Am J Cancer Res 2022;12(4):1919-1933
www.ajcr.us /ISSN:2156-6976/ajcr0140875

Original Article

Characterizing the mutational landscape of MM and its precursor MGUS

Akanksha Farswan[1], Anubha Gupta[1], Lingaraja Jena[2], Vivek Ruhela[3], Gurvinder Kaur[2], Ritu Gupta[2]

[1]SBILab, Department of ECE, Indraprastha Institute of Information Technology-Delhi (IIIT-Delhi), New Delhi 110020, India; [2]Laboratory Oncology Unit, Dr. B.R.A. IRCH, AIIMS, New Delhi 110029, India; [3]Department of Computational Biology, IIIT-Delhi, New Delhi 110020, India

Received November 30, 2021; Accepted March 2, 2022; Epub April 15, 2022; Published April 30, 2022

Abstract: Mutational Signatures and Tumor mutational burden (TMB) have emerged as prognostic biomarkers in cancer genomics. However, the association of TMB with overall survival (OS) is still unknown in newly diagnosed multiple myeloma (NDMM) patients. Further, the change in the mutational spectrum involving both synonymous and non-synonymous mutations as MGUS progresses to MM is unexplored. This study addresses both these aspects via extensive evaluation of the mutations in MGUS and NDMM. WES data of 1018 NDMM patients and 61 MGUS patients collected from three different global regions were analyzed in this study. Single base substitutions, mutational signatures and TMB were inferred from the variants identified in MGUS and MM patients. The cutoff value for TMB was estimated to divide patients into low TMB and high TMB (hypermutators) groups. This study finds a change in the mutational spectrum with a statistically significant increase from MGUS to MM. There was a statistically significant increase in the frequency of all the three categories of variants, non-synonymous (NS), synonymous (SYN), and others (OTH), from MGUS to MM ($P<0.05$). However, there was a statistically significant rise in the TMB values for TMB_NS and TMB_SYN only. We also observed that 3' and 5'UTR mutations were more frequent in MM and might be responsible for driving MGUS to MM via regulatory binding sites. NDMM patients were also examined separately along with their survival outcomes. The frequency of hypermutators was low in MM with poor OS and PFS outcome. We observed a statistically significant rise in the frequency of C>A and C>T substitutions and a statistically significant decline in T>G substitutions in the MM patients with poor outcomes. Additionally, there was a statistically significant increase in the TMB of the patients with poor outcome compared to patients with a superior outcome. A statistically significant association between the APOBEC activity and poor overall survival in MM was discovered. These findings have potential clinical relevance and can assist in designing risk-adapted therapies to inhibit the progression of MGUS to MM and prolong the overall survival in high-risk MM patients.

Keywords: Multiple myeloma, monoclonal gammopathy of undetermined significance, NGS, exome sequencing, tumor mutation burden, progression, mutational landscape

Introduction

Multiple Myeloma (MM) is a malignancy of abnormal plasma cells in the bone marrow where the progression of the disease is driven by numerous factors, including immune surveillance, microenvironment, and therapeutic agents. Monoclonal gammopathy of undetermined significance (MGUS) is a benign precursor state of MM characterized by lack of end-organ damage [1] and less than 10% of plasma cells in the bone marrow. MGUS may progress to asymptomatic or symptomatic multiple myeloma with a rate of nearly 1% per year [2], where MM is characterized by severe clinical problems such as bone fractures, anaemia, renal failure, and hypercalcemia. With the advent of Next Generation Sequencing technology, it has become easier to study the DNA of a patient and unearth the genetic causes of the disease. Multiple studies involving exome and genome data of MM have been performed to understand the genomic abnormalities driving tumor progression in MM. It is well established that the primary events in MM are either hyperdiploidy, i.e., trisomy of chromosomes 3, 5, 7, 9, 11, 15, 19 and/or 21 or non-hyperdiploidy involving translocations affecting the genes

encoding immunoglobulin (Ig) heavy chains (IGH)-mainly t(4;14), t(6;14), t(11;14), t(14;16) and t(14;20) [3]. Primary events are then followed by multiple secondary events promoting tumor progression. However, it has also been observed and validated that the genetic aberrations peculiar to MM are also present during the premalignant state of MGUS, where they do not show any clinical symptoms related to MM [4, 5]. It is, therefore, worthwhile to thoroughly investigate the mutational landscape of the genomic alterations affecting MGUS as well as MM. Though multiple studies have been performed to study the MGUS to MM progression [6-8], the landscape of the mutational patterns of the MGUS and MM largely remains unexplored. The study of the changing mutational spectrum of the MGUS as it advances to MM will provide more insight into the disease biology. Further, it will help identify the clinically relevant vital biomarkers that can assist in controlling the progression of MGUS to MM.

Mutational signatures have emerged as critical biomarkers in cancer genomics, with profound pathogenic, prognostic, and therapeutic implications. Multiple mutational events occur in a tumor, while only a few of these mutations are actual drivers of cancer. However, exploring the entire landscape of coding and non-coding mutations helps reveal the mutational signatures characteristics of the specific cancer types. For example, CG>AT transversion is associated with lung cancer [9], and CG>TA is associated with skin cancer [10, 11]. Various mutational signatures have been discovered based on the 96 possible combinations of the single base substitutions and their trinucleotide contexts. These signatures are linked with the defects of DNA repair mechanisms, ageing, UV exposure, and others, thereby validating the role of the mutational processes in shaping the genomic continuum of each cancer type [12-15]. Further, tumor mutational burden (TMB) has become a prominent biomarker of response to immunotherapy and is being explored for its association with overall survival, particularly in solid tumors. TMB is determined as the number of mutations identified per megabase. It has been observed that cancers with a high TMB load of greater than 10 mut/Mb have a better chance of responding to drugs called immune checkpoint inhibitors (ICIs). The primary function of ICIs is to activate the immune system better to recognize cancer cells [16] and act upon them. As a result, a high tumor mutational burden (TMB) has been increasingly associated with superior overall survival in ICI-treated patients. Multiple studies are now being conducted to discover the cancers with high TMB that respond best to ICIs and, thus, prolong the survival of cancer patients. In addition, the association of TMB with survival in non-ICI-treated patients has also been explored. It has been observed that high TMB was associated with poor prognosis and overall survival in the absence of immunotherapy, as opposed to ICI-treated patients in whom high TMB was associated with prolonged survival [17].

Synonymous mutations, earlier designated as silent mutations, were mostly ignored in cancer genomics due to their inability to alter the amino acid of the resultant protein [18]. However, they have the capability of changing the protein expression and function owing to their impact on RNA stability, RNA folding [19] or splicing [20], translation [21], or co-translational protein folding. Multiple studies have corroborated that natural selection is present in synonymous mutations [21-23], contrary to earlier studies that denied the role of selective pressure in synonymous mutations [24]. Various genome-wide association studies conducted in recent times have also confirmed the association of synonymous SNPs to human disease risk and other complex traits. Therefore, the role of synonymous mutations in the disease biology of MGUS and MM should be examined as it could lead to significant prognostic and clinical implications.

