Recurrence associations of a colorectal stemness gene set adapted to bulk expression in three cohorts

Renjia Wu

renjiawu4@gmail.com

Abstract This study assessed whether a fixed bulk adaptation of a published colorectal cancer stemness gene set shows consistent recurrence-related associations in stage II–III colon cancer. This exploratory analysis included 701 patients and 185 events from GSE39582, GSE17536 and The Cancer Genome Atlas colon adenocarcinoma cohort (TCGA-COAD). A fixed 45-gene stemness score (SCS45) and nine-marker comparator used within-sample expression ranks. Separate Cox models adjusted for age, sex and stage; six primary comparisons underwent Benjamini-Hochberg correction. SCS45 hazard ratios per one within-cohort standard deviation were 1.023 (95% confidence interval 0.853–1.228), 0.678 (0.484–0.949) and 1.169 (0.650–2.103), respectively. Associations were not consistent across cohorts, and no primary comparison met the adjusted 0.05 threshold. The nominal GSE17536 association (P = 0.024) had an adjusted P value of 0.142. Alternative scoring changed some estimates. These cohorts provide no consistent evidence of recurrence prognostic value for this bulk adaptation. Endpoint differences and limited events constrain interpretation; the findings neither establish absence of an association nor invalidate the original single-cell assay.

Keywords colon cancer; stemness; gene expression; recurrence; secondary analysis; prognostic association

1. Introduction

Single-cell expression signatures can describe cellular states, but their interpretation after transfer to bulk tumour measurements requires separate evaluation. Lin and colleagues reported a 50-gene colorectal cancer stemness signature using single-cell data [1]. Applying these genes to bulk tissue changes the measurement context because expression reflects multiple cell populations, while a different scoring operation may change the measured phenotype. Stromal contributions to putative intrinsic stemness signatures have been demonstrated [2].

Recent colorectal studies have combined single-cell and bulk data to derive prognostic models or identify therapeutic candidates [3, 4]. A signature that characterises stemness in single cells does not by itself establish the prognostic performance of a different bulk adaptation. This study therefore examined whether a fixed subset of an existing gene set, scored in bulk tumours without outcome-based gene selection or classifier training, shows consistent recurrence-related associations across independently collected cohorts.

2. Methods

2.1 Data sources and eligibility

The analysis used Gene Expression Omnibus (GEO) cohorts GSE39582 and GSE17536, described by Marisa et al. and Smith et al., respectively [5, 6]. The Cancer Genome Atlas colon adenocarcinoma cohort (TCGA-COAD) expression data were obtained through the University of California Santa Cruz (UCSC) Xena platform and linked to the TCGA Pan-Cancer Clinical Data Resource (TCGA-CDR) [7, 8]. Data were accessed on 24–25 September 2026. Eligibility required a primary colon tumour, stage II or III, complete fixed-gene coverage, a valid recurrence-related event indicator, positive follow-up, and complete age and sex. One primary TCGA profile per patient was retained by lexicographic sample identifier without consulting outcomes. The final cohorts contained 453, 111 and 137 patients, respectively.

GSE39582 was reduced from 585 deposited records to 469 stage II–III cases and then to 453 with valid recurrence status and positive follow-up. GSE17536 was reduced from 177 records to 114 stage II–III cases and then to 111 with valid follow-up. In TCGA-COAD, 514 expression profiles were screened, leaving 471 primary

ISSN 3106-5775 (Print) ISSN 3106-7034 (Online)

tumours, 458 unique patients after profile selection, 305 stage II–III cases and 137 with valid DFI and positive follow-up.

Endpoints were deposited relapse-free survival (RFS) in GSE39582, disease-free survival (DFS; the event field specifies cancer recurrence) in GSE17536, and disease-free interval (DFI) in TCGA-CDR. RFS in the GSE39582 publication starts at surgery, whereas TCGA-CDR DFI starts at diagnosis among patients considered disease-free after initial treatment [5, 8]. Cohorts were analysed separately, using all available deposited follow-up without pooling endpoints or estimating absolute calendar-time risks. Dataset screening, endpoint definitions and eligibility flows are detailed in Supplementary Material.docx (Sections B2-B3 and Tables C1-C3).

