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.

A four-gene ferro-aging-related score is associated with survival in hepatocellular carcinoma

Zunqiang Dai
Qingdao Hospital, Peking University People's Hospita
Sheng Chen
Qingdao Hospital, Peking University People's Hospita
Lina Mao
The First Affiliated Hospital with Nanjing Medical University
Pengcheng Zhao
The First Affiliated Hospital with Nanjing Medical University
Yanbin Dong
yang18361269628@163.com

The First Affiliated Hospital with Nanjing Medical University

Research Article

Keywords: hepatocellular carcinoma, ferro-aging, gene expression, overall survival, prognostic association, public cohort analysis

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11436800/v1

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

Additional Declarations: No competing interests reported.

A four-gene ferro-aging-related score is associated with survival in hepatocellular carcinoma

Zunqiang Dai[1,†], Sheng Chen[1,†], Lina Mao[2], Pengcheng Zhao[2], Yanbin Dong[2,*]

[1] Department of Thoracic Surgery, Qingdao Hospital, Peking University People's Hospital, Qingdao 266000, China

[2] Department of Emergency and Critical Care Medicine, the First Affiliated Hospital with Nanjing Medical University, Nanjing 210029, China

† Zunqiang Dai and Sheng Chen contributed equally to this work.

Correspondence: Yanbin Dong; email: yang18361269628@163.com

Article type: Analysis

Abstract

Ferro-aging-related transcriptional features provide a candidate basis for studying hepatocellular carcinoma prognosis, but their association with survival requires evaluation across cohorts. We derived a four-gene score comprising *HMOX1*, *TFRC*, *SOCS2* and *E2F1* in 363 patients from The Cancer Genome Atlas liver cancer cohort and examined two public cohorts. Higher scores were associated with shorter overall survival in TCGA-LIHC, GSE14520 and GSE16757. Hazard ratios per standard deviation were 1.84 (95% confidence interval 1.52–2.22), 1.43 (1.15–1.78) and 1.39 (1.03–1.88), respectively. A conditional bootstrap procedure yielded a mean held-out concordance index of 0.657, with empirical 2.5th–97.5th percentiles of 0.588–0.724. This procedure retained the initial candidate screen and did not repeat the complete model-development process. Associations with public senescence gene sets varied, including essentially no correlation with SenMayo. Molecular and single-cell analyses provided exploratory context, while predicted drug-response differences varied by compound. These retrospective findings support a cross-cohort prognostic association under cohort-dependent preprocessing. They motivate independent evaluation with fixed preprocessing and clinical covariates, without establishing a causal ferro-aging mechanism or treatment benefit.

Keywords: hepatocellular carcinoma; ferro-aging; gene expression; overall survival; prognostic association; public cohort analysis

Introduction

Hepatocellular carcinoma (HCC) exhibits considerable heterogeneity in its clinical course and molecular features. Its course reflects both the tumour and the underlying liver disease, while genomic and transcriptomic profiling identifies substantial variation among tumours [1, 2]. Gene-expression scores can summarize aspects of this variation and test their association with survival in retrospective cohorts. However, a survival association alone does not show that a score measures a particular biological process, improves decisions beyond clinical variables or predicts response to treatment. These distinctions are especially relevant when a score is derived from a biologically selected gene list and then interpreted through several computational analyses of the same cohort.

Ferroptosis is an iron-dependent form of regulated cell death [3]. Recent primate aging research described ferro-aging as a chronic iron–lipid peroxidation program associated with cellular senescence [4, 5]. That study also reported experimental effects of ACSL4 inhibition, including inhibition by vitamin C, in its aging models. This provides a rationale for examining ferro-aging-related transcriptional features in cancer. It does not establish that the same process drives HCC progression or that vitamin C would benefit patients selected by a tumour expression score. The transfer from an aging-related gene list to a cancer prognosis analysis therefore requires separate evaluation of the survival association and its biological interpretation.

Ferroptosis-related prognostic signatures have already been reported in HCC [6, 7], and recent work has also examined genes related to both senescence and ferroptosis in this disease [8]. The relevant question is consequently whether a specified candidate list yields a reproducible survival association and how that association relates to existing expression programs. Public senescence gene sets, including SenMayo, offer external definitions against which to examine this relationship [9]. Agreement with some sets and disagreement with others can help delimit the meaning of a score, although neither pattern is a direct measurement of cellular senescence in tumour tissue.

Here, we developed an expression score from a 95-gene candidate list in The Cancer Genome Atlas liver hepatocellular carcinoma cohort (TCGA-LIHC). The retained coefficients were then examined in the public cohorts GSE14520 and GSE16757 [10, 11]. We then compared its survival information with gene-set representations of published ferroptosis signatures, evaluated relationships with public aging and senescence sets, and described selected copy-number, methylation and immune associations. Single-cell expression, virtual knockout and predicted drug-response analyses were used to generate hypotheses about the score's context. The primary objective was to assess a cross-cohort association with overall survival, while keeping mechanistic and therapeutic interpretations separate from the evidence supplied by these observational data.

Methods

Study design and data sources

This retrospective computational study reused public bulk and single-cell expression datasets. TCGA-LIHC was used for score development, and GSE14520 and GSE16757 were used to examine the retained coefficients in additional cohorts. TCGA expression was

obtained from the UCSC Xena Toil resource, with survival information from the TCGA clinical outcome resource [12, 13, 14]. The local source expression file was TcgaTargetGtex_rsem_gene_tpm, and survival data were taken from Survival_SupplementalTable_S1_20171025_xena_sp. Overall survival was the common endpoint assessed across the three bulk-expression cohorts. No new participants were recruited, samples collected or experimental interventions performed for this analysis.

Candidate genes and TCGA expression preprocessing

The candidate list comprised the 95 genes in the Ferroaging gene set worksheet of the published supplementary workbook mmc3.xlsx [4]. The related primate study describes a set comprising 39 genes overlapping general aging signatures and 56 additional ferro-aging-related genes [4]. For the present analysis, the workbook was the operational source of membership (Supplementary Table S1). The saved human mapping used gene-symbol annotation through org.Hs.eg.db; it did not document a completed one-to-one cross-species orthology analysis. Seventy-one mapped symbols were present in the processed TCGA expression matrix.

Ensembl identifiers were stripped of version suffixes and mapped to human gene symbols. When multiple expression rows mapped to the same symbol, the row with the highest mean expression was retained. Primary-tumour and adjacent-normal samples were identified by TCGA sample-type codes 01 and 11. The input Toil values were already log transformed; genes were retained if their mean value on the supplied log2(TPM + 0.001) scale was at least 0.5. The resulting expression matrix comprised 11,206 genes across 419 tumour and adjacent-normal samples. Differential expression was estimated using limma linear models and empirical Bayes moderation [15], with tumour/normal group as the model term. The analysis treated the groups as unpaired and did not include a patient-pair term. Candidate genes were selected using Benjamini–Hochberg-adjusted $P < 0.05$ and $|\log2 \text{ fold change}| > 1$; the selected candidates are listed in Supplementary Table S2.

Score derivation and survival association

The 17 differentially expressed candidate genes underwent univariable Cox screening at nominal $P < 0.1$, yielding a nine-gene pool. Model selection used a LASSO Cox model with ten-fold cross-validation and the lambda.min penalty, followed by a multivariable Cox model and bidirectional stepwise reduction [16]. The random seed was set to 42 before model selection and cross-validation. LASSO fitting used standardized expression in the 363 samples with positive, non-missing survival time and non-missing event status. The retained coefficients were then used to calculate the score:

$$\text{Score} = 0.210475497478991 \times \text{z}(\text{HMOX1}) + 0.204299015473899 \times \text{z}(\text{TFRC}) - 0.392831710797402 \times \text{z}(\text{SOCS2}) + 0.188153864735699 \times \text{z}(\text{E2F1})$$

Here, z(gene) denotes centred expression divided by its sample SD in the relevant scoring population. For the saved TCGA risk table, expression was restandardized across all 369 tumour samples. Samples at or above the median score were labelled high before excluding the six samples without eligible survival times. This differs from the 363-sample standardization used for coefficient fitting. We report the recorded risk table and distinguish these operations rather than treating them as a single fixed preprocessing rule. Continuous-score Cox HRs were calculated per SD of the score within each eligible survival cohort. The

post-selection gene-level estimates in Table 1 do not account for model-selection uncertainty. The HRs in Table 2 are unadjusted score associations, not effects adjusted for clinical covariates.

