MSI is an established biomarker for immune checkpoint blockade (ICB). Interestingly, while MSI tumors were predominantly C5, 45% in our cohort and 47% in TCGA/ACRG fell into other ecotypes (Figures 3D and 3E). Using PredictIO,[40] a pan-cancer gene expression signature of ICB response, C5 MSI tumors had significantly higher predicted response than other MSI tumors across cohorts (Figure 3J), suggesting that ecotype refines immunotherapy stratification beyond MSI alone.

Proteogenomic programs underpin tumor ecotypes

To define the molecular programs underlying GC ecotypes, we integrated transcriptional hallmark signatures; inferred activities

(H) GO BP terms enriched among prioritized methylation drivers.
(I) Boxplots of protein abundance in tumors ($n = 159$) vs. NATs ($n = 30$) for nine genes involved in cell-cell junction organization.
(J) Boxplots comparing tumors ($n = 159$) and NATs ($n = 30$) for CLDN18.1 methylation and mRNA, CLDN18.2 mRNA and protein, and CDH1 and VIM protein expression.
(K) Heatmap of CDH1 protein, EMT signature scores, CLDN18.2 protein, and CLDN18.1 mRNA across tumors and NATs. CLDN18.1 mRNA and protein correlated with EMT score (Pearson correlation).
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$, n.s. non-significant. Also see Figure S2 and Table S2.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS

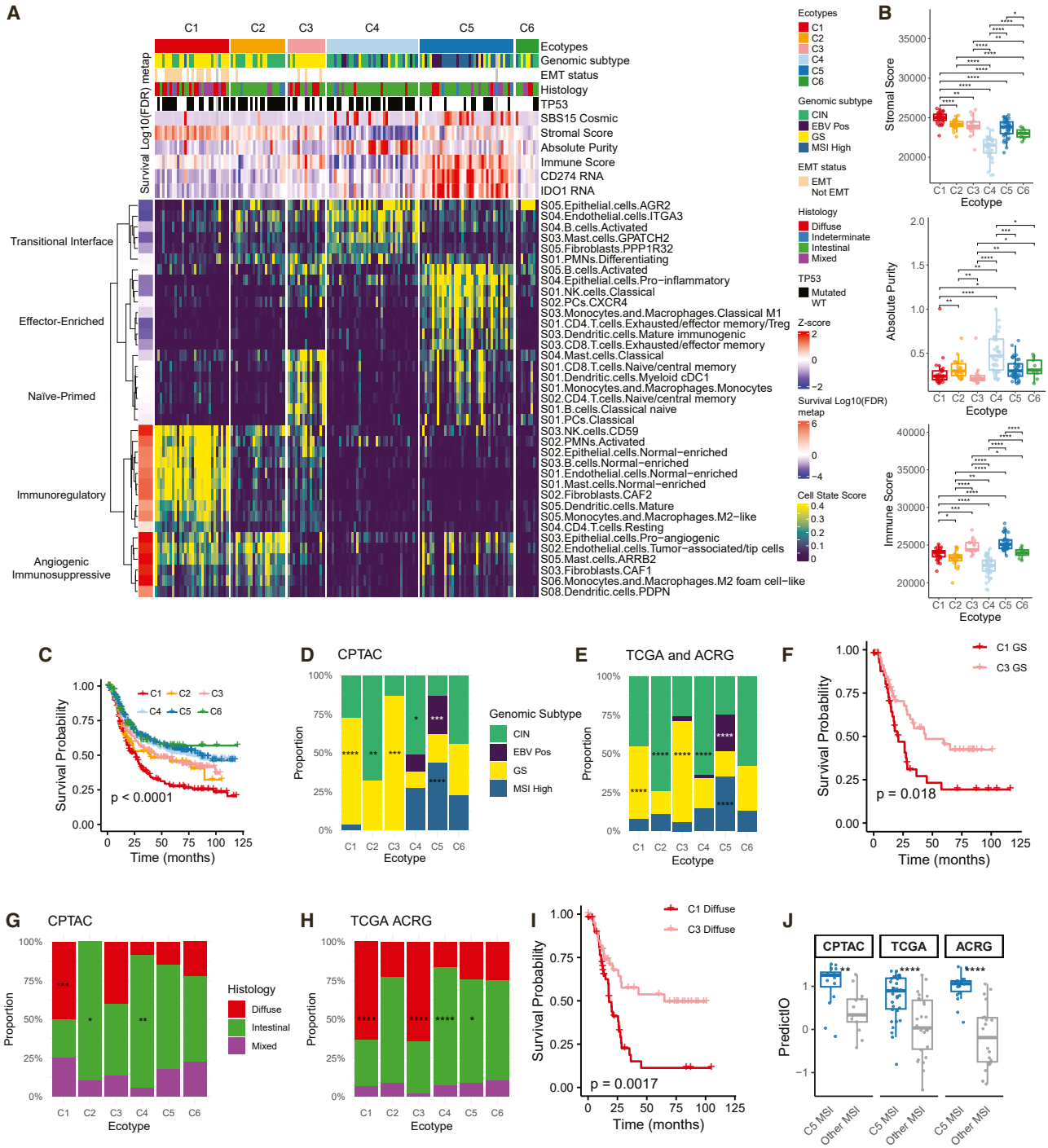


Figure 3. TME-defined tumor ecotypes stratify genomic and histological subtypes

(A) Six tumor ecotypes identified by consensus clustering of cell state scores. Meta-analysis of Cox regression survival associations from TCGA, ACRG, GSE15459, GSE26899, and GSE26901 cohorts is shown for each cell state (left).

(B) Boxplots comparing purity, immune, and stromal scores across ecotypes (WRS; *n* = C1: 30, C2: 22, C3: 15, C4: 37, C5: 38, C6: 9).

(C) Kaplan-Meier survival analysis of the tumor ecotypes using pooled TCGA, ACRG, GSE15459, GSE26899, and GSE26901 data (*n* = C1: 184, C2: 70, C3: 130, C4: 303, C5: 256, C6: 74).

(D and E) Stacked bar plots showing genomic subtype distribution by ecotype in the CPTAC (D) and TCGA/AGRG (E) cohorts, with enrichment assessed by Fisher's exact test.

(F) Kaplan-Meier survival analysis comparing C1 vs. C3 GS tumors in TCGA and ACRG (C1: 43, C3: 45).

(legend continued on next page)

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

of transcription factors (TFs), kinases, and miRNAs; multi-layer PTM profiling; and XL-MS-based protein interaction mapping.

Hallmark analysis revealed distinct ecotype-specific programs (Figure 4A). C1 was enriched for stromal pathways, including EMT, angiogenesis, and hypoxia, consistent with its stroma-rich phenotype and ACRG EMT subtype association. C4 was dominated by proliferative and biosynthetic signatures, reflecting a highly proliferative epithelial state. C5 exhibited broad activation of immune-related programs alongside suppression of metabolic pathways, supporting its immune-inflamed phenotype.

Regulator activity analysis showed the strongest differences in C1, C4, and C5 (Figures 4B–4D). Using an independent GC single-cell RNA-seq dataset[7] for orthogonal annotation, we mapped lineage enrichment of key regulators (Figures S4A and S4B; STAR Methods). C1 showed activation of fibroblast-specific TFs, including *FOXF1*, *MYOCD*, and *TCF21*, consistent with a stroma-driven transcriptional program. In contrast, its elevated kinases displayed broader lineage expression, suggesting intercellular signaling. C4 was characterized by epithelial-stabilizing miRNAs (including the miR-200 family), epithelial-enriched *E2F2* expression, and activation of proliferation-associated kinases (CDK1/2/7, AURKA, and PRKDC), supporting a coordinated proliferative program with epithelial enrichment. C5 showed activation of immune-related TFs, including STATs, IRFs, and NF-κB family members, many enriched in immune cells.

PTM analysis uncovered extensive ecotype-specific regulatory events not captured by mRNA or protein level, including 4,818 phosphorylation sites, 259 acetylation sites, 631 ubiquitination sites, and 1,587 intact glycopeptides (Figure 4E; Table S4). Among annotated PTM sites, many showed changes consistent with known functional consequences. For example, in C1, elevated activating phosphorylation of DEAF1 (T432) and PAK2 (S197) paralleled increased inferred TF and kinase activity, increased inhibitory acetylation of IDH2 (K106) suggested post-translational suppression of enzymatic function, and reduced degradative ubiquitination of ANXA1 (K312) corresponded to elevated protein abundance (Figures 4F–4I). In contrast, most ecotype-associated PTM events remain functionally uncharacterized, representing a largely unexplored regulatory layer. For example, the site-specific glycoform LRP1-N1575 N4H5F0S1G0 was elevated in C5 despite reduced LRP1 protein abundance (Figure 4J).

XL-MS profiling identified hundreds of significantly regulated intra- or inter-protein interactions associated with ecotype, EMT status, or tumor-NAT differences (Figure S4C). While most interaction changes tracked with protein abundance, some occurred independently of protein expression. For example, an intra-protein crosslink in EEF2, supported by structural mapping (Figure S4D), was elevated in C5 despite stable protein levels (Figure 4K). We next constructed a GC-specific protein interaction network using XL-Ranker[41] to resolve peptide-mapping ambiguity (see examples in Figure 1K), yielding 1,248 protein-coding genes connected by 1,185 high-confidence interactions (Figure S4E; Table S4, STAR Methods). Hierarchical module analysis using NetSAM[42] identified 102 network modules organized into four levels (Figure 4L; Table S4), spanning metabolism, immunity, gene regulation, and cytoskeleton and cell structure. Module-level differential analysis revealed ecotype-, EMT-, and tumor-NAT-associated alterations, with C1 and EMT tumors sharing numerous affected modules (Figure S4F). For instance, a chromatin-associated module (L2M9, Figure 4M) was elevated in tumors and stratified by ecotypes, showing the highest levels in good-prognosis ecotypes (C4–C6) and the lowest in poor-prognosis C1. In contrast, a cytoskeletal module (L1M6, Figure 4M) was broadly reduced in tumors but relatively preserved in C1, consistent with its tissue-maintaining TME features. An immune-associated module (L2M7, Figure 4M) was broadly upregulated in tumors, with the strongest elevation in the immune-infiltrated C5. While several genes in L2M7 are known to be expressed in immune cells, all interactions within this module were newly identified by XL-MS, highlighting our ability to uncover network context beyond curated databases.

Metabolic programs underpin tumor ecotypes

To define metabolic programs associated with tumor ecotypes, EMT status, and tumor-NAT differences, we integrated metabolomic, transcriptomic, proteomic, and PTM data. Comparative analysis identified 220 metabolites associated with ecotypes, EMT status, and tumor-NAT differences (Figure 5A). In C5, quinolinate and kynurenine were elevated, consistent with activation of the immunosuppressive tryptophan-kynurenine pathway in immune-infiltrated tumors.[43] C4 tumors showed increased thymidine and thymine, reflecting heightened nucleotide demand in proliferative states. In contrast, glycerol-3-phosphate was elevated in C1 and EMT tumors, consistent with the lipid metabolic reprogramming during EMT.[44]

Across 84 metabolites with quantifiable precursors, the maximum precursor-metabolite correlation per metabolite had a median Spearman correlation of 0.43 (range −0.22 to 1.0, Figure 5B), indicating substantial enzymatic regulation beyond substrate availability. To systematically identify candidate regulators beyond precursor effects, we applied MetaSage,[45] a machine learning framework that models the abundance of each metabolite's abundance by integrating precursor levels with mRNA, protein, and PTM measurements of directly associated anabolic and catabolic enzymes (STAR Methods, Figures 5C–5E). Of the 220 differential metabolites shown in Figure 5A, 114 had quantifiable predictors, and 51 were robustly predicted (Pearson $r > 0.5$, $p < 0.05$, Figure S5A; Table S5). For 25 metabolites, model predictions correlated more strongly with measured abundance than any individual precursor (Figure 5F),

(G and H) Stacked bar plots showing histological subtype distribution by ecotype in the CPTAC cohort (G) and TCGA/AGRG (H) cohorts. Fisher's test odds ratio >1.

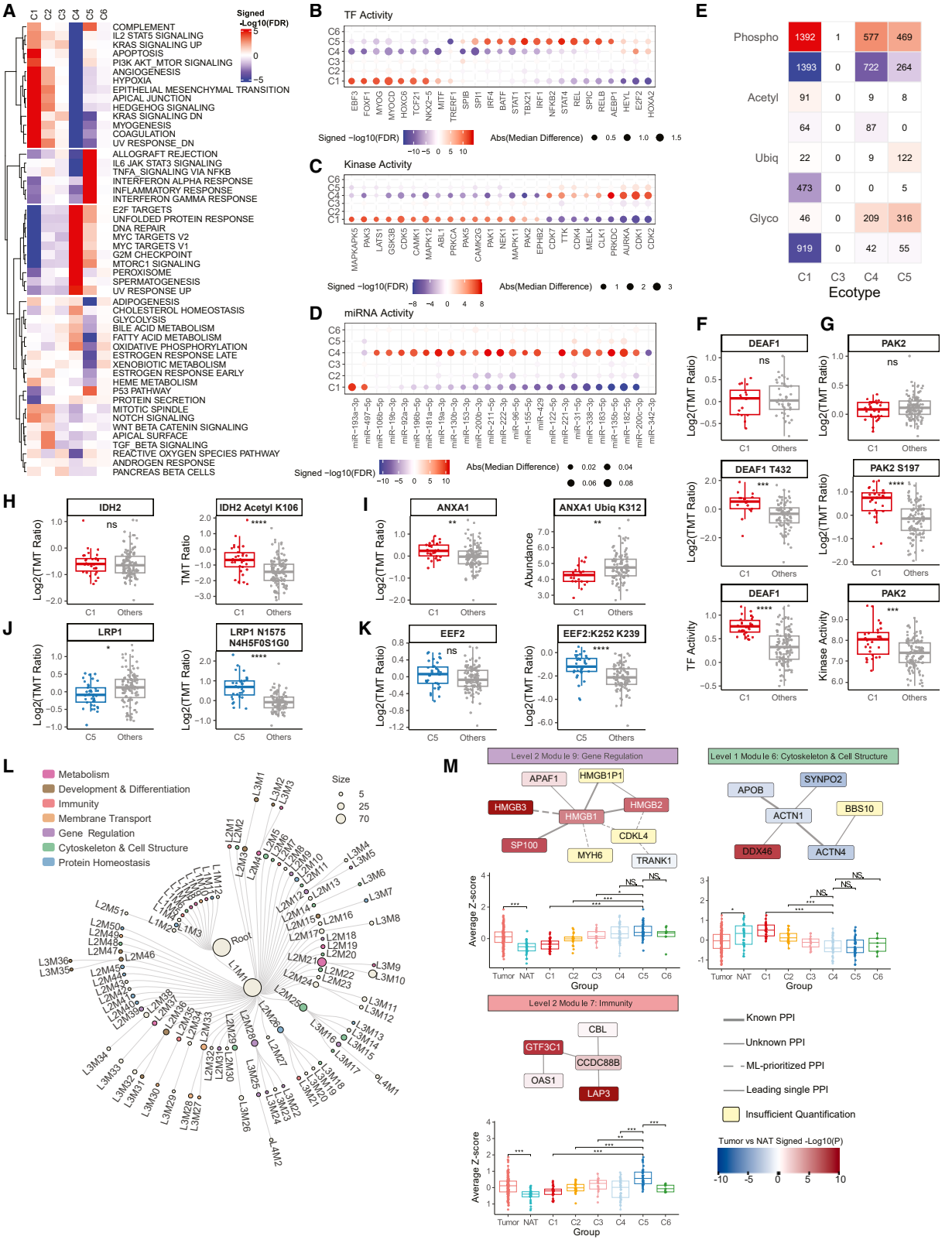
(I) Kaplan-Meier survival analysis comparing C1 vs. C3 diffuse tumors in TCGA and ACRG (C1: 50, C3: 30).

(J) PredictIO immune checkpoint response scores for C5 MSI vs. other MSI tumors in the CPTAC, TCGA, and ACRG cohorts. (WRS; n = CPTAC: 16 vs. 13, TCGA: 35 vs. 26, ACRG: 18 vs. 21).

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$. Also see Figure S3 and Table S3.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS



(legend on next page)

suggesting contributions from enzymatic regulation. Feature importance analysis yielded a regulatory network linking 196 important features to 51 well-predicted metabolites (Figure S5B; STAR Methods). Precursor metabolites accounted for 26% of these predictors, RNA and protein expression of associated enzymes each contributed 22%, and PTMs accounted for 30% (Figure 5G).

IDO1 emerged as the top predictor of formylkynurenine abundance, consistent with its role in catalyzing tryptophan conversion to formylkynurenine[46] (Figure 5H). IDO1 mRNA and protein were elevated in ecotype C5 (Figures 3A, 5I, and 5J), aligning with increased formylkynurenine and decreased tryptophan (Figures 5C and S5C). Supporting this axis, AFMID, which mediates downstream formylkynurenine metabolism, decreased in C5 (Figure S5D). Single-cell RNA-seq data[7] showed IDO1 enrichment in dendritic cells (Figure S5E), implicating the TME in regulating tryptophan metabolism.

Ubiquitination at IDO1 K389 was also associated with formylkynurenine levels (Figure 5E) and was elevated in C5 (Figure 5K), correlating positively with IDO1 mRNA and protein abundance (Figure 5L), suggesting that K389 ubiquitination does not promote degradation in this context. Structural mapping places K389 within the C-terminal K-helix, stabilized by surrounding helices and the EF-loop (Figure 5M), potentially limiting its accessibility.[47] Prior studies show that structural exposure of this region enables K389 ubiquitination-mediated degradation,[47] supporting its role as a conditional degron and a potential therapeutic vulnerability in GC.

Finally, correlating dysregulated metabolites with five TME programs (Figure 3A) identified program-specific metabolic signatures (Figure 5N). Formylkynurenine, kynurenine, and quinolinate were strongly associated with the Immune Effector-Enriched program, reinforcing IDO1-mediated tryptophan metabolism in immune-active TMEs. In contrast, glycolysis-feeding monosaccharides, including fructose, galactose, glucose, and mannose, were associated with the Immunoregulatory program, suggesting adaptation to hypoxic, stroma-rich niches. Pathway enrichment analysis further supported these associations (Figure 5O).

Tumor microbiome associations with subtypes and molecular programs

Using WGS and RNA-seq data processed through a comprehensive and rigorously filtered microbial analysis pipeline (STAR Methods), we profiled tumor-associated microbiota, identifying diverse bacteria, fungi, and viruses (Figure 6A). Of 1,506 initially detected species, removal of likely contaminants and false-positives[48,49] yielded 315 high-confidence, tissue-enriched species (Figure S6A) supported by concordant DNA and RNA evidence (Figure S6B). Across tumor samples, *Streptococcus* and *Prevotella* were most abundant, followed by other gastrointestinal-associated taxa such as *Lactobacillus*, *Fusobacterium*, and *Helicobacter*[50] (Figure S6C). EBV was enriched in 13 samples (Figure 6A), exactly matching the independently determined EBV subtype (Figure S1E), with a median abundance of 275 reads per million (~10.6 viral copies per tumor cell).

Community analysis of tumor microbiomes identified an EBV-only cluster and four additional multi-species microbial communities, designated as the *Helicobacter pylori*, *Lactobacillus-Candida albicans*, Pathobiont, and *Veillonella* clusters (Figure 6B; Table S6). The Pathobiont cluster was the most abundant (Figure S6D). Overall, tumors segregated into five "tumor enterotypes"[51,52] largely dominated by individual microbial clusters (Figure 6C).

In 27 matched tumor-NAT pairs, the *H. pylori* and *Veillonella parvula* clusters were enriched in NATs (Figure S6E). At the species level, *H. pylori* was enriched in NATs ($p = 5.2 \times 10^{-3}$), whereas EBV was enriched in tumors ($p = 1.22 \times 10^{-3}$) (Figure 6D), supporting a model in which *H. pylori* contributes to GC initiation and EBV to tumor progression.[53–55]

Microbial composition varied by genomic subtype. MSI tumors showed higher bacterial load and increased relative abundance of the Pathobiont cluster (Figures S6F and S6G). Species-level analysis confirmed MSI-associated taxa previously reported in colorectal cancer, including *Fusobacterium animalis*,[56] *Parvimonas micra*,[57] and *Selenomonas sputigena*,[58] and newly associated *Streptococcus anginosus* (Figure S6H).

Microbial composition also differed across tumor ecotypes. C4 and C5 tumors had higher bacterial load (Figure 6E) and were enriched for the Pathobiont cluster (Figure 6F). Species-level analysis revealed enrichment of *S. anginosus* and other Pathobiont cluster species in C4, while C5 was enriched with *Lactobacillus gasseri* and *Limosilactobacillus vaginalis* (Figure 6G). Additionally, *S. anginosus* was negatively associated with EMT status. Within the MSI genomic subtype, C4 tumors showed higher relative abundance of the Pathobiont cluster

Figure 4. Proteogenomic programs underpin tumor ecotypes

(A) Heatmap of proteomics-derived Hallmark pathway activity across ecotypes. Colors indicate signed $-\log_{10}$ FDR from one-vs.-rest WRSs.

(B–D) Top 25 transcription factors (B), kinases (C), and miRNAs (D) with the strongest differential activity across one-vs.-rest comparisons (WRS). For miRNA, only those with concordant activity and abundance are shown.

(E) Heatmap of significantly regulated PTM sites by PTM type from one-vs.-rest ecotype comparisons (WRS); (top) upregulated sites; (bottom) downregulated sites ($|\text{signed} -\log_{10}\text{FDR}| > 2$), without concordant RNA or protein changes.

(F–J) Boxplots comparing protein, PTM, or inferred activity between a specific ecotype and others for (F) DEAF1 (protein, pT432, TF activity), (G) PAK2 (protein, pS197, kinase activity), (H) IDH2 (protein, acK106), (I) ANXA1 (protein, ubK312), and (J) LRP1 (protein, glycosylation N4H5F0S1G0 at N1575). (C1: 30 vs. 121; C5: 38 vs. 113).

(K) Boxplots of EEF2 protein abundance and the intra-protein cross-link K252-K239 in C5 vs. others (protein: 38 vs. 113; XL-MS: 36 vs. 101).