2.2 Gene mapping and scores

The original signature contains 50 genes. We fixed a 45-gene bulk adaptation (SCS45) because only these genes had retained unambiguous probes under the specified GPL570 mapping; fixing membership across cohorts allowed comparison of the same gene set without reselecting genes using outcomes. The same GPL570-defined set was retained in TCGA RNA-seq data to keep gene membership identical across cohorts. This is a platform-constrained bulk adaptation, not an equivalent reproduction of the original 50-gene single-cell score. Mapping the published gene table excluded probes with missing or multiple symbols and aggregated retained probes by their median [9]. RPL23, RPL10, RACK1, RPL13 and TPT1 were omitted for mapping reasons, not biological absence. Aliases were harmonised before scoring. TCGA log2-transformed transcripts per million, log2(TPM+1), were mapped with the supplied GENCODE v36 annotation and median aggregation.

Expression was ranked in ascending order within each sample, with average ranks for ties. Each score was the mean rank of its genes divided by the mapped background size. Rank-based single-sample scoring is established [10]. No new scoring method is proposed. The fixed nine-marker comparator comprised LGR5, ASCL2, EPHB2, PROM1, AXIN2, LEFTY1, RNF43, CD44 and SLC12A2 [1]. All required genes were present in each retained cohort. SCS45 is a bulk adaptation, not a numerical reproduction of the original single-cell AddModuleScore assay.

Alternative scoring used the common 18,237-gene background, the mean of gene-wise within-cohort Z values, or expression-matched control adjustment. Control adjustment used 24 expression bins and 100 sampled controls per signature gene, subtracting the mean of the unique control union. Seed 20260925 was the fixed calculation, and 20 consecutive seeds assessed stability. Full scoring definitions are provided in Supplementary Material.docx (Section B5); cohort-dependent Z and control-adjusted scores are not fixed clinical assays.

2.3 Analysis plan and statistical methods

The primary analysis for the three-cohort extension comprised six comparisons: the native-rank SCS45 and nine-marker scores in each cohort. The two microarray cohorts had already been analysed before the comparison family was expanded from four to six tests. The extension plan and amendments were recorded locally before TCGA score-outcome modelling. Thus, the study remained exploratory and was not externally preregistered. Alternative scoring, sex stratification, treatment and mismatch-repair (MMR) adjustment, the complete 50-gene TCGA score, and five-year TCGA follow-up restriction were sensitivity analyses outside the primary family.

Separate Cox models estimated hazard ratios (HRs) and nominal 95% confidence intervals (CIs) per one standard deviation (SD) higher score within the modelled cohort, adjusting for age per ten years, sex and stage III versus II [11]. Efron handling was used for tied events. Benjamini-Hochberg (BH) correction was applied to the six primary tests, with adjusted P values denoted q and a threshold of 0.05 [12]. This correction addresses only the defined six-test family; it cannot remove exploratory or selection limitations arising during analysis-plan development and does not make the study confirmatory. Sensitivity P values were descriptive. Rank-transformed Schoenfeld-residual tests assessed proportional hazards. No outcome-optimised threshold, pooled effect estimate or treatment-effect analysis was used.

Analyses used Python 3.12, pandas 2.3.3 and lifelines 0.30.3. An independent Efron partial-likelihood implementation checked the six primary HRs, with a maximum absolute difference below 0.00001 (Supplementary Material.docx, Sections B6 and D1). Additional_file_1.zip provides the original results, model coefficients, diagnostics and executable reproduction code. Reporting was guided by the Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) [13].

3. Results

The analysis included 701 patients and 185 events: 453 patients/136 events in GSE39582, 111/31 in GSE17536 and 137/18 in TCGA-COAD (Table 1). The sequential eligibility counts are reported in Section 2.1.