Motivated by the above discussion, an exhaustive investigation of the mutations altered in MGUS and MM was carried out in the present study. We explored the change in the mutational landscape as the disease progressed from the MGUS to MM. We found that the difference in the frequency of the single base substitution is significantly different in MGUS and MM. We have also analyzed the frequency of the different types of variants across MGUS and MM and found that few have changed significantly as the disease progressed from MGUS to MM. Further, we categorized MM patients into low TMB and high TMB (hypermutators) based on their overall survival data. We explored the impact of TMB on the frequency

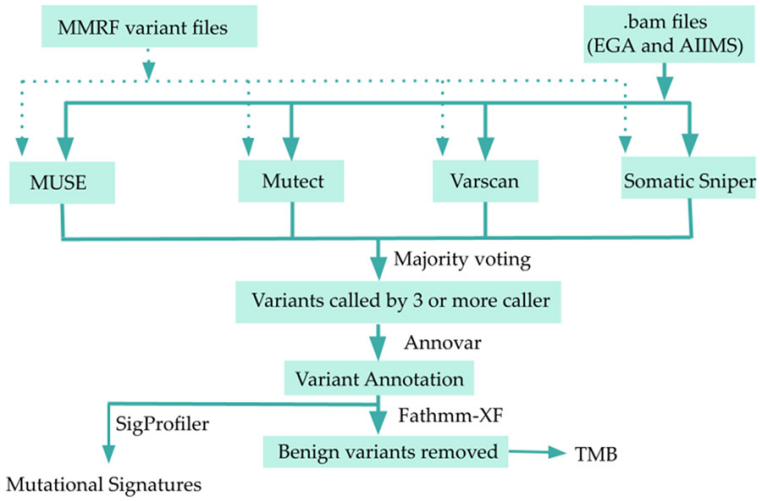


Figure 1. Workflow of the study and data analysis. Four different variant callers were used to identify variants in the MM and MGUS patients. Variants were finalized using the majority voting scheme. Variants were then annotated with Annovar for deducing TMB. Mutational signatures were inferred using Sigprofiler tool.

of single base substitutions and the different variant types across the low and high TMB groups of MM patients. The association of TMB with overall survival is still unknown in newly diagnosed multiple myeloma (NDMM) patients; therefore, we have correlated TMB with survival data and found that high TMB is linked with poor overall survival in NDMM patients.

Methods and materials

Datasets used in the study

The present study is based on the data of 1018 NDMM patients and 61 MGUS patients. Variant files generated from the exome data of 936 NDMM patients out of the total 1018 patients were obtained from the GDC portal via dbGaP authorized access (phs000748; phs000348). This data is a part of the MMRF CoMMpass study. Exome data of the remaining 82 NDMM patients were obtained from AIIMS, Delhi. In addition, exome data of 33 MGUS patients out of 61 patients was obtained from EGA (EGAD00001001901), and exome data of the remaining 28 patients was obtained from AIIMS, Delhi. Four variant callers, namely, MuSE [25], Mutect2 [26], VarScan2 [27], and Somatic-Sniper [28], was used for finding variants in patients from the MMRF CoMMpass study. Therefore, there were four vcf files corresponding to each variant caller for each patient. The workflow of the complete analysis is shown in **Figure 1**.

Analysis of exome data and the variants identified using the exome data

Exome data obtained from AIIMS and EGA was processed with a standard exome sequencing pipeline, and single nucleotide variants (SNVs) were extracted using MuSE, Mutect2, VarScan2, and Somatic-Sniper variant callers. SNVs were annotated using ANNOVAR [29] to gather the genomic information of the mutations, such as their variant type and the deleteriousness of the mutation, etc. FATHMM-XF [30] was used to remove the benign variants. The rest of the filtered variants were categorized into nonsynonymous (NS) variants, synonymous (SYN) variants, and other (OTH) variants. Exonic, nonsynonymous single nucleotide variants (snvs), ncRNA_exonic, stop gain, stop loss, start loss, splicing, frameshift insertion, and frameshift deletion were grouped in nonsynonymous variants. UTR3, synonymous single nucleotide variants (snvs), and UTR5 were grouped in synonymous variants. Non-frameshift insertion, non-frameshift deletion, non-frameshift substitution, intronic, intergenic ncRNA_intronic, upstream, downstream, unknown, and ncRNA_splicing were grouped in other variants.

Assessment of single base substitution, mutational signatures, and TMB

Variants identified by three or more callers were further processed to extract information on single base substitution and identify the mutational signatures present in the data. SigProfilerExtractor [31] was used to discover the single base substitutions and the mutational signatures in the MGUS and MM data. The etiology of the deduced signatures were found via the COSMIC v3.2 mutational signature database [32]. A total of six single base substitutions C>A, C>G, C>T, T>A, T>C, and T>G were identified. Tumor mutational burden (TMB) was calculated using the three different catego-

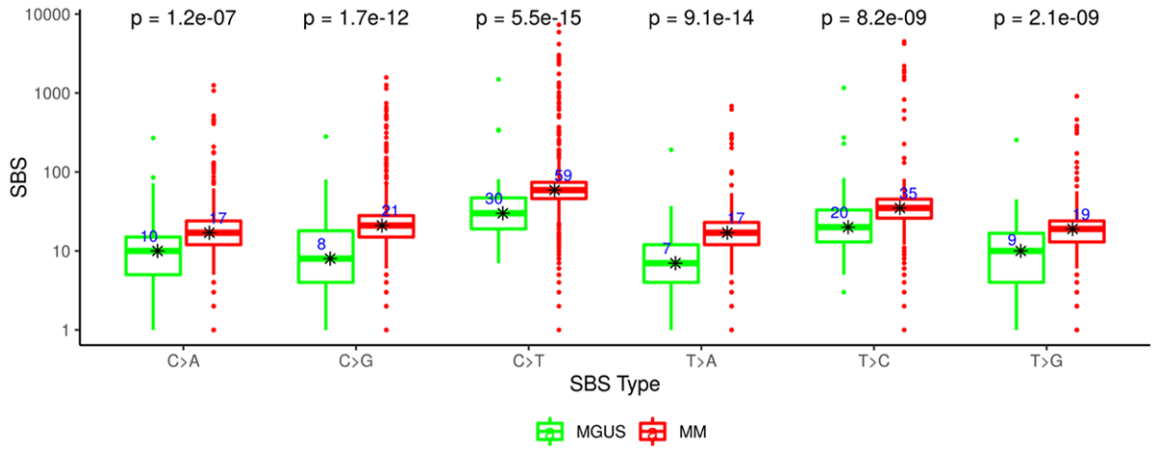


Figure 2. Boxplot shows the difference in the frequency of the single base substitutions between MGUS and MM patients. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not. For all the substitutions, there is significant variation in the frequency with *p*-values less than 0.05 between the two groups.

ries of variants-nonsynonymous (NS) variants, synonymous (SYN) variants, and other (OTH) variants. TMB was determined as described in [33]. TMB_NS, TMB_SYN, and TMB_OTH were estimated using nonsynonymous (NS) variants, synonymous (SYN) variants, and other (OTH) variants, respectively. Survival data were available for 832 (753+79) patients out of a total of 1018 NDMM patients, which were utilized to obtain the threshold values for TMB_NS, TMB_SYN, and TMB_OTH using the K-adaptive partitioning (KAP) algorithm [34] and Cutoff Finder [35].

