

A three-gene expression score is equivalent to a 25-gene panel as an independent prognostic factor in clear cell renal cell carcinoma

Bin Lian

drmi1o@gkd.edu.cn

Guangzhou Vocational University of Science and Technology <https://orcid.org/0000-0002-1477-9137>

Huiling Zhai

Guangzhou Vocational University of Science and Technology <https://orcid.org/0009-0001-4945-8392>

Jiayi Kuang

Guangzhou Vocational University of Science and Technology <https://orcid.org/0009-0005-0826-2341>

Research Article

Keywords: clear cell renal cell carcinoma, prognostic index, gene expression signature, external validation, mitoxyperilysis-related gene set

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11318310/v2>

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.

A three-gene expression score is equivalent to a 25-gene panel as an independent prognostic factor in clear cell renal cell carcinoma

Running title: A three-gene score equals a 25-gene panel in ccRCC

Authors: Bin Lian^{1,*}, Huiling Zhai¹, Jiayi Kuang¹

Affiliations: ¹ School of Health Sciences, Guangzhou Vocational University of Science and Technology, No. 1038 Guangcong 9th Road, Zhongluotan, Baiyun District, Guangzhou 510550, Guangdong, China

* Corresponding author: Bin Lian, School of Health Sciences, Guangzhou Vocational University of Science and Technology, No. 1038 Guangcong 9th Road, Zhongluotan, Baiyun District, Guangzhou 510550, Guangdong, China. E-mail: drmilo@gkd.edu.cn. ORCID iD: 0000-0002-1477-9137.

Abstract

Purpose. Whether a cell-death-derived signature is prognostic in clear cell renal cell carcinoma (ccRCC) and survives parsimony, transportability and size-matched benchmarks is untested. Mitoxyperilysis is a lytic cell-death modality driven by innate immune stimulation plus metabolic disruption.

Methods. From GSE235046 we assembled a 101-gene mouse panel, 87 evaluable in TCGA-KIRC (n = 508; 168 deaths); "MRG" denotes that selection history, not effectors. A 25-gene ridge index and a three-gene LASSO-Cox score were benchmarked against size-matched random panels, adjusted for clinical variables, frozen in two cohorts, and compared at 36 months by reclassification and decision curves.

Results. Forty-five of 87 genes tracked survival at $P < 0.05$, 32 after stage and grade adjustment. The 25-gene index reached 0.711 with no significant coefficient, its 1-SE compression 0.705 against 0.706. The score was independently prognostic (HR 1.609 per SD; $P = 1.02 \times 10^{-9}$), recovered 26.4% of the room above its null, and held in E-MTAB-1980 but inverted in GSE29609. At 36 months it raised the C-index +0.026 and the IDI +0.041 beyond a size-matched null, while the NRI (+0.34) did not and the net-benefit gain included zero. High-risk tumours showed SETD2/BAP1 enrichment.

Conclusions. Equivalence, not superiority: 22 further genes bought no measurable discrimination from 25 terms at 168 events. It marks a proliferation-adjacent, SETD2/BAP1-mutant state, not the mitoxyperilysis pathway; its clinical gain is modest and transportability rests on one positive and one negative cohort.

Keywords: clear cell renal cell carcinoma; prognostic index; gene expression signature; external validation; mitoxyperilysis-related gene set

1. Introduction

Kidney cancer accounts for roughly 2.2% of cancer diagnoses and 1.8% of cancer deaths worldwide, with incidence rising for three decades¹, and clear cell renal cell carcinoma (ccRCC) is its dominant subtype. ccRCC is defined by loss of the von Hippel–Lindau tumour suppressor, which releases the HIF transcription factors and rewires the cell onto a glycolytic, glutamine-dependent, lipid-accumulating programme^{2,3}. That biology has been unusually productive therapeutically — ccRCC is the tumour type in which the angiogenic and mTOR arms can be targeted to benefit^{2,4}, and the first in which checkpoint inhibition produced an overall survival advantage⁵ — yet risk stratification for localised disease still rests on stage, grade and a handful of clinical variables³, with wide

residual spread inside every stage group. The transcriptomic prognostic indices proposed to fill that gap are measured without a size-matched benchmark.

Ferroptosis, pyroptosis and cuproptosis have each been converted into transcriptomic indices that separate ccRCC patients by survival^{6–8}. Most share three weaknesses: the gene set is assembled by intersecting a literature-defined pathway with a differential-expression list, making the panel a description of the literature rather than of the tumour; the index often turns out to be a proliferation or inflammatory score in disguise; and the model is sized to the panel rather than to the number of events, so the reported discrimination is optimistic by an unquantified amount. Anchoring the gene set to a mechanistically defined, experimentally perturbed death pathway removes the first weakness, not the other two.

Mitoxyperilysis is such a pathway: combining innate immune stimulation with metabolic disruption — lipopolysaccharide plus carbon starvation — kills macrophages through a route independent of caspases, pyroptosis, PANoptosis, necroptosis, ferroptosis and oxeiptosis, with execution running from BAX/BAK1/BID-dependent mitochondrial oxidative stress to prolonged mitochondria–plasma membrane contact and then to lysis, and mTORC2 as the permissive switch^{9,10}. A mitoxyperilysis-related signature is already published in colorectal cancer¹¹, which establishes that such a gene set can be assembled from bulk transcriptomes but not that the pathway operates in tumour cells, and means we cannot claim priority. ccRCC is a well-motivated setting, though the argument is a rationale rather than an established mechanism here: both inputs the pathway requires are near-constitutive, VHL loss supplying the metabolic-disruption arm^{2,3} and the macrophage- and T-cell-rich infiltrate the innate immune arm^{5,12}, and its permissive node is the target of an approved drug⁴. Whether those two arms coincide in the same cell at the same time is not established here.

We therefore derive an index from the full 87 genes and test it against adjustment, a proliferation benchmark and internal optimism, then treat it as a hypothesis to be compressed rather than defended — comparing the full and compressed models under three nested estimates of optimism, against randomised panels of equal size, and, newly here, on reclassification, decision curves and a size-matched null for those metrics. **Because the panel does not contain the pathway's execution arm, nothing here tests whether mitoxyperilysis operates in ccRCC, and no such claim is made.**

2. Materials and methods

Procedural detail — excluded-patient accounting, complete gene lists, and the provenance and probe mapping of every external cohort — is in the Supplementary Methods (SM1–SM8).

2.1 Data and cohort assembly

Expression, clinical and somatic mutation data for TCGA Pan-Cancer Atlas cohort *kirc_tcga_pan_can_atlas_2018* were obtained through the cBioPortal API^{13,14} in October 2026, with the parallel GDC re-derivation (*ccrcc_tcga_gdc*), expression being the RNA-Seq V2 median all-sample *z*-score profile. Inclusion required a primary tumour sample, non-missing overall survival and non-missing expression for every panel gene; follow-up was estimated by reverse Kaplan–Meier¹⁵. The inherited pipeline drew its sample set from the study's `_sequenced` list, which enumerates samples with mutation data rather than the RNA-seq sample list; removing that restriction yields **508 patients with 168 deaths (33.1%)**, recovering **110 evaluable patients (21.7%)**, all lacking mutation data and all more advanced than those retained (stage III–IV 46.4% vs 39.2%; death rate 42.7% vs 30.4%). Covariates were age, sex, stage (I–II vs III–IV) and grade (G1–G2 vs G3–G4), complete for all 508 patients.

2.2 The MRG panel

The panel derives from GSE235046, in which mouse macrophages were exposed to lipopolysaccharide plus carbon starvation with or without the mTOR inhibitor Torin; a gene entered only if it responded to dual stimulation and the

response was reversed by Torin. "MRG" denotes membership of that set and nothing more: the screen produced **101 mouse genes**, ortholog mapping by symbol left **92**, and **87 were present and non-missing** in the TCGA-KIRC matrix. The framework's description of the panel as "91 genes" containing 15 mechanistic anchors is wrong on both counts — 91 exists at no point on the chain, and of twelve named anchors only TXNRD1, OSGIN2 and SLC7A1 are present.

2.3 Models and the parsimony profile

Genes were standardised within the cohort. Univariate Cox models were fitted for all 87 genes with Benjamini–Hochberg control¹⁶, and the screen repeated with stage III–IV and grade G3–G4 forced into every gene's model. Model A used the 25 genes with the smallest univariate *P* values under an L2 penalty of $\lambda = 1.0$, chosen to keep a 25-term fit numerically stable at 168 events; patients were split at the median of the linear predictor and discrimination is Harrell's C-index¹⁷. Model B used a LASSO-Cox path over 80 penalty values^{18,19}; rather than choose a size by eye, we took, for every distinct number of non-zero coefficients, the best 10-fold cross-validated C-index at that size — a parsimony profile. The pre-specified 1-SE rule (profile maximum minus one standard error defines a band; the smallest model inside it is preferred¹⁹) gives the three-gene model carried forward. LASSO coefficients were used only to select genes; retained genes were refitted unpenalised, and those are the frozen weights. Stability selection repeated the whole selection in 200 bootstrap resamples, recording each gene's frequency π against a pre-specified $\pi \geq 0.75$ ²⁰.

2.4 Adjustment and molecular analyses

Both models were entered into a multivariable Cox model with age, sex, stage III–IV and grade G3–G4, and the score's contribution over the clinical variables alone was tested by a likelihood-ratio test on one degree of freedom; no stepwise selection was performed. A 20-gene proliferation score was the mean within-cohort *z*-score of its members; patients were clustered on the 25 Model A genes by k-means. Immune infiltration was estimated by a marker-gene mean *z*-score across 11 features — a proxy, not a deconvolution — and reported three ways per feature: the unadjusted Mann–Whitney test, the same test within each stage stratum, and the risk-group term of a linear model containing stage III–IV, age and sex with HC3 standard errors. Mutation calls for 14 recurrent ccRCC driver genes were retrieved over the sequenced sample list, non-synonymous only, and compared between risk groups by two-sided Fisher exact test with Benjamini–Hochberg control (**278** patients). Tumour purity was the TCGA Pan-Cancer Atlas ABSOLUTE consensus estimate, matched by barcode and added to the adjusted Cox model for each score under a nested likelihood-ratio test.