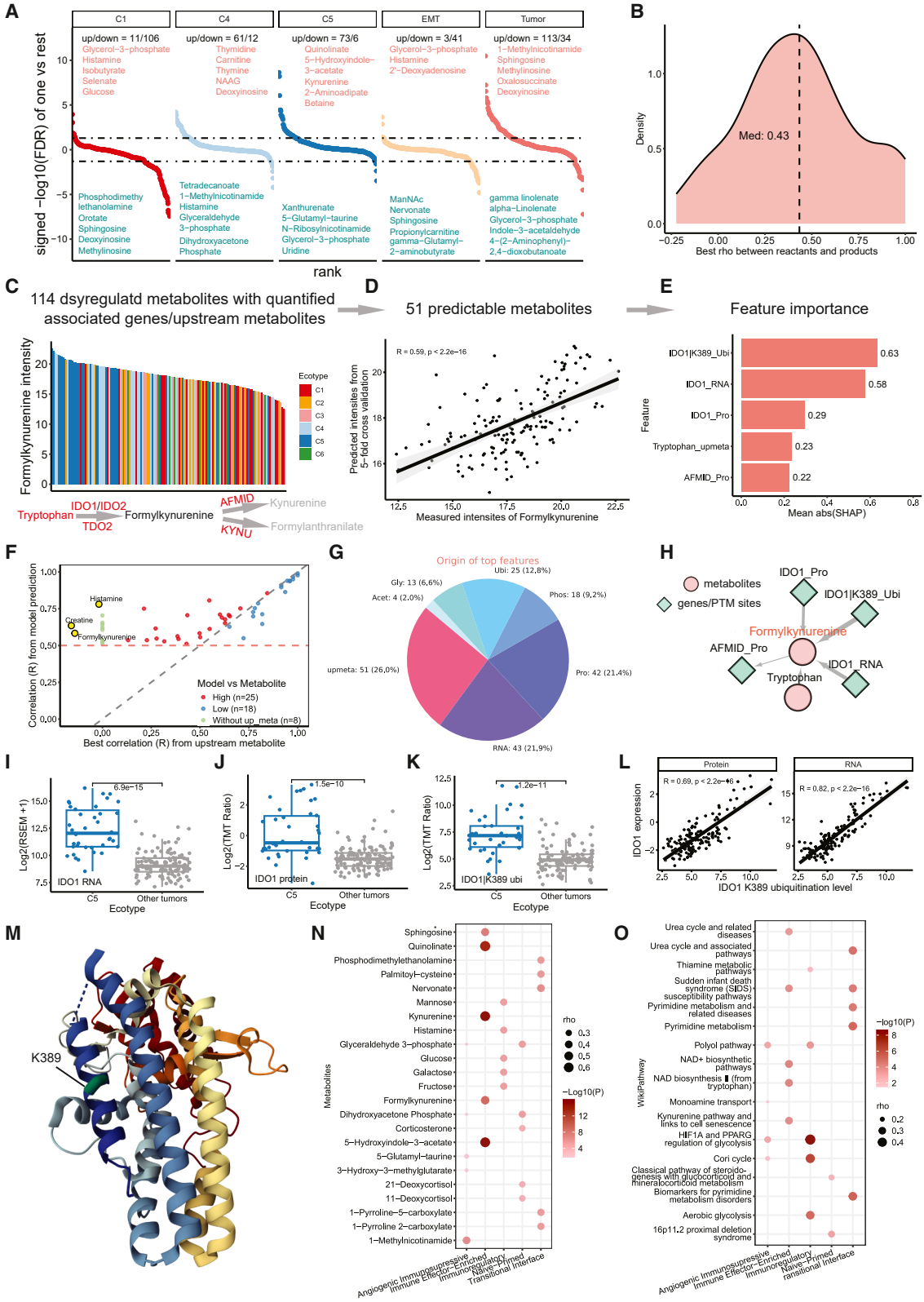
(L) Hierarchical organization of modules derived from the XL-MS PPI network.

(M) Representative modules from (L) showing differential abundance across ecotypes and tumor vs. NAT. (Top) Module networks colored by tumor vs. NAT expression; edge styles denote interaction groups determined by the XL-Ranker. (Bottom) Boxplots of module abundance across tumor, NAT, and ecotypes (tumor: 114, NAT: 29, C1: 26, C2: 20, C3: 14, C4: 32, C5: 36, C6: 9).

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. Also see Figure S4 and Table S4.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

than C5 (Figure 6H), suggesting that low immune infiltration may favor Pathobiont colonization.

To explore functional associations, we first assessed tumor glycosylation profiles, given their role in host-microbe interactions.[59–61] Tumors enriched for the Pathobiont cluster showed decreased levels of fucose-sialic glycans and elevated levels of high-mannose glycans; the latter have been linked to tumor proliferation[62] (Figure 6I).

Integrative metabolic and proteomic analysis revealed the most widespread associations with the Pathobiont cluster (false discovery rate [FDR] <0.1) (Figure 6J). Proteomics-based pathway enrichment, controlling for subtype and ecotype, linked the Pathobiont cluster to metabolic reprogramming, RB signaling, DNA mismatch repair, hypoxia, NF-κB mediated inflammation, and 5-fluorouracil (5-FU) chemoresistance (Figure 6K), many of which are supported by colorectal cancer studies.[63–66] Metabolite analysis associated the Pathobiont cluster with histidine metabolism (Figure 6K), including elevated N-formiminoglutamic acid (FIGLU) (Figure 6L), a biomarker of folate deficiency. As folate deficiency contributes to 5-FU resistance,[67,68] this association suggests a microbial link to chemoresistance. Consistently, thymidylate synthase (TYMS), a mediator of 5-FU resistance,[69] was also associated with the Pathobiont cluster (Figure 6M).

Therapeutic target prioritization in the context of GC heterogeneity

Because GC exhibits extensive genomic, molecular, and microenvironmental heterogeneity, traditional tumor-normal comparisons may overlook subgroup-specific targets. We, therefore, performed systematic outlier analysis[70] to identify proteins and PTMs aberrantly activated in tumor subsets (STAR Methods) and mapped them across genomic subtypes, ecotypes, and cell-of-origin derived from published single-cell RNA-seq data.[7]

Seventy-seven percent of tumors showed outlier overexpression of at least one of 23 established antibody- and T cell-based targets, including ERBB2 (HER2),[1,71] TACSTD2 (TROP2),[72] CEACAM5,[73] MUC1,[74] and FOLR1[75] (Figures 7A and S7A; Table S7). We also identified additional surface proteins with recurrent high outlier expression (Figure 7A), including COL12A1, COL5A1, and VCAN, whose mRNA levels were associated with poor survival in prior cohorts[4,5,37,38] (Figure S7B), expanding the repertoire of candidate surface targets.

Using intact glycoproteomics data, we identified site-specific-glycoforms with recurrent high outlier expression (Figures 7B; Table S7). Among site-specific-glycoforms with ≥15 high outliers, high-mannose species were significantly enriched (Figure 7C), indicating incomplete N-glycan biosynthesis[76] and potentially linked to downregulation of mannose-removing enzyme MAN2C1 and the fucosylation and sialylation enzymes FUT1, FUT2, ST6GALNAC1, and ST3CAL6 in tumors (Figure S7C). At the gene level, 340 genes harbored at least one high site-specific-glycoform outlier. CEACAM5 ranked fourth, with 70% of tumors showing glycosylation outliers and many other antibody drug targets also showed extensive glycopeptide outliers (Figures 7D and S7D; Table S7). Notably, COL1A2 and COL5A2 showed modest protein outliers (12 and 14 samples) but extensive site-specific-glycoform outliers (70 and 69 samples, Figure 7E), highlighting that glycan remodeling can markedly expand and refine the therapeutic target landscape beyond protein abundance alone.

We next prioritized intracellular targets by integrating outlier analysis with CRISPR knockout data from GC cell lines in DepMap[77] to identify overexpressed or hyperactivated dependencies (Figure 7F; STAR Methods), as intracellular targets generally require tumor cell dependency for small-molecule action. We identified 21 candidates, including CDK6, a target of approved oncology drugs[78]; 12 targets of experimental drugs; and seven additional proteins with small-molecule potential, linking aberrant activation to functional vulnerability.

Mapping prioritized targets to ecotype, genomic subtype, and lineage context revealed a broad repertoire of context-specific therapeutic opportunities (Figures 7G and 7H). C1- and C2-enriched targets (e.g., ASPN, MFAP4, OGN, COL14A1, DCN, PRELP, SERPINE2, COL3A1, BGN, COL5A1, and LTBP1/2) were predominantly expressed in fibroblasts. In contrast, C4- and C5-enriched targets exhibited broader lineage distribution. Epithelial-predominant targets included LAMC2 and EPCAM

Figure 5. Integrative analysis of metabolic dysregulation and model-based regulatory inference

(A) Differentially abundant metabolites identified by one-vs.-rest comparisons (C1, C4, or C5 vs. other tumors; EMT vs. non-EMT tumors; and tumors vs. NATs). The top five significantly up- or downregulated metabolites in each comparison are labeled (red, green).

(B) Distribution of the strongest Spearman correlation between each significantly altered metabolite in (A) and its quantified upstream reactants. Dashed line indicates the median.

(C) Formylkynurenine intensity across ecotypes. Associated enzymes and upstream/downstream metabolites are shown below; regulatory enzymes and metabolites used as MetaSage input are highlighted in red.

(D) Pearson correlation between measured and model-predicted formylkynurenine intensities.

(E) SHAP (Shapley Additive Explanations) feature importance scores for input features contributing to formylkynurenine intensity prediction.

(F) Scatterplot comparing, for each metabolite, the Pearson correlation between model-predicted and measured abundance (*y* axis) and the strongest correlation between measured abundance and any quantified upstream metabolite (*x* axis). Red and blue indicate metabolites for which the model correlation is higher and lower than the upstream correlation, respectively; green denotes metabolites without quantified upstream reactants.

(G) Origin of top predictive features: upstream reactants (red) and gene-derived features (blue), with different blue shades indicating distinct gene-based feature groups.

(H) Network centered on formylkynurenine, displaying its top predictive features.

(I–K) Abundance of IDO1 RNA (I), protein (J), and K389 ubiquitination (K) in C5 vs. other ecotypes (I: 38 vs. 113; J: 38 vs. 121; K: 38 vs. 107).

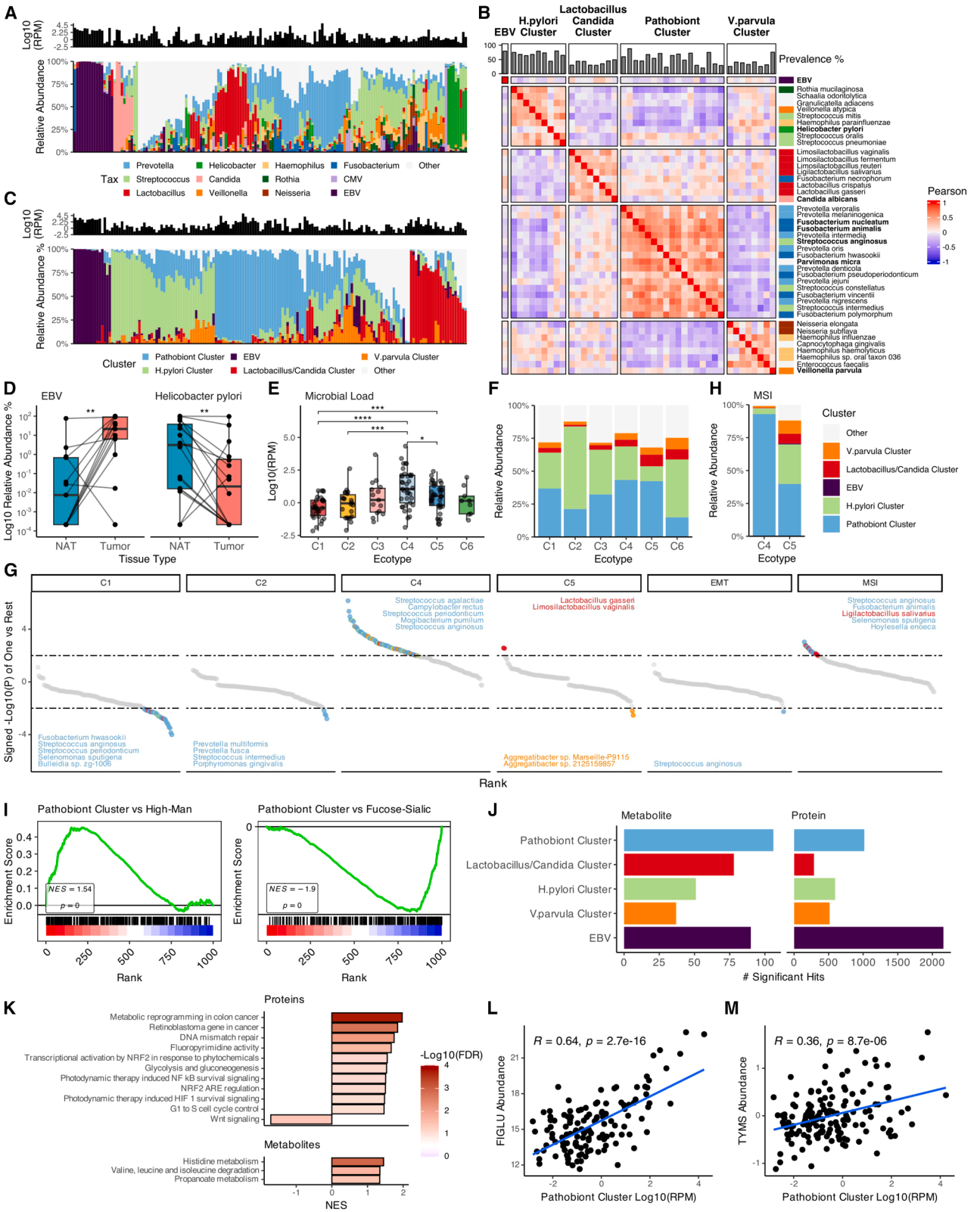
(L) Spearman's correlation between IDO1 K389 ubiquitination and IDO1 RNA/protein expression.

(M) Structural location of K389 on IDO1.

(N and O) Top five positively correlated metabolites and enriched metabolic pathways identified for each cell state community based on Spearman correlation. Also see Figure S5 and Table S5.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

(C4), LAMA3 (C5), TSPAN8 (MSI), and TROP2 (CIN). Interestingly, ERBB2, enriched in CIN and C4, was expressed at the mRNA level in both epithelial and fibroblast cells (Figures 7H and S7E). Together, these results demonstrate that therapeutic vulnerabilities in GC are ecotype and lineage dependent, extending beyond epithelial tumor cells to stromal compartments and underscoring the need for ecosystem-informed targeting strategies.

DISCUSSION

This study integrates tumor ecotype classification with deep multi-omics characterization to functionally characterize GC ecosystems. While ecotype stratification and multi-omics profiling have each been reported, their integration enables direct linkage of microenvironmental architecture to regulatory programs and therapeutic vulnerabilities.

Ecotype classification is significantly associated with patient survival and stratifies genomically and histologically similar tumors into biologically and clinically distinct groups. GS and diffuse-type tumors segregate into C1 and C3 ecotypes with markedly different survival outcomes. Among MSI tumors, those classified as C5 show higher predicted responsiveness to ICB than other MSI tumors, which may help explain the modest response rates observed in unselected MSI GC patients treated with pembrolizumab.[79]

Multi-omics integration reveals the functional architecture underlying ecotypes. By jointly analyzing transcriptomic, proteomic, PTM, PPI, metabolomic, and microbiome layers, we identify coordinated regulatory programs not apparent from RNA alone. For example, integration across layers identifies activation of the IDO1-kynurenine pathway in the immune-associated C5 ecotype, linking immune state to metabolic immune evasion and supporting a potential mechanism of resistance to ICB. Elevated ubiquitination of IDO1 at K389 in C5 correlates with expression and appears non-degradative under basal conditions; however, structural modulation can convert this site into a degradative signal,[47] highlighting the therapeutic potential of structure-guided IDO1 degraders in this context.

We extend prior therapeutic prioritization strategy beyond protein overexpression and kinase hyperactivation.[28] Intact glycoproteomics expands the surface target landscape, enabling glycoform-informed refinement of antibody-based therapies. XL-MS provides residue-level resolution of protein interactions, creating opportunities to target specific interaction interfaces rather than entire proteins,[80] potentially improving selectivity. Integration with external single-cell references addresses a key limitation of bulk proteomics by resolving the cellular origin of candidate targets, thereby supporting ecotype- and cell type-specific therapeutic strategies that extend beyond epithelial tumor cells to stromal and immune compartments. Finally, tumor-resident microbiome analysis links microbial communities to tumor immune and metabolic states, suggesting that microbial composition may serve both as a biomarker for treatment stratification and as a modifiable therapeutic axis.

Multi-omics integration also functionally contextualizes genomic and epigenomic alterations and refines associated therapeutic vulnerabilities. A clinically salient example is CLDN18.2. Although currently selected by protein expression alone, CLDN18.2-high tumors span distinct EMT states. Integration of methylation and EMT-associated features provides additional context that may reflect differential therapeutic accessibility, refining target interpretation beyond single-marker assessment.

In summary, by integrating tumor ecotype classification with deep multi-omics characterization, we define GC ecosystems in which microenvironmental architecture is directly linked to regulatory programs and therapeutic vulnerabilities. This framework refines tumor stratification beyond genome-centric models and provides a generalizable strategy for decoding complex tumor ecosystems in precision oncology.

Limitations of the study

This study has several limitations. First, the discovery cohort was derived from a predominantly Eastern European population. Although ecotype patterns were validated across independent cohorts with diverse geographic representation, broader generalization of quantitative associations will require multi-ancestry multi-omics studies. Second, the cohort lacked sufficient long-term survival data. Survival analyses were, therefore, performed using external public datasets without detailed treatment annotations, precluding treatment-type-stratified analyses and formal assessment of ecotype-treatment interactions. Third, the ecotypes were inferred from integrated molecular features rather than

Figure 6. Characterization of gastric tissue-associated microbiome and association with tumor features

(A) Microbial composition of decontaminated WGS primary tumors (n = 159). (Top) Total microbial load per sample (reads per million [RPM]). (Bottom) Relative abundance of selected high-abundance taxa.

(B) Pearson correlation heatmap of species abundances and microbial clusters; top eight species per cluster and pathobionts (bold) are shown.

(C) Microbial cluster composition: total microbial load (top) and cluster abundance (bottom).

(D) Paired tumor-normal analysis showing enrichment of EBV in tumors (n = 13) and *H. pylori* in NATs (n = 16).

(E) Tumor microbial load ($\log_{10}$RPM) by ecotype (C1: 30, C2: 22, C3: 15, C4: 38, C5: 38, C6: 9).

(F) Median relative abundance of microbial clusters ($\log_{10}$RPM) by ecotype.

(G) Relative abundance of median microbial cluster abundance ($\log_{10}$RPM) for C4-MSI (n = 10) and C5-MSI (n = 16) tumors.

(H) Species enrichment by ecotype and subtype using one-vs.-rest prevalence analysis. The top five significantly enriched or depleted species ($p < 0.01$) are colored by microbial cluster.

(I) Mountain plots showing enrichment and depletion of high-mannose and fucose-sialic glycan types in association with the Pathobiont cluster.

(J) Number of significant cluster-metabolite (top) and cluster-protein (bottom) Pearson correlations ($p < 0.05$), calculated after blocking for ecotype and subtype.

(K) Significantly enriched pathways identified from proteins (top) and metabolites (bottom) associated with Pathobiont cluster abundances.

(L) Scatterplot of Pathobiont cluster abundance ($\log_{10}$RPM) and FIGLU abundance.

(M) Scatterplot of Pathobiont cluster abundance and TYMS expression. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$. See also Figure S6 and Table S6.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS

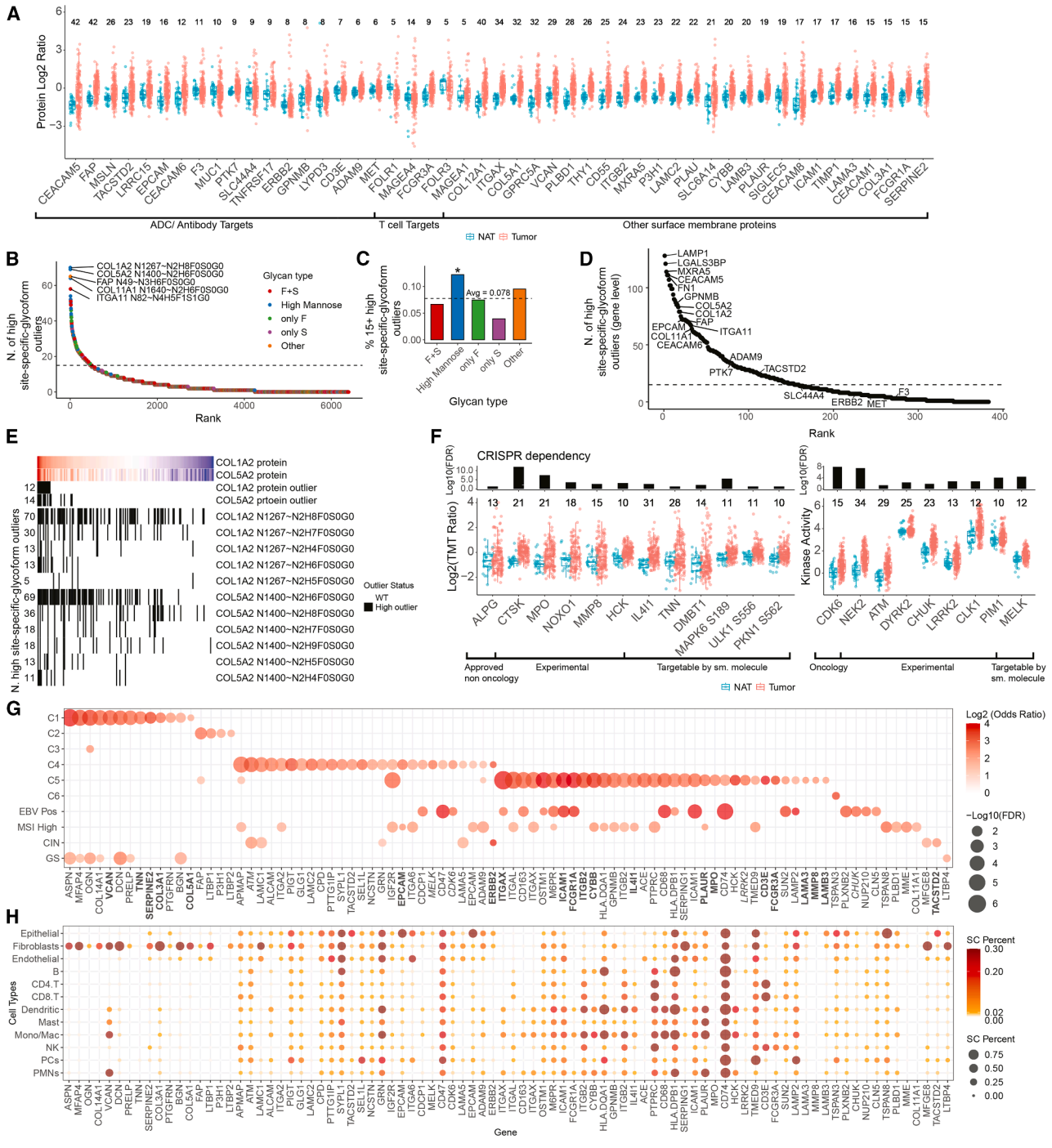


Figure 7. Therapeutic target discovery in the context of GC heterogeneity

(A) Boxplots of ADC targets, antibody targets, T cell targets (≥5 high protein outliers), and other non-essential gene surface proteins (≥15 high protein outliers). Numbers indicate high-outlier samples (n = 159 tumor, 30 normal).

