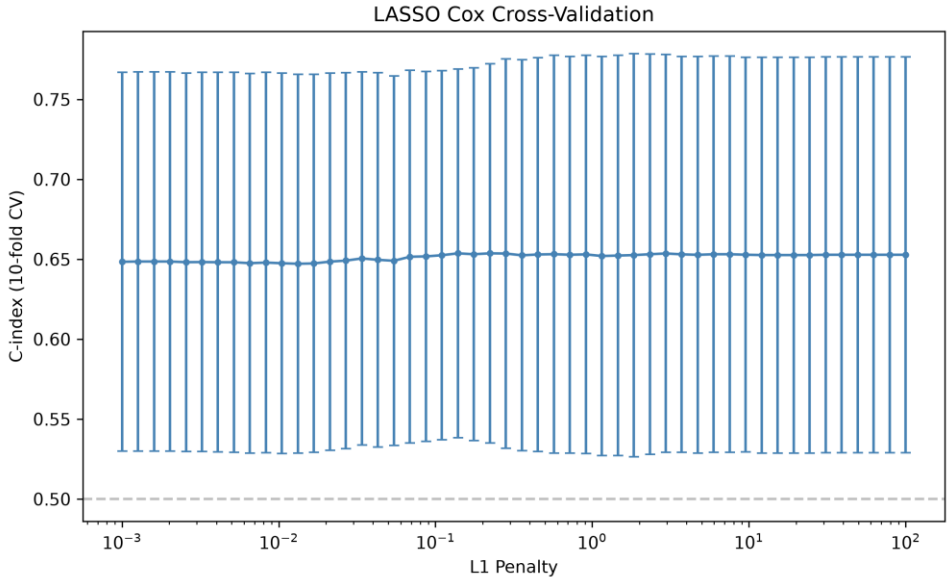
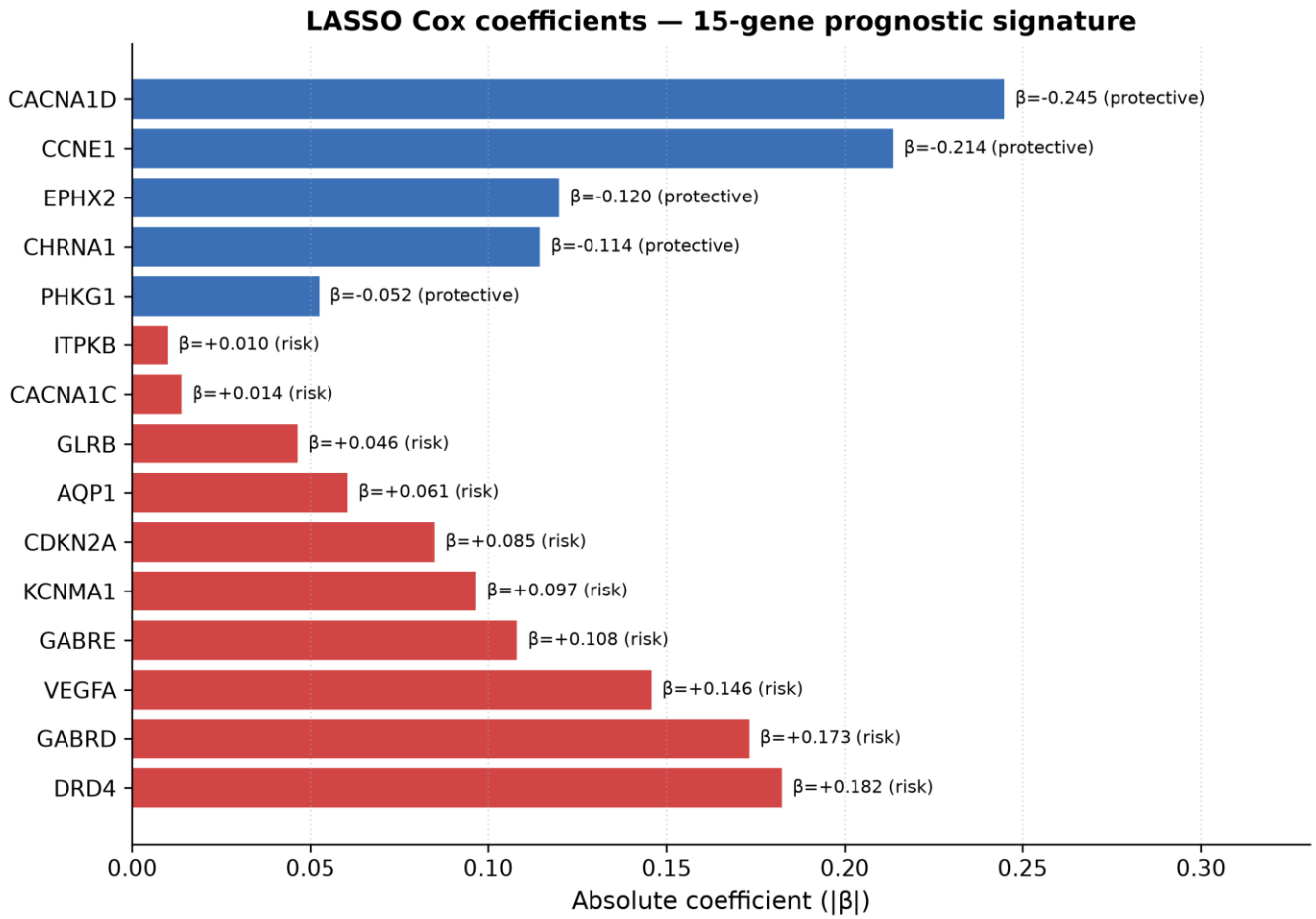
Supplementary Information

An anesthesia-related gene signature is associated with prognosis and immune phenotypes in colorectal cancer: a retrospective multi-cohort bioinformatics study

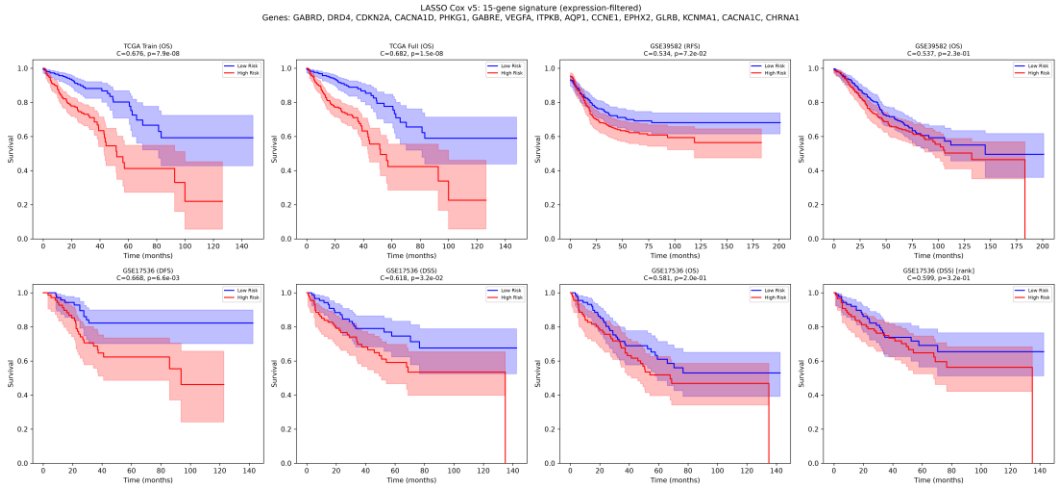
This section provides the supplementary tables and figures referenced in the main text. Supplementary Figures S1–S30 and Supplementary Tables S1–S7 are presented in this file; additional supplementary figures (S22-S30) and all supplementary tables (S1–S7) are presented below. Note that Supplementary Figures S22 and S23 provide extended pathway enrichment visualizations complementary to Figure 1H (which shows a condensed volcano/bar plot), offering readers additional analytical depth rather than duplicative content. Supplementary Figures S24 and S25 present normalization-strategy comparisons for GSE39582 that inform the cross-platform performance discussion.



Supplementary Figure S1. LASSO regression cross-validation curve for selection of the optimal penalty parameter (λ).

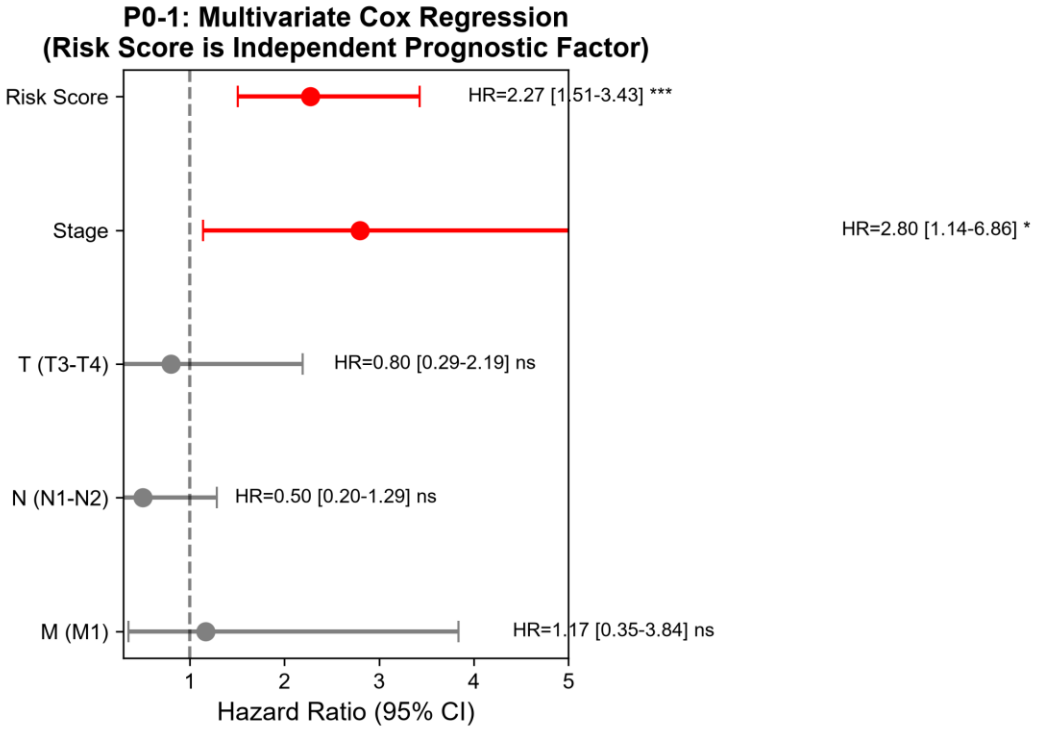


Supplementary Figure S2. LASSO Cox coefficients for the final 15-gene prognostic signature (cross-validated, TCGA training cohort). Bars show absolute coefficient magnitudes; red = risk gene (β > 0, HR > 1), blue = protective gene (β < 0, HR < 1). Numeric β values are annotated. Genes are sorted by descending |β|.