Table 1. Characteristics of the analysed cohorts

Characteristic	GSE39582	GSE17536	TCGA-COAD
Patients	453	111	137
Events	136	31	18
Stage II / III	253 / 200	55 / 56	85 / 52
Male / female	255 / 198	53 / 58	71 / 66
Age, years	69 (59-76.9)	67 (56-74)	65 (57-74)
Endpoint	RFS	DFS	DFI
Gene background	20845	20845	59426

Age is median (interquartile range); other clinical entries are counts. RFS, relapse-free survival; DFS, disease-free survival; DFI, disease-free interval; TCGA-COAD, The Cancer Genome Atlas colon adenocarcinoma cohort. These recurrence-related endpoints are not interchangeable. Gene background is the number of mapped features used for native ranking. No included patient lacked an adjustment covariate.

Per one within-cohort SD higher SCS45, covariate-adjusted HRs were 1.023 (95% CI 0.853–1.228; P = 0.805) in GSE39582, 0.678 (0.484–0.949; P = 0.024) in GSE17536 and 1.169 (0.650–2.103; P = 0.603) in TCGA-COAD (Figure 1). The wide TCGA CI spans substantial inverse and positive associations, reflecting limited precision with 18 events for four fitted coefficients; its P value alone cannot establish the presence or absence of an association. Comparator HRs were 1.059 (0.881–1.274), 0.757 (0.549–1.044) and 1.037 (0.619–1.737), respectively.

None of the six primary comparisons met the BH-adjusted 0.05 threshold. SCS45 q values were 0.890, 0.142 and 0.890, respectively; the smallest q across all six comparisons was 0.142. The GSE17536 association was therefore nominally significant only. SCS45-comparator Spearman correlations were 0.715, 0.796 and 0.600, respectively.

The cohort-specific estimates should not be read as a direct comparison of prognostic strength. The SCS45 point estimate was near the null in GSE39582, below one in GSE17536 and above one in TCGA-COAD, but the confidence intervals and endpoint definitions differ. No pooled estimate was calculated; without harmonised endpoints and a formal heterogeneity assessment, opposite point-estimate directions alone do not establish statistical or biological heterogeneity.