(B) Site-specific-glycoforms from non-essential surface proteins ranked by high-outlier frequency (as in A). Dashed line indicates the 15-outlier cutoff.

(C) Ratio of site-specific-glycoforms with ≥15 high outliers by glycan type; dashed line, mean. Enriched glycan type is marked (*Fisher's exact test, p = 0.0005, odds ratio >1).

(D) Non-essential surface proteins ranked by the number of samples with high site-specific-glycoform outliers; top genes, top glycoforms' genes, and targets from (A) are labeled; dashed line ≥15 outliers.

(E) Heatmap of protein- and glycoform-level high-outlier samples for COL1A2 and COL5A2, ordered by COL1A2 protein abundance (n = 159).

(legend continued on next page)

definitive spatial entities. Future single-cell and spatial studies will be needed for experimental validation. Fourth, while technologies such as XL-MS and deep PTM profiling expand biological insight, continued improvements in sensitivity, coverage, and data analysis are needed to maximize utility in clinical samples. Finally, as a resource-oriented, hypothesis-generating study, functional validation of specific therapeutic hypotheses was beyond scope and will require dedicated investigation.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Bing Zhang (bing.zhang@bcm.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Genomic and transcriptomic data files can be accessed via the Genomic Data Commons (GDC) Data Portal: https://portal.gdc.cancer.gov/projects/CPTAC-3. Raw proteomics data files can be accessed at the Proteomic Data Commons: https://pdc.cancer.gov, with the following accession numbers: Proteome PDC000614, phosphoproteome PDC000615, glycoproteome PDC000616, acetylome PDC000622, metabolome PDC000621, protein-protein interaction PDC000626, and ubiquitylome PDC000617. Processed data matrices utilized for this publication can be accessed via LinkedOmics: https://www.linkedomics.org/data_download/CPTAC-GC, where a brief description of the data normalization method for each dataset is also included.
- The software and code used in this study are referenced in their corresponding STAR Methods sections and the key resources table.
- Additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

CONSORTIUM

Elleine Allapitan, Eunkyung An, Matthew L. Anderson, Oliver F. Bathe, Jasmin Bavarva, Oleksandr Bolba, Brittney Borresen, Melissa Borucki, Shuang Cai, M. Constanza Camargo, Steven A. Carr, Daniel W. Chan, Lijun Chen, Wenrong Chen, Andra Ciocan, Reese Crispen, Diwakar Davar, Saravana Mohan Dhanasekaran, Li Ding, Anders B. Dohlman, Marcin Domagalski, Yongchao Dou, Nathan J. Edwards, John M. Elizarraras, Pınar Ö. Eser, Rafael Fonseca, Victoria Fulidou, Yifat Geffen, Gad Getz, Michael A. Gillette, Charles A. Goldthwaite, Jr., Vladislav Golubkov, Ramaswamy Govindan, David I. Heiman, Julian Hess, Tara Hiltke, Galen Hostetter, Ralph H. Hruban, Yingwei Hu, Lan Huang, Eric J. Jaehnig, Kevin X. Jiang, Fenglong Jiao, Liyuan Jiao, Ramon U. Jin, Karen A. Ketchum, Pratik Khare, Iga Kołodziejczak-Guglas, Chandan Kumar-Sinha, Anne Le, Toan Le, Jonathan T. Lei, Qing Kay Li, Yuxing Liao, T. Mamie Lih, Tao Liu, Zaoxing Liu, Yin Lu, Avi Ma'ayan, Rashna Madan, D.R. Mani, Fernanda Martins Rodrigues, Peter McGarvey, Mehdi Mesri, Parham Minoo, Paul Morenkov, Dawid Murawa, Alexey I. Nesvizhskii, Chelsea J. Newton, Kristen Nyce, Timothy M. Nywening, Lindsey K. Olsen, Gilbert S. Omenn, Akhilesh Pandey, Kristine Parikyan, Amanda G. Paulovich, Chien-Wei Peng, Paul D. Piehowski, Alexander Pilozzi, Olga Potapova, Sergiy Revin, Ana I. Robles, Henry Rodriguez, Dan Rohrer, Sara R. Savage, Rostyslav Semikov, Paul W. Shafer, Bohdana Shkurupii, Zhiao Shi, Yvonne Shutack, Yanling Sun, Zhenyu Sun, Ratna R. Thangudu, Mathangi Thiagarajan, Tsanko Tsankov, Jeffrey W. Tyner, Negin Vatanian, Miodrag Vucic, Chenwei Wang, Pei Wang, Yuefan Wang, Bart O. Williams, Maciej Wiznerowicz, Clinton Yu, Nengneng Yu, Bing Zhang, Hui Zhang, Kevin Zhang, Tongwu Zhang, Xu Zhang, Huili Zhu.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute, United States, Clinical Proteomic Tumor Analysis Consortium (CPTAC) grants U24CA271079, U24CA271076, U24CA270823, U24CA271012, U24CA271037, U24CA271075, U24CA271114, U01CA271402, U01CA271407, U01CA271410, and U01CA271412. A.B.D.'s work was partially supported by the Damon Runyon Cancer Research Foundation, United States (DRG-2504-23). This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), United States. The contributions of the NIH authors were made as part of their official duties as NIH federal employees, follow agency policy requirements, and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

AUTHOR CONTRIBUTIONS

Sample processing and QC, G.H., C.J.N., A.I.R., and M.T.; data generation, Y.W., F.J., L.J., P.K., H. Zhu, L.C., T.-S.M.L., C.Z., and Z. Sun; data management, L.K.O., S. Cai, and K.A.K.; data analysis, Y.W., L.K.O., F.J., C.W., K.X.J., Y.D., Y.H., W.C., J.M.E., N.Y., T.-S.M.L., P.O.E., F.M.R., Z. Shi, C.Y., D.I.H., Y.L., P.W.S., S. Choi, J.M.C., S.R.S., E.J.J., J.T.L., Y.S., C.-W.P., P. Morenkov, Y.G., J.H., D.R.M., Z.L., A.B.D., H. Zhang, and B.Z.; data interpretation, Y.W., L.K.O., F.J., C.W., K.X.J., Y.D., Y.H., W.C., J.M.E., C.K.-S., Q.K.L., G.S.O., P. Minoo, M.C.C., R.H.H., A.B.D., R.U.J., L.H., H. Zhang, and B.Z.; figures, Y.W., L.K.O., F.J., C.W., K.X.J., Y.D., Y.H., W.C., J.M.E., and N.Y.; manuscript writing and editing, Y.W., L.K.O., F.J., C.W., K.X.J., Y.D., W.C., J.M.E., N.Y., P.O.E., D.I.H., K.Z., G.S.O., A.I.R., M.C.C., R.H.H., M.T., A.B.D., R.U.J., L.H., H. Zhang, and B.Z.; supervision, D.R.M., L.D., G.G., A.L., A.I.R., M.M., E.A., Z.L., M.T., A.B.D., R.U.J., L.H., D.W.C., H. Zhang, and B.Z.; funding acquisition, D.R.M., L.D., G.G., D.W.C., H. Zhang, and B.Z.

DECLARATION OF INTERESTS

A.B.D. is an inventor on US patent application no. PCT/US2023/029634. A.I.N. is the founder of Fragmatics and serves on the scientific advisory boards of Protai Bio, Infinitopes, and Mobilion Systems. B.Z. received research funding from AstraZeneca and consulting fee from Merck and Inotiv. G.G. receives research funds from IBM, Pharmacyclics/Abbvie, Bayer, Genentech, Calico, Ultima Genomics, Inocras, Google, Kite, and Novartis and is also an inventor on patent applications filed by the Broad Institute related to MSMuTect, MSMutSig, POLYSOLVER, SignatureAnalyzer-GPU, MSEye, MinimuMM-seq, and DLBclass. G.G. is a founder and consultant and holds privately held equity in Scorpion Therapeutics. He is a founder of and holds privately held equity in Predicta Biosciences and holds privately held equity in Antares Therapeutics. G.S.O. serves on boards of several early-stage biopharma companies, including spin-outs from the University of Michigan. None is targeting gastric cancers. J.H. is currently an employee of Predicta Biosciences and has privately held equity in the company. M.L.A. has received research grants from GSK, Regeneron, Sutro Biopharma, and Astra Zeneca and has consulted for GSK and Caris Life Sciences, LLC. M.R. is a member of the scientific advisory board of Universal DX. P.S. received consulting fees from Stereo Biotherapeutics. R.F. is a consultant

(F) Boxplots of putative intracellular targets with ≥10 high outliers and corresponding CRISPR dependency. Numbers indicate high-outlier samples. Dependency is measured as signed −log10(FDR) from one-tailed *t* tests of gene knockout effects on GC cell-line growth (DepMap). Only activating phosphorylation sites are shown; all genes are non-essential (n = 159 tumor, 30 normal).
(G) Ecotype and genomic subtype enrichment of targets from (A, D, and F). Font indicates data type: bold, protein abundance; italics, kinase activity; regular, site-specific glycoforms. Up to 15 significant genes per category are shown (Fisher's exact test, FDR <0.05; n = 157 genomic subtypes, 151 ecotypes).
(H) Percent of cells expressing targets from (G) across 12 cell types from a published single-cell RNA dataset (n = 23). Also see Figure S7 and Table S7.

for AbbVie, Adaptive, Amgen, Apple, BMS/Celgene, GSK, Janssen, Karyopharm, Pfizer, RA Capital, Regeneron, and Sanofi; is a member of the scientific advisory board of Caris Life Sciences; is on the board of directors for Antengene; and has a patent for FISH in MM for ~$2,000/year. The graphical abstract and Figure 1 include icons generated from BioRender.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During preparation of this work, the authors used ChatGPT to improve clarity and subsequently reviewed and edited the text, taking full responsibility for the final content.

STAR★METHODS

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

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Human subjects
- METHOD DETAILS
  - Specimen acquisition
  - Sample homogenization and aliquoting for multi-omics analyses
  - Whole genome sequencing (WGS)
  - Whole exome sequencing (WES)
  - RNA sequencing
  - Ubiquitinome profiling (DIA-based)
  - Tandem mass tag based global proteomic and PTMs analyses
  - XL-MS sample preparation and data generation
  - Metabolomic profiling
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Harmonized genome alignment
  - Standardized reference for proteogenomics data processing
  - WGS data analysis
  - WES data analysis
  - XL-MS data processing
  - Inferred molecular phenotypes
  - Molecular subtype analysis
  - Significantly mutated genes
  - Data matrices used for analysis
  - Missing value exclusion criteria for proteomic based features
  - RNA-protein/PTM and protein-PTM correlation
  - Ecotype classification by consensus clustering
  - Analysis of previously published RNA datasets
  - Analysis of previously published single cell RNA-Seq dataset
  - Cell state and ecotype survival analysis
  - Clinical features associated with ecotypes
  - Ecotype prediction using XGBoost model
  - PredictIO analysis
  - Enhanced diffusion-based data pipeline for survivability prediction
  - XGBoost-based pipeline for metabolic regulator prioritization
  - Microbiome analysis
  - High outlier analysis
  - Pathway enrichment analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Additional resources

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2026.102756.

Received: September 22, 2025
Revised: January 19, 2026
Accepted: March 19, 2026

REFERENCES

1. Zeng, Y., and Jin, R.U. (2022). Molecular pathogenesis, targeted therapies, and future perspectives for gastric cancer. Semin. Cancer Biol. *86*, 566–582. https://doi.org/10.1016/j.semcancer.2021.12.004.
2. Sundar, R., Nakayama, I., Markar, S.R., Shitara, K., van Laarhoven, H.W.M., Janjigian, Y.Y., and Smyth, E.C. (2025). Gastric cancer. Lancet *405*, 2087–2102. https://doi.org/10.1016/S0140-6736(25)00052-2.
3. Thrift, A.P., Wenker, T.N., and El-Serag, H.B. (2023). Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention. Nat. Rev. Clin. Oncol. *20*, 338–349. https://doi.org/10.1038/s41571-023-00747-0.
4. Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209. https://doi.org/10.1038/nature13480.
5. Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.M., Ting, J.C., Wong, S.S., Liu, J., Yue, Y.G., Wang, J., Yu, K., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat. Med. *21*, 449–456. https://doi.org/10.1038/nm.3850.
6. Zhang, M., Hu, S., Min, M., Ni, Y., Lu, Z., Sun, X., Wu, J., Liu, B., Ying, X., and Liu, Y. (2021). Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. Gut *70*, 464–475. https://doi.org/10.1136/gutjnl-2019-320368.
7. Kim, J., Park, C., Kim, K.H., Kim, E.H., Kim, H., Woo, J.K., Seong, J.K., Nam, K.T., Lee, Y.C., and Cho, S.Y. (2022). Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. npj Precis. Oncol. *6*, 9. https://doi.org/10.1038/s41698-022-00251-1.
8. Kumar, V., Ramnarayanan, K., Sundar, R., Padmanabhan, N., Srivastava, S., Koiwa, M., Yasuda, T., Koh, V., Huang, K.K., Tay, S.T., et al. (2022). Single-Cell Atlas of Lineage States, Tumor Microenvironment, and Subtype-Specific Expression Programs in Gastric Cancer. Cancer Discov. *12*, 670–691. https://doi.org/10.1158/2159-8290.Cd-21-0683.
9. Sun, K., Xu, R., Ma, F., Yang, N., Li, Y., Sun, X., Jin, P., Kang, W., Jia, L., Xiong, J., et al. (2022). scRNA-seq of gastric tumor shows complex intercellular interaction with an alternative T cell exhaustion trajectory. Nat. Commun. *13*, 4943. https://doi.org/10.1038/s41467-022-32627-z.
10. Wang, R., Song, S., Qin, J., Yoshimura, K., Peng, F., Chu, Y., Li, Y., Fan, Y., Jin, J., Dang, M., et al. (2023). Evolution of immune and stromal cell states and ecotypes during gastric adenocarcinoma progression. Cancer Cell *41*, 1407–1426.e9. https://doi.org/10.1016/j.ccell.2023.06.005.
11. Ma, H., Srivastava, S., Ho, S.W.T., Xu, C., Lian, B.S.X., Ong, X., Tay, S.T., Sheng, T., Lum, H.Y.J., Abdul Ghani, S.A.B., et al. (2025). Spatially Resolved Tumor Ecosystems and Cell States in Gastric Adenocarcinoma Progression and Evolution. Cancer Discov. *15*, 767–792. https://doi.org/10.1158/2159-8290.Cd-24-0605.
12. Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., et al. (2018). A proteomic landscape of diffuse-type gastric cancer. Nat. Commun. *9*, 1012. https://doi.org/10.1038/s41467-018-03121-2.
13. Mun, D.G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., et al. (2019). Proteogenomic Characterization of Human Early-Onset Gastric Cancer. Cancer Cell *35*, 111–124.e10. https://doi.org/10.1016/j.ccell.2018.12.003.
14. Shi, W., Wang, Y., Xu, C., Li, Y., Ge, S., Bai, B., Zhang, K., Wang, Y., Zheng, N., Wang, J., et al. (2023). Multilevel proteomic analyses reveal molecular diversity between diffuse-type and intestinal-type gastric cancer. Nat. Commun. *14*, 835. https://doi.org/10.1038/s41467-023-35797-6.
15. Wang, Y., Jin, R.U., Xu, J., Lin, D.C., Sun, Z., Xu, Y., Li, Q.K., and Zhang, H. (2025). Harnessing technologies to unravel gastric cancer heterogeneity. Trends Cancer *11*, 753–769. https://doi.org/10.1016/j.trecan.2025.04.011.
16. Mertins, P., Tang, L.C., Krug, K., Clark, D.J., Gritsenko, M.A., Chen, L., Clauser, K.R., Clauss, T.R., Shah, P., Gillette, M.A., et al. (2018).

Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nat. Protoc. *13*, 1632–1661. https://doi.org/10.1038/s41596-018-0006-9.

17. Luca, B.A., Steen, C.B., Matusiak, M., Azizi, A., Varma, S., Zhu, C., Przybyl, J., Espín-Pérez, A., Diehn, M., Alizadeh, A.A., et al. (2021). Atlas of clinically distinct cell states and ecosystems across human solid tumors. Cell *184*, 5482–5496.e28. https://doi.org/10.1016/j.cell.2021.09.014.
18. Saito-Adachi, M., Hama, N., Totoki, Y., Nakamura, H., Arai, Y., Hosoda, F., Rokutan, H., Yachida, S., Kato, M., Fukagawa, A., and Shibata, T. (2023). Oncogenic structural aberration landscape in gastric cancer genomes. Nat. Commun. *14*, 3688. https://doi.org/10.1038/s41467-023-39263-1.
19. Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V., and Sullivan, M. (2012). PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res. *40*, D261–D270. https://doi.org/10.1093/nar/gkr1122.
20. Sun, S., Hu, Y., Ao, M., Shah, P., Chen, J., Yang, W., Jia, X., Tian, Y., Thomas, S., and Zhang, H. (2019). N-GlycositeAtlas: a database resource for mass spectrometry-based human N-linked glycoprotein and glycosylation site mapping. Clin. Proteomics *16*, 35. https://doi.org/10.1186/s12014-019-9254-0.
21. Zhang, W., Tan, X., Lin, S., Gou, Y., Han, C., Zhang, C., Ning, W., Wang, C., and Xue, Y. (2022). CPLM 4.0: an updated database with rich annotations for protein lysine modifications. Nucleic Acids Res. *50*, D451–D459. https://doi.org/10.1093/nar/gkab849.
22. Li, Y., Dou, Y., Da Veiga Leprevost, F., Geffen, Y., Calinawan, A.P., Aguet, F., Akiyama, Y., Anand, S., Birger, C., Cao, S., et al. (2023). Proteogenomic data and resources for pan-cancer analysis. Cancer Cell *41*, 1397–1406. https://doi.org/10.1016/j.ccell.2023.06.009.
23. Ungureanu, D., Saharinen, P., Junttila, I., Hilton, D.J., and Silvennoinen, O. (2002). Regulation of Jak2 through the ubiquitin-proteasome pathway involves phosphorylation of Jak2 on Y1007 and interaction with SOCS-1. Mol. Cell Biol. *22*, 3316–3326. https://doi.org/10.1128/mcb.22.10.3316-3326.2002.
24. Liu, X., Li, J., Deng, J., Zhao, J., Zhao, G., Zhang, T., Jiang, H., Liang, B., Xing, D., and Wang, J. (2023). Targeting Trop2 in solid tumors: a look into structures and novel epitopes. Front. Immunol. *14*, 1332489. https://doi.org/10.3389/fimmu.2023.1332489.
25. Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., Stark, C., Willems, A., Boucher, L., Leung, G., Kolas, N., Zhang, F., et al. (2021). The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci. *30*, 187–200. https://doi.org/10.1002/pro.3978.
26. Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., et al. (2021). Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell *184*, 3022–3040.e28. https://doi.org/10.1016/j.cell.2021.04.011.
27. Luck, K., Kim, D.K., Lambourne, L., Spirohn, K., Begg, B.E., Bian, W., Brignall, R., Cafarelli, T., Campos-Laborie, F.J., Charloteaux, B., et al. (2020). A reference map of the human binary protein interactome. Nature *580*, 402–408. https://doi.org/10.1038/s41586-020-2188-x.
28. Savage, S.R., Yi, X., Lei, J.T., Wen, B., Zhao, H., Liao, Y., Jaehnig, E.J., Somes, L.K., Shafer, P.W., Lee, T.D., et al. (2024). Pan-cancer proteogenomics expands the landscape of therapeutic targets. Cell *187*, 4389–4407.e15. https://doi.org/10.1016/j.cell.2024.05.039.
29. Bieging, K.T., Mello, S.S., and Attardi, L.D. (2014). Unravelling mechanisms of p53-mediated tumour suppression. Nat. Rev. Cancer *14*, 359–370. https://doi.org/10.1038/nrc3711.
30. Kang, J., Brajanovski, N., Chan, K.T., Xuan, J., Pearson, R.B., and Sanij, E. (2021). Ribosomal proteins and human diseases: molecular mechanisms and targeted therapy. Signal Transduct. Target. Ther. *6*, 323. https://doi.org/10.1038/s41392-021-00728-8.
31. Ola, R., Künzel, S.H., Zhang, F., Genet, G., Chakraborty, R., Pibouin-Fragner, L., Martin, K., Sessa, W., Dubrac, A., and Eichmann, A. (2018). SMAD4 Prevents Flow Induced Arteriovenous Malformations by Inhibiting Casein Kinase 2. Circulation *138*, 2379–2394. https://doi.org/10.1161/circulationaha.118.033842.
32. Hsu, T.Y.T., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Echeverria, G.V., Sun, T., Kurley, S.J., Tyagi, S., et al. (2015). The spliceosome is a therapeutic vulnerability in MYC-driven cancer. Nature *525*, 384–388. https://doi.org/10.1038/nature14985.
33. Nakayama, I., Qi, C., Chen, Y., Nakamura, Y., Shen, L., and Shitara, K. (2024). Claudin 18.2 as a novel therapeutic target. Nat. Rev. Clin. Oncol. *21*, 354–369. https://doi.org/10.1038/s41571-024-00874-2.
34. Zeng, Y., Lockhart, A.C., and Jin, R.U. (2024). The preclinical discovery and development of zolbetuximab for the treatment of gastric cancer. Expert Opin. Drug Discov. *19*, 873–886. https://doi.org/10.1080/17460441.2024.2370332.
35. Dottermusch, M., Krüger, S., Behrens, H.M., Halske, C., and Röcken, C. (2019). Expression of the potential therapeutic target claudin-18.2 is frequently decreased in gastric cancer: results from a large Caucasian cohort study. Virchows Arch. *475*, 563–571. https://doi.org/10.1007/s00428-019-02624-7.
36. Mathias-Machado, M.C., de Jesus, V.H.F., Jácome, A., Donadio, M.D., Aruquipa, M.P.S., Fogacci, J., Cunha, R.G., da Silva, L.M., and Peixoto, R.D. (2024). Claudin 18.2 as a New Biomarker in Gastric Cancer-What Should We Know? Cancers (Basel) *16*, 679. https://doi.org/10.3390/cancers16030679.
37. Ooi, C.H., Ivanova, T., Wu, J., Lee, M., Tan, I.B., Tao, J., Ward, L., Koo, J.H., Gopalakrishnan, V., Zhu, Y., et al. (2009). Oncogenic pathway combinations predict clinical prognosis in gastric cancer. PLoS Genet. *5*, e1000676. https://doi.org/10.1371/journal.pgen.1000676.
38. Oh, S.C., Sohn, B.H., Cheong, J.H., Kim, S.B., Lee, J.E., Park, K.C., Lee, S.H., Park, J.L., Park, Y.Y., Lee, H.S., et al. (2018). Clinical and genomic landscape of gastric cancer with a mesenchymal phenotype. Nat. Commun. *9*, 1777. https://doi.org/10.1038/s41467-018-04179-8.
39. Yu, N., Wang, Y., Olsen, L.K., Zhang, B., Zhang, h., and Liu, Z. (2025). TabSyM: A Generative Pipeline for Small Multi-Cohort Omics Tabular Data. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.14.664738.
40. Bareche, Y., Kelly, D., Abbas-Aghababazadeh, F., Nakano, M., Esfahani, P.N., Tkachuk, D., Mohammad, H., Samstein, R., Lee, C.H., Morris, L.G.T., et al. (2022). Leveraging big data of immune checkpoint blockade response identifies novel potential targets. Ann. Oncol. *33*, 1304–1317. https://doi.org/10.1016/j.annonc.2022.08.084.
41. Chen, W., Elizarraras, J.M., Jiao, F., Huang, L., and Zhang, B. (2025). XL-Ranker: A Computational Workflow for Prioritizing Protein-Protein Interactions from Cross-Linking Mass Spectrometry Data. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.18.665625.
42. Shi, Z., Wang, J., and Zhang, B. (2013). NetGestalt: integrating multidimensional omics data over biological networks. Nat. Methods *10*, 597–598. https://doi.org/10.1038/nmeth.2517.
43. Seo, S.K., and Kwon, B. (2023). Immune regulation through tryptophan metabolism. Exp. Mol. Med. *55*, 1371–1379. https://doi.org/10.1038/s12276-023-01028-7.
44. Din, Z.U., Cui, B., Wang, C., Zhang, X., Mehmood, A., Peng, F., and Liu, Q. (2025). Crosstalk between lipid metabolism and EMT: emerging mechanisms and cancer therapy. Mol. Cell. Biochem. *480*, 103–118. https://doi.org/10.1007/s11010-024-04995-1.
45. Wang, C., Elizarraras, J.M., and Zhang, B. (2025). MetaSage: Machine Learning-Based Prioritization of Metabolic Regulators from Multi-Omics Data. Preprint at bioRxiv. https://doi.org/10.1101/2025.07.09.663994.

