haematologica
Journal of the Ferrata Storti Foundation

# Shared high-risk transcriptional programs across multiple myeloma stages associate with disease progression and poor outcome

*by Claudio Cerchione, Nour Alassi, Delia Cangini, Sarah Tassinari, Asia Spedaluzzi, Irene Azzali, Sofia Cortesi, Giacomo Feliciani, Alice Rossi, Sonia Ronconi, Michela Ceccolini, Michela Tonelli, Gerardo Musuraca, Nicola Normanno, Emanuela Scarpi, Giorgia Simonetti, Giovanni Martinelli, Paul G. Richardson and Matteo Marchesini*

*Received: July 1, 2026.*
*Accepted: November 6, 2026.*

*Citation: Claudio Cerchione, Nour Alassi, Delia Cangini, Sarah Tassinari, Asia Spedaluzzi, Irene Azzali, Sofia Cortesi, Giacomo Feliciani, Alice Rossi, Sonia Ronconi, Michela Ceccolini, Michela Tonelli, Gerardo Musuraca, Nicola Normanno, Emanuela Scarpi, Giorgia Simonetti, Giovanni Martinelli, Paul G. Richardson and Matteo Marchesini. Shared high-risk transcriptional programs across multiple myeloma stages associate with disease progression and poor outcome. Haematologica. 2026 Nov 12. doi: 10.3324/haematol.2026.301504 [Epub ahead of print]*

*Publisher's Disclaimer.*
*E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication.*
*E-publishing of this PDF file has been approved by the authors.*
*After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval, the final version of the manuscript will then appear in a regular issue of the journal.*
*All legal disclaimers that apply to the journal also pertain to this production process.*

# Shared high-risk transcriptional programs across multiple myeloma stages associate with disease progression and poor outcome

Claudio Cerchione[1, #, *], Nour Alassi[2, 3, #], Delia Cangini[1], Sarah Tassinari[2], Asia Spedaluzzi[2], Irene Azzali[4], Sofia Cortesi[1], Giacomo Feliciani[5], Alice Rossi[6], Sonia Ronconi[1], Michela Ceccolini[1], Michela Tonelli[7], Gerardo Musuraca[1], Nicola Normanno[8,9], Emanuela Scarpi[4], Giorgia Simonetti[2], Giovanni Martinelli[3], Paul G. Richardson[10], Matteo Marchesini[2, *].

[1] Hematology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[2] Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[3] Department of Medical and Surgical Sciences (DIMEC), Alma Mater Studiorum-Università di Bologna, Bologna, Italy.

[4] Biostatistics and Clinical Trials Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[5] Medical Physics Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[6] Radiology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[7] U.O. Genetica Medica, Laboratorio Unico-AUSL della Romagna, Pievesestina di Cesena, Italy.

[8] Scientific Directorate, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy.

[9] Center for Advanced Molecular Diagnostics in Oncology, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Roma, Italy

[10] Department of Medical Oncology, Dana-Farber Cancer Institute, Jerome Lipper Center for Multiple Myeloma Research, Harvard Medical School, Boston, MA, United States

[#] contributed equally

* corresponding author

**Correspondence:**

Claudio Cerchione MD, PhD

Hematology Unit, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", via Piero Maroncelli 40, 47014, Meldola (FC), Italy; tel: +39 0543 739100; email: claudio.cerchione@irst.emr.it

Matteo Marchesini, PhD

Biosciences Laboratory, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", via Piero Maroncelli 40, 47014, Meldola (FC), Italy; tel: +39 0543 739100; email: matteo.marchesini@irst.emr.it

**Running Title:** High-risk Transcriptional Programs in Myeloma

**Key words:** Multiple myeloma, smoldering multiple myeloma, high-risk disease, gene expression profiling, SKY92 gene signature, prognosis.

**Acknowledgments**

The authors thank all participants in this study and their families. The authors thank Silvia D'ambrosi and Rowan Kuiper for scientific oversight and SkylineDx (Lichtenauerlaan 40, 3062 ME, Rotterdam, the Netherlands) for SKY92 test supply.

**Funding**

This work was supported by Pfizer Competitive Research Grant RFP: Pre-clinical & Translational Research in Multiple Myeloma (GRG_#70107607 to G.M. and M.M.); European Union - Next Generation EU, Mission 4, Component 2, Investment 3.3 (Ministerial Decree 117/2023), CUP J33C22001500009 (to N.A.); Affari e Solidarietà (#DON25-229 to M.M.).

**Conflict of Interest Statement**

The authors declare no conflicts of interest.

**Authorship Contributions**

Conceptualization and design: C.C., N.A. and M.M.; clinical oversight and sample collection and clinical data curation: A.R., D.C., S.C., G.M., S.R., M.C., G.MA. and C.C.; sample processing: S.T. and A.S.; cytogenetic analyses and data curation: M.T.; methodology and experiment execution: N.A., M.M. and S.T.; data analysis: N.A., G.S., G.F. and M.M.; writing and editing: C.C., N.A., G.S., G.MA., P.G.R. and M.M.; data curation: N.A., I.A., E.S. and M.M.; supervision: M.M. All the authors read and discussed the manuscript.

**Data sharing statement:** Data are available from the corresponding authors upon reasonable request.

**Declaration of AI Use:** During the preparation of this work, the authors used ChatGPT to improve the English language and readability. The authors reviewed and edited the content as needed and take full responsibility for the final content of the publication. ChatGPT was not used in data collection, analysis, or drawing scientific conclusions.

**Main:**

The therapeutic landscape of multiple myeloma (MM) has changed profoundly over the past two decades, resulting in substantially improved outcomes for newly diagnosed (NDMM) patients, particularly in the transplant-eligible setting. However, approximately 10-15% of patients do not benefit from this opportunity and their clinical outcome remains poor [1]. This subset of patients likely shares underlying biological and clinical features of a more aggressive course and increased likelihood of early disease progression, aligning with a high-risk (HR) myeloma profile. Early identification of HR patients remains an unmet clinical need, and addressing this gap would facilitate risk-adapted therapeutic strategies and more personalized clinical management. The International Myeloma Society and International Myeloma Working Group (IMS/IMWG) have made substantial efforts to define HR disease through risk stratification systems based on clinical, biochemical and DNA-based features [2]. Interphase fluorescence *in situ* hybridization (iFISH) remains the gold standard for detecting cytogenetic abnormalities associated with HR. However, limitations related to the lack of standardization, inter-laboratory variability, and delayed or incomplete results highlight the need for further optimization [3].

In parallel, the need to better capture the biological heterogeneity underlying HR MM has led to the implementation of plasma cell-derived gene expression profiling (GEP) as a complementary prognostic tool to DNA-based diagnostics, capable of identifying HR patients who may be missed by conventional risk assessment. GEP signatures, such as IFM15, UAMS70/80, and SKY92, are based on the differential expression of a defined set of genes and have demonstrated the ability to distinguish aggressive disease, both at diagnosis (NDMM) and in the relapsed/refractory (RRMM) setting [4-7], supporting the development of standardized, clinically applicable GEP-based prognostic assays. However, the prognostic value of GEP profiling has been less extensively investigated in the early asymptomatic phase of the disease (SMM, smoldering myeloma).