Statistical analysis

Wilcoxon rank-sum test was used to determine if the change in the frequencies of the single base substitutions and the different types of variants is statistically significant between the MGUS and MM. Unpaired Wilcoxon rank-sum was applied because the data did not follow the normality distribution and was unpaired.

Results

Frequency of single base substitutions (SBS) increases significantly from MGUS to MM

There was an increase in the median and mean frequency of the single base substitutions from MGUS to MM. The change in the frequency was statistically significant with *p*-values less than 0.05 for all the substitutions according to the Wilcoxon rank-sum test (**Figure 2**). C>T substitution was observed with the highest frequency in MGUS and MM, increasing the median value from 30 to 59. T>C substitution was next, with an increase in the median value from 20 (MGUS) to 35 (MM). T>A was observed with the lowest frequency in MGUS and MM, increasing the median value from 7 to 17.

Calculation of threshold values for the SBS and comparison between the high and low-frequency MM groups

Due to the availability of survival data for 832 MM patients, threshold values for the substitutions were inferred. K-adaptive partitioning (KAP) algorithm and Cutoff Finder were used to deduce the thresholds. **Table 1**, Supplementary Table 3 and Supplementary Figures 5 and 6 show the cut-off values estimated for the different types of substitutions for PFS and OS. Similar cut-offs were deduced by the two tools, i.e. KAP and Cutoff Finder. The higher of the two cut-offs obtained via KAP were selected for C>T, T>C, C>G, C>A, T>G, and T>A substitutions and were 99, 12, 37, 28, 6, and 32, respectively. The patients were then organized into two groups, one with SBS values less than the selected cut-offs and the other one with SBS values greater than the chosen cut-offs. Kaplan Meier (KM) curves corresponding to the two groups revealed that there was a significant difference in the survival patterns of the two groups of patients for the substitutions, C>T, C>G, C>A, and T>A. However, cut-offs obtained for T>C and T>G substitutions yielded a significantly poor outcome for the group with values less than the selected cut-offs.

Table 1. The table shows the cut-offs obtained for the six different types of substitutions via KAP

SBS	Min	Median	Max	PFS cutoff	OS cutoff	Manual cut-off	Frequency (≤, >)	PFS *p*-value	OS *p*-value
C>A	0	17	1251	26	28*	-	712, 120	0.00025	5.13E-06
C>G	0	21	1575	37*	34	-	763, 69	0.026	2.20E-04
C>T	1	59	7315	79	99*	-	750, 82	0.001	4.80E-06
T>A	0	17	684	5	32*	-	784, 48	0.01	0.005
T>C	0	35	4498	12	11	80*	816, 16	0.19	0.01
T>G	0	19	915	6	6	41*	804, 28	0.018	0.007

Two cut-offs were obtained for each SBS, one using PFS and the other using OS. The higher of the two cut-offs and the patients were then organized into two groups, one with SBS values less than the selected cut-offs and the other one with SBS values greater than the selected cut-offs. KM analysis showed that there was a significant difference in the survival patterns of the two groups of patients for the substitutions, C>T, C>G, C>A, and T>A. However, cut-offs obtained for T>C and T>G substitutions did not yield a significant difference in the survival curves. Therefore, cutoffs were manually deduced for the two substitutions where the KM curve has the maximum separability. *Denotes selected cutoffs.

Therefore, cut-offs were manually deduced for T>C and T>G substitutions where the KM curve has the maximum separability and was found to be 80 and 41, respectively. Univariate and multivariate hazard analysis was also done using the selected cut-offs via KAP, as shown in the Supplementary Table 5. The hazard ratio for all the substitutions was greater than 1 in the univariate analysis, demonstrating that an increase in the frequency of these substitutions correlated with an enhanced risk in MM patients. Univariate analysis revealed that C>T substitution had the most significant impact (*p*-value <0.05) on the overall survival (OS) owing to the highest hazard ratio followed by T>C and C>A while T>G had the most significant impact (*p*-value <0.05) on PFS followed by C>T and C>A. However, only C>A was significant in multivariate analysis with *p*-values less than 0.05 (0.04 for PFS and 0.03 for OS).

Comparison of mutational signature profiles between MGUS and MM

A total of 29 and 61 SBS signatures were extracted from the mutation data of MGUS and NDMM patients, respectively. Union of 29 and 61 signatures resulted in 66 unique signatures. Signatures SBS37, SBS49, and SBS55 were found only in MGUS. However, their frequency is low as they were found in a single sample in MGUS (1/61=1.6%). SBS49 and SBS55 signatures are possible sequencing artifacts, and the proposed etiology of signature 37 is unknown according to the COSMIC v3.2 mutational signature database. Further, 37 signatures were discovered only in MM. However, 7 out of 37 were mutated in more than 1% MM samples. They include SBS6, SBS7d, SBS9, SBS17b, SBS19, SBS40, and SBS42. The rest of the 30 signatures were found in less than 1% MM samples and include SBS7c, SBS8, SBS10d, SBS14, SBS20, SBS21, SBS22, SBS23, SBS25, SBS26, SBS27, SBS28, SBS30, SBS32, SBS33, SBS34, SBS35, SBS36, SBS39, SBS41, SBS43, SBS46, SBS47, SBS50, SBS52, SBS53, SBS57, SBS86, SBS88, and SBS89. SBS27, SBS43, SBS46, SBS47, SBS50, SBS52, SBS53, and SBS57 are possible sequencing artifacts, as described previously. Clock-like signatures SBS1 and SBS5 were present in both MGUS and MM. Defective DNA mismatch repair signatures SBS15 and SBS44 were present in both MGUS and MM while SBS6, SBS14, SSB20, SBS21, SBS26 were present only in MM. SBS2 and SBS13 are associated with the activity of the AID/APOBEC family of cytidine deaminases and were found in both MGUS and MM. MM patients with APOBEC signatures were investigated further using survival data. APOBEC signature was present in 27 out of 177 MM patients with poor OS outcome and 52 out of 655 MM patients with superior OS outcome. Fisher's exact test revealed a statistically significant association between the APOBEC activity and poor overall survival in MM (*p*-value =0.0056). However, there was no significant association between APOBEC activity and progression-free survival (*p*-value =0.9). KM curves showed a significant difference (*p*-value =1.8e-4) in the overall survival pattern of MM patients with and without APOBEC activity (Supplementary Figure 2). SBS84 and SBS85 are related to indirect effects of activation-induced cytidine deaminase (AID) induced somatic mutagenesis in

