

A preregistered test that outcome-permuted data also pass: mathematical coupling in a pan-cancer analysis of proliferation adjustment in prognostic gene expression studies

SOON WOO NAM

drswnam@catholic.ac.kr

The Catholic University of Korea Incheon St. Mary's Hospital <https://orcid.org/0000-0001-5562-4775>

Research Article

Keywords: Preregistration, Mathematical coupling, Permutation test, Prognostic biomarkers, Gene expression profiling

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11486198/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Preregistration fixes an analysis before the data are examined, but not whether the registered test can fail. I report a case from prognostic gene expression research, where, in a survey of hepatocellular carcinoma studies, no survival model was adjusted for tumor proliferation.

Methods

For every expressed gene, two Cox models for overall survival were fitted, without and with a standardized proliferation score. A registered plan predicted that b , the slope across genes of the log shift in hazard ratio on correlation with the score, would track L , the log hazard ratio of the score, across cohorts. The test used 20 public cohorts (4,986 patients, 14 cancer types). Post hoc, I permuted survival within all twenty cohorts and refitted the registered test 100 times; separately, in the observed data, I counted genes passing Benjamini-Hochberg control under three covariate sets.

Results

All three registered hypotheses were confirmed: coefficient -0.84 (95% CI -0.90 to -0.78), intercept -0.009 (-0.030 to 0.012), sign rule met in 20 of 20 cohorts. In 100 null replicates, in which nothing is prognostic, all three criteria were met in 92 (95% CI 85 to 96), below the 95% fixed in advance, but too often for the confirmation to support confounding. The observed coefficient lay at the upper edge of the null distribution. In 11 observed cohorts, 22,315 genes passing with age and sex fell to 15,988 when stage was added and to 7,638 when the score was added; stage alone removed most genes in some cohorts.

Conclusions

A preregistered hypothesis can be confirmed largely by construction. Running the registered pipeline on outcome-permuted data, and registering the values its criteria take, would have exposed this before any result was seen. Lists of prognostic genes depend on the covariate set, which should be stated and varied.

Highlights

- All three tested preregistered hypotheses were confirmed in 20 public cohorts
- Outcome-permuted data met all three criteria in 92 of 100 null replicates
- The relation was built in: both quantities came from the same models
- Running a pipeline on permuted outcomes before registering would expose this
- Adding a proliferation score removed two thirds of significant genes overall

1. Introduction

Prognostic gene expression studies in public cohorts often screen candidate genes against survival with a Cox model. Reporting guidance asks that every variable in the final model be reported [1], and prognostic factor research should assess a candidate beyond established prognostic factors [2]. Tumor proliferation carries prognostic information in many cancers [3, 4], and because it correlates with the expression of a large part of the transcriptome, the apparent association of a gene can partly reflect it [5]. In a previously reported sample of recent hepatocellular carcinoma studies [6], none of the 41 papers with a reachable full text adjusted for proliferation (95% CI 0.0% to 8.6%; the other 9 of 50 were not analyzed), and 28 did not state their covariates in the text. Lists of candidate genes are unstable across patient samples [7, 8], and observational associations shift with the choice of adjustment set [9]; how far lists of candidate genes depend on the covariate set in the same patients has received less attention.

Preregistration is recommended for analyses of existing data [10–12], where analytic flexibility can produce almost any result [13]. A registration does not ensure that the registered test could fail. A test that a hypothesis passes whether or not it is true is not a severe test [14], and preregistration alone does not make a test diagnostic [15]. Clinical research has long recognized one source of such tests, mathematical coupling, in which a quantity is related to another computed from the same data [16, 17]. Related safeguards exist: blind analysis fixes an analysis on data whose key result is hidden [18, 19], and negative controls test associations that should be absent [20].

This paper reports a case in which a registered test was confirmed by construction. I registered a prediction about how adjustment for proliferation moves prognostic gene estimates across cancer types, tested it in twenty cohorts, and found it confirmed in every respect. I then reran the same pipeline on data in which survival had been permuted, so that nothing could be prognostic, and its criteria were met again. I show how such a run can calibrate registered criteria, and report what remained: how much lists of significant genes depend on the covariate set.

2. Materials and methods

2.1. Data and cohorts

This is a secondary analysis of public, de-identified data, so ethical approval and consent were not required. Twenty cohorts were analyzed (Supplementary Table S1). Set A comprises four cancer types of The Cancer Genome Atlas that the first plan had excluded for size and that met the lower thresholds of the second, obtained from UCSC Xena [21] as HiSeqV2 matrices (log2 of normalized count + 1) with survival from the pan-cancer clinical data resource [22]. Set B comprises sixteen Gene Expression Omnibus series, obtained with GEOquery [23]. The registered inclusion criteria required a transcriptome-wide matrix, overall survival time and status, age, a recorded anatomic extent, reduced to binary stage under a rule recorded per cohort, and at least 100 patients and 30 deaths with complete data. All 287 screened platform files are listed with a decision and a reason in the deposit. Survival fields, the direction of the status code (one series codes the event as zero) and the stage rule were recorded before any

expression matrix was analyzed, and four errors in my own field selection were corrected before any analysis, with the uncorrected scripts deposited.

