Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Eight-Gene Prognostic Signature Capturing EGFR-Driven, Immune-Cold Biology in Cutaneous Melanoma

Shafi Rehman

shafirehman@hotmail.com

Khyber Medical College https://orcid.org/0000-0002-7262-0300

Research Article

Keywords: Prognostic signature, Melanoma, EGFR

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11731164/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.

An Eight-Gene Prognostic Signature Capturing EGFR-Driven, Immune-Cold Biology in Cutaneous Melanoma

ABSTRACT

Background: Cutaneous melanoma remains the deadliest skin cancer despite advances in targeted therapy and immunotherapy. About 14–27% of melanomas carry mutations in the NF1 tumor suppressor gene, and these patients have limited options once immunotherapy fails. Recent work has shown that NF1 loss drives EGFR upregulation and EGFR-dependent growth in patient-derived models, that NF1-mutant melanomas form immune-cold meta-niches with

impaired antigen presentation and CD8 T cell exclusion, and that they harbor a proliferative, CDC20-dependent transcriptional program amenable to combination targeting. However, no bulk-tissue, multi-cohort gene expression signature has yet operationalized this EGFR-driven, immune-cold biology in association with clinical outcomes, and it remains unknown whether such a signature tracks with NF1 mutation status or captures a broader EGFR-driven phenotype.

Methods: From 242 NF1/RAS pathway genes, we built a signature using TCGA-SKCM bulk RNA-seq. WGCNA identified genes co-expressed with *NF1*; LASSO Cox regression selected the final signature. We evaluated it in four independent GEO cohorts and characterized its biology with GSVA, immune deconvolution, and single-cell RNA-seq (GSE72056), and assessed associations with NF1 mutation status and clinical features.

Results: The final 8-gene signature (PLA2G2D, FASLG, RASGRP1, FLT3, FGF1, PLA2G5, FGF2, EGFR) stratified patients into High and Low Risk groups with significantly different overall survival (log-rank $p = 1.4\times10^{-10}$). EGFR was the only risk-factor gene (coefficient = +0.118); the remaining genes were protective. The signature was independent of age and stage in multivariate Cox analysis (HR = 5.32, 95% CI 3.55–7.99, $p < 0.001$; C-index = 0.72), with time-dependent AUCs of 0.72 (1-year), 0.73 (2-year), and 0.71 (3-year). GSVA showed that all top differentially active pathways were immune-related and downregulated in High-Risk patients (e.g., T cell receptor signaling, antigen presentation, NK cell cytotoxicity; FDR $< 10^{-28}$). Immune deconvolution confirmed depletion of CD8 T cells, CD4 T cells, NK cells, and cytolytic activity in High-Risk tumors. Single-cell analysis localized protective genes to immune cells (RASGRP1, 26.4%; FASLG, 9.8%), whereas EGFR was tumor-intrinsic and expressed in only 2.7% of cells. The signature was directionally consistent in three of four external cohorts (HR > 1), with a pooled meta-analysis HR of 2.25 (95% CI 1.03–4.89, $p = 0.041$; $I^2 = 7.6\%$). It did not

correlate with NF1 mutation status ($p = 0.12$; $n = 72$), consistent with the possibility that the signature captures a broader EGFR-driven phenotype rather than NF1 loss per se. This analysis was underpowered to detect a modest association.

Conclusions: We present a bulk-tissue 8-gene signature that operationalizes the EGFR-driven, immune-cold biology recently defined mechanistically, spatially, and therapeutically in NF1-mutant melanoma. The signature captures a broader EGFR-driven phenotype that is not restricted to NF1-mutant tumors, as it did not correlate with NF1 mutation status in our cohort. It stratifies survival in TCGA-SKCM, is independent of stage, captures an immune-cold phenotype concordant with spatial findings, and is directionally consistent in external cohorts. These findings generate testable hypotheses for selecting patients for EGFR inhibitor and immunotherapy combinations and provide a bulk-computable biomarker for prospective testing. The signature is not yet a validated clinical assay, but it is a biologically anchored, hypothesis-generating tool that operationalizes the EGFR-driven, immune-cold phenotype and is immediately testable in existing RNA-seq cohorts.

1. INTRODUCTION

Cutaneous melanoma remains the deadliest form of skin cancer, with an estimated 100,000 new cases and 8,000 deaths annually in the United States alone [1]. The advent of targeted therapies (BRAF and MEK inhibitors) and immune checkpoint inhibitors (ICIs) have dramatically improved outcomes for many patients [2, 3]. However, a substantial fraction of patients either do

not respond to these therapies or develop resistance, leaving them with few effective options. Genomic classification of cutaneous melanoma has identified four major subtypes based on driver mutations: BRAF-mutant (~52%), NRAS-mutant (~28%), NF1-mutant (14–27%), and triple wild-type [4]. Among these, NF1-mutant melanoma represents a particularly challenging subgroup. Patients with NF1-mutant melanoma have been reported to have worse disease-specific and overall survival in some studies [5], although this finding has not been consistently replicated. They also respond poorly to MEK inhibition [6,7,8], which has shown only modest activity in RAS-pathway-driven melanomas. While ICIs are effective in a subset of NF1-mutant patients (37% in one cohort) [9], those who do not respond have no effective targeted treatment alternatives [10].

1.1 NF1 Biology and the EGFR Connection

NF1 encodes neurofibromin, a GTPase-activating protein that negatively regulates RAS by promoting the conversion of active RAS-GTP to inactive RAS-GDP [11]. Loss of NF1 leads to hyperactivation of the RAS-RAF-MAPK and PI3K-AKT pathways, driving cell proliferation, survival, and tumor progression [12]. While this mechanism is well-established, it has been unclear why NF1-mutant melanoma is less sensitive to MEK inhibition than other RAS-pathway-driven melanomas.

A recent study by Ibrahim et al. addressed this question directly [13]. Using 32 patient-derived short-term cultures (STCs) and multi-omics approaches (RNA-seq, ATAC-seq, H3K4me3, H3K27Ac), they discovered that NF1 loss leads to transcriptional upregulation of EGFR, resulting in EGFR-driven ERK and AKT hyperactivation. Critically, they demonstrated that

EGFR is highly expressed and active in NF1-mutant melanoma cells (P = 0.0007 by IHC); that genetic or pharmacological EGFR inhibition blocks NF1-mutant melanoma growth; that EGFR inhibitors (cetuximab, afatinib) shrink NF1-mutant xenografts; that NF1-mutant cells have decreased IC50 for erlotinib (EGFR inhibitor) and increased IC50 for trametinib (MEK inhibitor); and that high EGFR expression correlates with worse survival in TCGA-SKCM. Ibrahim et al. concluded that EGFR inhibition represents a rational therapeutic strategy for NF1-mutant melanoma, irrespective of co-occurring BRAF or NRAS mutations. However, they explicitly acknowledged a key limitation that they only used patient-derived models, which could not account for potential effects on and interactions with the immune microenvironment.

1.2 The Immune Context: A Rapidly Evolving Picture

In a companion spatial multi-omics study, Ibrahim et al. profiled 42 melanoma cores by Xenium spatial transcriptomics (1,003,604 cells) and 56 tissues by 45-plex CODEX spatial proteomics (1,052,420 cells) [17]. They identified 12 recurrent meta-niches (MNs) and showed that NF1-mutant melanomas are enriched in immunosuppressive MNs dominated by cancer-associated fibroblasts (CAFs), M2-like macrophages, and regulatory T cells, and depleted of MNs enriched in cytotoxic $CD8^+$ T cells. NF1 loss was associated with impaired antigen-presentation programs (reduced HLA, TAP1, TAP2, CANX, CIITA) and reduced $CD8^+$ T cell infiltration. Critically, EGFR signaling emerged as a central feature of NF1-deficient melanoma ecosystems: EGFR expression correlated inversely with $CD8^+$ T cell abundance ($r_s = -0.622$) and with antigen-presentation genes, and was highest in tumors that did not respond to anti-PD-1 therapy. EGFR inhibition restored antigen-presentation programs and enhanced antitumor immune responses in a syngeneic NF1-knockdown model resistant to ICIs.

In parallel, an earlier study from Jour et al. characterized the genomic and transcriptomic landscape of NF1-mutant melanoma in 112 patients [10]. That work identified a proliferative, cell-cycle-driven program (MKI67, CDC20, CEP55), LY6E upregulation in tumor cells, and VEGFA upregulation in the microenvironment, and proposed CDC20 inhibition combined with MEK inhibition as a rational targeted regimen. It did not, however, provide a bulk-tissue prognostic signature, nor did it address the immune context of the EGFR-driven phenotype.

1.4 The Research Gap

Despite this body of work, three gaps remain. First, no bulk-tissue signature specifically captures the EGFR-driven, immune-cold phenotype of NF1-mutant melanoma in a form computable from standard RNA-seq and testable across independent cohorts. The Ibrahim 2025 study is preclinical, the Ibrahim 2026 study is spatial and single-cohort, and the Jour 2023 study is targeted-therapy-focused; none provides such a tool. A bulk-computable signature would allow this biology to be interrogated in the thousands of existing melanoma cohorts and could serve as an enrichment tool for eventual patient selection. Second, whether a bulk-tissue signature of EGFR-driven biology tracks with NF1 mutation status across independent cohorts remains unresolved. NF1 mutations are associated with aggressive disease, but their prognostic value alone is modest, and whether a signature capturing EGFR-driven biology tracks with NF1 mutation status or reads out a broader phenotype is unknown. Third, many published melanoma signatures are neither validated across multiple independent cohorts nor grounded in a specific biological mechanism [19,20,21]. A prognostic signature grounded in NF1/RAS biology and tested across multiple cohorts would therefore be a useful hypothesis-generating tool.

1.5 Our Approach and Rationale

We hypothesized that a gene expression signature derived from NF1/RAS pathway genes would capture the EGFR-driven biology described by Ibrahim et al. (2025, 2026) at the bulk-tissue level; be associated with survival in human tumors, independent of stage; reveal the immune context of EGFR-driven tumors — a question the 2025 study could not address and the 2026 study addressed spatially; and show directional consistency across independent cohorts, as a first step toward validation. To test this, we built a base gene set from the KEGG Ras signaling pathway (which includes NF1) plus literature-curated melanoma genes; used WGCNA to identify genes co-expressed with NF1 in TCGA-SKCM; applied LASSO Cox regression to select a minimal prognostic signature; assessed the signature in four independent GEO cohorts; characterized the biology using GSVA, immune deconvolution, and single-cell RNA-seq; and tested associations with NF1 mutation status and clinical features. Our goal was to provide a bulk-tissue, hypothesis-generating signature that operationalizes the EGFR-driven, immune-cold biology established by the NYU group and could serve as an enrichment tool for future biomarker-driven trials.

2. METHODS

2.1 Gene Set Construction

The KEGG 2021 Human "Ras signaling pathway" gene set was downloaded from Enrichr [14]. Fifteen literature-curated melanoma genes (CDC20, MKI67, LY6E, PTPN11, RASA2, RASSF2, CCND1, CDK4, CDK2, EGFR, KIT, VEGFA, TFEB, TFE3, SYT1) were added [22,23,24];

several of these were informed by the prior NF1-mutant melanoma literature [18]. The union yielded 242 unique genes **(Supplementary Table S2)**

2.2 TCGA-SKCM Data Processing

TCGA-SKCM STAR-counts data (TCGA-SKCM.star_counts.tsv) ; log2(count+1)-transformed) and clinical data (TCGA-SKCM.clinical.tsv) were obtained from the UCSC Xena browser https://xenabrowser.net/ . Ensembl IDs were converted to gene symbols using org.Hs.eg.db. Duplicate symbols were collapsed by highest mean expression. Low-expression genes (mean ≤ 1) were removed. The final matrix contained 18,432 genes across 471 samples.

2.3 WGCNA

Weighted Gene Co-expression Network Analysis was performed on the top 5,000 most variable genes (plus NF1) using the WGCNA R package [15]. A signed network was constructed with soft-thresholding power selected by scale-free topology fit ($R^2 > 0.85$). Modules were detected using dynamic tree cutting (minClusterSize = 30) and merged at cutHeight = 0.15. Module-trait correlation was computed between module eigengenes and NF1 expression **(Supplementary Table S1)**. Genes with module membership (kME) > 0.4 in significant modules ($|r| > 0.3$, $p < 0.05$) were intersected with the base gene set to yield the final candidate pool **(Supplementary Table S3)**.

2.4 LASSO Cox Regression

Univariate Cox regression was performed for each candidate gene. Genes with $p < 0.1$ (or top 20 if fewer than 5) entered LASSO. Ten-fold cross-validated LASSO Cox regression (glmnet, alpha

= 1) selected the final signature at lambda.min. Risk scores were computed as the linear predictor (link). Patients were stratified by median risk score into High and Low Risk groups.

2.5 Survival Analysis

Kaplan-Meier curves were generated with survminer; log-rank tests assessed significance. Time-dependent ROC (timeROC) computed 1-, 2-, and 3-year AUCs. Multivariate Cox regression adjusted for age and stage. A nomogram was built with rms; calibration was assessed by bootstrap (B = 200). Decision curve analysis (DCA) evaluated clinical utility.

