



## Genome-Wide Somatic Alterations in Multiple Myeloma Reveal a Superior Outcome Group

Samur Mehmet Kemal, Anil Aktas Samur, Mariateresa Fulciniti, Raphael Szalat, Tessa Han, Masood Shamma, Paul Richardson, Florence Magrangeas, Stéphane Minvielle, Jill Corre, et al.

### ► To cite this version:

Samur Mehmet Kemal, Anil Aktas Samur, Mariateresa Fulciniti, Raphael Szalat, Tessa Han, et al..  
Genome-Wide Somatic Alterations in Multiple Myeloma Reveal a Superior Outcome Group. *Journal of Clinical Oncology*, 2020, 38, pp.JCO2000461. 10.1200/JCO.20.00461 . inserm-02913784

**HAL Id: inserm-02913784**

**<https://inserm.hal.science/inserm-02913784>**

Submitted on 10 Aug 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genome-Wide Somatic Alterations in Multiple Myeloma Reveal a Superior Outcome Group

Mehmet Kemal Samur, PhD<sup>1,2,3</sup>; Anil Aktas Samur, PhD<sup>1,2</sup>; Mariateresa Fulciniti, PhD<sup>3</sup>; Raphael Szalat, MD, PhD<sup>3</sup>; Tessa Han, BS<sup>1</sup>; Masood Shammash, PhD<sup>3</sup>; Paul Richardson, MD<sup>3</sup>; Florence Magrangeas, PhD<sup>4</sup>; Stephane Minvielle, PhD<sup>4</sup>; Jill Corre, PharmD, PhD<sup>5</sup>; Philippe Moreau, MD<sup>4</sup>; Anjan Thakurta, PhD<sup>6</sup>; Kenneth C. Anderson, MD<sup>3</sup>; Giovanni Parmigiani, PhD<sup>1,2</sup>; Hervé Avet-Loiseau, MD, PhD<sup>5</sup>; and Nikhil C. Munshi, MD<sup>3,7</sup>

**PURPOSE** Multiple myeloma (MM) is accompanied by heterogeneous somatic alterations. The overall goal of this study was to describe the genomic landscape of myeloma using deep whole-genome sequencing (WGS) and develop a model that identifies patients with long survival.

**METHODS** We analyzed deep WGS data from 183 newly diagnosed patients with MM treated with lenalidomide, bortezomib, and dexamethasone (RVD) alone or RVD + autologous stem cell transplant (ASCT) in the IFM/DFCI 2009 study (ClinicalTrials.gov identifier: [NCT01191060](https://clinicaltrials.gov/ct2/show/study?term=NCT01191060)). We integrated genomic markers with clinical data.

**RESULTS** We report significant variability in mutational load and processes within MM subgroups. The timeline of observed activation of mutational processes provides the basis for 2 distinct models of acquisition of mutational changes detected at the time of diagnosis of myeloma. Virtually all MM subgroups have activated DNA repair–associated signature as a prominent late mutational process, whereas APOBEC signature targeting C>G is activated in the intermediate phase of disease progression in high-risk MM. Importantly, we identify a genomically defined MM subgroup (17% of newly diagnosed patients) with low DNA damage (low genomic scar score with chromosome 9 gain) and a superior outcome (100% overall survival at 69 months), which was validated in a large independent cohort. This subgroup allowed us to distinguish patients with low- and high-risk hyperdiploid MM and identify patients with prolongation of progression-free survival. Genomic characteristics of this subgroup included lower mutational load with significant contribution from age-related mutations as well as frequent *NRAS* mutation. Surprisingly, their overall survival was independent of International Staging System and minimal residual disease status.

**CONCLUSION** This is a comprehensive study identifying genomic markers of a good-risk group with prolonged survival. Identification of this patient subgroup will affect future therapeutic algorithms and research planning.

*J Clin Oncol* 38. © 2020 by American Society of Clinical Oncology

Creative Commons Attribution Non-Commercial No Derivatives 4.0 License 

## INTRODUCTION

Multiple myeloma (MM), is a plasma cell malignancy that demonstrates significant genomic and clinical heterogeneity even at the time of diagnosis.<sup>1-3</sup> Although disease outcomes have significantly improved over the last decade because of new treatment options, it remains incurable. Genomic characterization using whole-exome or targeted sequencing of large numbers of primary samples from newly diagnosed patients with MM has revealed potential driver mutations and identified several high-risk features.<sup>4-7</sup> In contrast, initial whole-genome sequencing (WGS) in MM<sup>6,8</sup> has profiled smaller patient cohorts and focused mainly on coding changes, therefore lacking the statistical power and analytical methodologies to robustly delineate the genomic landscape of somatic alterations in MM.

The majority of somatic mutations arise in the non-coding region of the genome<sup>9,10</sup>; however, little is known about their function.<sup>9-15</sup> Moreover, the processes contributing to the somatic architecture of MM have also not been evaluated in detail. WGS has now made it possible to interrogate causal processes underlying genomic heterogeneity in newly diagnosed MM. This heterogeneity in MM,<sup>16-18</sup> as in other cancers, reflects changes in cellular pathways in the MM clone, as well as the tumor microenvironment, which promote MM cell growth and drug resistance, as well as overcoming immune surveillance.

Here, we report the initial results of interrogation of the MM genome in uniformly treated patients. Our study uses deep WGS (median tumor depth, 75×) from patients newly diagnosed with MM enrolled in a phase III clinical trial. Integration of these data provides the landscape of genomic aberrations and mutational

## ASSOCIATED CONTENT

### Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on June 5, 2020 and published at [ascopubs.org/journal/jco](https://ascopubs.org/journal/jco) on July 20, 2020; DOI <https://doi.org/10.1200/JCO.20.00461>

## CONTEXT

### Key Objective

To define the genomic markers of long-term survival in multiple myeloma (MM).

### Knowledge Generated

Comprehensive large whole-genome sequencing data showed that mutations in subclones were enriched by DNA repair–associated processes. We show that patients with low genomic scar score and chromosome 9 gain have a superior outcome in MM and that traditional risk markers are not adequate to identify this subgroup of patients.

### Relevance

A number of previous studies have identified various clinical, biologic, and genomic changes indicative of high-risk disease; however, this is the first comprehensive study to our knowledge describing genomic markers of a good-risk group with prolonged survival. Identification of the good-risk patients will affect and inform both clinical research and patient care.

processes that contribute to MM development and identifies a patient subgroup with superior outcome, validated in an independent data set.

## METHODS

### Patient Samples

All samples from patients with MM were collected from the IFM/DFCI 2009<sup>19</sup> clinical trial (ClinicalTrials.gov identifier: [NCT01191060](#)) after written informed consent, and clinical and genomic data were de-identified in accordance with the Declaration of Helsinki. Myeloma cells were purified from bone marrow by CD138+ cell selection, and control DNA originated from peripheral blood mononuclear cells. Multiple cytogenetic loci were also evaluated with fluorescence in situ hybridization including for t(4;14), t(14;16), and del17p; International Staging System (ISS), progression-free survival (PFS) and overall survival (OS) data were collected for all patients. The MM Research Foundation (MMRF) CoMMpass (ClinicalTrials.gov identifier: [NCT01454297](#)) data set was used to validate findings (Appendix, online only).

### DNA Sequencing and Processing

We performed 150-bp paired-end sequencing on HiSeqX10 genome analyzers. The average sequence coverage was 75× for tumor samples and 35× for germline DNA (Appendix [Table A1](#), online only). Paired-end reads were aligned to the reference human genome, and Mutect2<sup>20</sup> was used to call single nucleotide variants (SNVs) and small insertions and deletions (indels). Copy number calls were analyzed using Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing,<sup>21</sup> and structural variants were analyzed using Manta.<sup>22</sup> SNVs and indels were annotated using Variant Effect Predictor<sup>23</sup> (Appendix). Mutational signatures were estimated by an R package, *signeR*<sup>24</sup>; contribution of each identified signature in subgroups and various clonal levels were then quantified using 1,000 bootstraps (Appendix).

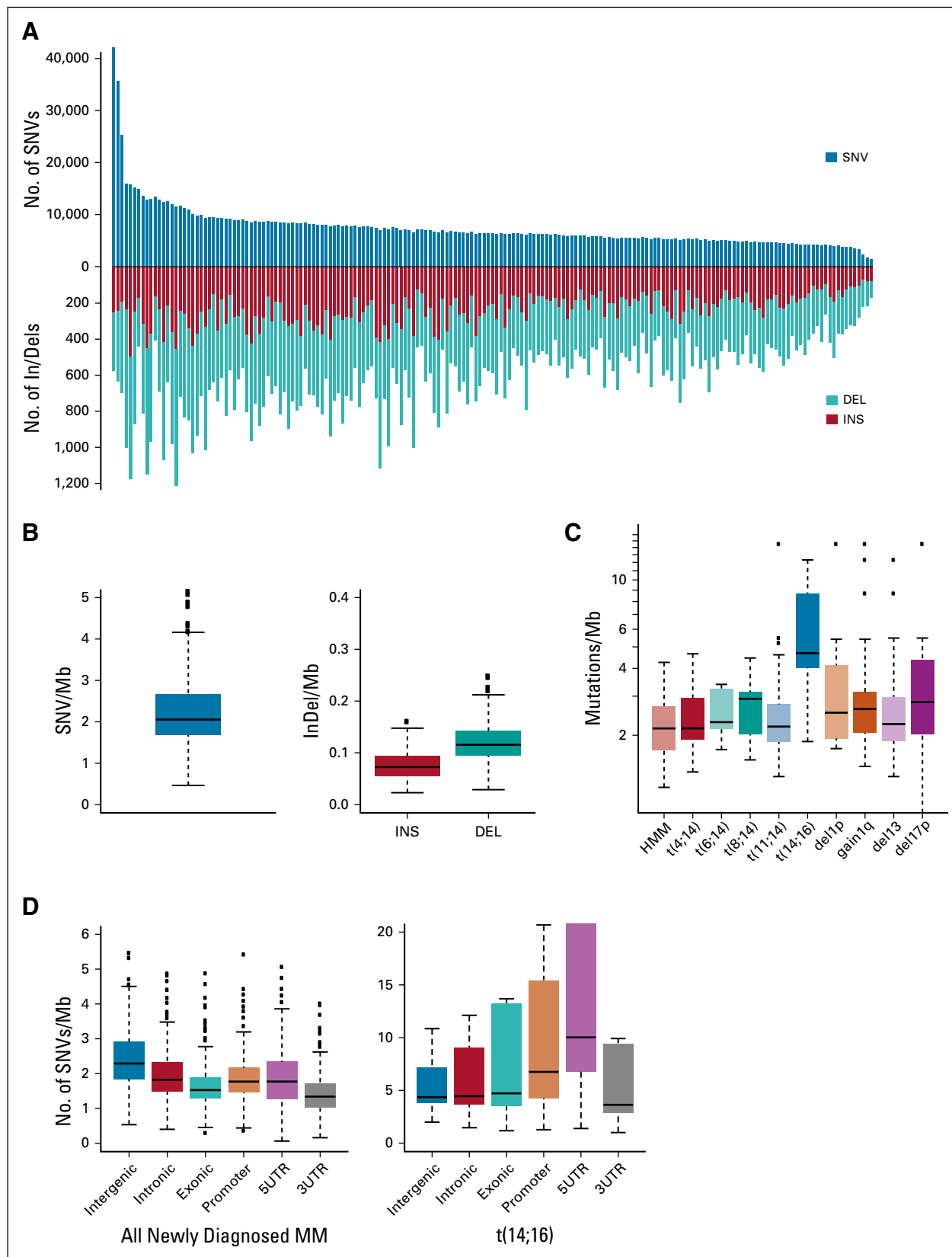
### Genomic Scar Score and Recursive Partitioning

Allele-specific copy-number alterations were used to calculate genomic scar score (GSS) with the scarHRD<sup>25</sup> package. Fifty-seven patients with total score  $\leq 5$  (minimum, 0; first quartile, 5; median, 9; mean, 10.53; third quartile, 16; maximum, 36) were considered as the low-GSS group. Associations between translocations, copy-number alterations, driver mutations, and GSS groups were calculated using Fisher's exact test. We used multivariate Cox regression to select variables and recursive partitioning to identify patient groups at different risk levels (Appendix). The Kaplan-Meier method was used to estimate time-to-event distributions, and statistical comparisons were done using log-rank tests (Appendix).

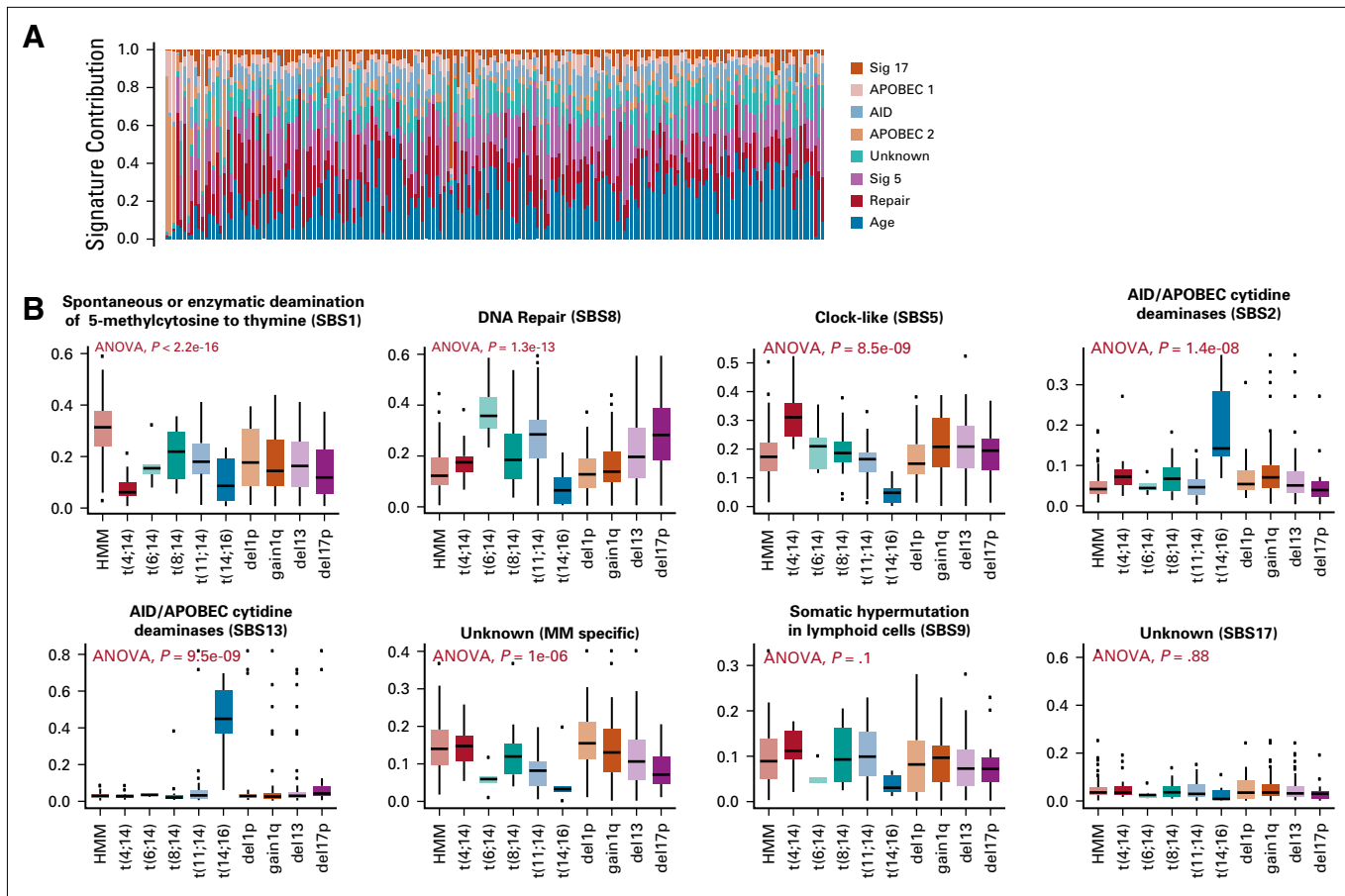
## RESULTS

### Mutational Burden in Newly Diagnosed MM

The WGS of CD138+ MM cells from 183 newly diagnosed patients identified an average of 7,343 SNVs (range, 1,431–42,108), 235 small insertions (range, 71–495), and 376 deletions per patient (range, 89–765; [Figs 1A and 1B](#)). Mutational load was significantly variable between MM subgroups. Hyperdiploid MM (HMM) has lowest and t(14;16) MM has the highest mutational load ([Fig 1C](#);  $P = .004$ ). The frequency of mutations across genomic domains was variable, with the intergenic region having the highest mutation load (median, 2.28 per Mb [range, 0.53–13.11]) and the exonic as well as 3' untranslated region (UTR) with the lowest SNV burden per Mb (median SNV per Mb, 1.52 and 1.33, respectively;  $P < 1e-10$ ; [Fig 1D](#)). The intergenic region also has more small insertions and deletions (median, 0.74 and 0.12 per Mb, respectively), whereas the 5'UTR has the lowest indel load ( $P < 1e-10$ ). Overall, the majority of the MM subgroups showed similar distribution of mutations across different genomic regions, except for the t(14;16) MM subset. This subgroup has significantly more SNVs per Mb in all regions, especially in 5'UTR and promoter regions (median, 6.75 and 10.02 SNVs per Mb,