To evaluate the performance of gene expression-derived cytogenetics and HR-associated signatures across MM disease stages in a real-world setting, we profiled consecutive SMM, NDMM and RRMM patients using plasma cell transcriptomic analysis.

Since February 2020, a total of n = 105 patients were enrolled and characterized using the SKY92-based prognostic assay MMprofiler (SkylineDx). The study was approved by the Institutional Ethics Committees (Prot. 5805/2019) and was conducted in accordance with the Declaration of Helsinki. Written informed consent was received from participants prior to inclusion in the study. Following $CD138^+$ plasma-cell selection, the $CD38^+$ purity threshold (≥80%) required for SKY92 analysis was met in n = 90 cases (85.7%), including n = 27 SMM, n = 39 NDMM and n = 24 RRMM (**Figure 1A** and **Supplemental Table 1**); samples failing this threshold due to insufficient plasma-cell recovery were excluded from downstream analysis. Failure rates were 18.2% (6 out of 33) in SMM, 7.1% (3 out of 42) in NDMM and 20% (6 out of 30) in RRMM.

The prevalence of SKY92 HR signature increased progressively across disease stages, from 19% in SMM (n = 5/27) to 41% in NDMM (n = 16/39) and 67% in RRMM (n = 16/24, *P* = 0.0018, **Figure 1B**). Moreover, in RRMM, there was a trend toward SKY92 HR signature prevalence among heavily pretreated patients (n = 15/20, 75% in patients that received three or more lines of treatment compared with those who had received fewer than three lines of treatment (n = 1/4, 25%; *P* = 0.090). We then analyzed the potential of the SKY92 test to intercept HR-associated cytogenetic markers by evaluating the concordance between SKY92-derived cytogenetic (virtual FISH or vFISH) predictions and conventional iFISH data (available for 48/90 cases, 53.3%; missing data reflected limitations of routine clinical testing, including insufficient material or plasma-cell recovery). In a global analysis across disease phases, high concordance was observed for t(4;14), t(14;16/20) and gain(1q) with

an accuracy of 97.9 (95% CI, 93.9-100), 85.2 (95% CI, 71.8-98.6) and 87.2 (95% CI, 77.7-96.8) respectively, decreasing to 66.7 (95% CI, 49.8-83.5) for del(17p), suggesting that this chromosomal abnormality may not be fully recapitulated by GEP-based projections (**Figure 1C** and **Supplemental Table 2**). Importantly, with slight differences, this trend was maintained in a MM phase-specific analysis (**Supplemental Table 2**). We next evaluated the prognostic potential of the SKY92 gene signature in the symptomatic MM setting, for which this test has been previously validated. Within the NDMM cohort, patients classified as high-risk by SKY92 (SKY92 HR) exhibited significantly worse overall survival (OS) compared with standard-risk patients [SKY92 SR; median OS 32.8 months versus not reached, $P = 0.0079$; **Figure 2A**]. A similar trend was observed in RRMM, although statistical significance was not reached, likely due to the limited cohort size (median OS 20.9 months versus not reached, $P = 0.093$; **Figure 2B**). These findings support the prognostic relevance of SKY92 in NDMM and suggest a consistent trend in RRMM. To further characterize the transcriptional structure associated with SKY92 risk in symptomatic MM, a linear model was applied to GEP data obtained from NDMM and RRMM samples, identifying a set of 822 probes (MAS5 $\log_2$ expression >7) associated with SKY92 status ($P < 0.01$ and absolute $\log_2$ fold change > 0.5). Principal component analysis (PCA) performed on this supervised gene set revealed a transcriptional landscape in which PC1 was strongly associated with SKY92 risk status by linear regression ($P = 9.26 \times 10^{-17}$; *BH-adjusted* $P = 9.26 \times 10^{-16}$), whereas none of PC2–PC10 showed a significant association (**Figure 2D** and **Supplemental Figure S1A**). Notably, in this PCA-defined transcriptional space NDMM and RRMM samples displayed a mixed distribution rather than forming distinct clusters (**Figure 2D**). To further examine whether the high-risk-associated transcriptional structure was consistent across symptomatic disease stages, we directly compared HR-NDMM and HR-RRMM across all probes retained after preprocessing. No individual probe remained differentially expressed after multiple-testing correction. Moreover, the HR–SR separation

along PC1 was comparable in NDMM and RRMM, with no evidence of a risk-by-stage interaction (*P* = 0.986). We next performed a gene set enrichment analysis (GSEA) combining NDMM and RRMM samples. The comparison of SKY92 HR versus SR patients revealed an HR-associated profile enriched in signatures related to disease progression and drug resistance in MM as well as cell proliferation [8] (E2F targets, NES = 1.68, *FDR* = 0.069 and G2M checkpoint, NES = 1.7, *FDR* = 0.118), oncogene-associated targets (MYC targets V1, NES = 1.69, *FDR* = 0.085 and V2, NES = 1.54, *FDR* = 0.131) and DNA damage repair [9] (DNA repair NES = 1.53, *FDR* = 0.128, and UV response, NES = 1.42, *FDR* = 0.209, **Figure 2C, Supplemental Figure S1B** and **Supplemental Table 2**). To independently assess the reproducibility of SKY92-associated transcriptional patterns, we analyzed the GSE87900 NDMM cohort (n=180; 24 HR and 156 SR)[10]. Projection onto the fixed IRST-derived PCA space showed significantly higher PC1 scores in HR than SR patients (mean, 6.20 versus −6.25; $P=4.84 \times 10^{-20}$). Hallmark GSEA identified nine pathways enriched in HR patients (FDR<0.25), eight of which were also enriched in our symptomatic MM cohort, including E2F targets, G2M checkpoint, MYC targets V1/V2, DNA repair, mitotic spindle, MTORC1 signaling, and unfolded protein response (**Figure 2E-F, Supplemental Figure S1C-D, and Supplemental Table 2**). These results support the reproducibility of SKY92-associated transcriptional programs in an independent NDMM cohort. Together, these findings identify specific transcriptional programs associated with SKY92 HR status in symptomatic MM and support a largely preserved high-risk-associated transcriptional structure across symptomatic disease stages. To explore whether the GEP-based risk assessment prediction could be anticipated to the asymptomatic phase, we projected SMM samples into the SKY92-enriched PCA manifold derived from symptomatic MM samples. As shown in **Figure 3A**, in this projection SMM samples do not form a distinct cluster along PC1 or PC2. Instead, they localize near the origin of PC1, indicating attenuation of the SKY92-associated covariance structure that separates HR and SR within symptomatic MM.

