Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Molecular Classification of Breast Cancer Using a Cuproptosis Related Gene Panel and a Robust REO Ensemble Classifier

Yuxin Shi

charlotte.shi123@gmail.com

Northwestern University https://orcid.org/0009-0004-1682-3746

Research Article

Keywords: Breast cancer, Cuproptosis, Molecular classification, Relative expression ordering, Machine learning, Tumor microenvironment, Precision oncology

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11598892/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Cuproptosis is a mitochondria-dependent form of regulated cell death. Its interplay with cancer metabolism can modulate tumor aggressiveness. We hypothesized that cuproptosis-related gene expression defines distinct breast cancer (BRCA) subtypes and that a relative expression ordering (REO) ensemble classifier built on a 19-gene panel (CRG19) yields a technically robust, clinically translatable tool.

Methods

We curated a mechanistically grounded 19-gene panel (CRG19) and analyzed primary BRCA RNA-seq samples from TCGA (n = 986, split into training n = 838 and testing n = 148). Consensus clustering (k = 3) identified molecular subtypes. From 171 pairwise gene relations, the top 100 discriminative REO pairs (chi-square ranking) were used to build an ensemble of four machine-learning models (SVM, Random Forest, Elastic Net, XGBoost) within a nested 5x5 cross-validation framework. Tumor microenvironment (CIBERSORTx), pathway activity (GSVA), and drug sensitivity (oncoPredict/CTRP2) were evaluated. External validation was performed on an independent cohort (GSE202203, n = 3,207).

Results

Consensus clustering defined three reproducible subtypes (A, B, and C) with distinct biology and clinical outcomes (p = 0.035). Subtype A was enriched for basal-like/HER2-enriched PAM50 classes, showing pronounced cell cycle activation and poorer early survival. Subtype B was predominantly luminal with active hormone signaling. Subtype C exhibited a quiescent state and the best long-term prognosis. The REO ensemble achieved high probability calibration and stability between training and test sets (KS statistics 0.12–0.14). Subtype A showed higher predicted chemosensitivity, whereas Subtypes B/C were predicted to respond better to endocrine therapy. In external validation, differential pathway patterns across subtypes were highly reproducible (all-pathway Spearman r = 0.838–0.925, weighted direction consistency = 0.947).

Conclusion

The CRG19 taxonomy defines biologically and clinically distinct breast cancer subtypes. The accompanying REO ensemble classifier provides a technically robust, cross-platform tool suitable for clinical risk stratification and therapeutic decision-making.

1. Introduction

Breast cancer represents a highly heterogeneous malignancy driven by distinct genetic, epigenetic, and metabolic alterations[1–3]. Although established molecular stratification systems such as PAM50 and multigene assays like Oncotype DX have substantially advanced risk prediction and treatment planning, substantial clinical and biological heterogeneity persists within current subtypes[4, 5]. Furthermore, conventional mRNA expression signatures frequently suffer from technical batch variation, platform dependency, and cohort-level normalization artifacts, hindering seamless point-of-care clinical translation[6, 7].

Recent breakthroughs have identified cuproptosis as a novel, mitochondria-dependent form of regulated cell death triggered by direct copper binding to lipoylated proteins within the tricarboxylic acid (TCA) cycle[8]. This binding induces cytotoxic lipoylated protein aggregation and loss of iron-sulfur cluster proteins, ultimately causing proteotoxic stress and cell death[8]. Copper is a dual-edge metal in oncology: Physiological levels support tumor growth by fueling ETC, TCA cycle, and antioxidant defenses, while Excess levels: induce cuproptosis, which can limit tumor growth if the tumor cannot adapt. In aggressive cancers, altered copper metabolism (e.g., upregulated copper transporters, high mitochondrial copper) can Promote cuproptosis under high copper stress, potentially reducing tumor aggressiveness or Enable resistance if tumors upregulate copper efflux or chaperones, allowing survival despite copper overload[9, 10]. Because metabolic rewiring and mitochondrial dependence are cardinal features of breast cancer subtypes[11–13], exploiting cuproptosis is a promising anticancer strategy. However, no systematic classification of breast cancer based on the coordinated expression of a comprehensive panel of cuproptosis-related genes exists, nor is there a practical tool for identifying such subtypes in clinical samples. This gap is not addressed by current genomic assays such as Oncotype DX and PAM50, or by most machine-learning models [14–17]. we hypothesized that cuproptosis-associated expression profiles define clinically distinct molecular taxonomies.

In this study, we curated a mechanistically grounded 19-gene cuproptosis panel (CRG19) spanning copper transport, protein lipoylation, TCA cycle targets, and adaptive stress response pathways. To ensure cross-platform robustness, we employed Relative Expression Ordering (REO)—a within-sample binary transformation method—and built an ensemble machine-learning classifier (SVM, Random Forest, and Elastic Net) within a nested 5×5 cross-validation framework. We systematically characterized the subtype-specific pathway activities, clinical survival, tumor microenvironment (TME) architectures, and drug vulnerabilities, followed by extensive external validation in an independent cohort (GSE202203, n = 3,207) and deployment as an interactive Shiny web application.

2. Materials and Methods

The overall analytical workflow and study design, encompassing data curation, subtype identification, REO classifier construction, and multi-dimensional validation, are schematically illustrated in Fig. 1.

2.1 Data Acquisition, Preprocessing, and Cohort Splitting

Primary breast tumor transcriptomic profiles and corresponding clinical annotations were retrieved from The Cancer Genome Atlas (TCGA-BRCA; n = 1,231). Inclusion criteria required female primary breast carcinoma, available non-zero survival follow-up (> 30 days), and complete panel expression. The final analytic cohort comprised n = 986 tumor samples. RNA-seq data were processed as Transcripts Per Million (TPM) and log2-transformed. The cohort was partitioned into a stratified training set (n = 838) and an independent testing set (n = 148), balanced for demographics and clinical stage. For independent cross-platform validation, transcriptomic and clinical data from GSE202203 (n = 3,207) were analyzed under an identical processing pipeline.

2.2 Curation of the CRG19 Gene Panel

A structured literature review prioritized mechanistically validated cuproptosis regulators. A 19-gene panel (CRG19) was established to cover four functional tiers: (1) Upstream copper transport/homeostasis (SLC31A1, ATP7A, ATP7B, MTF1); (2) Protein lipoylation cascade (FDX1, LIAS, LIPT1, LIPT2); (3) Lipoylated TCA cycle substrates (DLAT, DLD, DLST, DBT, GCSH); and (4) Metabolic flux and adaptive stress responses (PDHA1, PDHB, GLS, CDKN2A, NLRP3, NFE2L2).

2.3 Differential Expression, Enrichment Analysis, and Consensus Clustering

Differential expression between tumor and normal adjacent tissues was determined using Wilcoxon rank-sum tests. GO Biological Process and KEGG pathway over-representation analyses were performed using clusterProfiler with Benjamini-Hochberg (BH) false discovery rate (FDR) adjustment. Prognostic values were evaluated via univariate Cox proportional hazards regression and Kaplan-Meier analysis with optimal cutpoint determination.

Unsupervised consensus clustering was conducted on the training cohort (n = 838) using ConsensusClusterPlus (Pearson distance, 80% item resampling, 1,000 iterations). The optimal cluster number (k = 3) was selected based on the cumulative distribution function (CDF) area curve. Gene Set Variation Analysis (GSVA) quantified activity across canonical MSigDB KEGG pathways, with differential pathway activity identified using limma (FDR < 1e-8).

2.4 REO Feature Engineering and Ensemble Machine Learning

To establish a classifier resistant to technical batch effects, binary Relative Expression Ordering (REO) features were derived for all 171 pairwise gene combinations from the 19-gene panel. For pair (Gene A, Gene B), the feature value was assigned as 1 if Expression(Gene A) > Expression(Gene B) and 0 otherwise. Feature selection retained the top 100 discriminative REO pairs ranked by Chi-square test against consensus subtype labels.