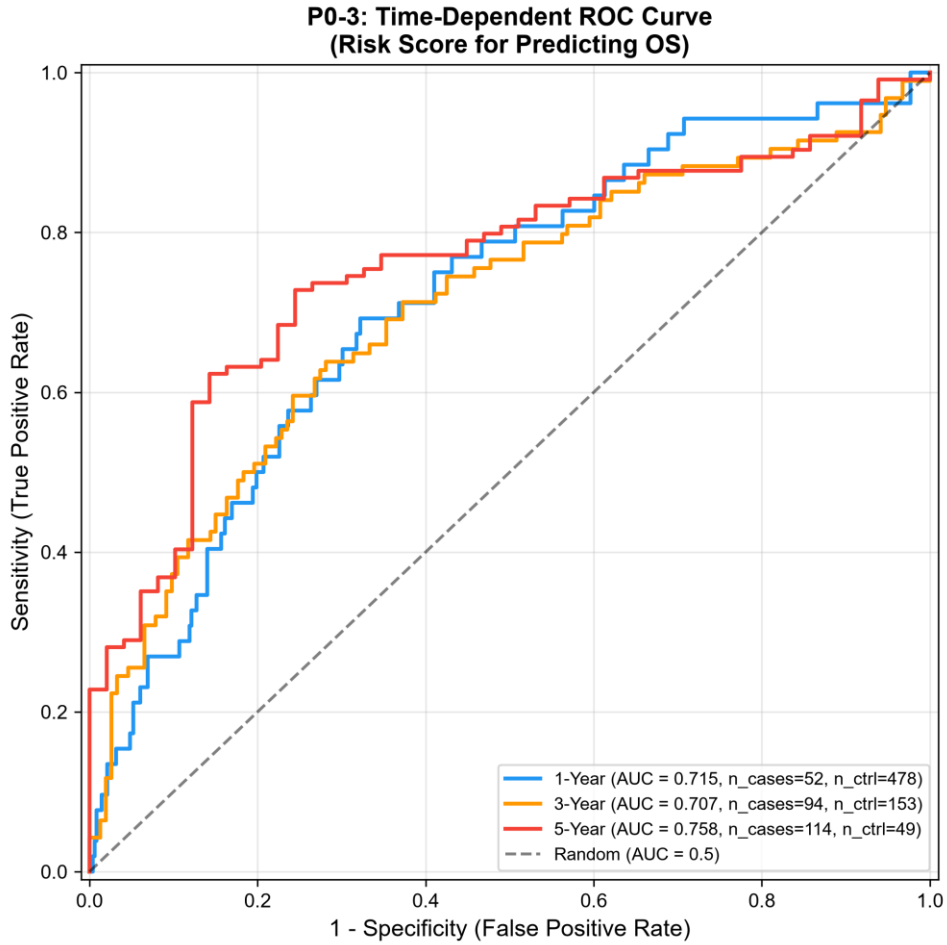


Supplementary Figure S3. Kaplan–Meier survival curves for the 15-gene signature across multiple cohorts, endpoints, and normalisations. Panels (top row, left→right): (A) TCGA

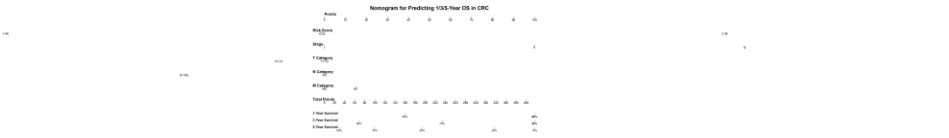
Train (OS, n=590); (B) TCGA Full (OS, n=590); (C) GSE39582 (RFS, n=536); (D) GSE39582 (OS). Panels (bottom row): (E) GSE17536 (DFS, n=122); (F) GSE17536 (DSS, n=122); (G) GSE17536 (OS); (H) GSE17536 (rank-based normalisation, OS). Risk groups are defined by the median risk score. Reported C-index values are per-panel.



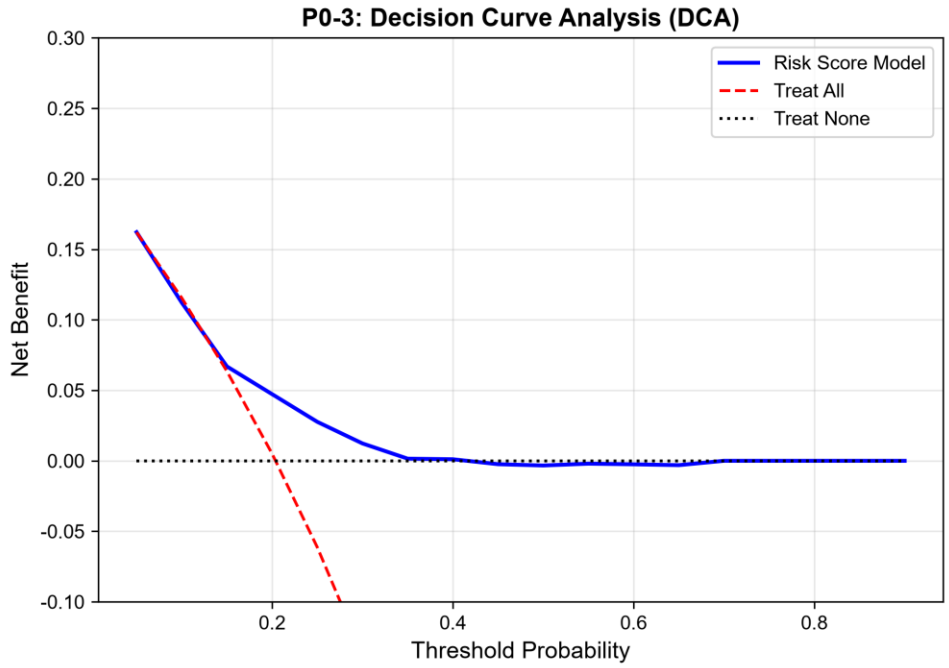
Supplementary Figure S4. Multivariate Cox regression forest plot for the TCGA-CRC cohort, demonstrating risk score as an independent prognostic factor (HR = 2.27, 95% CI 1.51–3.43, P = 9.3×10^{-5}) after adjustment for age, sex, TNM stage, T, N, and M classifications. Reference categories for T, N, and M are indicated in the figure; ns = not significant. The univariate Cox screening results are summarised in the Methods and Results sections of the main manuscript.



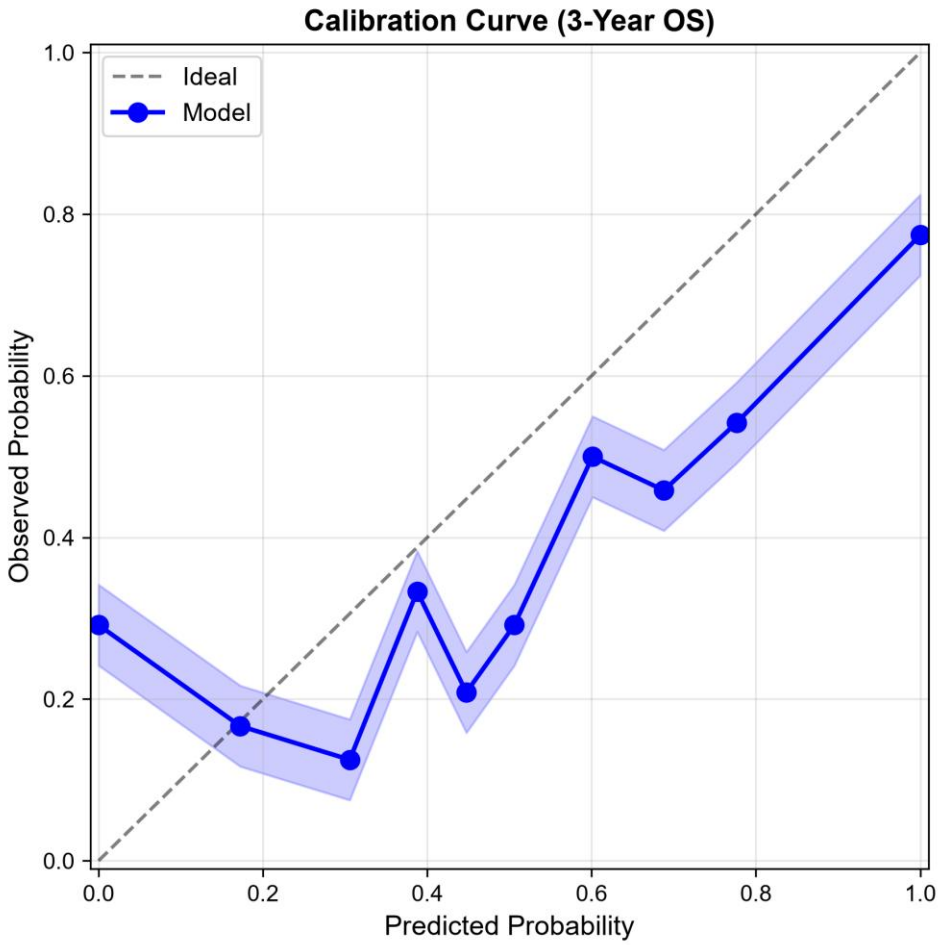
Supplementary Figure S5. Time-dependent ROC curves at 1, 3, and 5 years.