2.2. Within-cohort analysis

The proliferation score is the mean of the standardized expression of eight cell-cycle genes, *MKI67*, *TOP2A*, *CCNB1*, *PCNA*, *BUB1*, *CCNA2*, *AURKA* and *RRM2*, fixed in the earlier stages of this work [6] from the proliferation cluster reviewed in [4], and standardized within each cohort; at least six of the eight had to be present. Genes with a median and an interquartile range above zero entered, and the score genes were excluded. For each gene, two Cox models for overall survival were fitted: model 1 with the standardized gene, age, sex and stage, and model 2 adding the score. The shift is the factor by which the hazard ratio of the gene changes between the two models. The slope b of a cohort is the ordinary least squares slope, across genes, of the log shift on the Spearman correlation of the gene with the score. L is the log hazard ratio of the score in the cohort's model with age, sex and stage and no gene. Gene-level P values were adjusted within each cohort by the Benjamini-Hochberg procedure [24], with $q < 0.10$. Sex was dropped in single-sex cohorts and stage where it did not vary.

2.3. The two registrations

The first plan (doi:10.17605/OSF.IO/X5DCF) predicted that b would be negative in at least 90% of included cancer types of The Cancer Genome Atlas, with a median within-cohort R-squared of at least 0.50, and that the shift of a gene would agree across cancer types. The second plan (doi:10.17605/OSF.IO/N8DW6) was registered on the Open Science Framework at 09:02 Korea Standard Time on 21 September 2026, 27 minutes after the verdict on the first had been recorded, and before any of the twenty cohorts had been assembled for analysis. It regressed b on L across cohorts with three hypotheses: H1, a coefficient of L with a 95% CI below zero; H2, an intercept whose 95% CI includes zero; and H3, b of opposite sign to L in at least 80% of cohorts, refuted below 60%. A fourth item, H4, compared the coefficient descriptively, without a threshold, with the value of -0.956 recorded in the plan from the post hoc analysis that motivated it. The unweighted fit was primary; a fit weighted by the inverse squared standard error of L and an attenuation-corrected estimate were secondary.

2.4. Post hoc null checks

These analyses were not registered. Survival time and status were permuted together within a cohort, which keeps expression, the score, every gene-score correlation and the covariates, and removes every association with survival. A first check permuted TCGA-KIRP (large positive L) and TCGA-CESC (L near zero) 100 times each, treating each replicate as one unit of the regression of b on L . The main check reproduced the registered test itself: in each of 100 null replicates, survival was permuted within all twenty cohorts, L was refitted with age, sex and stage, b was recomputed from models 1 and 2, and the registered regression across the twenty was judged by H1, H2 and H3 as registered, H3 requiring at least 16 of 20 cohorts. Each cohort used a fixed subset of 1,500 genes, 150 per decile of the gene-score correlation, drawn once and reused; the unpermuted subset was compared with the full sweep as a control. Permutation seeds derived from 20261013 are recorded in the rule file. Before the main check,

three readings were written to file: the test could not discriminate if all three criteria were met in at least 95% of null replicates, and the observed coefficient and R-squared were indistinguishable from coupling if each lay inside the central 95% of its null distribution.

2.5. Post hoc comparison of covariate sets

In the same patients and genes, three models were compared: A, the gene with age and, where it varied, sex; B, model A with stage; and C, model A with the score. As a calibration, survival was permuted three times in each of TCGA-KIRP and TCGA-PAAD and models A and C were refitted. To compare the strength of the two covariates on one scale, the contribution of each to the linear predictor was taken as its absolute log hazard ratio times its standard deviation, which is 1 for the standardized score and the square root of $p(1 - p)$ for binary stage; for Set A, where p was not recorded, the upper bound of 0.5 was used, which favors stage. The models and the procedure were written to file before computation; the restriction to cohorts with at least 50 significant genes under model A was chosen when the results were tabulated.

2.6. Post hoc threshold check

Using the per-gene P values of the registered models 1 and 2 without refitting, I took the genes significant in model 1 and computed the median of their fractional rank by P value in model 2. Before the computation, a median of 0.05 or less was fixed as movement across the threshold only, 0.20 or more as substantive, and anything between as equivocal. Cohorts with at least 50 genes significant in model 1 were analyzed; this restriction, like that in section 2.5, was chosen when the results were tabulated.