46. Dolšak, A., Gobec, S., and Sova, M. (2021). Indoleamine and tryptophan 2,3-dioxygenases as important future therapeutic targets. Pharmacol. Ther. *221*, 107746. https://doi.org/10.1016/j.pharmthera.2020.107746.

47. Hennes, E., Lucas, B., Scholes, N.S., Cheng, X.F., Scott, D.C., Bischoff, M., Reich, K., Gasper, R., Lucas, M., Xu, T.T., et al. (2026). Monovalent Pseudo-Natural Products Supercharge Degradation of IDO1 by its native E3 KLHDC3 Nat. Chem. *18*, 585–596. https://doi.org/10.1038/s41557-025-02021-5.

48. Dohlman, A.B., Mendoza, D.A., Ding, S., Gao, M., Dressman, H., Iliev, I.D., Lipkin, S.M., and Shen, X. (2021). The cancer microbiome atlas: a pan-cancer comparative analysis to distinguish tissue-resident microbiota from contaminants. Cell Host Microbe *29*, 281–298.e285. https://doi.org/10.1016/j.chom.2020.12.001.

49. Gihawi, A., Ge, Y., Lu, J., Puiu, D., Xu, A., Cooper, C.S., Brewer, D.S., Pertea, M., and Salzberg, S.L. (2023). Major data analysis errors invalidate cancer microbiome findings. mBio *14*, e0160723–e0161623. https://doi.org/10.1128/mbio.01607-23.

50. Barko, P.C., McMichael, M.A., Swanson, K.S., and Williams, D.A. (2018). The Gastrointestinal Microbiome: A Review. J. Vet. Intern. Med. *32*, 9–25. https://doi.org/10.1111/jvim.14875.

51. Costea, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos, W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., et al. (2018). Enterotypes in the landscape of gut microbial community composition. Nat. Microbiol. *3*, 8–16. https://doi.org/10.1038/s41564-017-0072-8.

52. Yang, T.W., Lee, W.H., Tu, S.J., Huang, W.C., Chen, H.M., Sun, T.H., Tsai, M.C., Wang, C.C., Chen, H.Y., Huang, C.C., et al. (2019). Enterotype-based Analysis of Gut Microbiota along the Conventional Adenoma-Carcinoma Colorectal Cancer Pathway. Sci. Rep. *9*, 10923. https://doi.org/10.1038/s41598-019-45588-z.

53. Ignatova, E., Seriak, D., Fedyanin, M., Tryakin, A., Pokataev, I., Menshikova, S., Vakhabova, Y., Smirnova, K., Tjulandin, S., and Ajani, J.A. (2020). Epstein–Barr virus-associated gastric cancer: disease that requires special approach. Gastric Cancer *23*, 951–960. https://doi.org/10.1007/s10120-020-01095-z.

54. Lehr, K., Nikitina, D., Vilchez-Vargas, R., Steponaitiene, R., Thon, C., Skieceviciene, J., Schanze, D., Zenker, M., Malfertheiner, P., Kupcinskas, J., and Link, A. (2023). Microbial composition of tumorous and adjacent gastric tissue is associated with prognosis of gastric cancer. Sci. Rep. *13*, 4640. https://doi.org/10.1038/s41598-023-31740-3.

55. Liu, X., Shao, L., Liu, X., Ji, F., Mei, Y., Cheng, Y., Liu, F., Yan, C., Li, L., and Ling, Z. (2019). Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. EBioMedicine *40*, 336–348. https://doi.org/10.1016/j.ebiom.2018.12.034.

56. Hamada, T., Zhang, X., Mima, K., Bullman, S., Sukawa, Y., Nowak, J.A., Kosumi, K., Masugi, Y., Twombly, T.S., Cao, Y., et al. (2018). Fusobacterium nucleatum in Colorectal Cancer Relates to Immune Response Differentially by Tumor Microsatellite Instability Status. Cancer Immunol. Res. *6*, 1327–1336. https://doi.org/10.1158/2326-6066.CIR-18-0174.

57. Löwenmark, T., Li, X., Löfgren-Burström, A., Zingmark, C., Ling, A., Kellgren, T.G., Larsson, P., Ljuslinder, I., Wai, S.N., Edin, S., and Palmqvist, R. (2022). Parvimonas micra is associated with tumour immune profiles in molecular subtypes of colorectal cancer. Cancer Immunol. Immunother. *71*, 2565–2575. https://doi.org/10.1007/s00262-022-03179-4.

58. Younginger, B.S., Mayba, O., Reeder, J., Nagarkar, D.R., Modrusan, Z., Albert, M.L., and Byrd, A.L. (2023). Enrichment of oral-derived bacteria in inflamed colorectal tumors and distinct associations of Fusobacterium in the mesenchymal subtype. Cell Rep. Med. *4*, 100920. https://doi.org/10.1016/j.xcrm.2023.100920.

59. Abed, J., Emgård, J.E.M., Zamir, G., Faroja, M., Almogy, G., Grenov, A., Sol, A., Naor, R., Pikarsky, E., Atlan, K.A., et al. (2016). Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. Cell Host Microbe *20*, 215–225. https://doi.org/10.1016/j.chom.2016.07.006.

60. Lee, S.M., Donaldson, G.P., Mikulski, Z., Boyajian, S., Ley, K., and Mazmanian, S.K. (2013). Bacterial colonization factors control specificity and stability of the gut microbiota. Nature *501*, 426–429. https://doi.org/10.1038/nature12447.

61. Medley, B.J., Leclaire, L., Thompson, N., Mahoney, K.E., Pluvinage, B., Parson, M.A.H., Burke, J.E., Malaker, S., Wakarchuk, W., and Boraston, A.B. (2022). A previously uncharacterized O-glycopeptidase from Akkermansia muciniphila requires the Tn-antigen for cleavage of the peptide bond. J. Biol. Chem. *298*, 102439. https://doi.org/10.1016/j.jbc.2022.102439.

62. Boyaval, F., Dalebout, H., Van Zeijl, R., Wang, W., Fariña-Sarasqueta, A., Lageveen-Kammeijer, G.S.M., Boonstra, J.J., McDonnell, L.A., Wuhrer, M., Morreau, H., and Heijs, B. (2022). High-Mannose N-Glycans as Malignant Progression Markers in Early-Stage Colorectal Cancer. Cancers *14*, 1552. https://doi.org/10.3390/cancers14061552.

63. Zhang, S., Yang, Y., Weng, W., Guo, B., Cai, G., Ma, Y., and Cai, S. (2019). Fusobacterium nucleatum promotes chemoresistance to 5-fluorouracil by upregulation of BIRC3 expression in colorectal cancer. J. Exp. Clin. Cancer Res. *38*, 14. https://doi.org/10.1186/s13046-018-0985-y.

64. Kang, W., Jia, Z., Tang, D., Zhang, Z., Gao, H., He, K., and Feng, Q. (2019). Fusobacterium nucleatum Facilitates Apoptosis, ROS Generation, and Inflammatory Cytokine Production by Activating AKT/MAPK and NF-κB Signaling Pathways in Human Gingival Fibroblasts. Oxid. Med. Cell. Longev. *2019*, 1681972. https://doi.org/10.1155/2019/1681972.

65. Geng, F., Zhang, Y., Lu, Z., Zhang, S., and Pan, Y. (2020). Fusobacterium nucleatum Caused DNA Damage and Promoted Cell Proliferation by the Ku70/p53 Pathway in Oral Cancer Cells. DNA Cell Biol. *39*, 144–151. https://doi.org/10.1089/dna.2019.5064.

66. Fu, K., Cheung, A.H.K., Wong, C.C., Liu, W., Zhou, Y., Wang, F., Huang, P., Yuan, K., Coker, O.O., Pan, Y., et al. (2024). Streptococcus anginosus promotes gastric inflammation, atrophy, and tumorigenesis in mice. Cell *187*, 882–896.e17. https://doi.org/10.1016/j.cell.2024.01.004.

67. Saif, M.W., Makrilia, N., and Syrigos, K. (2010). CoFactor: Folate Requirement for Optimization of 5-Fluouracil Activity in Anticancer Chemotherapy. J. Oncol. *2010*, 934359. https://doi.org/10.1155/2010/934359.

68. Sethy, C., and Kundu, C.N. (2021). 5-Fluorouracil (5-FU) resistance and the new strategy to enhance the sensitivity against cancer: Implication of DNA repair inhibition. Biomed. Pharmacother. *137*, 111285. https://doi.org/10.1016/j.biopha.2021.111285.

69. Showalter, S.L., Showalter, T.N., Witkiewicz, A., Havens, R., Kennedy, E.P., Hucl, T., Kern, S.E., Yeo, C.J., and Brody, J.R. (2008). Evaluating the drug-target relationship between thymidylate synthase expression and tumor response to 5-fluorouracil. Is it time to move forward? Cancer Biol. Ther. *7*, 986–994. https://doi.org/10.4161/cbt.7.7.6181.

70. MacDonald, J.W., and Ghosh, D. (2006). COPA–cancer outlier profile analysis. Bioinformatics *22*, 2950–2951. https://doi.org/10.1093/bioinformatics/btl433.

71. Zhu, Y., Zhu, X., Wei, X., Tang, C., and Zhang, W. (2021). HER2-targeted therapies in gastric cancer. Biochim. Biophys. Acta. Rev. Cancer *1876*, 188549. https://doi.org/10.1016/j.bbcan.2021.188549.

72. Qiu, S., Zhang, J., Wang, Z., Lan, H., Hou, J., Zhang, N., Wang, X., and Lu, H. (2023). Targeting Trop-2 in cancer: Recent research progress and clinical application. Biochim. Biophys. Acta. Rev. Cancer *1878*, 188902. https://doi.org/10.1016/j.bbcan.2023.188902.

73. Tabernero, J., Bedard, P.L., Bang, Y.J., Vieito, M., Ryu, M.H., Fagniez, N., Chadjaa, M., Soufflet, C., Masson, N., and Gazzah, A. (2023). Tusamitamab Ravtansine in Patients with Advanced Solid Tumors: Phase I Study of Safety, Pharmacokinetics, and Antitumor Activity Using Alternative Dosing Regimens. Cancer Res. Commun. *3*, 1662–1671. https://doi.org/10.1158/2767-9764.Crc-23-0284.

74. Zhu, Y., Zhou, M., Kong, W., and Li, C. (2023). Antibody-drug conjugates: the clinical development in gastric cancer. Front. Oncol. *13*, 1211947. https://doi.org/10.3389/fonc.2023.1211947.

75. Kim, M., Pyo, S., Kang, C.H., Lee, C.O., Lee, H.K., Choi, S.U., and Park, C.H. (2018). Folate receptor 1 (FOLR1) targeted chimeric antigen receptor (CAR) T cells for the treatment of gastric cancer. PLoS One *13*, e0198347. https://doi.org/10.1371/journal.pone.0198347.

76. Lih, T.M., Cho, K.C., Schnaubelt, M., Hu, Y., and Zhang, H. (2023). Integrated glycoproteomic characterization of clear cell renal cell carcinoma. Cell Rep. *42*, 112409. https://doi.org/10.1016/j.celrep.2023.112409.

77. Pacini, C., Dempster, J.M., Boyle, I., Gonçalves, E., Najgebauer, H., Karakoc, E., van der Meer, D., Barthorpe, A., Lightfoot, H., Jaaks, P., et al. (2021). Integrated cross-study datasets of genetic dependencies in cancer. Nat. Commun. *12*, 1661. https://doi.org/10.1038/s41467-021-21898-7.

78. Voutsadakis, I.A. (2025). Development of CDK4/6 Inhibitors in Gastrointestinal Cancers: Biomarkers to Move Forward. Curr. Issues Mol. Biol. *47*, 454. https://doi.org/10.3390/cimb47060454.

79. Chao, J., Fuchs, C.S., Shitara, K., Tabernero, J., Muro, K., Van Cutsem, E., Bang, Y.J., De Vita, F., Landers, G., Yen, C.J., et al. (2021). Assessment of Pembrolizumab Therapy for the Treatment of Microsatellite Instability-High Gastric or Gastroesophageal Junction Cancer Among Patients in the KEYNOTE-059, KEYNOTE-061, and KEYNOTE-062 Clinical Trials. JAMA Oncol. *7*, 895–902. https://doi.org/10.1001/jamaoncol.2021.0275.

80. Nada, H., Choi, Y., Kim, S., Jeong, K.S., Meanwell, N.A., and Lee, K. (2024). New insights into protein-protein interaction modulators in drug discovery and therapeutic advance. Signal Transduct. Target. Ther. *9*, 341. https://doi.org/10.1038/s41392-024-02036-3.

81. Vasaikar, S.V., Straub, P., Wang, J., and Zhang, B. (2018). LinkedOmics: analyzing multi-omics data within and across 32 cancer types. Nucleic Acids Res. *46*, D956–D963. https://doi.org/10.1093/nar/gkx1090.

82. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B., et al. (2025). PubChem 2025 update. Nucleic Acids Res. *53*, D1516–D1525. https://doi.org/10.1093/nar/gkae1059.

83. Li, F., Chen, Y., Anton, M., and Nielsen, J. (2023). GotEnzymes: an extensive database of enzyme parameter predictions. Nucleic Acids Res. *51*, D583–D586. https://doi.org/10.1093/nar/gkac831.

84. Hu, Y., Schnaubelt, M., Chen, L., Zhang, B., Hoang, T., Lih, T.M., Zhang, Z., and Zhang, H. (2024). MS-PyCloud: A Cloud Computing-Based Pipeline for Proteomic and Glycoproteomic Data Analyses. Anal. Chem. *96*, 10145–10151. https://doi.org/10.1021/acs.analchem.3c01497.

85. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

86. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

87. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. *12*, R41. https://doi.org/10.1186/gb-2011-12-4-r41.

88. Garcia, M., Juhos, S., Larsson, M., Olason, P.I., Martin, M., Eisfeldt, J., DiLorenzo, S., Sandgren, J., Díaz De Ståhl, T., Ewels, P., et al. (2020). Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000Res. *9*, 63. https://doi.org/10.12688/f1000research.16665.2.

89. Hanssen, F., Garcia, M.U., Folkersen, L., Pedersen, A.S., Lescai, F., Jodoin, S., Miller, E., Seybold, M., Wacker, O., Smith, N., et al. (2024). Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. NAR Genom. Bioinform. *6*, lqae031. https://doi.org/10.1093/nargab/lqae031.

90. Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at arXiv. https://doi.org/10.48550/arXiv.1207.3907.

91. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat. Biotechnol. *31*, 213–219. https://doi.org/10.1038/nbt.2514.

92. Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Källberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P., and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. Nat. Methods *15*, 591–594. https://doi.org/10.1038/s41592-018-0051-x.

93. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122. https://doi.org/10.1186/s13059-016-0974-4.

94. Tan, V.Y.F., and Févotte, C. (2013). Automatic relevance determination in nonnegative matrix factorization with the β-divergence. IEEE Trans. Pattern Anal. Mach. Intell. *35*, 1592–1605. https://doi.org/10.1109/tpami.2012.240.

95. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. *38*, e164. https://doi.org/10.1093/nar/gkq603.

96. Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M.D., Wendl, M.C., and Ding, L. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics *30*, 1015–1016. https://doi.org/10.1093/bioinformatics/btt755.

97. Gao, Y., Wang, J., and Zhao, F. (2015). CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. Genome Biol. *16*, 4. https://doi.org/10.1186/s13059-014-0571-3.

98. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

99. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

100. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

101. Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat. Biotechnol. *42*, 293–304. https://doi.org/10.1038/s41587-023-01767-y.

102. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172. https://doi.org/10.1038/s41590-018-0276-y.

103. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat. Biotechnol. *37*, 773–782. https://doi.org/10.1038/s41587-019-0114-2.

104. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. *30*, 413–421. https://doi.org/10.1038/nbt.2203.

105. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast

universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

106. Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M.J., Blüthgen, N., and Saez-Rodriguez, J. (2018). Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat. Commun. *9*, 20. https://doi.org/10.1038/s41467-017-02391-6.

107. Sturm, G., Finotello, F., and List, M. (2020). Immunedeconv: An R Package for Unified Access to Computational Methods for Estimating Immune Cell Fractions from Bulk RNA-Sequencing Data. Methods Mol. Biol. *2120*, 223–232. https://doi.org/10.1007/978-1-0716-0327-7_16.

108. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinf. *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

109. Krug, K., Mertins, P., Zhang, B., Hornbeck, P., Raju, R., Ahmad, R., Szucs, M., Mundt, F., Forestier, D., Jane-Valbuena, J., et al. (2019). A Curated Resource for Phosphosite-specific Signature Analysis. Mol. Cell. Proteomics *18*, 576–593. https://doi.org/10.1074/mcp.TIR118.000943.

110. 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.

111. Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinović, S.M., Cheng, L.Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol. Cell. Proteomics *14*, 1400–1410. https://doi.org/10.1074/mcp.M114.044305.

112. Hsiao, Y., Zhang, H., Li, G.X., Deng, Y., Yu, F., Valipour Kahrood, H., Steele, J.R., Schittenhelm, R.B., and Nesvizhskii, A.I. (2024). Analysis and Visualization of Quantitative Proteomics Data Using FragPipe-Analyst. J. Proteome Res. *23*, 4303–4315. https://doi.org/10.1021/acs.jproteome.4c00294.

113. Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., and Nesvizhskii, A.I. (2017). MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat. Methods *14*, 513–520. https://doi.org/10.1038/nmeth.4256.

114. Yang, K.L., Yu, F., Teo, G.C., Li, K., Demichev, V., Ralser, M., and Nesvizhskii, A.I. (2023). MSBooster: improving peptide identification rates using deep learning-based features. Nat. Commun. *14*, 4539. https://doi.org/10.1038/s41467-023-40129-9.

115. The, M., MacCoss, M.J., Noble, W.S., and Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. J. Am. Soc. Mass Spectrom. *27*, 1719–1727. https://doi.org/10.1007/s13361-016-1460-7.

116. Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. *75*, 4646–4658. https://doi.org/10.1021/ac0341261.

117. Shteynberg, D.D., Deutsch, E.W., Campbell, D.S., Hoopmann, M.R., Kusebauch, U., Lee, D., Mendoza, L., Midha, M.K., Sun, Z., Whetton, A.D., and Moritz, R.L. (2019). PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. J. Proteome Res. *18*, 4262–4272. https://doi.org/10.1021/acs.jproteome.9b00205.

118. Yu, F., Haynes, S.E., and Nesvizhskii, A.I. (2021). IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. Mol. Cell. Proteomics *20*, 100077. https://doi.org/10.1016/j.mcpro.2021.100077.

119. da Veiga Leprevost, F., Haynes, S.E., Avtonomov, D.M., Chang, H.Y., Shanmugam, A.K., Mellacheruvu, D., Kong, A.T., and Nesvizhskii, A.I. (2020). Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat. Methods *17*, 869–870. https://doi.org/10.1038/s41592-020-0912-y.