Model training was performed within a nested 5×5 cross-validation framework combining Linear SVM, Random Forest (ranger), Elastic Net (glmnet), and Extreme Gradient Boosting (xgbTree). Ensemble class assignment was derived by averaging class-probability vectors across all four algorithms. Model stability was evaluated through 100 perturbation iterations (10% feature dropout, 5% bit-flip, 80% subsampling). Probability calibration was validated using Kolmogorov-Smirnov (KS) tests between training and testing predictions.

2.5 Cross-Cohort Reproducibility Validation

Cross-platform generalizability was tested in GSE202203 (n = 3,207). Differential pathway activity t-statistics across 658 KEGG pathways were derived for pairwise subtype contrasts. Concordance was evaluated using Spearman and Pearson correlation, cosine similarity, and weighted direction consistency.

2.6 Tumor Microenvironment and Drug Sensitivity Profiling

Immune cell infiltration fractions were estimated using CIBERSORTx (LM22 matrix, 1,000 permutations). Subtype differences in immune fractions were analyzed using Kruskal-Wallis tests followed by post-hoc pairwise Wilcoxon tests with BH FDR adjustment. In silico drug sensitivity (log2 IC50) across 45 candidate drugs was predicted using oncoPredict trained on CTRP2 reference models.

2.7 Clinical Associations and Web Deployment

Associations with PAM50 intrinsic subtypes, pathological histology, and clinical stage were evaluated using Chi-square tests, Cramér's V, and Kruskal-Wallis tests. The final REO ensemble pipeline was packaged as an interactive Shiny web application for single-sample subtype prediction.

3. Results

3.1 Expression Profile, Functional Enrichment, and Prognostic Value of CRG19

Functional enrichment confirmed that CRG19 genes are heavily involved in mitochondrial oxidative metabolism, acetyl-CoA/thioester synthesis, and the citrate cycle (Fig. 2A, 2B). In TCGA-BRCA, 18 of the 19 CRGs demonstrated significant dysregulation between tumor and normal tissues ($p < 0.05$; Fig. 2C). Core lipoylation genes (FDX1, LIAS, LIPT1) were

downregulated in tumors, whereas ATP7B and CDKN2A were upregulated. Kaplan-Meier survival analysis confirmed significant prognostic power for 18 CRGs (log-rank $p < 0.05$; Fig. 2D).

3.2 Identification of Three Molecular Subtypes by Consensus Clustering

Unsupervised consensus clustering on TCGA training data identified three distinct subtypes: Subtypes A, B, and C ($k = 3$; Fig. 3A–D). Overall survival differed significantly across subtypes (log-rank $p = 0.035$; Fig. 3E). GSVA profiling revealed distinct signaling architectures (Fig. 3F–H): Subtype A exhibited hyperactive cell cycle and DNA replication; Subtype B was enriched for estrogen/progesterone receptor signaling; and Subtype C displayed a quiescent signaling phenotype.

3.3 Construction and Evaluation of the REO Ensemble Classifier

The REO ensemble classifier achieved excellent probability calibration between training (n = 838) and testing (n = 148) sets (KS statistics 0.12–0.14; Fig. 4A–C). Survival curves demonstrated consistent trajectory ordering across consensus, predicted training, and predicted testing labels (Fig. 4D–F), with Subtype C demonstrating superior long-term prognosis (no mortality events beyond 5 years).

3.4 High Reproducibility in Independent External Dataset (GSE202203)

In GSE202203 (n = 3,207), pathway perturbation t-statistics correlated strongly across cohorts (Spearman $r = 0.838–0.925$, $p < 1e\text{-}174$; Fig. 5B–D). Top 20 pathways per contrast achieved 100% directional consistency (Fig. 5A), yielding an overall weighted direction consistency of 0.947.

3.5 Association with Clinicopathological Features and Stage Independence

Subtypes correlated strongly with PAM50 intrinsic classes (Cramér's V = 0.541; Fig. 6C, Table 1). Subtype A was enriched for Basal-like (60.1%) and HER2-enriched (16.3%) tumors; Subtype B was dominated by Luminal A (68.1%) and Luminal B (23.7%); and Subtype C comprised predominantly Luminal A (72.6%) with the lowest overall mortality (3.2%). Crucially, ML subtypes showed no significant association with clinical TNM stage ($p = 0.741$; Fig. 6C), demonstrating independent prognostic value.

3.6 Subtype-Specific Tumor Microenvironment Architecture

CIBERSORTx deconvolution demonstrated distinct immune microenvironments (Fig. 6A): Subtype A harbored a pro-inflammatory TME enriched for M1 macrophages and Tfh cells; Subtype B presented a 'cold' phenotype with reduced CD8 + T cells and increased M2 macrophages; and Subtype C exhibited an effector-suppressive mixed phenotype with elevated CD8 + T cells, plasma cells, and Tregs.

3.7 Distinct Therapeutic Vulnerabilities and Predicted Drug Sensitivity

oncoPredict profiling revealed striking therapeutic vulnerabilities (Table 2, Fig. 6B): Subtypes B and C showed superior predicted sensitivity to fulvestrant ($p = 3.99e\text{-}62$), whereas Subtype A displayed significantly heightened predicted sensitivity to chemotherapeutic agents including docetaxel, paclitaxel, vincristine, and topotecan ($p < 0.001$).

4. Discussion

In this study, we established a cuproptosis-based molecular taxonomy of breast cancer (BRCA) using a mechanistically curated 19-gene panel (CRG19) and developed a cross-platform robust machine-learning classifier based on Relative Expression Ordering (REO). Integrating data from TCGA-BRCA (n = 986) and a large external validation cohort (GSE202203, n = 3,207), we identified three distinct molecular subtypes (A, B, C) characterized by divergent biological signaling, clinical outcomes, microenvironmental architectures, and therapeutic vulnerabilities. Crucially, our REO ensemble classifier achieved high probability calibration, strong perturbation stability, and excellent cross-cohort reproducibility without relying on cross-sample normalization, highlighting its strong potential for clinical translation.

4.1 Biological Mechanism and the 'Metabolic Double-Edged Sword' Hypothesis

Cuproptosis represents a distinct, mitochondria-dependent form of regulated cell death triggered by direct copper binding to lipoylated proteins of the tricarboxylic acid (TCA) cycle—most notably dihydrolipoamide S-acetyltransferase (DLAT)—leading to cytotoxic protein aggregation, loss of iron-sulfur cluster proteins, and proteotoxic stress[8]. The 19 genes in our panel encompass all key functional cascade levels: copper transport (SLC31A1, ATP7A/B, MTF1)[18–20], protein lipoylation (FDX1, LIAS, LIPT1/2) [8, 21], TCA cycle substrates (DLAT, DLD, DLST, DBT, GCSH) [8, 21], and stress response regulators (CDKN2A, NLRP3, NFE2L2)[22–25].

A critical, paradoxical observation in our single-gene prognostic analysis was that high expression of core cuproptosis executioners (such as FDX1, LIAS, and LIPT1) correlated with poorer overall survival. At first glance, this appears counterintuitive for pro-cell-death genes. However, when interpreted through the lens of cancer metabolic reprogramming, this paradox unveils a 'metabolic double-edged sword' (or metabolic Achilles' heel) in breast cancer. Highly proliferative, aggressive tumors (such as Basal-like/TNBC tumors enriched in Subtype A) exhibit hyperactive cell-cycle progression and heavy reliance on mitochondrial TCA flux to generate bioenergetic precursors and biosynthetic intermediates[26, 27]. To sustain this rapid turnover, these cells upregulate FDX1 and protein lipoylation machinery (LIAS, LIPT1/2) to maintain DLAT/DLST activity[28]. While this metabolic dependency supports aggressive tumor growth, it simultaneously sensitizes Subtype A cells to copper overload and proteotoxic stress[8]. This explains why Subtype A tumors exhibit high susceptibility to cytotoxic agents (such as docetaxel and paclitaxel) and represents a strong biological rationale for targeting Subtype A tumors with copper ionophores (e.g., Elesclomol or Disulfiram) in combination with mitochondrial toxicants.