2.6 GSVA and Immune Deconvolution

GSVA (or ssGSEA) scored KEGG pathway activity per sample. Differential activity between High and Low Risk groups was tested by t-test with Benjamini-Hochberg FDR correction. Immune cell infiltration was quantified using 15 custom gene sets (CD8 T cells, CD4 T cells, Tregs, NK cells, B cells, monocytes, M1/M2 macrophages, dendritic cells, mast cells, neutrophils, endothelial cells, fibroblasts, PD-1 checkpoint, cytolytic activity) via ssGSEA.

2.7 External Cohorts

Four GEO cohorts were used: GSE65904 (n = 210), GSE54467 (n = 79), GSE22153 (n = 54), GSE19234 (n = 44). Platform annotation (GPL) files were used to map probes to gene symbols. Expression data were quantile-normalized. Risk scores were computed using TCGA-derived coefficients applied to z-scored expression. Cox regression and Kaplan-Meier analysis were performed per cohort. Meta-analysis used random-effects (REML) and Egger's test.

2.8 Single-Cell RNA-Seq

GSE72056 (Tirosh et al., 2016) was downloaded and parsed [16]. Cells with ≥ 200 detected genes were retained. Expression of signature genes was quantified as mean expression and percent expressed per gene.

2.9 NF1 Mutation Analysis

NF1 mutation annotation was obtained from cBioPortal's PanCancer Atlas melanoma study (skcm_tcga_pan_can_atlas_2018). The raw file contained 471 rows across 448 unique patients, including both -01 (normal) and -06 (tumor) aliquots; aliquots were collapsed by 12-character patient barcode. Variants were classified as Wild-Type, Not Profiled, Missense, Nonsense, Frameshift, or Splice, and binarized as Mutant (any non-WT, non-NP call) vs. Wild-Type. After intersecting with our 421-patient expression cohort, 72 patients (17.1%) harbored NF1 mutations.

Detailed results for all analyses, including gene set construction, WGCNA, LASSO Cox regression, GSVA, immune deconvolution, NF1 mutation analysis, single-cell expression, external validation, and meta-analysis, are provided in Supplementary Tables S1–S19.

2.10 Statistical Analysis

All analyses were performed in R 4.4.3. Two-sided $p < 0.05$ was considered significant. FDR correction used the Benjamini-Hochberg method.

3. RESULTS

3.1 WGCNA

WGCNA identified four modules. The turquoise module was significantly correlated with NF1 expression ($r = 0.353$, $p = 2.48 \times 10^{-15}$), while other modules showed weak or no correlation (grey: $r = 0.099$, $p = 0.032$; blue: $r = -0.070$, $p = 0.131$; brown: $r = 0.003$, $p = 0.952$) **(Table 1; Supplementary Table S1)**. Intersecting turquoise module genes (kME > 0.4) with the 242-gene base set yielded 27 candidate genes **(Supplementary Table S3).**

3.2 LASSO Cox Regression

Univariate Cox regression identified 19 genes with $p < 0.1$, all of which entered LASSO **(Supplementary Table S17)**. LASSO selected 8 genes at lambda.min (Figure 1A, B). Seven genes were protective (negative coefficients): PLA2G2D (−0.0533), FASLG (−0.0100), RASGRP1 (−0.0194), FLT3 (−0.0221), FGF1 (−0.0328), PLA2G5 (−0.0838), FGF2 (−0.0569). One gene was a risk factor: EGFR (+0.1179) **(Table 2; Supplementary Table S14).**

The eight genes fall into three broad categories: EGFR, the sole risk-factor gene and the direct mechanistic link to the NF1/EGFR axis described by Ibrahim et al. (2025); immune-related genes (FASLG, RASGRP1) consistent with the immune-cold phenotype described by Ibrahim et al. (2026); and growth factor/phospholipase genes (FGF1, FGF2, FLT3, PLA2G2D, PLA2G5) whose connection to NF1/EGFR biology is less direct. We acknowledge that this heterogeneity means the signature may capture a broader aggressive/immune-cold phenotype rather than being a pure readout of EGFR-driven biology. The dominance of EGFR (coefficient +0.1179, nearly double the next largest) nonetheless anchors the signature to the EGFR axis.

3.3 The Signature Stratifies Patients by Survival

High Risk patients (above median risk score) had significantly worse overall survival than Low Risk patients (log-rank $p = 1.41\times10^{-10}$; Figure 8.1). Median survival was 4.0 years vs. 12.35 years. Time-dependent AUCs were 0.716 (1-year), 0.733 (2-year), and 0.712 (3-year) (Figure 8.1).

3.4 Multivariate Cox regression

Multivariate Cox regression adjusted for age and stage confirmed the risk score as an independent predictor (HR = 5.32, 95% CI 3.55–7.99, $p = 6.8\times10^{-16}$) **(Table 3; Supplementary Tables S4, S11)**. Stage III (HR = 1.94, p = 0.006) and Stage IV (HR = 4.38, $p < 0.001$) were also significant. The C-index was 0.72. Notably, the risk score's HR increased from 4.67 (univariate; **Supplementary Table S5**) to 5.32 (multivariate), suggesting that the signature captures information partially masked by age and stage.

3.5 GSVA Analysis

GSVA revealed that all top 15 differentially active KEGG pathways were immune-related and downregulated in High-Risk patients (**Figure 4A, B**.; **Supplementary Table S6**). The most significant pathways included Th17 cell differentiation ($FDR = 5.8\times10^{-36}$), Toll-like receptor signaling ($FDR = 5.1\times10^{-33}$), T cell receptor signaling ($FDR = 1.3\times10^{-32}$), B cell receptor signaling ($FDR = 1.3\times10^{-32}$), NK cell cytotoxicity ($FDR = 2.0\times10^{-32}$), and antigen processing and presentation ($FDR = 1.4\times10^{-31}$). PD-L1/PD-1 checkpoint pathway was also downregulated ($FDR = 1.0\times10^{-28}$).

Immune deconvolution confirmed these findings: CD8 T cells (FDR = 3.3×10^{-34}), CD4 T cells (FDR = 8.2×10^{-36}), NK cells (FDR = 7.8×10^{-32}), B cells (FDR = 2.3×10^{-22}), and cytolytic activity (FDR = 1.5×10^{-29}) were all significantly lower in High-Risk tumors Figure **5A**; **Supplementary Table S7**). Our bulk-tissue signature therefore recapitulates, in a form computable from standard RNA-seq, the immune-cold phenotype that Ibrahim et al. (2026) demonstrated spatially in NF1-mutant melanoma [17]. The concordance between our bulk signature and their spatial meta-niche analysis supports the robustness of the EGFR/immune-cold axis and provides a clinically deployable readout. It is important to note that this finding does not establish causal direction: bulk transcriptomics cannot distinguish immune exclusion from immune dysfunction, and the single-cell analysis below provides only descriptive localization.

3.6 Single-Cell Analysis

Single-cell RNA-seq (GSE72056; 4,645 cells) revealed that the protective genes were expressed predominantly in immune cells: RASGRP1 in 26.35% of cells and FASLG in 9.84% of cells (**Figure 7A, B**.; **Supplementary Tables S16, S19**). In contrast, EGFR was expressed in only 2.71% of cells, consistent with a tumor-intrinsic origin. These findings indicate that the protective arm of the signature reflects immune infiltration, while the risk arm reflects tumor-intrinsic EGFR biology

3.7 External Cohorts Validation

In four GEO cohorts, the signature was directionally consistent in three (HR > 1) and statistically significant in one (GSE54467: HR = 6.71, p = 0.025) **(Table 4; Supplementary Table S18)**. Meta-analysis yielded a pooled HR of 2.25 (95% CI 1.03–4.89, p = 0.041; I^2 = 7.6%) (**Figure**

9D; **Supplementary Table S8**). C-indices in external cohorts ranged from 0.53 to 0.60, lower than the discovery cohort (C-index = 0.72). This attenuation is consistent with the platform, population, and treatment-era heterogeneity that characterizes public melanoma cohorts, and with the known difficulty of transporting bulk signatures across datasets. Egger's test showed no evidence of publication bias (p = 0.971; **Supplementary Table S9**), though the test was underpowered (k = 4). Leave-one-out analysis is provided in **Supplementary Table S10**. These findings support the biological relevance of the signature.

3.8 Relationship of Signature with NF1 Mutation Status

NF1 mutations were present in 72/421 (17.1%) patients in our primary expression cohort (**Figure 6A-D**; **Supplementary Tables S12, S13**). Risk scores did not differ significantly between NF1-mutant and wild-type tumors (t-test p = 0.122; Figure 6A; **Supplementary Table S13**), nor did overall survival (log-rank p = 0.111; Figure 6B). The NF1-mutant subgroup was biologically heterogeneous: 38 patients carried truncating driver variants (nonsense, frameshift, or splice-site), while 34 carried missense variants (Figure 6C; **Supplementary Table S12**). Truncating variants showed a trend toward worse survival than missense variants, though this did not reach significance (log-rank p = 0.274). Signature gene expression by NF1 mutation status is provided in **Supplementary Table S15** (Figure 6D).

Given the small size (n=72) and biological heterogeneity of the NF1-mutant subgroup, this analysis was underpowered to detect a modest association. The non-significant result should be interpreted as insufficient evidence rather than evidence of absence.

3.9 Clinical Correlations

High Risk patients were slightly older ($p = 0.025$) but did not differ by sex ($p = 0.25$). Risk scores varied by stage (ANOVA $p = 4.1\times10^{-6}$) and were significantly higher in dead vs. alive patients ($p = 1.4\times10^{-7}$) (**Figure 2A-D**.).

3.10 Decision Curve Analysis

DCA showed that the full model (risk score + stage + age) provided net benefit over "treat all" and "treat none" strategies at threshold probabilities between 0.05 and 0.20 (**Figure 3D**).

4. DISCUSSION

We present a bulk-tissue 8-gene signature that is biologically anchored to the EGFR-driven mechanism described by Ibrahim et al. (2025), stratifies overall survival in TCGA-SKCM, remains statistically significant after adjustment for stage and age, and captures an immune-cold microenvironment concordant with the spatial findings of Ibrahim et al. (2026). The signature is dominated by EGFR, the only risk-factor gene, which aligns with the discovery that NF1 loss drives EGFR upregulation and EGFR-dependent growth. It did not correlate with NF1 mutation status in our cohort, though this analysis was underpowered, suggesting that it may read out EGFR-driven biology rather than NF1 loss per se. External validation was directionally consistent across cohorts, though statistical significance was limited by cohort size and platform heterogeneity. We therefore position this work as a hypothesis-generating bridge between preclinical and spatial mechanism and bulk-tissue patient biology.

The dominance of EGFR in our signature (coefficient = +0.1179, almost double the next largest coefficient) is consistent with the findings of Ibrahim et al 2025., who demonstrated that EGFR is transcriptionally upregulated in NF1-mutant melanoma cells (among 1,168 upregulated transcripts), that its regulatory elements are more accessible (ATAC-seq) and active (H3K27Ac), that EGFR protein is overexpressed in patient tissues (P = 0.0007), and that EGFR inhibition blocks NF1-mutant melanoma growth in vitro and in vivo. Our observation that EGFR is the sole risk-factor gene in a bulk-tissue signature offers bulk-tissue, patient-level support for their mechanistic discovery: what they found mechanistically is reflected in the bulk transcriptome of patient tumors.

The remaining seven genes are biologically heterogeneous: FASLG and RASGRP1 are immune-related, and their protective coefficients are consistent with the immune-cold phenotype, in which loss of immune infiltration is associated with worse survival; FGF1, FGF2, and FLT3 are growth factor/receptor tyrosine kinase genes that may reflect stromal or autocrine signaling in the tumor microenvironment; and PLA2G2D and PLA2G5 are phospholipase genes whose connection to NF1/EGFR biology is less clear. This heterogeneity means the signature should be interpreted as capturing a broader aggressive/immune-cold phenotype anchored by EGFR, rather than as a pure EGFR-pathway signature.

Our most notable biological finding is that the EGFR-driven signature is associated with an immune-cold tumor microenvironment. All top 15 differentially active KEGG pathways were immune-related and downregulated in High-Risk patients, and immune deconvolution confirmed depletion of CD8 T cells, CD4 T cells, NK cells, B cells, and cytolytic activity. This is significant for three reasons. First, it provides a bulk-tissue, multi-cohort readout of the immune-cold phenotype that Ibrahim et al. (2026) demonstrated spatially; the concordance between an

independently derived bulk signature and their spatial meta-niche analysis supports the robustness of the EGFR/immune-cold axis. Second, it is consistent with the observation that a substantial fraction of NF1-mutant melanomas do not respond to immunotherapy [9, 10], though responses are heterogeneous, raising the hypothesis that EGFR-driven tumors may be intrinsically resistant to ICIs, which would need to be tested prospectively. Third, it supports the combination therapy hypothesis advanced by Ibrahim et al. (2025, 2026), in which EGFR inhibitors such as cetuximab or afatinib could kill tumor cells while immunotherapy reactivates the immune response; our signature could serve as a bulk-tissue enrichment tool for testing such strategies in existing cohorts. We emphasize that this is not the first patient-level evidence of the immune-cold phenotype, since the spatial work of Ibrahim et al. (2026) provides that, but it is a bulk-computable, multi-cohort operationalization applicable to existing RNA-seq datasets.