2.7. Software and deposit

Analyses used R 4.6.1 [25] with survival 3.8-6 [26]; figures and derived files were produced in Python from the deposited result files. Code, registered and post hoc rules, every result file and the run logs are deposited at Zenodo [27].

3. Results

3.1. The first registration was refuted

Ten cancer types met the inclusion rule of the first plan. The slope b was negative in five of the ten, and the median pairwise correlation of gene-level shifts between cancer types was -0.293 , so both hypotheses were refuted, as reported previously [6]. Ordering those cohorts by L separated negative from positive slopes; that post hoc reading became the second plan.

3.2. The second registration was confirmed

The twenty cohorts held 4,986 patients and 1,681 deaths in 14 cancer types on 12 platforms, and 459,827 gene fits entered the analysis. The hazard ratio of the score ranged from 0.81 to 2.65. The primary fit gave a coefficient of -0.842 (95% CI -0.900 to -0.784) and an intercept of -0.009 (-0.030

to 0.012), with R-squared 0.981, and the sign rule held in 20 of 20 cohorts (Table 1, Fig. 1A). H1, H2 and H3 were confirmed. The weighted estimate was -0.880 , and Set A and Set B gave -0.80 and -0.88 . For H4, the value of -0.956 lay outside the primary interval.

Table 1

The registered hypotheses applied to the observed cohorts and to 100 null replicates of the same twenty-cohort test.

Hypothesis	Criterion fixed in the registration	Observed: 20 cohorts	Null: 100 replicates of the twenty-cohort test
H1	Coefficient of L with a 95% CI below zero	-0.842 (-0.900 to -0.784)	Met in 100 of 100; median -0.971 (central 95%, -1.053 to -0.832)
H2	Intercept with a 95% CI including zero	-0.0090 (-0.0303 to 0.0123)	Met in 93 of 100; median 0.0018
H3	Sign of b opposite to L in at least 80% of cohorts; refuted below 60%	20 of 20 cohorts	Met in 99 of 100; median 19 of 20 cohorts
H1 to H3	All three met	Yes	92 of 100 (95% CI 85 to 96)
	Coefficient of L on the same 1,500-gene subsets as the null (no criterion)	-0.830	98 of 100 null coefficients at or below it
	R-squared of b on L (no criterion)	0.981	Median 0.974 (central 95%, 0.924 to 0.990)
H4	Descriptive comparison with the value -0.956 from the analysis that motivated the plan; no threshold	-0.956 lies outside the primary interval	Not applicable

The criteria are those of the registered plan, doi:10.17605/OSF.IO/N8DW6; the observed values are from its primary unweighted estimator. In each null replicate, survival time and status were permuted together within every one of the twenty cohorts, so that neither the proliferation score nor any gene carries prognostic information, and the registered regression of b on L across the twenty was refitted; each cohort used a fixed subset of 1,500 genes. With 100 replicates the null percentiles are resolved to about 0.01. b, slope of the log shift of a gene on its Spearman correlation with the proliferation score; L, log hazard ratio of the proliferation score; CI, confidence interval, exact for the share of replicates. Sources: results/CU04, CU05, CV01, DA03, DA04, DA06.

3.3. Permuted data met the same criteria

In the first check (two cohorts, 200 replicates), the permuted data already met all three criteria: the coefficient was -0.929 (-0.956 to -0.903), the intercept -0.002 (-0.007 to 0.003), and b had the sign opposite to L in 192 replicates (Fig. 1B; results/CW04, CW05). The null slope differed between TCGA-KIRP (-0.859) and TCGA-CESC (-1.074).

In the main check, every subset reproduced the deposited L exactly and the deposited b within 0.022 (results/DA01). Across 100 null replicates of the twenty-cohort test, H1 was met in 100, H2 in 93, H3 in 99, and all three in 92 (95% CI 85 to 96; Table 1). The first reading fixed beforehand, at least 95%, was therefore not met. Still, a test that confirms its hypothesis in 92 of 100 null replicates cannot support a confounding interpretation. The null coefficient had a median of -0.971 and a central 95% range of -1.053 to -0.832 . The observed -0.842 lay inside it, shallower than 96 of the 100 null values (Fig. 1C), so the second reading classed it as indistinguishable from coupling. Computed on the same 1,500-gene subsets as the null, however, the observed coefficient was -0.830 , shallower than 98 of the 100 and just outside the band (results/DA06). The null R-squared ranged from 0.924 to 0.990, and the observed 0.981 lay inside. With 100 replicates, the percentile of the observed value carries a Monte Carlo standard error of about 0.02, and the range of L within a null replicate (median 0.55) was narrower than the observed 1.18 (results/DA02, DA03, DA04), so whether the observed coefficient departs from coupling is unresolved.