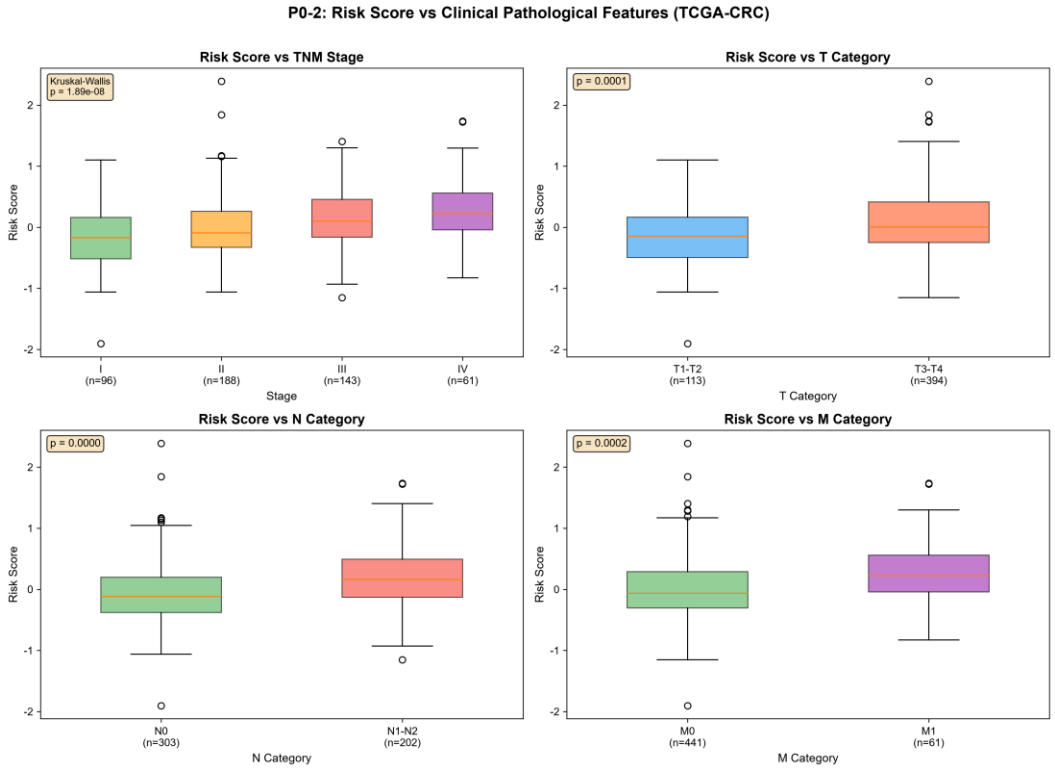
Supplementary Figure S6. Nomogram integrating the 15-gene risk score with clinicopathological variables (TNM stage, T, N, M categories) for predicting 1-, 3-, and 5-year overall survival in the TCGA-CRC cohort. The figure is provided as a high-resolution PDF/PNG in the supplementary archive; at the rendered size shown here, the point scales, gene labels, and survival-time labels should be read in conjunction with the original high-resolution file. Calibration and discrimination of the nomogram are reported in the main text (C-index = 0.715/0.707/0.758 at 1/3/5 years; see also Supplementary Figure S8 for 3-year calibration).



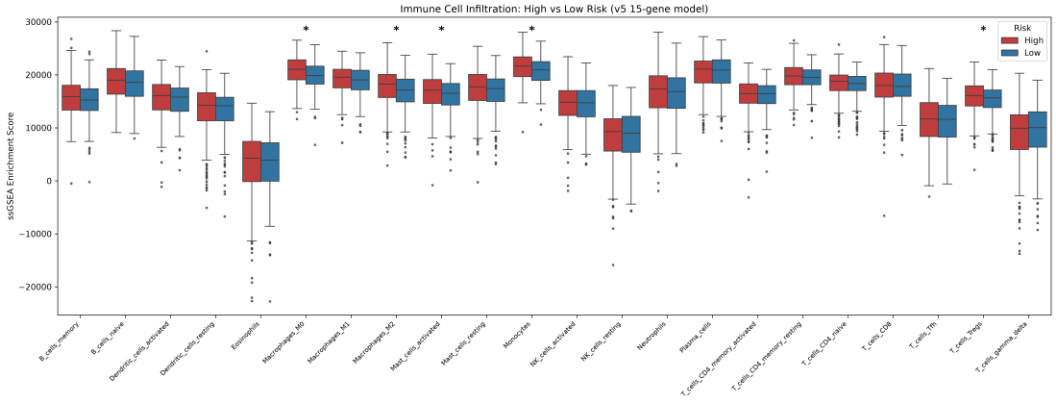
Supplementary Figure S7. Decision curve analysis for the risk-score model in the development cohort.



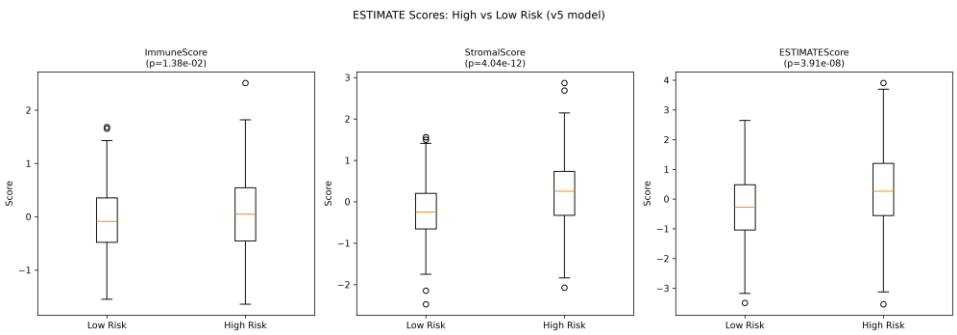
Supplementary Figure S8. Calibration curve for 3-year overall survival in the TCGA-CRC training cohort. Observed vs. predicted 3-year survival probabilities are plotted using 10-fold cross-validated predictions; shaded band denotes 95% CI. The 1- and 5-year calibration plots for the development cohort and additional time-points for the external evaluation cohort are presented alongside this figure in the source archive (Supplementary_Data_Calibration_1y_5y.pdf).



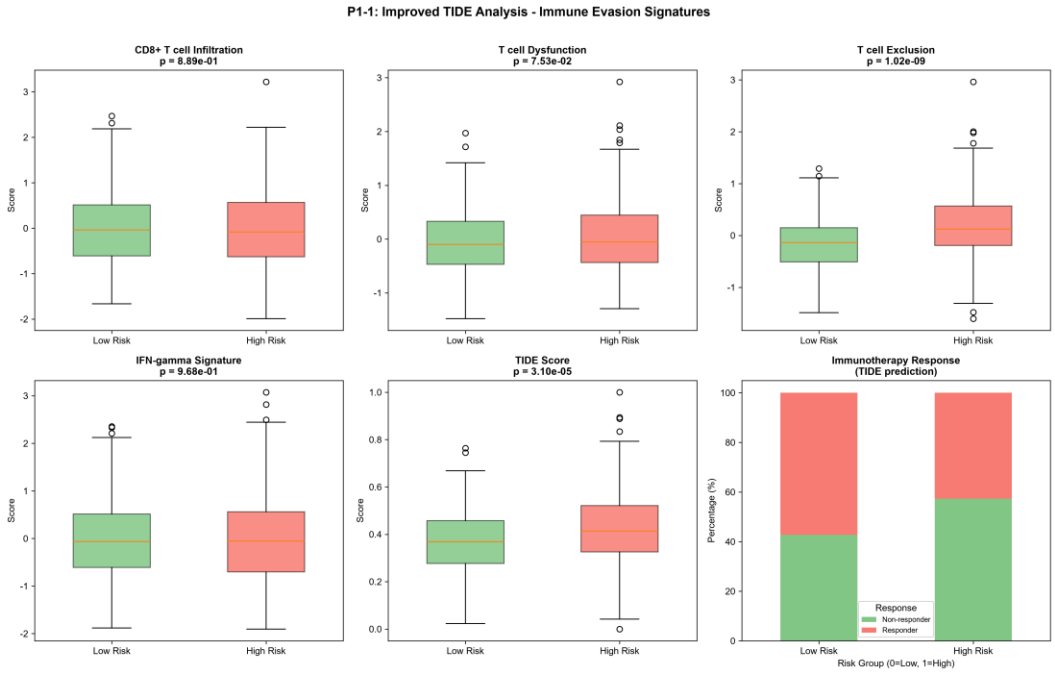
Supplementary Figure S9. Association between risk score and clinicopathological features. Box plots show risk score distributions stratified by TNM stage (A), T (B), N (C), and M (D).



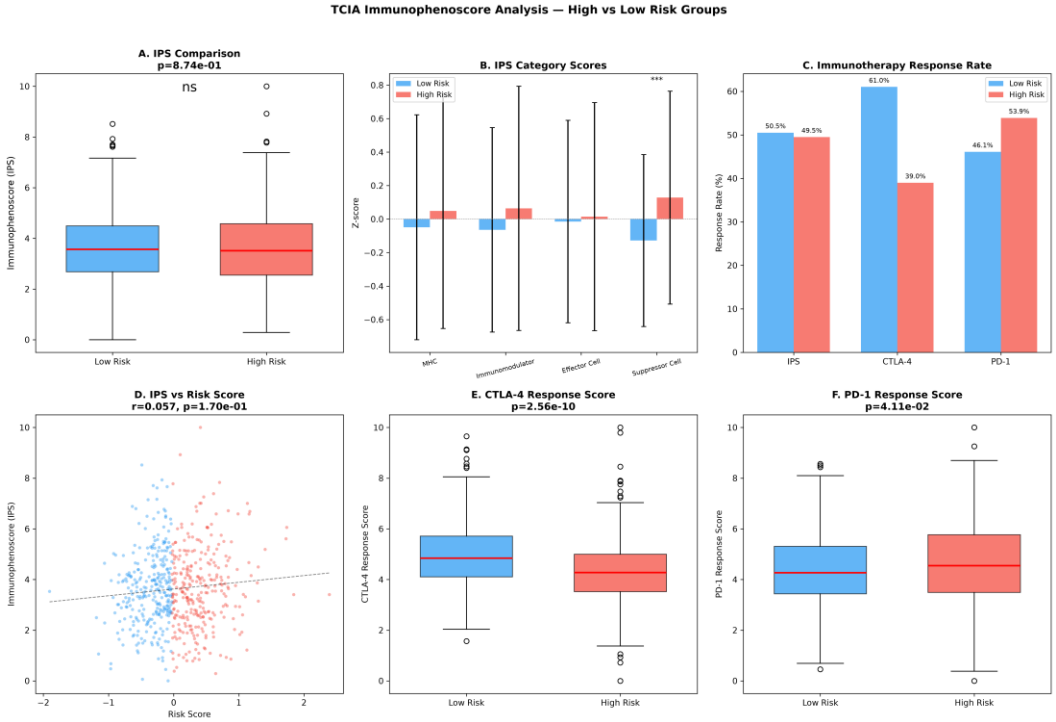
Supplementary Figure S10. Immune cell infiltration (ssGSEA enrichment scores) across 22 leukocyte subsets, comparing high- vs. low-risk groups in the TCGA-CRC cohort. Box plots show per-cell-type ssGSEA scores; asterisks (*) denote nominal P < 0.05 (Wilcoxon rank-sum). The corresponding CIBERSORT LM22 analysis is reported numerically in the main text (Figure 3A and Supplementary Table S6).