Our single-cell analysis revealed that the signature has two distinct arms. The protective arm, including RASGRP1 and FASLG, was expressed predominantly in immune cells (expressed in 26.35% and 9.84% of all cells analyzed, respectively, predominantly within immune cell clusters) and appears to mark immune infiltration. The risk arm, EGFR, was expressed in only 2.71% of cells and, based on its known biology and the absence of immune-cell expression, is likely tumor-intrinsic. The remaining genes (FGF1, FGF2, FLT3, PLA2G2D, PLA2G5) were not the focus of the single-cell analysis because their expression was low or their cell-type localization was ambiguous. Their contribution may reflect stromal/CAF biology, which is also a component of the immune-cold meta-niches described by Ibrahim et al. (2026), and this should be explored in future work. This dual nature may explain why the signature performs as a prognostic tool: it simultaneously captures tumor-intrinsic aggression and host immune response.

The signature did not correlate with NF1 mutation status (t-test p = 0.122; Fisher p = 0.195; log-rank p = 0.111) in the primary 421-patient cohort. This may seem surprising but is consistent with the mixed literature on NF1 mutation and prognosis. The NF1-mutant subgroup is small and biologically heterogeneous: of 72 mutants, only 38 carried truncating drivers, while 34 carried missense variants of varying significance. Our survival analysis by variant type showed a trend toward steeper survival decline for truncating variants than missense variants, which behaved similarly to wild-type, though this did not reach significance (log-rank p = 0.274), which may dilute any true association when pooled into a single mutant category. In addition, Ibrahim et al. (2025) identified pathogenic NF1 mutations across multiple domains, suggesting functional heterogeneity that a binary classification cannot capture. They also showed that EGFR is important across melanoma subtypes, not just NF1-mutants, and their spatial work [17] showed that EGFR expression, rather than NF1 mutation status per se, correlates with $CD8^+$ T cell exclusion. Our signature appears to capture this broader EGFR-driven biology and should be understood as an EGFR/immune signature rather than an NF1-mutation surrogate.

In external cohorts, the signature was directionally consistent in three of four cohorts (HR > 1) and statistically significant in one (GSE54467: HR = 6.71, p = 0.025). Meta-analysis yielded a pooled HR of 2.25 (95% CI 1.03–4.89, p = 0.041; I^2 = 7.6%), while C-indices in external cohorts ranged from 0.53 to 0.60, lower than the discovery cohort (C-index = 0.72). This attenuation is consistent with the platform, population, and treatment-era heterogeneity that characterizes public melanoma cohorts, and with the known difficulty of transporting bulk signatures across datasets; however, it also highlights the need for prospective validation in standardized cohorts. This is not unusual in cancer genomics, where cohorts differ by platform (Illumina, Affymetrix), population (some with acral melanoma), and treatment era (pre- versus post-immunotherapy).

The consistency of direction across cohorts, together with the biological concordance with the spatial findings of Ibrahim et al. (2026), supports the relevance of the signature.

Several melanoma prognostic signatures have been published, but many are neither biologically interpretable nor externally validated [19,20,21]. Our signature is distinctive in being grounded in NF1/RAS biology rather than agnostic data-mining from all genes, in having a clear biological interpretation (EGFR-driven, immune-cold), in being tested in four independent cohorts with the caveats noted above, and in being mechanistically supported by the Ibrahim et al. preclinical (2025) and spatial (2026) work. Relative to the three NYU papers, our contribution is best understood as the bulk-tissue, multi-cohort, prognostic layer. Jour et al. (2023) provided an NF1-mutant proliferation and immune signature (CDC20, MKI67, LY6E, VEGFA) and proposed CDC20 plus MEK combination therapy; we add a prognostic rather than purely proliferative signature, an immune-cold readout, and multi-cohort testing. Ibrahim et al. (2025) established NF1 loss as a driver of EGFR dependency and showed that EGFR inhibitors are active in preclinical models; we add a bulk-tissue prognostic readout of EGFR-driven biology and its immune context. Ibrahim et al. (2026) described the spatial immune-cold phenotype, showed that EGFR correlates with CD8 exclusion, and provided the rationale for EGFR inhibitor plus anti-PD-1 therapy; we add a bulk-computable, multi-cohort, survival-associated operationalization of that biology.

If validated, our signature could eventually be tested for stratifying patients by risk for intensified surveillance or adjuvant therapy, for enriching EGFR-driven tumors that may benefit from EGFR inhibitors as proposed by Ibrahim et al., for generating hypotheses about immunotherapy resistance in immune-cold High-Risk patients, and for informing the design of

combination therapy trials. Ibrahim et al. noted that patients with NF1-mutant melanoma have few treatment options if their tumors do not respond to ICIs [13,22,23,24]. A validated version of our signature could eventually help identify such patients, though prospective testing would be needed to determine whether it can guide treatment decisions. However, prospective validation, analytical validation, and predictive biomarker testing would all be required before clinical use. For this signature to move from hypothesis-generating to clinically actionable, several steps would be required: analytical validation to lock the assay platform and establish reproducibility across labs; prospective clinical validation in a cohort with standardized treatment and immunotherapy annotation; biological validation confirming that signature-High tumors are EGFR-dependent using functional assays such as EGFR inhibitor sensitivity in patient-derived models stratified by signature score; predictive validation testing whether the signature predicts response to EGFR inhibitors in a biomarker-stratified trial, as motivated by the work of Ibrahim et al.; and immune validation testing whether signature-High tumors are resistant to ICIs in a cohort with immunotherapy outcomes.

5. LIMITATIONS

This is a discovery and translation study that establishes a hypothesis-generating tool and enrichment biomarker, not a validated clinical assay. The principal limitation is that the signature has not yet been experimentally verified: all findings are computational and retrospective, derived in silico from TCGA-SKCM, four GEO cohorts, and one single-cell dataset, without laboratory confirmation. External validation was directionally consistent across cohorts, though statistical significance was constrained by cohort size and platform heterogeneity. Similarly, the absence of a significant association with NF1 mutation status reflects the limited power of the

NF1-mutant subgroup (n = 72) rather than evidence against the signature; larger sample sizes and prospective studies are required to confirm its validity.

6. FUTURE DIRECTIONS

Confirming the utility of this signature requires a defined sequence of laboratory and clinical validation. The logical first step is immunohistochemistry: staining patient tumor tissues for the protein products of the signature genes would confirm that the transcript-level signature is reflected at the protein level, and that the two arms of the signature — immune-localized protective genes and tumor-intrinsic EGFR — are spatially distinct in situ, as the single-cell data suggest. Functional studies should then move from correlation to causation on both arms. On the protective arm, RASGRP1 and FASLG should be knocked down in primary T cells to test whether loss of either gene impairs anti-tumor cytotoxicity and abolishes the survival benefit associated with the protective arm. On the risk arm, patient-derived models stratified by signature score should be tested for EGFR inhibitor sensitivity, with the prediction that signature-High tumors are preferentially EGFR-dependent. Demonstrating that the protective arm is causally immunologically active and the risk arm is causally EGFR-dependent would convert the signature from a statistical association into a mechanistically validated biomarker. The final step is a prospective cohort with standardized treatment and immunotherapy annotation to confirm its clinical validity and transportability.

7. CONCLUSION

We present a bulk-tissue 8-gene signature that connects the EGFR-driven mechanism described by Ibrahim et al. (2025) to patient-level biology in cutaneous melanoma. The signature stratifies survival in TCGA-SKCM, is independent of stage, and captures an immune-cold microenvironment concordant with the spatial findings of Ibrahim et al. (2026) — a finding that extends the preclinical work by addressing, at the bulk level, the immune limitation the 2025 study explicitly acknowledged. The signature does not correlate with NF1 mutation status, suggesting it reads out EGFR-driven biology more broadly. External validation is directionally consistent but statistically fragile.

Together with the mechanistic (Ibrahim et al. 2025), spatial (Ibrahim et al. 2026), and targeted-therapy (Jour et al. 2023) studies, our signature provides the bulk-tissue, multi-cohort prognostic layer. It is a hypothesis-generating tool that could enrich for the EGFR-driven, immune-cold tumors these studies describe — and it is computable from standard RNA-seq, making it immediately testable in existing cohorts. Bridging mechanistic discovery and clinical application will require prospective validation, functional testing, and predictive biomarker trials — a roadmap we outline here.

Declarations

Acknowledgements

Not applicable.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The author received no financial support for the design, conduct, analysis, or preparation of this manuscript.

Data availability

All datasets analyzed in this study are publicly available. No new primary data were generated.

- **TCGA-SKCM expression and clinical data:** UCSC Xena browser, https://xenabrowser.net/ (files: TCGA-SKCM.star_counts.tsv, TCGA-SKCM.clinical.tsv).
- **NF1 mutation annotation:** cBioPortal, skin cutaneous melanoma (TCGA, PanCancer Atlas 2018), https://www.cbioportal.org/ (file: skcm_tcga_pan_can_atlas_2018).
- **External validation cohorts:** NCBI Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/ — GSE65904 (n = 210), GSE54467 (n = 79), GSE22153 (n = 54), GSE19234 (n = 44).
- **Single-cell RNA-seq:** GSE72056 (Tirosh et al., 2016), NCBI GEO.
- **Base gene set:** KEGG 2021 Human "Ras signaling pathway," via Enrichr, https://maayanlab.cloud/Enrichr/.

Code availability: The analysis code (R scripts for WGCNA, LASSO Cox regression, GSVA, immune deconvolution, survival analysis, and meta-analysis) is available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable. This study is a secondary analysis of publicly available, de-identified datasets and did not involve human participants, human tissue, or animals. No ethical approval or informed consent was required. All original data sources were accessed in accordance with their respective terms of use and data access policies.

Consent for publication

Not applicable. This manuscript contains no individual person's data, images, or identifiable information.

Competing interests

The author declares no competing interests.

REFERENCES

[1] Rebecca L. Siegel, Angela N. Giaquinto, Ahmedin Jemal, Cancer statistics, 2024, *CA Cancer J Clin*. 74 (2024) 12–49. https://doi.org/10.3322/caac.21820.

[2] Brendan D. Curti, Mark B. Faries, Recent Advances in the Treatment of Melanoma, *N Engl J Med*. 384 (2021) 2229–2240. https://doi.org/10.1056/NEJMra2034861.

[3] Jason J. Luke, Keith T. Flaherty, Antoni Ribas, Georgina V. Long, Targeted agents and immunotherapies: optimizing outcomes in melanoma, *Nat Rev Clin Oncol*. 14 (2017) 463–482. https://doi.org/10.1038/nrclinonc.2017.43.

[4] Rehan Akbani, Kadir C. Akdemir, B. Arman Aksoy, Genomic Classification of Cutaneous Melanoma, *Cell*. 161 (2015) 1681–1696. https://doi.org/10.1016/j.cell.2015.05.044.

[5] Helena Cirenajwis, Martin Lauss, Henrik Ekedahl, NF1-mutated melanoma tumors harbor distinct clinical and biological characteristics, *Mol Oncol*. 11 (2017) 438–451. https://doi.org/10.1002/1878-0261.12050.

[6] Maija Kiuru, Klaus J. Busam, The NF1 gene in tumor syndromes and melanoma, *Lab Invest*. 97 (2017) 146–157. https://doi.org/10.1038/labinvest.2016.142.

[7] Ophélia Maertens, Bryan Johnson, Pablo Hollstein, Elucidating distinct roles for NF1 in melanomagenesis, *Cancer Discov*. 3 (2013) 338–349. https://doi.org/10.1158/2159-8290.CD-12-0313.

[8] Kari B. Wisinski, Yael Flamand, Melissa A. Wilson, Trametinib in Patients With NF1-, GNAQ-, or GNA11-Mutant Tumors: Results From the NCI-MATCH ECOG-ACRIN Trial (EAY131) Subprotocols S1 and S2, *JCO Precis Oncol*. 7 (2023) e2200421. https://doi.org/10.1200/PO.22.00421.

[9] Carl M. Thielmann, Eleftheria Chorti, Johanna Matull, NF1-mutated melanomas reveal distinct clinical characteristics depending on tumour origin and respond favourably to immune checkpoint inhibitors, *Eur J Cancer*. 159 (2021) 1–13. https://doi.org/10.1016/j.ejca.2021.09.035.

[10] George Jour, Irineu Illa-Bochaca, Milad Ibrahim, Genomic and Transcriptomic Analyses of NF1-Mutant Melanoma Identify Potential Targeted Approach for Treatment, *J Invest Dermatol*. 143 (2023) 444–455.e8. https://doi.org/10.1016/j.jid.2022.07.022.