Table 1 Retained coefficients and conditional estimates in the four-gene Cox model

Gene	Coefficient	HR	95% CI	Nominal P
HMOX1	0.210475	1.234	1.041–1.463	0.0152
TFRC	0.204299	1.227	0.998–1.507	0.0518
SOCS2	−0.392832	0.675	0.568–0.803	8.49×10^{-6}
E2F1	0.188154	1.207	0.979–1.488	0.0777

The model was fitted in 363 patients with 128 deaths. Gene expression was standardized within those patients for model fitting. Coefficients are rounded here; full precision is retained in the Methods equation and source table. HRs and CIs are conditional on the selected model and do not incorporate the preceding selection procedure. Final score calculation subsequently used the standardization populations described in Methods.

External cohort preprocessing and score reconstruction

For GSE14520, the GPL3921 series matrix with 445 samples was used. For GSE16757, the local series matrix identified GPL6102 and contained 100 samples. Probe identifiers were mapped to gene symbols, retaining the probe with the highest mean expression when several probes mapped to the same symbol. The selected probes for *HMOX1*, *TFRC*, *SOCS2* and *E2F1* were 203665_at, 208691_at, 203373_at and 204947_at in GPL3921, and ILMN_1800512, ILMN_1674243, ILMN_1798926 and ILMN_1800738 in GPL6102. All four genes contributed to the reconstructed scores in both cohorts.

Gene-level standardization used all expression columns in the selected platform matrix before samples with missing or non-positive survival times or missing event status were excluded. This included 225 tumour and 220 non-tumour columns in GSE14520. Survival was linked by GEO sample identifier using the supplied cohort-specific clinical files, and GSE14520 survival months were converted to days using 30.44 days per month. External high/low groups were assigned at or above versus below the median score after eligibility filtering. Coefficients were retained from the TCGA model and were not refitted in external cohorts. Fixed scores were reconstructed from the local matrices and retained coefficients to check the stored implementation. Reconstruction used the local GPL3921 annotation and an official GPL6102 annotation retrieved on 15 October 2026. Reconstructed values matched every stored eligible external-cohort score within an absolute tolerance of 10^{-8}. The newly retrieved annotation is a reproducibility supplement, not evidence of the exact annotation download used in the original run.

Table 2 Continuous-score associations with overall survival across cohorts

Cohort	Patients	Deaths	High/low groups	HR per SD	95% CI	Nominal P
TCGA-LIHC	363	128	180/183	1.84	1.52–2.22	2.76×10^{-10}
GSE14520	221	85	111/110	1.43	1.15–1.78	0.00114
GSE16757	100	43	50/50	1.39	1.03–1.88	0.0317

HRs are from unadjusted Cox models using the continuous risk score standardized within each eligible survival cohort. Group counts refer to the saved high/low classifications and are not the contrasts estimated by these HRs. The external scores retain the TCGA coefficients but use cohort-specific gene standardization and median grouping; a fixed cross-platform cutoff was not tested.

Conditional bootstrap assessment

The bootstrap procedure sampled the 363 eligible TCGA patients with replacement 1,000 times, with seed 42. The nine-gene candidate pool was fixed from the full-data screen. In each resample, gene expression was standardized using the resampled patients, LASSO selection was refitted by ten-fold cross-validation, and an unpenalized Cox model was fitted to the selected genes. The bootstrap-sample means and SDs were then applied to patients omitted from that resample. Concordance was evaluated directly from their predicted scores, with higher scores defined as greater risk, without fitting an additional model to held-out patients. The procedure did not repeat differential-expression screening, univariable screening or stepwise reduction. We report the mean held-out concordance and empirical 2.5th–97.5th percentile range across evaluable iterations. These summaries describe a conditional selection-and-refitting procedure, not full-pipeline validation of the final four-gene model.

Published gene-set comparisons and aging scores

Published ferroptosis signatures were represented by their gene memberships using single-sample gene-set enrichment analysis (ssGSEA) implemented with GSVA [17]. The Liang et al. list contained ACACA, ACSL3, CISD1, CARS, G6PD, GPX4, NQO1, NFS1, SLC7A11 and SLC1A5; the Hu et al. list contained CAPG, FLT3, G6PD, HAVCR1, *HMOX1*, IL33, MT3, SLC7A11, SRXN1 and STMN1 [6, 7]. Cox models with each gene-set score alone were compared with models additionally containing the risk score. Concordance differences were descriptive; nested-model likelihood-ratio tests assessed added model fit. These gene-set score comparisons did not reapply the original published model coefficients.

Public aging and senescence sets were obtained through msigdbr from MSigDB collections [18], including SenMayo [9], and intersected with the processed expression matrix. Sets were retained if at least five genes were represented in the processed matrix, and ssGSEA normalization was enabled. The saved score object retained the actual gene memberships used, from which 24 publicly defined sets were extracted for interpretation. Risk-score relationships were examined by Spearman correlation, and selected set scores were examined in per-SD Cox models and models adding the set score to the risk score. Previously calculated correlation and group-comparison FDR values retain the original 25-score testing family; removing an internal set from biological interpretation did not trigger

recalculation of the retained FDR values. Cox and likelihood-ratio P values were treated as nominal exploratory results unless a specific adjustment is stated. The historical MSigDB release was not recorded; reproducibility therefore depends on the archived membership lists rather than an assumed current release.

Copy-number and DNA methylation analyses

Gene-level thresholded GISTIC copy-number values for TCGA-LIHC were obtained from UCSC Xena. Copy-number states were defined as −2, homozygous deletion; −1, loss; 0, diploid; 1, gain; and 2, amplification. Any alteration was defined as |GISTIC score| ≥ 1, combining gains and losses. Risk-group differences in alteration proportions were examined using Fisher's exact test, and copy-number–expression relationships used Spearman correlation in samples with both measurements. Cox models also evaluated survival associations with copy number expressed as a continuous GISTIC score. The *E2F1* analyses had different denominators for the copy-number distribution, expression correlation and survival association (370, 362 and 356, respectively; Supplementary Table S7).

Illumina HumanMethylation450 beta values were linked to the signature genes using the retained probe annotation. Tumour/normal differences were assessed by Wilcoxon rank-sum tests, methylation–expression relationships by Spearman correlation, and survival associations by Cox models per SD of methylation. Tumour/normal FDR adjustment used the 67 tested gene-associated probe records, while Cox P values were adjusted within each gene's tested probe set. TSS-proximal probes were defined by a strand-aware distance between −1,500 and +1,500 bases. Among probes with negative expression correlation and nominal $P < 0.05$, cg00509772 had the smallest correlation P value for *E2F1*. Its recorded distance was +1,063 bases from the transcription start site. The retained annotation did not establish a promoter or CpG-island label for this probe. Probe selection and subsequent interpretation were performed in the same TCGA data, so the association was not treated as independent mechanistic validation.

Immune-related measurements and predicted drug response

The retained expression matrix contained two genes from the planned checkpoint panel, HAVCR2 and VSIR. High/low score groups were compared by Wilcoxon rank-sum tests with Benjamini–Hochberg adjustment across these two genes. TIDE output was linked to the existing patient risk groups [19]. The overall and component scores were summarized descriptively using group medians; no new TIDE significance tests were generated for this manuscript. A TIDE-derived CD274 output was not substituted for measured CD274 transcript expression.

Drug-response estimation used oncoPredict with the supplied GDSC2 expression and response training objects [20]. The recorded settings included empirical-Bayes batch correction, phenotype power transformation, removal of the lowest 20% of varying genes using the homogenized data setting, and a minimum of ten samples. Predictions were generated for 369 tumour samples and then restricted to the 363 samples with retained risk groups. The 16 drug records were compared using two-sided Wilcoxon rank-sum tests, with Benjamini–Hochberg adjustment across all 16 records. The two oxaliplatin identifiers were retained as distinct training-resource records. Per-patient predictions, group sizes and quantiles were preserved, allowing the numerical summaries to be checked independently of

the plotted images. Because the prediction workflow transforms response phenotypes, outputs are reported on the model's prediction scale and are not assigned micromolar units or interpreted as observed patient response.