This distribution suggests that the transcriptional structure associated with SKY92 risk in symptomatic MM is less pronounced in SMM. However, within SMM, projected PC1 scores were higher in SKY92 HR than SR samples (mean, 3.91 versus −3.61; difference, 7.52; 95% CI, 0.23–14.80; *P* = 0.044), indicating that the symptomatic MM-derived high-risk-associated transcriptional structure remained detectable within SMM. To further characterize the contribution of individual EMC92 signature probes to SKY92 risk classification in SMM, we calculated their weighted contributions to the HR–SR expression difference using the published EMC92 coefficients and reference standard deviations. *SFMBT1*, *ZBTB25* and *FGFR3* showed the largest absolute contributions (+0.246, +0.094 and +0.091, respectively). Leave-one-out sensitivity analysis confirmed the stability of these findings, with *SFMBT1* remaining the highest-ranking contributor in all 27 iterations and *ZBTB25* and FGFR3 ranking among the top three in 26/27 and 25/27 iterations, respectively (**Figure 3B-C and Supplemental Table 2**). These results indicate unequal contributions of individual EMC92 probes to SKY92 HR–SR separation in SMM, without establishing their independent prognostic significance. Notably, SKY92 HR SMM shared transcriptional features with symptomatic SKY92 HR, including enrichment of cell-proliferation programs (E2F targets, NES = 1.85, *FDR* = 4.18E$^{-04}$ and G2M checkpoint, NES = 1.85, *FDR* = 3.90E$^{-04}$) and DNA damage response (UV response NES = 1.56, *FDR* = 0.008). In addition, HR SMM showed enrichment of additional inflammatory and immune signaling programs, including TNFα/NF-κB, IL6-JAK-STAT3 and interferon responses, together with KRAS signaling (**Figure 3D, Supplemental Figure S1E-F** and **Supplemental Table 2**). The involvement of inflammatory signaling and KRAS/MAPK alterations in precursor disease evolution is consistent with previous studies[11,12]. Importantly, our findings extend these observations by identifying both shared and disease-stage-associated transcriptional programs within SKY92-defined high-risk disease across the MM continuum.

Nevertheless, SKY92 HR SMM patients demonstrated a significantly increased risk of progression to symptomatic MM compared with SR patients. In this cohort, the progression rate at 36 months was 70% in SKY92 HR SMM compared with 8.4% in the SR group (*P* = 0.038, **Figure 3E**). When considering only untreated patients, the progression rate at 36 months was 70% in the SKY92 HR group compared with 0% in the SR group (*P* = 0.173, **Supplemental Figure S1G**)**.** Together**,** these data suggest that SKY92-HR SMM may capture emerging aggressive biological features even though SMM transcriptomes do not fully recapitulate the SKY92 covariance patterns observed in symptomatic disease. An increasing number of studies recommend the application of SKY92-based tests for the identification of MM patients with HR disease [6] with a reported contribution in the risk-adapted therapeutic decision-making[13]. The findings of this pilot study further support the feasibility of implementing GEP-based risk profiling in routine clinical practice and suggest that transcriptomic risk assessment may also provide biologically relevant information in SMM, pending validation in larger independent cohorts. These observations are particularly relevant in light of recent evidence showing that the early identification and treatment of HR SMM can delay the progression to symptomatic MM, prevent end organ damage and improve clinical outcomes [14,15]. However, in this context, the identification of gene signatures more tailored to early asymptomatic disease and the reduction of input material requirements represent key priorities for advancing GEP-based technologies, improving both performance and feasibility in patients with low disease burden. Overall, our results support the existence of shared HR transcriptional programs across asymptomatic and symptomatic MM disease states and fuel the ongoing debate on the biological definition of SMM, especially given accumulating evidence of overlapping genomic features between NDMM and a subset of SMM[12,16].

**References**

1. Sonneveld P, Dimopoulos MA, Boccadoro M, et al. Daratumumab, bortezomib, lenalidomide, and dexamethasone for multiple myeloma. N Engl J Med. 2024;390(4):301-313.

2. Avet-Loiseau H, Davies FE, Samur MK, et al. International Myeloma Society/International Myeloma Working Group consensus recommendations on the definition of high-risk multiple myeloma. J Clin Oncol. 2025;43(24):2739-2751.

3. Lu X, Andersen EF, Banerjee R, et al. Guidelines for the testing and reporting of cytogenetic results for risk stratification of multiple myeloma: a report of the Cancer Genomics Consortium Plasma Cell Neoplasm Working Group. Blood Cancer J. 2025;15(1):86.

4. Decaux O, Lodé L, Magrangeas F, et al. Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the Intergroupe Francophone du Myélome. J Clin Oncol. 2008;26(29):4798-4805.

5. Kuiper R, Broyl A, de Knegt Y, et al. A gene expression signature for high-risk multiple myeloma. Leukemia. 2012;26(11):2406-2413.

6. Zhou X, Hofmann A, Engel B, et al. Combining SKY92 gene expression profiling and FISH (according to R2-ISS) defines ultra-high-risk Multiple Myeloma. Hemasphere. 2025;9(1):e70078.

7. Shaughnessy JD Jr, Zhan F, Burington BE, et al. A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. Blood. 2007;109(6):2276-2284.

8. Zicchedda B, Giannotta C, D'Agostino M, et al. Genomic and immune determinants of resistance to daratumumab-based therapy in relapsed refractory multiple myeloma. Blood Cancer J. 2024;14(1):117.

9. Marchesini M, Ogoti Y, Fiorini E, et al. ILF2 Is a regulator of RNA splicing and DNA damage response in 1q21-amplified multiple myeloma. Cancer Cell. 2017;32(1):88-100.e6.

10. Kuiper R, Zweegman S, van Duin M, et al. Prognostic and predictive performance of R-ISS with SKY92 in older patients with multiple myeloma: the HOVON-87/NMSG-18 trial. Blood Adv. 2020;4(24):6298-6309.

11. Botta C, Di Martino MT, Ciliberto D, et al. A gene expression inflammatory signature specifically predicts multiple myeloma evolution and patients survival. Blood Cancer J. 2016;6(12):e511.

12. Boyle EM, Deshpande S, Tytarenko R, et al. The molecular make up of smoldering myeloma highlights the evolutionary pathways leading to multiple myeloma. Nat Commun. 2021;12(1):293.

13. Biran N, Dhakal B, Niesvizky R, et al. Enhancing risk stratification and treatment decision in multiple myeloma with SKY92 gene expression profiling in real-world data. Br J Haematol. 2025;206(6):1642-1653.

14. Dimopoulos MA, Voorhees PM, Schjesvold F, et al. Daratumumab or active monitoring for high-risk smoldering multiple myeloma. N Engl J Med. 2025;392(18):1777-1788.

15. Mateos MV, Hernández MT, Salvador C, et al. Lenalidomide-dexamethasone versus observation in high-risk smoldering myeloma after 12 years of median follow-up time: A randomized, open-label study. Eur J Cancer. 2022;174:243-250.

16. Maura F, Bergsagel PL, Ziccheddu B, et al. Genomics define malignant transformation in myeloma precursor conditions. J Clin Oncol. 2026;44(3):188-199.