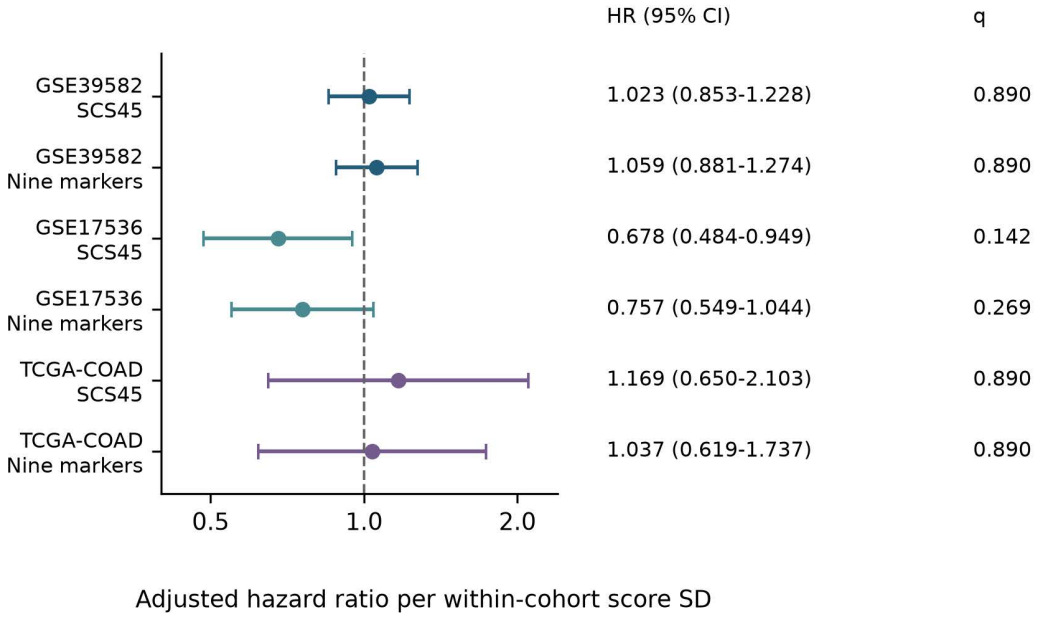


Figure 1. Primary recurrence-related associations. Separate Cox models adjust for age, sex and stage. Points show hazard ratios (HRs) per one within-cohort standard deviation (SD) higher score; bars show nominal 95% confidence intervals (CIs). The dashed line indicates HR = 1. q denotes Benjamini-Hochberg adjustment across

ISSN 3106-5775 (Print) ISSN 3106-7034 (Online)

all six primary comparisons. SCS45 is the fixed 45-gene bulk-adapted stemness score; Nine markers is the fixed nine-marker comparator. TCGA-COAD denotes The Cancer Genome Atlas colon adenocarcinoma cohort. Estimates are not pooled across endpoints.

Common-background ranking closely preserved SCS45 sample ordering (Spearman correlations 0.978–1.000). Mean-Z and control-adjusted scoring changed some estimates; in GSE17536, both 95% CIs included one (Figure 2). Primary score proportional-hazards diagnostic P values ranged from 0.206 to 0.898. Sex-stratified models, additional treatment/MMR adjustment in GSE39582, and the complete 50-gene and five-year analyses in TCGA-COAD also yielded SCS45 or full-signature CIs that included one, except for the sex-stratified GSE17536 model. That nominally significant GSE17536 sensitivity result was outside the six-test primary family. These analyses are descriptive and provide no independent confirmation; estimates are reported in Supplementary Material.docx (Section A, Tables A1-A3 and accompanying text).