Single-cell expression and virtual knockout

Processed GSE149614 counts and supplied cell annotations were used [21]. A downsample of 12,000 cells was drawn from 71,915 cells using seed 42. For descriptive expression summaries, counts were normalized to 10,000 counts per cell and log1p transformed. Cell-type mean expression was then standardized across cell types separately for each signature gene. The resulting heatmap values describe relative cell-type means; they are not raw expression averages or patient-level estimates.

We used scTenifoldKnk to model the consequences of virtual gene knockout [22]. Cells were selected using a hepatocyte annotation and a tumour-site label, yielding 2,300 cells from ten patients. These labels were not taken as independent proof of malignancy. Genes were ranked by mean log-normalized expression divided by (SD + 10^{-6}). The network combined the top 800 ranked genes with the candidate-list and signature genes. Intersecting this union with the count matrix yielded an 887-gene network. This is the implemented selection rule, rather than a high-variance-gene filter. Raw counts for the selected cells and genes were passed to scTenifoldKnk with *HMOX1* as the virtual knockout target. We used qc = FALSE and nCores = 1, leaving the remaining arguments at the package defaults. Adjusted perturbation P values were taken from the saved package output.

Fewer than five genes had adjusted $P < 0.05$, so exploratory GO biological-process enrichment used the top 20 genes ranked by perturbation P value. This ranked list included both the knockout target and non-significant genes. clusterProfiler was used with its annotated-gene background rather than the tested 887-gene network as the explicit universe [23]. The enrichment output represented 18 input genes and used an annotated background of 18,986 genes. Enrichment was treated as a ranked-list description and not as evidence for a validated downstream mechanism. ACSL4 was evaluated separately using the retained bulk expression, Spearman correlation and Cox outputs; it was not used to define the four-gene score or an internal functional module.

Statistical reporting and computational assistance

The primary unit for bulk survival analyses was the patient. Cell counts in the single-cell analysis were not treated as independent patient replication. Cox estimates are accompanied by their reported 95% CIs; these intervals do not incorporate model-selection uncertainty. All P values reported from the retained Wilcoxon and correlation analyses were two-sided. Benjamini–Hochberg adjustment was interpreted within each recorded testing family [24]. We distinguish nominal P values from adjusted values in the reported analyses. No hypothesis test or inferential model was refitted during manuscript preparation; only fixed-score reconstruction, source selection, counting and descriptive summaries were added. Values recorded as numerical zero for extremely small correlation P values were not reported as P = 0. Clinical-adjustment, calibration, decision-curve and tumour-mutational-burden results were not used to support the conclusions reported here because their inputs require further resolution. The supplied scripts use R; the recorded project environment identifies R 4.4.1, but a complete historical package-session record was not available.

Figures were redrawn from retained values using R 4.4.1, ggplot2 4.0.3 and patchwork 1.3.2. These are the documented figure-generation versions, not inferred historical analysis versions.

OpenAI Codex was used to assist with source organization, literature metadata checking and English drafting, figure-generation code and organization of source-data documentation. Numerical inferential results were taken from the retained analysis outputs, and no experimental observations or references were generated to fill missing evidence. The authors are responsible for the accuracy, scientific interpretation and final content of the manuscript.

Results

A four-gene score was derived from the ferro-aging-related candidate list

The processed TCGA-LIHC expression matrix contained 369 primary tumours and 50 adjacent-normal samples, with 11,206 genes retained after annotation and expression filtering. Of the 95 genes in the candidate list, 71 were represented in this matrix. The input and selection steps are summarized in Fig. 1a. Seventeen candidate genes met the differential-expression criteria of an adjusted P value below 0.05 and an absolute log2 fold change greater than 1. Survival analyses included 363 tumour samples with positive, non-missing overall-survival times and available event indicators; 128 deaths were recorded. One tumour sample lacked survival time and five had non-positive times.

Univariable screening, least absolute shrinkage and selection operator (LASSO) regression, and stepwise Cox regression retained *HMOX1*, *TFRC*, *SOCS2* and *E2F1*. Coefficients were positive for *HMOX1*, *TFRC* and *E2F1* and negative for *SOCS2* (Fig. 1b; Table 1). The selected-model confidence intervals for *TFRC* and *E2F1* included HR = 1, so retention does not imply nominal significance for every gene. The saved training groups comprised 180 high-score and 183 low-score patients under the original grouping rule described in Methods.

A higher continuous score was associated with shorter overall survival in TCGA-LIHC (Fig. 1c; Table 2). The hazard ratio (HR) per standard deviation (SD) was 1.84 (95% confidence interval [CI], 1.52–2.22; $P = 2.76 \times 10^{-10}$). A separate conditional bootstrap analysis produced 1,000 evaluable iterations and a mean held-out concordance index of 0.657 (Fig. 1d; Supplementary Table S4). The empirical 2.5th and 97.5th percentiles were 0.588 and 0.724. This procedure refitted LASSO selection and Cox coefficients within resamples but held the initial candidate screen fixed and did not repeat the final stepwise reduction. Its percentile range therefore describes the variability of that conditional procedure, rather than a confidence interval or an optimism-corrected performance estimate for the final four-gene model.

The score retained a survival association in two public cohorts

GSE14520 contributed 221 patients with eligible survival records and 85 deaths, and GSE16757 contributed 100 patients with 43 deaths. All four signature genes were represented in the probe mappings used for score reconstruction. Applying the retained TCGA coefficients yielded an HR per SD of 1.43 in GSE14520 (95% CI, 1.15–1.78; P = 0.00114). The corresponding HR in GSE16757 was 1.39 (95% CI, 1.03–1.88; P = 0.0317; Fig. 1c; Table 2). The corresponding high/low score groups contained 111/110 patients in

GSE14520 and 50/50 in GSE16757. These results support the direction of the survival association across the three cohorts.

The external analyses used cohort-specific expression standardization and median cutoffs. GSE14520 standardization included 225 tumour and 220 non-tumour samples before retaining 221 eligible patients; GSE16757 used all 100 local samples. TCGA coefficient fitting and final score calculation also used different standardization populations, as detailed in Methods and Supplementary Table S3. The cross-cohort results therefore support association under the recorded preprocessing, rather than validation of a fixed numerical cutoff for individual-patient deployment.