120. Chang H.Y., Deng Y., Li R., Avtonomov D., Wen B., Haynes S.E., Leprevost F.D.V., Zhang B., Yu F., Nesvizhskii A.I. Analysis of isobaric quantitative proteomic data using TMT-Integrator and FragPipe computational platform. Nat. Commun. 2026, accepted. https://doi.org/10.1038/s41467-026-70118-7

121. Guan, S., Price, J.C., Prusiner, S.B., Ghaemmaghami, S., and Burlingame, A.L. (2011). A data processing pipeline for mammalian proteome dynamics studies using stable isotope metabolic labeling. Mol. Cell. Proteomics *10*, 010728. https://doi.org/10.1074/mcp.M111.010728.

122. Clauser, K.R., Baker, P., and Burlingame, A.L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. Anal. Chem. *71*, 2871–2882. https://doi.org/10.1021/ac9810516.

123. Tsitsiridis, G., Steinkamp, R., Giurgiu, M., Brauner, B., Fobo, G., Frishman, G., Montrone, C., and Ruepp, A. (2023). CORUM: the comprehensive resource of mammalian protein complexes-2022. Nucleic Acids Res. *51*, D539–D545. https://doi.org/10.1093/nar/gkac1015.

124. Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. *51*, D638–D646. https://doi.org/10.1093/nar/gkac1000.

125. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.P., and Wang, L. (2014). CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics *30*, 1006–1007. https://doi.org/10.1093/bioinformatics/btt730.

126. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. Nature *505*, 495–501. https://doi.org/10.1038/nature12912.

127. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.

128. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573. https://doi.org/10.1093/bioinformatics/btq170.

129. Chen, T., and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery).

130. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

131. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

132. Wood, D.E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biol. *20*, 257. https://doi.org/10.1186/s13059-019-1891-0.

133. Elizarraras, J.M., Liao, Y., Shi, Z., Zhu, Q., Pico, A.R., and Zhang, B. (2024). WebGestalt 2024: faster gene set analysis and new support for metabolomics and multi-omics. Nucleic Acids Res. *52*, W415–W421. https://doi.org/10.1093/nar/gkae456.

134. Therneau, T.M., and Grambsch, P.M. (2000). Modeling Survival Data: Extending the Cox Model (Springer).

135. Alboukadel Kassambara, M.K., P. Biecek (2024). Survminer: Drawing Survival Curves Using 'ggplot2'.

136. Dewey, M. (2025). Metap: Meta-Analysis of Significance Values.

137. Wen, B., Jaehnig, E.J., and Zhang, B. (2022). OmicsEV: a tool for comprehensive quality evaluation of omics data tables. Bioinformatics *38*, 5463–5465. https://doi.org/10.1093/bioinformatics/btac698.

138. Wang, Y., Lih, T.M., Lee, J.W., Ohtsuka, T., Hozaka, Y., Mino-Kenudson, M., Adsay, N.V., Luchini, C., Scarpa, A., Maker, A.V., et al. (2025). Multi-omic profiling of intraductal papillary neoplasms of the pancreas reveals distinct patterns and potential markers of progression. Cancer Cell *43*, 2156–2170.E6. https://doi.org/10.1016/j.ccell.2025.08.001.

139. Cao, L., Huang, C., Cui Zhou, D., Hu, Y., Lih, T.M., Savage, S.R., Krug, K., Clark, D.J., Schnaubelt, M., Chen, L., et al. (2021). Proteogenomic characterization of pancreatic ductal adenocarcinoma. Cell *184*, 5031–5052.e26. https://doi.org/10.1016/j.cell.2021.08.023.

140. Sun, Z., Lih, T.M., Woo, J., Jiao, L., Hu, Y., Wang, Y., Liu, H., and Zhang, H. (2024). Improving Glycoproteomic Analysis Workflow by Systematic Evaluation of Glycopeptide Enrichment, Quantification, Mass Spectrometry Approach, and Data Analysis Strategies. Anal. Chem. *96*, 20481–20490. https://doi.org/10.1021/acs.analchem.4c04466.

141. Jiao, F., Yu, C., Wheat, A., Chen, L., Lih, T.S.M., Zhang, H., and Huang, L. (2024). DSBSO-Based XL-MS Analysis of Breast Cancer PDX Tissues to Delineate Protein Interaction Network in Clinical Samples. J. Proteome Res. *23*, 3269–3279. https://doi.org/10.1021/acs.jproteome.3c00832.

142. Wheat, A., Yu, C., Wang, X., Burke, A.M., Chemmama, I.E., Kaake, R.M., Baker, P., Rychnovsky, S.D., Yang, J., and Huang, L. (2021). Protein interaction landscapes revealed by advanced in vivo cross-linking-mass spectrometry. Proc. Natl. Acad. Sci. USA *118*, e2023360118. https://doi.org/10.1073/pnas.2023360118.

143. Jiao, F., Yu, C., Wheat, A., Wang, X., Rychnovsky, S.D., and Huang, L. (2022). Two-Dimensional Fractionation Method for Proteome-Wide Cross-Linking Mass Spectrometry Analysis. Anal. Chem. *94*, 4236–4242. https://doi.org/10.1021/acs.analchem.1c04485.

144. Yu, C., Huszagh, A., Viner, R., Novitsky, E.J., Rychnovsky, S.D., and Huang, L. (2016). Developing a Multiplexed Quantitative Cross-Linking Mass Spectrometry Platform for Comparative Structural Analysis of Protein Complexes. Anal. Chem. *88*, 10301–10308. https://doi.org/10.1021/acs.analchem.6b03148.

145. Dou, Y., Katsnelson, L., Gritsenko, M.A., Hu, Y., Reva, B., Hong, R., Wang, Y.T., Kolodziejczak, I., Lu, R.J.H., Tsai, C.F., et al. (2023). Proteogenomic insights suggest druggable pathways in endometrial carcinoma. Cancer Cell *41*, 1586–1605.e15. https://doi.org/10.1016/j.ccell.2023.07.007.

146. Vromman, M., Anckaert, J., Bortoluzzi, S., Buratin, A., Chen, C.Y., Chu, Q., Chuang, T.J., Dehghannasiri, R., Dieterich, C., Dong, X., et al. (2023). Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. Nat. Methods *20*, 1159–1169. https://doi.org/10.1038/s41592-023-01944-6.

147. Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. Nucleic Acids Res. *47*, D155–D162. https://doi.org/10.1093/nar/gky1141.

148. Cui, S., Yu, S., Huang, H.Y., Lin, Y.C.D., Huang, Y., Zhang, B., Xiao, J., Zuo, H., Wang, J., Li, Z., et al. (2025). miRTarBase 2025: updates to the collection of experimentally validated microRNA-target interactions. Nucleic Acids Res. *53*, D147–D156. https://doi.org/10.1093/nar/gkae1072.

149. Müller-Dott, S., Tsirvouli, E., Vazquez, M., Ramirez Flores, R.O., Badia-I-Mompel, P., Fallegger, R., Türei, D., Lægreid, A., and Saez-Rodriguez, J. (2023). Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Res. *51*, 10934–10949. https://doi.org/10.1093/nar/gkad841.

150. Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P., Yaffe, M.B., and Pawson, T. (2008). NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res. *36*, D695–D699. https://doi.org/10.1093/nar/gkm902.

151. Müller-Dott, S., Jaehnig, E.J., Munchic, K.P., Jiang, W., Yaron-Barir, T.M., Savage, S.R., Garrido-Rodriguez, M., Johnson, J.L., Lussana, A., Petsalaki, E., et al. (2025). Comprehensive evaluation of phosphoproteomic-based kinase activity inference. Nat. Commun. *16*, 4771. https://doi.org/10.1038/s41467-025-59779-y.

152. Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines. Cell Syst. *6*, 271–281.e7. https://doi.org/10.1016/j.cels.2018.03.002.

153. Tsubosaka, A., Komura, D., Kakiuchi, M., Katoh, H., Onoyama, T., Yamamoto, A., Abe, H., Seto, Y., Ushiku, T., and Ishikawa, S. (2023). Stomach encyclopedia: Combined single-cell and spatial transcriptomics reveal cell diversity and homeostatic regulation of human stomach. Cell Rep. *42*, 113236. https://doi.org/10.1016/j.celrep.2023.113236.

154. Bosi, C., Bartha, Á., Galbardi, B., Notini, G., Naldini, M.M., Licata, L., Viale, G., Mariani, M., Pistilli, B., Ali, H.R., et al. (2023). Pan-cancer analysis of antibody-drug conjugate targets and putative predictors of treatment response. Eur. J. Cancer *195*, 113379. https://doi.org/10.1016/j.ejca.2023.113379.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Primary tumor samples	See Experimental Model and Subject Details	N/A
Chemicals, peptides, and recombinant proteins		
Sodium chloride	Santa Cruz Biotechnology	Catalog: sc-295833
Tris(hydroxymethyl)aminomethane (Tris)	Invitrogen	Catalog: AM9855G
Ethylenediaminetetraacetic acid (EDTA)	Sigma	Catalog: E7889
Aprotinin	Sigma	Catalog: A6103
Leupeptin	Roche	Catalog: 11017101001
Phenylmethylsulfonyl fluoride	Sigma	Catalog: 93482
Sodium fluoride	Sigma	Catalog: S7920
Phosphatase Inhibitor Cocktail 2	Sigma	Catalog: P5726
Phosphatase Inhibitor Cocktail 3	Sigma	Catalog: P0044
Urea	Sigma	Catalog: U0631
PUGNAc	Sigma	Catalog: A7229
Sodium butyrate	Sigma	Catalog: B5887-1g
SAHA	Sigma	Catalog: SML0061-5MG
DUB Inhibitor V, PR-619	Sigma	Catalog: 662141-25MG
Benzonase® Nuclease	Sigma	Catalog: E1014-25KU
Dithiothreitol	Fisher Scientific	Catalog: 20291
Iodoacetamide	Fisher Scientific	Catalog: A3221
Lysyl endopeptidase (Lys-C), Mass Spectrometry Grade	Wako Chemicals	Catalog: 125-05061
Sequencing grade modified trypsin	Promega	Catalog: V511X
Formic acid	Fisher Scientific	Catalog: A117-50
Sep-Pak tC18 96-well Plate, 100 mg Sorbent per Well	Waters	Catalog: 186002321
Oasis MAX 1 cc Vac Cartridge, 30 mg Sorbent per Cartridge	Waters	Catalog: 186000366
TMTpro 16plex label reagent set	Fisher Scientific	Catalog: A44520
TMTpro-134C & TMTpro-135N label reagents	Fisher Scientific	Catalog: A52046
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Alfa Aesar	Catalog: J63218
Acetonitrile, Optima LC/MS	Fisher Scientific	Catalog: A955-4
Water, Optima LC/MS	Fisher Scientific	Catalog: W6-4
Anhydrous acetonitrile	Sigma	Catalog: 271004
Hydroxylamine solution	Sigma	Catalog: 467804
Ammonium hydroxide solution	Sigma	Catalog: 338818
Ni^{2+}-NTA agarose beads	QIAGEN	Catalog: 30410
Iron (III) chloride	Sigma	Catalog:451649
Trifluoroacetic acid (TFA)	Sigma	Catalog: 302031
Triethylammonium acetate buffer	Sigma	Catalog: 90358
Phosphate-Buffered Saline (PBS)	Fisher Scientific	Catalog: 10010-023
Ammonium Bicarbonate	Sigma	Catalog: A6141-500G
Magnesium Chloride	Innovating Science	Catalog: IS22098-500G

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Alkyne-A-DSBSO	Sigma	Catalog: 909688
NP-40	Sigma	Catalog: I3021-100ML
Tris (2-carboxyethyl) phosphine (TECP)	Sigma	Catalog: C4706-2G
Biotin Picolyl Azide	Vector Laboratories	Catalog: 1167-5
BTTAA	Vector Laboratories	Catalog: 1236-100
Copper sulfate	Sigma	Catalog: C1297-100G
Sodium L-ascorbate	Sigma	Catalog: A7631-25G
Sodium Phosphate Monobasic	Fisher Scientific	Catalog: S369-3
Sodium Phosphate Dibasic Anhydrous	Fisher Scientific	Catalog: S374-1
1M Tetraethylammonium bromide (TEAB) TMT buffer	Fisher Scientific	Catalog: 90114
Tandem mass tags – 10plex	Fisher Scientific	Catalog: 90406
TMT11-131C label reagent	Fisher Scientific	Catalog: A34807
Methanol	Fisher Scientific	Catalog: A452-4
AMPure XP beads	Beckman Coultier	A63881
8-base dual index sequences within the p5 and p7 primers	IDT	N/A (Custom product)
8-base index sequences	Roche	Catalog: KK8727
NovaSeq Cluster Amp Reagents DPX1, DPX2, DPX3	Illumina	NovaSeq 6000 Reagent Kit (300 cycles): Catalog Number: 20012854 (Single-Pack) Catalog Number: 20012856 (Ten-Pack)
Critical commercial assays		
TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero Gold	Illumina	Catalog: RS-122-2301
Infinium MethylationEPIC Kit	Illumina	Catalog: WG-317-1003
Nextera DNA Exosome Kit	Illumina	Catalog: 20020617
KAPA Hyper Prep Kit, PCR-free	Roche	Catalog: 07962371001
BCA Protein Assay Kit	Fisher Scientific	Catalog: 23225
PTMScan HS Acetl-Lysine Motif (Ac-K) Kit	Cell Signaling Technologies	Catalog: 46784S
PTMScan HS Ubiquitin/SUMO Remnant Motif (K-epsilon-GG) Kit	Cell Signaling Technologies	Catalog: 59322
MethylationEPIC Array Kits	Zymo Research	EZ-96 DNA Methylation Kit
MethylationEPIC_8v2_A1 Beadchip Array Kit	Illumina	20087709
XGen Hybridization and Wash Kit	IDT	1080584
Quant-It dsDNA quantification assay kit	Fisher Scientific	Catalog Q33130
KAPA HyperPrep Kit with Library Amplification product	KAPA	Catalog: KK8504
Library Quantification Kits	KAPA Biosystems	Catalog: 7960204001
KAPA Hyper Prep without Amplification Module	KAPA Biosystems	Catalog: KK8505
Sequencing-by-Synthesis Kits	Illumina	HiSeq X SBS Kit (300 cycles): Catalog Number: FC-501-2501 (Single-Pack) Catalog Number: FC-501-2521 (Ten-Pack) HiSeq X SBS Kit (500 cycles): Catalog Number: FC-502-2501 (Single-Pack) Catalog Number: FC-502-2521 (Ten-Pack)
HiSeq X Reagent PE Cluster Kit v2	Illumina	HiSeq X Ten Reagent Kit v2.5 (300 cycles): Single-Pack: FC-501-2501 Ten-Pack: FC-501-2521 HiSeq X Five Reagent Kit v2.5 (300 cycles) Single-Pack: FC-502-2501 Ten-Pack: FC-502-2521

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Stranded Total RNA Prep with RiboZero Gold	Illumina	20020599
NEXTflex Small RNA_seq Kit	Revvity	NOVA-5132-41
Deposited data		
Proteomic Data Commons	National Institutes of Health	PDC: https://pdc.cancer.gov/pdc/Under under accession number PDC000614, PDC000615, PDC000616, PDC000622, PDC000621, PDC000626, PDC000617
Genomic Data Commons	National Institutes of Health	GDC: https://gdc.cancer.gov/
The Cancer Imaging Archive	National Institutes of Health	TCIA: https://doi.org/10.7937/jw9a-8k71
LinkedOmics	Vasaikar et al.[81]	https://www.linkedomics.org/#/
PhosphositePlus	Hornbeck et al.[19]	https://www.phosphosite.org/homeAction.action
GENCODE V42	ENCODE consortium	https://www.gencodegenes.org/human/release_42.html
GSE15459	Ooi et al.[37]	https://www.ncbi.nlm.nih.gov/geo/
GSE26899	Oh et al.[38]	https://www.ncbi.nlm.nih.gov/geo/
GSE26901	Oh et al.[38]	https://www.ncbi.nlm.nih.gov/geo/
GSE62254	Cristescu et al.[5]	https://www.ncbi.nlm.nih.gov/geo/
Stomach Adenocarcinoma (TCGA, PanCancer Atlas)	TCGA[4]	https://www.cbioportal.org/
PubChem	Kim et al.[82]	https://pubchem.ncbi.nlm.nih.gov/
DepMap: CRISPR KO screen (combined)	DepMap Public 22Q4	https://depmap.org/portal/download/
Antibody Society	Antibody Society	https://www.antibodysociety.org/
Therapeutic target tiers	Savage et al.[28]	Table S2A in Savage et al.[28]
DepMap: Pan-cancer essential genes	DepMap Public 21Q4[77]	https://depmap.org/portal/download/
Metabolic Atlas (1.14.0)	Li et al.[83]	https://metabolicatlas.org/
Software and algorithms		
MS-PyCloud	Hu et al.[84]	https://github.com/huizhanglab-jhu/ms-pycloud
Samtools v1.2, v1.16.1	Li et al.[85]	https://www.htslib.org/
GATK v4.4.0.0, 4.0.5.1, and 4.1.4.1,	McKenna et al.[86]	https://github.com/broadgsa/gatk
GISTIC2 v2.0	Mermel et al.[87]	https://github.com/broadinstitute/gistic2
Sarek v3.4.0	Garcia et al.,[88] Hanssen et al.[89]	https://github.com/nf-core/sarek
FreeBayes v1.3.6	Garrison et al.[90]	https://github.com/freebayes/freebayes
MuTect v4.4.0.0	Cibulskis et al.[91]	https://github.com/broadinstitute/mutect
Strelka2 v2.9.10	Kim et al.[92]	https://github.com/Illumina/strelka
VEP v111.0	McLaren et al.[93]	https://github.com/Ensembl/ensembl-vep
Vcf2maf v1.6.21		https://github.com/mskcc/vcf2maf
Sig Profiler Matrix Generator R v1.0	Tan et al.[94]	https://github.com/AlexandrovLab/SigProfilerMatrixGeneratorR
annovar v04.16.2018	Want et al.[95]	https://annovar.openbioinformatics.org/en/latest/
MSIsensor	Niu et al.[96]	https://github.com/ding-lab/msisensor
CIRI v2.0.6	Gao et al.[97]	https://sourceforge.net/projects/ciri/
BWA v0.7.17-r1188	Li et al.[98]	http://bio-bwa.sourceforge.net/
RSEM v1.3.1	Li et al.[99]	https://deweylab.github.io/RSEM/
Bowtie2 v2.3.3	Langmead et al.[100]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Seurat v5	Hao et al.[101]	https://satijalab.org/seurat/
SingleR	Aran et al.[102]	https://github.com/dviraran/SingleR
EcoTyper	Luca et al.[17]	https://github.com/digitalcytometry/ecotyper

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CibersortX	Newman et al.[103]	https://cibersortx.stanford.edu/
DoAbsolute v2.2	Carter et al.[104]	https://github.com/ShixiangWang/DoAbsolute
ABSOLUTE v1.0.6	Carter et al.[104]	https://github.com/getzlab/ABSOLUTE
STAR v2.7.10a, v2.7.5b	Dobin et al.[105]	https://github.com/alexdobin/STAR
Picard v1.92	Broad Institute	https://github.com/broadinstitute/picard
PROGENy	Schubert et al.[106]	https://bioconductor.org/packages/release/bioc/html/progeny.html
immunedeconv v2.0.4	Sturm et al.[107]	https://github.com/omnideconv/immunedeconv
GSVA (v1.30.0)	Hänzelmann et al.[108]	https://bioconductor.org/packages/release/bioc/html/GSVA.html
ssGSEA2.0/PTM-SEA	Krug et al.[109]	https://github.com/broadinstitute/ssGSEA2.0
ESTIMATE	Yoshihara et al.[110]	https://bioinformatics.mdanderson.org/public-software/estimate/
Spectronaut v18.4	Bruderer et al.[111]	https://biognosys.com/software/spectronaut/
FragPipe v21.1	Hsiao et al.[112]	https://fragpipe.nesvilab.org/
MSFragger v4.0	Kong et al.[113]	https://msfragger.nesvilab.org/
MSBooster	Yang et al.[114]	https://github.com/Nesvilab/MSBooster
Percolator	The et al.[115]	https://github.com/percolator/percolator
ProteinProphet	Nesvizhskii et al.[116]	http://tools.proteomecenter.org/wiki/index.php?title=Main_Page
PTMProphet	Shteynberg et al.[117]	http://tools.proteomecenter.org/wiki/index.php?title=Main_Page
IonQuant v1.10.12	Yu et al.[118]	https://ionquant.nesvilab.org/
Philosopher v5.1.0	da Veiga Leprevost et al.[119]	https://philosopher.nesvilab.org/
TMT-Integrator-v1.0.10	Chang et al.[120]	http://tmt-integrator.nesvilab.org/
GPQuest v3.0	Hu et al.[84]	https://github.com/huizhanglab-jhu/ms-pycloud
PAVA	Guan et al.[121]	https://www.ncbi.nlm.nih.gov/pubmed/21937731
Protein Prospector	Clauser et al.[122]	https://prospector.ucsf.edu/prospector/mshome.htm
BioPlex	Huttlin et al.[26]	https://bioplex.hms.harvard.edu/index.php
BioGrid	Oughtred et al.[25]	https://thebiogrid.org
CORUM	Tsitsiridis et al.[123]	http://mips.helmholtz-muenchen.de/corum
STRING	Szklarczyk et al.[124]	https://string-db.org/
CrossMap v0.7.0	Zhao et al.[125]	http://crossmap.sourceforge.net/.
MutSig2CV v3.11	Lawrence et al.[126]	https://github.com/getzlab/MutSig2CV
Maftools v2.22	Mayakonda et al.[127]	https://github.com/PoisonAlien/maftools
ConsensusClusterPlus	Wilkerson et al.[128]	https://www.bioconductor.org/packages/release/bioc/html/ConsensusClusterPlus.html
XGBoost 2.0.2	Chen et al.[129]	https://xgboost.readthedocs.io/en/latest/
SAMtools	Danecek et al.[130]	https://github.com/samtools/
BEDTools v2.30.0	Quinlan et al.[131]	https://github.com/arq5x/bedtools2
Kracken2 v2.1.3	Wood et al.[132]	https://github.com/DerrickWood/kraken2
Fastp v0.23.4		https://github.com/OpenGene/fastp
Webgestalt 2024	Elizarraras et al.[133]	http://www.webgestalt.org/
NA12878 HapMap Controls	Coriell Institute	NA12878
Illumina GenomeStudio Software	Illumina	1.0
PredictIO	Bareche et al.[40]	https://predictio.ca/
Survival v3.8-3	Therneau et al.[134]	https://github.com/therneau/survival
Survminer	Kassambara et al.[135]	https://github.com/kassambara/survminer

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Metap	Dewey Michael[136]	https://cran.r-project.org/web/packages/metap/index.html
OmicsEV	Wen et al.[137]	https://github.com/bzhanglab/OmicsEV
TabSyM	Yu et al.[39]	https://github.com/Froot-NetSys/TabSyM Protected until publication, PAT: ghp_HPSbcFO36RYdEzmC2mLg YVys84v9SN1I6Kg4
MetaSage	Wang et al.[45]	https://github.com/bzhanglab/metasage
XL-Ranker	Chen et al.[41]	https://github.com/bzhanglab/xlranker/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human subjects

A total of 165 consented participants (99 males, 66 females aged 32–87 years) were prospectively recruited for this study by seven different tissue source sites (TSS) in Europe. Institutional review boards at the TSS reviewed protocols and consent documentation in accordance with the Clinical Proteomic Tumor Analysis Consortium (CPTAC) guidelines for study participation. Clinical data were obtained from TSS, quality checked and aggregated into an internal database called the CDR (Comprehensive Data Resource). Demographics, diagnosis, exposure, histopathologic information, treatment and patient outcome information were checked for quality and consistency before deposition into the Proteomic Data Commons (PDC) and Genomic Data Commons (GDC). All histologic and radiologic images can be accessed from The Cancer Imaging Archive (TCIA) Public Access portal. Samples with high blood content were removed from further analyses. The genotypic, clinical, geographical and other associated metadata is summarized in Table S1.