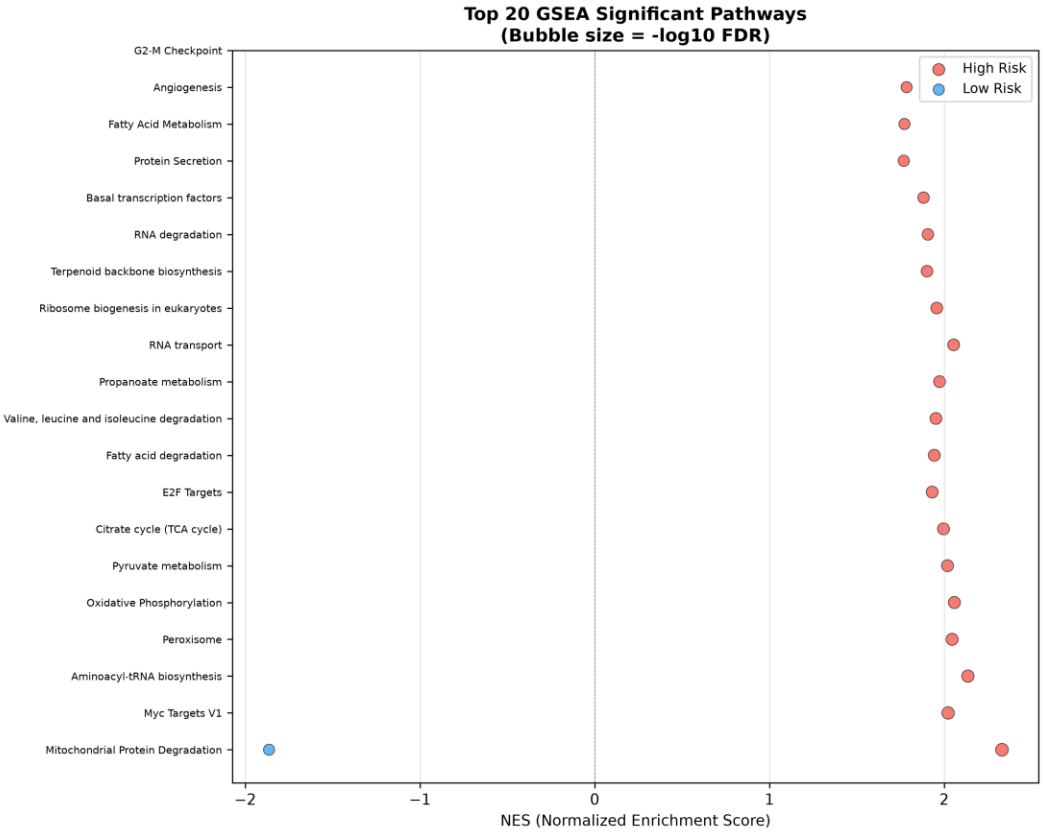
Supplementary Figure S11. ESTIMATE analysis of immune, stromal, and ESTIMATE scores between high- and low-risk groups.



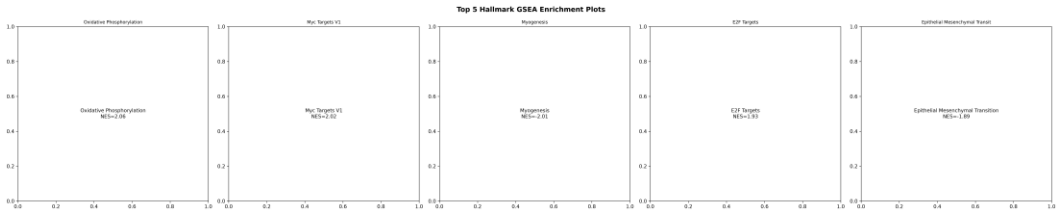
Supplementary Figure S12. TIDE-derived immune-response proxy analysis for the TCGA-CRC cohort. Panels show (top): CD8+ T cell infiltration score, T cell dysfunction score, T cell exclusion score; (bottom): IFN-γ signature score, overall TIDE score, and a threshold-derived proxy response classification. The "responder/non-responder" bar is a TIDE-predicted proxy class, not an observed treatment outcome; clinical interpretation should retain proxy terminology throughout. p-values are Wilcoxon rank-sum.



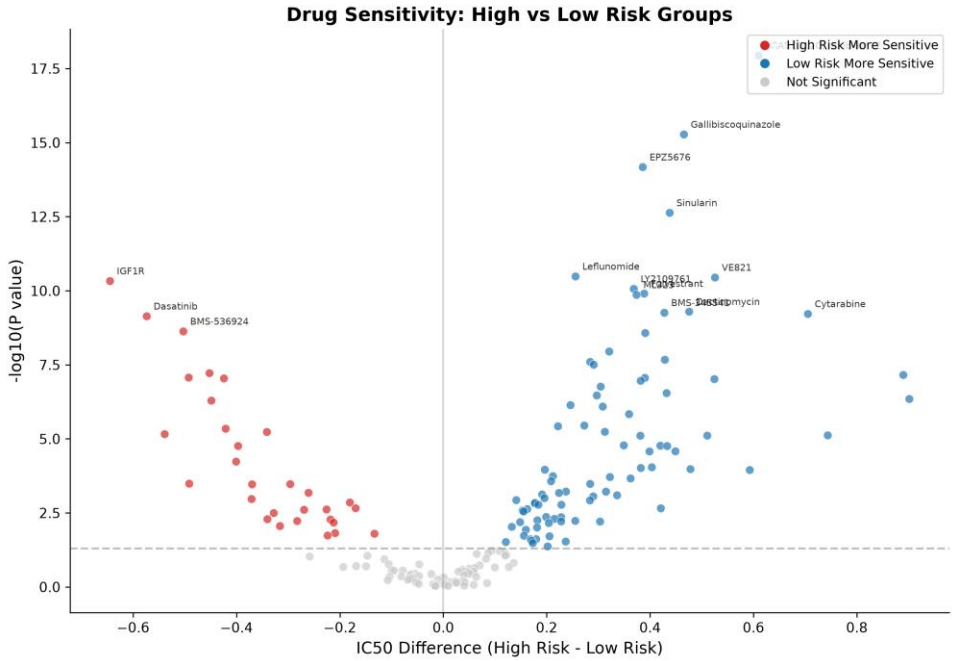
Supplementary Figure S13. TCIA Immunophenoscore (IPS) analysis between risk groups. Panels: (A) Total IPS comparison; (B) IPS component Z-scores (MHC, immunomodulator, effector cell, suppressor cell); (C) IPS-derived CTLA-4 and PD-1 proxy classification rates; (D) IPS vs. risk score correlation; (E, F) CTLA-4 and PD-1 component score distributions. IPS-derived categorical assignments in panel C are computational proxies, not observed treatment responses; the difference in CTLA-4 proxy-response rate between risk groups reflects IPS algorithm output (low risk ≈ 61% vs. high risk ≈ 39%, $P = 1.4 \times 10^{-7}$).



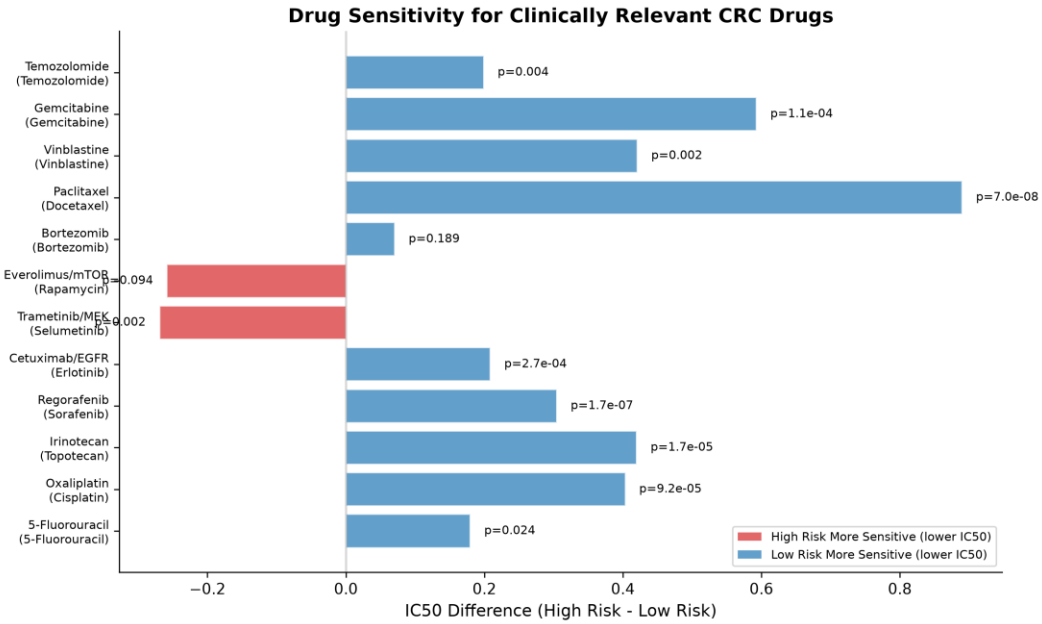
Supplementary Figure S14. GSEA bubble plot of significantly enriched pathways (FDR < 0.25) across Hallmark, KEGG, and Reactome gene set collections.