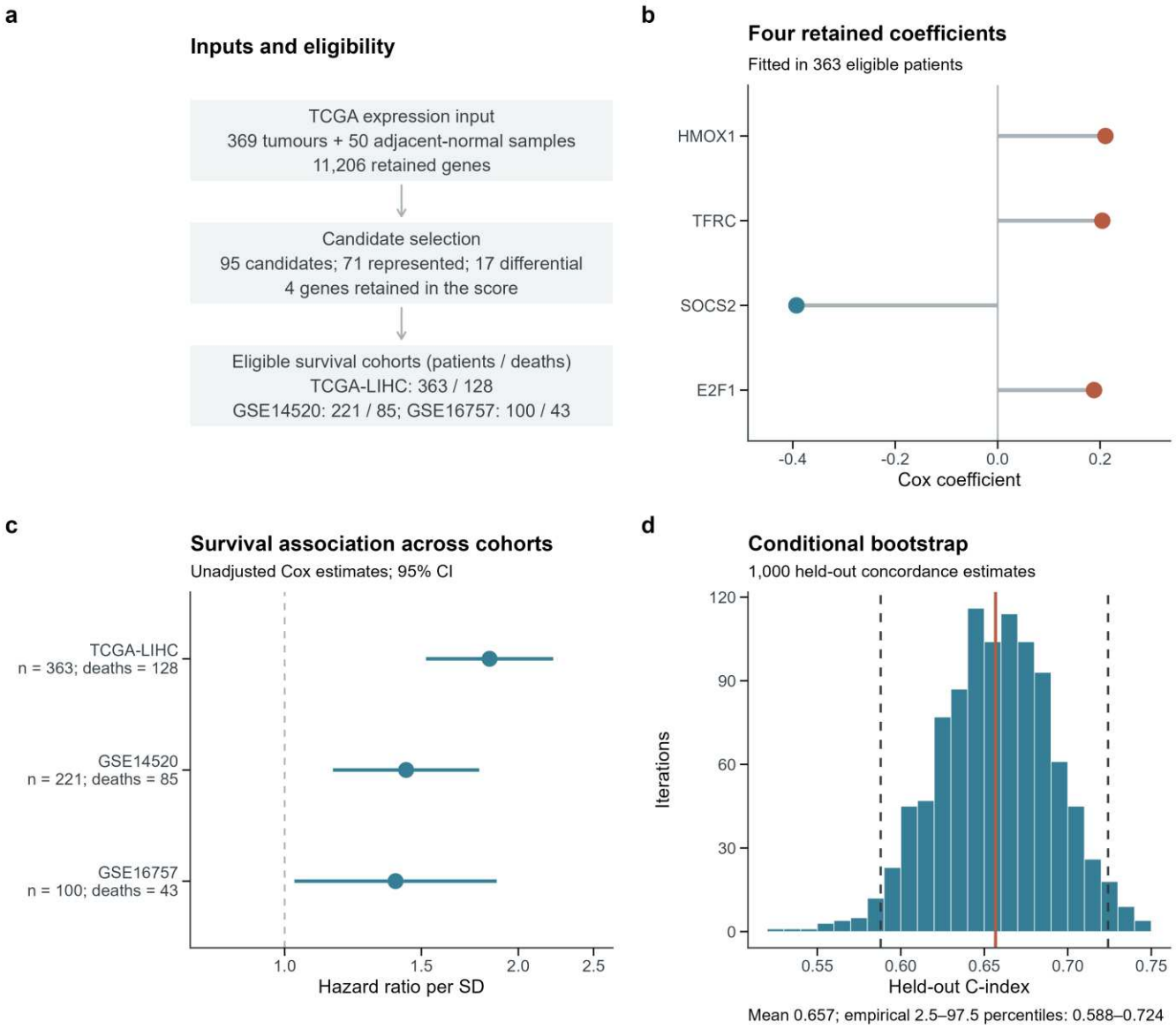

Fig. 1 Derivation and survival association of the four-gene score

a, Expression input, candidate selection and eligible survival cohorts. Adjacent-normal samples contribute to expression comparisons, not survival analyses. **b**, Retained Cox coefficients for *HMOX1*, *TFRC*, *SOCS2* and *E2F1*, fitted using standardized expression in 363 TCGA-LIHC patients. **c**, Unadjusted hazard ratios per standard deviation of the continuous score; horizontal lines show the reported 95% confidence intervals. TCGA-LIHC included 363 patients and 128 deaths; GSE14520 included 221 patients and 85 deaths; GSE16757 included 100 patients and 43 deaths. External scores retain TCGA coefficients

but use cohort-dependent standardization. The dashed reference denotes HR = 1. **d**, Histogram of held-out concordance from 1,000 conditional bootstrap iterations, using bins of width 0.01. The orange line denotes the mean; dashed lines denote empirical 2.5th and 97.5th percentiles. The procedure fixes the initial candidate screen and refits LASSO plus Cox coefficients without repeating stepwise reduction. These percentiles are not confidence limits for an unbiased performance estimate of the final four-gene model

Gene-set comparisons showed partial overlap with ferroptosis and senescence programs

For comparison with previous HCC studies, we represented two published ten-gene ferroptosis signatures as single-sample gene-set enrichment scores (Fig. 2a; Supplementary Table S5). In TCGA-LIHC, the concordance index increased from 0.666 for the Liang et al. gene-set score alone to 0.700 after adding the four-gene score; for the Hu et al. gene-set score, it increased from 0.660 to 0.681. The corresponding nested-model likelihood-ratio P values were 4.45×10^{-5} and 1.10×10^{-4}. This analysis compares apparent, within-cohort associations using summaries of gene-set expression. The likelihood-ratio tests compare model fit, and the reported changes in concordance do not establish statistical superiority over the original published coefficient-based models.

Relationships between the risk score and public aging and senescence sets were heterogeneous. The risk score correlated with the Reactome cellular-senescence score (Spearman ρ = 0.501; 132 represented genes; Fig. 2b). It also correlated with the Gene Ontology (GO) cellular-senescence score (ρ = 0.372; 77 genes; false discovery rate [FDR] = 9.48×10^{-13}). By contrast, its correlation with the represented SenMayo set was near zero (ρ = 0.00126; 73 genes; P = 0.981; Fig. 2c). The SenMayo score was also not associated with overall survival in its univariable Cox analysis (HR per SD, 0.94; 95% CI, 0.79–1.13; P = 0.519). Adding each of six examined public aging or senescence scores to the risk-score model gave nominal likelihood-ratio P values ranging from 0.218 to 0.615 (Supplementary Table S6). These findings indicate overlap with selected transcriptional programs without supporting interpretation of the score as a general measure of senescent-cell burden.

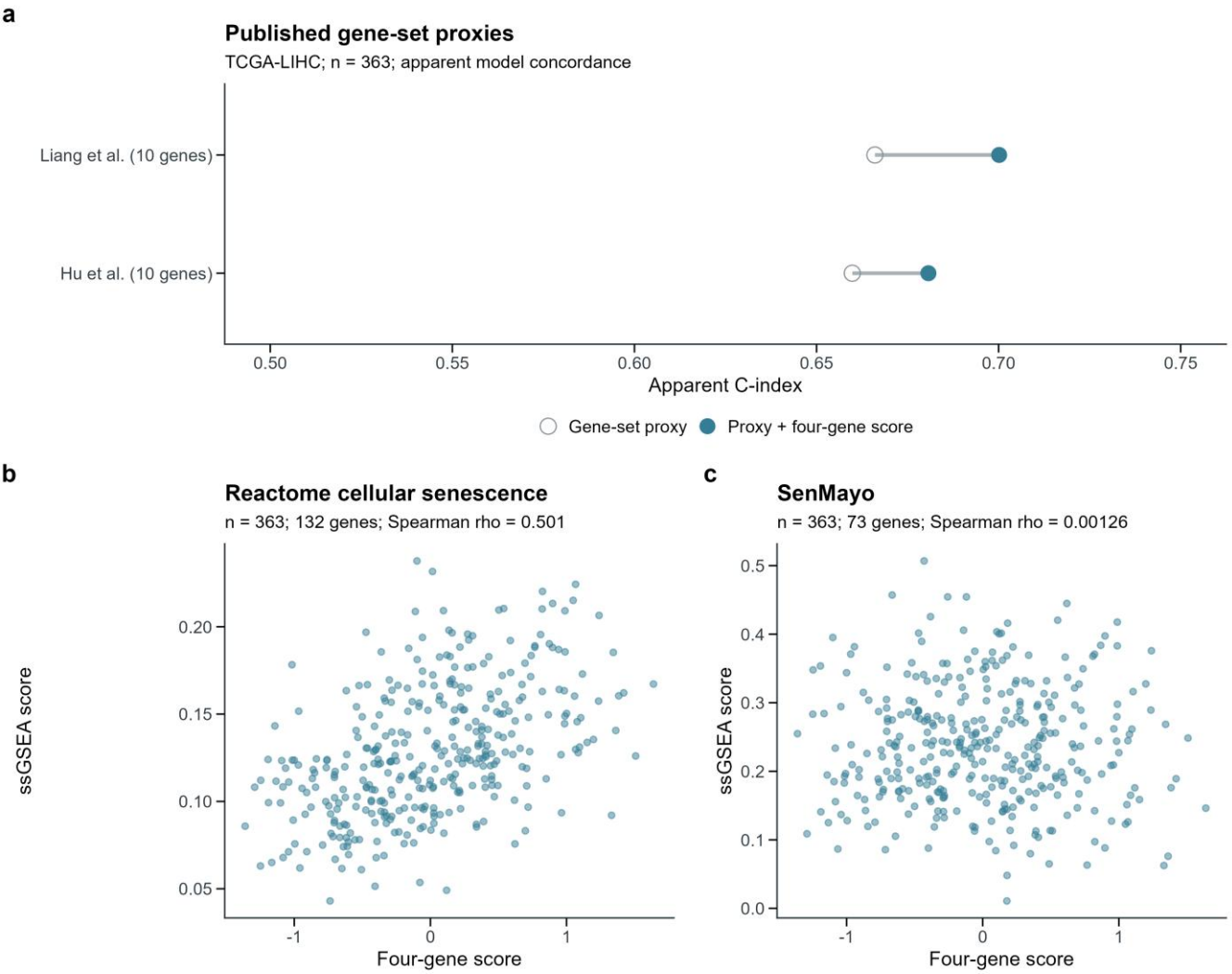

Fig. 2 Relationships with published gene-set summaries

a, Apparent concordance in 363 TCGA-LIHC patients for models containing the Liang et al. or Hu et al. ferroptosis gene-set proxy alone (open circles) or together with the four-gene score (filled circles). Horizontal segments connect the two model estimates for each gene set; they are not confidence intervals. Gene memberships were represented by ssGSEA; the published coefficient-based models were not reproduced. **b, c**, Relationships of the four-gene score with the represented Reactome cellular-senescence and SenMayo sets, comprising 132 and 73 genes, respectively. Each point represents one of the 363 patients; no points were removed for display. The annotated coefficients are Spearman correlations. The differing relationships limit interpretation of the score as a uniform measure of senescence. Confidence intervals for these correlations and apparent concordance estimates were not available in the retained outputs