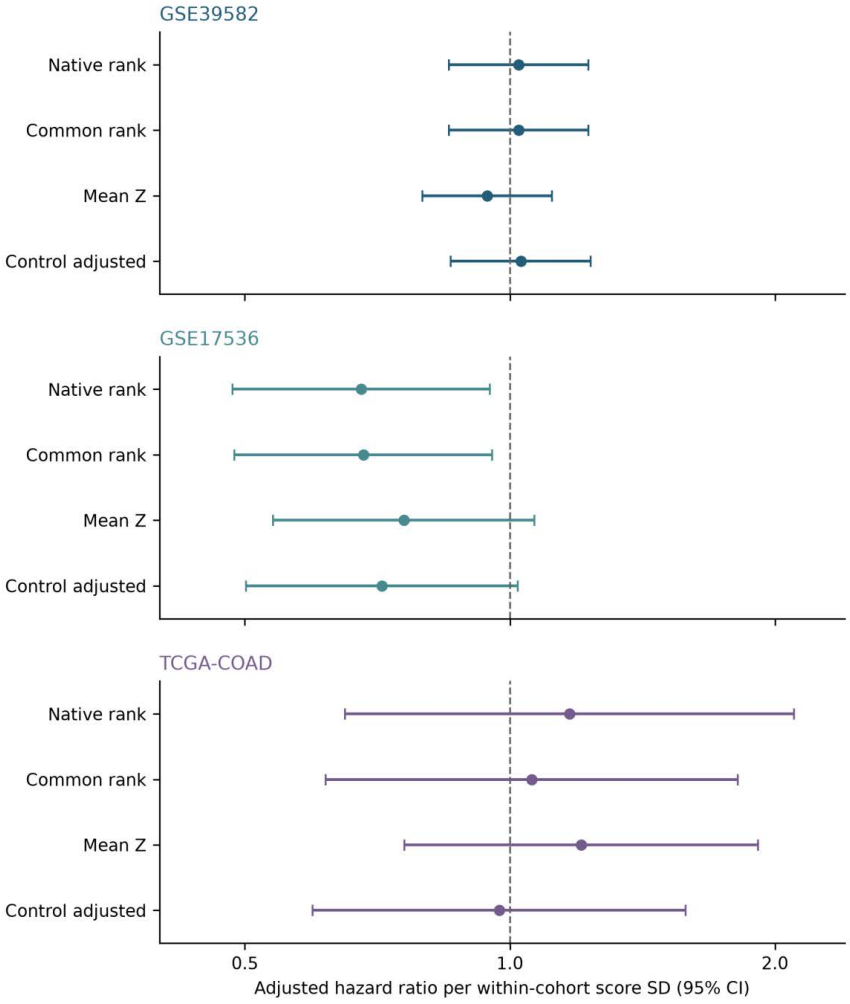


Figure 2. SCS45 associations under alternative scoring operations.

Points show hazard ratios (HRs) per one within-cohort standard deviation (SD) higher score; bars show nominal 95% confidence intervals (CIs). SCS45 is the fixed 45-gene bulk-adapted stemness score. Native rank uses each platform's mapped background; Common rank uses 18,237 shared genes. Mean Z averages gene-wise standardised expression; Control adjusted subtracts expression-matched controls using seed 20260925. Models adjust for age, sex and stage. TCGA-COAD denotes The Cancer Genome Atlas colon adenocarcinoma cohort. Native-rank estimates are shown for reference; the alternative-scoring analyses are outside the six-test primary family.

Across 20 control-selection seeds, SCS45 HRs ranged from 1.025 to 1.033 in GSE39582, 0.709 to 0.724 in GSE17536 and 0.957 to 0.993 in TCGA-COAD. The relatively narrow ranges indicate limited variation across seeds under this control-adjustment procedure, but do not establish equivalence between scoring methods. Results for both gene sets are provided in Supplementary Material.docx (Tables A4-A10), with full-precision records in Additional_file_1.zip (results/control_seed_sensitivity.csv).

4. Discussion

4.1 Interpretation

The GSE17536 SCS45 result was nominally significant (P = 0.024), but its BH-adjusted q = 0.142 did not meet the primary threshold. SCS45 point estimates were below one in that cohort and above one in the other two; none survived correction. The combined finding is a lack of consistent evidence for recurrence prognostic value of this bulk adaptation, not proof that no association exists. The TCGA estimate was imprecise and compatible with associations in either direction. Given the different endpoints and limited TCGA event count, differences between cohort estimates cannot be interpreted as evidence of statistical or biological heterogeneity. Non-significance does not establish equivalence.

Fixing gene membership allowed scoring choices to be examined without training a new predictor on outcomes. Common-background ranking and full-gene coverage in TCGA did not materially strengthen the evidence. Variation across the sampled control seeds was smaller than some changes between scoring operations, but this is not a formal test of differences between methods. Bulk cellular composition, treatment differences and assay context could contribute to the inconsistent estimates; stromal admixture was not directly measured and cannot be asserted as their explanation. Measurement correspondence and clinical performance require further evaluation before this adaptation can be interpreted as a recurrence prognostic marker.

The nine-marker score should be interpreted as a comparator rather than an independent validation signature. Four of its genes (ASCL2, LEFTY1, RNF43 and SLC12A2) also occur in SCS45, so the observed score correlations may partly reflect shared gene membership. This comparison can indicate whether a broader fixed set behaves differently from a compact marker set, but it does not establish the superiority of either score or independent prognostic validity.

4.2 Limitations

This retrospective, exploratory analysis incorporated known microarray results before TCGA extension planning and expansion of the multiplicity family. Correction across six tests does not remove this design limitation. Public cohorts are selected samples. TCGA DFI eligibility left only 18 events for a four-predictor model, limiting precision and stability. Treatment and MMR data were not consistently available; additional adjustment in GSE39582 also restricted the sample to 409 patients, so its change in estimate cannot be attributed solely to confounding control. Endpoint definitions, origins and follow-up differ, and competing risks were not harmonised.