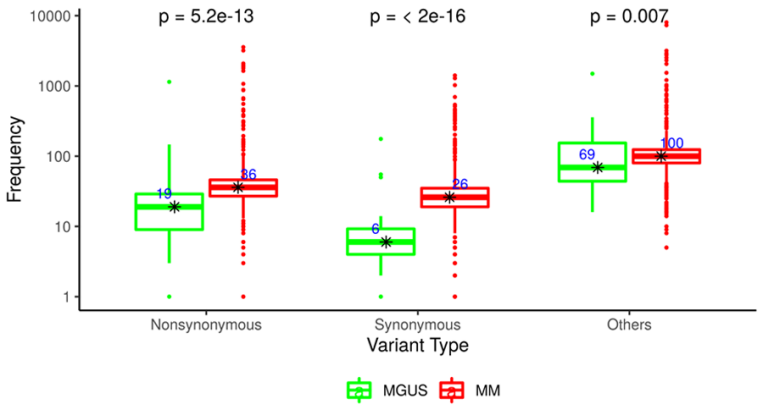


Figure 3. Boxplot showing the variation in the frequency of the three different categories of variants-Nonsynonymous, Synonymous, and Others between MGUS and MM. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not. There was a statistically significant variation in all the categories of variants with *p*-values less than 0.05.

lymphoid cells and were found in both MGUS and MM.

Frequency of the variants increases significantly from MGUS to MM

According to the Wilcoxon rank-sum test, there was a statistically significant increase in all the three categories of variants from MGUS to MM (**Figure 3**). The median value of nonsynonymous variants increased from 19 to 36 (*p*-value =5.2e-13) as the disease progressed from MGUS to MM. Median value of synonymous variants increased from 6 to 26 (*p*-value <2e-16) while that of other variants increased from 69 to 100 (*p*-value =0.007). Within the nonsynonymous category, there was a statistically significant increase in the nonsynonymous snv (*p*-value =2.9e-13) from 14 to 30 and stop-gain (*p*-value =0.016) variants from 0 to 2 as the disease progressed from MGUS to MM (**Figure 4A**). Within the synonymous category, there was a statistically significant increase in the UTR3 (*p*-value <2e-16) and UTR5 variants (*p*-value =2.7e-7) (**Figure 4B**). Within the other variant category, there was a statistically significant increase in the intronic, intergenic, and downstream variants (Supplementary Figure 1). The median value of UTR3 variants increased from 4 to 21, while that of UTR5 increased from 1 to 4.

Comparison of TMB values between MGUS and MM

Tumor mutational burden (TMB) was calculated using the three different categories of variants-nonsynonymous (NS), synonymous (SYN), and others (OTH). A statistically significant increase was observed for TMB_NS and TMB_SYN with *p*-values less than 0.05 (**Figure 5**). For TMB_OTH, the difference in the KM survival curve was not significant (**Figure 5**).

Calculation of TMB cut-offs and comparison between high and low TMB MM groups

Survival data were available for 832 MM patients. Hence, threshold values of TMB were calculated using the K-adaptive partitioning (KAP) algorithm and Cutoff Finder. Both the tools inferred almost the same cut-offs (**Table 2**, Supplementary Tables 1 and 2; Supplementary Figures 3 and 4). **Table 2** and Supplementary Table 1 reveal the different cut-offs obtained via KAP for progression-free survival (PFS) and overall survival (OS). For TMB_NS, 0.63 and 0.62 are the threshold values obtained via PFS and OS.

Similarly, for TMB_SYN, 0.55 and 0.52 are the threshold values obtained for PFS and OS. The patients were then organized into two groups, one with TMB values less than the selected cut-offs and the other one with TMB values greater than the chosen cut-offs. There was a significant difference (*p*-value <0.05) on the KM survival curves of the patients below 0.63/0.62 and above 0.63/0.62. There is a significant difference (*p*-value <0.05) on the KM survival curves of the patients below 0.55/0.52 and above 0.55/0.52. Univariate and multivariate hazard analysis was also done using the cut-offs via KAP, as shown in the Supplementary Table 4. Hazard ratios for TMB_NS, TMB_SYN and TMB_OTH were greater than 1 in both the univariate and multivariate analysis and indicate the enhanced risk associated with an increase in the mutation burden. Multivariate analysis showed the combined effect of the TMB values on the survival patterns where TMB_NS had the highest impact, followed by TMB_OTH and TMB_SYN, respectively.

MM patients with very high TMB_NS load and very low TMB_NS load were analyzed separately. Cut-off of 35 and 0.1 was deduced using the