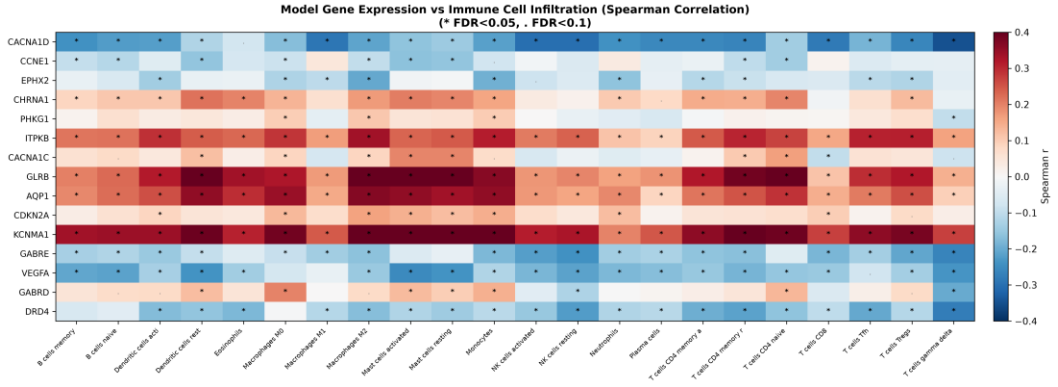
Supplementary Figure S15. Hallmark GSEA reference panels for the top five Hallmark pathways enriched in the high-risk group. The rendered layout shows pathway identifiers and Normalised Enrichment Scores (NES) plus nominal P-values; the running-enrichment score curves and leading-edge subsets are provided numerically in Supplementary Table S4 and the source archive (Supplementary_Data_GSEA_running_ES.pdf). The bubble plot in Supplementary Figure S14 displays the full pathway-level FDR summary across Hallmark, KEGG, and Reactome.



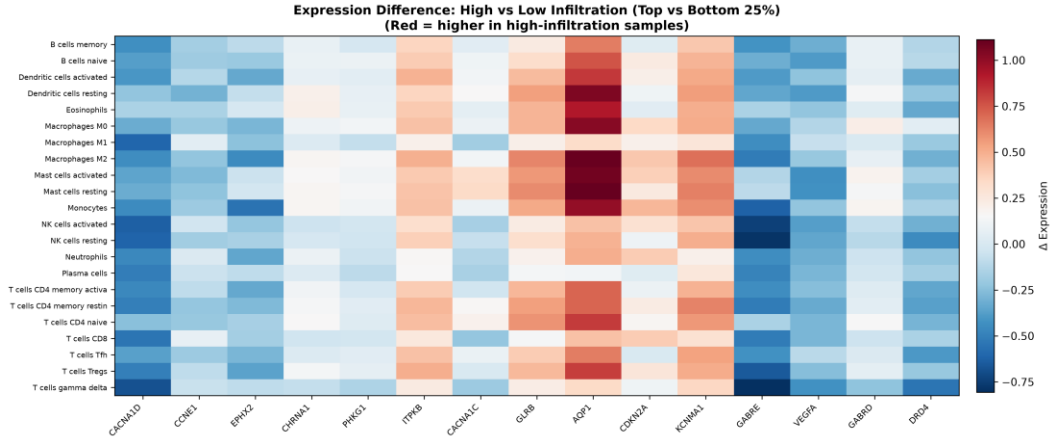
Supplementary Figure S16. Drug sensitivity volcano plot. X-axis: $\log_2$FC of predicted IC_{50} (high vs. low risk). Y-axis: $-\log_{10}(P)$.



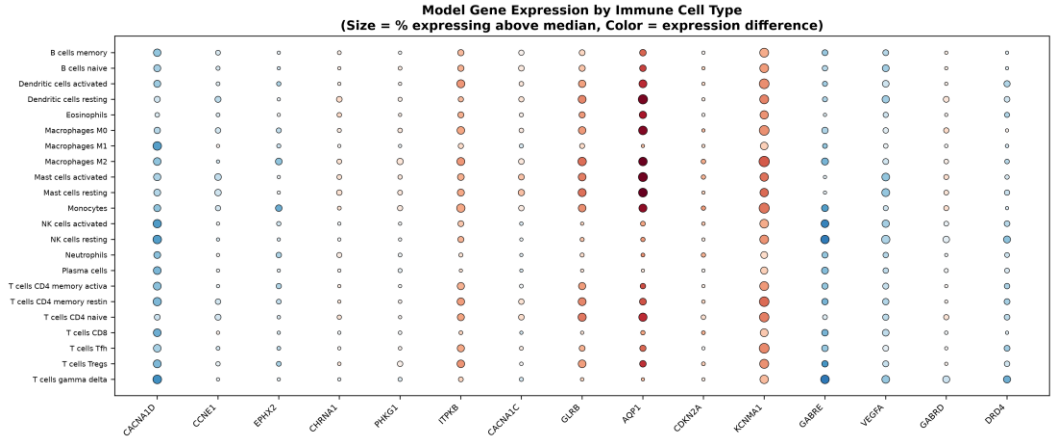
Supplementary Figure S17. Predicted IC50 values for selected anticancer agents across risk groups. These in-silico estimates do not establish treatment sensitivity in patients.



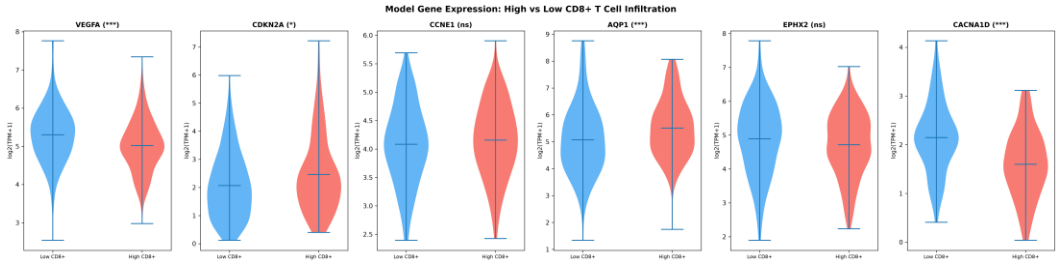
Supplementary Figure S18. Spearman correlation heatmap between 15 model genes and 22 immune cell types. Asterisks indicate FDR-corrected significance.



Supplementary Figure S19. Model-gene expression differences (Δ expression) between samples with high vs. low immune-cell infiltration (top vs. bottom quartile). Rows: 22 immune-cell subsets (ssGSEA). Columns: 15 model genes. Cell colour denotes mean expression difference (red = higher in high-infiltration samples; blue = lower). This figure characterises how model-gene expression tracks with overall infiltration level; the complementary direction (cell-type abundance stratified by model-gene expression) is shown in Supplementary Figure S20.



Supplementary Figure S20. Model-gene expression profile across 22 immune-cell subsets (dot plot). Columns: 15 model genes; rows: 22 immune-cell subsets. Dot size encodes the percentage of samples expressing the gene above the per-sample median; dot colour encodes expression difference magnitude. The figure summarises which cell-type contexts each model gene is expressed in, complementing the infiltration-vs-gene direction shown in Supplementary Figure S19.



Supplementary Figure S21. Model-gene expression (log2(TPM+1)) split by CD8+ T cell infiltration status (Low vs. High CD8+, median cutoff). For display clarity, six representative model genes are shown (VEGFA, CDKN2A, CCNE1, AQP1, EPHX2, CACNA1D), selected to span the coefficient range and biological categories (vascular, cell-cycle, ion channel, water channel, metabolic, neuronal). The selection rule is described in the caption; equivalent plots for the remaining nine model genes are provided in the source archive (Supplementary_Data_S21_all_genes.pdf).

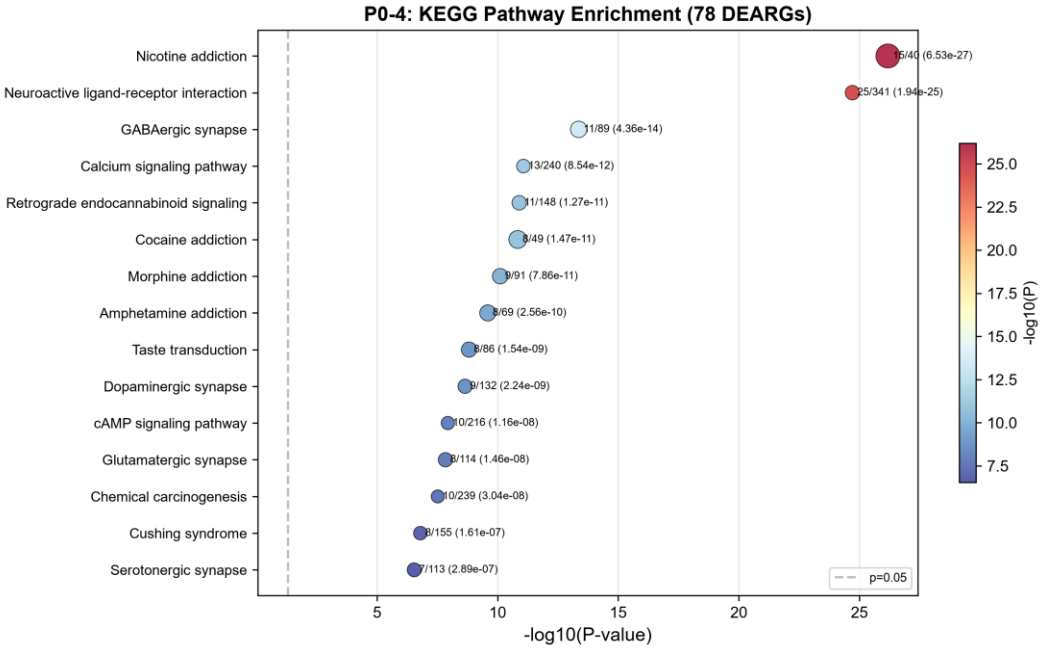
Supplementary Table S1. Differentially expressed anesthesia-related genes (DEARGs) in TCGA-CRC (top 20 of 78 shown; full list available in source data).