2.5 External validation

Four evaluations were run, and their order matters: three were pre-specified (GSE22541²¹, GSE29609, the GDC re-derivation), and the fourth, E-MTAB-1980, was added after all three had been executed, after the pre-specified cohort had been rejected on design grounds — it mixes pulmonary metastases with primary tumours and carries disease-free survival rather than overall survival — and after the GEO cohort had inverted. E-MTAB-1980 is the Sato et al. ccRCC series²², expression from the ArrayExpress matrix joined on sample label to survival from cBioPortal study *ccrcc_utokyo_2013*; GSE29609 was downloaded with its GPL1708 annotation, probes mapped to symbols²³; the GDC re-derivation re-processes the same patients²⁴ and is labelled a technical replication throughout. All three frozen models were applied to each usable cohort under cohort-internal standardisation with **no coefficient refitted and no gene reselected**.

2.6 Internal validation, benchmarks and clinical utility

Three estimates of optimism are reported in a ladder: **apparent; coefficients refitted inside folds** (gene set fixed, coefficients re-estimated in each training fold); and **whole procedure repeated inside folds** (the LASSO path, parsimony profile, 1-SE rule and refit all re-run on each training part). The third was produced for **both** pipelines — for Model A, its own univariate screen, $P < 0.05$ filter, top-25 cut and L2 fit — and both were evaluated on the same ten folds with a fixed seed. Discrimination was additionally summarised by out-of-fold time-dependent AUC

at 12, 36 and 60 months, and calibration assessed at 36 months by mapping each frozen score to absolute risk (SM6). Each model was benchmarked against its size-matched random null under the same protocol (500 draws each), the 25-gene null also redrawn **unpenalised** to separate shrinkage from panel size; because a null at 0.676 leaves less room above it than one at 0.599, the excess is also reported normalised — divided by the discrimination still available above the null mean. Clinical utility was assessed at 36 months by comparing a clinical model containing age, sex, stage III–IV and grade G3–G4 against the same model with each score added as a single standardised covariate. Category-free (continuous) NRI was computed from each patient's change in predicted 36-month mortality risk under inverse-probability-of-censoring weighting, with the censoring distribution estimated by reverse Kaplan–Meier; no categorical risk strata were pre-specified, and the IDI is the difference in discrimination slope. Net benefit used the decision-curve formulation for survival data, the event probability above each threshold estimated by Kaplan–Meier at the horizon, over thresholds of 0.05 to 0.60 in steps of 0.01 (SM8). Because NRI and IDI rise whenever any variable is added, every metric was also computed for 500 random panels handled exactly as the score is, the k genes entered individually and freely fitted; confidence intervals came from 1,000 patient-level bootstrap resamples in which both models were refitted at every draw.

2.7 Software

Analyses used Python 3.13 with lifelines 0.30.3, scikit-survival 0.28.0, scikit-learn and matplotlib, seeds fixed for every resampling procedure. TIDE, IMvigor210 and the GEPIA2 tumour-versus-normal panel²⁵ were **not** run, for the reasons given in SM7; immunotherapy-response prediction is therefore not claimed anywhere here. Every quantity was written to a machine-readable file before drafting, and the deposited per-patient tables are the reference against which the text was checked.

3. Results

3.1 The corrected cohort carries a dense, not sparse, survival signal

The corrected TCGA-KIRC cohort contains **508 patients with 168 deaths (33.1%)**: median age 60.0 years, 325 (64.0%) male, stage I/II/III/IV 247/54/124/83 and grade G1/G2/G3/G4 12/214/199/75; median follow-up by reverse Kaplan–Meier was **57.0 months** (Table 1). All 87 panel genes were evaluable in all 508 patients, and univariate Cox regression returned **45 of the 87** at $P < 0.05$ and **44 at $q < 0.05$** (Figure 1B), the strongest being MBOAT7 (HR 1.409 per SD, 95% CI 1.296–1.533; $q = 1.12 \times 10^{-13}$), BICRA, SPTBN1 (0.521, protective) and BMP1. Because a dense signal invites the objection that it is stage and grade in disguise, the screen was repeated with both forced into every gene's model: **32 genes remain at $P < 0.05$ and 27 at $q < 0.05$** , and the genes in the final score are the least affected (MBOAT7 $P = 8.8 \times 10^{-8}$; SPTBN1 8.8×10^{-7} ; BMP1 5.9×10^{-5}). That density organises the rest of the paper: an index can be built from almost any reasonable subset of this panel, so a high C-index says little about whether the model is any good.

3.2 Model A discriminates without any coefficient being defensible

Model A separated the cohort cleanly (apparent C-index **0.711**; median-split log-rank $P = 1.61 \times 10^{-11}$; high-risk hazard ratio 2.998, 95% CI 2.144–4.193; Figure 1C–D), but three properties argue against the fit: **none of the 25 coefficients is individually significant** (smallest $P = 0.108$), the events-per-fitted-term ratio is **6.72**, below the conventional threshold of ten^{26,27}, and the coefficients are small and tightly clustered. All 25 signs agree with their univariate estimates, which is why the model was taken forward as the comparator rather than discarded.

3.3 The parsimony profile and the three-gene score

Discrimination along the LASSO-Cox path rises from two genes (0.668) to three (0.686) and peaks at **four genes (C = 0.688)**, then declines monotonically — 0.681 at 25 genes and 0.623 at the unpenalised end (Figure 2A). This is not the usual flat curve: larger models are measurably worse. Under the 1-SE rule the band threshold is **0.6806**

and the smallest model inside the band is three genes — MBOAT7, SPTBN1 and BMP1, the top three of the univariate screen; the four-gene model adds BICRA, whose coefficient is not significant ($P = 0.307$), and differs in no conclusion.

The score is $0.1954 \times z(\text{MBOAT7}) - 0.4432 \times z(\text{SPTBN1}) + 0.2182 \times z(\text{BMP1})$ (Table 2). It reached an **apparent C-index of 0.697**, a 10-fold cross-validated C-index of **0.705 ± 0.070** and a hazard ratio per SD of 1.939 (95% CI 1.686–2.231; $P = 1.72 \times 10^{-20}$); its events-per-fitted-term ratio is **56.0** against 6.72 for the 25-gene fit, and all three coefficients are individually significant. **Independence.** Adjusted for age, sex, stage III–IV and grade G3–G4, the score retained a hazard ratio of **1.609** per SD (95% CI **1.381**–1.875; $P = 1.02 \times 10^{-9}$); age and stage III–IV were also independent, grade only narrowly (1.461, 1.003–2.128) and sex not at all, and adding the score raised the C-index from **0.750** to **0.775** (likelihood-ratio $\chi^2 = 35.58$ on one degree of freedom; $P = 2.45 \times 10^{-9}$). Stage alone reached 0.701 and grade 0.629, so the score alone (0.697) sits just below stage. Model A behaved almost identically (HR 1.627; C-index **0.782**), with one difference worth recording: grade is significant in the parsimonious models' adjusted fits ($P = 0.048$) and not in Model A's (0.117), which is what happens when 22 extra terms absorb part of the grade signal. The score is proliferation-adjacent but not a proliferation proxy: Model A correlated with a 20-gene proliferation score at Spearman $\rho = 0.393$ and the three-gene score at $\rho = 0.341$, proliferation alone reached only 0.610, and in joint models the 25-gene index dominated (HR 1.835) while proliferation did not reach significance ($P = 0.085$) — whereas the three-gene score kept its own strong effect (1.771) without absorbing proliferation ($P = 2.2 \times 10^{-4}$).

3.4 Internal validation and the randomised-panel benchmark

Apparent 0.711 (Model A), 0.697 (four-gene), **0.697** (three-gene) — the larger model appears to lead by 0.014. With **coefficients refitted inside folds** the three are 0.706 ± 0.061 , 0.700 ± 0.065 and **0.705 ± 0.070** : the apparent lead of 0.014 becomes 0.001, because a penalised fit whose coefficients are all non-significant is largely noise, and noise is not reproducible out of fold. Repeating the **whole procedure inside folds** gives **0.682 ± 0.065** , with a median selected size of 3 genes and six of ten folds returning exactly {MBOAT7, SPTBN1, BMP1}. Measured symmetrically at the same level, **the two pipelines returned 0.696 and 0.682**, a nominal difference of **+0.0140** that is not statistically distinguishable (paired t $P = 0.281$). All three models were calibrated at 36 months (predicted mortality 0.238 and 0.228 against **0.239** observed; Figure 4F), but **one result does not favour the parsimonious model**: out-of-fold time-dependent AUC at 12, 36 and 60 months was 0.758, 0.743 and 0.767 for Model A against 0.750, 0.725 and 0.729 for the three-gene score (Figure 4C). Time-dependent AUC weights the whole risk distribution rather than only its order, so this is a genuine advantage for the larger model, and it is why the claim is equivalence rather than superiority.

A C-index of 0.705 means nothing without knowing what an arbitrary model of the same size achieves in the same cohort (Figure 4D). Of 500 random **three-gene** panels, all 500 were fitted and evaluated: null mean **0.599** (SD **0.046**), maximum 0.697, and the observed 0.705 matched or beaten by **0 of 500**. Of 500 random **25-gene** panels fitted at Model A's penalty, all 500 were valid: null mean **0.676** (SD **0.014**), matched or beaten by **2 of 500**; refitting the same panels unpenalised gives a null mean of **0.658** (SD **0.016**), so the penalty raises the random-25 null by about one null standard deviation, and the observed 25-gene model is an excess of +0.048 over that null that no random panel reached. Comparing the two excesses on the same scale — as a fraction of the discrimination still available above the null mean — the three-gene score recovers **26.4%** of that room and the 25-gene score **9.3%**, or **14.0%** against the unpenalised null. The reading is not that the three-gene score is better — the paired bootstrap says it is not distinguishable — but that **the extra 22 genes add nothing that the extra panel size does not already provide**.