**Figure 1. Clinical, molecular and SKY92-based characterization of the multiple myeloma cohort. (A)** Circos plot showing clinical and molecular features of the study cohort (n = 90 patients). The outer lane classifies the disease stage (SMM = 27, NDMM = 39 and RRMM = 24, scale of green). SKY92 vFISH cytogenetics data including t(11;14), t(4;14), t(14;16/20) gain(1q), del(17p) and del(13q) (grey boxes) and gene expression signature of risk (SR = pale blue and HR = blue) are shown. IMWG 2/20/20 (SMM) and ISS (NDMM, RRMM) classes are depicted in pink and yellow scale respectively. Inner heatmaps show the biochemical levels of Lactate Dehydrogenase (LDH, blue), Albumin (ALB, brown), Creatinine (CRE, green), Hemoglobin (Hb, violet) and β2 Microglobulin (β2M). Dark red histograms indicate the percentage of plasma cell infiltration in the BM as assessed by BM biopsy analysis. MM isotype, sex and age are depicted in the three inner circles. White boxes indicate missing data. **(B)** Results of SKY92 test across disease stages calculated as a percentage of SKY92 HR detected samples on the total number of patients included in each category. **(C)** Diagnostic performance of the assay for detection of recurrent cytogenetic abnormalities. Bar graphs show sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall accuracy for detection of t(4;14), t(14;16)/t(14;20), gain(1q), and del(17p). Values are expressed as percentages. Error bars represent 95% confidence intervals (see **Supplemental Table 2** for related statistics).

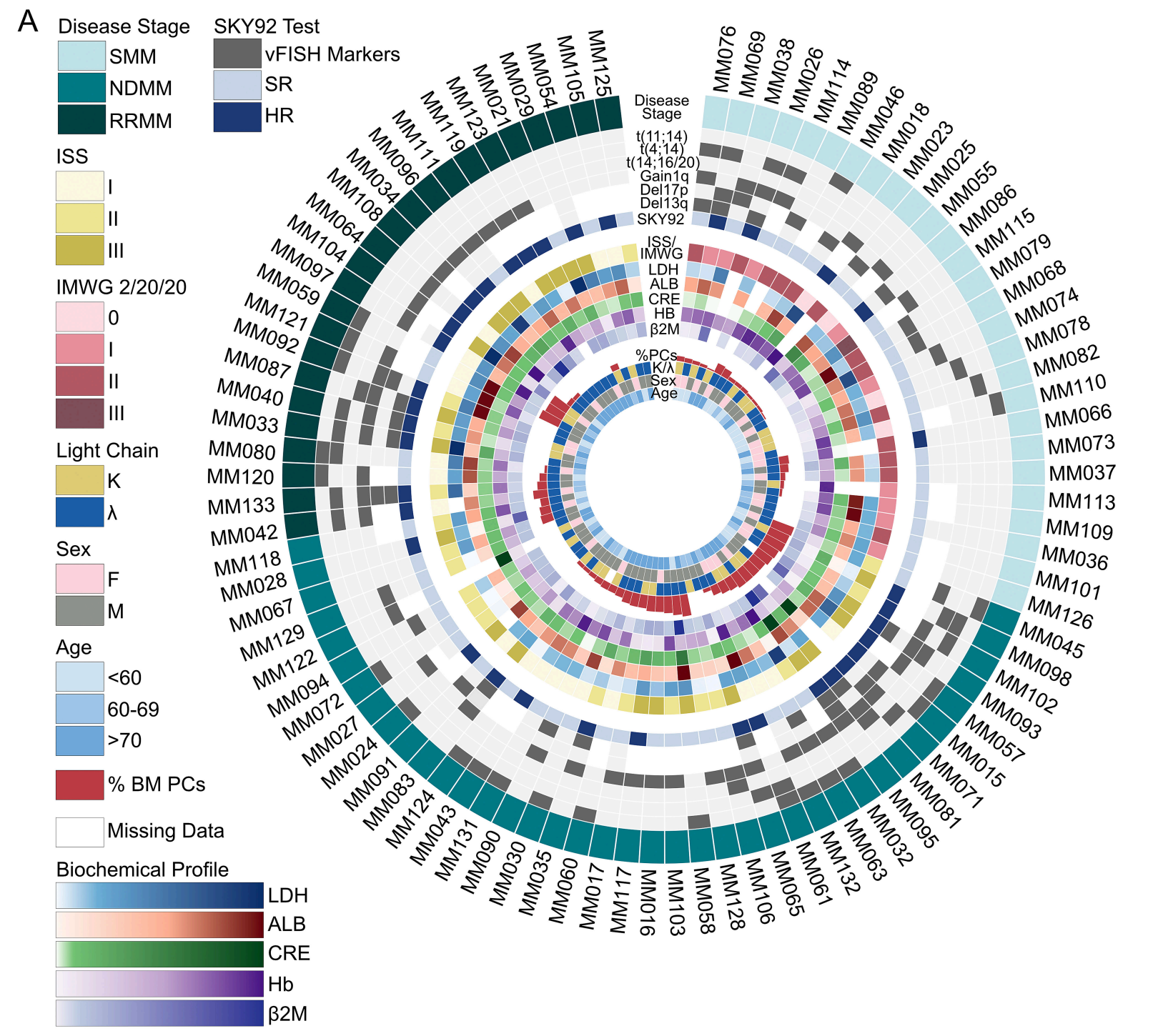
**Figure 2. Clinical and transcriptional features associated with SKY92 risk in symptomatic MM and validation in an independent NDMM cohort. (A)** Kaplan-Meier survival curves of SKY92 HR (n = 16) and SKY92 SR (n = 23) NDMM patients. X-axis, time elapsed since sample collection (months); y-axis probability of survival. Logrank *P value* and number of patients at risk are shown. **(B)** Kaplan-Meier stratification of overall survival (OS) by SKY92 risk classification (HR = 16 and SR = 8) in RRMM patients. X-axis, time elapsed since sample collection (months); y-axis probability of survival. Logrank *P value* and number