**FIG 1.** Overall and genomic region mutational loads for patients with newly diagnosed multiple myeloma (MM). (A) Barplot shows the total number of somatic alterations (y-axis) per patients. Single-nucleotide variants (SNVs) are shown on top panel, insertions (INS) and deletions (DEL) are stacked at the bottom panel for each patient, and samples are ordered from highest number of SNVs to lowest. (B) Number of SNVs and InDels identified per megabase of the MM genome. (C) Number of mutations per megabase (y-axis) among MM subgroups (x-axis). (D) Number of mutations per megabase (y-axis) among genomic regions for all MM (left) and t(14;16) subgroups (right). HMM, hyperdiploid multiple myeloma; UTR, untranslated region.



**FIG 2.** Genome-wide mutational signatures in newly diagnosed multiple myeloma (MM). (A) Eight mutational signatures identified by nonnegative matrix factorization using all samples. Contributions (y-axis) of each signature (color coded) are shown for each patient (x-axis). Samples ordered from highest number of single-nucleotide variants to lowest. (B) Contribution (y-axis) of mutational signatures (each panel) in each MM subgroup (x-axis).  $P$  values for analysis of variance (ANOVA) test that compares the differences among group means are given on the top. (C) Contribution (y-axis) of each mutational signature (color-coded lines) from clonal mutations to subclonal mutations (x-axis). (D) Relative changes in signature contributions (y-axis) from clonal to subclonal mutations (x-axis) for hyperdiploid multiple myeloma (HMM; left) and del13, t(4;14), t(6;14), and t(8;14) (right) subgroups. (E) Relative changes in signature contributions (y-axis) from clonal to subclonal mutations (x-axis) for del17p, gain1q, del1p, and t(11;14) subgroups. (F) Two MM mutational signature activation models from early to late stage of the disease. Top panel represents the sequence of the processes for subgroups listed in D, and bottom panel represents the model for subgroups that are given in E. SBS, single-base substitution.

respectively;  $P < .003$ ), 3.8- and 5.6-fold greater ( $P = .001$ ), respectively, compared with other MM samples (Fig 1D).

#### Mutational Processes Among MM Subgroups and Association With Clonality

The underlying mutational processes operative in MM were characterized by identifying the mutational patterns of 96 possible trinucleotides: C>T was the most common (median, 30%; range, 19.9%-63.4%) and C>G the least common (median, 7.2%; range, 4.5%-29.5%) type of mutation among the 6 possible options (Appendix Fig A1, online only). We applied an empirical Bayesian approach of nonnegative matrix factorization and identified 8 mutational signatures across all MM samples (Fig 2A; Appendix Fig A2A, online only). These signatures represented variable contribution of 5 separate mutational processes (Appendix

Fig A2B): (1) an endogenous mutational process initiated by spontaneous or enzymatic deamination of 5-methylcytosine to thymine (single-base substitution 1 [SBS1; median contribution, 24.6%]) observed as age related; (2) activity of the AID/APOBEC family of cytidine deaminases (SBS2 [median contribution, 4.5%]) and SBS13 [median contribution, 2.9%]; (3) somatic hypermutation in lymphoid cells (SBS9 [median contribution, 8.8%]); (4) dysregulated homologous recombination-based DNA damage repair or nucleotide excision repair (NER; SBS8 [median contribution, 16%]); and (5) processes with unidentified etiology (SBS5, SBS17 [median contribution, 17.6% and 3.5% respectively]).

We found that myeloma subgroups are driven by specific mutational processes. Relative contribution of the age-related mutational process (SBS1) was high in HMM; the

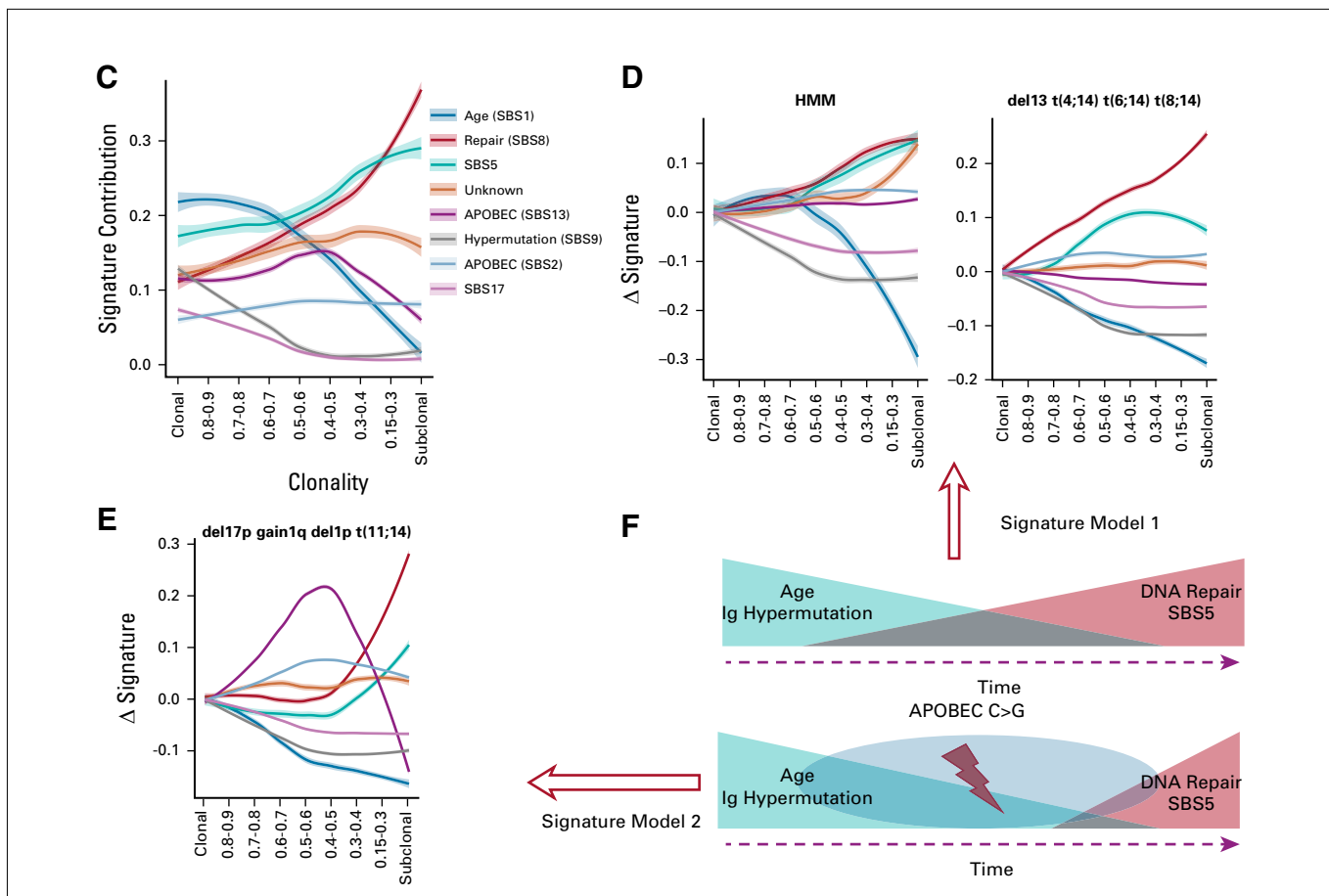


FIG 2. (Continued).

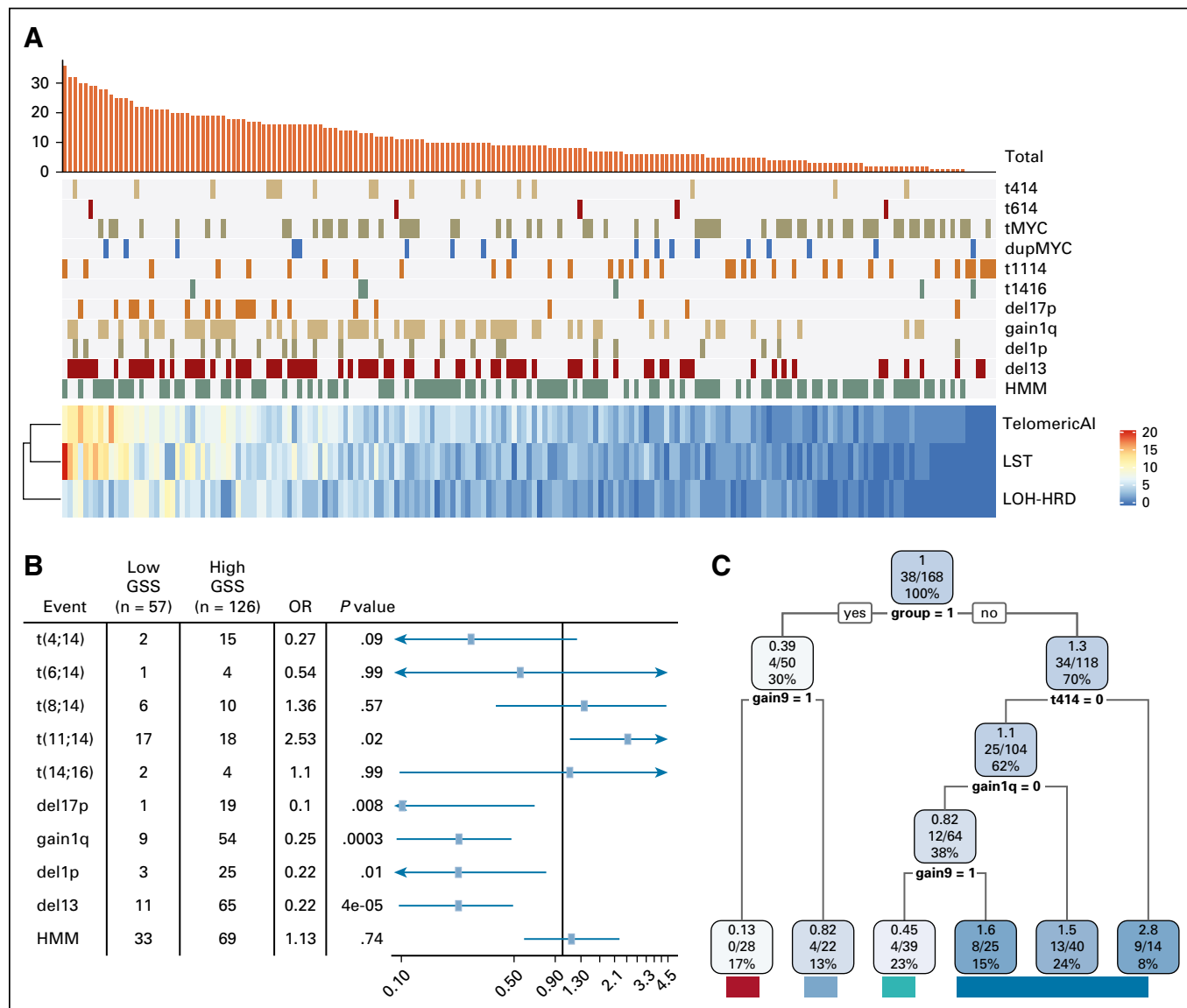
APOBEC-related mutational process was significantly higher in t(14;16) MM, and DNA repair was high in t(6;14), del17p, and t(11;14) MM (Fig 2B). SBS5, the clock-like signature with an unknown etiology, was significantly high in t(4;14) MM. Mutational processes associated with AID (SBS9) and unidentified etiology (SBS17) were constant across all MM subgroups (Fig 2B; Appendix Fig A3, online only).

To evaluate the time of activation of specific mutational processes in the development of MM, we analyzed the differences in signature utilization in clonal versus subclonal mutations. The age-related (SBS1) and SBS5 signatures (Fig 2C) were most frequently clonal, and hence affected early mutations. Although frequency of APOBEC signature associated with C>T mutations (SBS2) was constant at all clonality levels, APOBEC signature creating C>G mutations was significantly enriched in early to intermediate clones, especially in high-risk subgroups (Figs 2D and 2E; Appendix Figs A4-A6, online only). Importantly, mutations associated with homologous recombination (HR) and NER dysfunction were subclonal, and hence late mutations in virtually every subtype (Figs 2C-2E). Overall, MM subgroups showed 2 distinct patterns (Fig 2E). Both

patterns have high contribution of age-related mutations as an early event, with its decline by mid to late stage; however, in the high-risk group, the SBS13 (APOBEC) signature predominates in the middle phase of progression and then gets deactivated late in the disease process when DNA damage-associated mutations predominate. In contrast, DNA damage-associated mutations appear early and accumulate slowly over time in the standard-risk group (Fig 2F).