3.5 Clinical utility at 36 months: a discrimination gain and an IDI beyond a random panel, and an NRI that is not

By 36 months 106 of the 508 patients had died, 280 were alive with follow-up past the horizon and 122 were censored before it, so every metric carries inverse-probability-of-censoring weights (censoring survival at 36

months 0.725). Adding the three-gene score to the clinical model gave ΔC **+0.026** (bootstrap 95% CI +0.008 to +0.049), **IDI +0.041** (+0.006 to +0.087) and **NRI +0.34** (+0.08 to +0.61), with an unweighted NRI of +0.33 in the 386 patients whose 36-month status is known; Model A on the same footing gave ΔC +0.032 (+0.012 to +0.057), IDI +0.046 (+0.011 to +0.089) and NRI +0.41 (+0.12 to +0.69) (Supplementary Tables S24, S28). A random three-gene panel entered as three freely fitted terms in the same clinical model reached a mean NRI of **+0.207** (SD 0.128) and a mean IDI of **+0.017** (SD 0.011); the score's ΔC and IDI lie beyond that null — at the 99.8th and 95th percentiles of 500 draws — but **its NRI does not**, the observed +0.343 sitting at the 84th percentile, and a reclassification index that one in six arbitrary three-gene panels matches is not evidence of anything (Supplementary Tables S26, S28). We report the NRI and do not claim it. The benchmark bears on panel size in the opposite direction: fitting the 25 Model A genes freely placed them at the 68th percentile for NRI and the 62nd for IDI of the random-25 null, so on reclassification the 25 selected genes are no better than 25 arbitrary ones. Net benefit curves tell the same story with one further qualification: the score's curve lay above the clinical model's at 43 of 56 thresholds between 0.05 and 0.56, and above treat-all between 0.06 and 0.60 (Figure 3E), but with both models refitted in each of 1,000 bootstrap draws the mean gain was +0.003 at $p_t = 0.30$ and +0.007 at 0.40 and every interval included zero. The utility picture is therefore discrimination and IDI beyond a size-matched null, decision curves above the clinical model across most of the plausible threshold range, and a net-benefit increment this cohort cannot resolve.

3.6 Stability selection, molecular correlates and tumour purity

Stability selection succeeded in all 200 resamples (Figure 2B) but only **two genes reached $\pi \geq 0.75$** (MBOAT7 0.84; SPTBN1 0.82); BMP1 followed at 0.64 and no fourth gene exceeded 0.40. This is a negative result about the panel rather than about the model — the third gene is admitted instead because its coefficient is significant in the frozen multivariable fit — and it says that **this panel has no sparse solution to find**. Among **278 sequenced patients**, the two most frequent driver mutations did **not** differ between risk groups (VHL **56.8% vs 61.9%**; PBRM1 **46.8% vs 54.7%**) while **SETD2 (24.5% vs 10.1%, $q = 0.011$)** and **BAP1 (19.4% vs 7.9%, $q = 0.027$)** did (Figure 3D); both are chromatin-modifying tumour suppressors whose loss defines the worst-prognosis subgroup of the TCGA ccRCC taxonomy², and SETD2 reproduced under the three-gene score's own split. The VHL negative is the useful one: since VHL loss initiates essentially all ccRCC, a score tracking it would be tracking the disease rather than its aggressiveness.

Eight of 11 immune features were significantly **higher** in the Model A high-risk group, including regulatory T cells (Δ mean $z = +0.505$), CD8⁺ T cells (+0.460) and dendritic cells (+0.381), with the largest adjusted P being 2.1×10^{-3} ; M1 macrophages, the interferon signature and M2 macrophages did not differ (Figure 6A). Two readings are available and we do not resolve them: high-risk ccRCC may be an *inflamed but not rejecting* tumour, the phenotype immune-exclusion signatures were built to capture^{28,29}, or a marker-gene mean z -score may track a tumour-purity difference, since the high-risk group carries stage III–IV in **55.9%** of tumours against **25.6%**. Both confounders we could test were tested and neither explains it: all eight differences survive adjustment for stage, age and sex and hold within each stage stratum, and among the **479** patients with ABSOLUTE purity estimates purity was unassociated with survival once the score was present (HR 1.054; $P = 0.52$) and moved the three-gene hazard ratio only from 1.631 to 1.637. Because ABSOLUTE is genomic rather than transcriptomic, the correct conclusion is the weaker one. k-means clustering on the 25 Model A genes gave silhouettes of 0.239, 0.132 and 0.080 at $k = 2, 3$ and 4, which does not support describing a robust molecular classification.

3.7 External validation: the score held in one cohort and inverted in the other

In the Sato et al. series²² — **101 patients, 23 deaths** — the frozen coefficients produced $C =$ **0.773** for the 25-gene model and **0.688** for the three-gene score (Figure 5A), every coefficient coming out positive as fitted, the three-gene score carrying a per-SD hazard ratio of **1.70** (95% CI **1.17–2.46**) and the 25-gene index **2.20** (1.57–3.09). No coefficient was refitted and no gene reselected; 24 of 25 model genes could be mapped, and median overall survival was not reached in either risk group. The **GDC re-derivation** of the same cohort (368 patients, 94 deaths)

reproduced all three models ($C = 0.722, 0.709, 0.711$), establishing that the fitted quantities are not artefacts of one processing pipeline and unable to establish transportability, because the patients are the same patients. The same per-gene statistic computed on the **same 23 genes** in both independent cohorts returns opposite answers (Figure 5B): **19 of 23 genes (82.6%)** agree in direction in E-MTAB-1980 ($r = +0.7155$) against 5 of 23 (21.7%) in GSE29609 ($r = -0.6328$), where in the 39 patients with survival data and 17 deaths the frozen coefficients produced $C = 0.300$ for the 25-gene model and **0.351** for the three-gene score. In that matrix **35 of 44** panel and proliferation genes correlated negatively with Fuhrman grade, including CCNB2 (-0.752), ASPM (-0.701) and CDKN3 (-0.651) — genes for which a positive correlation with grade is not a matter of debate in ccRCC — so the corruption is specific to the proliferative compartment, which is the one our score draws on (Figure 5C–D). We therefore attribute the inversion to the deposited GSE29609 matrix rather than to genuine inverse biology, and deposit the full per-gene direction table so a reader can judge the attribution rather than take it. Two limits remain: the attribution is not falsifiable from within GSE29609 itself, and the external evidence here is one positive cohort and one negative.

3.8 What the three genes are, and what they are not

None of the three genes is a mitoxyperilysis effector. **MBOAT7** (risk-associated, HR 1.216) encodes a lysophospholipid acyltransferase that remodels phosphatidylinositol, with no assigned role in mitochondrial membrane lysis. **SPTBN1** (protective, 0.642) encodes β II-spectrin, the principal component of the membrane skeleton; it is the only member with a direct structural connection to the terminal event the pathway describes, and it points the opposite way to the pathway's logic, since more membrane skeleton is associated with better survival. **BMP1** (risk-associated, 1.244) encodes a secreted metalloprotease that processes procollagen — a matrix-remodelling gene rather than a cell-death gene. Two observations follow, and neither is in the paper's favour: the panel **does not contain the pathway's execution machinery at all**, so a score built from it cannot be a readout of that arm; and the three retained genes read more naturally as a **matrix-remodelling, membrane-turnover tumour state** than as a death-execution programme.

4. Discussion

We set out to test whether a mitoxyperilysis-related gene set carries prognostic information in ccRCC, and then to test the resulting models harder than cell-death signatures usually are. The panel carries information abundantly: 44 of 87 genes survived false-discovery control and 27 still did once stage and grade were forced into every gene's model. What we did not expect, and what became the substance of the paper, is that the information is so evenly distributed that model size stops mattering: the paired difference between the 25-gene and three-gene models was -0.0137 (95% CI -0.0353 to 0.0067), and at the whole-procedure level the 25-gene pipeline returned 0.696 against 0.682. Twenty-two extra coefficients bought no measurable discrimination in either direction; they did buy a large amount of unverifiable fitting — not one of the 25 coefficients individually significant, and an events-per-fitted-term ratio of 6.72 against 56.0.

The randomised-panel benchmark explains why, and this is the result we would most want a reader to take away. An arbitrary 25-gene panel drawn from the same 87 genes and penalised the same way reached 0.676 against 0.706 for our 25-gene model, an excess of $+0.030$, and the unpenalised redraw drops that null to 0.658; an arbitrary three-gene panel reached only 0.599 while our three-gene score reached 0.705, an excess of $+0.106$ with none of 500 random panels matching it. Panel size here is therefore not merely unhelpful — it is *self-concealing*: a large arbitrary panel scores well for reasons that have nothing to do with the genes in it, and a C-index of 0.7 looks like a success whether it was earned or assembled. The same instrument has to be applied to the clinical-utility metrics, because NRI and IDI rise whenever any variable is added and a reclassification result quoted without a size-matched reference is uninterpretable; here the reference separates the metrics cleanly — ΔC and IDI survive, NRI does not — and we report that asymmetry rather than the favourable half of it. The growth in discrimination over

clinical variables remains small, and the decision-curve increment, though consistent across most of the plausible threshold range, has a confidence interval that includes zero at every threshold examined.

The external evidence is one positive result and one negative one, and the pair is more informative than either alone. In the Sato et al. series the frozen weights validated without being touched — $C = 0.773$ and 0.688 , every coefficient at the correct sign, a hazard ratio of 1.70 per SD — which is a statement about transportability, not about fitting; in GSE29609 the same weights inverted, but on the identical 23-gene set E-MTAB-1980 reproduces the discovery cohort's per-gene direction at $r = +0.72$ with 19 of 23 genes concordant while GSE29609 reverses it at $r = -0.63$ with 5 of 23. That does not dispose of the negative result — it stays in the abstract rather than being demoted to a footnote — because one validating cohort is one cohort. The mutation result, by contrast, is the most concrete corroboration we have: after multiplicity control the high-risk group was enriched for SETD2 and BAP1, while VHL and PBRM1 showed no difference, which is what a score tracking aggressiveness rather than disease would do. The biological reading of the score is nonetheless narrower than the framework that motivated the study, and we state the gap rather than blur it: the panel is described in the project's own framework as a mechanism-anchored 91-gene set containing the pathway's execution arm, and it is neither. What it does contain, overwhelmingly, are matrix and cytoskeletal, chromatin and transcription, and lipid-metabolism genes, and the three selected genes fit that profile — the score detects an aggressive ccRCC state characterised by membrane and matrix turnover, proliferation and immune infiltration. The methodological lesson transfers: internal cross-validation could not separate a 25-gene model from a three-gene model (0.706 vs 0.705), neither could the apparent C-index once its optimism was removed, and neither could the procedure-level estimate (0.696 vs 0.682); measured alone, the parsimony pipeline's 0.682 would have been read as a fall from 0.705 . What separated the models was the benchmark against size-matched randomised panels — the only instrument that asks not whether a model predicts but whether it predicts *better than this much panel*.

Limitations. The external evidence is one positive cohort and one negative one, neither a replication, and if our attribution of the GSE29609 inversion to a matrix defect is wrong the score has an inverse association there; the GDC analysis re-processes the same patients and cannot test generalisation; GSE22541, the pre-specified cohort, was not used and the validating cohort was added afterwards, which is an addition to the protocol rather than a fulfilment of it; the parsimony claim is equivalence, not superiority, time-dependent AUC favouring the larger model at all three horizons; the third gene is not stability-selected; clinical utility is assessed at 36 months only, the NRI does not clear its size-matched null, the net-benefit gain includes zero at every threshold, and nothing here establishes that the score changes a treatment decision; the immune comparison is tested against stage and genomic tumour purity but not against a transcriptomic purity estimate or deconvolution; the panel is a mouse selection history holding no execution-arm member, and the expression matrix is a cBioPortal z-score matrix rather than raw counts.