METHOD DETAILS

Specimen acquisition

The tumor, NAT, and whole blood samples used in this study were prospectively collected for the CPTAC project. Treatment naive patients scheduled for surgical treatment of a tumor mass assumed to be a primary stomach adenocarcinoma with different anatomical locations were recruited. Patients who underwent cancer treatment more than ten years prior were included if the cancer was at a site other than the stomach. Only histopathologically defined adenocarcinomas of the subtypes: papillary, tubular, poorly cohesive-signet ring cell and mucinous were considered for analysis. Forty-one out of the 165 tumor samples had matched NATs that met the criteria for classification as acceptable normal tissue. For most individuals, three to four tumor specimens were collected. Each tissue specimen endured cold ischemia for less than 30 min prior to freezing in liquid nitrogen; the average ischemic time was 20 min from resection/collection to freezing. Specimens were flash-frozen in liquid nitrogen. Histologic sections obtained from top and bottom portions from each case were reviewed by a board-certified pathologist and a disease specific expert pathologist to confirm the assigned pathology. Samples with a tumor nuclei content greater than 60% were deemed acceptable. For mucinous gastric cancers, a rare histologic subtype, tumor nuclei content less than 60% was deemed acceptable given the acellular composition of this subtype. The top and bottom sections had to contain an average of less than 20% necrosis. Specimens were shipped overnight from the TSS to the biospecimen core resource (BCR) located at Van Andel Research Institute, Grand Rapids, MI, U.S., using a cryoport that maintained an average temperature of less than −140°C.

Sample homogenization and aliquoting for multi-omics analyses

At the biospecimen core resource, specimens were confirmed for pathology qualification and prepared for genomic, transcriptomic, proteomic, metabolomic, and PPI analyses. Selected specimens were cryopulverized using a Covaris CryoPREP instrument and material aliquoted for subsequent molecular characterization. Genomic DNA and total RNA were extracted and sent to the genome sequencing centers. The whole exome and whole genome DNA sequencing and methylation EPIC array analyses were performed at the Broad Institute, Cambridge. Total RNA and miRNA sequencing were performed at the University of North Carolina, Chapel Hill, NC. Material for proteomic and PTM analyses was sent to the Proteomic Characterization Center (PCC) at the Johns Hopkins University, Baltimore, MD. Metabolomic profiling was conducted at Gigantest, Baltimore. PPI characterizations were carried out at the University of California, Irvine.

Whole genome sequencing (WGS)

Preparation of libraries for the cluster amplification and sequencing

An aliquot of genomic DNA (350ng in 50μL) was used as the input into DNA fragmentation (also known as shearing). Shearing was performed acoustically using a Covaris focused-ultrasonicator, targeting 385bp fragments. Following fragmentation, additional size

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

selection was performed using an SPRI cleanup. Library preparation was performed using a commercially available kit provided by KAPA Biosystems (KAPA Hyper Prep without amplification module, product KK8505), and with palindromic forked adapters with unique 8-base index sequences embedded within the adapter (purchased from Roche). Following sample preparation, libraries were quantified using quantitative PCR (kit purchased from KAPA Biosystems), with probes specific to the ends of the adapters. This assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 2.2 nM and pooled into 24-plexes.

Cluster amplification and sequencing (HiSeqX and NovaSeq 6000)

Sample pools were combined with HiSeqX Cluster Amp Reagents EPX1, EPX2 and EPX3 into single wells on a strip tube using the Hamilton Starlet Liquid Handling system. Cluster amplification of the templates was performed according to the manufacturer's protocol (Illumina) with the Illumina cBot. Flowcells were sequenced on HiSeqX utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads. Output from Illumina software was processed by the Picard data-processing pipeline to yield CRAM or BAM files containing demultiplexed, aggregated aligned reads. All sample information tracking was performed by automated LIMS messaging. Sample pools were combined with NovaSeq Cluster Amp Reagents DPX1, DPX2 and DPX3 and loaded into single lanes of a NovaSeq 6000 S4 flowcell cell using the Hamilton Starlet Liquid Handling system. Cluster amplification and sequencing occured on NovaSeq 6000 Instruments utilizing sequencing-by-synthesis kits to produce 151bp paired-end reads. Output from Illumina software was processed by the Picard data-processing pipeline to yield CRAM or BAM files containing demultiplexed, aggregated aligned reads. All sample information tracking was performed by automated LIMS messaging.

Whole exome sequencing (WES)

Library construction

An aliquot of genomic DNA (100–250 ng in 50μL) was used as the input into DNA fragmentation (also known as shearing). Shearing is performed acoustically using a Covaris focused-ultrasonicator, targeting 150bp fragments. Library preparation is performed using a commercially available kit provided by KAPA Biosystems (KAPA HyperPrep Kit with Library Amplification product KK8504) and IDT's duplex UMI adapters. Unique 8-base dual index sequences embedded within the p5 and p7 primers (purchased from IDT) are added during PCR. Enzymatic clean-ups are performed using Beckman Coultier AMPure XP beads with elution volumes reduced to 30μL to maximize library concentration. Following library construction, library quantification is performed using the Invitrogen Quant-It broad range dsDNA quantification assay kit (Thermo Scientific Catalog: Q33130) with a 1:200 PicoGreen dilution. Following quantification, each library is normalized to a concentration of 35 ng/μL, using Tris-HCl, 10mM, pH 8.0. All steps performed during the library construction process and library quantification process are performed on the Agilent Bravo liquid handling system.

In-solution hybrid selection

After library construction, hybridization and capture are performed using the relevant components of IDT's XGen hybridization and wash kit and following the manufacturer's suggested protocol, with several exceptions. A set of 12-plex pre-hybridization pools are created. These pre-hybridization pools are created by equivolume pooling of the normalized libraries, Human Cot-1 and IDT XGen blocking oligos. The pre-hybridization pools undergo lyophilization using the Biotage SPE-DRY. Post lyophilization, custom exome bait (TWIST Biosciences) along with hybridization mastermix is added to the lyophilized pool prior to resuspension. Samples are incubated overnight. Library normalization and hybridization setup are performed on a Hamilton Starlet liquid handling platform, while target capture is performed on the Agilent Bravo automated platform. Post capture, a PCR is performed to amplify the capture material.

Preparation of libraries for cluster amplification and sequencing

After post-capture enrichment, library pools are quantified using qPCR (automated assay on the Agilent Bravo), using a kit purchased from KAPA Biosystems with probes specific to the ends of the adapters. Based on qPCR quantification, pools are normalized using a Hamilton Starlet to the required loading concentration. Up to 24 samples are sequenced per lane on Illumina's NovaSeq S4 sequencing technology.

Cluster amplification and sequencing

Cluster amplification of library pools was performed according to the manufacturer's protocol (Illumina) using Exclusion Amplification cluster chemistry and NovaSeq S4 flowcells. Flowcells were sequenced on Sequencing-by-Synthesis chemistry for NovaSeq S4 flowcells using paired 151bp runs.

DNA methylation profiling

The Illumina EPIC V2 methylation profiling array provides coverage of CpG islands, genes, and enhancers. We used the Zymo chemistry products to convert methylated cytosine regions of DNA to uracil through a 2-day bisulfite conversion where unmethylated cytosine remains. Then, sample processing was finished using Infinium MethylationEPIC using the iScan scanner systems and the GenomeStudio analysis software; the level of methylation was determined by measuring the level of fluorescence on methylated versus non-methylated sites per sample on each chip.

RNA sequencing

QA/QC of RNA analytes

Upon receipt, flash frozen GC RNA samples were assayed for RNA integrity, concentration, and fragment size. Samples for total RNA-Seq were quantified on a TapeStation system (Agilent, Inc. Santa Clara, CA) and RNA Integrity score (RIN) calculated. We targeted input concentrations greater than 500 ng/μL, although some samples were lower.

Total RNA-Seq library construction

For total RNA-Seq library construction we used Illumina Stranded Total RNA Prep with RiboZero Gold and bar-coded with individual tags following the manufacturer's instructions (Illumina, Inc. San Diego, CA). Libraries were prepared and then 24–92 samples were pooled on an automated liquid handling system to minimize variance. Quality control was performed at every step. Libraries were quantified for concentration, fragment size and distribution using a TapeStation system.

Total RNA sequencing

Indexed libraries were prepared and run on an Illumina NovaSeq 6000, paired end 100 base pairs to generate a minimum of 150 million reads per sample library with a target of greater than 90% mapped reads. In all but a few cases, all data was from the same sequencing run. A few samples needed additional read depth, which was provided with a secondary sequencing run. The raw Illumina sequence data were demultiplexed and converted to fastq files; adapter and low-quality sequences were quantified. Samples were assessed for informational quality by mapping reads to human genome reference (hg38), estimating total number of reads that mapped, amount of RNA mapping to coding regions, amount of rRNA in sample, number of genes expressed, and relative expression of housekeeping genes. The samples passing this QA/QC were then bioinformatically clustered with other expression data from similar and distinct tumor types to confirm expected expression patterns. The clustering results were subjected to expert review. Atypical samples were SNP typed to confirm source analyte. FASTQ files of all reads were then uploaded to the GDC repository and distributed to the analysis team.

miRNA sequencing

miRNA-Seq library construction used the NEXTflex Small RNA-Seq Kit (v3, PerkinElmer, Waltham, MA). Samples were barcoded with individual tags following the manufacturer's instructions. Libraries were prepared on Sciclone Liquid Handling Workstation. Quality control was performed at every step, and the libraries were quantified using a TapeStation system and an Agilent Bioanalyzer using the Small RNA analysis kit. Pooled libraries were then size selected according to NEXTflex Kit specifications using a Pippin Prep (Sage Science, Beverly, MA. Samples were then sequenced on an Illumina NextSeq2000 at 1 × 50 bp. Post-sequencing QA/QC evaluated the abundance and diversity of miRNA and the number of informative reads. After assessment it was determined whether top-off sequencing was required. All sample data that passed was then merged and provided to the GDC.

Protein digestion

The general procedure of tissue lysis sample preparation followed previous publications.[16] Each of cryo-pulverized GC tissue and NAT was lyzed in 8M urea lysis buffer (75 mM NaCl, 50 mM Tris (pH 8.0), 1 mM EDTA, 2 μg/mL aprotinin, 10 μg/mL leupeptin, 1 mM PMSF, 10 mM NaF, Phosphatase Inhibitor Cocktail 2 and Phosphatase Inhibitor Cocktail 3 [1:100 dilution], 20 μM PUGNAc, 10 mM Sodium Butyrate, 2 μM SAHA, 50 μM PR-619, and 500 units/mL Benzonase Nuclease) by vortexing for 20 s. The lyzed solution was kept on ice for 15 min. This procedure was repeated three times. Insoluble cell debris was removed by centrifugation at 16,000 x g for 15 min at 4°C. The protein solution was collected and measured by BCA assay (Pierce). The protein concentration was adjusted to 8 mg/mL with the lysis buffer, and 200 μL protein solution was used for further digestion reactions, including the steps of reduction, alkylation and digestion. Proteins were reduced and alkylated with dithiothreitol (6 mM, 37°C, 1h) and iodoacetamide (12 mM, room temperature (r.t.) in dark, 45 min), respectively. The alkylated proteins were diluted 1:3 with 50 mM Tris-HCl (pH 8.0) to reduce urea concentration to 2M and digested with LysC (enzyme to substrate ratio of 1 mAU:50 mg protein, r.t., 2h), followed by trypsin (enzyme to substrate ratio of 1:50, r.t., overnight). The proteolytic reaction was quenched by 50% of formic acid to adjust pH to less than 3. The digested solution was centrifuged at 16,000 x g for 15 min at 4°C, and the supernatant was desalted on Sep-Pak tC18 96-well plate. The eluted peptides were dried using Speed-Vac (Thermo Scientific).

Ubiquitinome profiling (DIA-based)

Enrichment of ubiquitinated peptides

A total of 300 μg of peptides (quantified by peptide-level BCA assay) of each sample was dissolved in 120 μL of 100 mM HEPES buffer (pH 8.5). Thirteen pooled peptide samples (300 μg of each sample) were generated by aliquoting from 165 GC tissues and 41 NATs for future TMTpro reference channel and underwent the enrichment process as well. The antibody-bead slurry was briefly spun and re-suspended to ensure uniformity, and 42.5 μL was transferred to a 1.7 mL microcentrifuge tube. After washing the beads twice with ice-cold 1× PBS, they were resuspended in 800 μL PBS and evenly aliquoted (100 μL each) into eight PCR tubes. The beads were then incubated with 300 μg of peptides per tube on an end-over-end rotator at 4°C for 2 h. Following incubation, the unbound peptides were collected for future TMTpro labeling. Beads were washed twice with 1× PBS and twice with IAP wash buffer. Bound peptides were eluted with 50 μL of IAP Elution Buffer (0.15% TFA) at room temperature while mixing, and this elution was repeated three times. The combined eluates were then loaded onto Evotip (Evosep Biosystems) according to the EvosepOne (Evosep Biosystems) protocol for ubiquitylation data independent analysis (DIA) analysis on timsTOF-HT (Bruker).

LC-MS/MS for ubiquitinated peptides DIA analysis

The LC-MS/MS system was built by EvosepOne-timsTOF-HT coupled system according to our previous publication.[138] Individual ubiquitinated peptide samples were loaded on Evotip and used 30 SPD LC method for seperation in Bruker column toaster (50°C) with PepSep C18 column of 15 cm × 150 μm, 1.5 μm (Bruker). The ubiquitinated peptide data was generated using the DIA-PASEF mode on timsTOF-HT with setting as: MS1 scan range of 100–1700 *m/z*; MS2 scan range of 342 (1/K0 range 0.7–0.8) - 1217 (1/K0 range 1.12–1.41)*m/z*, mass width 25.0 Da without mass overlap, 1 MS1 ramps, 20 MS/MS ramps, 1/K0 range of 0.70–1.45 V·s/cm^2, ramp time 85.0 ms.

Tandem mass tag based global proteomic and PTMs analyses

Tandem mass tag reagent TMTpro labeling of peptides

The unbound peptides from ubiquitylation enrichment of each sample were labeled with 18-plex TMTpro reagents (TMTpro 16plex label reagent set with TMTpro-134C & TMTpro-135N label reagents). The pooled samples were included in 13 TMTpro 18-plex sets as reference channels. All GC tissues and NATs were co-randomized to 13 TMTpro 18-plex sets. TMTpro reagents were dissolved in 500 μL anhydrous acetonitrile, and 40 μL of each channel TMTpro reagent was added to the corresponding aliquot of peptides. The labeling reaction was shaken at r.t. for 1h and quenched by 5% hydroxylamine at r.t. for 15 min. The TMTpro labeled peptides were desalted on Sep-Pak tC18 96-well plate and dried by Speed-Vac.

Basic reversed phase liquid chromatography (bRPLC) of peptide fractionation and sample preparation for global proteomic analyses

Each TMTpro set of the desalted labeled peptides were dissolved in 900 μL of 5 mM ammonium formate solution (pH 10) with 2% acetonitrile and fractionated onto a 4.6 mm × 250 mm Zorbax Extend-C18 analytical column with 3.5 μm beads (Agilent) lined up with an Agilent 1220 Series HPLC. Buffer A was 5 mM ammonium formate in 2% acetonitrile (pH 10), and buffer B was 5 mM ammonium formate in 90% acetonitrile. Peptides were separated and collected by using a non-linear gradient with 1 mL/minute: 0% buffer B (7 min), 0%–16% buffer B (6 min), 16%–40% buffer B (60 min), 40%–44% buffer B (4 min), 44%–60% buffer B (5 min), and then held at 60% buffer B for 14 min. The collected fractions were combined into 24 fractions as previously work,[16,139] and 2% of the each combined 24 fractions were aliquoted and dried for global proteomic analysis. The dried 24 samples for each set were resuspended in 3% acetonitrile containing 0.1% formic acid water solution and loaded on Evotip, analyzing by Evosep-Ascend system. The remaining sample was used for further enrichment.

Phosphopeptide enrichment

The remaining samples were further combined into 12 fractions before the phosphopeptide enrichment as previously described.[16,139] The enrichment was conducted by immobilized metal affinity chromatography (IMAC). In brief, Ni^{2+}-NTA agarose beads were conditioned and incubated with 10 mM $FeCl_3$ aqueous solution at r.t. for 1h to prepare Fe^{3+}-NTA agarose beads. Peptides from each fraction were dissolved in 80% acetonitrile containing 0.1% trifluoroacetic acid (TFA) and were incubated with 10 μL of the Fe^{3+}-IMAC beads for 30 min. Samples were then spun down at 1,000 x g for 1 min and the supernatant containing unbound peptides was collected. The beads were brought up in 80% acetonitrile with 0.1% trifluoroacetic acid and then loaded onto conditioned C-18 stage tips. The stage tip was washed by 80% acetonitrile containing 0.1% trifluoroacetic acid, followed by 1% formic acid. The flowthroughs were collected and combined with the enrichment unbound peptides for subsequent glycopeptide enrichment. The phosphopeptides were eluted from the Fe^{3+}-IMAC beads to conditioned C-18 stage tip with 70 μL of 500 mM dibasic potassium phosphate, pH 7.0 three times. The tip was washed twice with 1% formic acid, followed by elution of the phosphopeptides with 50% acetonitrile containing 0.1% formic acid twice. Phosphopeptides were dried down and resuspended in 3% acetonitrile, 0.1% formic acid for 2.5 injection and loaded on Evoptip to LC-MS/MS analysis.

Glycopeptide enrichment

Peptides not retained during phosphopeptide enrichment underwent desalting using a Sep-Pak tC18 96-well plate. For glycopeptide enrichment, OASIS MAX solid-phase extraction cartridges were used. The cartridges were sequentially conditioned with three washes of 1 mL acetonitrile, followed by three washes with 100 mM triethylammonium acetate buffer, three washes with water, and finally three washes with 95% acetonitrile containing 1% trifluoroacetic acid. Peptides were applied to the cartridge twice to maximize binding. To keep non-glycosylated peptides for acetylation enrichment, the cartridge was washed four times with 1 mL of 95% acetonitrile (1% TFA). Glycopeptides were subsequently eluted using 50% acetonitrile containing 0.1% TFA, dried under vacuum, and reconstituted to 2.5 injections in 3% acetonitrile containing 0.1% formic acid before LC-MS/MS analysis. The one injection peptides were loaded onto Evotip for TMTpro analysis on Ascend.

Enrichment of acetylated peptides

Peptides (three fractions combined to one fraction) of four fractions each set were reconstituted in 400 μL of 1× IAP bind buffer from the kit by pipetting and shaking to ensure complete dissolution. The pH was confirmed to be neutral (no lower than 7.0) and adjusted with Tris buffer if necessary. The sample was centrifuged at 10,000 × g for 10 min at 4°C. The antibody-bead slurry was briefly spun at 1,000 × g and gently pipetted to obtain a uniform suspension. A total of 31 μL of Ac-K bead slurry was transferred into a 2.0 mL microcentrifuge tube, followed by three washes with 1.65 mL of ice-cold 1× PBS buffer. The beads were then resuspended and aliquoted equally (400 μL each) into four 1.7 mL tubes. After removing the PBS, about 750 μg of peptide solution was added to each tube and incubated on an end-over-end rotator for 2 h at 4°C. The unbound peptides were collected for subsequent experiments. Beads were washed four times with 400 μL chilled HS IAP wash buffer, followed by two washes with 400 μL chilled LC-MS water. Peptides were

eluted with 50 μL of IAP Elution Buffer (0.15% TFA) at room temperature while mixing (1,750 rpm) for 10 min. This elution was repeated three times, and the combined eluates were pooled. Half of the final eluent was loaded onto Evotip for TMTpro analysis on Ascend.

LC-MS/MS for global proteomic and PTMs analyses

The TMTpro-labeled global peptides, phosphopeptide, glycopeptide, and acetylated peptide fractions were analyzed on Evosep-Orbitrap Ascend Tribrid Mass Spectrometer (Thermo Scientific). Approximately 1.0 μg TMTpro labeled peptides were separated on PepSep C18 column of 15 cm × 150 μm, 1.5 μm. The column was heated to 50°C and the LC gradient is 15 SPD. The data was collected in a data-dependent acquisition (DDA) mode with the settings for global proteomic samples: MS1 resolution – 120,000, mass range – 400 to 1,400 m/z, RF Lens – 60%, AGC Target – 1.2e6, Max injection time – auto, MIPS – peptide with most abundant peak, intensity threshold – 2.5e4, charge state include – 2–6, dynamic exclusion – 45 s, exclude after 1 time. The cycle time was set to 3 s, and within this 3 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 30,000, TurboTMT – All TMT Reagents, first mass (m/z) – 110, high-energy collision dissociation (HCD) activation energy– 35, isolation width (m/z) – 0.7, AGC Target – 5.0e4, Max injection time – 59 ms. The phosphoproteomic and acetylated proteomic samples were analyzed with the setting: MS1 resolution – 120,000, mass range – 400 to 1,400 m/z, RF Lens – 60%, AGC Target – 1.2e6, Max injection time – auto, MIPS – peptide with most abundant peak, intensity threshold – 1.0e4, charge state include – 2–6, dynamic exclusion – 45 s, exclude after 1 time. The cycle time was set to 3 s, and within this 3 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 30,000, TurboTMT – All TMT Reagents, first mass (m/z) – 110, high-energy collision dissociation (HCD) activation energy– 33, isolation width (m/z) – 0.7, AGC Target – 3.0e5, Max injection time – 59 ms. The glycoproteomic samples were analyzed by following the optimized setting[140]: MS1 resolution – 60,000, mass range – 350 to 2,000 m/z, RF Lens – 60%, AGC Target – 1.2e6, Max injection time – auto, MIPS – peptide with most abundant peak, intensity threshold – 1.0e4, charge state include – 2–8, dynamic exclusion – 45 s, exclude after 1 time. The cycle time was set to 3 s, and within this 3 s the most abundant ions per scan were selected for MS/MS in the orbitrap. MS2 resolution – 30,000, TurboTMT – All TMT Reagents, first mass (m/z) – 110, high-energy collision dissociation (HCD) activation energy– 25,35,45, isolation width (m/z) – 0.7, AGC Target – 6.0e5, Max injection time – 64 ms.