Conversely, Subtype C displayed a 'quiescent' metabolic state characterized by attenuated RTK/PI3K signaling, reduced cell cycle turnover, and low TCA flux. This reduced metabolic rate lowers basal mitochondrial ROS production and copper accumulation stress[29–31], allowing Subtype C tumors to exist in an indolent, metabolically dormant state. Consequently, patients in Subtype C experienced superior 5-year survival, highlighting how low cuproptosis stress aligns with long-term clinical stability.

4.2 Immunological Architecture and the Immune Equilibrium Hypothesis

Characterizing the tumor microenvironment (TME) across the CRG19 taxonomy provided profound insights into the cross-talk between cuproptosis metabolic signaling and antitumor immunity:

1. Subtype A (Pro-inflammatory / 'Hot' TME): Subtype A harbored significant enrichment of pro-inflammatory M1 macrophages and T follicular helper (Tfh) cells. Rapid proliferation and high metabolic turnover in Subtype A tumors generate extensive neoantigenic loads and damage-associated molecular patterns (DAMPs), driving M1 macrophage polarization and Tfh-mediated germinal center responses[32–34]. However, despite this pro-inflammatory milieu, the hyperactive cell-cycle velocity of Subtype A overrides immune control, necessitating combination strategies that pair immune checkpoint blockade or copper ionophores with intensive cytotoxic chemotherapy.

2. Subtype B (Immunologically 'Cold' TME): Subtype B, dominated by Luminal A/B tumors, exhibited an immunologically cold landscape characterized by depleted CD8+ T cells, enriched M2 macrophages, and elevated resting mast cells. Active estrogen receptor (ER) signaling in Luminal tumors is known to suppress NF-kB signaling and inhibit pro-inflammatory chemokine secretion, promoting M2 macrophage polarization and creating a immunosuppressive barrier[35, 36]. This explains Subtype B's resistance to chemotherapy and immunotherapy, supporting a therapeutic strategy focused on endocrine monotherapy (e.g., Fulvestrant) or ER degradation rather than toxic chemotherapy.

3. Subtype C (Immune Equilibrium / Effector-Suppressive Mixed TME): Subtype C exhibited concurrent infiltration of cytotoxic CD8+ T cells, plasma cells, and regulatory T cells (Tregs). This mixed infiltration pattern exemplifies the classic 'Immune Equilibrium' phase of cancer immunoediting: effector CD8+ T cells and plasma cells exert continuous immune surveillance that suppresses tumor outgrowth, while infiltration of Tregs prevents overt tissue destruction and blunts hyper-inflammation[37–39]. This balanced immune-tumor dynamic accounts for the indolent disease course and exceptional long-term survival observed in Subtype C.

4.3 Methodological Rigor of REO Feature Engineering and Biological Continuity

Translating transcriptomic classifiers into single-sample clinical assays (N-of-1 diagnostics) is notoriously difficult due to batch effects, platform disparities (RNA-seq vs microarrays vs RT-qPCR), and cohort-dependent normalization artifacts (e.g., ComBat or Z-score normalization)[40, 41]. We addressed this challenge by using Relative Expression Ordering (REO), which converts continuous mRNA expression into qualitative within-sample binary ranks (Gene A > Gene B). Because relative ranks are invariant to monotonic transformations, experimental noise, and batch variation, the top 100 REO pairs selected in our pipeline provided an exceptionally stable foundation for cross-cohort classification.

In nested 5x5 cross-validation and perturbation tests (dropout, bitflip, subsampling), our ensemble model demonstrated high probability calibration (KS statistics 0.12–0.14) and strong stability (> 0.90). Interestingly, model agreement was highest for Subtype B and lowest for Subtype C. Rather than representing a computational defect, this variation reflects fundamental biological continuity: Subtype C represents the quietest tail of the Luminal A spectrum. The boundary between Subtype B (active Luminal) and Subtype C (quiescent Luminal) represents a continuous gradient of luminal metabolic activity rather than a discrete genomic threshold. Recognizing this continuum is essential for clinical decision-making, where Subtype C represents candidates for treatment de-escalation.

4.4 Precision Medicine and Clinical Decision Tree Guidance

The CRG19 taxonomy offers a clear, actionable clinical decision framework for personalized breast cancer management:

- Subtype A (High Risk / Chemosensitive / High Metabolic Stress): Patients should be prioritized for intensive cytotoxic regimens (Docetaxel/Paclitaxel) and represent primary candidates for clinical trials evaluating copper ionophores (Elesclomol, Disulfiram) or mitochondrial complex inhibitors.

- Subtype B (Moderate Risk / Endocrine Sensitive / Immunologically Cold): Chemotherapy de-escalation is recommended to avoid unnecessary toxicity. First-line management should center on fulvestrant, selective estrogen receptor degraders (SERDs), or CDK4/6 inhibitor combinations.

- Subtype C (Low Risk / Quiescent / Superior Prognosis): Represents low-risk Luminal tumors eligible for therapy de-escalation, long-term endocrine surveillance, and avoidance of overtreatment.

To facilitate clinical translation, we packaged the entire pipeline into an interactive Shiny web application, enabling single-sample subtyping from raw gene expression profiles without computational overhead.

4.5 Limitations and Future Trajectories

Several limitations warrant consideration. First, while drug sensitivities were inferred using oncoPredict calibrated against CTRP2 reference models, prospective clinical validation and patient-derived xenograft (PDX) assays are required to confirm in vivo therapeutic responses[42]. Second, wet-lab functional assays (e.g., shRNA knockdown or CRISPR editing of FDX1, LIAS, and LIPT1 in BRCA cell lines) are needed to measure direct changes in mitochondrial respiration, DLAT lipoylation, and copper susceptibility. Third, translating the 19-gene panel to low-throughput diagnostic platforms (e.g., multiplex RT-qPCR or IVD chips) will require clinical validation.

4.6 Conclusion

In conclusion, the CRG19 molecular taxonomy establishes a biologically grounded, clinically predictive framework for breast cancer subtyping based on cuproptosis metabolism. Paired with a technically robust REO ensemble classifier, this approach bridges complex cancer genomics with clinical translation, offering a powerful tool for risk stratification, microenvironmental characterization, and precision oncology.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The transcriptomic and clinical datasets analyzed in this study are publicly available in open-access repositories:

- GEO Discovery & Validation Cohorts: RNA-sequencing matrices for GSE202203 were downloaded from the NCBI Gene Expression Omnibus (GEO) repository (https://www.ncbi.nlm.nih.gov/geo/).
- TCGA Validation Cohort: Primary transcriptomic profiles and clinical annotations for the TCGA-BRCA cohort were retrieved from the NCI Genomic Data Commons (GDC) Data Portal (https://portal.gdc.cancer.gov/) and cBioPortal for Cancer Genomics (https://www.cbioportal.org/).
- Therapeutic Sensitivity & Functional Gene Sets: Drug response matrices from the Cancer Therapeutics Response Portal v2 (CTRP2) were accessed via the Broad Institute Portal (https://portals.broadinstitute.org/ctrp/). Pathway collections for Gene Set Variation Analysis (GSVA) were curated from the Molecular Signatures Database (MSigDB v2025.1.Hs) (https://www.gsea-msigdb.org/gsea/msigdb).

Competing interests

The author declares no competing interests.