Conclusion

A three-gene expression score derived from a mitoxyperilysis-related gene set — MBOAT7, SPTBN1 and BMP1 — is independently associated with overall survival in clear cell renal cell carcinoma (HR 1.609 per SD after adjustment for age, sex, stage and grade; $P = 1.02 \times 10^{-9}$) and is not a proliferation proxy. Twenty-two further genes from the same panel add no measurable discrimination, and the reason is structural: the survival signal is dense, so an arbitrary 25-gene panel already reaches a cross-validated C-index of 0.676 , and the extra genes buy $+0.030$ where the three retained genes buy $+0.106$ over their own size-matched null. The larger model was unfittable in any defensible sense — 25 terms, 168 events, not one significant coefficient. The score marks a proliferation-adjacent, matrix-remodelling state enriched for SETD2 and BAP1 mutation and unseparated by VHL status, which is what one expects of an aggressiveness index rather than a disease index. It is not a readout of the mitoxyperilysis pathway: the panel does not contain the pathway's execution machinery. External transportability is demonstrated in one independent cohort and contradicted in another, which fails the very per-gene diagnostic the validating cohort passes. Clinical utility is measured and modest: a 36-month C-index increment of $+0.026$ and an

IDI beyond a size-matched null, with decision curves above the clinical model but a net-benefit gain whose interval includes zero. The contribution we would defend is the demonstration that in a dense-signal panel size can be cut five-fold at no measurable cost, and that the instrument revealing this is not the cross-validated C-index but the benchmark against size-matched randomised panels.

Declarations

Ethics approval and consent to participate. Not applicable. Only publicly available, de-identified data were analysed: TCGA-KIRC through cBioPortal and its GDC re-derivation, GEO accessions GSE29609 and GSE22541, the ArrayExpress matrix E-MTAB-1980 joined to the cBioPortal record for study `ccrcc_utokyo_2013`, and the TCGA Pan-Cancer Atlas ABSOLUTE tumour-purity calls. No new human or animal material was generated, no controlled-access tier was used, no re-identification was attempted, and no individual-level consent was required. The mouse experiment from which the panel was derived (GSE235046) was performed in the parent study and is cited rather than re-reported here.

Consent for publication. Not applicable.

Availability of data and materials. All primary data are public. TCGA-KIRC expression, clinical and mutation data were obtained from cBioPortal (study `kirc_tcga_pan_can_atlas_2018`) in October 2026, with the GDC re-derivation from study `ccrcc_tcga_gdc`; GSE29609 and GSE22541 from the NCBI Gene Expression Omnibus; and the tumour-purity estimates from the TCGA Pan-Cancer Atlas ABSOLUTE master calls. The validation cohort joins the ArrayExpress matrix **E-MTAB-1980** to the cBioPortal survival record **ccrcc_utokyo_2013** on sample label (`ccRCC-{i} ↔ ccrcc_{i}`), because that study carries mutations but no mRNA profile. Derived tables and the analysis code are archived at Zenodo under CC BY 4.0 (tables) and the MIT licence (code), DOI 10.5281/zenodo.23141875, where the record is versioned; every table behind this manuscript is listed in the Supplementary material index below.

Competing interests. The authors declare no financial or non-financial competing interests, including business or family interests.

Funding. No specific grant was received from any funding agency in the public, commercial or not-for-profit sectors.

Author contributions (CRediT). Bin Lian: conceptualisation; methodology; software; formal analysis; data curation; investigation; visualisation; writing — original draft; writing — review and editing; project administration; guarantor of the analysis. Huiling Zhai: investigation; validation; writing — review and editing. Jiayi Kuang: data curation; validation; writing — review and editing. All authors have read and approved the final manuscript.

Acknowledgements. Not applicable.

Declaration of generative AI and AI-assisted technologies in the writing process. An AI-assisted tool was used for language editing and to improve readability and concision. The authors reviewed and edited the content afterwards and take full responsibility for the publication.

Declaration on figure provenance. All figures were generated as vector graphics directly from the analysed data using Matplotlib; no generative AI image models were used and no third-party images reproduced.

Figure legends

Figure 1. Cohort assembly, the univariate screen and the 25-gene full-panel index (Model A). (A) Assembly of the corrected discovery cohort: rebuilding the matrix from the RNA-seq V2 profile rather than the `_sequenced` sample list moved the cohort from 398 patients with 121 deaths to 508 with 168, recovering 110 evaluable patients (21.7%). (B) Univariate Cox regression of all 87 evaluable genes, ordered by unadjusted P ; hazard ratios are per 1 SD of standardised expression, and red points denote Benjamini–Hochberg $q < 0.05$. (C) Distribution of the Model A linear predictor across the 508 patients, with the median split marked. (D) Kaplan–Meier overall survival by Model A median split (254 per group); median overall survival 53.9 months (high risk) versus not reached (low risk).

Figure 2. The parsimony profile, stability selection and the three-gene score. (A) Best 10-fold cross-validated C-index at each distinct model size along the LASSO-Cox path. The optimum is four genes; the 1-SE band threshold is 0.6806 (dashed line) and the smallest model inside the band is three genes. (B) Selection frequency π per gene over 200 bootstrap resamples; red bars exceed the pre-specified $\pi \geq 0.75$ threshold (dashed line). Only MBOAT7 (0.84) and SPTBN1 (0.82) clear it. (C) Frozen coefficients, refitted unpenalised: MBOAT7 +0.1954 (HR 1.216, 95% CI 1.085–1.362), SPTBN1 –0.4432 (HR 0.642, 0.524–0.786), BMP1 +0.2182 (HR 1.244, 1.097–1.410). (D) Kaplan–Meier overall survival by three-gene score median split (254 per group).

Figure 3. Clinical independence, proliferation adjacency, adjusted effect sizes, mutation landscape and 36-month decision curves. (A) Multivariable Cox model containing the three-gene score with age, sex, stage III–IV and grade G3–G4 ($n = 508$; 168 events); forest plot of hazard ratios (95% CI, log scale), red where the interval excludes 1. (B) Model A score against the 20-gene proliferation score; the proliferation coefficient was not significant in a joint model. (C) Score per SD after adjustment for all three models. (D) Mutation frequency of 14 recurrent ccRCC driver genes in the 278 sequenced patients, split at the median of the Model A score (139 per group); SETD2 (24.5% vs 10.1%; $q = 0.011$) and BAP1 (19.4% vs 7.9%; $q = 0.027$) survive multiplicity control while VHL (56.8% vs 61.9%) does not differ. (E) Decision curves for 36-month mortality: net benefit against threshold probability for treat-all (grey), treat-none (zero line), the clinical model (blue) and the clinical model plus the three-gene score (red). The score's curve is above the clinical model's at 43 of 56 thresholds between 0.05 and 0.56 and above treat-all between 0.06 and 0.60; the vertical axis is truncated, and treat-all leaves the frame above $p_t = 0.34$.

Figure 4. Internal validation: three nested estimates of optimism, fold-level model size, the randomised-panel benchmark, the GDC replication and 36-month calibration. (A) Apparent, coefficients-refitted-inside-folds and whole-procedure-repeated-inside-folds estimates of optimism. Model A 0.711 $\rightarrow$ 0.706 $\rightarrow$ 0.696; three-gene 0.697 $\rightarrow$ 0.705 $\rightarrow$ 0.682. (B) Genes selected inside each of the ten outer folds: 9, 3, 3, 3, 17, 4, 3, 4, 3, 3 (median 3). (C) Out-of-fold time-dependent AUC at 12, 36 and 60 months: Model A 0.758 / 0.743 / 0.767, four-gene 0.741 / 0.720 / 0.732, three-gene 0.750 / 0.725 / 0.729. (D) Cross-validated C-index of 500 random panels at $k = 3$ (null mean 0.599), $k = 25$ penalised as Model A (0.676) and $k = 25$ unpenalised (0.658). (E) GDC re-processing of the same cohort (368 patients, 94 deaths): $C = 0.722, 0.709$ and 0.711 . (F) Calibration at 36 months: Kaplan–Meier observed mortality against the absolute risk predicted by each frozen model, by quintile of predicted risk ($n = 508$, 168 deaths); the dashed line is perfect calibration.

Figure 5. External validation in two independent cohorts, and the matched diagnostic that separates a matrix defect from a biological contradiction. (A) Kaplan–Meier overall survival in the Sato et al. series (E-MTAB-1980 expression joined to the `ccrcc_utokyo_2013` survival record; 101 patients, 23 deaths, median follow-up 67.0 months) by the frozen three-gene score at the cohort-internal median: C-index = 0.688, log-rank $P = 5.11 \times 10^{-3}$, Cox HR 1.70 per SD. (B) Matched per-gene diagnostic on the same 23 genes in both cohorts: each gene's external-cohort univariate Cox coefficient against its discovery-cohort value. E-MTAB-1980 (red) agrees for 19 of 23 genes (82.6%); GSE29609 (grey) for 5 of 23 (21.7%). (C) GSE29609 scale diagnostic: 35 of the 44 mappable panel and proliferation genes correlated *negatively* with Fuhrman grade; the 20 most extreme are shown.

(D) Kaplan–Meier overall survival in GSE29609 (39 patients with survival data, 17 deaths) by the same frozen score: it inverts ($C = 0.351$).

Figure 6. Immune features, tumour purity and molecular subtyping. (A) Marker-gene mean z -scores for 11 immune features in the Model A high- and low-risk groups (254 per group); bars are the mean difference (high minus low risk), and labels give the raw Mann–Whitney P and the stage-adjusted P from a linear model with HC3 standard errors. Eight features were higher by both routes; M1 macrophages (+0.067) and the interferon signature (+0.027) were null on both. M2 macrophages are hatched because they are significant only in the adjusted mean model ($P = 0.037$), not on the rank test ($P = 0.915$). (B) Tumour purity (ABSOLUTE, 479 of 508 patients) by risk group (medians 0.55 vs 0.56; $P = 0.20$), and the score's hazard ratio before and after adding purity (1.631 $\rightarrow$ 1.637). (C) k-means solutions for $k = 2, 3$ and 4: silhouettes 0.239, 0.132 and 0.080; between-cluster log-rank P of 6.91×10^{-17} , 3.77×10^{-19} and 3.55×10^{-16} .