The reason lies in the construction. When the score is added, each gene's coefficient moves roughly in proportion to the coefficient the score takes, in a direction set by the gene's association with the score, as in the change-in-estimate account of confounding, whatever produced that coefficient, sampling error included. The registered relation would therefore hold under confounding, under a causal effect of proliferation and under no effect at all. The observed data carried the same mark: in 7 of the 20 cohorts L was smaller than its standard error, yet all 20 obeyed the sign rule. The registered attenuation correction, a regression dilution correction [28], assumes that the error in L is independent of b; here the sampling error in L enters b with the opposite sign, so its value (-1.05) is not reported as an estimate.

3.4. Gene lists depend on the covariate set

In six outcome-permuted replicates of two cohorts, no gene passed control under model A or model C, although six replicates cannot exclude a per-replicate false-positive rate of up to 39% (one-sided 95% bound). In the 11 cohorts with at least 50 significant genes under model A, the total of 22,315 fell to 15,988 when stage was added and to 7,638 when the score was added (Table 2, Fig. 2); the other nine cohorts were not analyzed.

Table 2
Genes passing Benjamini-Hochberg control under three covariate sets in the same patients.

Cohort	Cancer	A: age and sex	B: A + stage (kept)	C: A + proliferation score (kept)	Contribution of stage	Contribution of score
TCGA-KIRP	kidney papillary	4,364	1,917 (43.9%)	101 (2.3%)	≤ 1.12	1.15
TCGA-PAAD	pancreas	4,684	4,764 (101.7%)	56 (1.2%)	≤ 0.19	0.48
TCGA-CESC	cervix	326	238 (73.0%)	412 (126.4%)	≤ 0.37	0.05
GSE199633	breast	611	224 (36.7%)	0 (0.0%)	0.44	0.39
GSE68465	lung	493	72 (14.6%)	9 (1.8%)	0.54	0.26
GSE72094	lung	3,671	2,447 (66.7%)	330 (9.0%)	0.41	0.33
GSE136337	myeloma	542	355 (65.5%)	344 (63.5%)	0.33	0.40
GSE84433	stomach	687	524 (76.3%)	177 (25.8%)	0.30	0.16
GSE85047	neuroblastoma	3,603	2,135 (59.3%)	2,140 (59.4%)	1.47	0.90
GSE17536	colon	412	2 (0.5%)	985 (239.1%)	0.76	0.02
GSE42568	breast	2,922	3,310 (113.3%)	3,084 (105.5%)	0.71	0.42
Total		22,315	15,988 (71.6%)	7,638 (34.2%)		

Counts are genes with $q < 0.10$ within the cohort. Model A contains the gene, age and, where it varies, sex. Cohorts with at least 50 significant genes under model A are shown. The contribution of a covariate to the linear predictor is its absolute log hazard ratio times its standard deviation: 1 for the standardized score and the square root of $p(1 - p)$ for binary stage. For the cohorts of The Cancer Genome Atlas the proportion in the higher stage was not recorded, so the upper bound of 0.5 is used ($\leq$). The contribution of the score is from model C and therefore differs from L in Supplementary Table S1, which is estimated with stage in the model. In TCGA-PAAD the higher stage carried the lower hazard (log hazard ratio = 0.39); that group held 8 patients, and the raw stage labels were binned as intended (results/DA05). Not analyzed, with fewer than 50 significant genes under model A: TCGA-UCEC, GSE65858, GSE13507, GSE87371, GSE49997, GSE42127, GSE103479 and GSE162228; GSE50081 was not analyzed because stage does not vary. With survival permuted, no gene passed under model A or C in six replicates (results/CX02); six replicates cannot exclude a per-replicate false-positive rate of up to 39% (one-sided 95% bound). Sources: results/CX01, CY01, CY02, DA05.

Stage alone removed most genes in some cohorts, 85% in GSE68465 and more than 99% in GSE17536, so the loss was not specific to proliferation. With 11 cohorts, the contribution of a covariate showed only a moderate and imprecise association with the share of genes kept: the rank correlation was -0.40 for the score ($P = 0.22$) and -0.54 for stage ($P = 0.09$), where a magnitude of about 0.6 would have been needed for $P < 0.05$.

3.5. The threshold check was equivocal and partly constrained

In the ten cohorts with at least 50 genes significant in model 1, 15,986 genes passed without the score, 5,853 with it, and 3,958 in both; the other ten cohorts were not analyzed. The median, across cohorts, of the median fractional rank of the previously significant genes was 0.083 , between the two thresholds fixed in advance, so the verdict is equivocal. Across the nine cohorts that lost genes, a median of 17% of the lost genes stayed below an adjusted P value of 0.20 , and 24% were pushed above 0.50 . The rule was partly constrained: even with no change in rank, the median could not fall below 0.05 in five of the ten cohorts, so that verdict was unreachable there; the observed median was 1.0 to 9.0 times this floor (median 1.4 ; results/CZ03).