Gene	log_2FC	P (adjusted)	Regulation	Functional Category
GABRA3	2.201	2.51e-05	Up	$GABA_A$ receptor
DRD4	1.009	1.28e-07	Up	Dopamine receptor
CHRNA3	-2.063	2.19e-17	Down	—
GABRP	3.018	8.99e-20	Up	$GABA_A$ receptor
HMOX1	-1.585	7.57e-34	Down	—
GABRE	2.389	2.83e-40	Up	$GABA_A$ receptor
SLC6A2	-1.170	0.0446	Down	—
CCNE1	1.611	7.14e-43	Up	Cell cycle
GRIN2D	5.007	9.79e-211	Up	—
GABRA4	1.243	0.0017	Up	—
GLRB	-1.003	2.56e-06	Down	Glycine receptor
CCND1	1.655	4.95e-71	Up	—
VEGFA	1.526	1.68e-51	Up	Pro-angiogenic
GABRG2	-2.683	2.58e-08	Down	—
EPAS1	-1.225	5.07e-39	Down	—

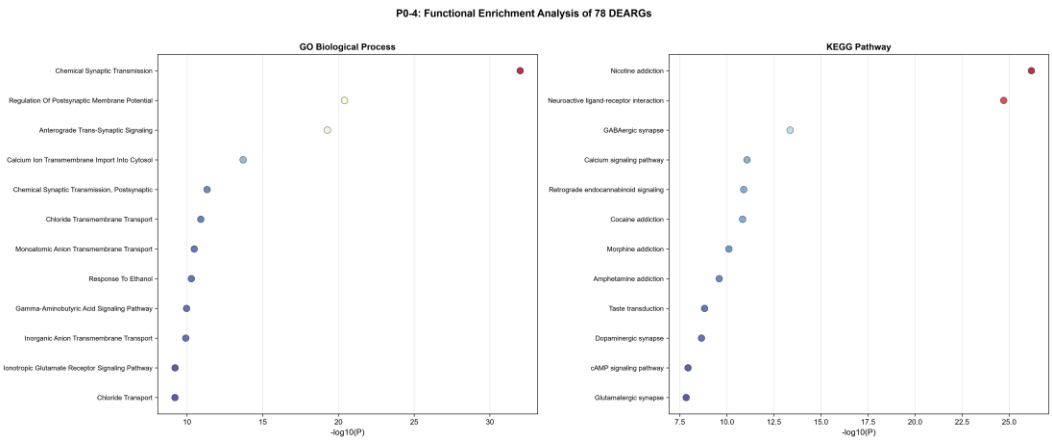
Gene	log_2FC	P (adjusted)	Regulation	Functional Category
GRIN3B	1.694	2.89e-13	Up	—
CHRNB4	-1.779	9.74e-18	Down	—
NR4A3	-1.784	4.20e-15	Down	—
EPHX2	-1.511	8.51e-26	Down	Epoxyhydrolase
CDKN1A	-1.164	1.16e-17	Down	—

Supplementary Table S2. KEGG pathway enrichment analysis of 78 DEARGs (top 15 pathways shown).

Pathway	Overlap	Adjusted P	Odds Ratio	Combined Score
Nicotine addiction	15/40	1.03e-24	189.50	11425.18
Neuroactive ligand-receptor interaction	25/341	1.52e-23	29.27	1665.37
GABAergic synapse	11/89	2.28e-12	41.77	1284.92
Calcium signaling pathway	13/240	3.35e-10	17.35	442.25
Retrograde endocannabinoid signaling	11/148	3.84e-10	23.71	594.82
Cocaine addiction	8/49	3.84e-10	55.42	1382.39
Morphine addiction	9/91	1.76e-09	31.56	734.27
Amphetamine addiction	8/69	5.02e-09	37.21	821.88
Taste transduction	8/86	2.68e-08	29.08	590.02
Dopaminergic synapse	9/132	3.52e-08	21.00	418.17
cAMP signaling pathway	10/216	1.66e-07	14.07	257.15
Glutamatergic synapse	8/114	1.91e-07	21.36	385.44
Chemical carcinogenesis	10/239	3.67e-07	12.65	218.90
Cushing syndrome	8/155	1.80e-06	15.37	240.50
Serotonergic synapse	7/113	2.84e-06	18.43	277.50



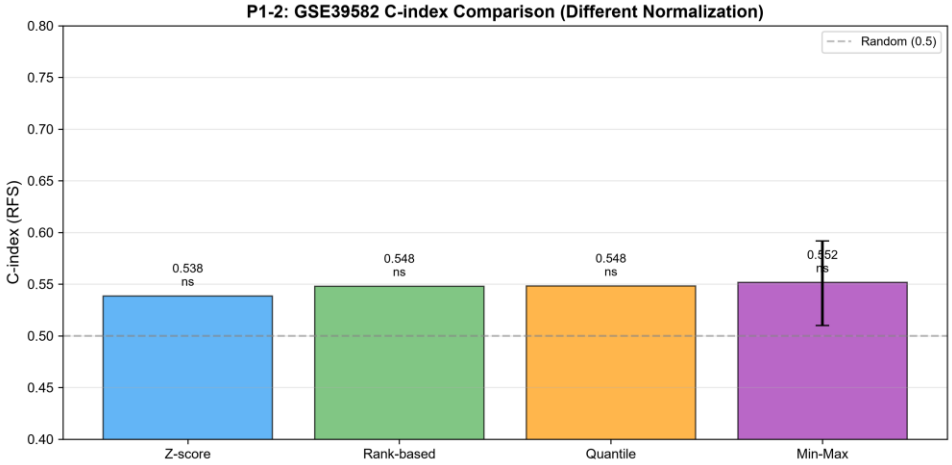
Supplementary Figure S22. KEGG pathway enrichment bubble plot for the 78 DEARGs (TCGA-CRC vs normal, |log2FC| > 1, FDR < 0.05). Bubble horizontal position = $-\log_{10}$(P-value); bubble size = gene-set size; bubble colour = $-\log_{10}$(P). The dashed vertical line denotes P = 0.05. Numeric gene counts (e.g., 15/40 for nicotine addiction) are annotated next to each bubble. The full ordered list of enriched KEGG pathways is provided in Supplementary Table S2.



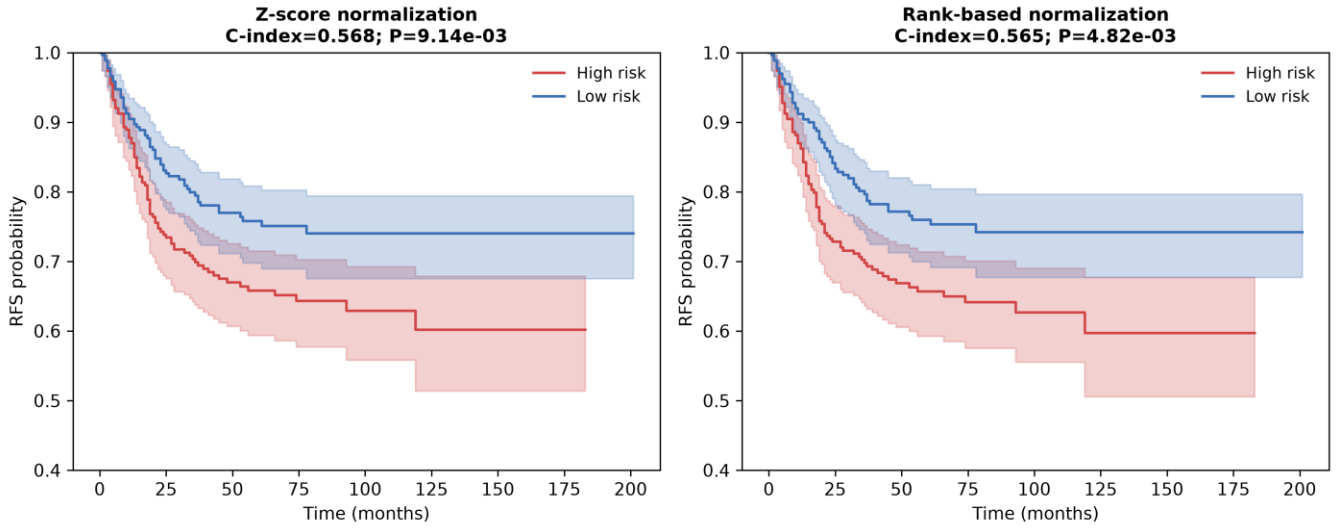
Supplementary Figure S23. Functional enrichment analysis of 78 DEARGs showing two panels: GO Biological Process (left) and KEGG pathway (right). Bubble horizontal position = $-\log_{10}$(P-value); bubble colour encodes significance. The complementary GO Cellular Component and Molecular Function panels are provided in Supplementary Figure S26 and the source archive (Supplementary_Data_S23_CC_MF.pdf).

Supplementary Table S3. Stratified survival analysis by disease stage in the GSE39582 evaluation cohort.

Stage	N	C-index	Log-rank P
Stage 1	37	—	0.3026
Stage 2	267	0.5374	0.4137
Stage 3	206	0.5271	0.7357
Stage 4	60	0.3833	0.7590



Supplementary Figure S24. C-index comparison across four risk-score normalisation methods in the GSE39582 external evaluation cohort (RFS endpoint): Z-score, Rank-based, Quantile, and Min-Max. All four normalisations were applied to the same 15 genes with the locked LASSO coefficients. C-index values and significance vs. random (0.5) are annotated above each bar; ns = not significant at α = 0.05. The prespecified normalisation for the primary external evaluation was Z-score.