Tables

Table 1. Baseline characteristics of the corrected TCGA-KIRC discovery cohort.

Characteristic	All analysed (n = 508)	Retained (n = 398)	Recovered (n = 110)
Deaths, n (%)	168 (33.1%)	121 (30.4%)	47 (42.7%)
Median follow-up, months (reverse Kaplan–Meier)	57.0	39.3 (naïve)	—
Median age, years	60.0	—	—
Male sex, n (%)	325 (64.0%)	—	—
Stage I / II / III / IV, n	247 / 54 / 124 / 83	—	—
Stage III–IV, n (%)	207 (40.7%)	156 (39.2%)	51 (46.4%)
Grade G1 / G2 / G3 / G4, n	12 / 214 / 199 / 75	—	—
Grade G3–G4, n (%)	274 (53.9%)	216 (54.3%)	58 (52.7%)
Panel genes evaluable / mutation data available, n	87 / 278	87 / 278	87 / 0
ABSOLUTE tumour purity available, n	479	450	29 unmatched
Median tumour purity (IQR)	0.560 (0.440–0.650)	—	—
Median overall survival, months (Model A split)	not reached (low) / 53.9 (high)	—	—

Table 2. The frozen three-gene score and the performance of the three models (n = 508, 168 deaths). Panel A: frozen weights, refitted unpenalised, applied to within-cohort standardised expression. Panel B: discrimination, events per fitted term (EPV), and the hazard ratio per SD before and after adjustment for age, sex, stage III–IV and grade G3–G4. The Model A cross-validated value is the coefficients-refitted-in-folds estimate, its gene set being fixed by an *a priori* rule.

Panel A — gene	Frozen β	HR (95% CI)	P	Panel B — model	Genes	Apparent C	Cross-validated C	EPV	Adjusted HR (95% CI)	Adjusted C
MBOAT7	+0.1954	1.216 (1.085–1.362)	7.49×10^{-4}	Model A (L2, $\lambda = 1.0$)	25	0.711	0.706	6.72	1.627 (1.406–1.882)	0.782
SPTBN1	−0.4432	0.642 (0.524–0.786)	1.80×10^{-5}	Parsimony optimum	4	0.697	0.700 ± 0.065	42.0	1.591 (1.371–1.846)	0.776
BMP1	+0.2182	1.244 (1.097–1.410)	6.38×10^{-4}	Parsimony 1-SE (carried forward)	3	0.697	0.705 ± 0.070	56.0	1.609 (1.381–1.875)	0.775

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2024;74(3):229-263. PMID: 38572751. doi:10.3322/caac.21834.
2. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. *Cell Rep* 2018;23(1):313-326.e5. PMID: 29617669. doi:10.1016/j.celrep.2018.03.075.
3. Kaelin WG Jr. The von Hippel-Lindau tumour suppressor protein: O₂ sensing and cancer. *Nat Rev Cancer* 2008;8(11):865-73. PMID: 18923434. doi:10.1038/nrc2502.
4. Motzer RJ, Escudier B, Oudard S, Hutson TE, Porta C, Bracarda S, et al. Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* 2008;372(9637):449-56. PMID: 18653228. doi:10.1016/S0140-6736(08)61039-9.
5. Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, et al. Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* 2015;373(19):1803-13. PMID: 26406148. doi:10.1056/NEJMoa1510665.
6. Du X, Cao H, Zhou YJ, Kong Q, Zhang X. Identification of ferroptosis-related gene signatures as a novel prognostic model for clear cell renal cell carcinoma. *Discov Oncol* 2025;16(1):456. PMID: 40178680. doi:10.1007/s12672-025-02202-1.
7. Chen J, Jiang R, Guan W, Cao Q, Tian Y, Dong K, et al. Novel model of pyroptosis-related molecular signatures for prognosis prediction of clear cell renal cell carcinoma patients. *Int J Med Sci* 2024;21(3):496-507. PMID: 38250606. doi:10.7150/ijms.88301.
8. Liu G, Li F, Ge Y, Shi Y, Ren F, Zhu L. Prognosis, Immune Microenvironment Infiltration and Immunotherapy Response in Clear Cell Renal Cell Carcinoma Based on Cuproptosis-related Immune Checkpoint Gene Signature. *J Cancer* 2023;14(17):3335-3350. PMID: 37928426. doi:10.7150/jca.88467.
9. Wang Y, Lu J, Carisey AF, Chadchan SB, Lee HW, Malireddi RKS, et al. Innate immune and metabolic signals induce mitochondria-dependent membrane lysis via mitoxyperiosis. *Cell* 2025;188(25):7155-7174.e25. PMID: 41317732. doi:10.1016/j.cell.2025.11.002.
10. Kayagaki N, Kornfeld OS, Lee BL, Stowe IB, O'Rourke K, Li Q, et al. NINJ1 mediates plasma membrane rupture during lytic cell death. *Nature* 2021;591(7848):131-136. PMID: 33472215. doi:10.1038/s41586-021-03218-7.
11. Zhang Y, Sun B, Lin J, Ding C, Zhou X, Xu X, et al. A mitoxyperilysis-related signature stratifies prognosis and identifies an aggressive colorectal cancer ecosystem with immune remodeling. *Front Cell Dev Biol* 2026;14:1851988. PMID: 42491158. doi:10.3389/fcell.2026.1851988.
12. Beuselinck B, Job S, Becht E, Karadimou A, Verkarre V, Couchy G, et al. Molecular subtypes of clear cell renal cell carcinoma are associated with sunitinib response in the metastatic setting. *Clin Cancer Res* 2015;21(6):1329-39. PMID: 25583177. doi:10.1158/1078-0432.CCR-14-1128.
13. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2(5):401-4. PMID: 22588877. doi:10.1158/2159-8290.CD-12-0095.

14. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6(269):pl1. PMID: 23550210. doi:10.1126/scisignal.2004088.
 15. Schemper M, Smith TL. A note on quantifying follow-up in studies of failure time. *Control Clin Trials* 1996;17(4):343-6. PMID: 8889347. doi:10.1016/0197-2456(96)00075-x.
 16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x.
 17. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87. PMID: 8668867. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.
 18. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997;16(4):385-95. PMID: 9044528. doi:10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3.
 19. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 2010;33(1):1-22. PMID: 20808728.
 20. Meinshausen N, Bühlmann P. Stability selection. *J R Stat Soc Series B Stat Methodol* 2010;72(4):417-473. doi:10.1111/j.1467-9868.2010.00740.x.
 21. Wuttig D, Baier B, Fuessel S, Meinhardt M, Herr A, Hoefling C, et al. Gene signatures of pulmonary metastases of renal cell carcinoma reflect the disease-free interval and the number of metastases per patient. *Int J Cancer* 2009;125(2):474-82. PMID: 19391132. doi:10.1002/ijc.24353.
 22. Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013;45(8):860-7. PMID: 23797736. doi:10.1038/ng.2699.
 23. Edeline J, Mottier S, Vigneau C, Jouan F, Perrin C, Zerrouki S, et al. Description of 2 angiogenic phenotypes in clear cell renal cell carcinoma. *Hum Pathol* 2012;43(11):1982-90. PMID: 22626276. doi:10.1016/j.humpath.2012.01.023.
 24. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* 2016;375(12):1109-12. PMID: 27653561. doi:10.1056/NEJMp1607591.
 25. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019;47(W1):W556-W560. PMID: 31114875. doi:10.1093/nar/gkz430.
 26. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373-9. PMID: 8970487. doi:10.1016/s0895-4356(96)00236-3.
 27. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165(6):710-8. PMID: 17182981. doi:10.1093/aje/kwk052.
 28. 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. *Nat Med* 2018;24(10):1550-1558. PMID: 30127393. doi:10.1038/s41591-018-0136-1.
 29. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature* 2018;554(7693):544-548. PMID: 29443960. doi:10.1038/nature25501.
-

Supplementary material — index

The article has six figures and two display tables; the item list that follows is a file index, not a display item. **Supplementary Tables S1–S28.** S1–S18 existed before the first submitted version of this manuscript; S19–S23 were added in a later revision, S20 and S22 being tables retired from an earlier main text. **S24–S28** carry the 36-month clinical-utility analysis shown as Figure 3E — NRI and IDI with a size-matched random-panel null, and the decision curves — together with the 1,000-resample bootstrap. A separate Supplementary Methods document (SM1–SM8) carries the procedural detail the article states only in summary; SM8 documents the clinical-utility protocol. All supplementary tables are supplied in a single Excel workbook, one worksheet per table and sub-table, each beginning with a header row; the workbook's first worksheet carries this index.

Item	Content	Origin
Supplementary Methods SM1–SM8	Procedural detail referenced from Section 2: excluded-patient accounting, panel construction and the anchor audit, model-fitting and scaling conventions, benchmarks and gene lists, external-cohort provenance and probe mapping, the internal-validation protocol, software, and the clinical-utility protocol	Existing SM1–SM7 / SM8 new
S1 / S1b	Panel construction and the mouse-to-human mapping chain (101 → 92 → 87), with the five non-evaluable orthologs named; anchor-gene audit of the twelve named mechanistic anchors	Existing
S2	Univariate Cox screen for all 87 genes with Benjamini–Hochberg q	Existing
S3	The 25 Model A coefficients, each paired with its univariate estimate and sign agreement	Existing
S4a / S4b	Four-gene and three-gene coefficients; the complete 80-point parsimony profile	Existing
S5	Per-patient scores, survival times and covariates for all 508 patients	Existing
S6	Stability-selection frequencies for all 87 genes across 200 bootstrap resamples	Existing
S7a / S7b	LASSO-Cox cross-validation path; reconciliation against the frozen by-size profile	Existing
S8a / S8b / S8c	Randomised-panel benchmark: draw counts, the unpenalised $k = 25$ null, and the full 500-draw distributions	Existing
S9a–S9e	Internal-validation ladder; whole-procedure fold-wise C; paired bootstrap; time-dependent AUC; genes selected per fold	Existing
S10a–S10c	GSE29609 per-gene direction, scale diagnostic and per-patient frozen scores	Existing
S11a / S11b	External-validation summary; GSE29609 scale summary	Existing
S12 / S13	Per-patient mutation calls for 14 driver genes; group comparison with exact Fisher tests and q values	Existing
S14 / S14b	Immune-feature group comparison; marker genes behind each feature	Existing
S15	k-means solutions, silhouettes and between-cluster log-rank tests	Existing
S16a–S16e	E-MTAB-1980 frozen-model summary, per-patient scores, per-gene direction, and the two-cohort matched diagnostic	Existing
S17	The full univariate screen with stage III–	Existing