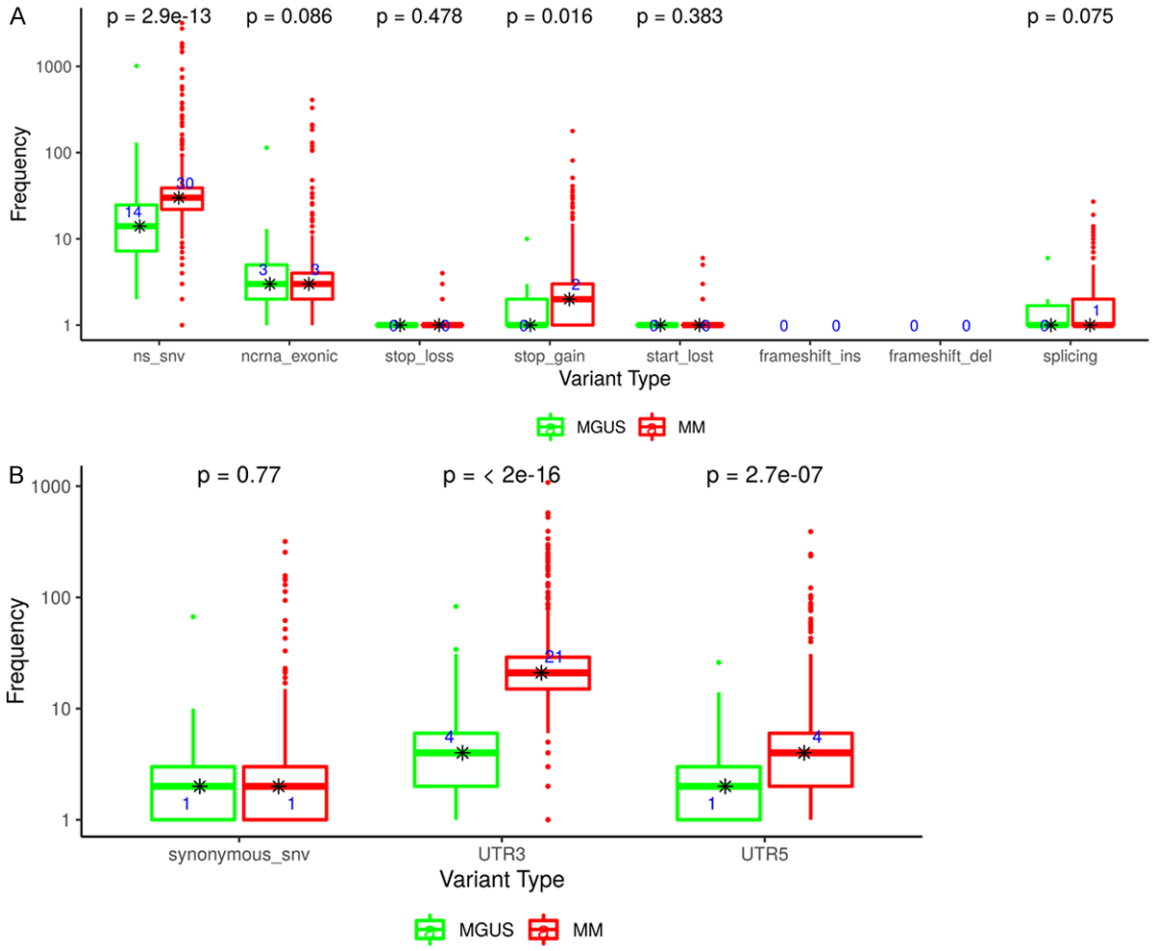


Figure 4. A. Boxplot showing the variation in the frequency of the variants under the nonsynonymous category. There was a statistically significant variation in the frequency of nonsynonymous_snv and stop_gain variants with *p*-values less than 0.05. B. Boxplot showing the variation in the frequency of the variants under the synonymous category. There was a statistically significant variation in the frequency of UTR3 and UTR5 variants with *p*-values less than 0.05. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not.

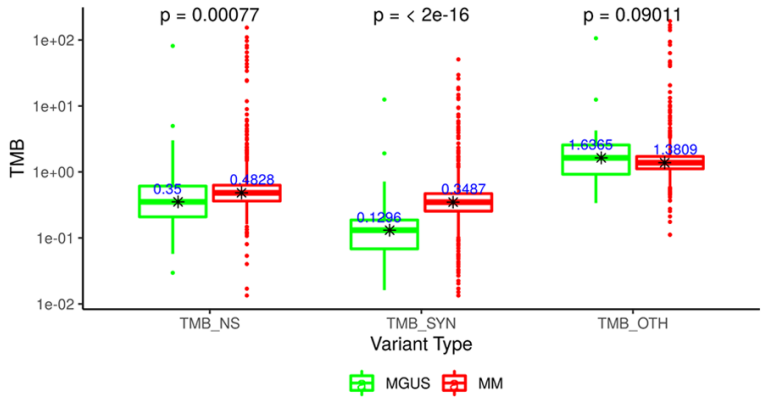


Figure 5. Boxplot reveals that the difference in the low TMB and high TMB groups is statistically significant with *p*-values less than 0.05 for TMB_NS and TMB_SYN. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not.

maximum separability on the KM survival curves. There were 822 patients with TMB_NS less than 35 and only 10 with TMB_NS greater than 35. There were 6 patients with TMB_NS less than 0.1 and 826 patients with TMB_NS greater than 0.1. A significant difference in the survival patterns of patients with TMB_NS less than 35 and greater than 35 were observed. For PFS, the observed *p*-value was 0.045, and for OS, the observed *p*-value was 0.022 (**Figure 6**). The

Table 2. The table shows the cut-offs obtained for TMB_NS and TMB_SYN via KAP. Two cut-offs were obtained, one using PFS and the other using OS. The two cut-offs are close to each other and KM analysis was done using both the cut-offs

	Min	Median	Max	KAP on PFS			KAP on OS		
				Cut-off (≤, >)	PFS	OS	Cut-off (≤, >)	PFS	OS
TMB_NS	0	0.496	154.2	0.63 (612, 220)	3.19E-07	3.52E-08	0.62 (611, 221)	3.90E-07	2.09E-08
TMB_SYN	0	0.3487	50.84	0.55 (703, 129)	4.12E-05	2.05E-08	0.52 (668, 164)	5.60E-04	3.50E-08

There was a significant difference (*p*-value <0.05) on the KM survival curves of the patients below and above the selected cut-offs.

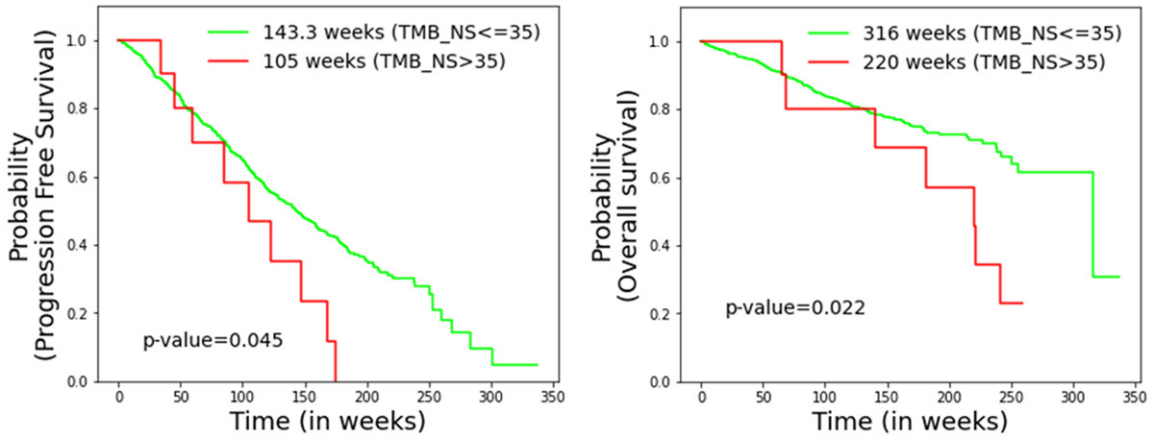


Figure 6. High TMB is associated with poor overall survival in NDMM patients. The difference in the overall survival probability between low and high TMB_NS is statistically significant with *p*-values 0.045 and 0.022 for PFS and OS respectively.