### GSS Is Associated With Superior Clinical Outcome in MM

On the basis of the observation that the DNA damage signature (SBS8) was most active in the late subclonal cell population in all MM subtypes, we evaluated the association between GSS (incorporating Loss of Heterozygosity–Homologous Recombination Deficiency [LOH-HRD], Large-Scale Transitions [LST], and Number of Telomeric Allelic Imbalances [AI]), mutational signatures, and clinical outcome. We used allele-specific copy-number profile to calculate each score (sum of LOH, LST, and Telomeric AI). The median GSS was 9 (range, 0-36; IQR, 11; Fig 3A). A low GSS ( $\leq 5$ , first quartile) was significantly higher in patients with t(11;14) compared with other subgroups (29.8% v 14.2%, respectively;



**FIG 3.** Low genomic scar score (GSS) predicts the superior outcome in multiple myeloma (MM). (A) Loss of Heterozygosity–Homologous Recombination Deficiency (LOH-HRD), Large-Scale Transitions (LST), and Number of Telomeric Allelic Imbalances (TelomericAI) scores are given in the bottom (rows) for each patient (columns). Red indicates higher scores. Known translocations and copy number alterations (CNAs) are given in the middle panel as present (indicated with different colors for different alterations) or absent. Total GSS is given on the top panel. (B) Forest plot shows the odds ratio (with 95 CI) of translocations and CNAs between low-GSS and high-GSS groups. Number of samples affected by each alteration in each group is shown in the second (low-GSS group) and third (high-GSS group) columns. *P* values from Fisher's exact test are given at the last column. (C) A decision tree from recursive partitioning analysis. For groups (low GSS with [w/] gain 9, low GSS without [w/o] gain 9, high GSS with t(4;14) or gain1q or no gain9 and the rest) frequencies in the IFM/DFCI cohort are given in each final node, with total number of samples and number of deaths at the end of the study. Relative risk for each terminal node is calculated compared with overall cohort. (D) Kaplan-Meier plot shows the overall survival probability for 4 groups defined by recursive partitioning in IFM/DFCI cohort. (E) Kaplan-Meier plot shows the overall survival probability for 4 groups in independent MM Research Foundation Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile validation cohort. (F) Kaplan-Meier plot shows the progression-free survival (PFS) for the patients in the low-GSS + gain9 group. Patients are divided by treatment arm. (G) Kaplan-Meier plots show the PFS for the patients in the high-GSS + t(4;14) or gain1q or no gain9 group. Patients are divided by treatment arm. ASCT, autologous stem cell transplant; HMM, hyperdiploid multiple myeloma; OR, odds ratio.

$P = .02$ ), whereas frequency of low GSS was significantly lower in del17p (1% v 15%;  $P$  value = .008), gain1q21 (15% v 42%;  $P$  value = .0003), del1p (5% v 19%;  $P$  value = .01), and del13 (19.3% v 51.5%;  $P$  value < .0001) subgroups (Fig 3B).

We next investigated the clinical impact of GSS. Patients in the low-GSS group had significantly longer median PFS (45.8 v 33.8 months; hazard ratio, 1.51; CI, 0.98 to 2.32;  $P = .055$ ) and OS (4-year OS, 96% v 77%; hazard ratio, 3.85; CI, 1.36 to 10.86;  $P = .006$ ) than other patients



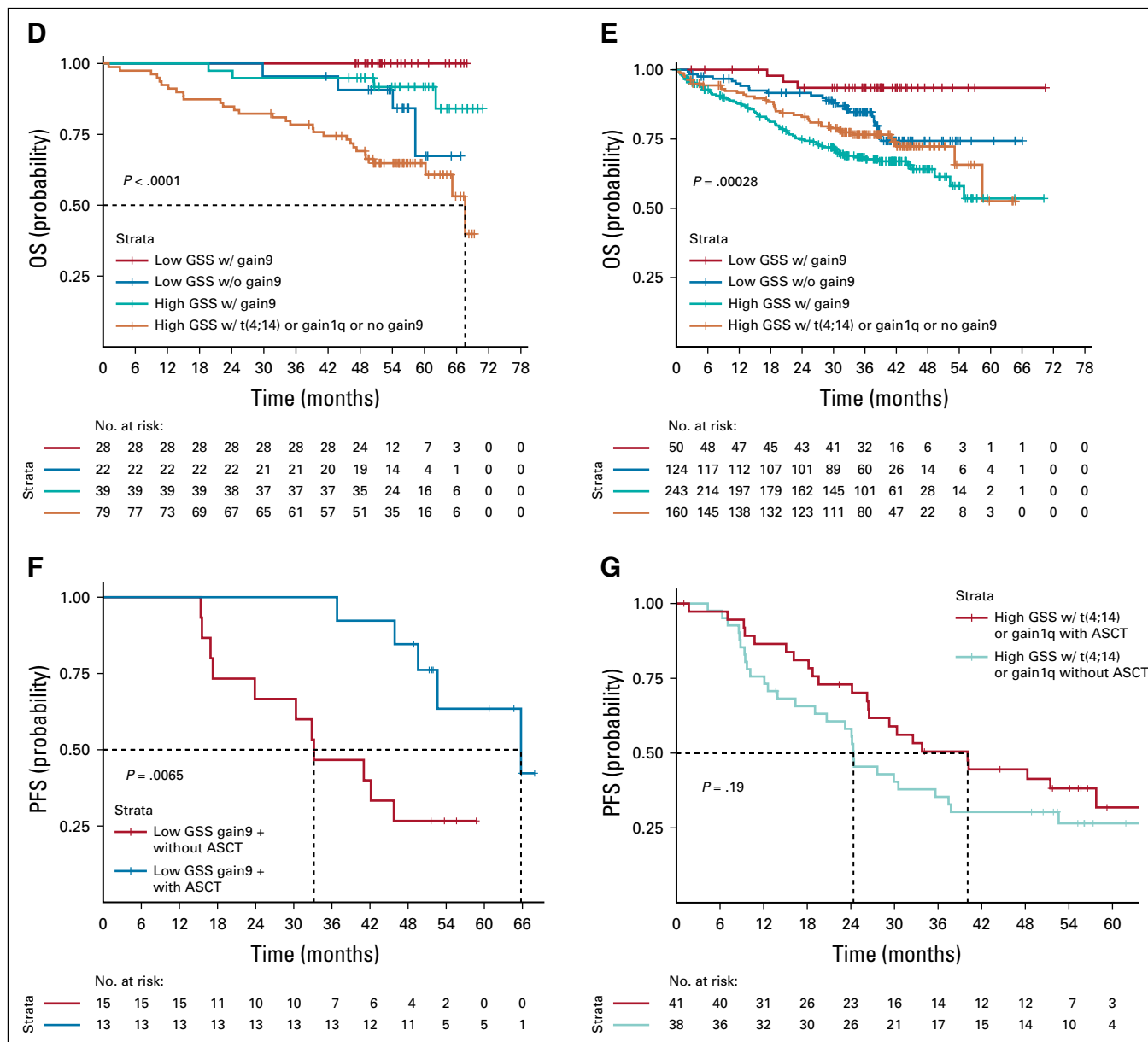


FIG 3. (Continued).

(Appendix Fig A7, online only). A recursive partitioning analysis incorporating other genomic risk features identified 6 branches (Fig 3C). Interestingly, patients with low GSS and chromosome 9 gain (gain9; 17% of all newly diagnosed cases) had a significantly superior outcome compared with other subgroups (OS probability was 100% in the IFM/DFCI 2009 data set; log-rank  $P$  value  $< .0001$ ; Fig 3D; Appendix Fig A8, online only). Moreover, patients with low GSS and no gain9 and patients with high GSS with gain9 (without t(4;14) and gain1q21) had intermediate outcome (OS probability 67% and 80%, respectively); and patients with high GSS and t(4;14) or gain1q21 or no gain9 had the worst OS (OS probability, 38%). We further validated the superior outcome in the low-

GSS group with gain9 using an independent data set of 577 patients (MMRF CoMMpass; 5-year OS probability, 93%; log-rank  $P$  value  $< .0001$ ; Fig 3E).

### Clinical and Genomic Differences Between Groups

In this study, patients were randomly assigned between lenalidomide, bortezomib, and dexamethasone (RVD) versus RVD plus high-dose therapy and autologous stem-cell transplantation. This gave us a unique opportunity to investigate the impact of therapy on outcome. We did not observe any difference in GSS on the basis of treatment arms. Moreover, distribution of patients between the 2 arms in each of the subgroups was similar (Table 1). In each of the subgroups, the OS was not significantly different on the



**TABLE 1.** Clinical Characteristics of Patients in Each of the 4 Subgroups Defined by Recursive Partitioning

Characteristic	Low GSS + gain9 (n = 28)	Log GSS + No gain9 (n = 22)	High GSS + gain9 (n = 39)	High GSS + t(4;14) or gain1q or no gain9 (n = 79)	P
<b>Clinical</b>					
Age, median (min-max)	58 (38-65)	60 (42-65)	59 (44-65)	60 (34-65)	.7
ISS1	12 (42.8)	6 (27.2)	14 (35.8)	27 (34.1)	.75
ISS2	13 (46.4)	9 (40.9)	15 (38.4)	35 (44.3)	.95
ISS3	3 (10.7)	6 (27.2)	9 (23)	17 (21.5)	.40
MRD (10e-6) + v –	8/7	10/4	17/8	21/19	.47
MRD (10e-5) + v –	5/10	7/7	16/9	18/22	.26
RVD v ASCT	14/13	15/7	18/21	41/38	.43
sCR or CR	18 (64.2)	12 (54.5)	14 (35.8)	43 (54.4)	.12
VGPR	6 (21.4)	6 (27.2)	18 (46.1)	22 (27.8)	.11
PR	4 (14.2)	4 (18.1)	6 (15.3)	12 (15.1)	.97
<b>Genomics</b>					
HMM	26 (92.4)	3 (13.6)	38 (97.4)	26 (32.9)	< 2.2e-16
t(4;14)	0	2 (9)	0	14 (17.7)	.001
t(6;14)	0	1 (4.5)	0	4 (5)	.39
t(11;14)	1 (3.5)	14 (63.6)	3 (7.6)	14 (17.7)	7.05e-07
t(14;16)	1 (3.5)	1 (4.5)	0	4 (5)	.56
t(MYC)	20 (71.4)	4 (18.1)	14 (35.8)	19 (24)	4.82e-05
dup(MYC)	3 (10.7)	2 (9)	7 (17.9)	4 (5)	.14
TP53+del17p	0	0	3 (7.6)	7 (8.8)	.28
TP53 Alone	0	0	0	1 (1.2)	.99
del17p Alone	0	1 (4.5)	1 (2.5)	8 (10.1)	.21
gain1q	6 (21.4)	3 (13.6)	0	49 (62)	3.66e-16
del1p	1 (3.5)	1 (4.5)	5 (12.8)	16 (20.2)	.09
del13	3 (10.7)	8 (36.3)	7 (17.9)	54 (68.3)	6.05e-10

NOTE. Data are presented as No. (%) unless otherwise noted.

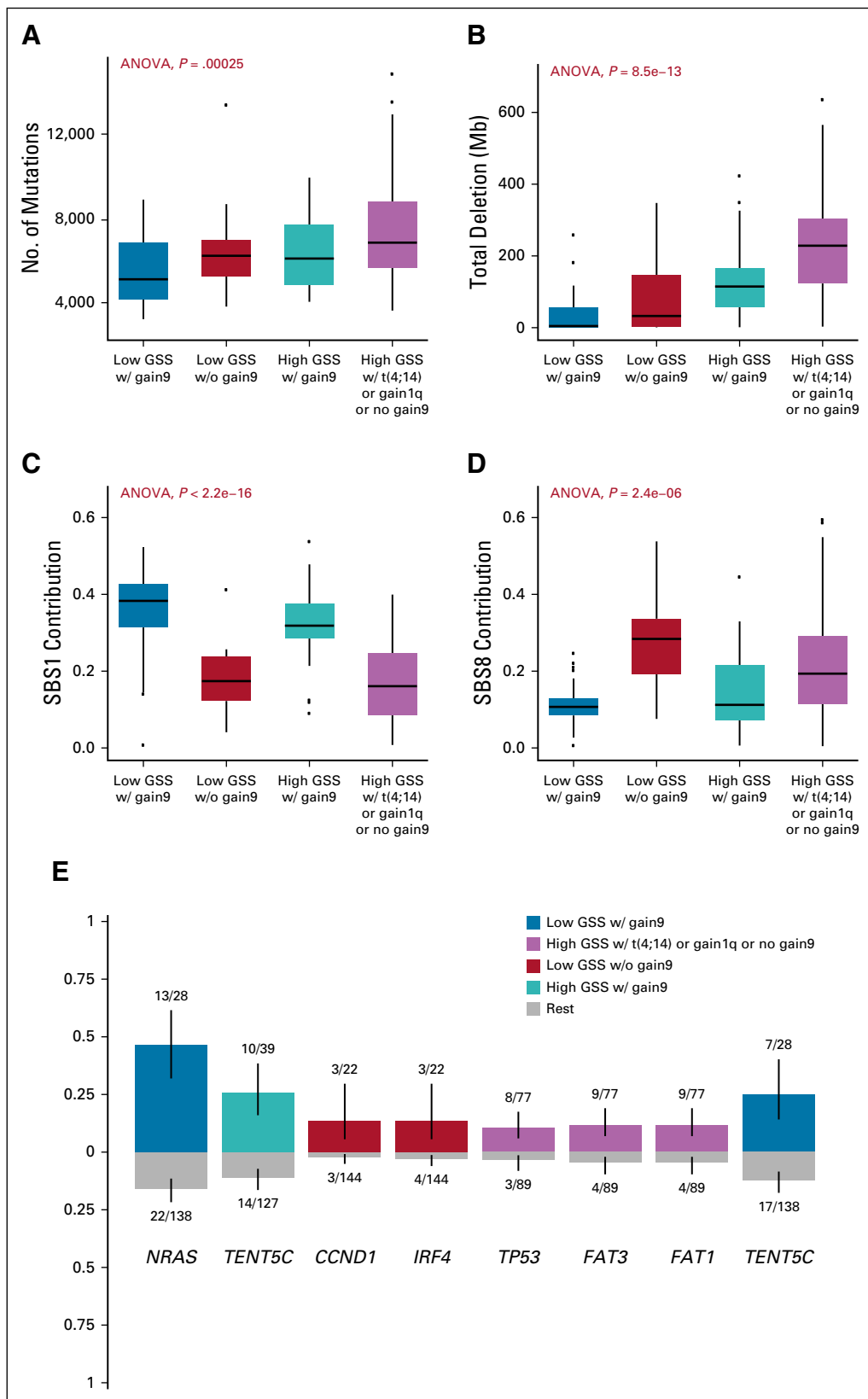
Abbreviations: ASCT, autologous stem cell transplant; CR, complete response; GSS, genomic scar score; HMM, hyperdiploid multiple myeloma; ISS, International Staging System; MRD, minimal residual disease; PR, partial response; RVD, lenalidomide, bortezomib, and dexamethasone; sCR, stringent complete response; VGPR, very good partial response.

basis of the arm to which patients were randomly assigned. Except for the good-risk group, PFS was also not significantly different between the 2 randomized arms. In the good-risk group, we observed a significantly superior PFS in favor of those undergoing transplant ( $P < .006$ ; Appendix Figs A3F, A3G, and A9, online only). Importantly, there was no statistically significant difference in ISS stage, treatment arms, clinical response, or achievement of minimal residual disease (MRD) negativity in each of the subgroups (Table 1).

To define the effect of genomic alterations on outcome in these groups, we compared their genomic characteristics. Hyperdiploidy is a frequent feature for low GSS with gain9 (92.4%) and high GSS with gain9 but not gain1q or t(4;14) (97.4%;  $P$  value < 2.2e-16). Although hyperdiploidy is associated with better prognosis, our further analysis of

patients with HMM showed that low GSS can separate patients with HMM into low- and high-risk HMM subgroups, with a substantial number of patients (69% in our data set and 81% in the MMRF data set) having inferior survival in our data set (log-rank  $P$  value = .0005) and in the MMRF data set (log-rank  $P$  value = .001; Appendix Fig A8). *MYC* translocations with any partner were also common in the low-risk group (71.4%;  $P$  value = 4.82e-05). The low-GSS with gain9 group also had significantly lower frequency of del13 and gain1q21 (Table 1).