	IV and grade G3–G4 forced into every gene's model	
S18	Immune comparison with stage stratification and HC3-adjusted linear models	Existing
S19	Discrimination and internal performance of the three models, with the events-per-fitted-term ratios	Existing
S20	Multivariable adjustment for age, sex, stage and grade, for all three models	Existing
S21	Tumour-purity sensitivity analysis: ABSOLUTE purity matched to 479 of 508 patients, added as a covariate to the adjusted Cox model for each score, with the nested likelihood-ratio tests and the missingness comparison	Existing
S22	Pre-specified protocol analyses that were not performed, and the final status of the pre-specified external validation	Existing
S23	36-month calibration of the three frozen models: predicted absolute risk against Kaplan–Meier observed mortality, by quintile of predicted risk	Existing
S24	36-month clinical utility: category-free NRI (IPCW and known-status versions), IDI, reclassification counts, ΔC and the decision-curve summary for the clinical model with and without each score	New
S25	Decision curve at 36 months, threshold by threshold (0.05–0.60 in steps of 0.01): net benefit for treat-all, treat-none, the clinical model, the clinical model plus each score, and the three-gene score alone	New
S26	Size-matched null benchmark for the utility metrics: 500 random three-gene and 500 random 25-gene panels, entered as freely fitted terms in the same clinical model, with the null mean, SD, 95th percentile, maximum and the fraction of draws matching or beating the observed value	New
S27	Per-draw null values behind S26 (500 draws for each of the two arms)	New
S28	Freely fitted arms matched to the null protocol, and the 1,000-resample bootstrap with both models refitted at every draw: NRI, IDI, ΔC and the Δ net benefit at five thresholds	New

Retired main-text tables and where their content now lives: Table 1 of the full report is main-text Table 1 here; the full report's Table 2 → S2; Table 3 → S3; Table 4 → S4a; Table 5 → S19; Table 6 → S20; Table 7 → S9a–S9e; Table 8 → S8a–S8c; Table 9a → S11a and S16a; Table 9b → S11b and S16d; Table 10 → S12/S13; Table 11 → S22. The main text contains 6 figures and 2 display tables.

Figures

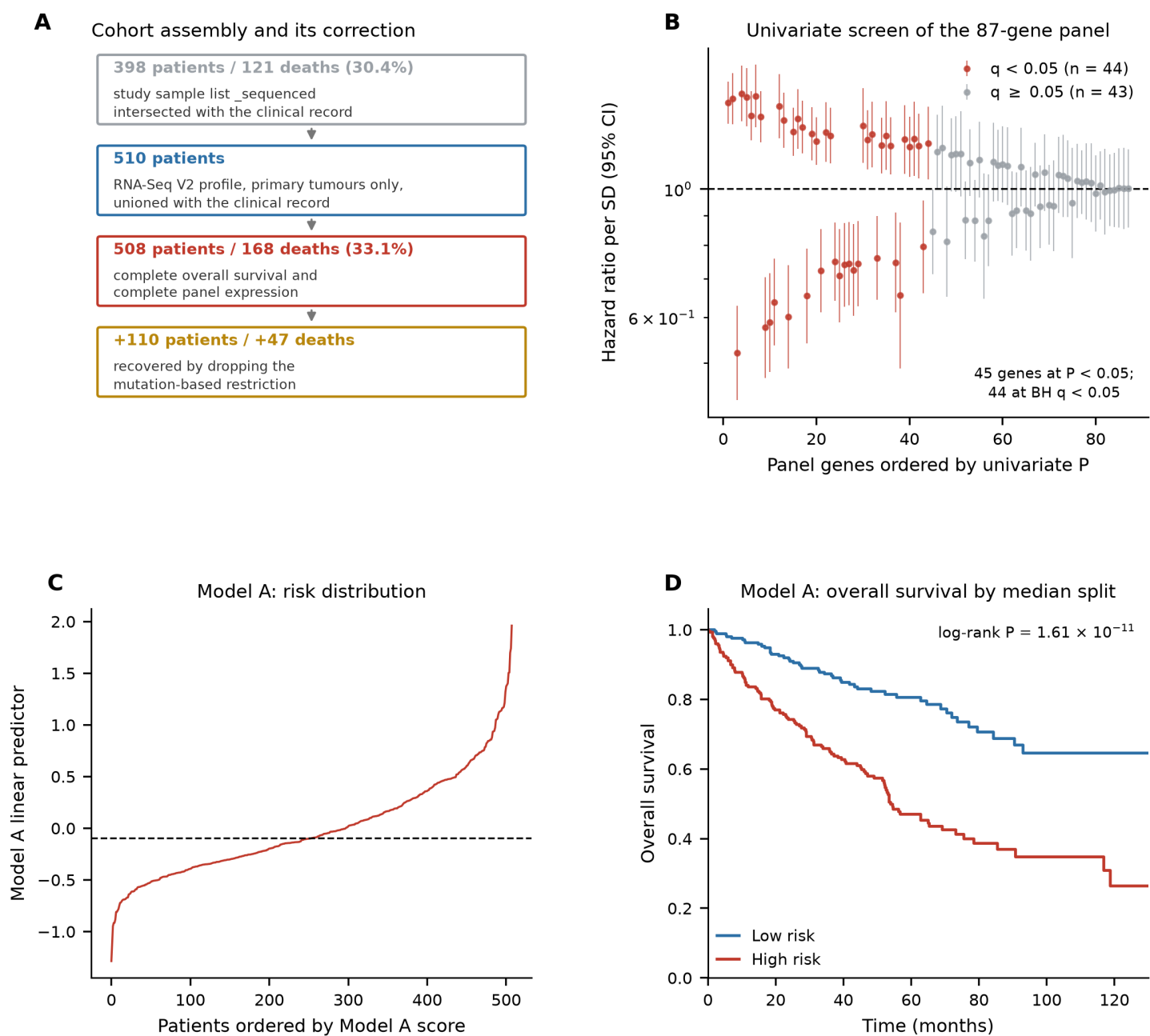


Figure 1

Figure 1. Cohort assembly, the univariate screen and the 25-gene full-panel index (Model A). (A) Assembly of the corrected discovery cohort: rebuilding the matrix from the RNA-seq V2 profile rather than the _sequenced sample list moved the cohort from 398 patients with 121 deaths to 508 with 168, recovering 110 evaluable patients (21.7%). (B) Univariate Cox regression of all 87 evaluable genes, ordered by unadjusted P; hazard ratios are per 1 SD of standardised expression, and red points denote Benjamini–Hochberg q < 0.05. (C) Distribution of the Model A linear predictor across the 508 patients,

with the median split marked. (D) Kaplan–Meier overall survival by Model A median split (254 per group); median overall survival 53.9 months (high risk) versus not reached (low risk).

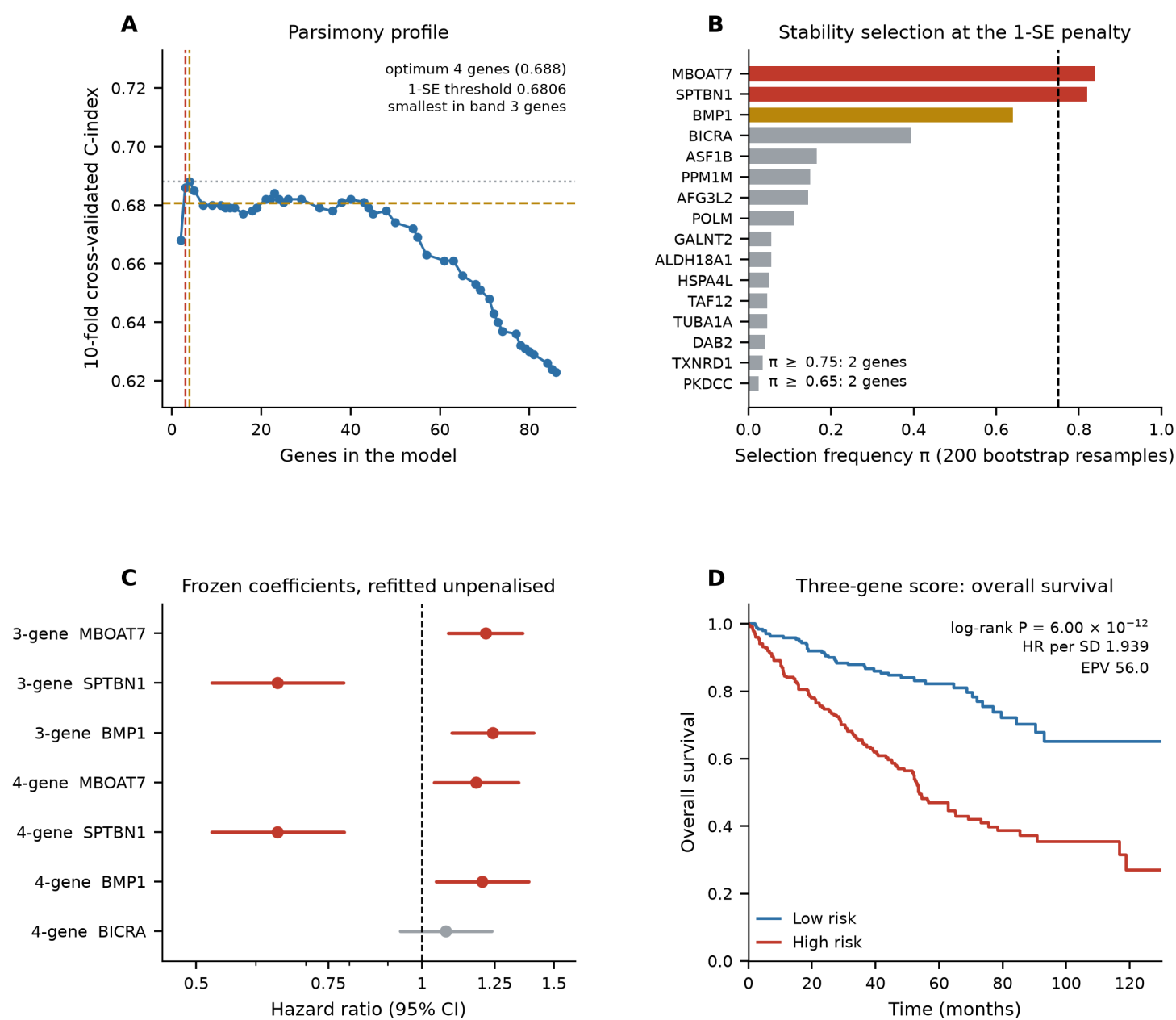


Figure 2

Figure 2. The parsimony profile, stability selection and the three-gene score. (A) Best 10-fold cross-validated C-index at each distinct model size along the LASSO-Cox path. The optimum is four genes; the 1-SE band threshold is 0.6806 (dashed line) and the smallest model inside the band is three genes. (B) Selection frequency π per gene over 200 bootstrap resamples; red bars exceed the pre-specified $\pi \geq 0.75$ threshold (dashed line). Only MBOAT7 (0.84) and SPTBN1 (0.82) clear it. (C) Frozen coefficients, refitted unpenalised: MBOAT7 +0.1954 (HR 1.216, 95% CI 1.085–1.362), SPTBN1 –0.4432 (HR 0.642, 0.524–

0.786), BMP1 +0.2182 (HR 1.244, 1.097–1.410). (D) Kaplan–Meier overall survival by three-gene score median split (254 per group).