Table 3. The table shows the median values of TMB and SBS for the two groups of MM patients, one where the death event was observed and the other where the death event was not observed

		Median (OS event =0)	Median (OS event =1)	*p-value*
SBS	C>A	17	18	0.018
	C>G	20	21	0.1205
	C>T	59	64	0.038
	T>A	17	16	0.07
	T>C	36	33	0.08
	T>G	19	17	0.02
TMB	TMB_NS	0.4828	0.5766	4.26E-07
	TMB_SYN	0.3487	0.4023	0.002
	TMB_OTH	1.341	1.5288	3.08E-04

Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not. For substitutions, C>A, C>T, and T>G, the frequency was statistically different between the two groups.

patients with TMB_NS greater than 35 are hypermutators, and the characteristics specific to these high-risk patients were examined thoroughly.

Comparison of TMB and SBS based on the overall survival event

Out of 832 MM patients for which survival data were available, 177 observed poor OS outcome while the rest of the 655 MM patients observed superior OS outcome. SBS and TMB values of the two groups were examined, and Wilcoxon rank-sum test was used to deduce if the change in the TMB and SBS values is statistically significant or not. The median SBS and TMB values for the two groups are shown in **Table 3**. There was a significant change (*p*-value <0.05) for SBS T>G, C>A, and C>T. An increase was observed in the C>A and C>T substitution values, while a decrease was observed in T>G substitutions. Further, there is a statistically significant difference in the TMB values of TMB_NS, TMB_SYN, and TMB_OTH, i.e. there was a considerable increase in the tumor mutational burden of the patients with poor outcome as compared to patients with a superior outcome.

Discussion

The fundamental goal of the study was to investigate the entire spectrum of the mutations altered in MGUS and MM, thereby identifying the critical factors responsible for the progression of the disease from MGUS to MM. In this study, we have explored the nonsynonymous and synonymous variants due to their impact on protein expression and function. First of all, variants were identified using four different variant callers to reduce the false positives from the study. Our approach ensured that the variants discovered in our research are the closest possible estimation of the true variants present in the MM and MGUS patients. Variants were then categorized into three main categories-nonsynonymous (NS), synonymous (SYN), and other (OTH) variants. TMB was calculated for each of the three categories of variants. This study reveals changes in the mutational spectrum from MGUS to MM. There was a statistically significant rise in the single base substitutions as the disease progressed from MGUS to MM (**Figure 2**). The frequency pattern of the substitutions in MM is similar to what was observed in a previous study [36]. The highest rise in the frequency was observed in C>T transitions, where the median almost doubled from 30 to 59. An increase in the C>T transitions in MM can be attributed to the overexpression of A3B, an APOBEC cytidine deaminase, that has an essential part in immunity against diseases [37]. Aberrant expression of A3B is known to be correlated with drug resistance, metastasis, and poor prognosis in breast cancer [38], lung cancer [39], and ovarian cancer [40]. Yamazaki et al. [37] proposed that A3B may promote disease progression and drug resistance in MM, which validates our observation of the hike in C>T transitions from MGUS to MM. The association of the frequency of substitutions in the MM patients and their survival outcome was further explored. Frequency of C>T, C>A, and T>G substitutions were significantly higher in MM patients with poor overall survival outcome as compared to MM patients with superior overall survival outcome (**Table 3**). However, in multivariate Cox Hazard analysis (Supplementary Table 5), only C>A transitions have a statistically significant impact on the survival outcome of MM patients.

In addition, SBS2 and SBS13 mutational signatures are linked to APOBEC activity reported in MM in multiple studies [41, 42]. APOBEC signatures were found in nearly 9.63% (98/1018) of the total MM patients, while they were primarily absent in MGUS patients (present in only 1 out of 61 MGUS patients). This finding suggests that ABOPEC activity may be responsible for the molecular mechanisms driving tumor progression from MGUS to MM. The association of ABOPEC activity with overall and progression-free survival in MM was also explored. There was a statistically significant association between the ABOPEC activity and poor overall survival in MM (p-value =0.0056). The KM survival analysis validated this, which yielded significant separation (p-value = 1.8e-4) in the OS curves of MM patients with and without APOBEC activity. Contrary to these findings, no significant association was found between PFS and APOBEC activity. Further, signatures SBS6, SBS14, SSB20, SBS21, SBS26 were found only in MM and are associated with defective DNA mismatch repair and microsatellite instability (MSI) as described previously. MSI has been reported in Multiple Myeloma [43]. However, its frequency is low (~10%) [44]. MSI has been observed to be an effective indicator of response to immunotherapy in solid tumors [45], like colorectal carcinoma [46]. Therefore, it is vital to look for these signatures in MM to help identify the high-risk MM patients in need of immunotherapy.

In the present study, synonymous mutations have been examined along with nonsynonymous mutations. Though synonymous mutations do not change the amino acid sequence of the resulting protein, they have a profound influence on RNA stability, RNA folding [19] or splicing [20], translation [21], or co-translational protein folding. Hence, their role in cancer progression cannot be ignored. There are three different variants categorized under synonymous-synonymous snvs, 3'UTRs and 5'UTRs. A statistically significant rise in the 3'UTR (p-value =2e-16) and 5'UTR (p-value =2.7e-7) mutations were observed from MGUS and MM. 3' untranslated region (UTR) are a part of mRNA containing regulatory binding sites that post-transcriptionally influence gene expression and may lead to disruption in critical pathways associated with different types of cancers. Multiple studies have demonstrated that 3'UTR variants are linked to the risk of developing tumor or tumor progression. Zhang

et al. [47] discovered that a polymorphism detected in the IL-1α 3'UTR of the miRNA-122-binding site was associated with the risk of epithelial ovarian cancer. A unique variant located in the 3'UTR was identified in the gene PCM1, which was significantly associated with ovarian cancer [48]. Recently, Melaiu et al. [49] evaluated the significance of germline genetic variants located within the 3'-untranslated region (polymorphic 3'UTR, i.e., p3UTR) of candidate genes involved in multiple myeloma. Their findings suggested that *IL10*-rs3024496 was associated with an increased risk of developing MM and worse overall survival in MM patients. They also observed that *IL10*-rs3024496 SNP might regulate the *IL10* mRNA expression and hence, could help in the stratification of MM patients in terms of risk progression and prognosis.

5'UTR regions are a part of mRNA, which regulates the protein expression by controlling the translation initiation. Hence, single nucleotide polymorphisms (SNPs) located at 5'UTR regions may alter the protein levels by regulating the mRNA translation efficiency, thereby disturbing consequential biological pathways. The role of 5'UTR variants in multiple cancers has been explored in previous studies. A 5'UTR variant was the driving factor leading to familial breast and ovarian cancer in two independent families [50]. 5'UTR SNP in the PLA2G2A gene was associated with PC metastasis [51]. Thus, it can be concluded that 3' and 5'UTR mutations are more frequent in MM and drive MGUS to MM via regulatory binding sites.