Patients with low GSS and gain9 had significantly lower mutational load compared with other groups (Fig 4A;  $P$  = .0002). The median aggregate loss of genomic material in the low-GSS and gain9 group was only 5.5 Mb; however, this was 42-fold higher (231 Mb) in the high-GSS with t(4;14) or gain1q21 or no gain9 subgroup (Fig 4B). The good-risk



**FIG 4.** Genomic differences among 4 groups defined by recursive partitioning. (A) Number of total mutations per patient in each group; (B) total genomic loss (deletion) in each subgroup; (C) contribution of single-base substitution 1 (SBS1; age related), and (D) SBS8 (DNA repair) mutational signatures; and (E) driver mutation enrichment analysis. ANOVA, analysis of variance; GSS, genomic scar score; w/, with; w/o, without.

group also had higher SBS1 (spontaneous deamination;  $P$  value =  $2.2 \times 10^{-16}$ ) and very low SBS8 (HR/NER signature;  $P$  value =  $2.4 \times 10^{-6}$ ) mutational signature usage (Figs 4C and 4D). The previously defined driver mutation frequency was also significantly different among the 4 groups (Fig 4E). Importantly, low GSS with gain9 had high NRAS mutation frequency compared with other groups (46% v 16%), and patients with high GSS and t(4;14) or gain1q21 or no gain9 showed frequent mutations in *TP53* (10%), as well as *FAT1* and *FAT3*.

## DISCUSSION

This large, deep WGS analysis in newly diagnosed MM incorporating clinical correlates provides a global picture of active somatic processes in MM and identifies a genomic segmentation for a subgroup of patients with long survival. The deep WGS provides a comprehensive catalog of genetic lesions in MM and an improved delineation of the subclonal structure. Although the early and thus clonal lesions are important to drive the initial malignant transformation, it is the later-acquired and thus subclonal lesions that may drive the ultimate disease behavior and patient outcome. Identification of a median of > 7,000 SNVs per patient allows for study of both early and late mutational processes operative in MM. Overall, the mutational load in newly diagnosed MM is intermediate compared with all other cancer genomes in International Cancer Genome Consortium or The Cancer Genome Atlas studies.<sup>26,27</sup> Among the genomic regions, more protected regions like exons<sup>28</sup> showed significantly fewer mutations compared with intergenic regions. However, the lowest mutational load observed in the 3'UTR region suggests that the mechanisms protecting active translated regions can also remain active to protect untranslated regions.

The large sample size and a representative pool of mutations that covers every region has allowed identification of active mutational processes.<sup>29</sup> Previous studies with clinical annotations mainly interrogated exons and found predominantly clock-like and APOBEC signatures.<sup>6,30-32</sup> Here we were able to extend our understanding of the mutational processes in MM subgroups and also evolution from clonal to subclonal mutations. Overall, contribution from signatures was significantly different among the MM subgroups. We report AID-signature (SBS9)-related mutations predominantly contributing to clonal mutations, while not contributing at the subclonal stages of the disease, in all MM subtypes. Although translocations and copy number alterations are considered as the initiating events, malignant transformation requires additional genomic changes, including *MYC* dysregulation and driver SNVs.<sup>1,18,33,34</sup> AID-caused mutations may have role in transformation, and AID may be responsible for some common driver mutations, as indicated before.<sup>35</sup> These results also validate our earlier study identifying the AID-driven process (SBS9) as an early event that is active at the smoldering stage.<sup>6</sup> Both the earlier

study investigating SMM-to-MM progression and the current report suggest that APOBEC and other signatures including SBS8 shape the landscape of the later stages of progressive MM. These results also confirm early reports of increased APOBEC activity with the development of MM.<sup>6,7,30,31</sup> Importantly, deeper WGS provides the required granularity to identify DNA repair signatures, HR and NER, as the dominant processes at the subclonal level in every subtype. The ongoing genomic instability driven by dysfunctional repair processes drives later disease progression and outcome.

To date, risk stratification analyses in MM have used copy number-based classification, gene expression profile, and integrated genomic analysis including exome sequencing with DNA mutations and indels<sup>1,2,17</sup> and have focused on defining a high-risk group with poor survival outcome. These high-risk disease subgroups are identified as double-hit,<sup>36</sup> revised ISS,<sup>37</sup> or translocations.<sup>38-40</sup> Moreover, del17p, del1p, gain1q21,<sup>40-43</sup> multiple driver gene mutations,<sup>7,44</sup> or changes in gene expression patterns identified with 70-gene or 92-gene signatures<sup>45,46</sup> also identify high-risk disease. However, none of these studies have further investigated patients with superior outcome. Our study, to our knowledge for the first time, focuses on patients with long survival to support future research planning, clinical studies, and patient care.

Interestingly, the superior outcome group identified here was independent of traditional clinical risk factors, such as ISS, response to treatment, and achievement of MRD negativity. Importantly, our study shows that, in addition to traditional risk markers, we can use genomic markers such as low GSS and gain9 to identify true low-risk groups. Traditionally, patients with HMM are considered to have superior outcome.<sup>42,47</sup> In our good-risk group, > 90% of the patients showed hyperdiploidy; however, only 31% of patients with HMM (19% in CoMMpass data) fall within our good-risk group. Thus, low GSS clearly distinguishes HMM into low- and high-risk subgroups. In our study, patients were uniformly treated as part of the IFM/DFCI study and randomly assigned to RVD alone versus RVD + ASCT. In the overall IFM/DFCI study, patients receiving RVD + ASCT had superior PFS,<sup>19</sup> and we here evaluated the impact of treatment arms on the outcome in our patient subgroups. Importantly, the OS in each of the subgroups was not significantly different on the basis of the arm to which the patients were assigned. PFS was also not significantly different between the 2 arms except for the good-risk group. In these patients, we observed a significantly superior PFS in favor of those undergoing transplant ( $P < .006$ ), which needs to be further validated in ongoing and future studies.

In clinical practice, the sequencing platforms (WGS and whole exome sequencing) or single nucleotide polymorphism array can be used to identify those patients with longer survival expectations as reported here. The current

targeted sequencing panels may not be adequate to identify GSS, because the coverage of the genome may not be sufficient. Identifying patients with a long survival provides important prognostic information and can help reassure the patients about the long-term outcome. Although therapeutic intervention should not yet be changed based on this information, use of high-dose therapy may be an important consideration in these patients. Clinical trials in this patient subgroup can further help tailor induction/consolidation treatment to improve or maintain the long survival while decreasing toxicity. For example, the utility of high-dose chemotherapy and duration of maintenance therapy can be evaluated in this patient subgroup. Our data further suggest the need to stratify patients with long-term survival in large phase III studies to balance the arms and

assure accurate assessment of novel therapeutics. In addition, clinical trials in this patient subgroup should not evaluate OS or even PFS as realistic end points but instead include MRD. Finally, delineation of this new subgroup can assure patients and caregivers alike of good prognosis and prolonged favorable outcome.

In conclusion, this is the first comprehensive genomic study, to our knowledge, that identifies a good-risk group in MM. This favorable subgroup comprises 17% of patients with NDMM with lower mutation burden, fewer deletion events, and driven by age-related signature, suggesting lower level of genomic instability. Identifying these good-risk patients will affect clinical research and inform therapeutic algorithms in the future.

## AFFILIATIONS

<sup>1</sup>Department of Data Sciences, Dana Farber Cancer Institute, Boston, MA

<sup>2</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

<sup>3</sup>Department of Medical Oncology, Dana Farber Cancer Institute, Harvard Medical School, Boston, MA

<sup>4</sup>Inserm UMR892, CNRS 6299, Université de Nantes, and Centre Hospitalier Universitaire de Nantes, Unité Mixte de Genomique du Cancer, Nantes, France

<sup>5</sup>University Cancer Center of Toulouse Institut National de la Santé, Toulouse, France

<sup>6</sup>Celgene Corporation, Summit, NJ

<sup>7</sup>VA Boston Healthcare System, Boston, MA

## CORRESPONDING AUTHOR

Mehmet Kemal Samur, PhD, Dana-Farber Cancer Institute, 450 Brookline Ave, Boston, MA 02215; e-mail: mehmet\_samur@dfci.harvard.edu.

## SUPPORT

Supported by National Institutes of Health Grants No. P01 CA155258 and P50 CA100707, Celgene Corporation, and VA Healthcare System Grant No. 5I01BX001584.

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT

Disclosures provided by the authors and data availability statement (if applicable) are available with this article at DOI <https://doi.org/10.1200/JCO.20.00461>.

## AUTHOR CONTRIBUTIONS

**Conception and design:** Mehmet Kemal Samur, Kenneth C. Anderson, Hervé Avet-Loiseau, Nikhil C. Munshi

**Financial support:** Kenneth C. Anderson, Nikhil C. Munshi

**Administrative support:** Nikhil C. Munshi

**Provision of study material or patients:** Jill Corre, Hervé Avet-Loiseau, Nikhil C. Munshi

**Collection and assembly of data:** Mehmet Kemal Samur, Anil Aktas Samur, Mariateresa Fulciniti, Raphael Szalat, Florence Magrangeas, Stephane Minvielle, Kenneth C. Anderson, Hervé Avet-Loiseau, Nikhil C. Munshi

**Data analysis and interpretation:** Mehmet Kemal Samur, Anil Aktas Samur, Raphael Szalat, Tessa Han, Paul Richardson, Jill Corre, Philippe Moreau, Anjan Thakurta, Kenneth C. Anderson, Giovanni Parmigiani, Hervé Avet-Loiseau, Nikhil C. Munshi

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

**Accountable for all aspects of the work:** All authors

## REFERENCES

- Manier S, Salem KZ, Park J, et al: Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* 14:100-113, 2017
- Perrot A, Corre J, Avet-Loiseau H: Risk stratification and targets in multiple myeloma: From genomics to the bedside. *Am Soc Clin Oncol Educ Book* 38: 675-680, 2018
- Robiou du Pont S, Cleyne A, Fontan C, et al: Genomics of multiple myeloma. *J Clin Oncol* 35:963-967, 2017
- Bolli N, Avet-Loiseau H, Wedge DC, et al: Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat Commun* 5:2997, 2014
- Bolli N, Biancon G, Moarii M, et al: Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia* 32:2604-2616, 2018
- Bolli N, Maura F, Minvielle S, et al: Genomic patterns of progression in smoldering multiple myeloma. *Nat Commun* 9:3363, 2018
- Walker BA, Mavrommatis K, Wardell CP, et al: Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood* 132:587-597, 2018 [Erratum: *Blood* 132:1461, 2018]
- Chapman MA, Lawrence MS, Keats JJ, et al: Initial genome sequencing and analysis of multiple myeloma. *Nature* 471:467-472, 2011
- Khurana E, Fu Y, Chakravarty D, et al: Role of non-coding sequence variants in cancer. *Nat Rev Genet* 17:93-108, 2016
- Zhang W, Bojorquez-Gomez A, Velez DO, et al: A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. *Nat Genet* 50:613-620, 2018
- Gloss BS, Dinger ME: Realizing the significance of noncoding functionality in clinical genomics. *Exp Mol Med* 50:97, 2018
- Piraino SW, Furney SJ: Beyond the exome: the role of non-coding somatic mutations in cancer. *Ann Oncol* 27:240-248, 2016
- Rahman S, Mansour MR: The role of noncoding mutations in blood cancers. *Dis Model Mech* 12:dmm041988, 2019

14. Zhu H, Uusküla-Reimand L, Isaev K, et al: Candidate cancer driver mutations in distal regulatory elements and long-range chromatin interaction networks. *Mol Cell* 77:1307-1321.e10, 2020
15. Hornshøj H, Nielsen MM, Sinnott-Armstrong NA, et al: Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival. *NPJ Genom Med* 3:1, 2018
16. Corre J, Cleyne A, Robiou du Pont S, et al: Multiple myeloma clonal evolution in homogeneously treated patients. *Leukemia* 32:2636-2647, 2018
17. Szalat R, Munshi NC: Genomic heterogeneity in multiple myeloma. *Curr Opin Genet Dev* 30:56-65, 2015 [Erratum: *Curr Opin Genet Dev* 37:158, 2016]
18. Aktas Samur A, Minvielle S, Shammas M, et al: Deciphering the chronology of copy number alterations in multiple myeloma. *Blood Cancer J* 9:39, 2019
19. Attal M, Lauwers-Cances V, Hulin C, et al: Lenalidomide, bortezomib, and dexamethasone with transplantation for myeloma. *N Engl J Med* 376:1311-1320, 2017
20. Cibulskis K, Lawrence MS, Carter SL, et al: Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31:213-219, 2013
21. Shen R, Seshan VE: FACETS: Allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 44:e131, 2016
22. Chen X, Schulz-Trieglaff O, Shaw R, et al: Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32:1220-1222, 2016
23. McLaren W, Gil L, Hunt SE, et al: The Ensembl Variant Effect Predictor. *Genome Biol* 17:122, 2016
24. Rosales RA, Drummond RD, Valieris R, et al: signer: An empirical Bayesian approach to mutational signature discovery. *Bioinformatics* 33:8-16, 2017
25. Sztupinski Z, Diossy M, Krzystanek M, et al: Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. *NPJ Breast Cancer* 4:16, 2018
26. Hoadley KA, Yau C, Hinoue T, et al: Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. *Cell* 173:291-304.e6, 2018
27. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium: Pan-cancer analysis of whole genomes. *Nature* 578:82-93, 2020
28. Gonzalez-Perez A, Sabarinathan R, Lopez-Bigas N: Local determinants of the mutational landscape of the human genome. *Cell* 177:101-114, 2019
29. Alexandrov LB, Kim J, Haradhvala NJ, et al: The repertoire of mutational signatures in human cancer. *Nature* 578:94-101, 2020
30. Maura F, Petljak M, Lionetti M, et al: Biological and prognostic impact of APOBEC-induced mutations in the spectrum of plasma cell dyscrasias and multiple myeloma cell lines. *Leukemia* 32:1044-1048, 2018
31. Walker BA, Wardell CP, Murison A, et al: APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat Commun* 6:6997, 2015
32. Hoang PH, Cornish AJ, Dobbins SE, et al: Mutational processes contributing to the development of multiple myeloma. *Blood Cancer J* 9:60, 2019
33. Keats JJ, Chesi M, Egan JB, et al: Clonal competition with alternating dominance in multiple myeloma. *Blood* 120:1067-1076, 2012
34. Affer M, Chesi M, Chen WG, et al: Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia* 28:1725-1735, 2014
35. Maura F, Rustad EH, Yellapantula V, et al: Role of AID in the temporal pattern of acquisition of driver mutations in multiple myeloma. *Leukemia* 34:1476-1480, 2020
36. Walker BA, Mavrommatis K, Wardell CP, et al: A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia* 33:159-170, 2019
37. Palumbo A, Avet-Loiseau H, Oliva S, et al: Revised International Staging System for multiple myeloma: A report from International Myeloma Working Group. *J Clin Oncol* 33:2863-2869, 2015
38. Barwick BG, Neri P, Bahlis NJ, et al: Multiple myeloma immunoglobulin lambda translocations portend poor prognosis. *Nat Commun* 10:1911, 2019
39. Sonneveld P, Avet-Loiseau H, Lonial S, et al: Treatment of multiple myeloma with high-risk cytogenetics: A consensus of the International Myeloma Working Group. *Blood* 127:2955-2962, 2016
40. Mikhael J, Ismaila N, Cheung MC, et al: Treatment of multiple myeloma: ASCO and CCO joint clinical practice guideline. *J Clin Oncol* 37:1228-1263, 2019
41. Perrot A, Lauwers-Cances V, Tournay E, et al: Development and validation of a cytogenetic prognostic index predicting survival in multiple myeloma. *J Clin Oncol* 37:1657-1665, 2019
42. Avet-Loiseau H, Li C, Magrangeas F, et al: Prognostic significance of copy-number alterations in multiple myeloma. *J Clin Oncol* 27:4585-4590, 2009
43. Weinhold N, Kirn D, Seckinger A, et al: Concomitant gain of 1q21 and MYC translocation define a poor prognostic subgroup of hyperdiploid multiple myeloma. *Haematologica* 101:e116-e119, 2016
44. Miller A, Asmann Y, Cattaneo L, et al: High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood Cancer J* 7:e612, 2017
45. Samur MK, Minvielle S, Gulla A, et al: Long intergenic non-coding RNAs have an independent impact on survival in multiple myeloma. *Leukemia* 32:2626-2635, 2018
46. Kuiper R, Broyl A, de Knecht Y, et al: A gene expression signature for high-risk multiple myeloma. *Leukemia* 26:2406-2413, 2012 [Erratum: *Leukemia* 28:1178-1180, 2014]
47. Kumar S, Fonseca R, Ketterling RP, et al: Trisomies in multiple myeloma: Impact on survival in patients with high-risk cytogenetics. *Blood* 119:2100-2105, 2012 [Erratum: *Blood* 123:1621, 2014]



**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST****Genome-Wide Somatic Alterations in Multiple Myeloma Reveal a Superior Outcome Group**

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [ascopubs.org/jco/authors/author-center](http://ascopubs.org/jco/authors/author-center).