[11] M. M. Daston, N. Ratner, Neurofibromin, a predominantly neuronal GTPase activating protein in the adult, is ubiquitously expressed during development, *Dev Dyn*. 195 (1992) 216–226. https://doi.org/10.1002/aja.1001950307.

[12] Moriah H. Nissan, Christine A. Pratilas, Alexis M. Jones, Loss of NF1 in cutaneous melanoma is associated with RAS activation and MEK dependence, *Cancer Res*. 74 (2014) 2340–2350. https://doi.org/10.1158/0008-5472.CAN-13-2625.

[13] Milad Ibrahim, Irineu Illa-Bochaca, George Jour, NF1 Loss Promotes EGFR Activation and Confers Sensitivity to EGFR Inhibition in NF1-Mutant Melanoma, *Cancer Res*. (2025). https://doi.org/10.1158/0008-5472.CAN-24-3904.

[14] Edward Y. Chen, Christopher M. Tan, Yan Kou, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, *BMC Bioinformatics*. 14 (2013) 128. https://doi.org/10.1186/1471-2105-14-128.

[15] Peter Langfelder, Steve Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinformatics*. 9 (2008) 559. https://doi.org/10.1186/1471-2105-9-559.

[16] Itay Tirosh, Benjamin Izar, Sanjay M. Prakadan, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq, *Science*. 352 (2016) 189–196. https://doi.org/10.1126/science.aad0501.

[17] Milad Ibrahim, Irineu Illa-Bochaca, Tara Muijlwijk, Genotype-driven tumor ecosystems drive immune evasion and immunotherapy resistance in melanoma, *Mol Cancer*. (2026). https://doi.org/10.1186/s12943-026-02781-9.

[18] George Jour, Irineu Illa-Bochaca, Milad Ibrahim, Genomic and Transcriptomic Analyses of NF1-Mutant Melanoma Identify Potential Targeted Approach for Treatment, *J Invest Dermatol*. 143 (2023) 444–455.e8. https://doi.org/10.1016/j.jid.2022.07.022.

[19] Kaushala Jayawardana, Sarah-Jane Schramm, Varsha Tembe, Identification, Review, and Systematic Cross-Validation of microRNA Prognostic Signatures in Metastatic Melanoma, *J Invest Dermatol*. (2015). https://doi.org/10.1038/JID.2015.355.

[20] Sarah-Jane Schramm, Anna E. Campain, Richard A. Scolyer, Review and cross-validation of gene expression signatures and melanoma prognosis, *J Invest Dermatol*. 132 (2012) 274–283. https://doi.org/10.1038/jid.2011.305.

[21] C. Piyawajanusorn, G. Ghislat, P. J. Ballester, Deconstructing biomarker generalisation failure in anti-PD1 cancer immunotherapy response: a multi-cohort framework, *Brief Bioinform*. 26 (2025) i8–i9. https://doi.org/10.1093/bib/bbaf631.010.

[22] George Jour, Irineu Illa-Bochaca, Milad Ibrahim, Genomic and Transcriptomic Analyses of NF1-Mutant Melanoma Identify Potential Targeted Approach for Treatment, *J Invest Dermatol*. 143 (2023) 444–455.e8. https://doi.org/10.1016/j.jid.2022.07.022.

[23] Helena Cirenajwis, Martin Lauss, Henrik Ekedahl, NF1-mutated melanoma tumors harbor distinct clinical and biological characteristics, *Mol Oncol*. 11 (2017) 438–451. https://doi.org/10.1002/1878-0261.12050.

[24] Eleazar Vega-Saenz de Miera, Kevin M. Giles, Yu-Hsin Yeh, CDC20 is a novel potential therapeutic target in NF1-mutant melanoma, *J Clin Oncol*. 38 (2020) e22075. https://doi.org/10.1200/JCO.2020.38.15_suppl.e22075.

Supplementary Tables

Supplementary Table S1. WGCNA module–trait correlations with NF1 expression. (Source file: module_trait_correlation.csv)

Supplementary Table S2. Base gene set (242 genes) with source annotation (KEGG_Ras, Literature, Both). (Source file: NF1_base_set_transparent.csv)

Supplementary Table S3. Twenty-seven candidate genes after intersecting the WGCNA turquoise module (kME > 0.4) with the base gene set. (Source file: NF1_final_candidate_pool.csv)

Supplementary Table S4. Multivariate Cox regression for clinical variables and risk score (formatted for manuscript). (Source file: NF1_forest_multivariate.csv)

Supplementary Table S5. Univariate Cox regression for clinical variables and risk score (formatted for manuscript). (Source file: NF1_forest_univariate.csv)

Supplementary Table S6. GSVA pathway activity results for all KEGG pathways. (Source file: NF1_GSVA_results.csv)

Supplementary Table S7. ssGSEA immune cell infiltration results (15 cell types/signatures). (Source file: NF1_immune_infiltration_results.csv)

Supplementary Table S8. Meta-analysis results (primary and sensitivity analyses). (Source file: NF1_meta_analysis.csv)

Supplementary Table S9. Egger's regression test for publication bias. (Source file: NF1_meta_egger_test.csv)

Supplementary Table S10. Leave-one-out sensitivity analysis for the meta-analysis. (Source file: NF1_meta_leave_one_out.csv)

Supplementary Table S11. Multivariate Cox regression results (raw output). (Source file: NF1_multivariate_cox.csv)

Supplementary Table S12. Distribution of NF1 variant types (frameshift, missense, nonsense, splice, wild-type). (Source file: NF1_mutation_class_distribution.csv)

Supplementary Table S13. Summary of NF1 mutation rate and association tests (t-test, Fisher, log-rank). (Source file: NF1_mutation_summary.csv)

Supplementary Table S14. Final 8-gene signature coefficients and direction (protective vs. risk factor). (Source file: NF1_signature_coefficients.csv)

Supplementary Table S15. Expression of the 8 signature genes by NF1 mutation status. (Source file: NF1_signature_genes_by_mutation.csv)

Supplementary Table S16. Single-cell expression summary for signature genes (mean expression, percent expressed, coefficient, direction). (Source file: NF1_single_cell_expression_summary.csv)

Supplementary Table S17. Univariate Cox regression results for all 27 candidate genes. (Source file: NF1_univariate_cox.csv)

Supplementary Table S18. External cohort validation results (HR, C-index, log-rank p per cohort). (Source file: NF1_validation_summary.csv)

Supplementary Table S19. Single-cell expression summary (raw output). (Source file: singlecell_signature_results.csv)

Table 1. Module-trait correlation with NF1 expression.

Module	Correlation	P-value
MEturquoise	0.353	2.48×10^{-15}
MEgrey	0.099	0.032
MEblue	−0.070	0.131
MEbrown	0.003	0.952

Table 2. Final 8-gene signature coefficients.

Gene	Coefficient	Direction
PLA2G2D	−0.0533	Protective
FASLG	−0.0100	Protective
RASGRP1	−0.0194	Protective
FLT3	−0.0221	Protective
FGF1	−0.0328	Protective
PLA2G5	−0.0838	Protective
FGF2	−0.0569	Protective
EGFR	+0.1179	Risk Factor

Table 3. Multivariate Cox regression.

Variable	HR	95% CI	P-value
age	1.010	0.999–1.022	0.067
Stage II	1.037	0.602–1.784	0.896
Stage III	1.939	1.209–3.111	0.006
Stage IV	4.381	2.174–8.829	<0.001
risk_score	5.323	3.547–7.988	<0.001

Table 4. External cohort results.

Study	N	Events	HR	95% CI	P-value	C-index
GSE65904	210	102	1.86	0.56–6.20	0.315	0.527
GSE54467	79	48	6.71	1.28–35.27	0.025	0.604
GSE22153	54	47	2.69	0.46–15.68	0.271	0.593
GSE19234	44	24	0.70	0.11–4.67	0.717	0.537

FIGURE LEGENDS

Figure 1: Construction of the Prognostic Risk Model using LASSO Regression

(A) LASSO Cross-Validation: Partial likelihood deviance plot used to determine the optimal penalty parameter (λ) for the LASSO regression. The x-axis represents the log(λ), and the y-axis represents the partial likelihood deviance. The top axis shows the number of non-zero coefficients (genes) at each log(λ). The dotted vertical lines indicate the optimal λvalues selected by minimum criteria and 1-standard error criteria.

(B) LASSO Coefficient Paths: Visualization of the coefficient shrinkage as log(λ) increases. Each line represents a specific gene's coefficient trajectory. The vertical dotted line corresponds to the optimal λchosen in panel A, where a subset of genes remains with non-zero coefficients, forming the final risk signature.

Figure 2: Clinical Correlation of the Risk Score

(A) Age by Risk Group: Violin plots comparing patient age between the Low-Risk (green) and High-Risk (red) groups. A two-sample T-test indicates a statistically significant difference ($p = 0.025$), with the High-Risk group tending to be older.

(B) Risk Score by Sex: Violin plots showing the distribution of risk scores between female and male patients. The T-test indicates no significant difference in risk score based on sex ($p = 0.25$).

(C) Risk Score by Stage: Boxplots illustrating risk score distribution across different pathological stages (I-IV). One-way ANOVA reveals a highly significant difference across stages ($p = 4.1e-06$), suggesting the risk score correlates with disease progression.

(D) Risk Score by Survival Status: Violin plots comparing risk scores between patients who survived (Status 0, green) and those who died (Status 1, red). A significant T-test ($p = 1.4e-07$) indicates that higher risk scores are strongly associated with mortality.

Figure 3: Independent Prognostic Value and Clinical Utility

(A & B) Univariate and Multivariate Cox Regression Analysis: Forest plots and tables summarizing the Hazard Ratios (HR) for various clinical variables (age, stage, risk score). In both univariate (A) and multivariate (B) analyses, the risk score remains an independent prognostic factor ($HR > 4.6, p < 0.001$).

(C) Nomogram: A predictive nomogram integrating age, stage, and risk score to estimate 1-, 3-, and 5-year overall survival probabilities for individual patients. Points are assigned to each variable to calculate a total score.

(D) Decision Curve Analysis (DCA): A DCA plot comparing the standardized net benefit of different models (Risk Score, Stage, and a combined model) against "Treat All" and "Treat None" strategies for 1-year mortality. The combined model (green line) demonstrates the highest net benefit across a wide range of threshold probabilities.

Figure 4: Functional Enrichment Analysis

(A) Heatmap of Top 30 Pathways: A heatmap showing the relative enrichment (Z-score) of the top 30 KEGG pathways in High-Risk (red bar) versus Low-Risk (green bar) groups. Red indicates high enrichment, while blue/grey indicates low enrichment.

(B) Top 15 Differentially Active KEGG Pathways: A bar chart displaying the top 15 significantly downregulated pathways in the High-Risk group, ranked by $-\log_{10}(P\text{-value})$. Notably, pathways related to immune function (e.g., Th17 cell differentiation, Toll-like receptor signaling, T cell receptor signaling) are prominently suppressed in the High-Risk group.

Figure 5: Immune Microenvironment Infiltration

Violin plots comparing the enrichment scores of various immune cell types (e.g., CD8+ T cells, Tregs, Macrophages M1/M2, NK cells) and immune-related signatures (Immune checkpoint PD1, Cytolytic activity) between Low-Risk (green) and High-Risk (red) groups. Most immune cell types and cytolytic activity are significantly decreased in the High-Risk group (indicated by asterisks for significance levels), suggesting an immunosuppressive phenotype associated with high risk.

Figure 6: NF1 Mutation Analysis

(A) Risk Score by NF1 Mutation Status: Violin plot showing no significant difference ($p = 0.12$) in risk scores between patients with NF1 mutations versus wild-type.

(B) Overall Survival by NF1 Mutation Status: Kaplan-Meier curves comparing overall survival between NF1 mutant and wild-type patients. The difference is not statistically

significant (Log-rank $p = 0.11$).

(C) OS by NF1 Variant Type: Kaplan-Meier curves stratifying patients by specific NF1 variant types (truncating, missense, wild-type). No significant difference was observed (Log-rank $p = 0.27$).

(D) Signature Gene Expression by NF1 Status: Bar chart comparing the mean expression of the signature genes between NF1 mutant and wild-type patients. Expression levels appear relatively similar across groups for most genes.

Figure 7: Single-Cell Validation of Signature Genes

(A) Bar Charts: The top panel shows the mean expression levels of the signature genes in melanoma single cells (GSE72056 dataset), while the bottom panel shows the percentage of cells expressing each gene. RASGRP1 and FGF2 show the highest expression levels.

(B) Heatmap: A heatmap displaying the expression profile of the signature genes across individual melanoma single cells. The color scale ranges from dark blue (low expression) to red (high expression), illustrating the heterogeneity of gene expression at the single-cell level.

Figure 8.1 and 8.2: Model Validation and Meta-Analysis

(a) LASSO Cross-Validation: Re-display of the LASSO model selection confirming the optimal $\lambda$for the gene signature.