Funding

Not applicable

Authors' contributions

YS: Substantial contribution to the conception, design of the work, the acquisition, analysis, and interpretation of data, manuscript draft.

Acknowledgements

The author would like to acknowledge Vijayakrishna Gadi, MD, PhD (University of Illinois Chicago) for providing valuable clinical and scientific insights and for offering constructive feedback on earlier versions of this manuscript. The author also extends sincere thanks to Mr. Valentin Torres (biology educator, Walter Payton College Prep) for encouragement and thoughtful guidance throughout the development of this project.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT and DeepSeek to synthesize literature and improve language and readability. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Guo L, Kong D, Liu J et al (2023) Breast cancer heterogeneity and its implication in personalized precision therapy. Exp Hematol Oncol 12:3. https://doi.org/10.1186/s40164-022-00363-1
2. Testa U, Castelli G, Pelosi E (2020) Breast Cancer: A Molecularly Heterogenous Disease Needing Subtype-Specific Treatments. Med Sci 8:18. https://doi.org/10.3390/medsci8010018
3. Lüönd F, Tiede S, Christofori G (2021) Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. Br J Cancer 125:164–175. https://doi.org/10.1038/s41416-021-01328-7
4. Haywood SC, Gupta S, Heemers HV (2022) PAM50 and beyond: when will tissue transcriptomics guide clinical decision-making? Eur Urol Focus 8:916–918. https://doi.org/10.1016/j.euf.2022.08.010
5. Wallrabenstein T, Chiru ED, Sonderegger M et al (2026) Comparison of a Novel MRNA-Based Breast Cancer Subtyping Assay with PAM50 and Oncotype DX in Estrogen Receptor-Positive/HER2-Negative Breast Cancer: An Exploratory Age-Stratified Retrospective Cohort Study. Curr Issues Mol Biol 48:886. https://doi.org/10.3390/cimb48090886
6. Kreitmann L, D'Souza G, Miglietta L et al (2024) A computational framework to improve cross-platform implementation of transcriptomics signatures. https://doi.org/10.1016/j.ebiom.2024.105204. eBioMedicine 105:
7. Borisov N, Buzdin A (2022) Transcriptomic Harmonization as the Way for Suppressing Cross-Platform Bias and Batch Effect. Biomedicines 10:2318. https://doi.org/10.3390/biomedicines10092318
8. Tsvetkov P, Coy S, Petrova B et al (2022) Copper induces cell death by targeting lipoylated TCA cycle proteins. Science 375:1254–1261. https://doi.org/10.1126/science.abf0529
9. Xie J, Yang Y, Gao Y, He J (2023) Cuproptosis: mechanisms and links with cancers. Mol Cancer 22:46. https://doi.org/10.1186/s12943-023-01732-y

10. Jiang D, Zhuang L, Koong AC, Gan B (2026) Cuproptosis in cancer: from molecular mechanisms to therapeutic intervention. Trends Cancer 12:275–286. https://doi.org/10.1016/j.trecan.2025.12.002
11. Menegollo M, Bentham RB, Henriques T et al (2024) Multistate Gene Cluster Switches Determine the Adaptive Mitochondrial and Metabolic Landscape of Breast Cancer. Cancer Res 84:2911–2925. https://doi.org/10.1158/0008-5472.CAN-23-3172
12. Iqbal MA, Siddiqui S, Smith K et al (2023) Metabolic stratification of human breast tumors reveal subtypes of clinical and therapeutic relevance. https://doi.org/10.1016/j.isci.2023.108059. iScience 26:
13. Ripoll C, Roldan M, Ruedas-Rama MJ et al (2021) Breast Cancer Cell Subtypes Display Different Metabolic Phenotypes That Correlate with Their Clinical Classification. Biology 10:1267. https://doi.org/10.3390/biology10121267
14. Zhang B, Xie L, Liu J et al (2023) Construction and validation of a cuproptosis-related prognostic model for glioblastoma. Front Immunol 14. https://doi.org/10.3389/fimmu.2023.1082974
15. Sha R, Dong X, Yan S et al (2024) Cuproptosis-related genes predict prognosis and trastuzumab therapeutic response in HER2-positive breast cancer. Sci Rep 14:2908. https://doi.org/10.1038/s41598-024-52638-8
16. Wang Y, Wang Y, Zhang Y (2020) Identification of prognostic signature of non–small cell lung cancer based on TCGA methylation data. Sci Rep 10:8575. https://doi.org/10.1038/s41598-020-65479-y
17. Wu Y, Fan Y, Dong X et al (2025) A multi-omics prognostic model of cuproptosis affects the prognosis of stomach adenocarcinoma. Clin Epigenetics 17:99. https://doi.org/10.1186/s13148-025-01894-0
18. Balamurugan K, Egli D, Hua H et al (2007) Copper homeostasis in Drosophila by complex interplay of import, storage and behavioral avoidance. EMBO J 26:1035–1044. https://doi.org/10.1038/sj.emboj.7601543
19. Kim B-E, Nevitt T, Thiele DJ (2008) Mechanisms for copper acquisition, distribution and regulation. Nat Chem Biol 4:176–185. https://doi.org/10.1038/nchembio.72
20. Lutsenko S (2010) Human copper homeostasis: a network of interconnected pathways. Curr Opin Chem Biol 14:211–217. https://doi.org/10.1016/j.cbpa.2010.01.003
21. Tang D, Chen X, Kroemer G (2022) Cuproptosis: a copper-triggered modality of mitochondrial cell death. Cell Res 32:417–418. https://doi.org/10.1038/s41422-022-00653-7
22. Li J, Yang X, Yin C et al (2025) CDKN2A, a key gene in copper-induced cell death model, influencing melanoma invasion and apoptosis. Discov Oncol 16:246. https://doi.org/10.1007/s12672-025-01992-8
23. Sharpless NE, Sherr CJ (2015) Forging a signature of in vivo senescence. Nat Rev Cancer 15:397–408. https://doi.org/10.1038/nrc3960
24. Swanson KV, Deng M, Ting JP-Y (2019) The NLRP3 inflammasome: molecular activation and regulation to therapeutics. Nat Rev Immunol 19:477–489. https://doi.org/10.1038/s41577-019-0165-0
25. Tonelli C, Chio IIC, Tuveson DA (2018) Transcriptional Regulation by Nrf2. Antioxid Redox Signal 29:1727–1745. https://doi.org/10.1089/ars.2017.7342
26. Chen Z, Liu Y, Lyu M et al (2025) Classifications of triple-negative breast cancer: insights and current therapeutic approaches. Cell Biosci 15:13. https://doi.org/10.1186/s13578-025-01359-0
27. Ensenyat-Mendez M, Llinàs-Arias P, Orozco JIJ et al (2021) Current Triple-Negative Breast Cancer Subtypes: Dissecting the Most Aggressive Form of Breast Cancer. Front Oncol 11. https://doi.org/10.3389/fonc.2021.681476
28. Dreishpoon MB, Bick NR, Petrova B et al (2023) FDX1 regulates cellular protein lipoylation through direct binding to LIAS. J Biol Chem 299:105046. https://doi.org/10.1016/j.jbc.2023.105046
29. Bresnick AR, Backer JM (2019) PI3Kβ—A Versatile Transducer for GPCR, RTK, and Small GTPase Signaling. Endocrinology 160:536–555. https://doi.org/10.1210/en.2018-00843