4. Discussion

A registered prediction about how adjustment for proliferation moves prognostic gene estimates was confirmed in twenty cohorts, and the same criteria were met in 92 of 100 permuted replicates of the same test, in which nothing was prognostic. The confirmation of H1 to H3 therefore cannot be taken as evidence of confounding by proliferation. The coefficient itself lay near the upper edge of the null distribution, but that distribution was built from a narrower range of L and a different weighting of cohorts, so this analysis cannot settle whether anything beyond coupling is present. The relation is built in by the change in estimate that any added covariate produces, a form of mathematical coupling [16], akin to the coupling between change and initial value [17].

The registration fixed the analysis but could not ensure that the test could fail [14, 15]. The remedy borrows from blind analysis [18, 19]: before registering, run the registered pipeline with the outcome permuted, record the value each criterion takes, and register those values beside the thresholds. Registered Reports already ask for outcome-neutral checks, such as positive controls, that show a hypothesis can be tested [29], and specification curve analysis judges an observed pattern against permuted or bootstrapped under-the-null data [30]; the step proposed here applies the same idea to the registered criteria themselves. Permutation removes every association with the outcome, so the run reveals nothing about the result. The check does not flag every criterion: for the counts of significant genes, the permuted data returned none (section 3.4), so a registered hypothesis about those counts would pass it. A criterion met under the null must be rewritten as a comparison with the null distribution, computed in every unit of the registered test. Any hypothesis relating an estimate to an adjustment estimated in the same patients is a candidate for this check.

The list of genes passing control for multiple testing in a public cohort depends heavily on the covariate set, in ways that the strength of the added covariate does not reliably predict and that are not specific to proliferation. Such lists vary between patient samples [7, 8], and they also vary between reasonable models in the same patients. Fewer genes under adjustment is not evidence of a better list: where proliferation mediates a gene's association, adjustment removes a real one. A report should state its covariate set, as reporting guidance already asks [1], show how the candidate list changes under alternative sets including a proliferation score, and report the hazard ratio of that score in its own cohort.

The limitations are substantial. The null checks used 100 replicates and fixed subsets of 1,500 genes, and on matched subsets the observed coefficient fell just outside the null band, so a larger run is needed to settle that comparison. All analyses after the registered test are post hoc; their main rules were written to file before computation, but after the registered result had been seen, and the restriction to cohorts with at least 50 significant genes was not fixed in advance. The first registration and the literature survey were posted earlier as a preprint [6], and the survey covers one tumor type and was coded by one person. Stage entered as a binary variable, the score is a construct rather than a measured proliferative index, and all cohorts are bulk tissue. The work rests on public data and has one author.

5. Conclusions

A preregistered hypothesis was confirmed in twenty cancer cohorts, and its criteria were met again in 92 of 100 permuted replicates, in which nothing was prognostic, because the relation it tested is built into the estimates. Rerunning a registered pipeline on outcome-permuted data before registration, and registering the values its criteria take, would expose such a test. Lists of prognostic genes from public cohorts depend substantially on the covariate set, which should be stated and varied.

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

During the preparation of this work the author used Claude (Anthropic) to help write and debug the analysis and figure scripts, to cross-check every reported number against the deposited result files, and to draft and edit the text. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the published article. The study design, the registered hypotheses and decision rules, the interpretation and the conclusions are the author's own.

Declarations

CRedit authorship contribution statement

Soon Woo Nam: Conceptualization, Methodology, Software, Formal analysis, Data curation, Visualization, Writing – original draft, Writing – review and editing.

Declaration of competing interest

The author has no competing interests to declare.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

All data analyzed are publicly available. Expression, clinical and survival data of The Cancer Genome Atlas are available from UCSC Xena (<https://xenabrowser.net>), and the Gene Expression Omnibus series under the accession numbers in Supplementary Table S1. Code, registered rules, post hoc rules, every result file and the run logs are deposited at Zenodo (version 2.6.1, doi:10.5281/zenodo.23340017; concept doi:10.5281/zenodo.22685965). The registered plans are at doi:10.17605/OSF.IO/X5DCF and doi:10.17605/OSF.IO/N8DW6.

Prior posting

An earlier manuscript reporting the first registration and the literature survey was posted as a preprint [6]. The second registration, the null check and the comparison of covariate sets are reported here for the first time.

Acknowledgments

None.