E2F1 expression was associated with copy number and TSS-proximal methylation

Among 370 tumours with copy-number data, *E2F1* showed low-level gain in 30.27%, amplification in 1.08% and loss in 2.43% (Fig. 3a). Risk-group comparisons included 356 samples with eligible survival data and copy-number measurements. Any non-diploid *E2F1* state occurred in 43.75% versus 24.44% of the high- and low-score groups, respectively (nominal Fisher's exact $P = 1.38 \times 10^{-4}$). The alteration comparison pooled copy-number gains and losses within each risk group. *E2F1* copy number correlated positively with its expression in 362 paired samples ($\rho = 0.293$; $P = 1.28 \times 10^{-8}$; Fig. 3b). The continuous copy-

number survival model included 356 patients and 125 deaths. Its HR per GISTIC unit was 1.27 (95% CI, 0.91–1.77; P = 0.161; Fig. 3c), providing no nominally significant survival association.

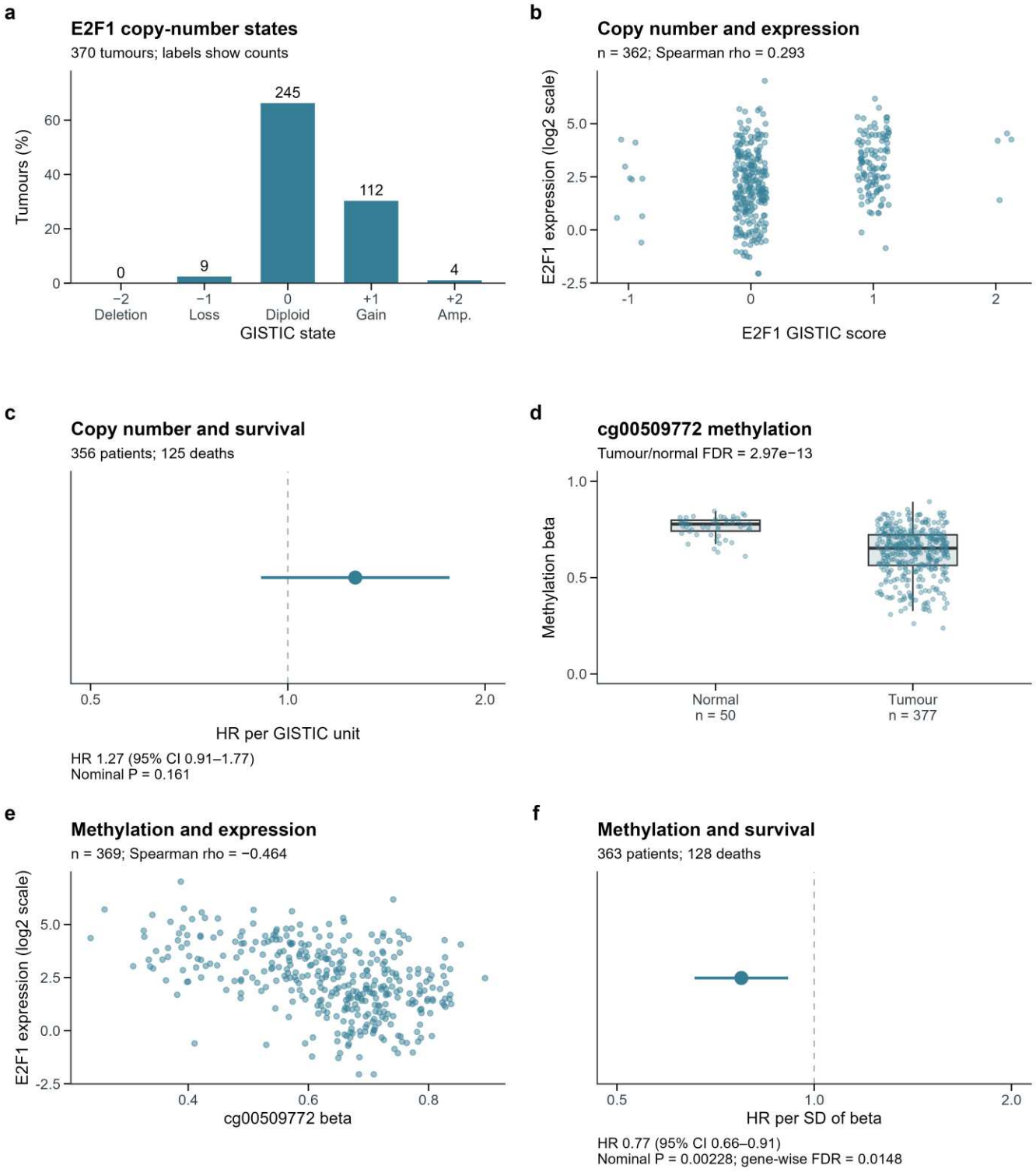

Fig. 3 E2F1 copy-number and TSS-proximal methylation associations

a, GISTIC states in 370 tumours; bar heights show percentages and labels show counts. **b**, Copy number and *E2F1* expression in 362 paired samples. Horizontal jitter separates observations at the same GISTIC state without changing expression values. **c**, Continuous copy-number Cox estimate per GISTIC unit in 356 patients with 125 deaths. **d**, cg00509772 methylation in 377 tumours and 50 adjacent-normal samples. Boxes show the median and first/third quartiles; whiskers reach the most extreme observations within 1.5 interquartile

ranges. All observations, including those beyond whiskers, are shown with horizontal jitter; none were removed as outliers. The Wilcoxon comparison uses Benjamini–Hochberg adjustment across the 67 tested gene-associated probe records. **e**, Methylation–expression relationship in 369 paired tumours. Correlation annotations in **b** and **e** are Spearman coefficients. **f**, Methylation Cox estimate per standard deviation of beta in 363 patients with 128 deaths; the FDR uses the tested *E2F1* probe family. In **c** and **f**, points show HRs, horizontal lines show reported 95% confidence intervals and dashed lines denote HR = 1. The probe lies +1,063 bases from the annotated transcription start site and was selected by expression association within a ±1,500-base window. These associations do not establish promoter silencing or a causal regulatory pathway

The *E2F1*-associated probe cg00509772 was located +1,063 bases from the annotated transcription start site in strand-aware coordinates. Mean methylation beta values were 0.631 in 377 tumours and 0.762 in 50 adjacent-normal samples (difference, −0.131; FDR = 2.97×10^{-13}; Fig. 3d). Methylation at this probe was negatively correlated with *E2F1* expression ($\rho = -0.464$; $P = 4.17 \times 10^{-21}$; 369 paired samples; Fig. 3e). Higher methylation was associated with a lower hazard in the probe-level survival model (HR per SD, 0.77; 95% CI, 0.66–0.91; P = 0.00228; gene-wise FDR = 0.0148; 363 patients; Fig. 3f; Supplementary Table S8). Because the probe was selected for its expression association within a transcription-start-site window, these results are exploratory associations and do not demonstrate promoter silencing or a causal copy-number–methylation–expression pathway.