30. Su WY, Tian LY, Guo LP et al (2023) PI3K signaling-regulated metabolic reprogramming: From mechanism to application. Biochim Biophys Acta Rev Cancer 1878:188952. https://doi.org/10.1016/j.bbcan.2023.188952
31. Han B, Lin X, Hu H (2024) Regulation of PI3K signaling in cancer metabolism and PI3K-targeting therapy. Transl Breast Cancer Res 5:33. https://doi.org/10.21037/tbcr-24-29
32. Chen R, Zou J, Liu J et al (2025) DAMPs in the immunogenicity of cell death. Mol Cell 85:3874–3889. https://doi.org/10.1016/j.molcel.2025.09.007
33. Jiang J, Yan Y, Yang C, Cai H (2025) Immunogenic Cell Death and Metabolic Reprogramming in Cancer: Mechanisms, Synergies, and Innovative Therapeutic Strategies. Biomedicines 13:950. https://doi.org/10.3390/biomedicines13040950
34. Yanai H, Hangai S, Taniguchi T (2021) Damage-associated molecular patterns and Toll-like receptors in the tumor immune microenvironment. Int Immunol 33:841–846. https://doi.org/10.1093/intimm/dxab050
35. Rothenberger NJ, Somasundaram A, Stabile LP (2018) The Role of the Estrogen Pathway in the Tumor Microenvironment. Int J Mol Sci 19:611. https://doi.org/10.3390/ijms19020611
36. Chakraborty B, Byemerwa J, Krebs T et al (2023) Estrogen Receptor Signaling in the Immune System. Endocr Rev 44:117–141. https://doi.org/10.1210/endrev/bnac017
37. Dikiy S, Rudensky AY (2023) Principles of regulatory T cell function. Immunity 56:240–255. https://doi.org/10.1016/j.immuni.2023.01.004
38. McKay A, Cruz JA, Taylor I et al Decoding Treg diversity and dysfunction to advance Treg-based therapies in autoimmune and inflammatory diseases. Front Immunol 17:1870050. https://doi.org/10.3389/fimmu.2026.1870050
39. Goswami TK, Singh M, Dhawan M et al Regulatory T cells (Tregs) and their therapeutic potential against autoimmune disorders – Advances and challenges. Hum Vaccines Immunother 18:2035117. https://doi.org/10.1080/21645515.2022.2035117
40. Kreitmann L, D'Souza G, Miglietta L et al (2024) A computational framework to improve cross-platform implementation of transcriptomics signatures. https://doi.org/10.1016/j.ebiom.2024.105204. eBioMedicine 105:
41. Li X, Cai H, Wang X et al (2019) A rank-based algorithm of differential expression analysis for small cell line data with statistical control. Brief Bioinform 20:482–491. https://doi.org/10.1093/bib/bbx135
42. Wang X, Zhu J, Li L et al (2024) Utility of patient-derived xenografts to evaluate drug sensitivity and select optimal treatments for individual non-small-cell lung cancer patients. Mol Med 30:209. https://doi.org/10.1186/s10020-024-00934-4

Tables

Table 1
Cohort Demographics and Clinical Features by Cuproptosis Subtype (TCGA-BRCA, n = 986)

Variable	A	B	C	P-value
age at diagnosis	56 (28–89)	58.5 (26–89)	61.5 (30–89)	0.064
days to last follow up	773.5 (-7– 7777)	780 (0–8008)	879 (0–8556)	0.239
race				< 0.001
american indian or alaska native	1 (0.4%)			
asian	21 (7.6%)	30 (5.1%)	8 (6.5%)	
black or african american	78 (28.3%)	50 (8.5%)	31 (25%)	
not reported	17 (6.2%)	63 (10.8%)	2 (1.6%)	
white	159 (57.6%)	443 (75.6%)	83 (66.9%)	
ethnicity				0.005
Unknown	2 (0.7%)	5 (0.9%)	1 (0.8%)	
hispanic or latino	5 (1.8%)	21 (3.6%)	7 (5.6%)	
not hispanic or latino	235 (85.1%)	452 (77.1%)	109 (87.9%)	
not reported	34 (12.3%)	108 (18.4%)	7 (5.6%)	
metastasis at diagnosis				0.057
No	85 (30.8%)	226 (38.6%)	36 (29%)	
Yes	9 (3.3%)	8 (1.4%)	1 (0.8%)	
NA	182 (65.9%)	352 (60.1%)	87 (70.2%)	
stage				0.011
Stage_I	34 (12.3%)	107 (18.3%)	26 (21%)	
Stage_II	184 (66.7%)	319 (54.4%)	69 (55.6%)	
Stage_III	46 (16.7%)	140 (23.9%)	28 (22.6%)	
Stage_IV	7 (2.5%)	9 (1.5%)	1 (0.8%)	
NA	5 (1.8%)	11 (1.9%)		
pathology				< 0.001
IDC	170 (61.6%)	269 (45.9%)	39 (31.5%)	
ILC	7 (2.5%)	92 (15.7%)	18 (14.5%)	
Mixed	13 (4.7%)	61 (10.4%)	7 (5.6%)	
Other	29 (10.5%)	65 (11.1%)	10 (8.1%)	
NA	57 (20.7%)	99 (16.9%)	50 (40.3%)	
subtype pam50				< 0.001
Basal	166 (60.1%)	5 (0.9%)	1 (0.8%)	

Variable	A	B	C	P-value
Her2	45 (16.3%)	27 (4.6%)	6 (4.8%)	
LumA	23 (8.3%)	399 (68.1%)	90 (72.6%)	
LumB	29 (10.5%)	139 (23.7%)	21 (16.9%)	
Normal	13 (4.7%)	16 (2.7%)	6 (4.8%)	
vital status				0.016
Alive	243 (88%)	517 (88.2%)	120 (96.8%)	
Dead	33 (12%)	69 (11.8%)	4 (3.2%)	

Table 2
Top Predicted Differential Drug Sensitivities across Subtypes

drug	median			p	fdr	cohen_d			pairwise_p		
	A	B	C			AB	AC	BC	AB	AC	BC
fulvestrant	20.852	20.495	20.746	< 0.001	< 0.001	1.188	0.250	-1.064	< 0.001	0.023	< 0.001
bosutinib	16.098	16.699	16.225	< 0.001	< 0.001	-0.873	-0.159	0.811	< 0.001	0.405	< 0.001
docetaxel	6.976	9.501	11.254	< 0.001	< 0.001	-0.868	-1.237	-0.714	< 0.001	< 0.001	< 0.001
vincristine	9.131	11.925	11.491	< 0.001	< 0.001	-1.156	-0.801	0.248	< 0.001	< 0.001	0.018
alisertib	16.730	17.835	17.881	< 0.001	< 0.001	-0.883	-0.602	0.151	< 0.001	< 0.001	0.278
topotecan	11.892	14.242	12.739	< 0.001	< 0.001	-0.975	-0.194	0.814	< 0.001	0.125	< 0.001
paclitaxel	8.715	11.932	10.746	< 0.001	< 0.001	-0.991	-0.660	0.282	< 0.001	< 0.001	0.006
teniposide	14.010	16.224	13.974	< 0.001	< 0.001	-0.759	0.014	0.819	< 0.001	0.791	< 0.001
lapatinib	18.622	19.243	18.510	< 0.001	< 0.001	-0.653	0.047	0.837	< 0.001	0.651	< 0.001
pazopanib	19.078	19.641	19.227	< 0.001	< 0.001	-0.511	-0.146	0.561	< 0.001	0.304	< 0.001