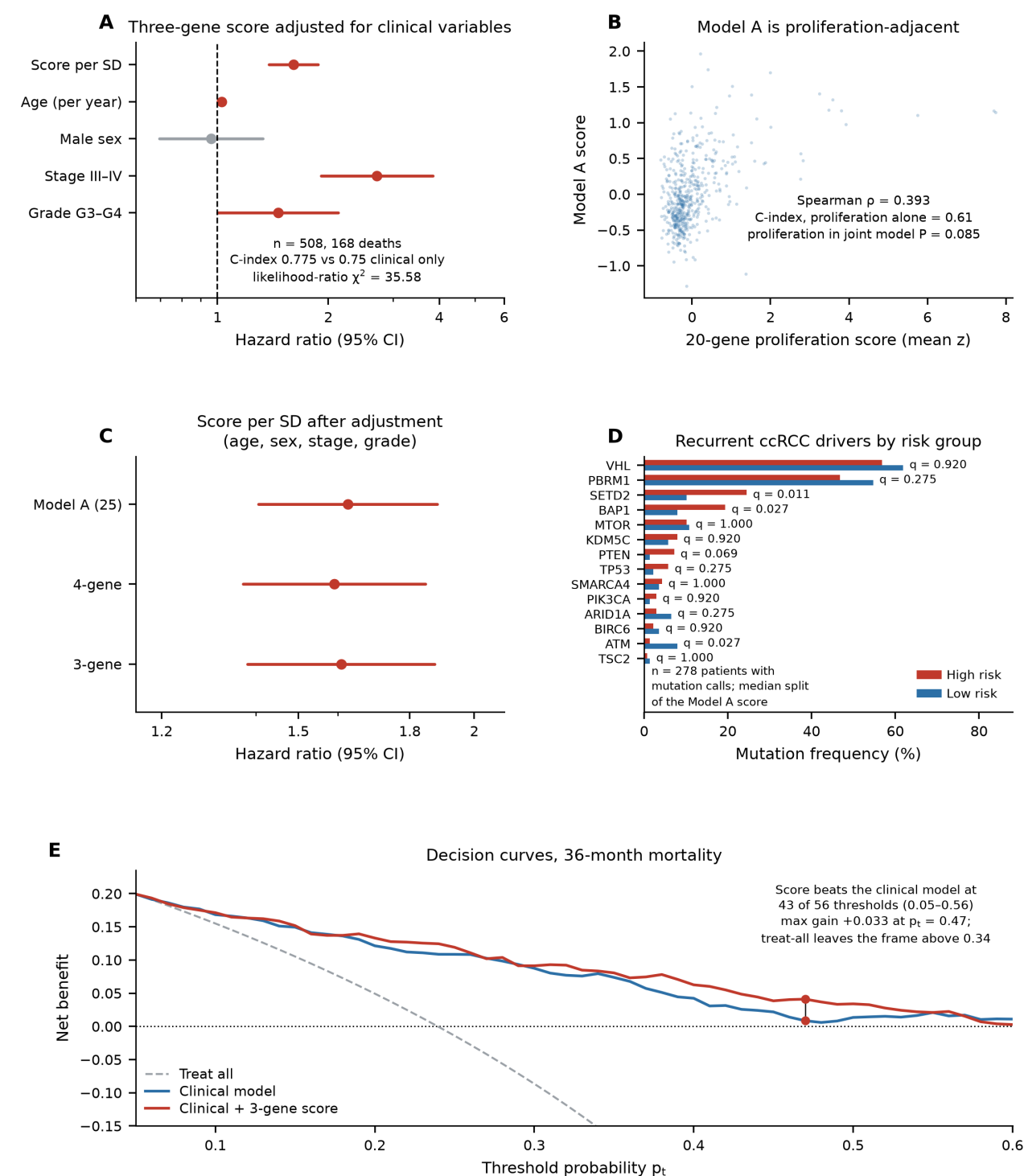


Figure 3

Figure 3. Clinical independence, proliferation adjacency, adjusted effect sizes, mutation landscape and 36-month decision curves. (A) Multivariable Cox model containing the three-gene score with age, sex,

stage III–IV and grade G3–G4 (n = 508; 168 events); forest plot of hazard ratios (95% CI, log scale), red where the interval excludes 1. (B) Model A score against the 20-gene proliferation score; the proliferation coefficient was not significant in a joint model. (C) Score per SD after adjustment for all three models. (D) Mutation frequency of 14 recurrent ccRCC driver genes in the 278 sequenced patients, split at the median of the Model A score (139 per group); SETD2 (24.5% vs 10.1%; q = 0.011) and BAP1 (19.4% vs 7.9%; q = 0.027) survive multiplicity control while VHL (56.8% vs 61.9%) does not differ. (E) Decision curves for 36-month mortality: net benefit against threshold probability for treat-all (grey), treat-none (zero line), the clinical model (blue) and the clinical model plus the three-gene score (red). The score's curve is above the clinical model's at 43 of 56 thresholds between 0.05 and 0.56 and above treat-all between 0.06 and 0.60; the vertical axis is truncated, and treat-all leaves the frame above $p_t = 0.34$.

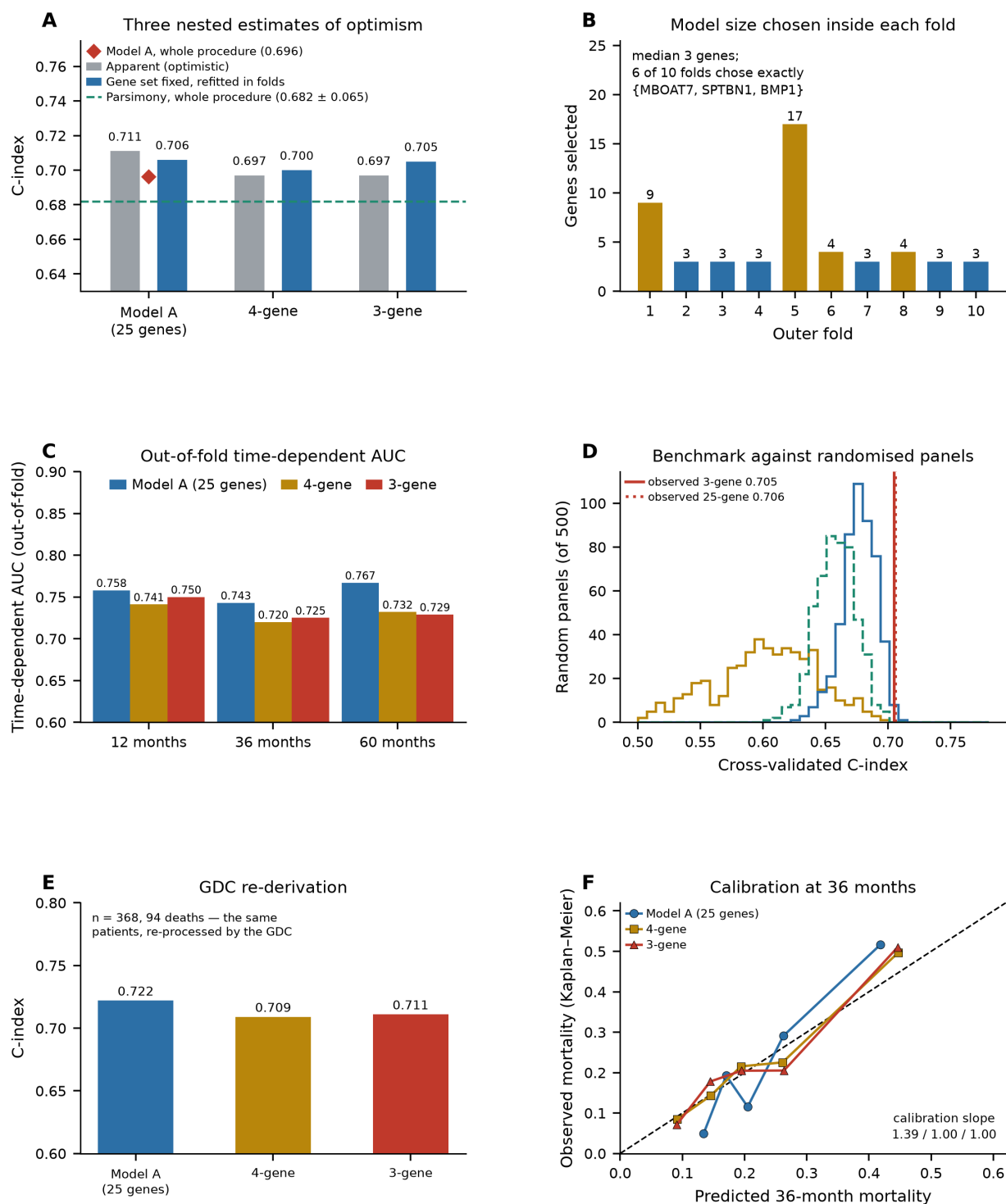


Figure 4

Figure 4. Internal validation: three nested estimates of optimism, fold-level model size, the randomised-panel benchmark, the GDC replication and 36-month calibration. (A) Apparent, coefficients-refitted-inside-folds and whole-procedure-repeated-inside-folds estimates of optimism. Model A 0.711 → 0.706 → 0.696; three-gene 0.697 → 0.705 → 0.682. (B) Genes selected inside each of the ten outer folds: 9, 3, 3, 3, 17, 4, 3, 4, 3, 3 (median 3). (C) Out-of-fold time-dependent AUC at 12, 36 and 60 months: Model A

0.758 / 0.743 / 0.767, four-gene 0.741 / 0.720 / 0.732, three-gene 0.750 / 0.725 / 0.729. (D) Cross-validated C-index of 500 random panels at $k = 3$ (null mean 0.599), $k = 25$ penalised as Model A (0.676) and $k = 25$ unpenalised (0.658). (E) GDC re-processing of the same cohort (368 patients, 94 deaths): $C = 0.722, 0.709$ and 0.711 . (F) Calibration at 36 months: Kaplan–Meier observed mortality against the absolute risk predicted by each frozen model, by quintile of predicted risk ($n = 508, 168$ deaths); the dashed line is perfect calibration.