XL-MS sample preparation and data generation

Cross-linking, protein extraction and tryptic digestion

15∼35 mg of cryo-pulverized GC tissues and NAT were suspended in cross-linking buffer (20 mM HEPES (4-(2-Hydroxyethyl)piperazine-1-ethane-sulfonic acid), 150 mM NaCl, 1.5 mM $MgCl_2$, pH 7.8) containing 2 mM Alkyne-A-DSBSO. The cross-linking reaction was carried out for 1 h with rotation at 37°C and then quenched with 50 mM ammonium bicarbonate for 15 min at room temperature. The cross-linked tissues were lysed in UA lysis buffer (8 M urea, 150 mM NaCl, 50 mM Tris (pH 8.0), 0.5mM DTT, 0.5% NP-40) by probe sonication. The supernatant was clarified by centrifugation at 14,000 g and protein concentration was measured by Bradford assay. The extracted proteins were digested using FASP (filter aided sample preparation) procedure as previously described.[141] Cross-linked proteins were transferred onto a 30 KDa FASP centrifugal filter and washed with 8 M urea buffer. The denatured proteins were reduced by 2 mM TCEP for 30 min and alkylated with 10 mM iodoacetamide for 30 min in dark. After washing with 25 mM ammonium bicarbonate, proteins were reconstituted in 60 μL of 8 M urea in 25 mM ammonium bicarbonate. Lys-C (enzyme to protein ratio of 1:100) was added to the solution and the mixture was incubated at 37°C for 4 h. Then, the concentration of urea was reduced to 1.5 M for trypsin digestion (enzyme to protein ratio of 1:50) at 37°C overnight. The peptide digests were extracted, desalted by Sep-Pak C18 cartridge and stored in −80°C freezer prior to subsequent "click"-based labeling and enrichment.

Click chemistry-based cross-link enrichment

Click labeling of BPA (biotin picolyl azide) to Alkyne-A-DSBSO cross-linked peptides was performed as previously described.[141,142] Briefly, BPA, BTTAA, $CuSO_4$, and sodium ascorbate were added to peptide digests (40 μL) to obtain a final concentration of 1.2 mM, 1 mM, 500 μM, and 5 mM, respectively. The mixture was then rotated in the dark at r.t. for 2 h. After labeling, binding buffer (25 mM sodium phosphate, 150 mM NaCl, pH 7.5) was added to a final volume of 200 μL and rotated with 250 μL of streptavidin beads for 2 h at room temperature. After washing with binding buffer twice, the DSBSO cross-linked peptides were eluted from the beads through acid cleavage using 10% formic acid at 37°C overnight and dried using Speed-Vac.

TMT labeling of cross-linked peptides

Equal amount of cross-linked peptides from each tissue sample were labeled with 11-plex TMT reagents according to the manufacturer's protocol. Briefly, peptides (20 μg) were dissolved in 25 μL of 100 mM TEAB (pH 8.5). 0.8mg of TMT reagents were dissolved in 41 μL of anhydrous acetonitrile (Sigma), and 3 μL of each TMT reagent was added to the corresponding aliquot of peptides together with 7 μL of anhydrous acetonitrile. The reaction was incubated at r.t. for 1h with shaking and quenched with 5% hydroxylamine at r.t. for 15 min. The labeled peptides were desalted using Sep-Pak C18 cartridge (Waters) and dried using Speed-Vac. In this study, 150 GC tissues and 40 NATs were co-randomized to create 19 TMT 11-plex sets. A reference sample was created by pooling peptides from 109 GC to 20 NAT tissues (representing ∼68% of the sample cohort), which was included as a reference for all TMT 11-plex experimental sets.

SEC-HpHt fractionation of cross-linked peptides

Cross-linked peptides from each TMT set were dissolved in 30% acetonitrile/0.1% TFA and fractionated on a Superdex 30 Increase 3.2/300 column with an Agilent 1260 series HPLC. The two primary fractions with inter-linked peptides (F23-25min, F25-27min) were collected, dried, dissolved in 160 μL of ammonia water (pH 10) and subjected to HpHt fractionation separately.[143]

For HpHt separation, a pipette tip (200 μL) was first blocked with a layer of C8 membrane (Empore 3M). After filling with 5 mg of C18 solid phase (3 μm, Durashell, Phenomenex). The tip was balanced with 90 μL of methanol, 90 μL of acetonitrile and 90 μL of ammonia water (pH 10) sequentially. Then, each fraction from SEC was loaded onto one tip and centrifuged at 1,200 rpm until the liquid level was close to the beads. After washing with another 90 μL of ammonia water (pH 10), peptides were eluted with increasing percentage of acetonitrile in ammonia water (6%, 9%, 12%, 15%, 18%, 21%, 25%, 30%, 35%, and 50%). The 25%, 30%, 35% and 50% fractions were combined to 6%, 9%, 12% and 21% fractions, respectively. The peptides were dried for LC-MS/MS analysis.

LC-MS^n analysis for cross-linked peptides

Cross-linked peptides were analyzed by LC-MS^n using an UltiMate 3000 RSLC coupled with an Orbitrap Fusion Lumos mass spectrometer.[141] Samples were loaded onto a 50 cm × 75 μm Acclaim PepMap C18 column and separated over a 180 min gradient of 4%–25% acetonitrile at a flow rate of 300 nL/min. The targeted CID-MS3 acquisition method based on the mass difference (Δ = 182.0071, $C_5H_{10}O_3S_2$) between the characteristic ion pairs of DSBSO cross-linked peptides in MS^2 (i.e., $\Delta = \alpha_T\text{-}\alpha_A = \beta_T\text{-}\beta_A$, α/β_A means the peptide fragment is modified with a alkene (A) moiety, α/β_T means the peptide fragment is modified with a thiol (T) moiety) was used to select MS^2 fragment ions for MS^3 sequencing. Ions with charge of 4+ to 8+ in the MS^1 scan were selected for MS^2 analysis. MS^1 and MS^2 scans were acquired in the Orbitrap whereas MS^3 scans were detected in the ion trap. For MS^2 scans, the resolution was set to 30,000, the AGC target 1e5, the precursor isolation width was 1.6 m/z, and the maximum injection time was 70 ms for CID. The CID-MS^2 normalized collision energy was 23%. For MS^3 scans, CID was used with a collision energy of 35%, the AGC target was set to 2×10^4, and the maximum injection time was set to 150 ms. TMT quantitation on cross-linked peptides was accomplished using the Synchronous Precursor Selection (SPS) MS^3 method.[144]

Metabolomic profiling

Metabolites extraction

Tumor specimens from each experimental group underwent simultaneous extraction using a standardized procedure. Tissue samples were first weighed, then homogenized using a Bead Mill Homogenizer in an extraction solvent comprising 99% LC-MS-grade acetonitrile with 1% formic acid. Following homogenization, samples underwent centrifugation, after which the resulting supernatant was transferred to extraction plates equipped with specialized sorbent filters for concurrent protein precipitation and phospholipid elimination. Using a Positive Pressure-96 processor, the metabolite-containing solution was filtered through the sorbent material, yielding purified metabolite extracts in MS-compatible plates. The final extract volume was concentrated to 150 μL in preparation for instrumental analysis.

Metabolites analysis

Mass spectrometric analysis employed a Thermo Scientific IQX Mass Spectrometer interfaced with a Vanquish UHPLC platform. Sample temperature was maintained at 4°C during chromatographic separation, with 2 μL injection volumes. Chromatographic separation utilized reversed-phase methodology on a Discovery HSF5 column (Sigma) over a 15-min analytical window. The mobile phase system comprised aqueous and organic components: 0.1% formic acid in MS-grade water and 0.1% formic acid in acetonitrile, respectively.

Instrument calibration preceded sample analysis to maintain analytical sensitivity and measurement precision. Metabolite quantification involved chromatographic peak area integration with subsequent normalization to individual sample masses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Harmonized genome alignment

WGS, WES, RNA-Seq sequence data were harmonized by NCI Genomic Data Commons (GDC; https://gdc.cancer.gov/about-data/gdc-data-harmonization), which included alignment to GDC's hg38 human reference genome (GRCh38.d1.vd1) and additional quality checks, as previously described.[145] GDC-aligned BAMs were used for all downstream genomic data processing to ensure reproducibility.

Standardized reference for proteogenomics data processing

The GENCODE V42 Basic (CHR) annotation and matched reference genome (GRCh38.p13) were selected for multi-omics data processing. To facilitate human understanding of the Ensembl gene IDs, we used corresponding HUGO gene symbols as display names. When multiple Ensembl gene IDs are mapped to the same gene symbol, each ID was assigned a unique gene name by combining the gene symbol and Ensembl gene ID. For PAR_Y genes (e.g., ENSG00000002586_PAR_Y), a tag "_PAR_Y" was added to the end of gene names (e.g., CD99_PAR_Y).

Matched protein database, gencode.v42.pc_translations.fa, was downloaded from GENCODE. Only proteins from coding transcripts in GENCODE V42 Basic (CHR) annotation were retained and others were discarded. To remove redundant protein sequences, protein sequences were grouped into 50,827 groups with the same sequences. Only one representable protein was selected for each protein sequence group.

Please cite this article in press as: Wang et al., A 15-layer multi-omics analysis of gastric cancer ecotypes provides therapeutic insights, Cell Reports Medicine (2026), https://doi.org/10.1016/j.xcrm.2026.102756

WGS data analysis

WGS data processing

WGS analysis was performed as previously described.[145] WGS data of patient matched tumor and blood normal samples were analyzed using the Getz Lab's production hg38 WGS characterization pipeline on wolF, a workflow manager to run dockerized workflows and dynamically allocate hardware resources to sequential tasks. The wolF characterization pipeline integrates quality control, and mutation (SSNV and indel) and absolute allelic copy number calling for hg19 and hg38 WGS and whole-exome sequencing data. Due to the large number of computational resources required to efficiently process cancer whole genomes, we ran these analysis pipelines on an elastic high-performance computing (HPC) cluster on Google Cloud VMs, comprising thousands of CPU cores.

The DNA Sequence Quality Control module, at the head of the pipeline, employs (i) GATK4's CalculateContamination (Ver GATK 4.1.4.1) tool to calculate the fraction of reads coming from cross-sample contamination and (ii) GATK4 Picard tools (ver GATK 4.0.5.1) to validate the BAM files and collect multiple classes of metrics that can be used to evaluate sequencing data quality. The pipeline's Somatic Copy Number Analysis module combines callers ReCapSeg, based on GATK4 Best Practices Workflow (ver GATK 4.1.4.1) for discovery of allele-specific copy-number alterations, with the Getz Lab's new caller HapASeg.

Gene and focal level somatic copy number calling

Segment level somatic copy number alterations were predicted using paired tumor and blood normal from WGS data. Then GITSIC2 with threshold of ±0.3 was applied to identify gene wise gain or loss of copy number. The standardized reference was used to map the copy number information to genes.

WES data analysis

Somatic variant calling

We implemented Sarek (v3.4.0) for processing WES data, utilizing FreeBayes (v1.3.6), Mutect2 (gatk v4.4.0.0), and Strelka2 (v2.9.10) for identifying somatic mutations. VCF outputs from somatic mutation calling tools were left aligned and merged based on a two-out-of-three-tools rule. The merged results were annotated using VEP (v110.0) and transferred to maf format by Vcf2maf (v1.6.21).

Mutational signatures

The R package SigProfilerMatrixGeneratorR (version 1.0) was used to call mutation signatures from somatic mutation data.[94] The maximum number of signatures was set to 10 and the number of NMF replicates was set to 100. The activity scores of SigProfilerMatrixGenetator suggested decomposed solutions were used as signature scores. Only single base substitutions signatures were included in our analysis.

Microsatellite instability prediction

MSI scores were calculated by MSIsensor (https://github.com/ding-lab/msisensor) and interpreted as the percentage of microsatellite sites (with deep enough sequencing coverage) that have a lesion. Samples with an MSIscore >5 were classified as ''MSI-High'' and the rest will be classified as ''MSS'' based on a bimodal distribution of MSIscore on this cohort.

DNA methylation analysis

We downloaded processed probe level beta values which are methylated to unmethylated signal intensities from GDC. All loci in EPIC Manifest file infinium-methylationepic-v-1-0-b5-manifest-file-csv.zip were reannotated using annovar (v 04.16.2018) with the standardized reference.[95] For downstream integrated analysis, CpG islands annotated as upstream (1kb upstream of transcription start site) and UTR'5 were included for coding genes. For noncoding genes, only CpG islands annotated as upstream were included. Then, the gene-level methylation was derived by averaging these probe-level methylation beta values.

RNA quantification and circular RNA prediction

The hg38 reference genome and standardized annotation were used for the RNA-Seq data analysis. First, CIRI (v2.0.6) was used to call circular RNA with default parameters, based on a published large-scale benchmarking study demonstrating high sensitivity of CIRI2 among well-established circular RNA detection tools.[146] BWA (version 0.7.17-r1188) was used as the mapping tool. The cutoff of supporting reads for circular RNA was set to 10, representing a conservative choice to further enhance detection precision in light of benchmarking evidence that higher back-spliced junction read counts are associated with improved precision.[146] Then we used a pseudo-linear transcript strategy to quantify gene and circular RNA expression 117. In brief, for each sample, linear transcripts of circular RNAs were extracted and 75bp (read length) from the 3′ end was copied to the 5′ end. The modified transcripts were called pseudo-linear transcripts. Transcripts of linear genes were also extracted and mixed with pseudo-linear transcripts. RSEM (version 1.3.1) with Bowtie2 (version 2.3.3) as the mapping tool was used to quantify gene and circular RNA expression based on the mixed transcripts. After quantification, the upper quantile method was applied for normalization. The normalized matrix was log2-transformed and separated into gene and circular RNA expression matrices.

miRNA-seq data analysis

miRNA isoform quantification data were obtained from the GDC. Isoform-level expression, measured in reads per million (RPM), was then aggregated into mature miRNA expression based on the miRBase v21 annotation.[147] Isoforms with 3′ or 5′ ends deviating by more than 2 bp from the annotated locations were excluded from the aggregation.

Ubiquitinated proteome DIA data processing

The.d files generated from timsTOF-HT were analyzed by Spectronaut (version 18.4) with directDIA approach. The Pulsar search settings for the ubiquitinated peptide search were: peptide length range of 7–52 amino acids, the fixed modification of carbamidomethylation on cysteine (C), the variable modifications of glygly on lysine (K), oxidation on methionine (M), and acetylation on protein

N-terminal. The library generation settings were followed BGS factory settings against human database (GENCODEV42; 51,689 entries) with all PSM FDR, peptide FDR and protein FDR below 0.01. The peptides were identified and quantified at modified peptide level without cross-run normalization. Then we median centered the data. Finally, the site level matrix was generated by averaging peptides mapping to the same ubiquitination site.

TMTpro data processing

MS/MS spectra were processed using the FragPipe (v21.1) pipeline. Specifically, the spectra were searched using MSFragger (v4.0) against GENCODE V42 protein database. For the analysis of the whole proteome, MS/MS spectra were searched with a precursor-ion mass tolerance of 20 ppm, fragment mass tolerance of 20 ppm, and isotope errors (−1/0/1/2/3). Fixed modifications were set as cysteine carbamidomethylation (+57.0215) and lysine TMT labeling (+304.20715). Variable modifications were set as methionine oxidation (+15.9949), N-terminal protein acetylation (+42.0106) and TMT labeling of peptide N terminus and serine residues. The searching allowed for up to two missed cleavage sites. For the analysis of phosphoproteomic data, the variable modifications also included phosphorylation (+79.9663) of serine, threonine, and tyrosine residues. For the acetylomics data, 4 missed cleavage sites were allowed and variable modifications also included (+42.0105) and carbamylation (+43.0058) on lysine residues.

For global proteomics, MSBooster and Percolator were used with default settings to rescore PSMs using deep learning prediction to predict retention time and spectra.[114] Only Percolator was used for rescoring for phosphoproteomic and acetylomic data. Protein inference was done using ProteinProphet with max ppm difference 2,000,000 and minimum probability of 50%. PTMProphet was used to localize the PTM sites for phosphoproteomic and acetylomic data. For the PSMs that passed filtration, MS1 intensities of the corresponding precursor-ions were extracted using Philosopher (v5.1.0) for global and phosphoproteomic data and IonQuant (v1.10.12) for the acetylomics data.

Then TMT-Integrator was used to obtain the gene level quantification and modification site quantification for the phosphoproteomic and acetylomics data. TMT integrator selects the best PSMs by ensuring all PSMs 1) are TMT labeled 2) have reference channel intensity >0, 3) precursor ion purity >50%, 4) summed MS2 intensity across all channels is $\geq$5% (2.5% for phosphorylation and acetyl enriched data) 5) PSM with the highest summed MS2 intensity among all PSMs identifies the same peptide ion in the fraction 6) Does not map to contaminate proteins 7) PSMs grouped by gene or modified site and intensities fall within Q1-1.5xIQR and Q3+1.5xIQR and 8) contains phosphorylation or acetylation for PTM data. For selected PSMs, all channel intensities are divided by the reference channel, log2 transformed, and median centered to obtain normalized intensities.

PTM site mapping

Representative isoforms were selected using the method described.[22] PTM sites were re-aligned to the selected isoforms if at least one identified peptide assigned to that gene matched the primary isoform sequence. The site position was given relative to the selected isoform. If the peptide matched multiple locations in the protein sequence, then the first location was used. The sites that could not be mapped were discarded. Similarly, only glycopeptides that mapped to the primary isoform were included.

Glycoproteomic data processing

Intact N-linked glycopeptides were identified and quantified using GPQuest 3.0,[84] a specialized software for glycopeptide analysis. The identification workflow involved detecting MS/MS spectra containing oxonium ions (notably m/z 204.0966)—diagnostic fragments from glycan moieties—to select candidate glycopeptide spectra (“oxo-spectra”) after TMT reporter ion removal These spectra were searched against a database of tryptic glycosite-containing (motif of N[^P][ST]) peptide sequences—ranging from 7 to 50 amino acids in length, allowing up to two missed cleavages, with fixed modifications of TMTpro on N-term and Lys and carbamidomethylation on cysteine residues, and variable oxidation on methionine—derived from 51,689 protein entries in GENCODE v42, along with a library of 695 N-linked glycan compositions reported in pGlyco. Spectra underwent preprocessing steps including reporter ion removal, de-noising, deisotoping, oxonium ion evaluation, and glycan type prediction. The top 300 peaks of each processed spectrum were matched to indexed peptide fragments to generate candidate peptide hits, which were scored using Morpheus scoring that accounts for peptide, glycopeptide, and isotopic fragment matches. The best-scoring peptide was assigned to each spectrum, and the corresponding glycan was identified by matching the mass gap between the peptide and precursor ion to glycan masses. Final identifications with an FDR of glycopeptide-spectrum matchings (GPSMs) < 1% were retained, with mass tolerances set at 10 ppm for precursors and 20 ppm for fragments. For quantification, the relative abundance of each intact N-linked glycopeptide was derived from the median log2 ratio of its GPSMs, normalized across samples, and scaled to an estimated absolute abundance using a reference sample. Identified glycopeptides were further analyzed for enriched pathways and site-specific glycosylation patterns. Lastly, site-specific-glycoform level data was summarized by taking median of glycopeptides mapping to the same glycoform.

XL-MS data processing

Identification of cross-linked peptides

MS^3 spectra was extracted by PAVA (UCSF) and subjected to Protein Prospector (v.6.3.5) for database searching (using Batch-Tag against GENCODE.V42.basic.CHR. contaminants_removed with a random concatenated decoy database). The mass tolerances were set as ±20 ppm for parent ions and 0.6 Da for fragment ions. Trypsin was set as the enzyme with three maximum missed cleavages allowed. Cysteine carbamidomethylation and TMT at protein N-terminus were set as static modifications. A maximum of four variable modifications were also allowed, including TMT at lysine, methionine oxidation, N-terminal acetylation, and N-terminal conversion of glutamine to pyroglutamic acid. Two additional modifications, alkene (C_3H_2O, +54.0106 Da) and unsaturated thiol

($C_8H_{12}S_2O_4$, +236.0177 Da) were added for uncleaved lysines, corresponding to remnant moieties of DSBSO after cross-link cleavage.[141]

The in-house software XL-Tools was used to automatically identify, summarize and validate cross-linked peptides based on Protein Prospector database search results and MS^n data. Only cross-linked peptides with a length of 6 amino acids or more were retained. The false discovery rate for cross-link identification was calculated based on a target-decoy approach and the FDR at total CSMs level was 2.03%. PPI analysis and structural mapping were derived based on DSBSO XL-MS data. BioPlex (https://bioplex.hms.harvard.edu/),[26] BioGrid (https://thebiogrid.org/),[25] CORUM (http://mips.helmholtz-muenchen.de/corum),[123] and STRING (https://string-db.org/)[124] were used to annotate PPIs and map protein complexes. Mapping of cross-links to high-resolution structures of protein complexes was performed using PyMOL as previously described.[141–143]

Mapping peptide crosslinks to PPIs

To reduce redundancy and ambiguity in peptide pair-to-protein pair assignments from XL-MS data, we applied XL-Ranker,[41] a prioritization workflow to select the most probable interactions for PPIs with XL-MS and global proteomic data. Peptide pairs detected by XL-MS data were first constructed into a bipartite graph connecting peptide pair nodes and protein pair nodes based on matching sequences. Distinct subnetworks, known as connected components, were identified. Within each component, a greedy set cover algorithm selected a minimal set of protein pair groups necessary to represent all peptides pairs, creating a parsimonious protein pair list. If a subnetwork resulted in only one protein pair selected, it was marked as a "leading single" pair. For components with multiple candidate protein pairs, an XGBoost model ranked them based on protein abundance and known PPI status. The highest-scored pair was labeled "leading multiple ML-prioritized", while the others were labeled as "not ML-prioritized. Only the "leading single" and "leading multiple ML-prioritized" pairs were selected to construct the final PPI network, thus improving specificity by resolving ambiguous peptide pair mappings.