Figures

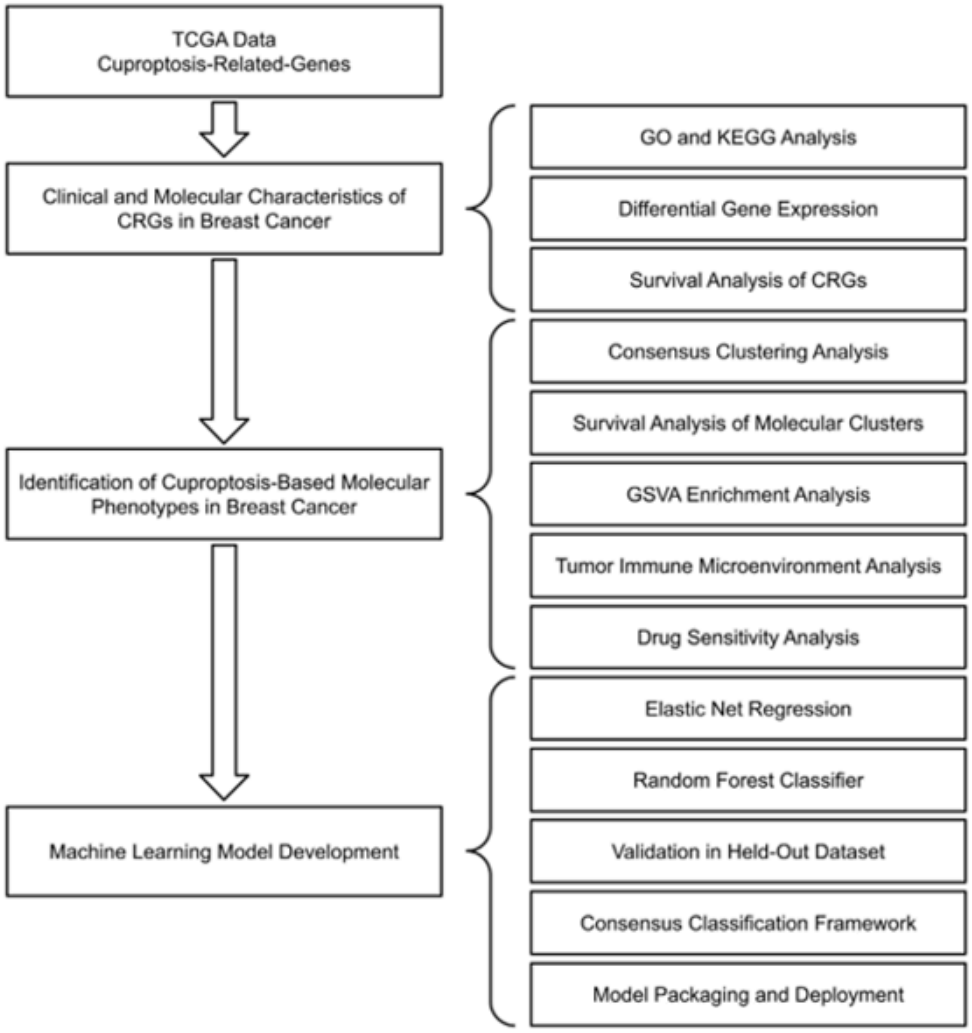


Figure 1

Overall study design and analytical workflow. . Schematic overview illustrating curation of the CRG19 panel, CRG19 profiling, consensus clustering, machine-learning ensemble pipeline, cross-cohort validation, and downstream clinical/TME/drug-sensitivity evaluations. TCGA:The Cancer Genome Atlas; CRG: cuproptosis-related genes; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSVA, Gene Set Variation Analysis.

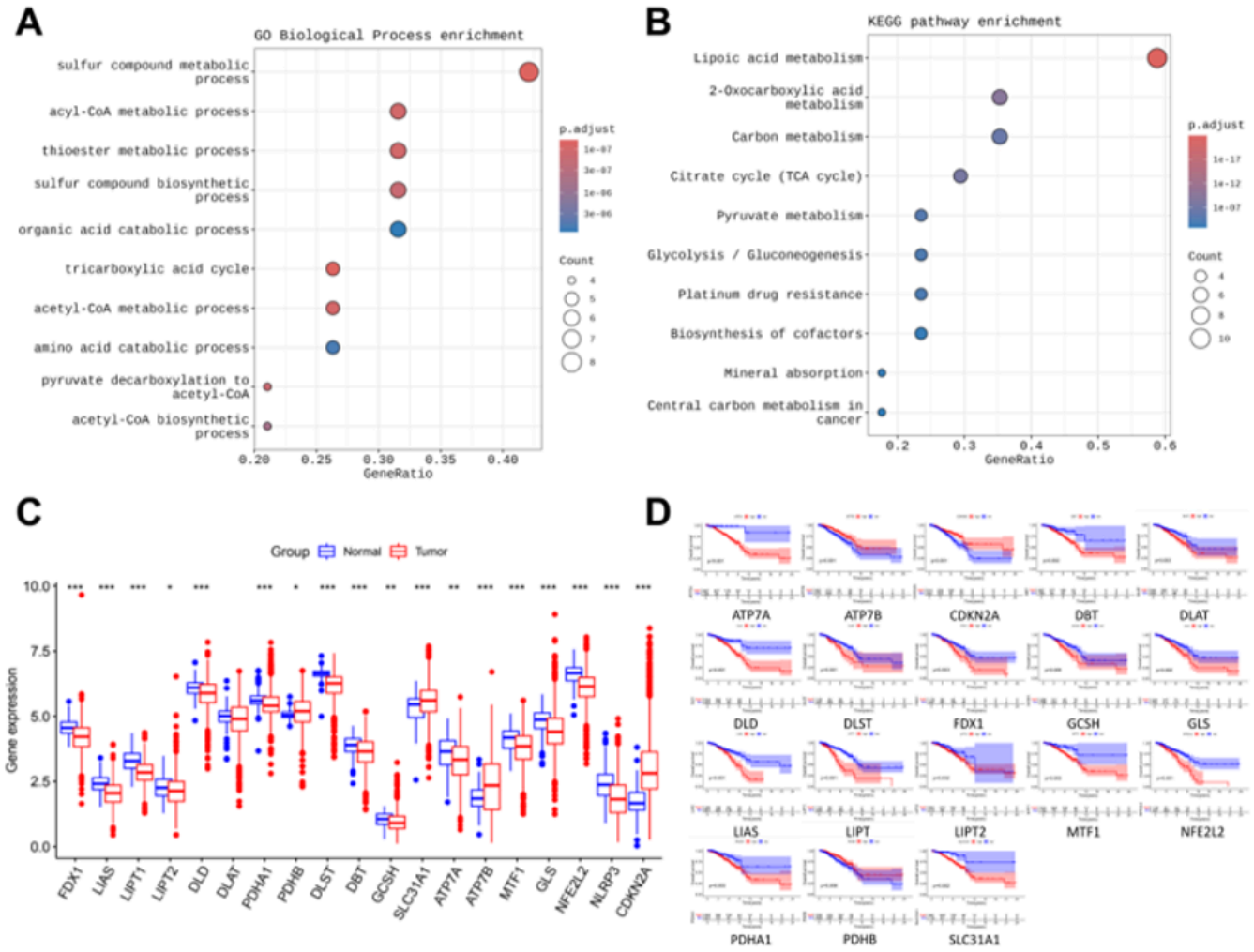


Figure 2

Molecular landscape and prognostic value of the CRG19 cuproptosis panel. (A) GO Biological Process enrichment analysis. (B) KEGG pathway enrichment analysis. (C) Differential expression of CRG19 genes between tumor and normal tissues in TCGA-BRCA (*p < 0.05, **p < 0.01, ***p < 0.001). (D) Kaplan-Meier survival curves for individual CRG19 genes with optimal dichotomization cutpoints.