References

1. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM (2005) Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* 97:1180–1184. <https://doi.org/10.1093/jnci/dji237>
2. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA et al (2013) Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 10:e1001380. <https://doi.org/10.1371/journal.pmed.1001380>
3. Starmans MHW, Krishnapuram B, Steck H, Horlings H, Nuyten DSA, van de Vijver MJ et al (2008) Robust prognostic value of a knowledge-based proliferation signature across large patient microarray studies spanning different cancer types. *Br J Cancer* 99:1884–1890. <https://doi.org/10.1038/sj.bjc.6604746>
4. Whitfield ML, George LK, Grant GD, Perou CM (2006) Common markers of proliferation. *Nat Rev Cancer* 6:99–106. <https://doi.org/10.1038/nrc1802>
5. Venet D, Dumont JE, Detours V (2011) Most random gene expression signatures are significantly associated with breast cancer outcome. *PLoS Comput Biol* 7:e1002240. <https://doi.org/10.1371/journal.pcbi.1002240>

6. Nam SW (2026) Omitting tumor proliferation from prognostic gene expression models: a preregistered cross-cancer test, a transcriptome-wide measurement, and a survey of how often the omission is made. *Res Square [Preprint]*. <https://doi.org/10.21203/rs.3.rs-11273325/v1>
7. Ein-Dor L, Kela I, Getz G, Givol D, Domany E (2005) Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 21:171–178. <https://doi.org/10.1093/bioinformatics/bth469>
8. Michiels S, Koscielny S, Hill C (2005) Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 365:488–492. [https://doi.org/10.1016/S0140-6736\(05\)17866-0](https://doi.org/10.1016/S0140-6736(05)17866-0)
9. Patel CJ, Burford B, Ioannidis JPA (2015) Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J Clin Epidemiol* 68:1046–1058. <https://doi.org/10.1016/j.jclinepi.2015.05.029>
10. Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018) The preregistration revolution, *Proc. Natl. Acad. Sci. U. S. A.* 115 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
11. Weston SJ, Ritchie SJ, Rohrer JM, Przybylski AK (2019) Recommendations for increasing the transparency of analysis of preexisting data sets. *Adv Methods Pract Psychol Sci* 2:214–227. <https://doi.org/10.1177/2515245919848684>
12. van den Akker OR, Weston SJ, Campbell L, Chopik WJ, Damian RI, Davis-Kean PE et al (2021) Preregistration of secondary data analysis: a template and tutorial, *Meta-Psychol.* 5 MP.2020.2625. <https://doi.org/10.15626/MP.2020.2625>
13. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci* 22:1359–1366. <https://doi.org/10.1177/0956797611417632>
14. Mayo DG (2018) *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/9781107286184>
15. Szollosi A, Kellen D, Navarro DJ, Shiffrin R, van Rooij I, Van Zandt T et al (2020) Is preregistration worthwhile? *Trends Cogn Sci* 24:94–95. <https://doi.org/10.1016/j.tics.2019.11.009>
16. Archie JP Jr. (1981) *Ann Surg* 193:296–303. <https://doi.org/10.1097/00000658-198103000-00008>. Mathematic coupling of data: a common source of error
17. Tu Y-K, Gilthorpe MS (2007) Revisiting the relation between change and initial value: a review and evaluation. *Stat Med* 26:443–457. <https://doi.org/10.1002/sim.2538>
18. MacCoun R, Perlmutter S (2015) Blind analysis: hide results to seek the truth. *Nature* 526:187–189. <https://doi.org/10.1038/526187a>
19. Dutilh G, Sarafoglou A, Wagenmakers E-J (2021) Flexible yet fair: blinding analyses in experimental psychology. *Synthese* 198:5745–5772. <https://doi.org/10.1007/s11229-019-02456-7>
20. Lipsitch M, Tchetgen Tchetgen E, Cohen T (2010) Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology* 21:383–388. <https://doi.org/10.1097/EDE.0b013e3181d61eeb>

21. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A et al (2020) Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 38:675–678. <https://doi.org/10.1038/s41587-020-0546-8>
22. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD et al (2018) An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 173:400–416e11. <https://doi.org/10.1016/j.cell.2018.02.052>
23. Davis S, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23:1846–1847. <https://doi.org/10.1093/bioinformatics/btm254>
24. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
25. [software] R, Core Team (2026) R: a language and environment for statistical computing, version 4.6.1, R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
26. Therneau TM (2026) A package for survival analysis in R, R package version 3.8-6, <https://CRAN.R-project.org/package=survival>
27. [dataset] SW, Nam (2026) A preregistered test that outcome-permuted data also pass: mathematical coupling in a pan-cancer analysis of proliferation adjustment in prognostic gene expression studies, version 2.6.1. Zenodo. <https://doi.org/10.5281/zenodo.23340017>
28. Frost C, Thompson SG (2000) Correcting for regression dilution bias: comparison of methods for a single predictor variable. *J R Stat Soc Ser Stat Soc* 163:173–189. <https://doi.org/10.1111/1467-985X.00164>
29. Chambers CD, Tzavella L (2022) The past, present and future of Registered Reports. *Nat Hum Behav* 6:29–42. <https://doi.org/10.1038/s41562-021-01193-7>
30. Simonsohn U, Simmons JP, Nelson LD (2020) Specification curve analysis. *Nat Hum Behav* 4:1208–1214. <https://doi.org/10.1038/s41562-020-0912-z>