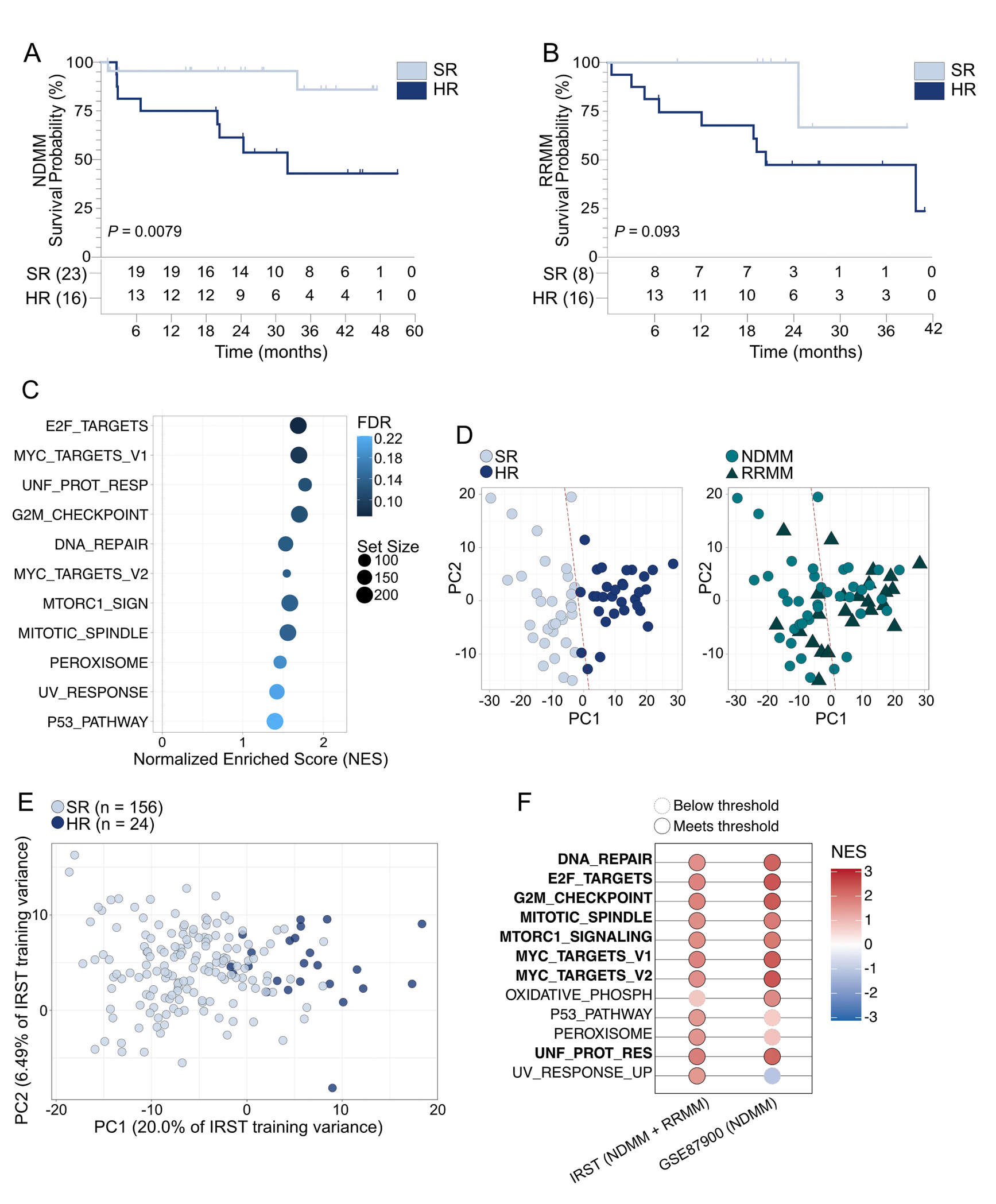
of patients at risk are shown. **(C)** Dot plot showing significantly enriched Hallmark gene sets in SKY92 HR (n = 32) compared with SR (n = 31) patients with symptomatic MM (NDMM and RRMM), as detected by GSEA (FDR q < 0.25). The x-axis shows the normalized enrichment score (NES); dot size represents gene-set size and color indicates the FDR. **(D)** PCA score plots of PC1 versus PC2 based on a supervised gene set associated with SKY92 risk status in symptomatic MM (NDMM and RRMM). After exclusion of low-expression genes (mean MAS5 $\log_2$ expression < 7), genes associated with SKY92 HR versus SR status were selected using raw P < 0.01 and absolute $\log_2$ fold change > 0.5, yielding 822 probes. Left: samples are stratified by SKY92 risk classification (SR and HR). Right: the same samples are displayed within the same PCA space according to disease stage (NDMM and RRMM). **(E)** Projection of 180 newly diagnosed multiple myeloma (NDMM) samples from the independent GSE87900 cohort (24 SKY92 high-risk and 156 standard-risk) onto the PCA space derived from the IRST symptomatic MM cohort. Projection was performed using the original 822 probe sets and the corresponding IRST training means, standard deviations, and PCA loadings, without refitting the PCA. **(F)** Comparison of Hallmark enrichment profiles between the IRST symptomatic MM cohort (pooled NDMM and RRMM; 32 HR and 31 SR) and GSE87900. NES values are shown for pathways enriched in HR samples in at least one cohort at FDR < 0.25. Solid outlines indicate pathways meeting this threshold, whereas dashed outlines indicate pathways below the threshold. Eight pathways were enriched in HR samples in both cohorts. The corresponding GSEA results and cross-cohort comparison are provided in **Supplemental Table 2**.

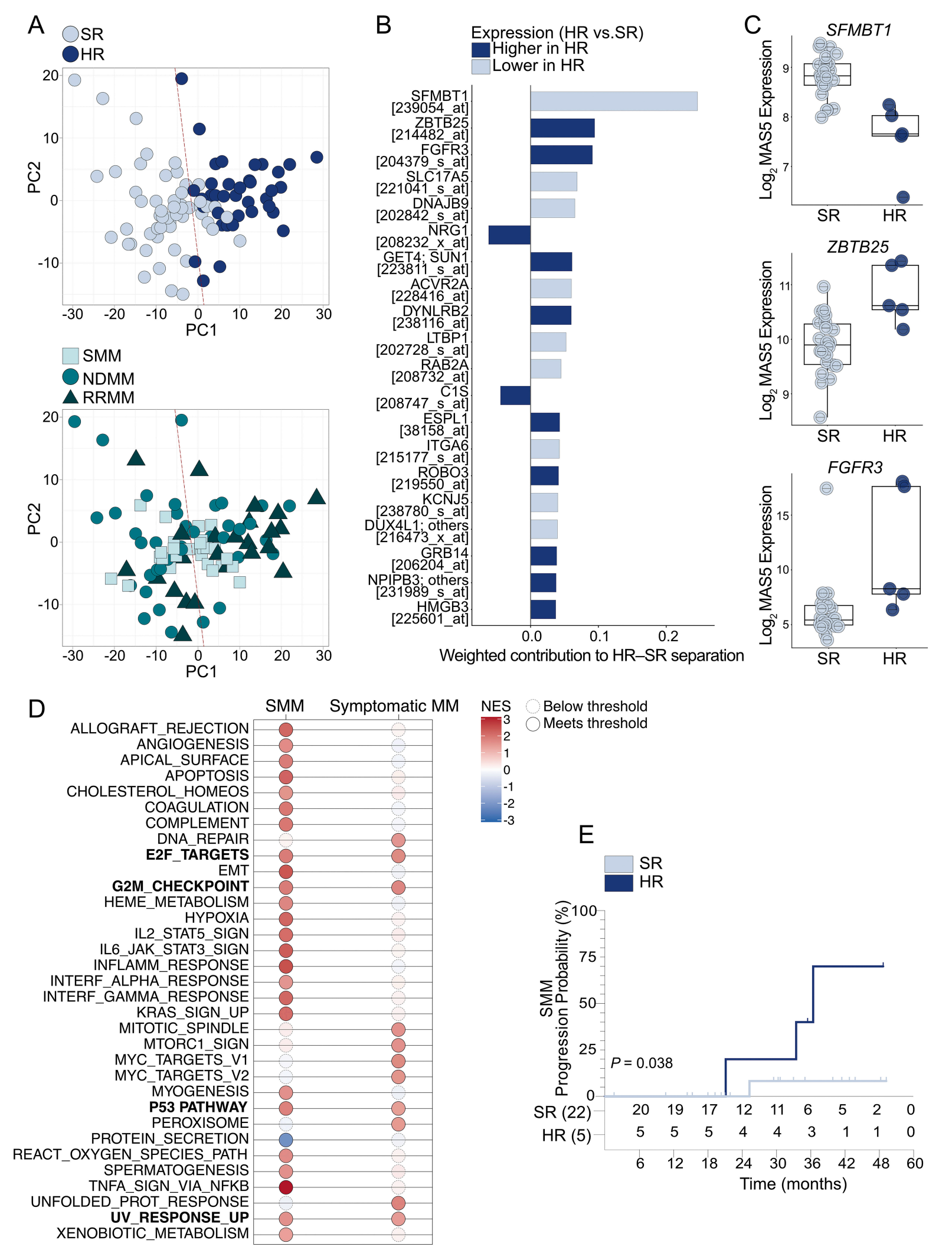
**Figure 3. Transcriptional features associated with SKY92 risk in smoldering multiple myeloma and progression to symptomatic disease. (A)** PCA score plots of PC1 versus PC2 based on a supervised gene set associated with SKY92 risk status in symptomatic MM (NDMM and RRMM). After exclusion of low-expression genes (mean MAS5 $\log_2$ expression