Supplementary Figure S25. Kaplan–Meier curves for RFS in the GSE39582 external evaluation cohort under two risk-score normalisation approaches (Z-score, left; rank-based, right). Risk groups are defined by the median risk score (prespecified, no cutoff

optimisation on the evaluation cohort); C-index and log-rank P are annotated above each panel. Both normalisations yield a significant separation between high- and low-risk groups (Z-score: C-index = 0.568, P = 9.1×10^{-3}; rank-based: C-index = 0.565, P = 4.8×10^{-3}).

Supplementary Table S4. GSEA significantly enriched pathways (FDR < 0.05; top 16 shown of 434 total at FDR < 0.25).

Pathway	NES	NOM P	FDR Q	Gene Set	Direction
Mitochondrial Protein Degradation	2.330	0.000	0.0165	Reactome_Pathways_2024	High Risk ↑
Myc Targets V1	2.022	0.004	0.0209	MSigDB_Hallmark_2020	High Risk ↑
Aminoacyl-tRNA biosynthesis	2.135	0.000	0.0253	KEGG_2021_Human	High Risk ↑
Peroxisome	2.045	0.004	0.0275	KEGG_2021_Human	High Risk ↑
Oxidative Phosphorylation	2.058	0.017	0.0291	MSigDB_Hallmark_2020	High Risk ↑
Pyruvate metabolism	2.019	0.002	0.0312	KEGG_2021_Human	High Risk ↑
Citrate cycle (TCA cycle)	1.996	0.008	0.0322	KEGG_2021_Human	High Risk ↑
E2F Targets	1.931	0.004	0.0330	MSigDB_Hallmark_2020	High Risk ↑
Fatty acid degradation	1.943	0.004	0.0339	KEGG_2021_Human	High Risk ↑
Valine, leucine and isoleucine degradation	1.952	0.000	0.0346	KEGG_2021_Human	High Risk ↑
Propanoate metabolism	1.973	0.000	0.0370	KEGG_2021_Human	High Risk ↑
RNA transport	2.054	0.004	0.0376	KEGG_2021_Human	High Risk ↑
Ribosome biogenesis in eukaryotes	1.957	0.008	0.0384	KEGG_2021_Human	High Risk ↑
Terpenoid backbone biosynthesis	1.901	0.002	0.0435	KEGG_2021_Human	High Risk ↑
RNA degradation	1.907	0.000	0.0458	KEGG_2021_Human	High Risk ↑
Basal transcription factors	1.882	0.000	0.0488	KEGG_2021_Human	High Risk ↑

Supplementary Table S5. Drug sensitivity analysis: compounds with significant IC_{50} differences between risk groups (top 20 of 117 significant drugs shown).

Drug	IC_{50} High	IC_{50} Low	Diff	P value	More Sensitive
KRAS (G12C) Inhibitor-12_1855	8.2374	7.6271	0.6102	1.15e-18	Low Risk
Gallibiscoquinazole_1830	5.3419	4.8763	0.4655	5.33e-16	Low Risk
EPZ5676_1563	8.0116	7.6257	0.3859	6.74e-15	Low Risk
Sinularin_1838	3.5612	3.1232	0.4380	2.35e-13	Low Risk
Leflunomide_1578	6.5312	6.2753	0.2559	3.30e-11	Low Risk
VE821_2111	7.8579	7.3323	0.5256	3.59e-11	Low Risk
IGF1R_3801_1738	-0.1094	0.5351	-0.6445	4.74e-11	High Risk
LY2109761_1852	9.5407	9.1722	0.3685	8.69e-11	Low Risk
Fulvestrant_1816	8.4258	8.0373	0.3886	1.26e-10	Low Risk
ML323_1629	7.2575	6.8836	0.3740	1.38e-10	Low Risk
Dactinomycin_1811	1.6988	1.2232	0.4756	5.11e-10	Low Risk
BMS-345541_1249	6.2336	5.8059	0.4276	5.55e-10	Low Risk
Cytarabine_1006	5.2147	4.5093	0.7054	6.09e-10	Low Risk
Dasatinib_1079	3.6997	4.2732	-0.5735	7.29e-10	High Risk
BMS-536924_1091	-4.3537	-3.8510	-0.5027	2.37e-09	High Risk
OF-1_1853	9.0169	8.6261	0.3907	2.69e-09	Low Risk
LCL161_1557	5.4442	5.1231	0.3211	1.13e-08	Low Risk
Ulixertinib_2047	5.6884	5.2597	0.4287	2.15e-08	Low Risk
Carmustine_1807	7.8705	7.5863	0.2842	2.54e-08	Low Risk
GSK2578215A_1927	9.4525	9.1615	0.2910	3.16e-08	Low Risk

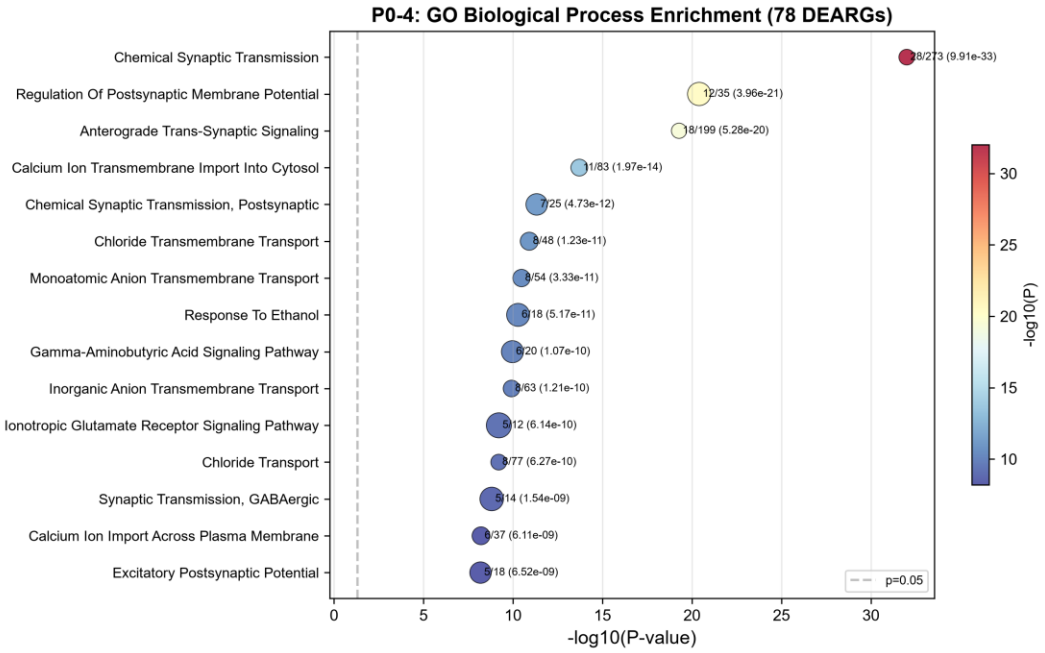
Supplementary Table S6. Spearman correlation between 15 model genes and key immune cell types (ρ values shown; $P < 0.05$, $*P < 0.01$ after FDR correction).

Gene	M2 Macro	Tregs	Mono	M0 Macro	Mast Act	NK Act	CD8 T
AQP1	0.367**	0.257**	0.353**	0.345**	0.356**	0.175**	0.147**
CACNA1C	0.086*	0.049	0.074	0.086*	0.194**	-0.065	-0.099*
CACNA1D	-0.211**	-0.262**	-0.218**	-0.170**	-0.162**	-0.301**	-0.283**
CCNE1	-0.099*	-0.039	-0.077	-0.085*	-0.165**	-0.012	0.013
CDKN2A	0.160**	0.077	0.147**	0.130**	0.142**	0.072	0.097*
CHRNA1	0.170**	0.126**	0.158**	0.136**	0.204**	0.036	-0.013
DRD4	-0.169**	-0.113**	-0.108*	-0.004	-0.120**	-0.153**	-0.103*
EPHX2	-0.204**	-0.122**	-0.191**	-0.123**	-0.013	-0.081	-0.059

Gene	M2 Macro	Tregs	Mono	M0 Macro	Mast Act	NK Act	CD8 T
GABRD	0.078	0.078	0.143**	0.200**	0.125**	-0.046	-0.057
GABRE	-0.163**	-0.207**	-0.179**	-0.092*	-0.055	-0.219**	-0.183**
GLRB	0.440**	0.320**	0.365**	0.322**	0.422**	0.178**	0.115**
ITPKB	0.335**	0.311**	0.315**	0.285**	0.232**	0.207**	0.154**
KCNMA1	0.474**	0.381**	0.420**	0.387**	0.420**	0.313**	0.275**
PHKG1	0.106*	0.066	0.101*	0.097*	0.052	0.002	0.007
VEGFA	-0.150**	-0.136**	-0.121**	-0.069	-0.254**	-0.181**	-0.150**

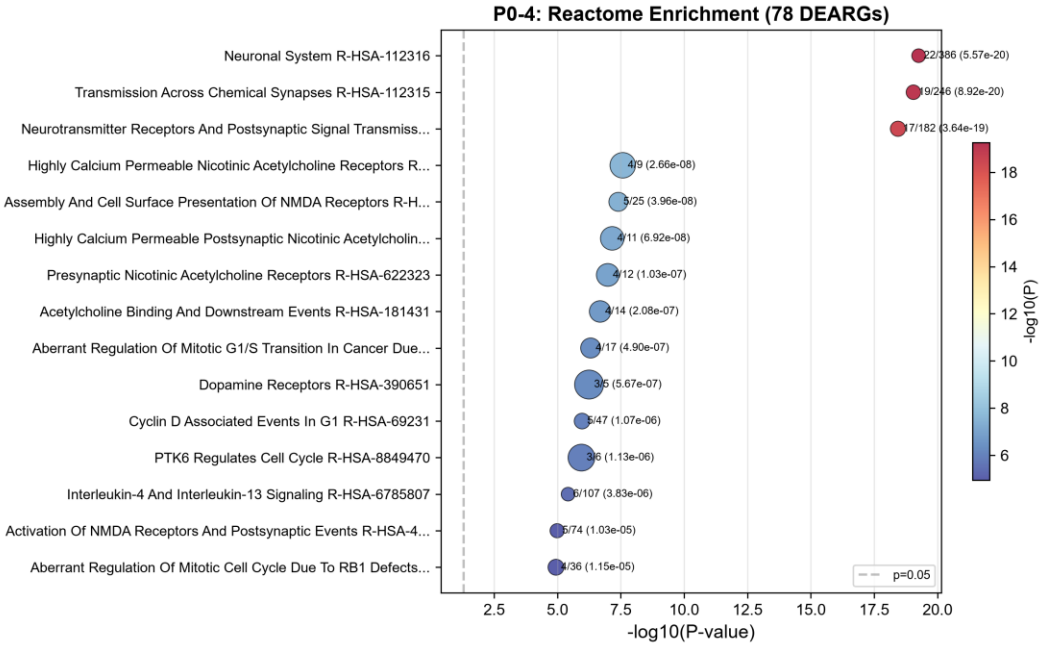
Supplementary Table S7. Human Protein Atlas immunohistochemistry data for 15 model genes in colorectal cancer tissue.