Figures

Fig. 1

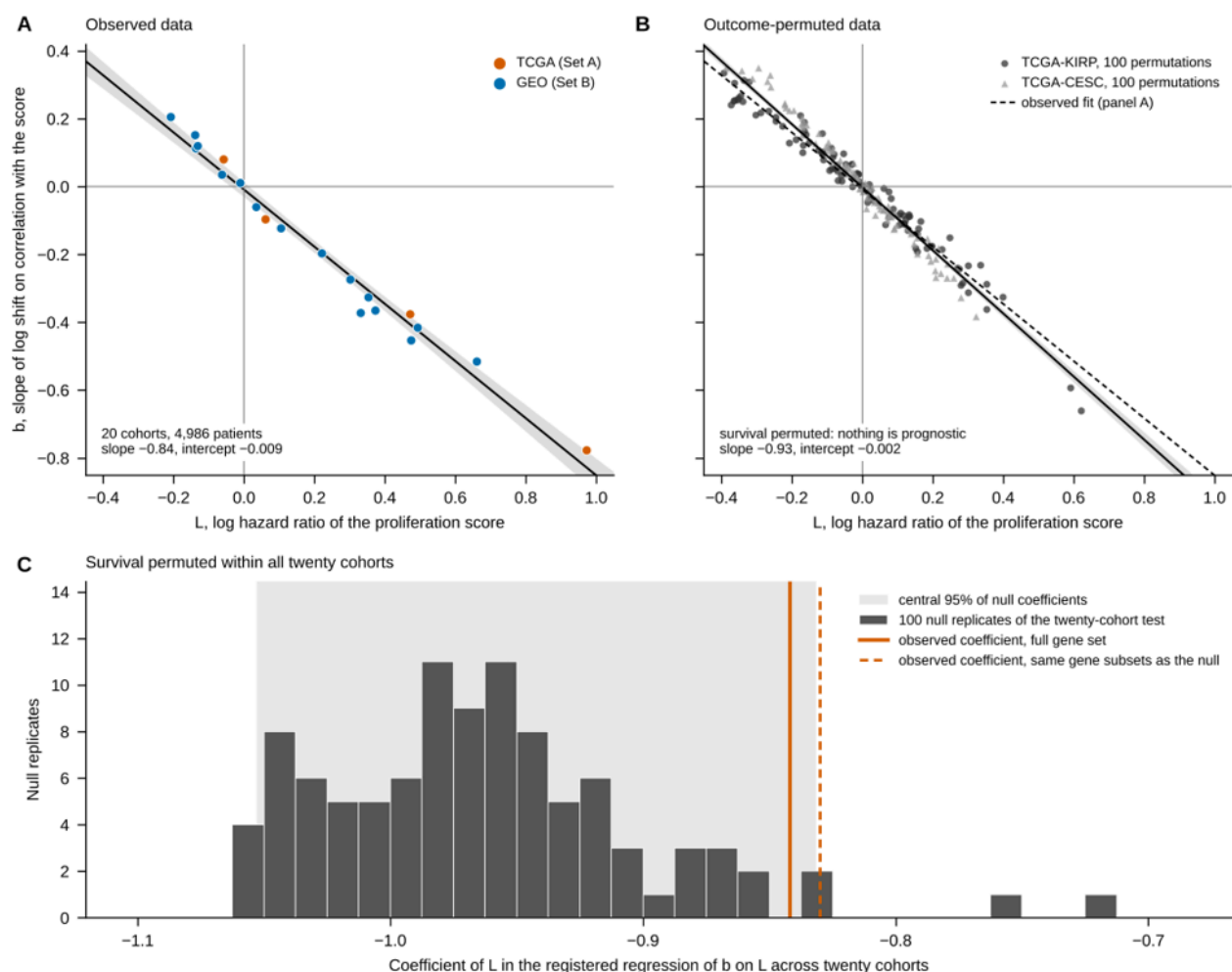


Figure 1

The registered relation in observed and in outcome-permuted data. (A) For each of the 20 cohorts, b , the slope of the log shift in the hazard ratio of a gene on its Spearman correlation with the proliferation score, against L , the log hazard ratio of the score; the line is the registered primary fit with its 95% confidence band. Vermilion, cohorts of The Cancer Genome Atlas; blue, Gene Expression Omnibus series. (B) The same quantities in 200 replicates in which survival time and status were permuted within TCGA-KIRP (dark gray circles) and TCGA-CESC (light gray triangles), so that nothing is prognostic. Each replicate used a fixed subset of 1,500 genes. The solid line is the fit across replicates with its 95% confidence band, and the dashed line repeats the fit of panel A. (C) The coefficient of L from the registered regression across the twenty cohorts in