(b) Risk Factor Heatmap: A heatmap showing the expression of the 8 signature genes (rows) across patients (columns) sorted by Risk Type. It also indicates the direction of the risk factor (red = risk, blue = protective).

(c) Kaplan-Meier Curve (Discovery Cohort): The primary survival analysis showing that the

High-Risk group has significantly worse overall survival than the Low-Risk group (Log-rank $p = 1.41 \times 10^{-10}$).

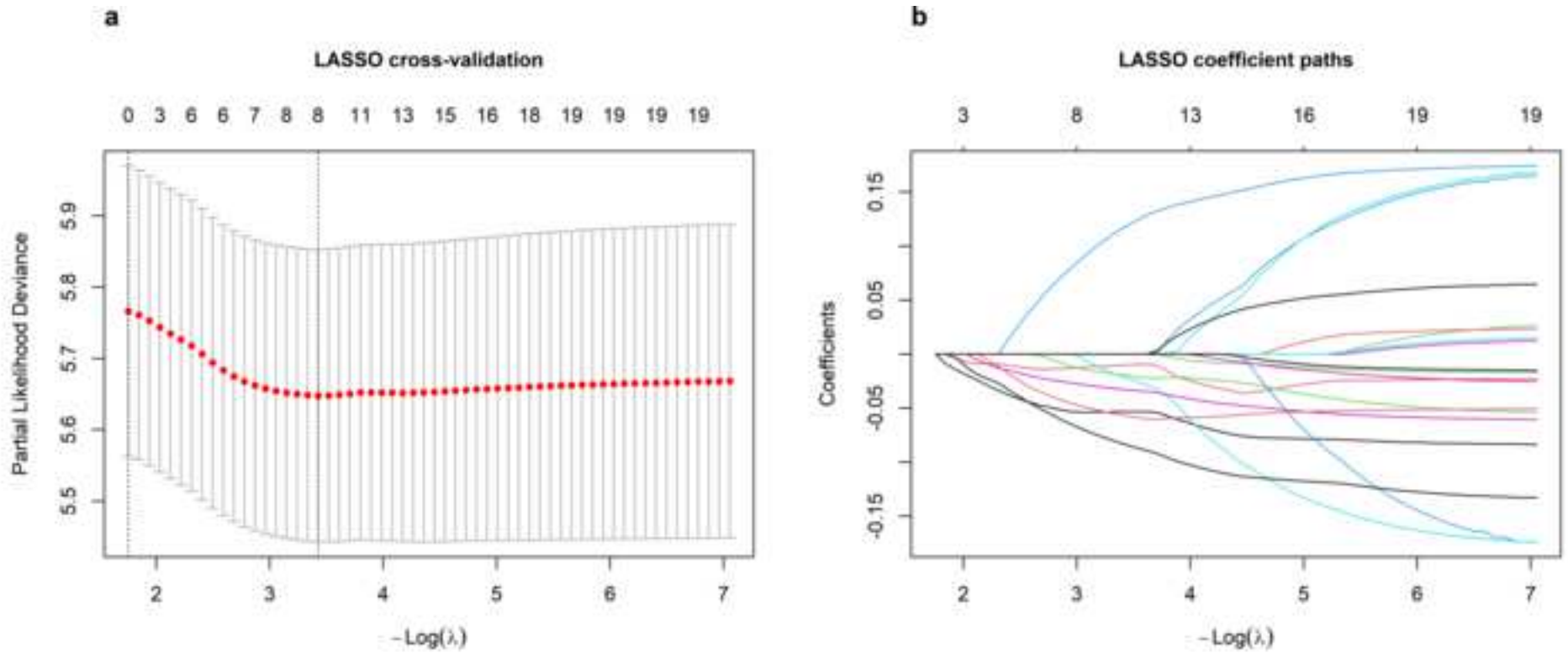
(d) Time-dependent ROC: ROC curves evaluating the predictive accuracy of the risk model for 1-, 2-, and 3-year survival. The AUC values range from 0.712 to 0.733, indicating good predictive performance.

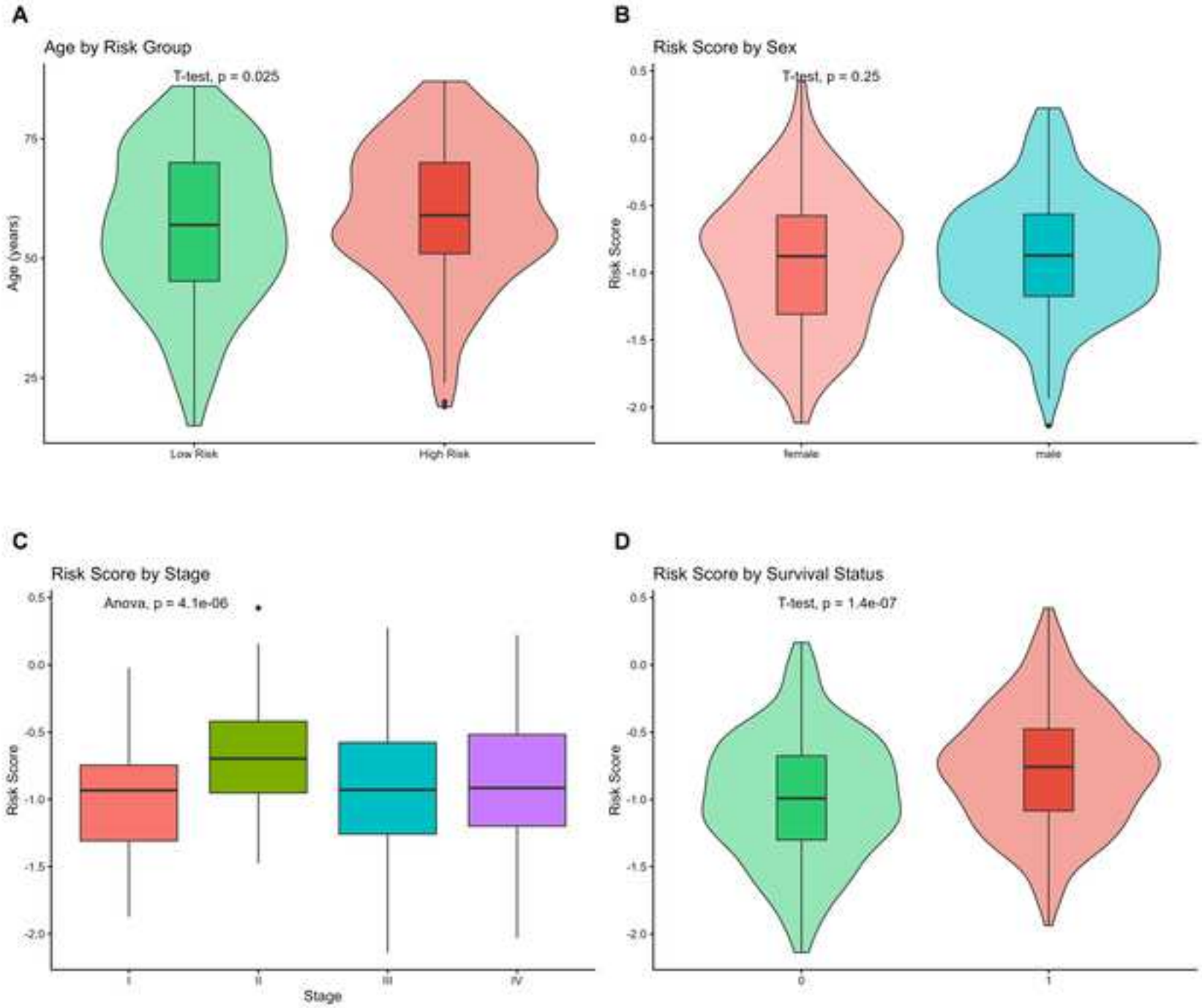
(Bottom Panels) External Validation & Meta-Analysis: Kaplan-Meier curves for four independent external datasets (GSE65904, GSE54467, GSE22153, GSE19234) validating the risk model. Below these, forest plots show the results of a meta-analysis combining these datasets, demonstrating a pooled Hazard Ratio (HR) of 2.247 (95% CI: 1.032–4.892), confirming the prognostic value of the signature.

Figure 9: Summary of Model Validation

This figure consolidates key validation results:

- **Top Left:** Discovery cohort overall survival Kaplan-Meier curve.
- **Top Right:** Time-dependent ROC curves for the TCGA-SKCM dataset.
- **Panel c:** LASSO cross-validation plot.
- **Panel d:** Forest plot of the meta-analysis across the four external validation cohorts (GSE65904, GSE54467, GSE22153, GSE19234), showing a pooled HR of 2.247, reaffirming the model's robust prognostic capability.


a
LASSO cross-validation
0 3 6 6 7 8 8 11 13 15 16 18 19 19 19 19
Partial Likelihood Deviance
5.5 5.6 5.7 5.8 5.9
2 3 4 5 6 7
−Log(λ)
b
LASSO coefficient paths
3 8 13 16 19 19
Coefficients
-0.15 -0.05 0.05 0.15
2 3 4 5 6 7
−Log(λ)


A
Age by Risk Group
T-test, p = 0.025
Age (years)
75
50
25
Low Risk
High Risk
B
Risk Score by Sex
T-test, p = 0.25
Risk Score
0.5
0.0
-0.5
-1.0
-1.5
-2.0
female
male
C
Risk Score by Stage
Anova, p = 4.1e-06
Risk Score
Stage
I
II
III
IV
D
Risk Score by Survival Status
T-test, p = 1.4e-07
Risk Score
0
1

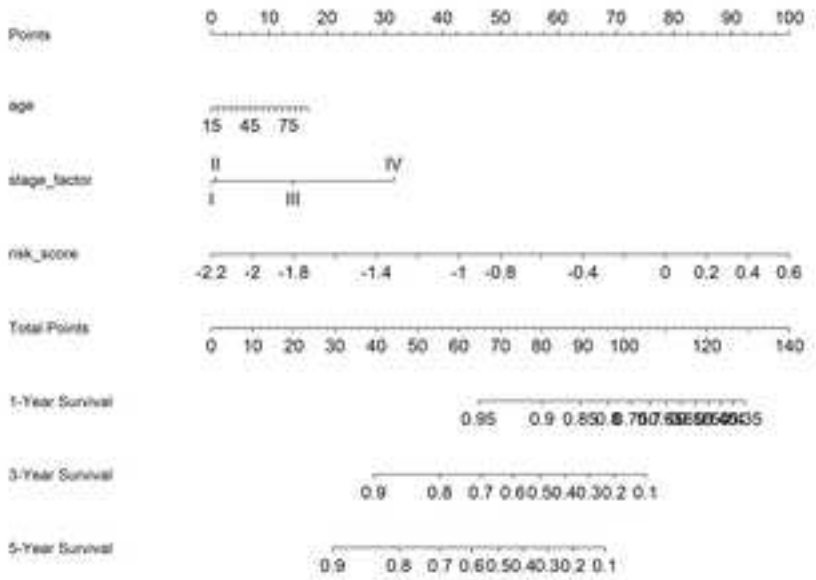
A

Variable	HR (95% CI)	P-value
age	1.014 (1.003-1.025)	0.010
stage_factorII	1.6 (0.943-2.716)	0.081
stage_factorIII	1.882 (1.176-3.012)	0.008
stage_factorIV	3.77 (1.884-7.541)	<0.001
risk_score	4.673 (3.157-6.917)	<0.001