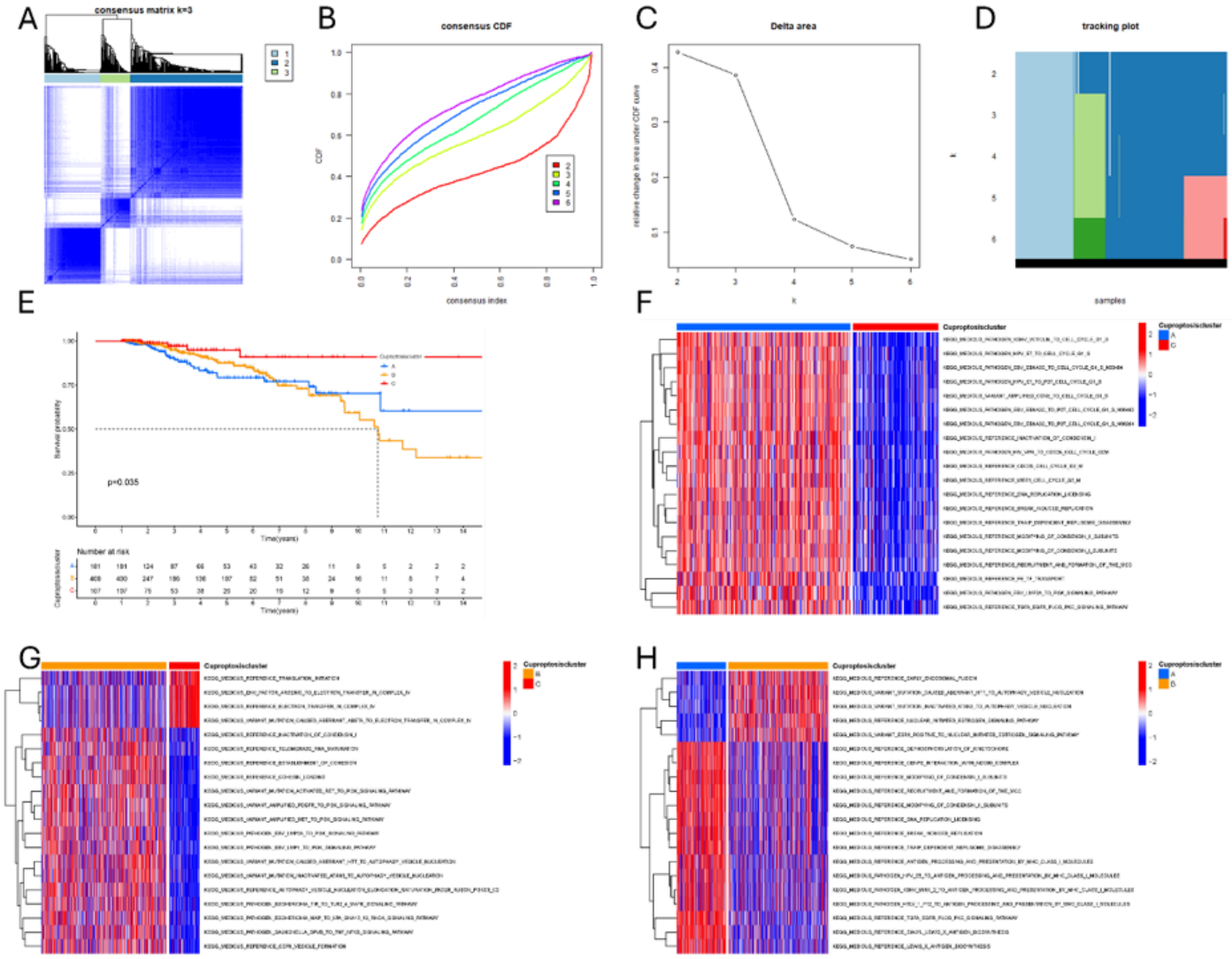


Figure 3

Consensus clustering of TCGA-BRCA into three cuproptosis molecular subtypes. (A–D) Unsupervised consensus clustering matrix and CDF curves for k = 3. (E) Kaplan–Meier overall survival analysis among Subtypes A, B, and C (log-rank p = 0.035). (F–H) GSVA pathway activity heatmaps illustrating top differentially enriched signaling programs across subtypes.

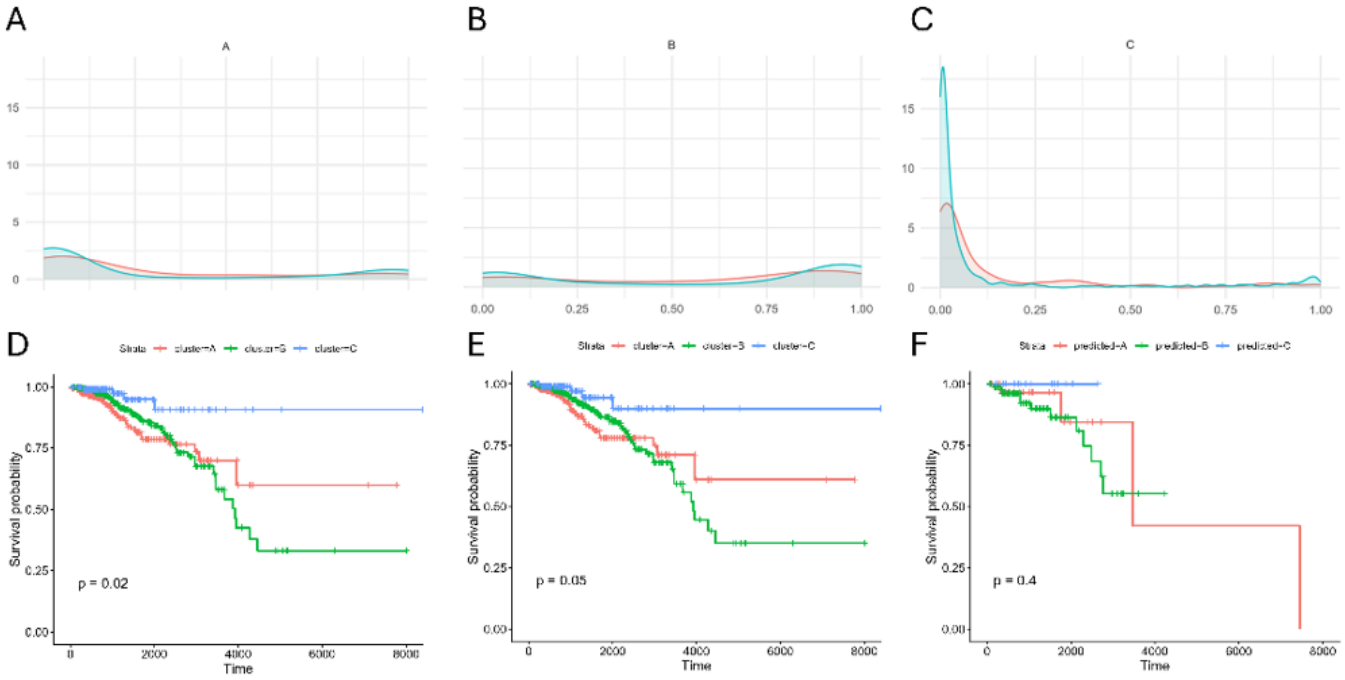


Figure 4

Development, probability calibration, and prognostic performance of the REO ensemble classifier. (A–C) Class-probability distribution and Kolmogorov-Smirnov calibration tests between training (n = 838) and testing (n = 148) cohorts. (D–F) Kaplan-Meier survival curves comparing consensus labels (training), predicted training labels, and predicted testing labels.