each of 100 null replicates in which survival was permuted within every cohort (gray bars), with the central 95% of the null coefficients shaded and the observed coefficient marked (solid vermilion line); the dashed vermilion line marks the observed coefficient computed on the same 1,500-gene subsets as the null. Sources: results/CV01, CU04, CW02, CW04, DA03, DA06.

Fig. 2

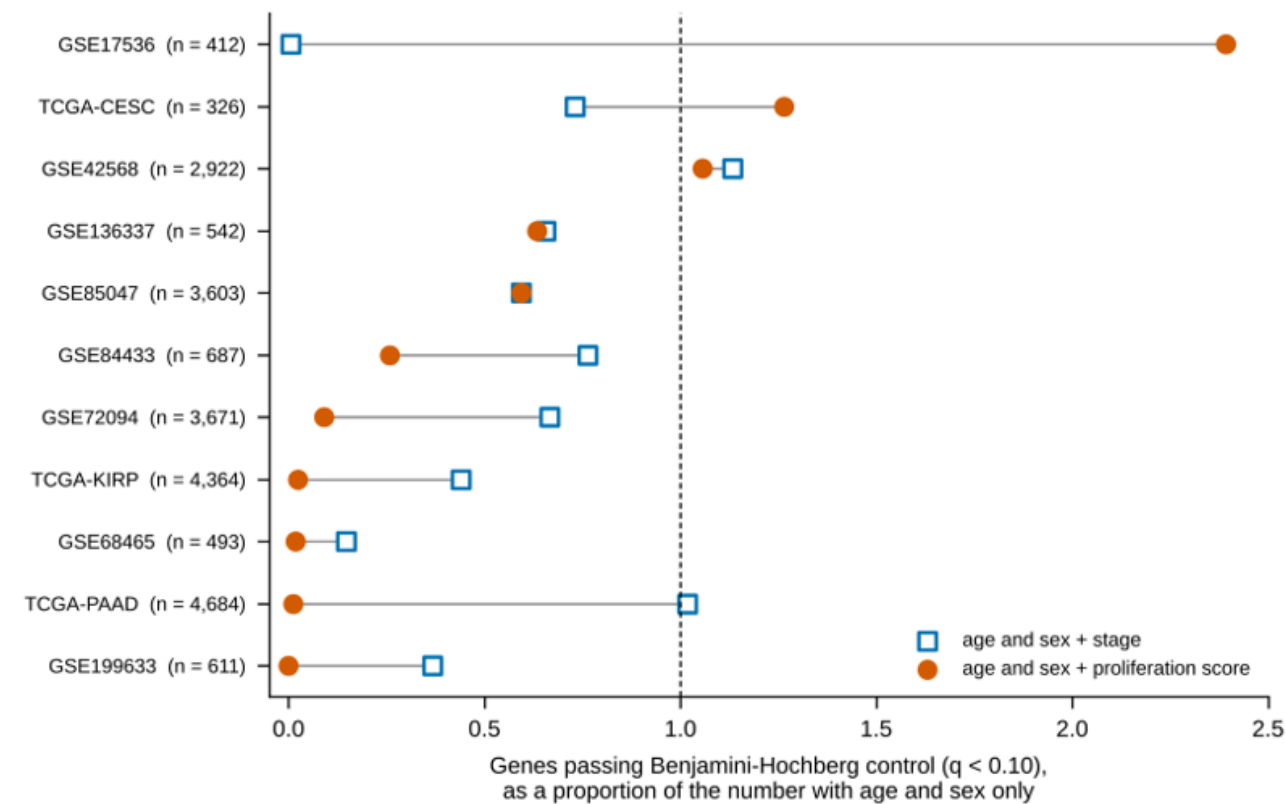


Figure 2

Genes passing Benjamini-Hochberg control ($q < 0.10$) when stage (open squares) or the proliferation score (filled circles) is added to a model with age and sex, as a proportion of the number passing with age and sex only (n), in the 11 cohorts with at least 50 such genes. The dashed line marks no change. Cohorts are ordered by the proportion kept with the score. Source: results/CY02.

Supplementary Files

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

- [Nam2026preprintSupplementaryTableS1.docx](#)