TMB has become a prominent biomarker of enhanced responsiveness to immunotherapy and better outcomes. High TMB is often associated with longer survival after treatment with immune checkpoint inhibitors (ICIs) [16]. However, in non-ICI-treated patients, high TMB was associated with poor prognosis and overall survival in many cancer types [17]. Correlation of high TMB with response to targeted immunotherapies has been established in solid tumors [52, 53]. High somatic mutation and neoantigen loads have been correlated with reduced PFS in MM [54]. However, the association of TMB with overall survival is still unknown in newly diagnosed multiple myeloma (NDMM) patients. Patients with very high TMB_NS values were further analyzed to examine the relation of TMB with OS. These are known as hypermutators and are high-risk patients. Hypermutators demonstrated a significant poor overall survival (p-value =0.022) and poor progression-free survival (p-value =0.045) as compared to non-hypermutators (TMB_NS≤ 35) (**Figure 6**). The median overall survival of hypermutators was 220 weeks compared to 316 weeks of non-hypermutators, while the median progression-free survival of hypermutators was 105 weeks compared to 143.3 weeks non-hypermutators. Mutational signatures SBS1, SBS5, and SBS54 were observed in hypermutators and death events in 7 out of 10 hypermutators. DBS4, DBS5, DBS9, DBS10, and DBS11 are the mutational signatures reflective of double base substitutions (DBS) and were found to be present in hypermutators. On the contrary, no DBS signatures were found in low TMB patients (TMB_NS< 0.1; n=6). SBS1 and SBS5 were present in low TMB patients, including SBS7a, SBS17b, SBS27, SBS51, and SBS86. Our study establishes that the frequency of hypermutators is low in the MM population, and hypermutators are associated with poor OS and poor PFS outcome. Since TMB is a predictor of enhanced responsiveness to immunotherapy, hypermutators may be treated with immunotherapy drugs such as Daratumumab/Elotuzumab [55], Isatuximab [56], and Belantamab Mafodotin [57] to improve their overall survival.

In conclusion, the present study reveals the factors responsible for disease progression from MGUS to MM and poor survival outcome in MM via a detailed investigation of the mutations present in MGUS and MM. The entire landscape of the mutational spectrum involving both synonymous and nonsynonymous mutations was examined. This study finds a change in the mutational spectrum with a statistically significant increase from MGUS to MM. There was a statistically significant increase in the frequency of all the three categories of variants-non-synonymous, synonymous, and others from MGUS to MM (P<0.05). However, there was a statistically significant rise in the TMB values for TMB_NS and TMB_SYN only. We also observed that 3' and 5'UTR mutations were more frequent in MM and might be responsible for driving MGUS to MM via regulatory binding sites. In addition, NDMM patients were also examined separately along with their survival outcome. 10 out of 832 NDMM patients had TMB_NS values greater

than 35 and were designated as hypermutators. It could be concluded that the frequency of hypermutators was low in MM with poor OS and PFS outcome. We also observed a statistically significant rise in the frequency of C>A and C>T substitutions and a statistically significant decline in T>G substitutions. There was a statistically significant increase in the tumor mutational burden of the patients with poor outcome as compared to patients with a superior outcome. Further, a statistically significant association between the APOBEC activity and poor overall survival in MM was discovered. A limitation of the current study is that the number of MGUS patients is significantly less than the number of MM patients. Comparison with a larger dataset of MGUS patients can substantiate the study findings of the significant increase in the mutational frequencies from MGUS to MM. A coherent analysis of evolving mutational landscapes and cancer signatures could assist in designing therapies to impede the transformation of benign MGUS to malignant MM. Additionally, a systematized comparison of high-risk MM patients with low-risk MM patients can aid in identifying the risk factors responsible for disease progression and ultimately guide towards a personalized cure, thereby improving the overall survival of MM patients. A significant rise in 3' and 5'UTR mutations from MGUS to MM was observed in our study. A detailed investigation of these mutations might help understand the mechanism of the progression of MGUS to internecine MM and may be explored in future studies.

Availability of data and materials

Mutation data of 936 MM patients was obtained via dbGaP (MMRF CoMMpass study; phs000748; phs000348), while the remaining 82 patients' exome data was obtained from AIIMS. Exome data of 33 MGUS patients out of 61 patients was obtained from EGA (EGAD-00001001901), and data of the remaining 28 patients was obtained from AIIMS. Variant files of MM patients from the MMRF CoMMpass study were downloaded from the GDC portal via dbGaP authorized access.

Acknowledgements

This work was supported by a grant from the Department of Biotechnology, Govt. of India [Grant: BT/MED/30/SP11006/2015] and Department of Science and Technology, Govt. of India [Grant: DST/ICPS/CPS-Individual/2018/279(G)]. Authors acknowledge dbGaP (Project #18964) for providing authorized access to the MM datasets (phs000748 and phs000-348). Funding support for the phs000348 study was provided by the Multiple Myeloma Research Foundation in collaboration with the Multiple Myeloma Research Consortium. Assistance with data generation, processing, and analysis was provided by the Broad Institute Genome Sequencing, Genetic Analysis, and Biological Samples Platforms. The datasets used for the analyses described in this work were obtained from dbGaP through dbGaP accession number phs000348.v1.p1. Data of study phs000748 were generated as part of the Multiple Myeloma Research Foundation CoMMpass [SM] (Relating Clinical Outcomes in MM to Personal Assessment of Genetic Profile) study (www.themmrf.org). We also acknowledge EGA (EGAD00001001901) for providing authorized access to the MGUS data. The authors would also like to thank the Centre of Excellence in Healthcare, IIIT-Delhi for support in their research. This work was supported by a grant from the Department of Biotechnology, Govt. of India [Grant: BT/MED/30/SP11006/2015] and the Department of Science and Technology, Govt. of India [Grant: DST/ICPS/CPS-Individual/2018/279(G)]. The funding bodies had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data used in the study and had final responsibility for the decision to submit for publication.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Ritu Gupta, Laboratory Oncology Unit, Dr. B.R.A. IRCH, AIIMS, New Delhi 110029, India. E-mail: drritugupta@gmail.com; drritu.laboncology@aiims.edu; Dr. Anubha Gupta, SBILab, Department of ECE, IIIT-Delhi, Core Member, Centre of Excellence in Healthcare, IIIT-Delhi Member, Infosys Centre for AI, IIIT-Delhi, New Delhi 110020, India. E-mail: anubha@iiitd.ac.in

References

[1] Kyle RA, Therneau TM, Rajkumar SV, Larson DR, Plevak MF, Offord JR, Dispenzieri A,

Katzmann JA and Melton LJ 3rd. Prevalence of monoclonal gammopathy of undetermined significance. N Engl J Med 2006; 354: 1362-1369.

[2] Kyle RA, Therneau TM, Rajkumar SV, Offord JR, Larson DR, Plevak MF and Melton LJ 3rd. A long-term study of prognosis in monoclonal gammopathy of undetermined significance. N Engl J Med 2002; 346: 564-569.

[3] Manier S, Salem K, Glavey SV, Roccaro AM and Ghobrial IM. Genomic Aberrations in Multiple Myeloma. Cancer Treat Res 2016; 169: 23-34.