< 7), genes associated with SKY92 HR versus SR status were selected using raw $P < 0.01$ and absolute $\log_2$ fold change > 0.5, yielding 822 probes. The PCA space was derived from symptomatic MM samples, with SMM samples excluded from feature selection and subsequently projected into the resulting 822-probe transcriptional space. Top: samples stratified by SKY92 risk classification (SR and HR); the dashed red line indicates the main direction of separation. Bottom: the same samples displayed within the same PCA space according to disease stage (SMM, NDMM and RRMM). **(B)** Top 20 EMC92 probes ranked by the absolute magnitude of their weighted contributions to the difference between SKY92 high-risk (HR, n = 5) and standard-risk (SR, n = 22) SMM samples. Contributions were calculated as the published EMC92 coefficient multiplied by the mean HR-SR $\log_2$ expression difference, divided by the corresponding reference standard deviation from the original H65 training cohort (n = 290; without APEX correction). Positive and negative values indicate the direction of the weighted contribution to HR-SR separation, whereas bar colors indicate higher or lower probe expression in HR relative to SR samples. The individual contributions of all 92 EMC92 probes are provided in **Supplemental Table 2. (C)** Individual $\log_2$-transformed MAS5 expression values of *SFMBT1, ZBTB25* and *FGFR3,* the three highest-ranking contributors, in HR and SR SMM samples. Probe contributions were ranked descriptively; no statistical significance testing was performed on the weighted contributions. **(D)** Comparison of Hallmark GSEA profiles in SMM and symptomatic MM. Hallmark gene-set enrichment analysis (GSEA) profiles from separate SKY92 HR-versus-SR comparisons performed within SMM (5 HR, 22 SR) and symptomatic MM (pooled NDMM and RRMM; 32 HR, 31 SR). The 33 pathways meeting the significance threshold in either analysis are shown; pathways meeting the significance threshold in both analyses are highlighted in bold. For each disease-stage cohort, normalized enrichment scores (NES) were derived from the corresponding within-cohort HR-versus-SR comparison; positive NES indicates enrichment in SKY92 HR samples and negative NES enrichment in SKY92 SR samples. Solid outlines

indicate pathways meeting the significance threshold in the corresponding analysis, whereas dashed outlines indicate pathways below the threshold. SMM GSEA used gene-set permutations with *FDR* < 0.05, whereas symptomatic MM GSEA used phenotype permutations with *FDR* < 0.25. Because different permutation methods and significance thresholds were used to account for differences in cohort size, NES magnitudes and FDR values are not directly comparable between the two analyses. No direct statistical comparison between SMM and symptomatic MM was performed. **(E)** Kaplan-Meier curves of progression to symptomatic MM in SKY92 HR (n = 5) and SKY92 SR (n = 22) SMM patients. X-axis, time elapsed since sample collection (months); y-axis, probability of progression to symptomatic MM. Logrank *P* value and number of patients at risk are shown.


A
Disease Stage
SMM
NDMM
RRMM
SKY92 Test
vFISH Markers
SR
HR
ISS
I
II
III
IMWG 2/20/20
0
I
II
III
Light Chain
K
λ
Sex
F
M
Age
<60
60-69
>70
% BM PCs
Missing Data
Biochemical Profile
LDH
ALB
CRE
Hb
β2M
Disease Stage
t(11;14)
t(4;14)
t(14;16/20)
Gain1q
Del17p
Del13q
SKY92
ISS/ IMWG
LDH
ALB
CRE
HB
β2M
%PCs
K/λ
Sex
Age
MM076
MM069
MM038
MM026
MM114
MM089
MM046
MM018
MM023
MM025
MM055
MM086
MM115
MM079
MM068
MM074
MM078
MM082
MM110
MM066
MM073
MM037
MM113
MM109
MM036
MM101
MM126
MM045
MM098
MM102
MM093
MM057
MM015
MM071
MM081
MM095
MM032
MM063
MM132
MM061
MM065
MM106
MM128
MM058
MM103
MM016
MM117
MM017
MM060
MM035
MM030
MM090
MM131
MM043
MM124
MM083
MM091
MM024
MM027
MM072
MM094
MM122
MM129
MM067
MM028
MM118
MM042
MM133
MM120
MM080
MM033
MM040
MM087
MM092
MM121
MM059
MM097
MM104
MM064
MM108
MM034
MM096
MM111
MM119
MM123
MM021
MM029
MM054
MM105
MM125
B
SKY92 SR
SKY92 HR
Cohort Patients (%)
100
75
50
25
0
82%
59%
33%
18%
41%
67%
SMM
NDMM
RRMM
C
Sensitivity
Specificity
Positive Predictive Value
Negative Predictive Value
Accuracy
% Value
100
75
50
25
0
t(4;14)
t(14;16)/(14;20)
Gain(1q)
Del(17p)


A
NDMM
Survival Probability (%)
100
75
50
25
0
SR
HR
P = 0.0079
SR (23) 19 19 16 14 10 8 6 1 0
HR (16) 13 12 12 9 6 4 4 1 0
6 12 18 24 30 36 42 48 60
Time (months)
B
RRMM
Survival Probability (%)
SR
HR
P = 0.093
SR (8) 8 7 7 3 1 1 0
HR (16) 13 11 10 6 3 3 0
6 12 18 24 30 36 42
Time (months)
C
E2F_TARGETS
MYC_TARGETS_V1
UNF_PROT_RESP
G2M_CHECKPOINT
DNA_REPAIR
MYC_TARGETS_V2
MTORC1_SIGN
MITOTIC_SPINDLE
PEROXISOME
UV_RESPONSE
P53_PATHWAY
FDR
0.22
0.18
0.14
0.10
Set Size
100
150
200
0 1 2
Normalized Enriched Score (NES)
D
SR
HR
NDMM
RRMM
PC2
PC1
20
10
0
-10
-30 -20 -10 0 10 20 30
E
SR (n = 156)
HR (n = 24)
PC2 (6.49% of IRST training variance)
PC1 (20.0% of IRST training variance)
10
0
-20 -10 0 10 20
F
Below threshold
Meets threshold
NES
3
2
1
0
-1
-2
-3
DNA_REPAIR
E2F_TARGETS
G2M_CHECKPOINT
MITOTIC_SPINDLE
MTORC1_SIGNALING
MYC_TARGETS_V1
MYC_TARGETS_V2
OXIDATIVE_PHOSPH
P53_PATHWAY
PEROXISOME
UNF_PROT_RES
UV_RESPONSE_UP
IRST (NDMM + RRMM)
GSE87900 (NDMM)