Immune associations and predicted drug responses varied by measure and drug

HAVCR2 expression was higher in the high-score group, with median log-scale expression values of 0.986 and 0.139 in the high- and low-score groups, respectively (FDR = 4.22×10^{-9}). VSIR did not meet the FDR threshold of 0.05 in the same two-gene comparison (FDR = 0.0713; Fig. 4a). Annotated Tumor Immune Dysfunction and Exclusion (TIDE) output was available for all 363 patients with retained risk groups. Descriptive medians were higher in the high-score group for overall TIDE (−0.023 versus −0.135) and exclusion (0.860 versus 0.807). The myeloid-derived suppressor-cell score median was also higher (0.0471 versus 0.0255). The dysfunction-score median was lower in the high-score group (−0.606 versus −0.508; Fig. 4b). These computational measures describe distinct immune-related features and should not be combined into a uniform claim of T-cell dysfunction or resistance to immunotherapy.

Predicted drug-response values were available for 16 drug records in all 363 patients, including 180 high-score and 183 low-score patients (Fig. 4c; Supplementary Table S9). The high-score group had lower predicted values for paclitaxel (FDR = 2.07×10^{-10}), 5-fluorouracil (FDR = 0.00649) and sorafenib (FDR = 0.0367). Higher values occurred for erlotinib (FDR = 5.61×10^{-5}), gefitinib (FDR = 1.02×10^{-4}) and irinotecan (FDR = 0.00220). Both retained oxaliplatin records also had higher predicted values in this group (FDR = 0.00131 and 0.0273). Gemcitabine and cisplatin showed no group difference at the FDR threshold of 0.05 (FDR = 0.435 and 0.949). The values are predictions on the model's response scale, not measured patient IC50 values. They support drug-specific computational hypotheses, without demonstrating general chemotherapy resistance or clinical treatment benefit.

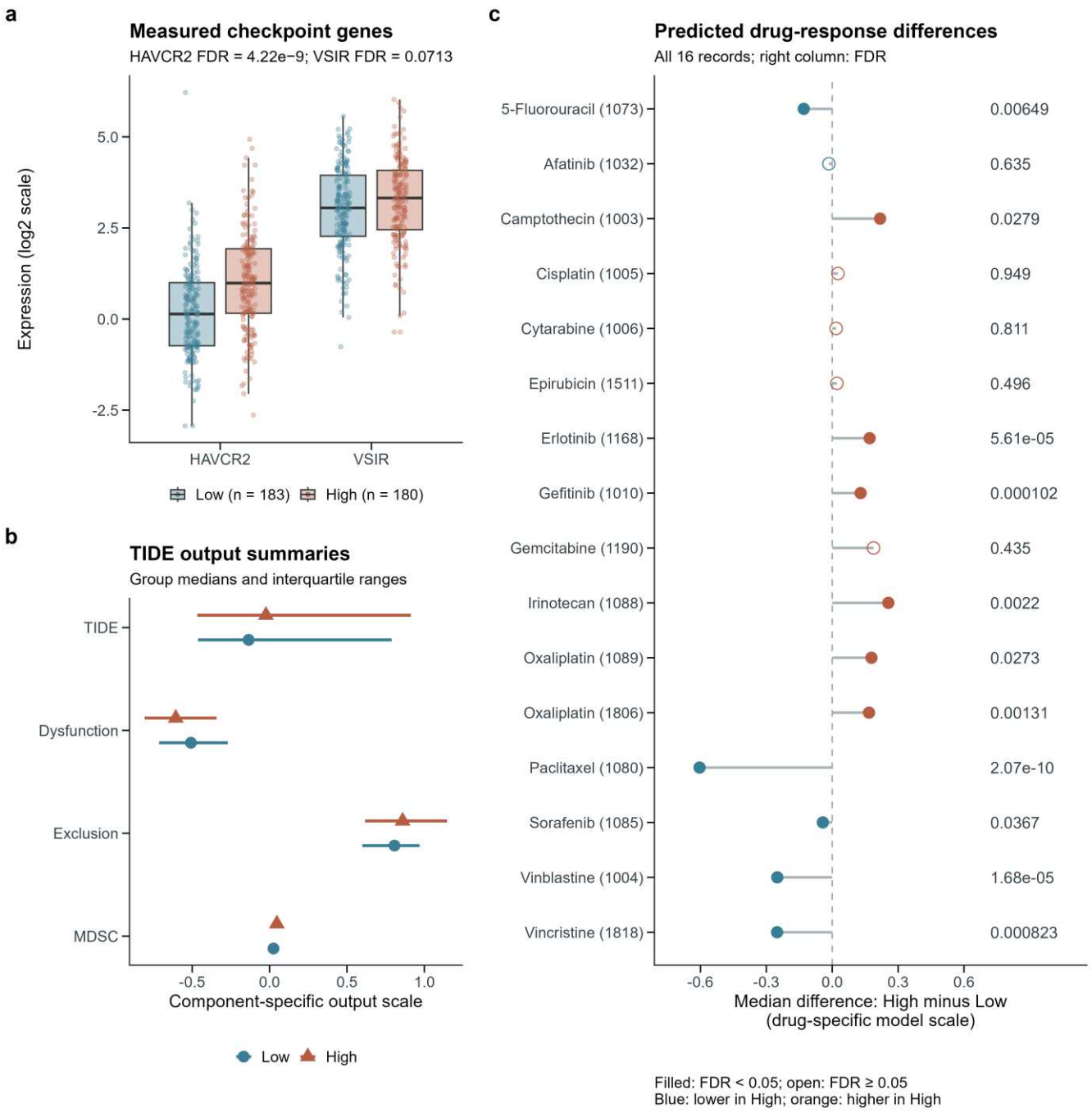

Fig. 4 Immune measurements and drug-specific computational predictions

a, HAVCR2 and VSIR expression in low-score (blue; n = 183) and high-score (orange; n = 180) patients. Boxes show medians and first/third quartiles; whiskers reach the most extreme observations within 1.5 interquartile ranges. All patient values are shown with horizontal jitter, including values beyond whiskers. Wilcoxon P values were adjusted across the two genes. **b**, TIDE and component summaries in the same patients. Circles and triangles show low- and high-score medians; horizontal lines show interquartile ranges, not confidence intervals. These are descriptive summaries without newly calculated significance tests. Each component retains its own output scale. **c**, Differences between high- and low-score median predictions for all 16 drug-resource records, ordered alphabetically. Each record has 180 and 183 evaluable predictions, respectively. Negative/blue values indicate lower predictions in the high-score group; positive/orange values indicate higher predictions. Filled/open circles indicate FDR below/at or above 0.05; exact adjusted values appear at right. Wilcoxon P values were adjusted across 16 records, and both oxaliplatin training

identifiers remain separate. The horizontal segments connect zero to each difference and do not represent uncertainty. These drug-specific transformed prediction scales have no assigned micromolar units; their magnitudes do not rank clinical efficacy across drugs. Predictions do not establish observed patient response or treatment benefit

Single-cell analysis and virtual knockout provided exploratory cellular context

We examined single-cell expression in GSE149614 using its public processed matrix and annotations [21]. The analysis used a reproducible downsample of 12,000 cells from the supplied matrix of 71,915 cells, representing 10 patients and 21 samples. The downsample contained 4,278 T/NK cells, 3,437 hepatocytes, 2,664 myeloid cells, 632 endothelial cells, 608 B cells and 381 fibroblasts (Fig. 5a). In the saved heatmap values, which standardize each gene's mean expression across cell types, *HMOX1* and *TFRC* were highest in myeloid cells, *SOCS2* in endothelial cells and *E2F1* in hepatocytes (Fig. 5b; Supplementary Table S10). These descriptive cell-type differences indicate that the bulk score may combine signals from several cellular compartments; they do not assign all four genes to a single tumour-cell program.

A virtual *HMOX1* knockout used 2,300 tumour-site cells annotated as hepatocytes and an 887-gene network. Apart from *HMOX1* itself, APOC1 was the only gene with an adjusted perturbation P value below 0.05 (adjusted $P = 3.44 \times 10^{-8}$; Fig. 5c). The associated exploratory Gene Ontology analysis used the 20 highest-ranked genes because fewer than five genes passed the significance threshold. That ranked list included non-significant genes as well as the knockout target (Supplementary Table S11). The result is a network-model prediction in a pooled cell subset, rather than experimental evidence of *HMOX1*-dependent regulation or a confirmed downstream pathway in patients.