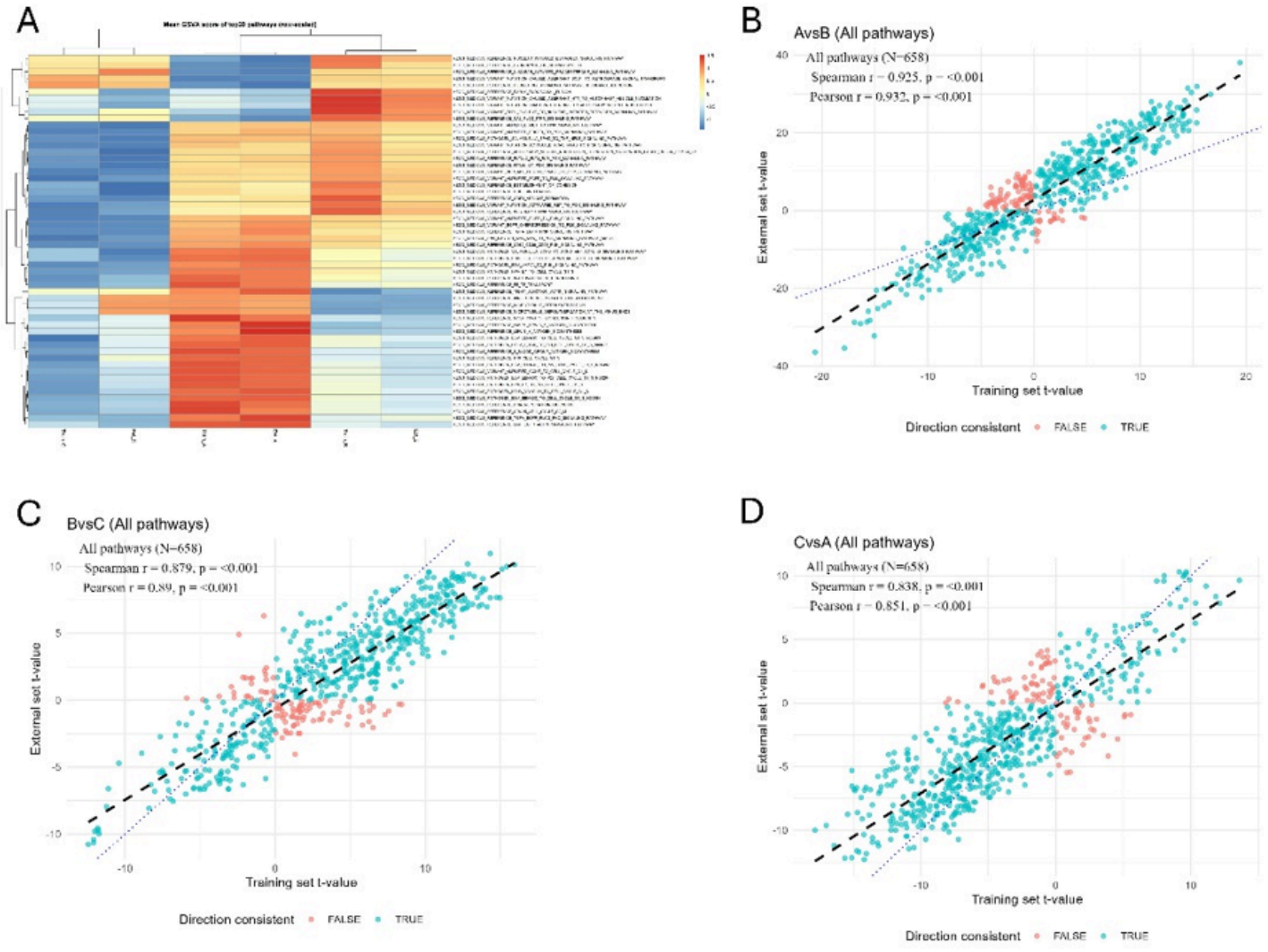


Figure 5

External validation of subtype-specific pathway signatures in GSE202203 (n = 3,207). (A) Heatmap of row-scaled mean GSVA scores for top 20 discriminative pathways across cohorts. (B–D) Scatter plots comparing GSVA t-statistics across 658 KEGG pathways for contrasts A vs. B, B vs. C, and C vs. A (Spearman r = 0.838–0.925, all p < 1e-174).

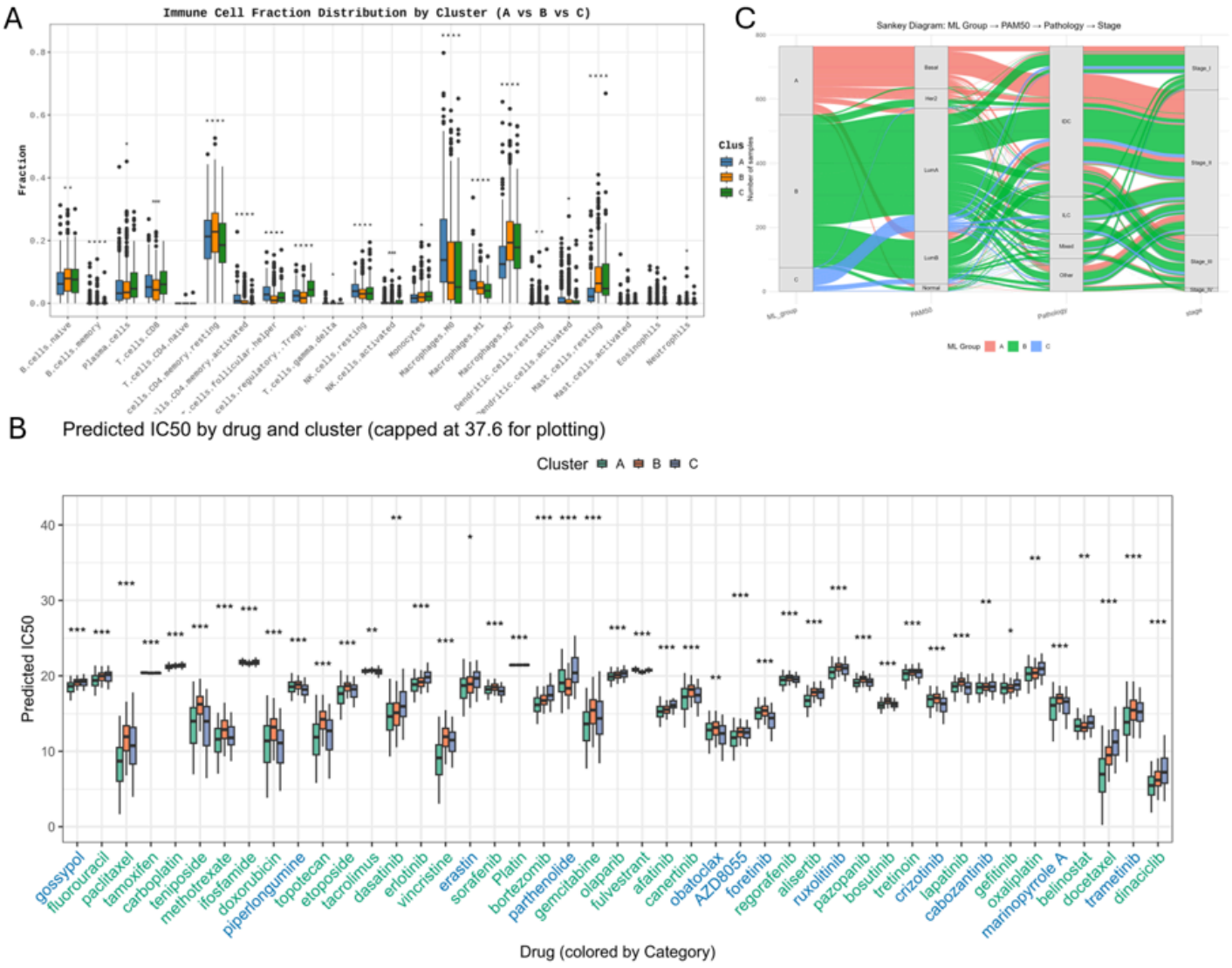


Figure 6

Tumor microenvironment, drug sensitivity, and clinical associations of cuproptosis subtypes. (A) CIBERSORTx immune cell infiltration fractions across Subtypes A, B, and C. (B) Predicted drug sensitivity (log2 IC50) distributions for key chemotherapeutic and endocrine agents. (C) Sankey diagram illustrating flow relationships between ML subtypes, PAM50 intrinsic subtypes, histological pathology, and clinical stages.