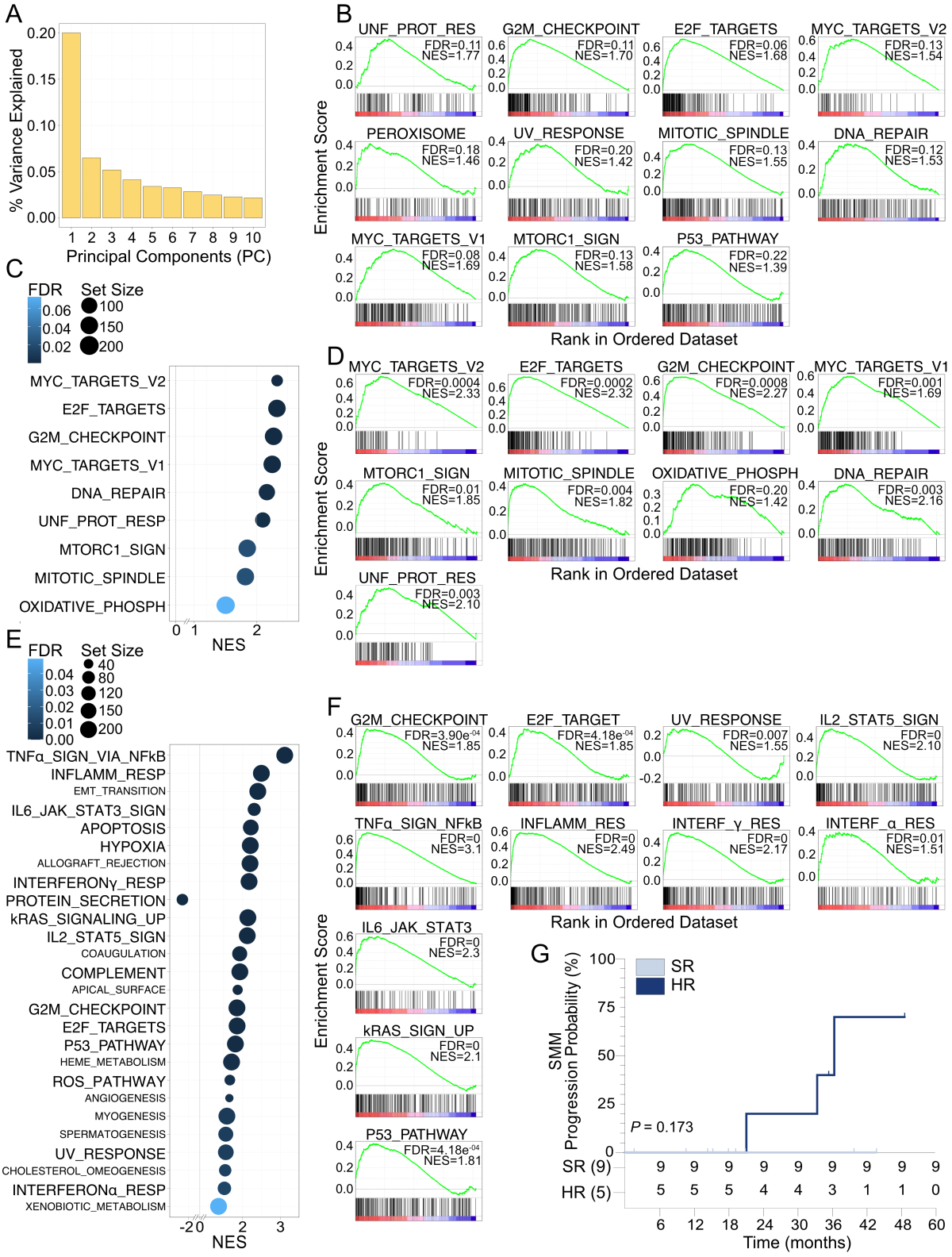

A
SR
HR
PC2
PC1
SMM
NDMM
RRMM
B
Expression (HR vs.SR)
Higher in HR
Lower in HR
SFMBT1 [239054_at]
ZBTB25 [214482_at]
FGFR3 [204379_s_at]
SLC17A5 [221041_s_at]
DNAJB9 [202842_s_at]
NRG1 [208232_x_at]
GET4; SUN1 [223811_s_at]
ACVR2A [228416_at]
DYNLRB2 [238116_at]
LTBP1 [202728_s_at]
RAB2A [208732_at]
C1S [208747_s_at]
ESPL1 [38158_at]
ITGA6 [215177_s_at]
ROBO3 [219550_at]
KCNJ5 [238780_s_at]
DUX4L1; others [216473_x_at]
GRB14 [206204_at]
NPIPB3; others [231989_s_at]
HMGB3 [225601_at]
Weighted contribution to HR–SR separation
C
SFMBT1
ZBTB25
FGFR3
Log2 MAS5 Expression
SR
HR
D
SMM
Symptomatic MM
NES
Below threshold
Meets threshold
ALLOGRAFT_REJECTION
ANGIOGENESIS
APICAL_SURFACE
APOPTOSIS
CHOLESTEROL_HOMEOS
COAGULATION
COMPLEMENT
DNA_REPAIR
E2F_TARGETS
EMT
G2M_CHECKPOINT
HEME_METABOLISM
HYPOXIA
IL2_STAT5_SIGN
IL6_JAK_STAT3_SIGN
INFLAMM_RESPONSE
INTERF_ALPHA_RESPONSE
INTERF_GAMMA_RESPONSE
KRAS_SIGN_UP
MITOTIC_SPINDLE
MTORC1_SIGN
MYC_TARGETS_V1
MYC_TARGETS_V2
MYOGENESIS
P53 PATHWAY
PEROXISOME
PROTEIN_SECRETION
REACT_OXYGEN_SPECIES_PATH
SPERMATOGENESIS
TNFA_SIGN_VIA_NFKB
UNFOLDED_PROT_RESPONSE
UV_RESPONSE_UP
XENOBIOTIC_METABOLISM
E
SR
HR
SMM Progression Probability (%)
P = 0.038
SR (22)
HR (5)
Time (months)