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians ([Open Payments](#)).

**Paul Richardson**

**Consulting or Advisory Role:** Celgene, Janssen, Takeda, Karyopharm Therapeutics, Oncopeptides, Sanofi, Jazz Pharmaceuticals, Secura Bio

**Research Funding:** Celgene (Inst), Takeda (Inst), Bristol Myers Squibb (Inst), Oncopeptides (Inst)

**Jill Corre**

**Honoraria:** Amgen, Celgene, Janssen, Sanofi

**Consulting or Advisory Role:** Celgene

**Travel, Accommodations, Expenses:** Sanofi

**Philippe Moreau**

**Honoraria:** Celgene, Takeda, Novartis, Janssen-Cilag, Amgen, GSK

**Consulting or Advisory Role:** Celgene, Takeda, Janssen, Amgen, GSK

**Anjan Thakurta**

**Employment:** Celgene/Bristol Myers Squibb

**Stock and Other Ownership Interests:** Celgene/BMS

**Patents, Royalties, Other Intellectual Property:** Patents related to use of Celgene/BMS drugs in cancer and identification of targets and patient segments

**Kenneth C. Anderson**

**Stock and Other Ownership Interests:** C4 Therapeutics, OncoPep

**Consulting or Advisory Role:** Celgene, Millennium, Gilead Sciences, Bristol Myers Squibb, Janssen Oncology, Sanofi, Tolero Pharmaceuticals, Precision Biosciences

**Patents, Royalties, Other Intellectual Property:** C4 Therapeutics, OncoPep

**Giovanni Parmigiani**

**Leadership:** Phaeno Biotech

**Stock and Other Ownership Interests:** CRA Health

**Consulting or Advisory Role:** Biogen, Konica Minolta

**Patents, Royalties, Other Intellectual Property:** Patent: Genetic alterations in malignant gliomas; Copyright: BayesMendel software (Inst)

**Expert Testimony:** Natera

**Travel, Accommodations, Expenses:** Konica Minolta

**Nikhil C. Munshi**

**Stock and Other Ownership Interests:** OncoPep

**Consulting or Advisory Role:** Celgene, Takeda, Janssen, OncoPep, AbbVie, Adaptive Biotechnologies, Amgen, BeiGene, Karyopharm Therapeutics, Bristol-Myers Squibb

**Patents, Royalties, Other Intellectual Property:** OncoPep

No other potential conflicts of interest were reported.



## APPENDIX

## Patient Samples

All samples from patients with multiple myeloma (MM) were collected from the IFM/DFCI 2009 clinical trial, a phase III, multicenter, randomized, open-label study designed to evaluate the clinical benefit from the drug combination lenalidomide, bortezomib, and dexamethasone (RVD) without immediate high-dose therapy (HDT) followed by lenalidomide maintenance (arm A) versus RVD plus HDT and peripheral blood stem cell transplant followed by lenalidomide maintenance (arm B), after written informed consent was obtained and samples and genomic and de-identified clinical data managed in accordance with the Declaration of Helsinki. After bone marrow collection from 183 patients, CD138+ selection was performed to purified myeloma cells from bone marrow. DNA was extracted from CD138+ purified bone marrow cells. All patient samples with symptomatic and progressive multiple myeloma on the basis of International Multiple Myeloma Working Group criteria were collected at diagnosis. Constitutional control DNA originated from peripheral blood mononuclear cells. Purity of the CD138+ fraction was assessed by anti-CD138 immunocytochemistry post sorting, and only samples that passed quality control were sequenced. At diagnosis, the median age for patients was 58 years (range, 34-65 years), and 58.5% and 41.5% were male and female, respectively. International Staging System (ISS) stage distributions from stage 1 to 3 were 32%, 48.5%, and 19.5%. Multiple cytogenetic loci were also evaluated with fluorescence in situ hybridization (FISH), including t(4;14), t(14;16), and del17p, and ISS, PFS, and OS were collected for patients.

## DNA Sequencing and Processing

Short insert genomic libraries were constructed, flowcells prepared, and sequencing clusters generated according to Illumina protocols. We performed 150-bp paired-end sequencing on HiSeqX10 genome analyzers. The average sequence coverage was 75× for tumor samples and 35× for germline DNA. Paired-end reads were aligned to the reference human genome GRCh38 using BWA-mem. Duplicated reads were marked and base quality score recalibration was performed using MarkDuplicates and ApplyBQSR with GATK4.

Mutect2 was used to call single-nucleotide variants (SNVs) and small insertions and deletions (indels). All germline samples were first analyzed with Mutect2, and results were combined with CreateSomaticPanelOfNormals to create Panel of Normals (PoN). PoN was used to filter sites and the population variant resource containing allele-specific frequencies from gnomAD and matched germline samples to filter alleles. Raw mutation calls were then filtered using FilterMutectCalls from GATK4, and only the mutations PASS all the filters applied by FilterMutectCalls and had at least 10× coverage for both tumor and normal samples extracted with bcftools for further analysis. Homozygous/heterozygous deletions, copy-neutral loss-of-heterozygosity (LOH), allele-specific gains/amplification, as well as ploidy and purity of each sample were analyzed using FACETS (Fraction and Allele-Specific Copy Number Estimates from Tumor Sequencing). Structural variants were analyzed using Manta. SNVs and indels were annotated using Variant Effect Predictor from Ensembl.

## FISH Analysis

Sorted plasma cells were fixed in Carnoy's fixative and stored at -20°C until hybridization. After slide preparation, they were denatured in 70% formamide for 5 minutes and dehydrated in 70%, 85%, and 100% ethanol series. The probes specific for t(4;14), del17p, and t(14;16) were purchased from Abbott Molecular (Abbott Park, Illinois) and denatured separately for 5 minutes at 75°C. After denaturation, the probes were dropped on the plasma cells and hybridized overnight at 37°C. Then, coverslips were removed and the slides were washed for 2 minutes in 2×SSC 0.1% Triton at 75°C.

## SNV Signatures

Mutational signatures were estimated by an R package signer, which uses an empirical Bayesian treatment of the nonnegative matrix factorization model. Single-base substitutions were mapped onto

trinucleotide sequences by including the 5' and 3' neighboring base context to construct a 96 × 183 matrix of mutation counts. The optimal number of mutational signatures was estimated by considering a saddlepoint approximation to the Bayesian information criterion, and Cosine similarity was used to determine the Cosmic v3 single-base substitution signatures that were closest to the detected novel signatures. Contribution of each identified signature in subgroups and various clonal levels was then quantified using 1,000 bootstraps with a multiple linear regression model with the caveat that any coefficient must be > 0. We have modified the approach explained in deconstructSig to be able to measure the estimation errors for each signature.

## Genomic Scar Score

Allele-specific copy number alterations were used to calculate genomic scar score (GSS) with scarHRD R package. Total score is calculated as the sum of Homologous Recombination Deficiency–Loss of Heterozygosity (HRD-LOH: number of 15-Mb–exceeding LOH regions which do not cover the whole chromosome), Large-Scale Transitions (LST: chromosomal break between adjacent regions of at least 10 Mb, with a distance between them ≤ 3 Mb), and Number of Telomeric Allelic Imbalances (telomeAI: number AIs that extend to the telomeric end of a chromosome). Fifty-seven patients, those having a total score ≤ 5 (minimum, 0; first quartile, 5; median, 9; mean, 10.53; third quartile, 16; maximum, 36) were considered as the low-GSS group. Associations between translocations, copy number alterations, driver mutations, and GSS groups were calculated using Fisher's exact test.

## Recursive Partitioning

We used recursive partitioning analysis to identify patient groups at different risk levels. First, we used multivariate Cox model in R using "survival" package for variable selection. GSS, IgH translocations, deletion 13, 1p, 17p, gain1q, and gains in hyperdiploid chromosomes as well as mutations in 10 driver mutations detected by "dndscv" were used for variable selection. Variables reached to significance level of 0.1 were then used for recursive partitioning with "rpart" package in R. We evaluated the stability of the tree construction using "ipred" package. Brier score using out-of-bag estimator with 1,000 bootstrapping.

## Validation Data Set

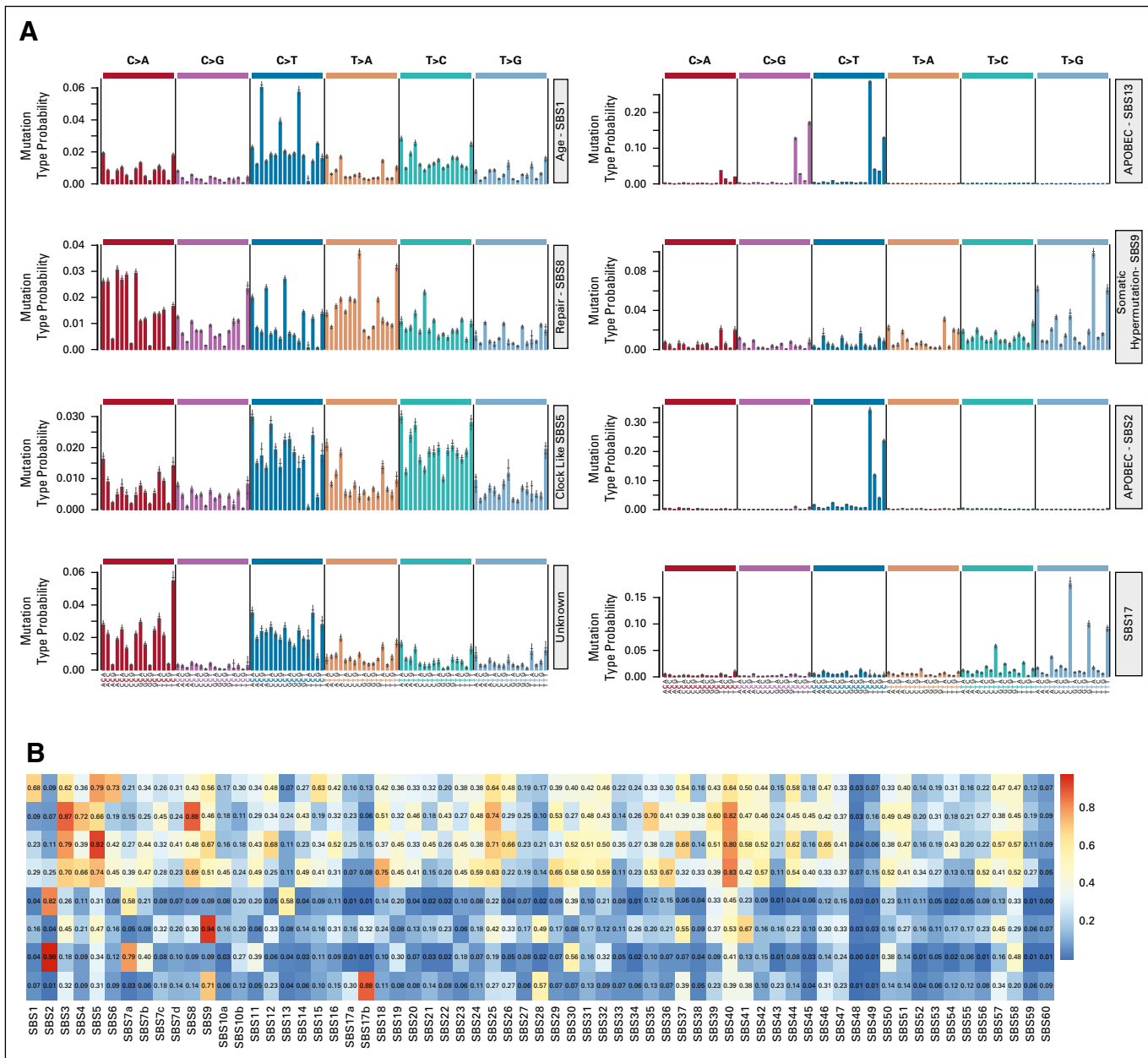
The Multiple Myeloma Research Foundation (MMRF) CoMMpass (Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile) data set was used to validate findings. Genomic data for patients with newly diagnosed MM in the CoMMpass cohort were downloaded from dbGaP (Accession Number: phs000748.v7.p4). GSS for each newly diagnosed patient was calculated using allele-specific copy number profiles generated from whole-exome sequencing data. Other clinical and genomic variables were integrated with GSS using "MMRF\_CoMMpass\_IA13\_STAND\_ALONE\_SURVIVAL.csv," "MMRF\_CoMMpass\_IA12a\_LongInsert\_Canonical\_Ig\_Translocations.txt," and "MMRF\_CoMMpass\_IA12a\_CNA\_LongInsert\_FISH\_CN\_All\_Specimens.txt" files downloaded from MMRF Researcher Gateway (<https://research.themmr.org>).