ACSL4 was analysed separately because it was not a member of the 95-gene candidate list. Its expression was higher in tumour than adjacent-normal samples (Wilcoxon $P = 3.05 \times 10^{-12}$; 369 versus 50 samples). It also correlated positively with the risk score ($\rho = 0.235$; $P = 5.90 \times 10^{-6}$; Fig. 5d). In a joint Cox model, the ACSL4 HR per SD was 0.97 (95% CI, 0.80–1.17; P = 0.723). Adding ACSL4 changed the apparent concordance index from 0.6792 to 0.6795 (Fig. 5e; Supplementary Table S12). Thus, the observed expression relationship did not provide evidence of appreciable additional prognostic information in this cohort. No vitamin C intervention was evaluated in these HCC data.



Fig. 5 Exploratory cellular context, virtual perturbation and ACSL4 analysis

a, Composition of the saved 12,000-cell downsample from 71,915 input cells, representing ten patients and 21 samples. **b**, Mean normalized expression standardized across cell types separately for each signature gene; colours represent the saved gene-wise z scores, without additional scaling or clustering. These values show relative cell-type means, not absolute expression. **c**, The top 20 genes ranked by saved perturbation P value after virtual *HMOX1*

knockout in an 887-gene network. The network used 2,300 tumour-site cells annotated as hepatocytes from ten patients. The triangle denotes the target; the filled circle denotes APOC1, the only non-target gene with adjusted $P < 0.05$; open circles denote the remaining ranked genes. Perturbation distance is shown on a logarithmic axis. Cells are not independent patient replicates, and virtual knockout is computational. **d**, ACSL4 expression and the four-gene score in 363 patients; each point represents a patient, and the annotation reports Spearman correlation. **e**, Apparent concordance before and after adding ACSL4 to the score model; estimates are descriptive and have no supplied uncertainty interval. The displayed joint-model ACSL4 HR and 95% CI provide its conditional association estimate. No vitamin C intervention was evaluated

Discussion

This study identified a four-gene expression score with a consistent direction of association with overall survival across TCGA-LIHC and two public HCC cohorts. The external-cohort HRs were smaller than the discovery-cohort estimate but remained above 1. The results support a retrospective prognostic association under the preprocessing used in each cohort. A transportable clinical prediction rule remains unestablished because expression standardization and group thresholds depended on each cohort. The reported association models also did not establish independence from tumour stage, liver function or treatment.

The biological label attached to the candidate list needs particular care. Ferro-aging was proposed from experimental aging research involving iron accumulation, lipid peroxidation and ACSL4 [4, 5]. The present score was instead selected for its relationship with survival in bulk HCC expression data. The source list motivated gene selection, but selection by survival association can retain features of proliferation, tissue composition or other correlated processes. The positive relationships with some senescence sets, together with the near-zero SenMayo correlation, argue against treating the resulting score as an established measure of senescent-cell burden. Equally, the SenMayo result in this filtered HCC matrix does not invalidate that gene set in the tissues and experimental settings in which it was developed [9].

Previous HCC ferroptosis signatures and senescence–ferroptosis analyses limit any claim that these biological themes are new to HCC [6, 7, 8]. Our comparisons address a narrower question: whether the four-gene score contains survival information beyond two published gene lists represented as enrichment scores. The in-sample improvements in model fit are compatible with partially non-overlapping information. Because the original published coefficients and preprocessing were not reproduced, they do not demonstrate that the present model outperforms those published models. A fair predictive comparison would require common eligibility rules, prespecified preprocessing and evaluation in data not used for any model selection or refitting.

The *E2F1* analyses provide molecular context without resolving direction of causation. Positive copy-number–expression correlation and inverse methylation–expression correlation are compatible with multiple determinants of *E2F1* expression. Yet the continuous copy-number survival analysis was inconclusive, and the methylation probe was selected using its association with expression. These features preclude a claim that copy-number gain and promoter hypomethylation jointly drive the score or explain its survival association. The appropriate interpretation is that selected genomic measurements covary with a score component and merit independent study.

Cellular composition offers another possible contributor to the bulk signal. The single-cell summaries placed *HMOX1* and *TFRC* most strongly in the myeloid compartment, *SOCS2* in endothelial cells and *E2F1* in hepatocytes within the analysed downsample. This observation motivates cell-type-aware analyses, but it does not quantify how much of the bulk survival association is attributable to composition. The virtual knockout likewise offers a specific hypothesis involving *HMOX1* and APOC1. Its limited number of adjusted-significant genes, pooled-cell design and ranked-list enrichment fallback constrain broader pathway claims. Experimental perturbation and patient-level replication would be needed to test this hypothesis.

The immune and drug analyses should be interpreted at the level of their individual outputs. TIDE models aspects of immune dysfunction and exclusion [19], and oncoPredict estimates drug response from cell-line screening data [20]. Neither supplies observed treatment outcomes for the patients analysed here. The opposing directions among drug predictions, and the differing directions of TIDE component scores, are therefore relevant findings rather than inconsistencies to be compressed into a general resistance phenotype. They do not support choosing a drug, excluding immunotherapy or expecting vitamin C benefit from the four-gene score. Similarly, the ACSL4–score correlation did not translate into an appreciable increase in the recorded concordance index, which limits its contribution to the present prognostic argument.

Limitations

Several features of the study constrain its interpretation and reproducibility. First, the study used retrospective public cohorts and has no prospective clinical or experimental validation. The survival analyses included 363 TCGA-LIHC patients, 221 GSE14520 patients and 100 GSE16757 patients; the single-cell analysis represented ten patients. Second, the candidate list, differential-expression filter and univariable screen were selected before the conditional bootstrap. The bootstrap did not repeat the entire discovery procedure, and its empirical percentile range cannot be used as uncertainty for an unbiased final-model performance estimate. Third, the TCGA fitting and scoring populations differed, and GSE14520 standardization included non-tumour samples before survival filtering. Reproducible reconstruction of the recorded scores does not remove these design limitations. Fourth, the analysis did not establish proportional-hazards adequacy or incremental value beyond a fully specified clinical model. Fifth, transcriptomic gene-set scores, cell-line-trained drug predictions and pooled single-cell network perturbations have distinct biological and statistical units. Their convergence cannot substitute for an experimental mechanism or a treatment-response study. The exact historical releases of some annotation and gene-set resources also remain incompletely documented, although retained gene memberships, source tables and the reconstructed external probe mappings enable targeted reproduction.

Conclusions

The current evidence supports further evaluation of the *HMOX1–TFRC–SOCS2–E2F1* score as a survival-associated transcriptional feature in HCC. The next validation should fix preprocessing and thresholds before evaluating an independent cohort, assess added information beyond clinical predictors and examine cell-type contributions at the patient level. Until those conditions are met, the score is best understood as a retrospective association with exploratory molecular context.

Declarations

Ethics approval and consent to participate

This study exclusively analysed publicly available data. No new participants were recruited and no new human samples were collected. No new ethical approval was required for this secondary analysis.

Funding

No funding was received for this study.

Competing interests

The authors declare no competing interests.

Author contributions

Zunqiang Dai: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft.

Sheng Chen: Conceptualization, Data curation, Validation, Writing – review & editing.

Lina Mao: Data curation, Investigation.

Pengcheng Zhao: Formal analysis, Visualization.

Yanbin Dong: Supervision, Resources, Writing – review & editing.

Declarations

Data availability

This study analysed publicly available datasets. TCGA-LIHC gene-expression data are available from the UCSC Xena Toil TcgaTargetGtex_rsem_gene_tpm matrix. Survival outcomes are available in the TCGA PanCanAtlas Survival_SupplementalTable_S1_20171025_xena_sp, and the corresponding TCGA-LIHC clinical annotations are available from UCSC Xena.

The external bulk-expression cohorts are available in the Gene Expression Omnibus under GSE14520 and GSE16757. Dataset-specific files comprise the GSE14520-GPL3921 series matrix, GPL3921 annotation and GSE14520 clinical supplement, and the GSE16757 series matrix, GPL6102 annotation and GSE16757 clinical outcome file. The provenance of the GPL6102 annotation used for score reconstruction is described in Methods. Public processed single-cell data are available under GSE149614, including the count matrix and updated cell metadata.