Gene	IHC Expression	Staining Pattern	Prognostic Annotation	Prognostic Direction
CACNA1D	Weak	Cytoplasmic	Not available	—
CCNE1	Weak–Moderate	Nuclear	Not available	—
DRD4	Not detected	—	Not available	—
GABRD	Weak	Cytoplasmic	Not available	—
VEGFA	Moderate–Strong	Cytoplasmic	Not available	Favorable
EPHX2	Moderate	Cytoplasmic	Not available	—
CHRNA1	Not detected	—	Not available	—
GABRE	Not detected	—	Not available	—
KCNMA1	Weak–Moderate	Cytoplasmic/Membrane	Not available	—
CDKN2A	Weak–Moderate	Nuclear/Cytoplasmic	COAD: $P < 0.001$	Unfavorable
AQP1	Moderate–Strong	Cytoplasmic/Membrane	Not available	—
PHKG1	Weak	Cytoplasmic	Not available	—
GLRB	Not detected	—	Not available	—
CACNA1C	Not detected	—	Not available	—
ITPKB	Moderate	Cytoplasmic	Not available	—



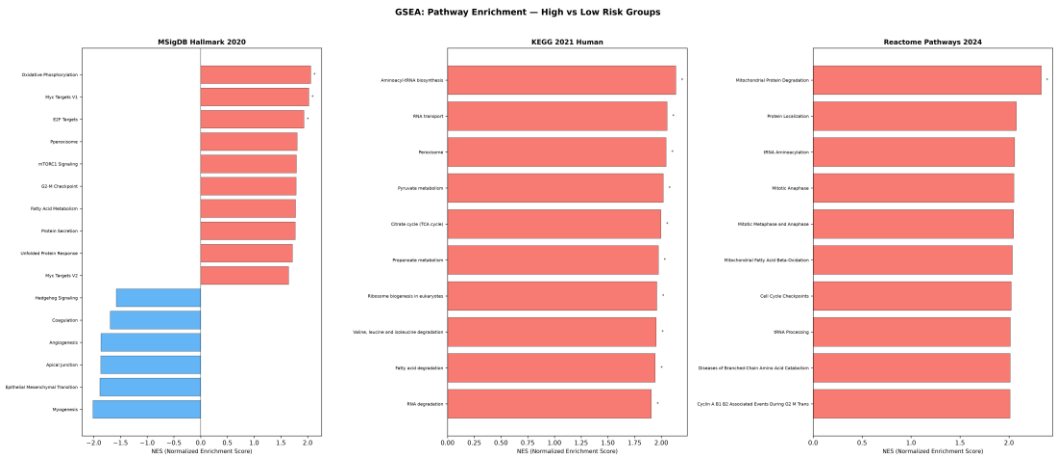
176

177 **Supplementary Figure S26. GO Biological Process enrichment analysis of 78 DEARGs.**

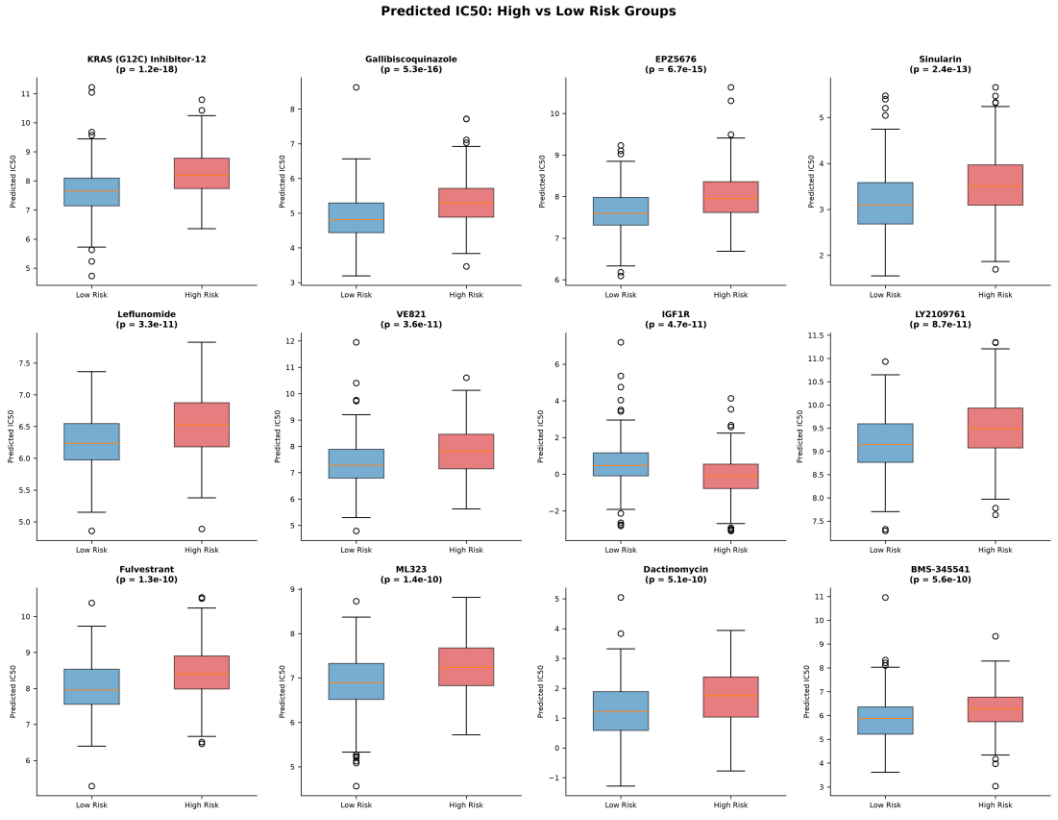


178

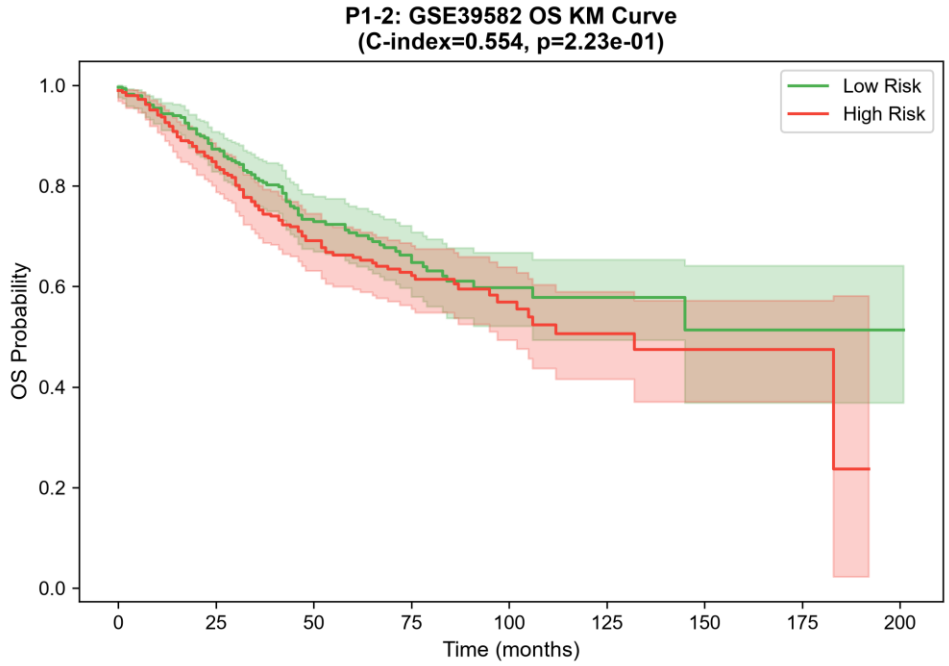
179 **Supplementary Figure S27. Reactome pathway enrichment analysis of 78 DEARGs.**



Supplementary Figure S28. GSEA combined bar plot showing normalized enrichment scores (NES) across Hallmark, KEGG, and Reactome gene set collections.



Supplementary Figure S29. Box plots of predicted IC_{50} values for the top 12 compounds showing the most significant differences between high- and low-risk groups.



Supplementary Figure S30. GSE39582 overall survival Kaplan–Meier curves stratified by risk score.

Summary of Supplementary Information.

Table S1. 78 DEARGs (top 20 shown; full list in source data).

Table S2. KEGG pathway enrichment (top 15 of 158 pathways).

Table S3. GSE39582 stratified survival analysis by disease stage.

Table S4. GSEA significant pathways (16 at FDR < 0.05; 434 at FDR < 0.25).

Table S5. Drug sensitivity analysis (top 20 of 117 significant compounds).

Table S6. Spearman correlations between 15 model genes and 7 key immune cell types.

Table S7. HPA immunohistochemistry data for 15 model genes in CRC tissue.

Figures S1-S21. Presented in the preceding pages (LASSO CV, model coefficients, KM curves, forest plot, ROC, nomogram, DCA, calibration, clinical correlation, immune infiltration, ESTIMATE, TIDE, IPS, GSEA, drug-sensitivity, and gene-immune association analyses).

Figures S22-S30. Presented in the preceding pages (functional enrichment, GSE39582 performance checks, GSEA, and predicted drug-sensitivity analyses).