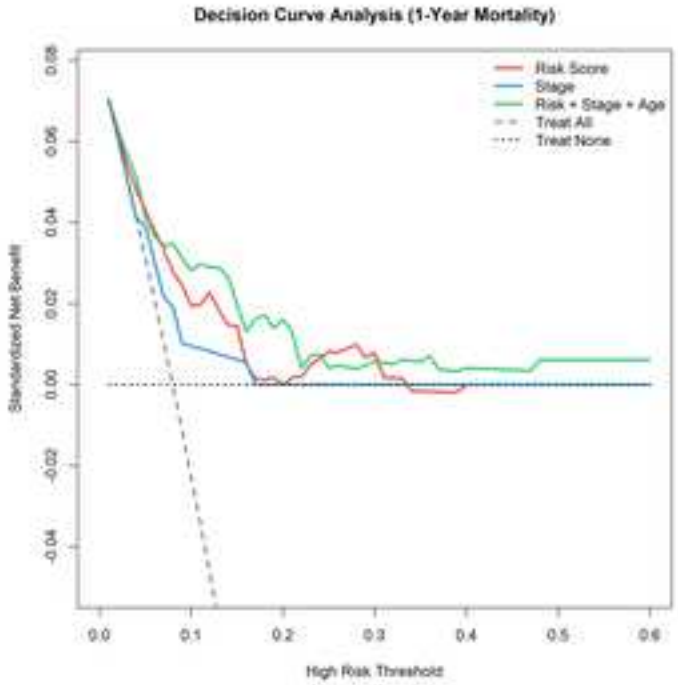
B

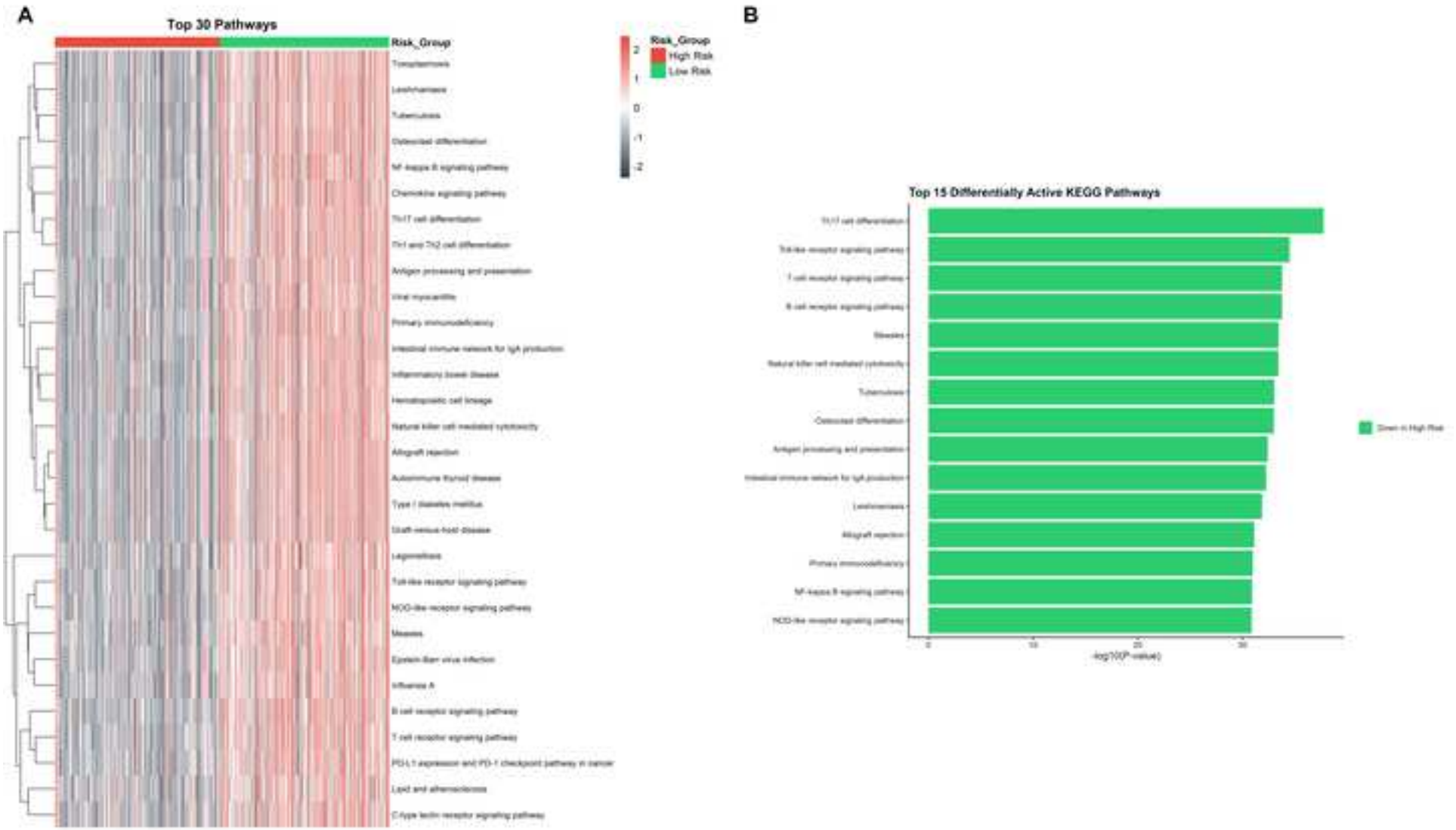
Variable	HR (95% CI)	P-value
age	1.01 (0.999-1.022)	0.067
stage_factorII	1.037 (0.602-1.784)	0.898
stage_factorIII	1.939 (1.209-3.111)	0.006
stage_factorIV	4.381 (2.174-8.829)	<0.001
risk_score	5.323 (3.547-7.988)	<0.001