Metabolome data processing

Analytical data processing utilized Gigantest's Laboratory Information Management System (LIMS) alongside Python and R programming environments. Machine learning algorithms facilitated initial peak detection, followed by manual verification of each chromatographic peak to ensure integration accuracy.

Quality control and sample exclusion based on quantitative data tables

We applied OmicsEV to the protein, PTM, and RNA-Seq data tables for quality control.[137] Distribution normalization and feature count consistency were confirmed across TMT plexes. However, we found that TMT 13 had a distinct batch effect from the PCA plots and correlation heatmap. This is likely due to the fact samples with high blood content were grouped in TMT 13. As an independent evaluation, we performed UMAP analysis for RNA-Seq, miRNA, global proteomics, phosphoproteomics, acetylomics, glycoproteomics, ubiquitomics, and the XL-MS data. The input to the UMAP excluded any measurements with missing values. In the UMAPs, TMT 13 samples consistently showed separation from other batches across multiple omics layers, confirming the batch effect identified in OmicsEV. As a result, 10 tumor and 6 NAT samples from TMT 13 were removed.

For TMT-labeled datasets (proteomics, phosphoproteomics, glycoproteomics, and acetylomics), data were reprocessed without TMT 13. OmicsEV analysis repeated after TMT 13 exclusion confirmed the removal of batch effects. One additional NAT sample (C3L.06421.N) was excluded because it clustered with tumor samples and lacked clear pathological annotation after expert review.

Inferred molecular phenotypes

PROGENy pathway activity

The PROGENy R package was applied to the log2 transformed RSEM mRNA matrix to estimate activity of 11 cancer related pathways: EGFR, Hypoxia, JAK/STAT, MAPK, NFkB, PI3K, TGFb, TNFa, Trail, VEGF, p53.[106]

Immune deconvolution

The R package immunedeconv (V2.0.4) was used to do the immune deconvolution using the RNA expression data (TPM).[128] Cibersort and xCell were selected in our analysis among the 7 deconvolution methods in immunedeconv.

Inferred gene sets and miRNA activity scores

All scores were inferred by single sample gene set enrichment analysis (ssGSEA) method from the GSVA R package.[108] The immune signatures are from MSigDB.[109] Proteomics data was used to infer KEGG pathway activity and hallmark score. RNA-Seq was used to infer Hallmark score and all other gene sets. Targets of miRNAs were downloaded from the miRNA targets database miRTarBase and only the miRNA/target pairs with strong experimental evidence were retained.[148] miRNA target sets with fewer than 10 genes were removed. The -log2 transformed ssGSEA score was used as the miRNA activity score.

Transcription factor activity score

Transcription factor activity was performed on the log2 transformed RSEM RNA-Seq data using the univariate linear model from the decoupleR (2.10.0) package and the CollecTRI human transcription factor target network in accordance with previous benchmarking analysis.[149]

Tumor microenvironment score

The ESTIMATE scores reflecting the overall immune and stromal infiltration were calculated by the R package ESTIMATE using the normalized RNA expression data (RSEM).[99]

ABSOLUTE tumor purity and ploidy

The DoAbsolute R package (V2.2) with ABSOLUTE (V1.0.6) was used to infer tumor purity and ploidy from somatic mutations and WGS-based CNA data. The parameters min.mut.af and max.as.seg.count were set to 0.02 and 5000, respectively, with all other parameters at default values. These settings were selected based on empirical evaluation of mutation and CNA calling results and to account for improved detection of low–allele-fraction variants and shorter CNA segments enabled by increased sequencing depth.

Kinase activity score

Known kinase-substrate relationships (KSRs) were obtained from the PhosphoSitePlus database (Downloaded July 2023) and the training dataset of GPS 5.0. Additionally, predicted KSRs with scores greater than 5.0 were sourced from NetworKIN (Downloaded Sept. 2023).[150] This is in accordance with previous benchmarking efforts which found well curated substrates combined with predicted substrates supported by network interactions from NetworKin perform the best in inferring kinase activity.[151] Substrate positions were remapped based on phosphorylated peptide sequences and gene symbols to ensure alignment with our reference database. When a phosphorylated peptide matched multiple protein isoforms from the same gene, all corresponding isoforms were considered substrates of the associated kinase. In line with previous benchmarking results,[151] kinase activity was inferred using PTM-SEA[109] with $\log_2$ (TMT ratio) phosphoproteomic data that were not normalized to protein abundance (i.e., without subtracting protein-level signals) as input. Phosphorylation sites missing in more than 50% of samples were excluded, and the following parameters were used: weight = 0, statistic = ‘‘area.under.RES’’, output.score.type = ‘‘NES’’, nperm = 1000, and min.overlap = 5.

Molecular subtype analysis

EBV positive sample determination

VirDetect was used to detect virus-driven reads in RNA-Seq Data. In detail, STAR (v2.7.10a) was used to map RNA-Seq reads to the human hg38 reference genome. Unmapped reads were mapped to the masked human virus sequences using the same tool. Picard (v1.92) was used to count reads mapped to each virus. Samples with more than 4000 reads were identified as EBV positive.

Copy number classification

The copy number subtypes were characterized by CNA focal level changes following the TCGA classification method.[4] In detail, tumors were clustered based on thresholded focal level amplifications and deletions reported by GISTIC2.0 in all_lesions.conf_99.txt file. Euclidean distance using Ward’s method was used to cluster tumor samples in R.

TCGA subtype classification

TCGA subtypes were assigned following the TCGA method in the following order: EBV-positive samples first, then MSI-high, followed by chromosome instability, and finally the remaining samples were labeled as genomically stable.[4]

ACRG subtype classification

The ACRG subtype was characterized using the ACRG gene signature score and as described in the ACRG GC paper.[5] Using the RNA-Seq expression in our cohort we calculated the average expression of the EMT and MSI signatures. The signature thresholds for defining EMT and MSI high samples were defined using qqplots for each signature’s distribution. Next, the *TP53* signature score was derived from the two gene RNA-Seq signature (MDM2 and CDKN1A). We used an AUC-ROC curve of the *TP53* signature with *TP53* mutations and assigned the threshold using the optimal threshold. Remaining samples were classified as MSS. Only the EMT subtype was used for downstream analysis.

Significantly mutated genes

WES maf file was split by MSI high and MSS tumors. The HG38 based maf file was then mapped to HG19 using CrossMap (v0.7.0).[125] Significantly mutated genes were then identified for each maf file using MutSig2CV (3.11) $q < 0.1$ and genes names were mapped back to HG38.

Frequency of significantly mutated genes in TCGA STAD was determined by downloading TCGA maf file using the R package maftools (v2.22) along with the MC3 project.[152] The oncoprints were generated using maftools.

Data matrices used for analysis

All analyses were performed using the processed data matrices described above and deposited on LinkedOmics, unless otherwise specified in the STAR Methods. *Z* score was used to make scales comparable when plotting RNA and protein quantification in the same plot. *Z* score was calculated with the R scale function, using measurements across all samples under comparison as input.

Missing value exclusion criteria for proteomic based features

No data imputation was performed to avoid introducing artificial signals or bias into downstream analyses. Instead, missingness was addressed through feature-level filtering to ensure analytical robustness. For *cis/trans* analysis of significantly mutated genes in MSS tumors, CNA driver and methylation driver analyses, and outlier analysis, proteomic-based features were filtered to exclude features with more than 50% missing values. For differential one-vs-rest ecotype analyses, at least five values were required for both groups.

RNA-protein/PTM and protein-PTM correlation

Gene level RNA and protein data was correlated together and with all quantifiable PTMs using Spearman’s correlation. At least 20 samples with complete measurements were required.

Ecotype classification by consensus clustering

The 71 cell state scores were generated by EcoTyper[17] using TPM RNA-Seq for our cohort and default parameters. Ecotypes were obtained through consensus clustering using all 71 cell states with ConsensusClusterPlus (v1.7), applying default settings except for maxK = 10, reps = 100, distance = "spearman", innerLinkage = "ward.D", and seed = 123. 6 clusters were selected based on the elbow plot from the ConsensusClusterPlus output. The 6 clusters were further evaluated with silhouette scores. Silhouette scores were calculated from Spearman distance matrix using silhouette function from Cluster (2.1.8.1) R package.

Analysis of previously published RNA datasets

GSE15459,[37] GSE26899,[38] GSE26901,[38] and GSE62254 (ACRG)[5] were downloaded from GEO. Clinical information was downloaded from supplemental tables from the studies' publications. GSE62254 was normalized by RMA (Robust Multiarray Average) and we ensured other cohorts were log2 transformed and quantile normalized. TCGA TPM and RSEM RNA-Seq data, along with clinical data, were downloaded from cBioPortal for the Stomach Adenocarcinoma (PanCancer Atlas) cohort. For Lauren subtyping for TCGA, classifications were downloaded from cBioPortal for the original 2014 paper[4] as the PanCancer cohort does not have Lauren classification. RNA data were used to generate the 71 cell state scores with EcoTyper[17] using default parameters in recovery mode. For TCGA, TPM data were used to compute the cell state scores, while all other analyses involving TCGA data were performed using log2-transformed RSEM values.

Analysis of previously published single cell RNA-Seq dataset

The count matrices in CellRanger output format were downloaded from GEO (GSE150290). Data for each sample were read and combined into one Seurat object without integration. There are 23 tumor samples and 29 normal samples. Standard practice was done for quality control (expressed gene larger than 200 and smaller than 2500 and percentage of mitochondria gene <5%), log normalization, selecting variable 2000 features, running PCA, clustering and UMAP. Top 15 principal components were selected for clustering and running UMAP.

SingleR R package was used to assign major cell types (epithelial, fibroblast, endothelial, B/plasma, T, mast, monocyte/macrophage) with a gastric ref.[153] Clusters were dominated by one major cell type and all cells in the cluster were assigned to corresponding cell types. To align to the same 12 cell types for Ecotype analysis, immune cells were reassigned by using the LM22 signature matrix by grouping the cell types to B cells, PCs, CD8 T cells, CD4 T cells, NK cells, Monocytes/Macrophages, Dendritic cells, Mast cells and Neutrophils following the EcoTyper framework.[17]

Cell state and ecotype survival analysis

Survival analysis of the cell states scores was done using coxph function using the survival (3.8–3) R package. For each of the 5 cohorts (TCGA, ACRG, GSE15459, GSE26899, and GSE2690), a univariate model was fitted. Samples were filtered to only include patients with an overall survival of at least 1 month. *p* values for cell states from each cohort were summarized to meta *p* values using the sumz method in the R package metap (v1.12) as described previously.[28] For the ecotype survival analysis, all cohorts were combined, and a single univariate model was fitted for the ecotypes. Survival curves were plotted with R package survminer (0.0.5). We then further assessed the survival of the ecotypes by fitting a multivariate model with clinically recorded gender, age, stage and Lauren subtype type were used as covariates. Age was split by the median into $\geq$64 or <64, stage was split into stage I/II and III/IV, and Lauren subtype was split into diffuse or intestinal.

Clinical features associated with ecotypes

For categorical variables such as pathologic stage, clinically recorded gender, *H. pylori* status, and tumor site association with ecotype was determined by fisher's exact test. Pathologic stage was grouped as I, II, III, or IV. For numerical features such as age ecotypes were compared pairwise with Wilcoxon rank-sum test.

Ecotype prediction using XGBoost model

The 71 EcoTyper[17] cell states from our cohort were used to train a classifier for our six ecotypes using the XGBoost (v2.0.2) algorithm. Feature selection was performed using the Minimum Redundancy Maximum Relevance (mRMR) method, and the performance of different feature sets was evaluated through 5-fold cross-validation. The feature set with the highest AUROC was selected to build the final prediction model, which was subsequently applied to classify samples from the 5 previous GC cohorts into the six ecotypes.

PredictIO analysis

Immune checkpoint inhibitor response scores from PredictIO were obtained by uploading RNA datasets for this study, TCGA, and ACGR to their website.

Enhanced diffusion-based data pipeline for survivability prediction

To predict survival outcomes in the CPTAC cohort, we leveraged a modular learning pipeline[39] that integrates generative AI-based data augmentation, task-aware sampling, and multi-source domain adaptation. The framework consists of three key components: (i) a diffusion-based generative model that synthesizes RNA-Seq data with limited survival information, (ii) a task-aware sampling

mechanism that selects task-relevant synthetic subsets, and (iii) a domain adaptation module that aligns multiple cohorts, including both original and synthetic sources, into a unified representation space. A comprehensive description of the pipeline, including architectural details and implementation specifics, is provided in a companion study.[39]

Each component of the pipeline addresses a specific challenge associated with survival prediction in CPTAC. First, data generation using Denoising Diffusion Probabilistic Models (DDPM) was applied to expand data availability across five source datasets, including TCGA, ACRG, GSE15459, GSE26899, and GSE26901, generating synthetic samples based on cell-state score features derived from transcriptomic data. Second, a downstream survivability model was trained using a Multi-Domain Adversarial Network (MDAN), which integrates the original and synthetic source cohorts to predict 3-year survival as a binary classification task. Model outputs correspond to class probabilities, and model training was performed using a cross-entropy loss function. The MDAN explicitly learned feature distributions from the five source cohorts and their synthetic counterparts and adapted these representations to the target CPTAC cohort through unsupervised representation learning using the 124 unlabeled CPTAC samples. During downstream MDAN training, a task-aware sampling strategy was applied to select synthetic samples most relevant for CPTAC prediction. Because the majority of CPTAC samples lack long-term follow-up information and only 27 samples had available 3-year survival outcomes, these labeled samples were reserved as a held-out validation set.

Model performance was assessed through cross-validation on the source datasets and independent validation on the held-out CPTAC samples (Figure S3G). Together, these components form a cohesive validation framework for evaluating ecotype-associated survival signals under low-resource, high-variance conditions.

XGBoost-based pipeline for metabolic regulator prioritization

We used MetaSage,[45] an XGBoost-based pipeline to prioritize metabolic regulators. The human genome-scale metabolic model (GEM) was obtained from the Metabolic Atlas (https://metabolicatlas.org/, version 1.14.0),[83] which includes 13,024 metabolic reactions involving 8,363 metabolites and 2,920 genes. PubChem Compound IDs were used to map the dysregulated metabolites identified in this study to the GEM.

For each dysregulated metabolite, we compiled mRNA, protein, and PTM levels for all known enzymes involved in its synthesis or degradation, as well as precursor metabolite levels when available. Only genes quantified in either RNA-Seq or proteomic datasets were retained. Subsequently, the RNA and protein expression levels of these genes, along with all quantified PTM sites, were extracted from our multi-omics data and used as input features. Upstream metabolites (i.e., reactants) were also included as additional features if they were detected in the metabolomics dataset.

For each dysregulated metabolite with at least 1 input feature, an XGBoost regression model was trained to predict its intensity using the selected features. 5-fold cross-validation was repeated 10 times to obtain robust predicted intensities. Pearson correlation coefficients were calculated between the average predicted and measured metabolite intensities. Metabolites with a correlation coefficient $R > 0.5$ and a p-value <0.05 were considered predictable.

For each predictable metabolite, a final model was trained on the full dataset still using XGBoost, with Shapley values computed during training to assess feature importance. The average absolute Shapley value was calculated for each feature. Default XGBoost parameters were used, except for n_estimators = 100 and random_state = 1 through 10 for the repeated cross-validations.

For network visualization, associated genes were categorized into three groups based on the reactions they participated in.

1 Upstream genes: genes involved in reactions that produce the dysregulated metabolites;
2 Downstream genes: genes involved in reactions that consume the dysregulated metabolites;
3 Bi-directional genes: genes involved in reversible reactions involving the dysregulated metabolites.

In addition to the top-ranked feature with the highest average absolute Shapley value, other features were retained if their contribution was at least 10% of that of the top feature. For each metabolite, a maximum of five features were selected.

Microbiome analysis

Microbial classification from CPTAC NGS data

BAM files pre-aligned to hg38 for WGS and RNA Sequencing (RNA-Seq) were obtained from CPTAC. All BAM files were downloaded from the NCI Genomic Data Commons (GDC) API, where they were loaded and processed on the Kraken High Performance Cluster at Dana-Farber Cancer Institute. Matched WGS Blood-Derived Normal (BDN) and Primary Tumor (PT) BAM files were processed for 159 patients. 159 Primary Tumor and 27 Solid Tissue Normal (STN) BAM files were processed for RNA-Seq. SAMtools' flagstats (v1.16.1) was used to obtain total paired primary read counts prior to host subtraction. The first round of host subtraction used SAMtools (v1.16.1) to remove all properly paired reads (f −3) with an alignment score greater than 35 (AS > 35), with an exception for reads aligned to the viral genomes included in the hg38 decoy genome. Paired reads were extracted using bedtools (v2.30.0). Quality filtering was performed with fastp (0.23.4) with the following parameters (–detect_adapter_for_pe -D -n 10 -y -Y 20). WGS quality-passed paired reads were aligned to a custom human reference generated from the Telomere-to-Telomere human genome assembly (T2T-CHM13v2.0) and GATK PathSeq host reference database using bwa-mem (v0.7.17-r1188) with default parameters. GATK Pathseq host references included Immuno Polymorphism Database IMGT/HLA, NCBI UniVec, Gencode (human v25), and GenBank accessions KY503218-KY5808060. RNA-Seq quality-passed paired reads were aligned to T2T-CHM13v2.0 using STAR

(v2.7.5b) with default parameters. Properly paired reads with an alignment score greater than 35 were removed with SAMtools (v1.16.1) and unaligned paired reads were extracted for microbial classification. Kraken2 (v2.1.3) was used to classify putative microbial reads and estimate microbial abundances. Reads were aligned to the Kraken2 PlusPF database using Kraken2 default parameters in paired mode. Microbial count abundances from Kraken2 were normalized to total primary reads to obtain RPM for analysis in R version 4.2.1.

Decontamination of CPTAC microbial profiles using blood-derived normal and matched RNA-Seq samples

Contamination is a pressing issue in tissues with low microbial biomass, including tumor tissues. Tumor-resident species in the CPTAC cohort were decontaminated identified through two underlying biological assumptions: 1) Tissue-resident species are likely to be significantly more prevalent in the tumor samples than compared to blood samples, and 2) Microbiota living in tumor tissues are expected to be found in WGS and RNA-Seq data from matched samples. Tissue-resident species detected in WGS samples are significantly likely to be detected in their matched RNA-Seq sample.

With the first assumption of blood to be a microbial sterile site, WGS BDN samples were used as a negative control in the CPTAC cohort. Species that had a prevalence of less than 20% of WGS BDN (reads >1) samples and were significantly prevalent in WGS PT (reads >3) samples compared to BDN by a one-sided fisher's exact test ($q < 0.05$) were classified as likely tumor-resident species.

Total RNA-Seq of the CPTAC cohort has allowed for high quality detection of functional microbes in the gastric tumor environment. Therefore, relevant tissue-associated microbes should be detected from matched WGS and RNA-Seq samples of gastric tumors. Jaccard similarity test was performed on each species to determine the significant likelihood of presence in matched WGS (reads >3) and RNA-Seq (reads >1) PT samples.

The overlapping tumor-resident species identified from 1) WGS Tumor vs. Blood comparison and 2) WGS vs. RNA Tumor comparison were then used to create a conservative set of tissue-resident species for downstream analysis. Microbial counts of genera were recalculated using a mixture model accounting for the proportion of tissue-resident species to contaminant species. This method was used to re-estimate the tumor-associated microbial load of higher order taxonomic ranks for each patient WGS and RNA sample.

Enrichment analysis of high-mannose and fucose-sialic glycan types in association with Pathobiont cluster

Enrichment was calculated using a decreasing ranked list of signed -$\log_{10}$(P) between glycopeptides and microbial cluster abundances.

High outlier analysis

Therapeutic target classification was obtained from our previous work.[28] It was expanded to specify if targets were known ADC/Antibody or T cell targets,[154] Antibody Society (downloaded 02-29-24), and chimeric antigen receptors.[154] Essential genes defined by DepMap (21Q4)[77] were removed. Targets with both ADC/Antibody and T cell targets were classified as ADC/Antibody. High protein, site-specific-glycoform, gene level glycosylation abundance, and phosphosite abundance outliers were defined using a threshold of 3 MAD above mean of NAT samples and 2-fold higher than 90th percentile of NAT samples. Fold change was calculated based on the $\log_2$(TMT ratio) data included in the LinkedOmics data freeze. ADC targets, antibody targets, and T cell targets with at least 5 high protein outliers were highlighted, and other non-essential gene surface proteins with at least 15 high protein outliers were highlighted. For site-specific-glycoform outliers, a sample is considered a high outlier for a protein if any site-specific-glycoform for the protein is a high outlier. We highlight site-specific-glycoform outliers with at least 15 high outlier samples. Kinase activity high outliers were identified using a threshold of 3 MAD above mean of NAT samples. For internal target identification, high outliers were further filtered to only include targets with CRISPR KO dependency. CRISPR KO data from combined Broad and Sanger studies (CRISPR_gene_effect.csv) was downloaded from DepMap (21Q4)[77] and filtered to only include GC cell lines. CRISPR dependencies were determined from one-tailed *t* test of gene knockout effects (FDR <0.05). High outlier phosphorylation was filtered to only include activating sites from PhosphoSitePlus (Downloaded on March 2022). We highlight internal targets that have at least 10 high outlier samples.

Pathway enrichment analysis

All pathway enrichment analysis was done using WebGestalt 2024.[133]

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using R unless explained otherwise. Details of the statistical tests used, sample sizes, and statistical parameters such as definition of center and dispersion and precision measures are reported in the main text, figure legends, STAR Methods, and supplementary information. P-values were adjusted by the Benjamini-Hochberg procedure.

Additional resources

Details on CPTAC program initiatives, teams, partners, and governing body committees are provided on the CPTAC website (https://proteomics.cancer.gov/programs/cptac).