## Other Statistical Analyses

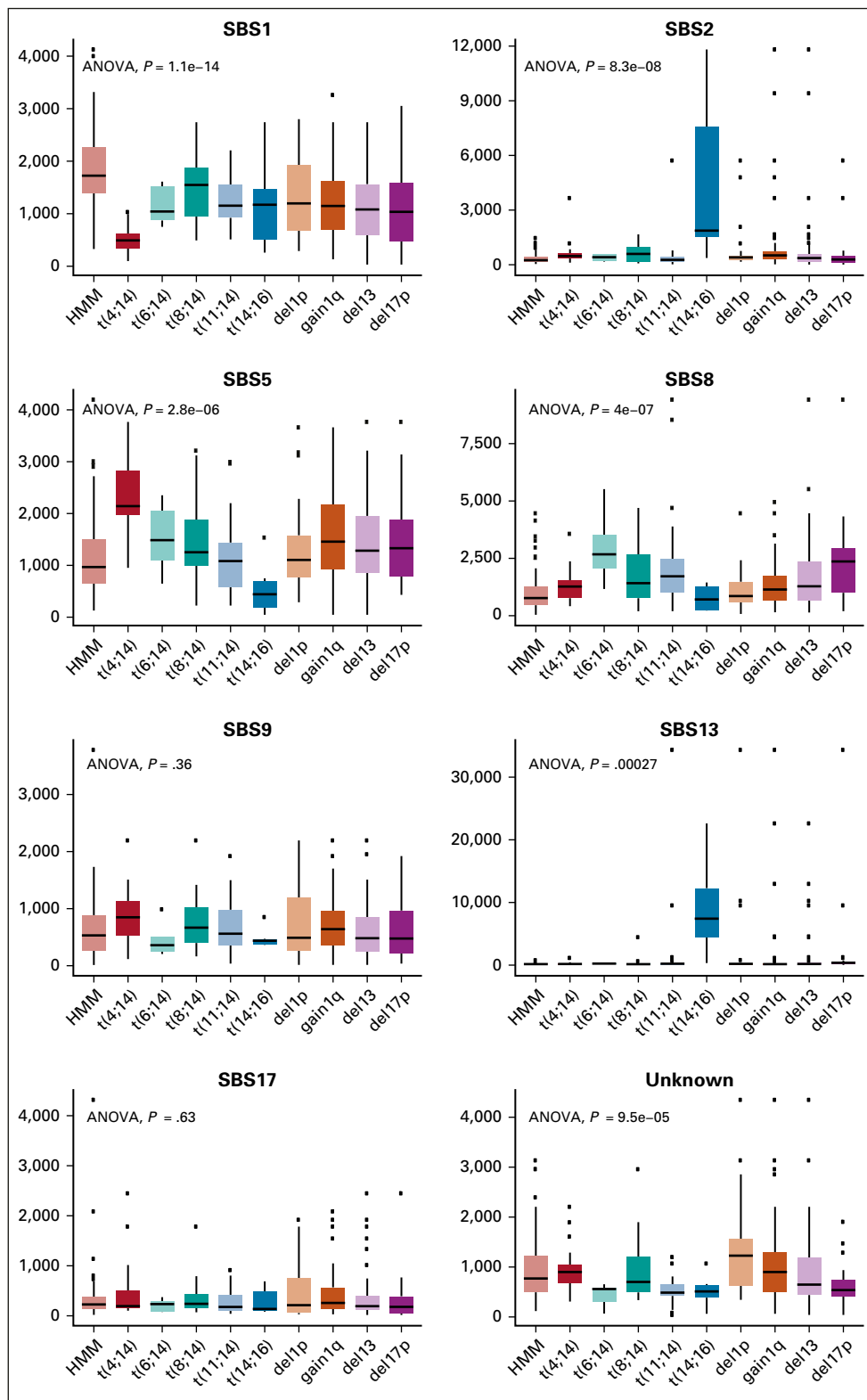
All other analyses were completed in the R programming language. Survival analysis and Kaplan-Meier plots were generated using survival and survminer packages. Data preparation was done using readr, readxl, bedr, and maftools packages, and color schemas for the figures were selected using RColorBrewer. GSS visualization with clinical annotation was prepared using "ComplexHeatmap"; "ggpubr" and "ggplot2" were used for other figures. Driver mutations were identified using "dndscv." Mutational signature analysis was completed using signeR, deconstructSigs, and BSgenome.Hsapiens.UCSC.hg38. All these packages were supported with custom scripts. The Kaplan-Meier method was used to estimate time-to-event distributions, and statistical comparisons were done using log-rank tests, with the null hypothesis being all groups have the same outcome.



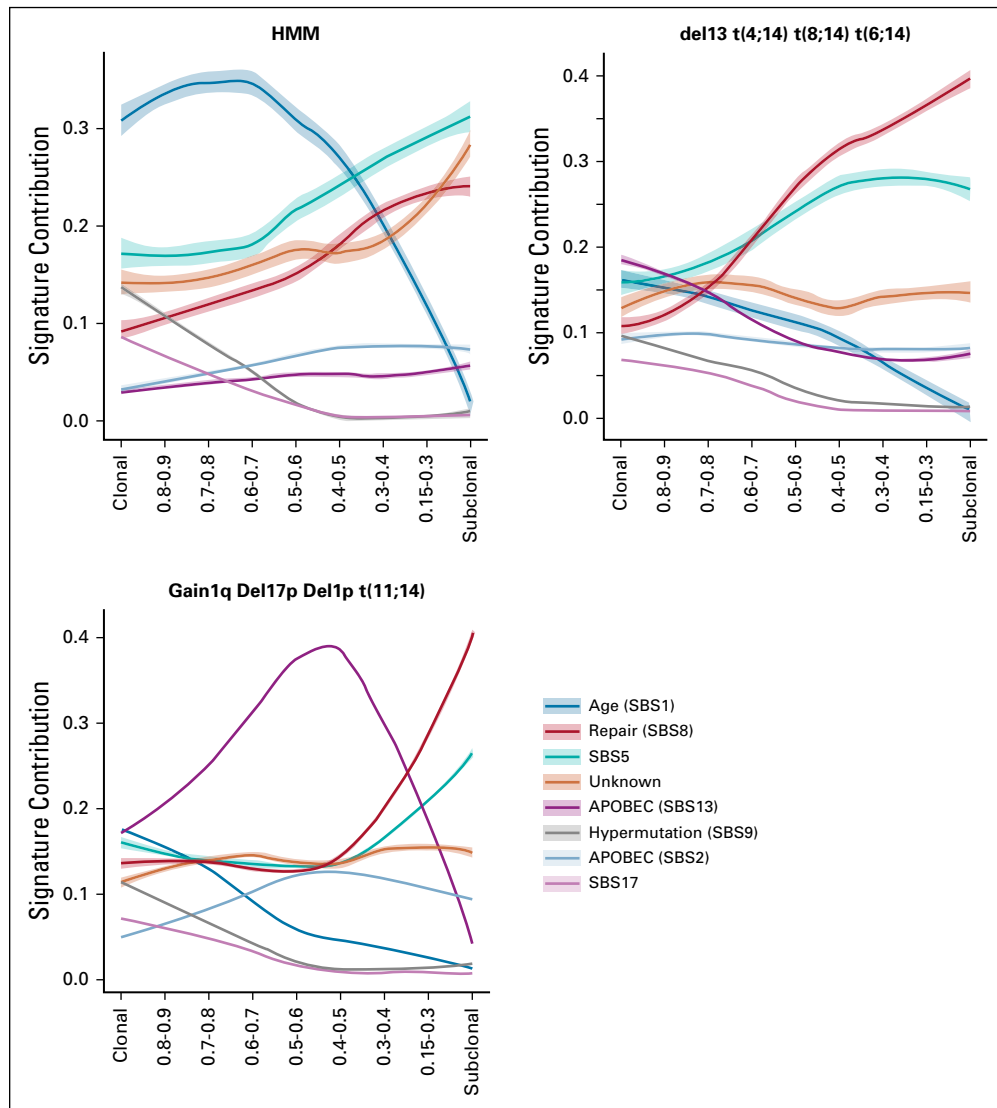




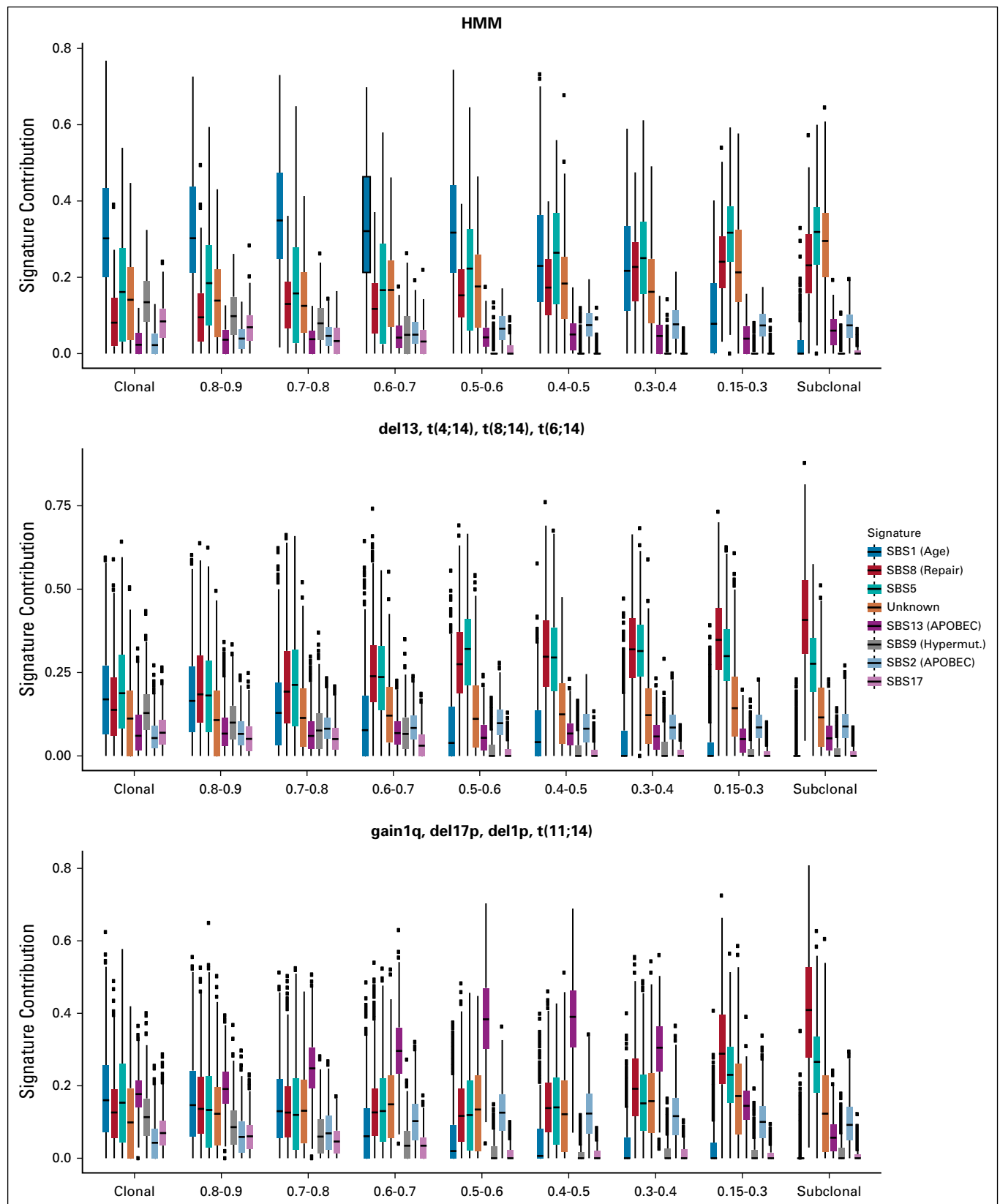
**FIG A2.** Genome-wide mutational signatures of 183 patients with newly diagnosed multiple myeloma (MM). (A) Eight mutational signatures identified by nonnegative matrix factorization using all samples. Contribution of each trinucleotide context (y-axis) is shown, and error bars are given. Each signature is shown in a separate panel. (B) Cosine distance between 8 mutational signatures extracted from the MM cohort and single-base substitution (SBS) signatures provided on Cosmic Mutational Signature V3.



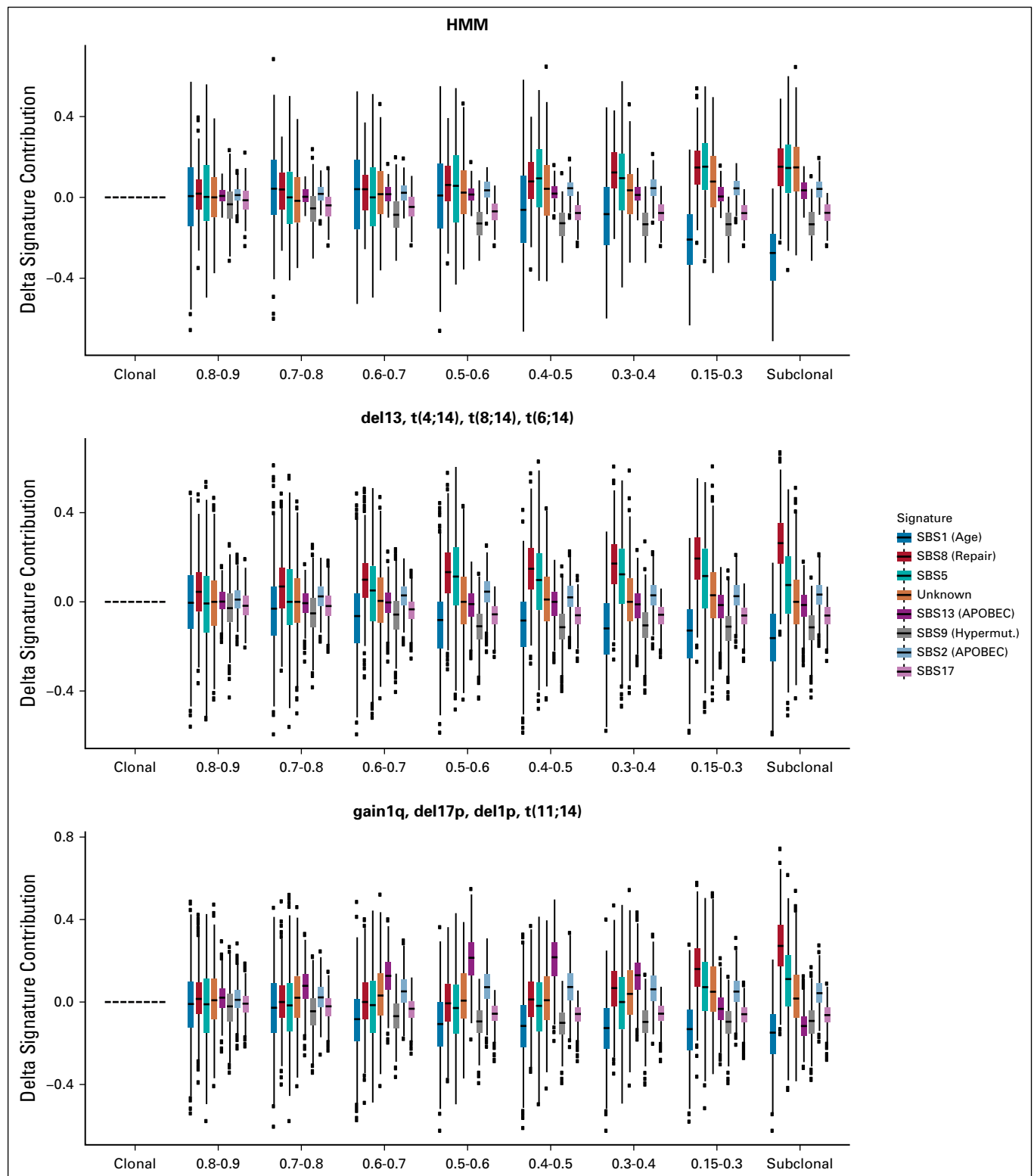
**FIG A3.** Absolute contribution (No. of mutations, y-axis) of 8 mutational signatures to multiple myeloma (MM) subgroups (x-axis). Each signature is shown in panels and ordered by overall contribution to MM. ANOVA, analysis of variance; HMM, hyperdiploid multiple myeloma; SBS, single-base substitution.