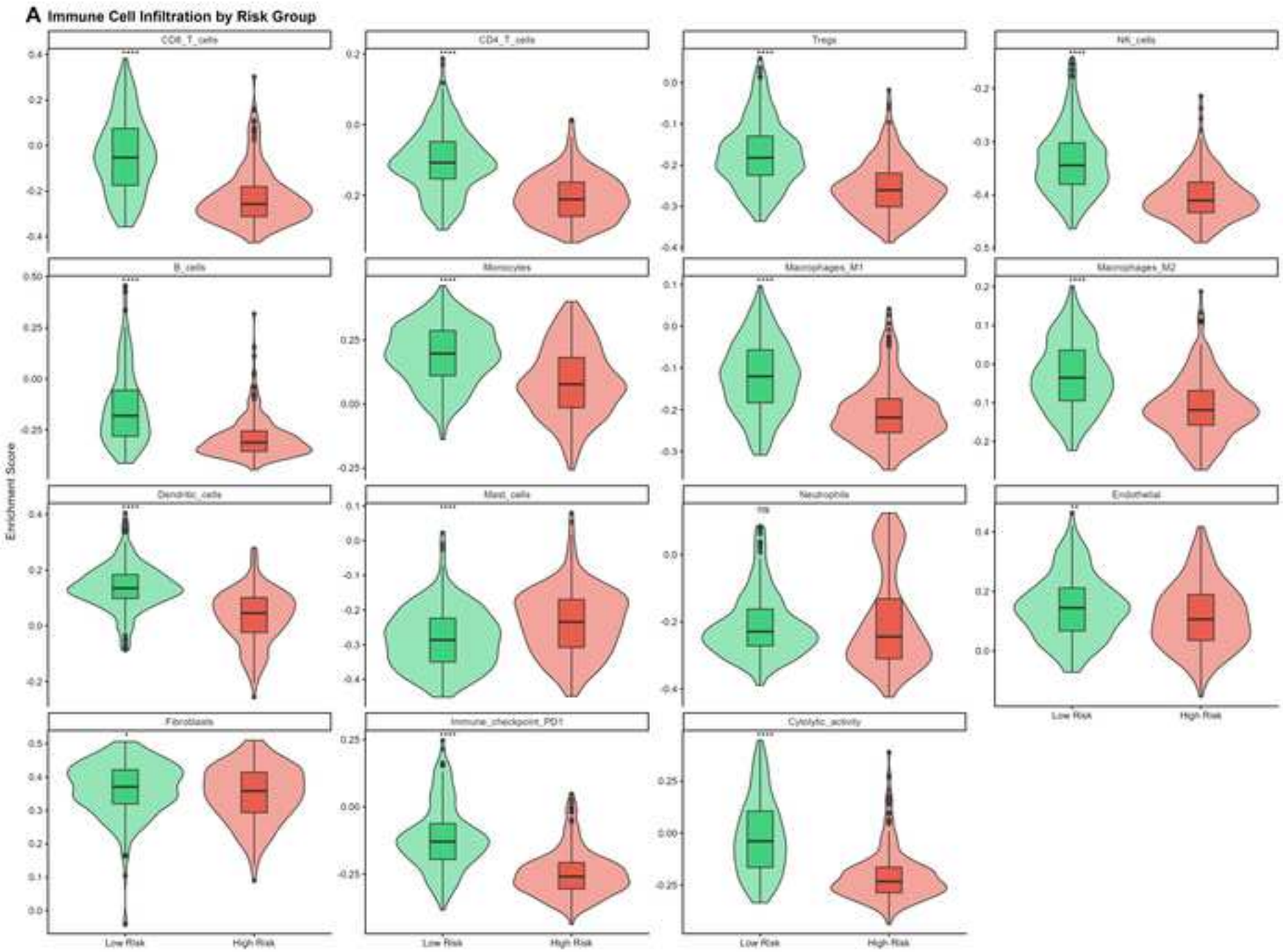
C

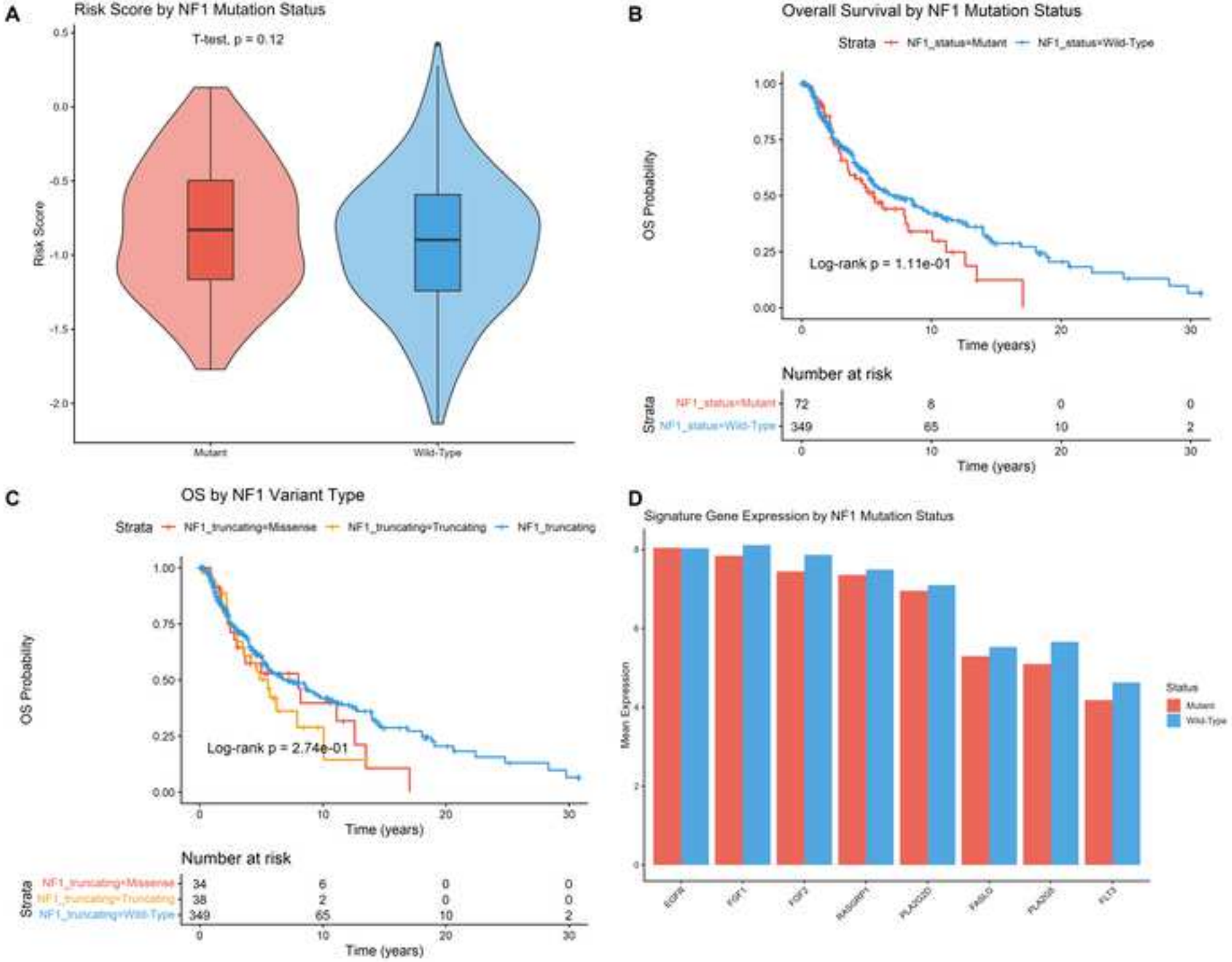


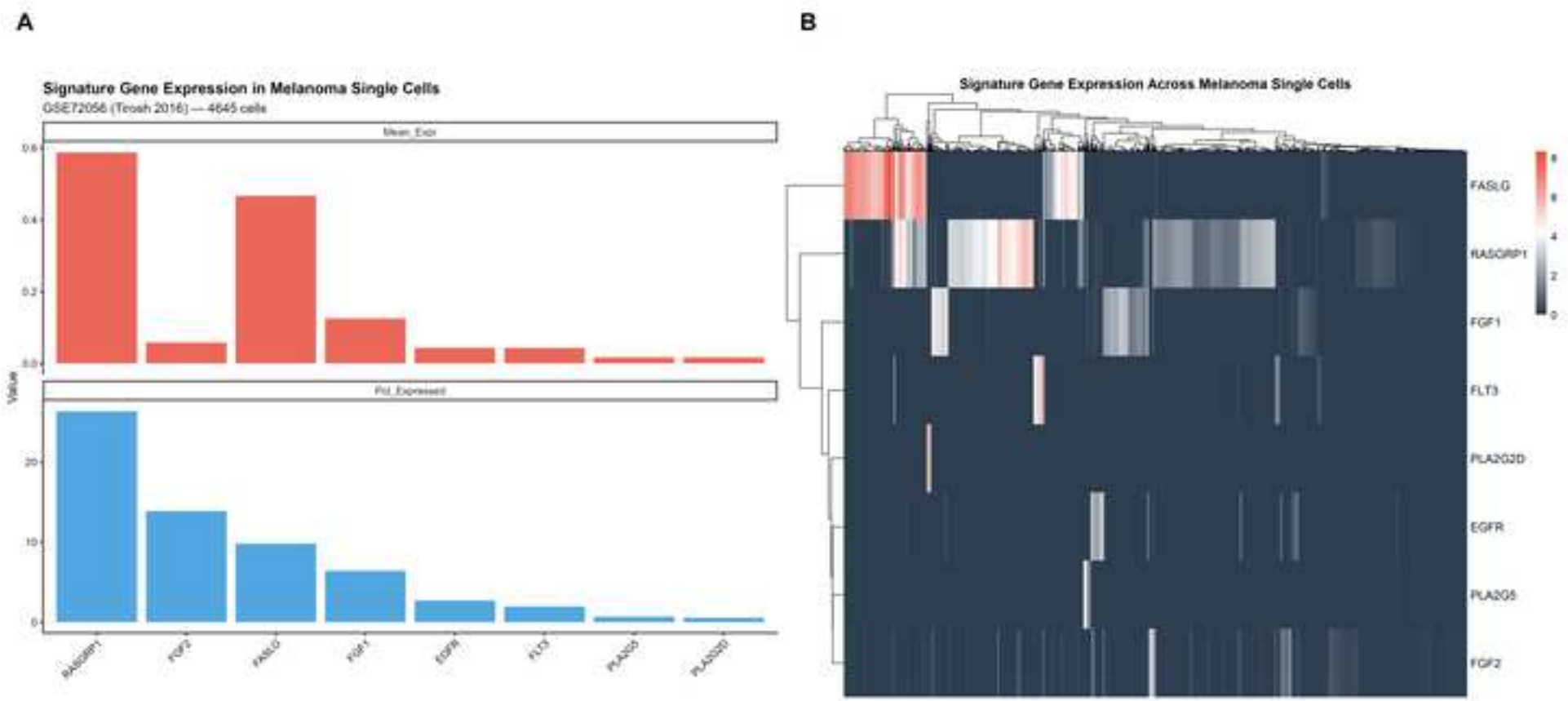
D

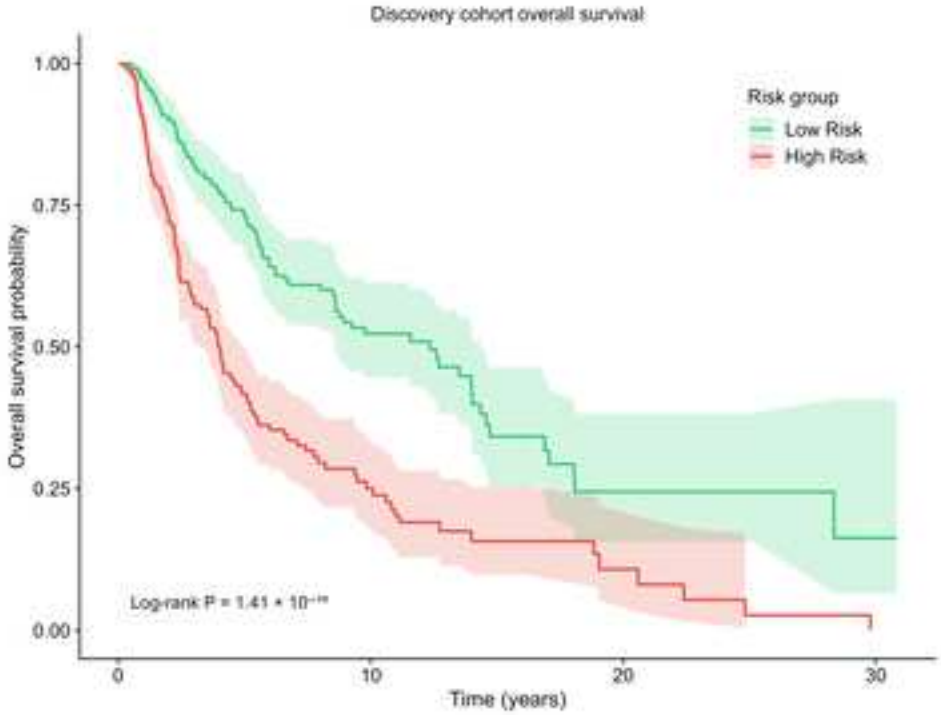


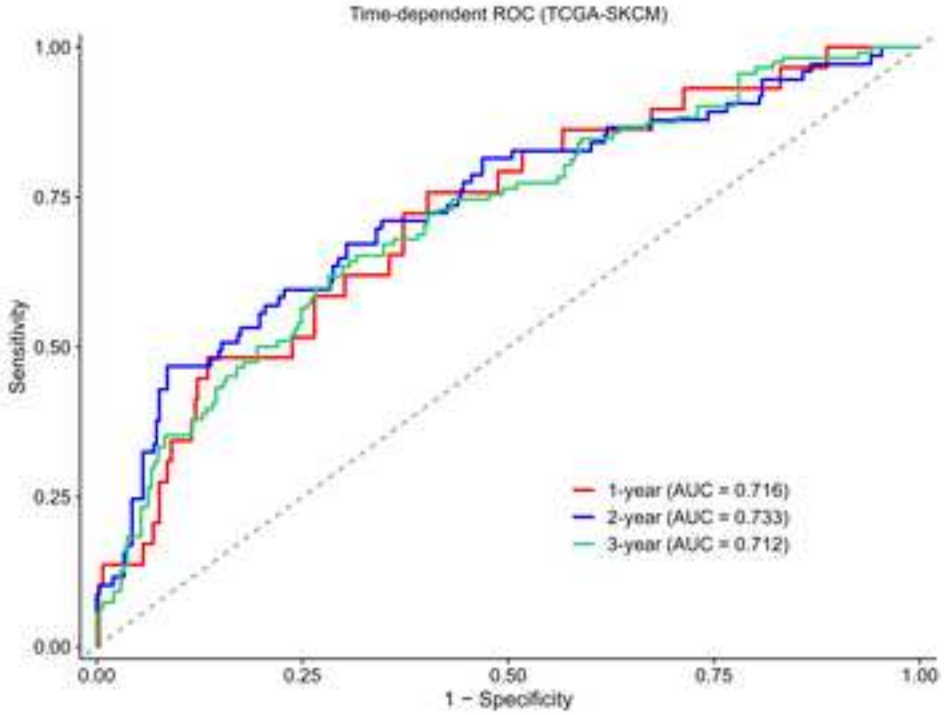

A
Top 30 Pathways
Risk_Group
High Risk
Low Risk
2
1
0
-1
-2
Tuberculosis
Osteoclast differentiation
NF-kappa B signaling pathway
Chemokine signaling pathway
Th17 cell differentiation
Th1 and Th2 cell differentiation
Antigen processing and presentation
Primary immunodeficiency
Intestinal immune network for IgA production
Inflammatory bowel disease
Hematopoietic cell lineage
Natural killer cell mediated cytotoxicity
Allograft rejection
Autoimmune thyroid disease
Graft-versus-host disease
Legionellosis
Toll-like receptor signaling pathway
NOD-like receptor signaling pathway
Measles
Epstein-Barr virus infection
Influenza A
B cell receptor signaling pathway
T cell receptor signaling pathway
Lipid and atherosclerosis
C-type lectin receptor signaling pathway
B
Top 15 Differentially Active KEGG Pathways
Th17 cell differentiation
Toll-like receptor signaling pathway
T cell receptor signaling pathway
B cell receptor signaling pathway
Measles
Natural killer cell mediated cytotoxicity
Tuberculosis
Osteoclast differentiation
Antigen processing and presentation
Intestinal immune network for IgA production
Leishmaniasis
Allograft rejection
Primary immunodeficiency
NF-kappa B signaling pathway
NOD-like receptor signaling pathway
0
10
20
30
-log10(P-value)
Down in High Risk


A Immune Cell Infiltration by Risk Group
CD8_T_cells
CD4_T_cells
Tregs
NK_cells
B_cells
Monocytes
Macrophages_M1
Macrophages_M2
Dendritic_cells
Mast_cells
Neutrophils
Endothelial
Fibroblasts
Immune_checkpoint_PD1
Cytolytic_activity
Enrichment Score
Low Risk
High Risk

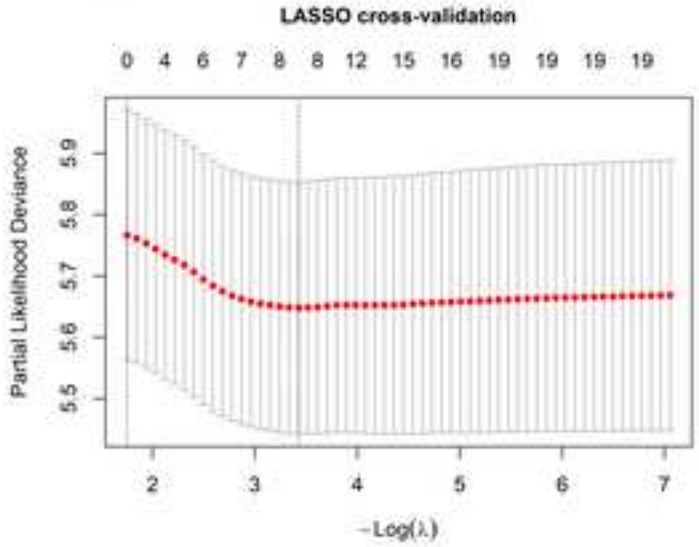

A
Risk Score by NF1 Mutation Status
T-test, p = 0.12
Risk Score
Mutant
Wild-Type
B
Overall Survival by NF1 Mutation Status
Strata
NF1_status=Mutant
NF1_status=Wild-Type
OS Probability
Log-rank p = 1.11e-01
Time (years)
Number at risk
72
8
0
0
349
65
10
2
C
OS by NF1 Variant Type
Strata
NF1_truncating=Missense
NF1_truncating=Truncating
NF1_truncating
OS Probability
Log-rank p = 2.74e-01
Time (years)
Number at risk
34
6
0
0
38
2
0
0
349
65
10
2
D
Signature Gene Expression by NF1 Mutation Status
Mean Expression
Status
Mutant
Wild-Type
EGFR
FGF1
FGF2
RASGRP1
PLAGL2
FABP3
PLAG1
FLT3


A
Signature Gene Expression in Melanoma Single Cells
GSE72056 (Tirosh 2016) — 4645 cells
Mean_Expr
Pct_Expressed
Value
RASGRP1
FGF2
FASLG
FGF1
EGFR
FLT3
PLA2G5
PLA2G2D
B
Signature Gene Expression Across Melanoma Single Cells
FASLG
RASGRP1
FGF1
FLT3
PLA2G2D
EGFR
PLA2G5
FGF2

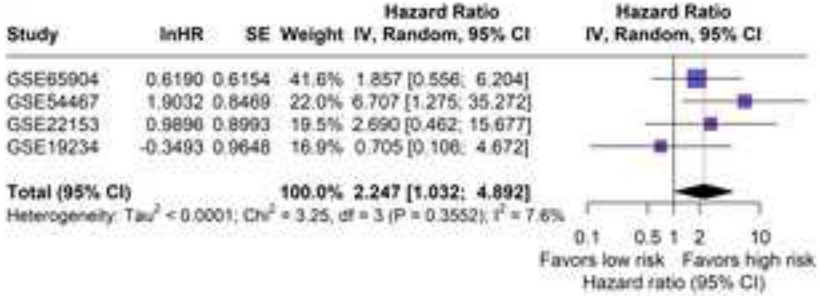

Discovery cohort overall survival
Risk group
Low Risk
High Risk
Overall survival probability
1.00
0.75
0.50
0.25
0.00
Log-rank P = 1.41 × 10⁻⁴⁴
0
10
20
30
Time (years)