**Supplemental Table 1 (related to Figure 1).** Cohort description: biochemical and genetic characteristics.

| ***Disease Status n (%)*** | *SMM 27 (30.0)* | *NDMM 39 (43.3)* | *RRMM 24 (26.7)* |
|---|---|---|---|
| *Male, n (%)* | 13 (48.1) | 27 (69.2) | 16 (66.7) |
| *Female, n (%)* | 14 (51.9) | 12 (30.8) | 8 (33.3) |
| ***Age (years), median (range)*** | 64.0 (36-83) | 69 (52-91) | 70 (56-88) |
| **ISS/ *IMWG 2/20/20*** | | | |
| *Stage 0 (2/20/20 only)* | 4 (14.8) | NA | NA |
| *Stage I* | 9 (33.3) | 11 (28.2) | 10 (41.7) |
| *Stage II* | 13 (48.1) | 15 (38.4) | 5 (20.8) |
| *Stage III* | 1 (3.7) | 11 (28.2) | 9 (37.5) |
| *NA, n (%)* | 0 | 2 (5.1) | 0 (0) |
| ***BM infiltration (%), median (range)*** | 15 (5-40) | 55 (15-95) | 35 (10-90) |
| ***Albumin (g/L), median (range)*** | 41 (38-49) | 37 (28-51) | 41 (31-49) |
| *> 35 (G/L), n (%)* | 22 (100) | 27 (70) | 22 (91.7) |
| *< 35 (G/L), n (%)* | 0 (0) | 11 (30) | 2 (8.3) |
| *NA, n (%)* | 5 (18.5) | 1 (2.6) | 0 (0) |
| ***B2M (mg/L), median (range)*** | 2.3 (1.3-13.1) | 3.7 (1.1-76.6) | 3.7 (1.9-20.5) |
| *< 3 (mg/L), n (%)* | 14 (63.6) | 10 (27) | 8 (34.8) |
| *> 3 (mg/L), n (%)* | 8 (36.4) | 27 (73) | 15 (65.2) |
| *NA, n (%)* | 5 (18.5) | 2 (5.1) | 1 (4.2) |
| ***LDH (U/L), median (range)*** | 152 (124-352) | 160 (94-376) | 209 (115-499) |
| *< 225 (U/L), n (%)* | 19 (86.4) | 34 (94.5) | 15 (65.2) |
| *> 225 (U/L), n (%)* | 3 (13.6) | 2 (5.5) | 8 (34.8) |
| *NA, n (%)* | 5 (18.5) | 3 (7.7) | 1 (4.2) |
| ***Creatinine (mg/dl), median (range)*** | 0.9 (0.2-3.3) | 1.1 (0.5-8.6) | 1 (0.7-2.2) |
| > 1.20 *(mg/dl), n (%)* | 5 (20.8) | 17 (43.6) | 8 (33.3) |
| < 1.20 *(mg/dl), n (%)* | 19 (79.1) | 22 (56.4) | 16 (66.7) |
| *NA, n (%)* | 3 (11.1) | 0 (0) | 0 (0) |
| ***Hemoglobin (g/dl), median (range)*** | 12.9 (9-14.9) | 10.7 (7.5-15) | 11.6 (8.8-15.3) |

| | | | |
|---|---|---|---|
| 13.5-17 *(g/dl), n (%)* | 8 (30.7) | 6 (15.4) | 3 (12.5) |
| < 13.5 *(g/dl), n (%)* | 18 (69.3) | 33 (84.6) | 21 (87.5) |
| *NA, n (%)* | 1 (3.7) | 0 (0) | 0 (0) |
| ***SKY92 Valid QC (CD38$^{+}$ ≥80% following CD138$^{+}$ enrichment), n (%)*** | 27 (81.8) | 39 (92.8) | 24 (80) |
| ***SKY92 vFISH Cytogenetics, n (%)*** | | | |
| *t (4;14)* | 5 (18.5) | 5 (12.8) | 5 (20.8) |
| *NA* | 0 (0) | 0 (0) | 0 (0) |
| *t (11;14)* | 2 (7.4) | 15 (38.4) | 5 (20.8) |
| *NA* | 0 (0) | 0 (0) | 0 (0) |
| *t (14;16)/t (14;20)* | 1 (3.7) | 4 (10.2) | 1 (4.1) |
| *NA* | 0 (0) | 0 (0) | 0 (0) |
| *gain (1q)* | 8 (29.6) | 23 (58.9) | 17 (70.8) |
| *NA* | 0 (0) | 0 (0) | 0 (0) |
| *del (17p)* | 8 (47.1) | 5 (25.0) | 2 (14.3) |
| *NA* | 10 (37.0) | 19 (48.7) | 10 (41.6) |
| *del (13q)* | 3 (17.6) | 8 (40.0) | 4 (28.6) |
| *NA* | 10 (37.0) | 19 (48.7) | 10 (41.6) |
| ***SKY92 GE Signature, n (%)*** | | | |
| *High Risk* | 5 (18.5) | 16 (41) | 16 (67) |
| *Standard Risk* | 22 (81.5) | 23 (59) | 8 (33) |

SMM = smoldering multiple myeloma; NDMM = newly diagnosed multiple myeloma; RRMM = relapse/refractory multiple myeloma; ISS = International Staging System (stages I-III); IMWG 2/20/20 = International Myeloma Working Group risk classification for SMM (stages 0-III); BM = bone marrow; B2M = β2 microglobulin; LDH = lactate dehydrogenase; QC = quality control; vFISH = virtual FISH; GE = gene expression; NA = not available.




**Supplemental Figure S1 (related to Figures 2 and 3): Transcriptional features associated with SKY92 risk in symptomatic and smoldering multiple myeloma. (A)** Scree plot displaying the percentage of total variance explained by the first 10 principal components (PCs) based on the 822-probe supervised gene set associated with SKY92 risk

status in symptomatic MM (NDMM and RRMM). PC1 explains the largest proportion of variance (~20%), followed by PC2 (~7%) and PC3 (~5%), with progressively smaller contributions from subsequent components. **(B)** GSEA running score plots of significantly enriched Hallmark gene sets from the Molecular Signatures Database (MSigDB) in SKY92 HR (n = 32) compared with SR (n = 31) patients with symptomatic MM (NDMM and RRMM; *FDR* < 0.25). Normalized enrichment score (NES) and FDR q-value are indicated for each gene set. Statistics and enriched gene lists are included in **Supplemental Table 2**. **(C)** Hallmark GSEA comparing SKY92 HR and SR samples in GSE87900. The dot plot shows normalized enrichment scores for the nine pathways enriched in HR samples at *FDR* < 0.25; dot size represents gene-set size and color indicates FDR. **(D)** Representative enrichment plots of HR-associated Hallmark pathways in GSE87900, showing NES and FDR values. **(E)** Dot plot showing significantly enriched Hallmark gene sets in SKY92 HR (n = 5) compared with SR (n = 22) SMM patients, as detected by GSEA (*FDR* < 0.05). The x-axis shows the NES; dot size represents gene-set size and color indicates the FDR. **(F)** GSEA running score plots of significantly enriched Hallmark gene sets from the MSigDB in SKY92 HR (n = 5) compared with SR (n = 22) SMM patients (*FDR* < 0.05). NES and FDR are indicated for each gene set. Statistics and enriched gene lists are included in **Supplemental Table 2**. **(G)** Kaplan–Meier curves of progression to symptomatic MM in untreated SKY92 HR (n = 5) and SKY92 SR (n = 9) SMM patients. X-axis, time elapsed since sample collection (months); y-axis, probability of progression to symptomatic MM. Log-rank P value and number of patients at risk are shown.

**Supplemental Table 2 provided as Excel file. Additional statistical and transcriptional analyses.** The Excel file contains three sections**:** ***(i)*** vFISH Performance, diagnostic performance of virtual FISH predictions across multiple myeloma cohorts; ***(ii)*** GSEA Analysis, Hallmark gene set

enrichment analysis results across the investigated cohorts; and ***(iii)*** Individual Genes Score SMM, individual probe-level contributions of the 92 EMC92 probes to SKY92 risk classification in SMM.