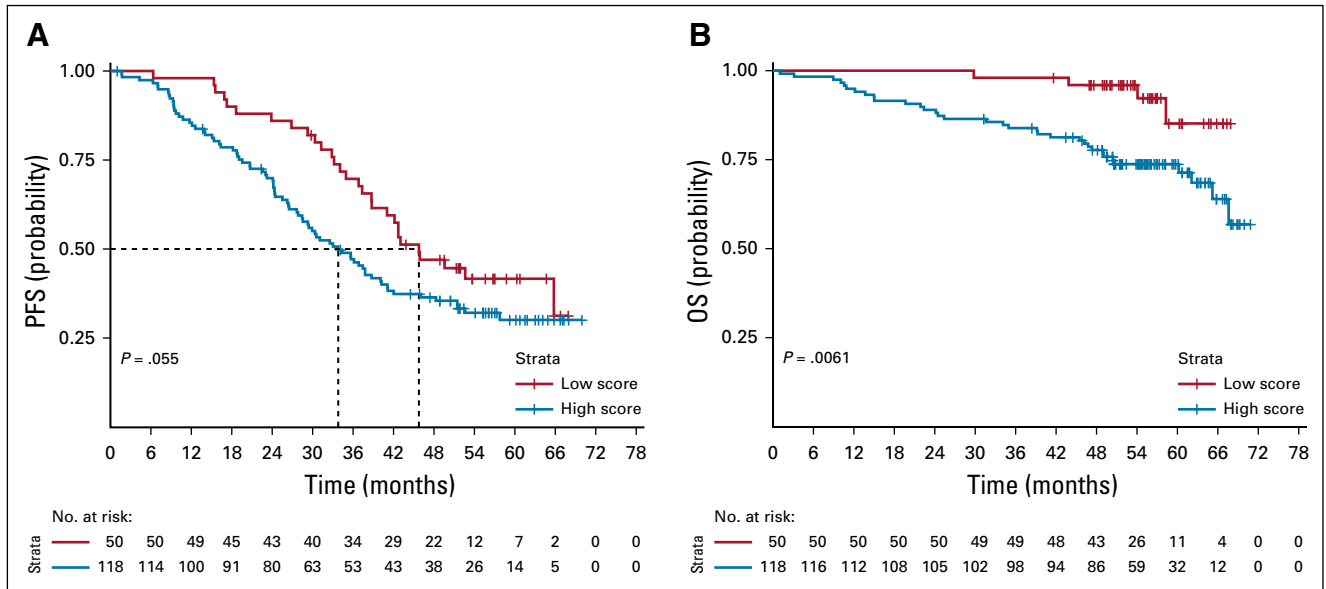
**FIG A4.** Contribution of each mutational signature in multiple myeloma subgroups from clonal to subclonal mutations. HMM, hyperdiploid multiple myeloma; SBS, single-base substitution.



**FIG A5.** Contribution of each mutational signature in multiple myeloma subgroups from clonal to subclonal mutations. Box plots are clustered for signatures at clonality levels and use the same color codes. Each box plot at clonality level shows the distribution of signature contribution among the patients belonging to each subcategory (panels from top to bottom). HMM, hyperdiploid multiple myeloma; SBS, single-base substitution.

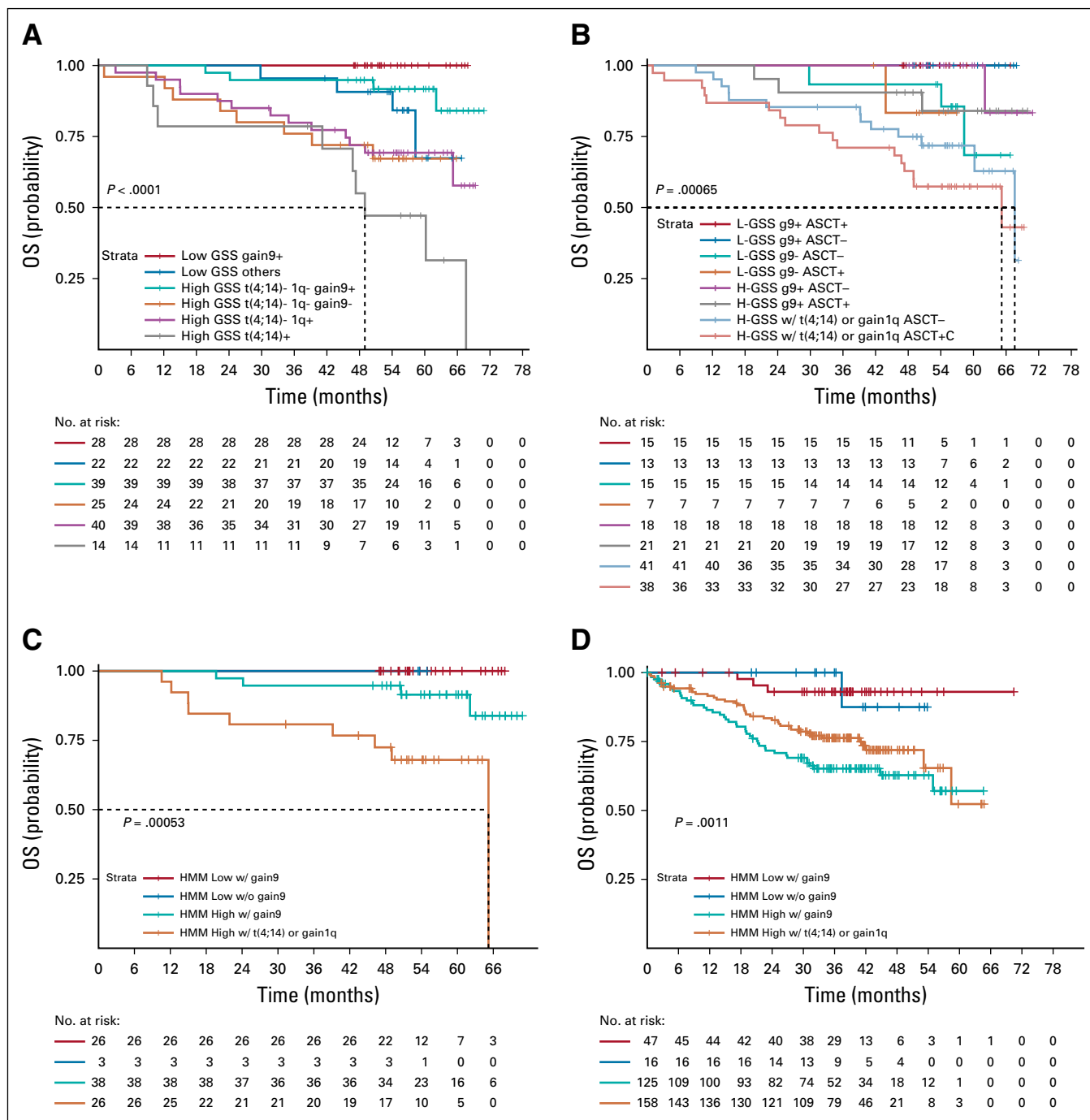


**FIG A6.** Relative changes of each mutational signature contribution in multiple myeloma subgroups from clonal to subclonal mutations. Box plots are clustered for signatures at clonality levels and use the same color codes. Each box plot at clonality level shows the relative contribution (relative to clonal) of mutational signatures among the patients belonging to each subcategory (panels from top to bottom). HMM, hyperdiploid multiple myeloma; SBS, single-base substitution.

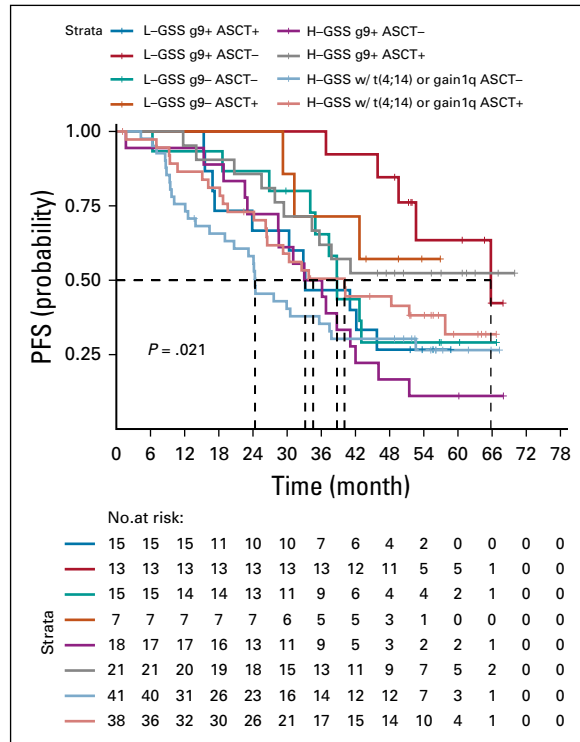


**FIG A7.** Genomic scar score (GSS) is associated with progression-free survival (PFS) and overall survival (OS). (A) PFS probability of low-GSS (red) and high-GSS (blue) groups. (B) OS probability of low-GSS (L-GSS; red) and high-GSS (H-GSS; blue) groups. ASCT, autologous stem cell transplant.





**FIG A8.** Genomic scar score (GSS) is associated with overall survival (OS). (A) OS probability of all 6 segments defined by recursive partitioning. (B) Kaplan-Meier plot shows the OS probability for 4 groups defined by recursive partitioning in IFM/DFCI cohort. Patients in each group divided by treatment arm they were assigned to. (C) Kaplan-Meier plot shows the OS probability for patients with hyperdiploid multiple myeloma (HMM) in 4 groups defined by recursive partitioning in the IFM/DFCI cohort. (D) Kaplan-Meier plot shows the overall survival probability for patients with HMM in 4 groups defined by recursive partitioning in the MM Research Foundation CoMMpass data set. w/, with; w/o, without.



**FIG A9.** Kaplan-Meier plots show progression-free survival (PFS) probability for patients with multiple myeloma in 4 groups defined by recursive partitioning in the IFM/DFCI cohort. Patients in each group are divided by the treatment arm to which they were assigned. ASCT, autologous stem cell transplant; H-GSS, high genomic scar score; L-GSS, low genomic scar score.

**TABLE A1.** Quality Control Measurements Including Depth, Adapter %, Contamination %, Duplication %, Estimated Library Size, and Mean Insert Size for 183 Patients With Newly Diagnosed Multiple Myeloma

Sample ID	Adapter (%)	Contamination (%)	Duplication (%; library average)	Estimated Library Size (library average)	Mean Coverage (raw)	Mean Insert Size (library average)	Mean Read Length	Reads Aligned in Pairs (%)	Total Reads
12210	0.2	2.21	14.943	7,530,126,030	79.94	319	151	99.8	2,071,465,878
12530	0.6	0.33	21.126	3,953,486,228	64.13	320	151	99.8	1,824,801,848
12290	0.3	0.25	18.098	6,262,839,037	78.12	345	151	99.7	2,068,498,840
11764	0.2	0.13	14.722	6,997,201,971	74.82	330	151	99.7	1,915,013,832
11837	0.5	0.12	19.899	5,073,243,566	70.28	369	151	99.7	1,889,692,842
11884	0.2	0.77	18.155	5,899,322,217	70.09	364	151	99.8	1,767,545,106
11977	0.4	23.66	19.789	6,845,872,024	119.27	378	151	99.7	3,169,846,114
12150	0.2	0.66	20.254	6,126,095,744	87.34	336	151	99.8	2,312,446,878
11745	0.4	0.32	17.902	5,495,807,791	64.88	350	151	99.7	1,667,950,458
12202	0.4	0.21	20.679	4,187,222,654	66.19	311	151	99.8	1,866,670,876
12466	0.3	0.13	18.635	7,328,574,061	80.2	332	151	99.8	2,179,155,040
12255	0.2	0.13	22.384	5,851,523,012	82.11	370	151	99.7	2,210,886,114
12280	0.2	1.01	13.795	7,856,132,332	166.58	367	151	99.7	4,049,850,010
12560	0.2	0.5	18.362	6,211,788,979	69.03	364	151	99.7	1,853,596,536
12226	0.3	1.78	16.627	10,863,616,668	131.71	358	151	99.6	3,405,369,818
12430	0.3	1.5	16.977	8,253,177,912	98.32	364	151	99.6	2,499,972,758
12163	0.2	0.23	17.724	7,641,443,994	98	364	151	99.7	2,499,422,010
12579	0.4	0.16	14.721	7,752,655,511	80.47	351	151	99.7	1,997,202,402
12607	0.1	0.36	18.366	5,760,295,811	65.14	353	151	99.7	1,699,814,856
12921	0.3	0.19	14.548	7,379,683,876	77.38	375	151	99.6	1,910,798,246
12708	0.4	0.12	19.97	5,712,532,048	77.17	355	151	99.7	2,019,283,722
12941	0.6	0.15	20.149	4,839,990,301	78.29	343	151	99.7	2,116,592,356
12993	0.1	0.34	21.318	6,530,569,115	88	387	151	99.6	2,372,301,468
13017	0.4	1.46	18.592	5,131,450,265	67	342	151	99.7	1,784,167,450
13070	0.5	17.5	21.772	5,063,756,842	75.69	336	151	99.8	2,066,177,002
13400	0.5	0.13	25.013	5,121,568,883	99.68	352	151	99.8	2,820,695,624
13413	0.4	0.17	22.229	4,109,155,485	69.54	343	151	99.8	1,913,422,734
13414	0.3	0.41	14.188	6,132,096,795	71.73	362	151	99.6	1,774,783,620
13563	0.3	24.34	19.157	6,646,959,559	79.93	374	151	99.7	2,089,484,536
13578	0.1	0.29	18.512	7,092,606,434	97.22	358	151	99.6	2,520,260,436
13655	0.5	0.38	18.126	6,722,527,732	77.84	379	151	99.6	2,031,542,098
13630	0.3	0.54	14.637	8,788,179,697	86.36	358	151	99.6	2,147,943,734
13634	0.3	0.16	14.584	8,736,463,103	85.96	365	151	99.6	2,172,313,072
13599	0.6	0.21	19.714	5,209,629,968	70.43	349	151	99.7	1,840,892,782
14069	0.6	21.62	16.731	6,200,868,167	99.66	340	151	99.7	2,602,817,688
14219	0.4	0.18	23.879	3,988,498,553	83.45	340	151	99.8	2,353,768,368
14246	0.2	0.31	19.832	7,491,038,610	103.22	372	151	99.6	2,672,301,088
14270	0.3	0.46	19.127	6,299,377,733	85.2	342	151	99.8	2,216,271,374
14338	0.3	0.23	18.574	6,655,387,440	72.9	347	151	99.7	1,949,005,954
14345	0.4	0.17	20.754	4,637,776,352	73.07	349	151	99.8	1,966,184,132
14420	0.2	0.66	19.869	5,684,676,143	69.17	369	151	99.6	1,817,646,440

(continued on following page)

**TABLE A1.** Quality Control Measurements Including Depth, Adapter %, Contamination %, Duplication %, Estimated Library Size, and Mean Insert Size for 183 Patients With Newly Diagnosed Multiple Myeloma (continued)