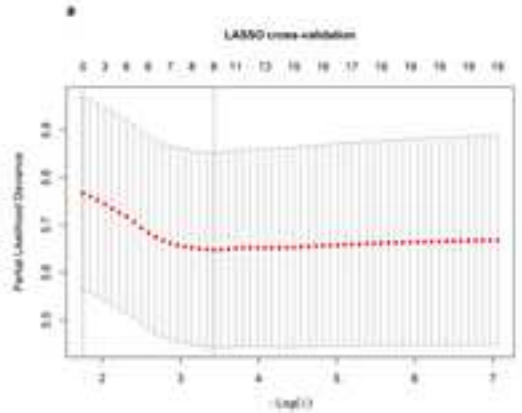

Time-dependent ROC (TCGA-SKCM)
Sensitivity
1.00
0.75
0.50
0.25
0.00
1-year (AUC = 0.716)
2-year (AUC = 0.733)
3-year (AUC = 0.712)
0.00
0.25
0.50
0.75
1.00
1 − Specificity

c

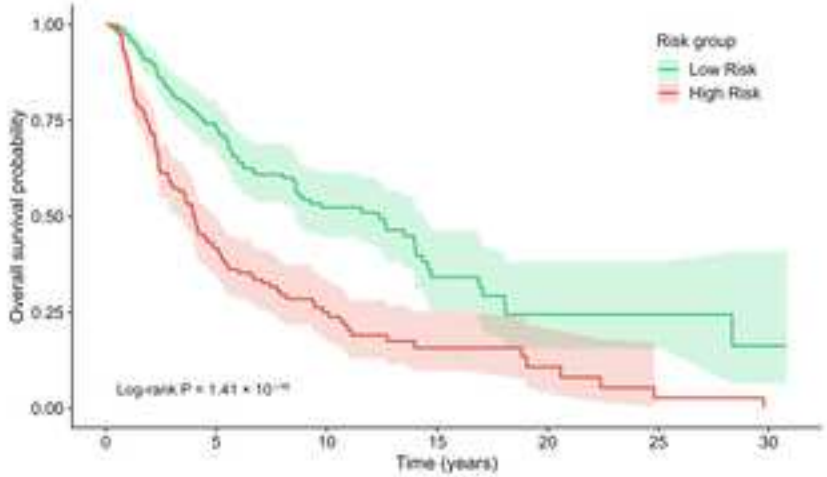

LASSO cross-validation
0 4 6 7 8 8 12 15 16 19 19 19 19
Partial Likelihood Deviance
5.5
5.6
5.7
5.8
5.9
2
3
4
5
6
7
−Log(λ)

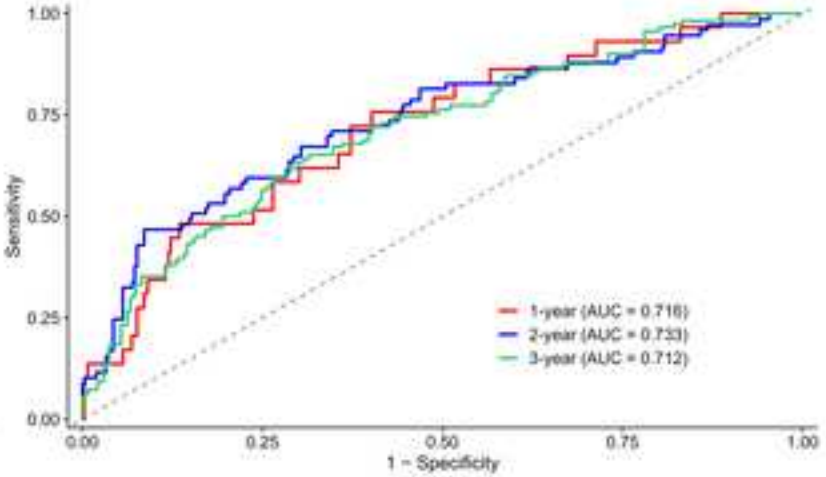
d

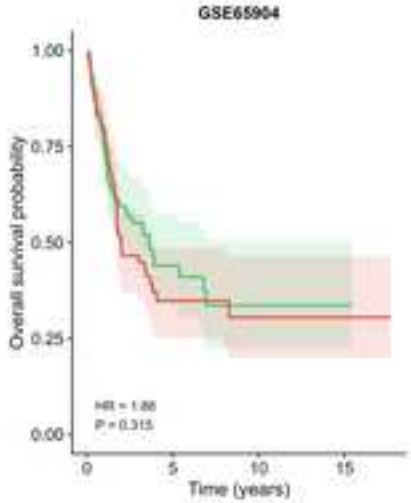

Study
lnHR
SE
Weight
Hazard Ratio
IV, Random, 95% CI
Hazard Ratio
IV, Random, 95% CI
GSE65904 0.6190 0.6154 41.6% 1.857 [0.556; 6.204]
GSE54467 1.9032 0.8469 22.0% 6.707 [1.275; 35.272]
GSE22153 0.9896 0.8993 19.5% 2.690 [0.462; 15.677]
GSE19234 -0.3493 0.9648 16.9% 0.705 [0.106; 4.672]
Total (95% CI) 100.0% 2.247 [1.032; 4.892]
Heterogeneity: Tau² < 0.0001; Chi² = 3.25, df = 3 (P = 0.3552); I² = 7.6%
0.1
0.5
1
2
10
Favors low risk
Favors high risk
Hazard ratio (95% CI)

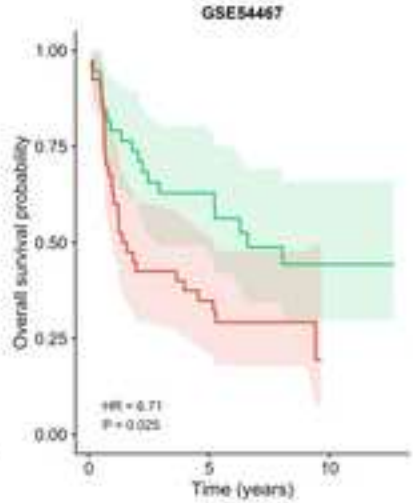

LASSO cross-validation
Partial Likelihood Deviance
Log(λ)

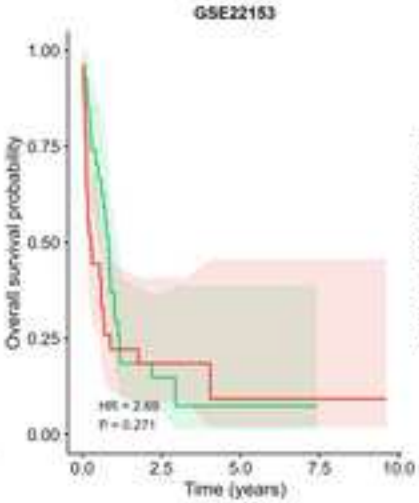

Risk_Type
Low Risk
High Risk

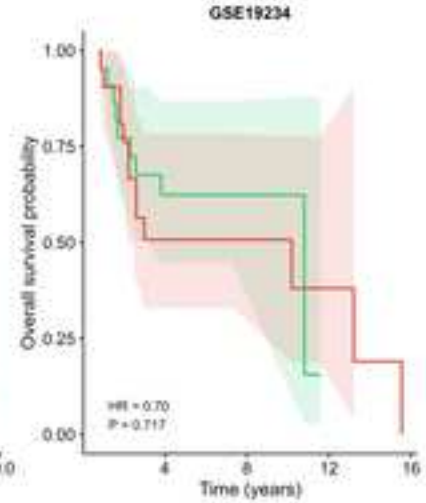

Risk group
Low Risk
High Risk
Overall survival probability
Log-rank P = 1.41 × 10⁻⁴⁶
Time (years)

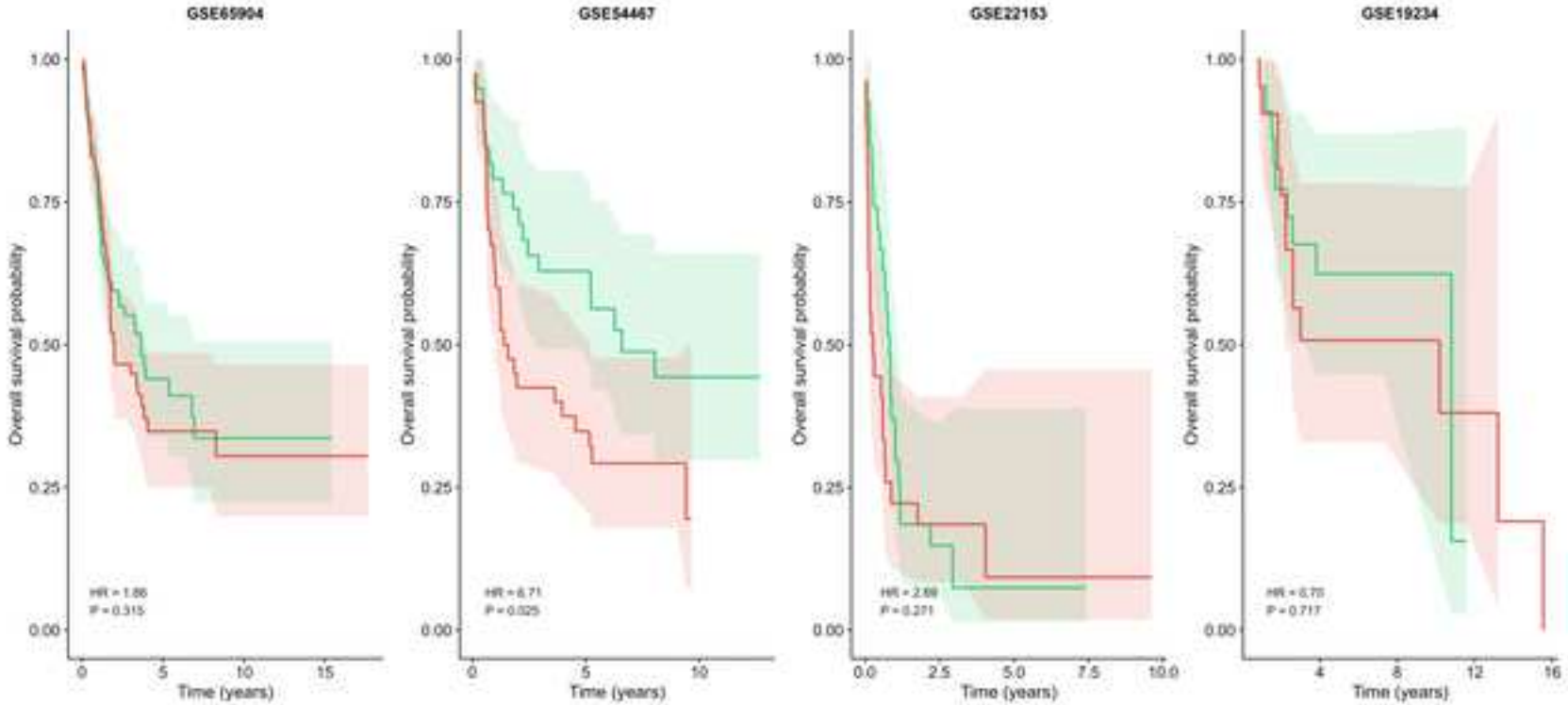

Sensitivity
1 − Specificity
1-year (AUC = 0.716)
2-year (AUC = 0.733)
3-year (AUC = 0.712)


GSE65904
Overall survival probability
HR = 1.88
P = 0.315
Time (years)


GSE54467
Overall survival probability
HR = 0.71
P = 0.025
Time (years)


GSE22153
Overall survival probability
HR = 2.69
P = 0.271
Time (years)


GSE19234
Overall survival probability
HR = 0.70
P = 0.717
Time (years)


GSE65904
GSE54467
GSE22153
GSE19234
Overall survival probability
Time (years)
HR = 1.86
P = 0.315
HR = 6.71
P = 0.025
HR = 2.69
P = 0.271
HR = 0.70
P = 0.717

e


Study
lnHR
SE
Weight
Hazard Ratio IV, Random, 95% CI
GSE65904 0.6190 0.6154 41.6% 1.857 [0.556; 6.204]
GSE54467 1.9032 0.8469 22.0% 6.707 [1.275; 35.272]
GSE22153 0.9896 0.8993 19.5% 2.690 [0.462; 15.677]
GSE19234 -0.3493 0.9648 16.9% 0.705 [0.106; 4.672]
Total (95% CI) 100.0% 2.247 [1.032; 4.892]
Heterogeneity: Tau² = 0.0001; Chi² = 3.25, df = 3 (P = 0.3552); I² = 7.8%
0.1 0.5 1 2 10
Favors low risk Favors high risk
Hazard ratio (95% CI)

f


Study
lnHR
SE
Weight
Hazard Ratio IV, Random, 95% CI
GSE65904 0.6190 0.6154 50.1% 1.857 [0.556; 6.204]
GSE54467 1.9032 0.8469 26.4% 6.707 [1.275; 35.272]
GSE22153 0.9896 0.8993 23.5% 2.690 [0.462; 15.677]
Total (95% CI) 100.0% 2.845 [1.212; 6.681]
Heterogeneity: Tau² = 0; Chi² = 1.51, df = 2 (P = 0.4700); I² = 0.0%
0.1 0.5 1 2 10
Favors low risk Favors high risk
Hazard ratio (95% CI)

Supplementary Files

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

- moduletraitcorrelation.csv
- NF1basesettransparent.csv
- NF1finalcandidatepool.csv
- NF1forestmultivariate.csv
- NF1forestunivariate.csv
- NF1GSVAresults.csv
- NF1immuneinfiltrationresults.csv
- NF1metaanalysis.csv
- NF1metaeggertest.csv
- NF1metaleaveoneout.csv
- NF1multivariatecox.csv
- NF1mutationclassdistribution.csv
- NF1mutationsummary.csv
- NF1signaturecoefficients.csv
- NF1signaturegenesbymutation.csv
- NF1singlecellexpressionsummary.csv
- NF1univariatecox.csv
- NF1validationsummary.csv
- singlecellsignatureresults.csv