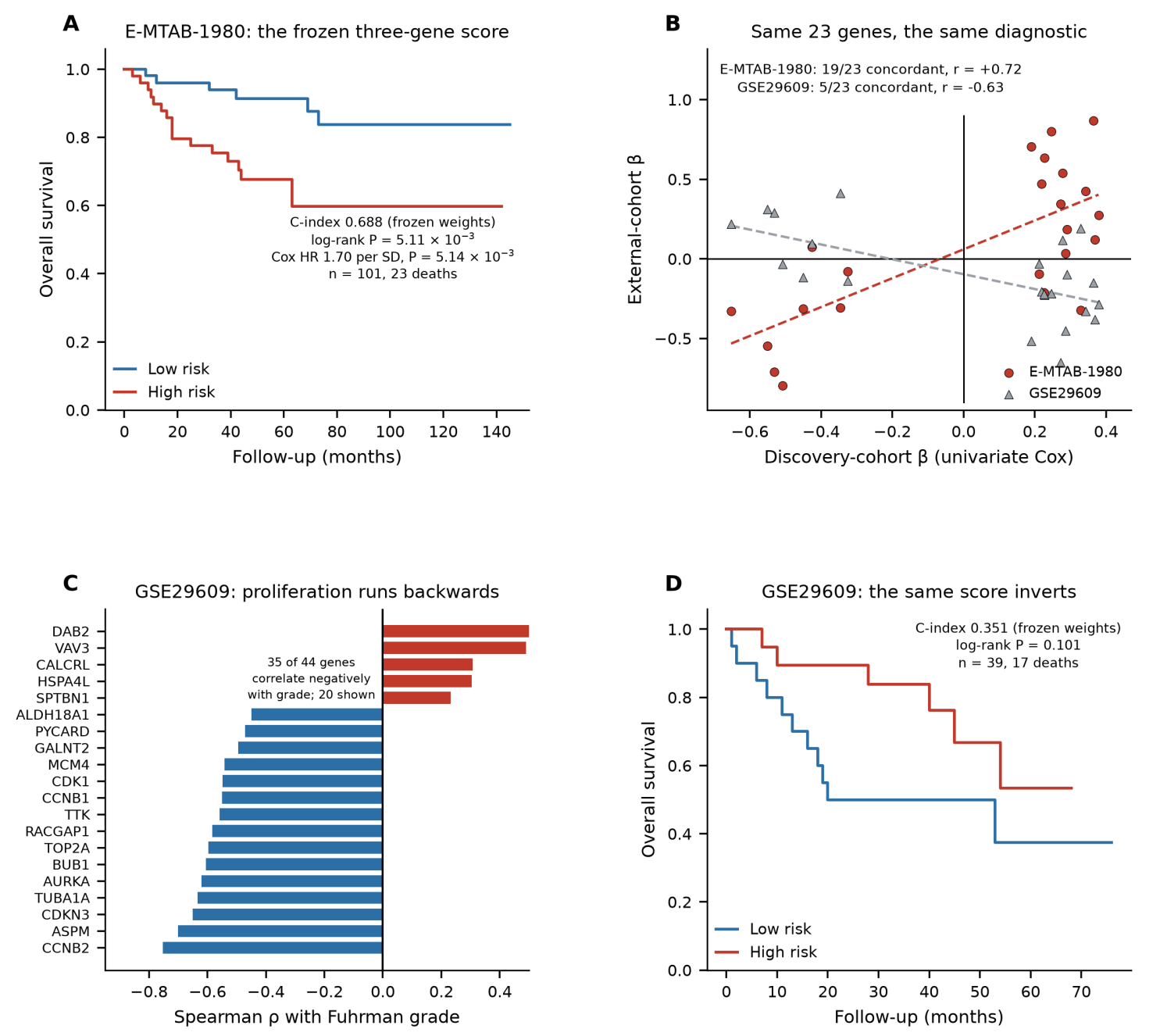


Figure 5

Figure 5. External validation in two independent cohorts, and the matched diagnostic that separates a matrix defect from a biological contradiction. (A) Kaplan–Meier overall survival in the Sato et al. series (E-MTAB-1980 expression joined to the ccrcc_utokyo_2013 survival record; 101 patients, 23 deaths, median follow-up 67.0 months) by the frozen three-gene score at the cohort-internal median: C-index = 0.688, log-rank $P = 5.11 \times 10^{-3}$, Cox HR 1.70 per SD. (B) Matched per-gene diagnostic on the same 23 genes in both cohorts: each gene's external-cohort univariate Cox coefficient against its discovery-cohort value. E-MTAB-1980 (red) agrees for 19 of 23 genes (82.6%); GSE29609 (grey) for 5 of 23 (21.7%). (C) GSE29609 scale diagnostic: 35 of the 44 mappable panel and proliferation genes correlated negatively with Fuhrman grade; the 20 most extreme are shown. (D) Kaplan–Meier overall survival in GSE29609 (39 patients with survival data, 17 deaths) by the same frozen score: it inverts ($C = 0.351$).

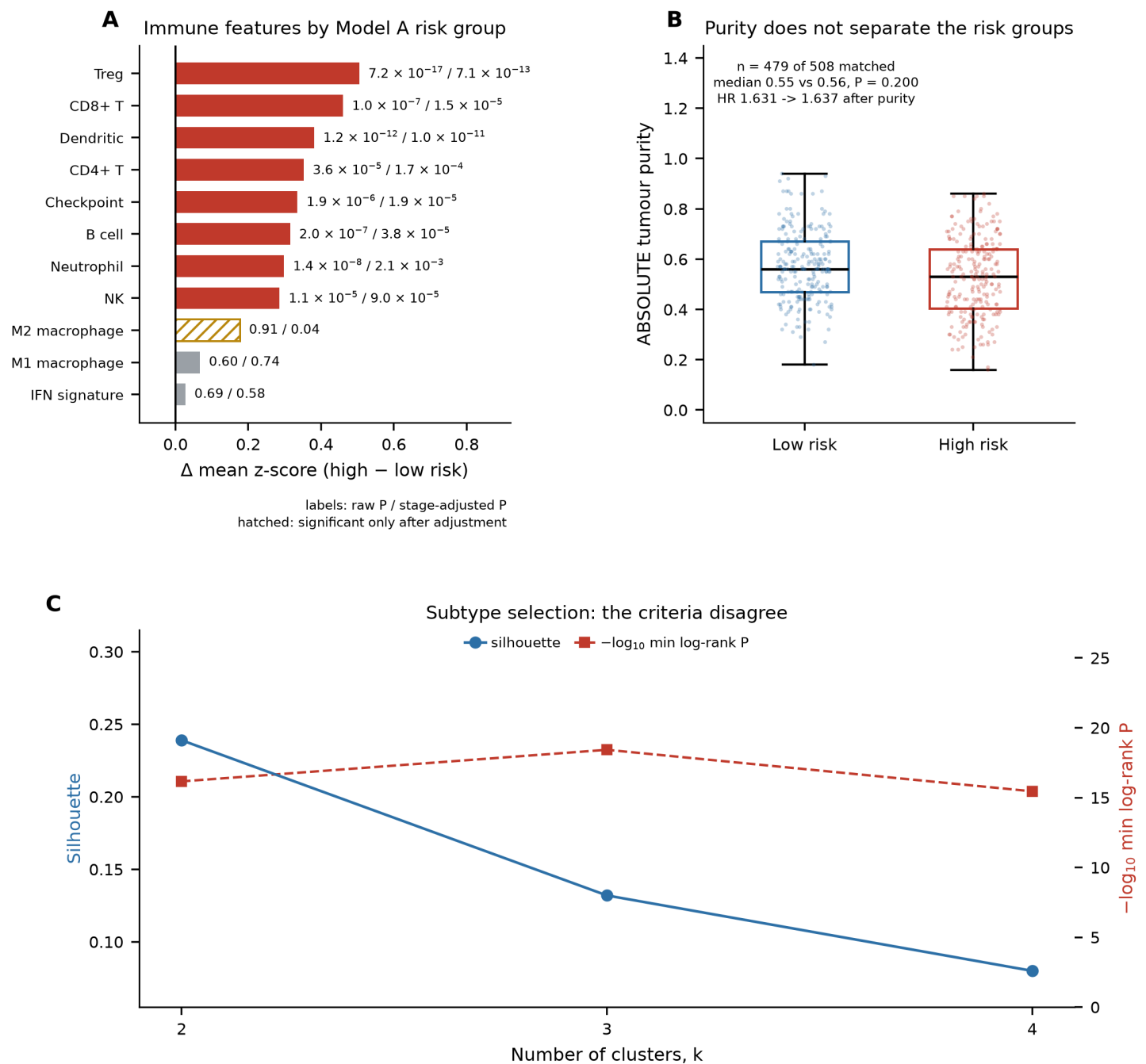


Figure 6

Figure 6. Immune features, tumour purity and molecular subtyping. (A) Marker-gene mean z-scores for 11 immune features in the Model A high- and low-risk groups (254 per group); bars are the mean difference (high minus low risk), and labels give the raw Mann–Whitney P and the stage-adjusted P from a linear model with HC3 standard errors. Eight features were higher by both routes; M1 macrophages (+0.067) and the interferon signature (+0.027) were null on both. M2 macrophages are hatched because they are significant only in the adjusted mean model (P = 0.037), not on the rank test (P = 0.915). (B) Tumour purity (ABSOLUTE, 479 of 508 patients) by risk group (medians 0.55 vs 0.56; P = 0.20), and the

score's hazard ratio before and after adding purity ($1.631 \rightarrow 1.637$). (C) k-means solutions for $k = 2, 3$ and 4: silhouettes 0.239, 0.132 and 0.080; between-cluster log-rank P of 6.91×10^{-17} , 3.77×10^{-19} and 3.55×10^{-16} .

Supplementary Files

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

- [SupplementaryTablesS1S28.xlsx](#)
- [TRIPODChecklist.docx](#)
- [SupplementaryMethods.docx](#)