TCGA-LIHC thresholded GISTIC2 copy-number data, HumanMethylation450 beta values and the corresponding 450K probe annotation are available from UCSC Xena. The hg19 refGene coordinates used for transcription-start-site annotation are available from UCSC.

The 95-gene candidate list was derived from the Ferroaging gene set worksheet in the published supplementary workbook mmc3.xlsx, associated with https://doi.org/10.1016/j.cmet.2026.02.010. The candidate list and human-symbol mapping used in this study are provided in Supplementary Table S1, and the two published comparator gene lists are provided in Supplementary Table S5. Source records for all 24 public MSigDB gene sets, including SAUL_SEN_MAYO and REACTOME_CELLULAR_SENESCENCE, are listed in resource_catalogue.tsv within Supplementary Data 1. The exact represented gene memberships are provided in Supplementary Table S6. The original MSigDB release was not recorded; the archived memberships define the gene-set inputs used in the retained analysis.

The oncoPredict GDSC2 training objects, GDSC2_Expr (RMA Normalized and Log Transformed).rds and GDSC2_Res.rds, are available in DataFiles.zip through the OSF c6tfx file record. Checkpoint-expression, TIDE and per-patient drug-prediction outputs are provided in Supplementary Table S9.

Processed data supporting the reported results are provided in Supplementary Data 1 (Supplementary_Data_1.zip). This file contains Supplementary Tables S1-S12 as 26 CSV files, 24 CSV source-data files for Figures 1-5, a resource catalogue, file indexes, checksums and a README. These data include gene lists, cohort-specific eligible clinical outcomes and risk scores, conditional-bootstrap results, gene-set memberships, copy-number and methylation results, immune-related outputs, drug predictions, single-cell summaries and virtual-knockout results.Direct download URLs for every resource are listed in resource_catalogue.tsv in Supplementary Data 1. Data-related enquiries may be directed to the author, Sheng Chen (chensheng0722@126.com).

Code availability

The analysis scripts are publicly available at https://github.com/rabbit0722/ferro-aging-hcc-signature. The version recorded for this manuscript is commit 994dc9cc9540a62de0599026604af5093c153fcc. The TIDE implementation is available from the TIDEpy repository. Other software and annotation resources are indexed in resource_catalogue.tsv within Supplementary Data 1. Limitations in the historical software-version records are described in Methods.

Ethics approval

Not applicable. This study involved secondary computational analyses of publicly available, de-identified datasets, including existing cell-line screening data. No human participants were recruited, no human tissues were collected or handled, and no animal experiments or new in vitro experiments were performed. Ethics committee approval was therefore not required for the present analysis.

Consent to participate

Not applicable. The study used publicly available, de-identified data and did not involve the direct participation of identifiable individuals. Informed consent to participate was not required for this secondary analysis.

Consent to publish

Not applicable. The manuscript contains no identifiable personal data, images or other materials requiring individual consent for publication.

Supplementary material caption

Supplementary Data 1 (Supplementary_Data_1.zip). Source data, supplementary tables and resource catalogue. This archive contains Supplementary Tables S1-S12 (26 CSV files), 24 CSV source-data files for Figures 1-5, and a catalogue of the public datasets, gene sets, annotations and computational resources used in the study. The supplementary-table and figure-source indexes describe the individual files, fields and row counts. A README explains the package structure and provenance limitations, and a file manifest provides SHA-256 checksums.

References

1. Forner A, Reig M, Bruix J. Hepatocellular carcinoma. The Lancet. 2018;391:1301–1314. https://doi.org/10.1016/s0140-6736(18)30010-2

2. Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. Cell. 2017;169:1327–1341.e23. https://doi.org/10.1016/j.cell.2017.05.046

3. Dixon SJ, Lemberg KM, Lamprecht MR, Skouta R, Zaitsev EM, Gleason CE, et al. Ferroptosis: An Iron-Dependent Form of Nonapoptotic Cell Death. Cell. 2012;149:1060–1072. https://doi.org/10.1016/j.cell.2012.03.042

4. Liu L, Zheng Z, You W, Yang P, Wen Y, Qiao Y, et al. Vitamin C inhibits ACSL4 to alleviate ferro-aging in primates. Cell Metabolism. 2026;38:673–693.e17. https://doi.org/10.1016/j.cmet.2026.02.010

5. Liu L, Zheng Z, You W, Yang P, Wen Y, Qiao Y, et al. Vitamin C inhibits ACSL4 to alleviate ferro-aging in primates. [Correction] Cell Metabolism. 2026;38:2146. https://doi.org/10.1016/j.cmet.2026.08.011

6. Liang JY, Wang DS, Lin HC, Chen XX, Yang H, Zheng Y, et al. A Novel Ferroptosis-related Gene Signature for Overall Survival Prediction in Patients with Hepatocellular Carcinoma. International Journal of Biological Sciences. 2020;16:2430–2441. https://doi.org/10.7150/ijbs.45050

7. Hu Y, Hao C, Dong C, Tao P, Wang J, Lu Q. Identification of a novel ferroptosis-related gene signature in hepatocellular carcinoma: clinical significance, tumor microenvironment, drug sensitivity, and gene landscape analysis. Discover Oncology. 2025;16:1298. https://doi.org/10.1007/s12672-025-03123-9

8. Jiao C, Wang YP. TP53, HIF1A, and CDKN2A in Hepatocellular Carcinoma: Roles in Senescence, Ferroptosis, and Prognosis. Clinical and Translational Gastroenterology. 2026;17:e01002. https://doi.org/10.14309/ctg.0000000000001002

9. Saul D, Kosinsky RL, Atkinson EJ, Doolittle ML, Zhang X, LeBrasseur NK, et al. A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues. Nature Communications. 2022;13:4827. https://doi.org/10.1038/s41467-022-32552-1

10. Roessler S, Jia HL, Budhu A, Forgues M, Ye QH, Lee JS, et al. A Unique Metastasis Gene Signature Enables Prediction of Tumor Relapse in Early-Stage Hepatocellular Carcinoma Patients. Cancer Research. 2010;70:10202–10212. https://doi.org/10.1158/0008-5472.can-10-2607

11. Kim SM, Leem SH, Chu IS, Park YY, Kim SC, Kim SB, et al. Sixty-five gene-based risk score classifier predicts overall survival in hepatocellular carcinoma. Hepatology. 2012;55:1443–1452. https://doi.org/10.1002/hep.24813

12. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nature Biotechnology. 2020;38:675–678. https://doi.org/10.1038/s41587-020-0546-8

13. Vivian J, Rao AA, Nothaft FA, Ketchum C, Armstrong J, Novak A, et al. Toil enables reproducible, open source, big biomedical data analyses. Nature Biotechnology. 2017;35:314–316. https://doi.org/10.1038/nbt.3772

14. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;173:400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052

15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43:e47. https://doi.org/10.1093/nar/gkv007

16. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. Journal of Statistical Software. 2011;39:1–13. https://doi.org/10.18637/jss.v039.i05

17. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinformatics. 2013;14:7. https://doi.org/10.1186/1471-2105-14-7

18. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics. 2011;27:1739–1740. https://doi.org/10.1093/bioinformatics/btr260

19. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nature Medicine. 2018;24:1550–1558. https://doi.org/10.1038/s41591-018-0136-1

20. Maeser D, Gruener RF, Huang RS. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. Briefings in Bioinformatics. 2021;22:bbab260. https://doi.org/10.1093/bib/bbab260

21. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. Nature Communications. 2022;13:4594. https://doi.org/10.1038/s41467-022-32283-3

22. Osorio D, Zhong Y, Li G, Xu Q, Yang Y, Tian Y, et al. scTenifoldKnk: An efficient virtual knockout tool for gene function predictions via single-cell gene regulatory network perturbation. Patterns. 2022;3:100434. https://doi.org/10.1016/j.patter.2022.100434

23. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. 2021;2:100141. https://doi.org/10.1016/j.xinn.2021.100141

24. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology. 1995;57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData1.zip