Sample ID	Adapter (%)	Contamination (%)	Duplication (%; library average)	Estimated Library Size (library average)	Mean Coverage (raw)	Mean Insert Size (library average)	Mean Read Length	Reads Aligned in Pairs (%)	Total Reads
14464	0.4	0.62	14.991	5,672,259,764	65.84	334	151	99.7	1,710,634,048
14720	0.3	0.19	21.243	5,058,352,886	76.02	365	151	99.7	2,072,103,950
14742	0.2	0.62	20.993	4,994,763,539	76.24	368	151	99.8	1,993,037,386
14782	1.6	0.12	22.05	4,460,729,964	71.84	346	151	99.6	1,983,057,992
14800	0.4	2.99	15.583	7,348,038,974	75.8	352	151	99.6	1,934,891,610
14832	0.2	0.12	19.239	5,923,643,645	71.88	381	151	99.6	1,882,947,728
14380	0.2	0.98	16.312	8,599,412,307	93.68	354	151	99.7	2,390,550,314
14909	0.2	1.66	14.989	6,306,338,049	71.47	362	151	99.7	1,798,567,856
14922	0.5	0.38	24.496	5,201,424,089	97.93	333	151	99.7	2,876,144,102
15016	0.4	23.94	24.001	5,892,542,558	121.52	326	151	99.8	3,473,979,454
15127	0.4	2.56	21.303	5,422,445,161	75.01	371	151	99.6	1,951,957,766
15128	0.8	0.27	19.749	4,455,117,096	65.89	315	151	99.8	1,825,924,210
15190	0.3	0.41	17.866	6,559,363,646	64.04	379	151	99.6	1,610,366,254
15202	0.2	1.86	21.69	5,065,869,203	75.26	413	151	99.6	1,952,333,144
15212	0.2	0.1	17.328	6,100,174,332	63.8	344	151	99.8	1,631,481,058
15228	0.9	0.43	19.768	4,748,156,739	70.55	312	151	99.7	1,944,946,290
15251	0.1	50	14.647	8,633,800,418	179.78	383	151	99.6	4,334,407,474
15297	0.3	0.48	24.536	5,431,540,164	88.23	350	151	99.7	2,466,561,290
15321	0.2	0.09	18.267	6,312,162,470	74.39	351	151	99.7	1,909,864,416
15359	0.5	14.79	19.83	6,698,729,094	88.43	342	151	99.7	2,377,734,156
15492	0.4	3.61	19.872	5,425,175,783	67.39	348	151	99.7	1,781,607,752
15528	0.6	0.14	14.463	8,548,570,163	82.55	333	151	99.6	2,135,617,916
15563	0.4	0.22	23.252	4,665,768,205	75.42	349	151	99.7	2,105,224,768
15585	0.4	0.17	15.641	7,076,649,021	85.45	336	151	99.6	2,200,532,168
15597	0.4	0.12	19.727	5,787,173,456	78.46	368	151	99.7	2,049,635,314
15598	0.1	0.14	18.373	7,159,889,783	77.24	373	151	99.6	2,004,582,622
15677	0.4	0.2	23.142	4,240,608,746	71.74	367	151	99.7	1,949,858,008
15678	0.3	0.1	16.807	5,759,946,048	72.02	374	151	99.6	1,817,453,716
15765	0.2	0.21	22.148	5,785,058,792	82.35	369	151	99.7	2,192,215,234
15801	0.7	21	19.459	5,805,826,329	74.15	313	151	99.7	2,040,679,518
15805	0.3	0.17	15.463	8,735,385,326	94.02	353	151	99.5	2,369,431,090
15822	0.1	0.41	17.767	7,222,720,897	75.71	413	151	99.6	1,901,287,668
15898	0.4	1.45	20.128	6,673,774,828	89.71	335	151	99.8	2,417,572,006
15915	1.2	0.1	20.658	4,706,332,318	73.19	327	151	99.7	2,017,707,144
15923	0.2	16.38	18.331	9,296,856,822	101.27	340	151	99.7	2,634,581,170
15956	0.3	0.26	19.381	5,362,826,405	68.54	349	151	99.7	1,831,376,028
15998	0.3	0.16	21.615	5,465,950,425	87.47	345	151	99.8	2,377,902,620
15999	0.1	0.14	20.338	6,911,323,576	89.28	375	151	99.7	2,311,414,860
16085	0.3	0.16	23.166	5,679,217,220	91.97	394	151	99.7	2,476,610,604
16124	0.3	0.14	20.634	4,977,361,253	73.77	352	151	99.7	1,988,433,348
16125	0.4	0.13	20.025	4,965,735,390	71.82	350	151	99.6	1,899,275,468

(continued on following page)

**TABLE A1.** Quality Control Measurements Including Depth, Adapter %, Contamination %, Duplication %, Estimated Library Size, and Mean Insert Size for 183 Patients With Newly Diagnosed Multiple Myeloma (continued)

Sample ID	Adapter (%)	Contamination (%)	Duplication (%; library average)	Estimated Library Size (library average)	Mean Coverage (raw)	Mean Insert Size (library average)	Mean Read Length	Reads Aligned in Pairs (%)	Total Reads
16159	0.2	0.75	23.257	5,540,134,133	82.53	391	151	99.6	2,201,781,484
16183	0.3	0.12	18.304	9,972,808,533	119.09	373	151	99.6	3,029,067,790
16212	0.5	0.36	20.936	5,989,262,477	96.26	352	151	99.7	2,584,922,326
16216	1.2	0.68	19.929	4,996,035,346	63.25	318	151	99.7	1,822,431,366
16217	0.3	0.15	22.063	5,124,189,203	82.59	348	151	99.7	2,251,285,776
16244	0.3	0.15	22.913	5,257,478,365	85.39	345	151	99.7	2,365,089,880
16245	0.5	0.67	23.975	3,511,579,532	64.15	346	151	99.8	1,802,527,880
16259	0.2	0.25	15.732	5,834,198,668	66.18	348	151	99.7	1,675,501,836
16288	0.4	0.36	20.486	6,658,171,968	93.2	340	151	99.7	2,557,561,406
16390	0.3	0.2	19.465	4,685,702,479	78.06	350	151	99.8	2,023,943,860
16404	0.2	0.16	22.142	4,188,594,305	70.95	348	151	99.7	1,914,639,754
16462	0.4	0.12	20.751	4,288,996,535	66.91	344	151	99.7	1,786,888,124
16479	0.3	0.37	19.752	5,133,831,444	73.45	381	151	99.7	1,933,204,794
16494	0.1	0.99	16.101	8,194,747,090	71.52	398	151	99.6	1,780,878,304
16527	0.6	0.22	20.48	4,937,761,241	66.02	344	151	99.7	1,755,149,014
16544	0.3	1.16	23.147	3,791,675,109	72.54	342	151	99.8	1,969,967,246
16572	0.4	20.45	17.39	7,015,022,621	71.79	358	151	99.7	1,841,985,208
12541_D	0.3	0.34	17.413	6,001,762,988	173.04	366	151	99.7	4,346,427,506
12674_D	0.6	0.18	23.942	3,703,945,836	71.93	341	151	99.7	2,019,245,360
12839_D	0.3	0.64	22.926	4,911,343,573	83.82	340	151	99.7	2,337,643,708
12893_D	0.2	0.21	15.264	8,913,184,332	85.15	314	151	99.8	2,251,633,948
13738_D	0.3	0.69	23.025	4,667,662,753	71.42	377	151	99.7	1,961,472,742
14498_D	0.1	1.77	21.522	5,882,050,457	92.95	362	151	99.8	2,541,904,056
14783_D	0.1	0.2	19.842	8,415,144,086	110.15	381	151	99.7	2,827,136,694
15415_D	3.2	0.41	20.674	4,933,283,543	73.54	330	151	99.6	2,088,426,908
15582_D	0.1	0.71	15.393	7,628,378,472	75.19	358	151	99.7	1,889,179,208
15803_D	0.3	0.16	21.358	4,548,743,241	74.11	365	151	99.8	1,960,476,960
16045_D	0.3	0.67	18.963	5,243,502,639	73.33	357	151	99.8	1,884,105,388
16327_D	0.2	0.18	19.963	6,170,035,603	79.37	403	151	99.6	2,108,623,044
12994_D	0.3	0.56	20.188	5,106,780,039	69.65	345	151	99.7	1,894,894,312
13629_D	1	0.28	21.781	3,696,325,829	64.94	302	151	99.8	1,898,507,330
14688_D	0.2	0.18	18.688	6,729,944,964	74.64	350	151	99.8	1,976,792,956
14766_D	0.2	0.57	23.091	4,521,840,392	89.16	354	151	99.8	2,446,019,252
14868_D	0.4	0.52	19.251	5,038,772,620	66.99	297	151	99.7	1,911,714,142
15697_D	0.2	0.16	20.159	6,401,029,398	88.53	335	151	99.7	2,383,684,960
15959_D	0.5	0.1	19.922	4,897,491,613	69.12	346	151	99.8	1,822,199,020
11957	0.3	0.11	15.073	7,545,259,335	95.14	382	151	99.7	2,308,456,482
12386	0.3	0.11	12.678	9,186,056,437	96.23	389	151	99.6	2,272,464,282
12339	0.4	0.87	15.932	5,445,654,746	84.85	366	151	99.7	2,077,207,926
12605	0.2	0.12	11.952	8,931,267,108	94.77	388	151	99.7	2,217,578,608
12628	0.2	1.87	9.112	9,789,835,721	66.66	395	151	99.6	1,494,930,230

(continued on following page)

**TABLE A1.** Quality Control Measurements Including Depth, Adapter %, Contamination %, Duplication %, Estimated Library Size, and Mean Insert Size for 183 Patients With Newly Diagnosed Multiple Myeloma (continued)

Sample ID	Adapter (%)	Contamination (%)	Duplication (%; library average)	Estimated Library Size (library average)	Mean Coverage (raw)	Mean Insert Size (library average)	Mean Read Length	Reads Aligned in Pairs (%)	Total Reads
12675	0.2	0.89	9.482	9,527,715,192	65.54	404	151	99.4	1,523,283,418
12707	0.1	0.29	13.985	4,697,571,637	69.69	419	151	99.5	1,642,052,088
12906	0.3	0.26	11.279	8,367,517,203	73.67	399	151	99.6	1,715,893,582
16310	0.3	0.16	17.7	5,697,850,669	98.11	382	151	99.7	2,501,996,948
16328	0.2	0.13	13.595	5,372,572,579	74.04	365	151	99.7	1,805,954,176
16342	0.3	0.89	21.037	3,885,359,466	108.24	338	151	99.7	2,874,569,788
16356	0.2	0.45	17.518	3,384,598,813	69.89	362	151	99.7	1,807,068,978
16359	0.3	0.4	13.267	5,356,486,768	71.83	415	151	99.4	1,730,640,102
16405	0.1	1.37	18.026	3,972,666,236	93.34	391	151	99.6	2,304,017,556
16427	0.1	1.24	12.991	6,181,039,577	72.16	384	151	99.7	1,764,091,112
16439	0.2	0.11	22.839	3,020,252,220	87.32	390	151	99.6	2,356,144,558
16440	0.6	0.27	13.624	7,416,252,620	80.25	382	151	99.7	1,964,335,188
16443	0.2	0.45	15.384	4,982,462,622	67.61	394	151	99.7	1,659,627,932
16468	0.3	0.11	10.526	8,019,063,933	62.24	349	151	99.7	1,487,384,214
14767_D	0.5	0.1	13.624	6,037,634,356	74	389	151	99.6	1,794,086,450
15284_D	0.2	0.2	11.933	6,090,972,825	66.16	343	151	99.7	1,603,083,136
15318_D	0.1	2.02	15.678	4,623,821,467	81.21	383	151	99.6	2,020,833,974
15495_D	0.3	0.3	12.214	6,786,224,485	70.74	366	151	99.7	1,664,161,728
15555_D	0.2	0.15	12.064	6,850,157,775	71.21	376	151	99.6	1,710,139,094
15596_D	0.2	9.04	17.495	5,445,685,656	85.74	359	151	99.7	2,238,235,160
15941_D	0.3	2.17	12.793	5,767,743,512	75.54	388	151	99.6	1,794,431,652
16036_D	0.3	0.11	13.07	7,543,840,571	69.33	374	151	99.7	1,709,410,814
16433_D	0.6	0.35	12.245	5,760,631,779	64.7	352	151	99.7	1,587,399,792
16506_D	0.1	0.68	12.666	9,591,973,365	105.9	391	151	99.6	2,480,823,118
11576_D	0.3	0.1	14.858	4,795,357,518	67.21	386	151	99.7	1,627,043,334
11840_D	0.5	0.16	15.321	5,035,428,061	70.65	387	151	99.5	1,714,528,822
11845_D	0.2	0.15	9.358	10,588,296,942	87.95	392	151	99.6	1,973,491,652
11883_D	0.1	0.58	13.082	5,712,152,182	83.83	392	151	99.6	1,962,948,168
12162_D	0.5	0.82	12.903	6,024,305,533	76.25	393	151	99.6	1,819,357,340
12531_D	0.2	0.13	18.255	3,571,943,927	89.99	367	151	99.7	2,248,625,708
12844_D	0.4	0.4	12.018	7,266,921,260	66.74	335	151	99.7	1,635,236,212
12886_D	0.4	0.29	15.138	5,244,346,265	77.35	407	151	99.5	1,861,412,842
12930_D	0.2	0.13	12.226	8,410,323,910	82.39	383	151	99.7	1,953,150,734
13510_D	0.1	0.64	14.934	5,153,555,556	82.52	376	151	99.7	2,014,043,684
13532_D	0.3	0.09	14.759	4,444,948,215	70.17	387	151	99.6	1,698,902,704
13577_D	0.3	0.37	15.843	4,826,441,870	88.68	433	151	99.3	2,161,124,706
13739_D	0.1	0.29	9.99	8,419,622,169	77.68	361	151	99.6	1,771,194,588
13856_D	0.1	0.22	12.428	4,513,073,074	64.37	408	151	99.5	1,492,451,368
14112_D	0.2	0.45	14.703	6,408,465,334	107.51	410	151	99.5	2,579,907,660
14534_D	0.1	0.69	15.367	3,528,747,485	68.06	396	151	99.6	1,625,250,576
15124	0.2	0.47	9.503	7,628,834,467	66.13	385	151	99.6	1,512,252,290

(continued on following page)

**TABLE A1.** Quality Control Measurements Including Depth, Adapter %, Contamination %, Duplication %, Estimated Library Size, and Mean Insert Size for 183 Patients With Newly Diagnosed Multiple Myeloma (continued)

Sample ID	Adapter (%)	Contamination (%)	Duplication (%; library average)	Estimated Library Size (library average)	Mean Coverage (raw)	Mean Insert Size (library average)	Mean Read Length	Reads Aligned in Pairs (%)	Total Reads
15229	0.1	0.32	9.272	8,678,642,234	72.02	396	151	99.7	1,614,222,688
15474	0.4	0.5	14.561	5,975,905,539	81.23	364	151	99.6	2,033,958,952
15903	0.3	0.25	14.12	6,743,361,274	69.17	367	151	99.7	1,714,941,158
13179	0.6	0.13	16.31	4,814,886,757	83.62	362	151	99.7	2,114,163,588
13231	0.1	0.18	10.796	8,096,417,639	80.27	404	151	99.5	1,858,254,482
13533	0.2	0.13	12.909	5,034,312,974	66.51	380	151	99.6	1,621,171,946
13671	0.2	0.86	10.922	6,239,751,266	66.69	363	151	99.7	1,542,148,614
13682	0.1	0.83	10.819	9,586,920,193	71.67	444	151	99.3	1,673,448,392
14134	0.1	0.66	12.193	5,458,562,104	64.25	368	151	99.7	1,522,773,688
13785	0.4	10.37	13.934	6,161,932,103	83.91	380	151	99.6	2,052,690,720
13884	0.1	0.12	9.811	6,893,473,303	67.95	407	151	99.5	1,538,318,724
14040	0.1	11.8	13.916	4,358,266,282	70.99	372	151	99.7	1,710,100,664
14271	0.1	0.64	11.903	8,078,111,745	85.7	396	151	99.5	2,010,242,514
14402	0.4	17.16	13.68	5,182,884,215	80.04	375	151	99.6	1,939,438,442
14446	0.2	0.58	13.332	5,026,374,049	69.25	331	151	99.8	1,739,037,490
14666	0.1	0.33	10.097	11,016,800,931	82.29	406	151	99.7	1,879,920,620
14835	0.2	0.66	11.254	6,280,007,665	63.74	377	151	99.6	1,481,756,768
13967	0.1	0.45	12.308	7,045,250,839	73.04	401	151	99.6	1,700,323,712
13710	0.4	1.6	14.394	5,329,833,247	75.8	384	151	99.7	1,875,795,894