The 45-gene adaptation changes coverage and scoring relative to the source assay. SCS45 measures a relative expression score calculated from bulk tissue under the present rules; it cannot be equated with cancer-stem-cell abundance or the original single-cell stemness phenotype. Older array annotation was not sequence-reannotated, and platform backgrounds differ. No cell-type deconvolution, functional stem-cell experiment, discrimination, calibration or clinical utility assessment was performed. Within-cohort standardisation supplies no deployable threshold. These results neither invalidate the original single-cell study nor exclude clinically relevant associations that the available cohorts are too limited to detect.

5. Conclusion

The 45-gene stemness score was not consistently associated with recurrence across the three stage II–III colon cancer cohorts after adjustment for age, sex and stage. None of the six primary comparisons remained significant after Benjamini–Hochberg correction, and some estimates changed when different scoring methods were used. The results therefore provide insufficient evidence to support this bulk-adapted score as a recurrence prognostic marker. They do not rule out an association or challenge the original single-cell signature. Studies with more recurrence events and comparable endpoint definitions are needed to assess the score further and clarify what it measures in bulk tumour tissue.

References

[1] Lin K, Chowdhury S, Zeineddine MA, Zeineddine FA, Hornstein NJ, Villarreal OE, et al. Identification of Colorectal Cancer Cell Stemness from Single-Cell RNA Sequencing. Molecular Cancer Research. 2024;22(4):337-346. https://doi.org/10.1158/1541-7786.MCR-23-0468

[2] Kreis J, Aybey B, Geist F, Brors B, Staub E. Stromal Signals Dominate Gene Expression Signature Scores That Aim to Describe Cancer Cell-intrinsic Stemness or Mesenchymality Characteristics. Cancer Research Communications. 2024;4(2):516-529. https://doi.org/10.1158/2767-9764.CRC-23-0383

[3] Chen IH, Chang KF, Chao CC, Lin CH, Chang CH, Ko CC, et al. An integrated bulk and single-cell transcriptomic analysis reveals stemness-driven immune regulation and therapeutic vulnerability in colorectal cancer. Journal of Cancer. 2026;17(4):740-758. https://doi.org/10.7150/jca.132694

[4] Li G, Wang D, Chen G. Single-cell RNA sequencing defines prognostic subtypes and identifies AIF1L as a therapeutic target in colorectal cancer. BMC Cancer. 2025;25:1885. https://doi.org/10.1186/s12885-025-15241-2

[5] Marisa L, de Reyniès A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLOS Medicine. 2013;10(5):e1001453. https://doi.org/10.1371/journal.pmed.1001453

[6] Smith JJ, Deane NG, Wu F, Merchant NB, Zhang B,

ISSN 3106-5775 (Print) ISSN 3106-7034 (Online)

Jiang A, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010;138(3):958-968. https://doi.org/10.1053/j.gastro.2009.11.005

[7] Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nature Biotechnology. 2020;38(6):675-678. https://doi.org/10.1038/s41587-020-0546-8

[8] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell. 2018;173(2):400-416.e11. https://doi.org/10.1016/j.cell.2018.02.052

[9] Lin K, et al. Supplementary Table 3 from Identification of Colorectal Cancer Cell Stemness from Single-Cell RNA Sequencing [dataset]. AACR Figshare. 2024. https://doi.org/10.1158/1541-7786.25523139

[10] Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of molecular phenotypes. BMC Bioinformatics. 2018;19:404. https://doi.org/10.1186/s12859-018-2435-4

[11] Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological). 1972;34(2):187-202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

[12] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological). 1995;57(1):289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

[13] McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM; Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics. Reporting recommendations for tumor marker prognostic studies. Journal of Clinical Oncology. 2005;23(36):9067-9072. https://doi.org/10.1200/JCO.2004.01.0454