[4] Chapman MA, Lawrence MS, Keats JJ, Cibulskis K, Sougnez C, Schinzel AC, Harview CL, Brunet JP, Ahmann GJ, Adli M, Anderson KC, Ardlie KG, Auclair D, Baker A, Bergsagel PL, Bernstein BE, Drier Y, Fonseca R, Gabriel SB, Hofmeister CC, Jagannath S, Jakubowiak AJ, Krishnan A, Levy J, Liefeld T, Lonial S, Mahan S, Mfuko B, Monti S, Perkins LM, Onofrio R, Pugh TJ, Rajkumar SV, Ramos AH, Siegel DS, Sivachenko A, Stewart AK, Trudel S, Vij R, Voet D, Winckler W, Zimmerman T, Carpten J, Trent J, Hahn WC, Garraway LA, Meyerson M, Lander ES, Getz G and Golub TR. Initial genome sequencing and analysis of multiple myeloma. Nature 2011; 471: 467-472.

[5] Fonseca R, Blood EA, Oken MM, Kyle RA, Dewald GW, Bailey RJ, Van Wier SA, Henderson KJ, Hoyer JD, Harrington D, Kay NE, Van Ness B and Greipp PR. Myeloma and the t(11;14)(q13;q32); evidence for a biologically defined unique subset of patients. Blood 2002; 99: 3735-3741.

[6] Walker BA, Wardell CP, Melchor L, Brioli A, Johnson DC, Kaiser MF, Mirabella F, Lopez-Corral L, Humphray S, Murray L, Ross M, Bentley D, Gutiérrez NC, Garcia-Sanz R, San Miguel J, Davies FE, Gonzalez D and Morgan GJ. Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. Leukemia 2014; 28: 384-390.

[7] Dhodapkar MV. MGUS to myeloma: a mysterious gammopathy of underexplored significance. Blood 2016; 128: 2599-2606.

[8] Dutta AK, Fink JL, Grady JP, Morgan GJ, Mullighan CG, To LB, Hewett DR and Zannettino ACW. Subclonal evolution in disease progression from MGUS/SMM to multiple myeloma is characterised by clonal stability. Leukemia 2019; 33: 457-468.

[9] Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, Jones D, Lin ML, Beare D, Lau KW, Greenman C, Varela I, Nik-Zainal S, Davies HR, Ordoñez GR, Mudie LJ, Latimer C, Edkins S, Stebbings L, Chen L, Jia M, Leroy C, Marshall J, Menzies A, Butler A, Teague JW, Mangion J, Sun YA, McLaughlin SF, Peckham HE, Tsung EF, Costa GL, Lee CC, Minna JD, Gazdar A, Birney E, Rhodes MD, McKernan KJ, Stratton MR, Futreal PA and Campbell PJ. A small-cell lung cancer genome with complex signatures of tobacco exposure. Nature. 2010; 463: 184-190.

[10] Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, Edkins S, O'Meara S, Vastrik I, Schmidt EE, Avis T, Barthorpe S, Bhamra G, Buck G, Choudhury B, Clements J, Cole J, Dicks E, Forbes S, Gray K, Halliday K, Harrison R, Hills K, Hinton J, Jenkinson A, Jones D, Menzies A, Mironenko T, Perry J, Raine K, Richardson D, Shepherd R, Small A, Tofts C, Varian J, Webb T, West S, Widaa S, Yates A, Cahill DP, Louis DN, Goldstraw P, Nicholson AG, Brasseur F, Looijenga L, Weber BL, Chiew YE, DeFazio A, Greaves MF, Green AR, Campbell P, Birney E, Easton DF, Chenevix-Trench G, Tan MH, Khoo SK, Teh BT, Yuen ST, Leung SY, Wooster R, Futreal PA and Stratton MR. Patterns of somatic mutation in human cancer genomes. Nature 2007; 446: 153-158.

[11] Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, Greenman CD, Varela I, Lin ML, Ordóñez GR, Bignell GR, Ye K, Alipaz J, Bauer MJ, Beare D, Butler A, Carter RJ, Chen L, Cox AJ, Edkins S, Kokko-Gonzales PI, Gormley NA, Grocock RJ, Haudenschild CD, Hims MM, James T, Jia M, Kingsbury Z, Leroy C, Marshall J, Menzies A, Mudie LJ, Ning Z, Royce T, Schulz-Trieglaff OB, Spiridou A, Stebbings LA, Szajkowski L, Teague J, Williamson D, Chin L, Ross MT, Campbell PJ, Bentley DR, Futreal PA and Stratton MR. A comprehensive catalogue of somatic mutations from a human cancer genome. Nature 2010; 463: 191-196.

[12] Kasar S, Kim J, Improgo R, Tiao G, Polak P, Haradhvala N, Lawrence MS, Kiezun A, Fernandes SM, Bahl S, Sougnez C, Gabriel S, Lander ES, Kim HT, Getz G and Brown JR. Whole-genome sequencing reveals activation-induced cytidine deaminase signatures during indolent chronic lymphocytic leukaemia evolution. Nat Commun 2015; 6: 1-12.

[13] Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA, Menzies A, Martin S, Leung K, Chen L, Leroy C, Ramakrishna M, Rance R, Lau KW, Mudie LJ, Varela I, McBride DJ, Bignell GR, Cooke SL, Shlien A, Gamble J, Whitmore I, Maddison M, Tarpey PS, Davies HR, Papaemmanuil E, Stephens PJ, McLaren S, Butler AP, Teague JW, Jönsson G, Garber JE, Silver D, Miron P, Fatima A, Boyault S, Langerød A, Tutt A, Martens JW, Aparicio SA, Borg Å, Salomon AV, Thomas G, Børresen-Dale AL, Richardson AL, Neuberger MS, Futreal PA, Camp-

bell PJ and Stratton MR; Breast Cancer Working Group of the International Cancer Genome Consortium. Mutational processes molding the genomes of 21 breast cancers. Cell 2012; 149: 979-993.
[14] Davies H, Morganella S, Purdie CA, Jang SJ, Borgen E, Russnes H, Glodzik D, Zou X, Viari A, Richardson AL, Børresen-Dale AL, Thompson A, Eyfjord JE, Kong G, Stratton MR and Nik-Zainal S. Whole-genome sequencing reveals breast cancers with mismatch repair deficiency. Cancer Res 2017; 77: 4755-4762.
[15] Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, Campbell PJ, Vineis P, Phillips DH and Stratton MR. Mutational signatures associated with tobacco smoking in human cancer. Science 2016; 354: 618-622.
[16] Fusco MJ, West HJ and Walko CM. Tumor mutation burden and cancer treatment. JAMA Oncol 2021; 7: 316.
[17] Valero C, Lee M, Hoen D, Wang J, Nadeem Z, Patel N, Postow MA, Shoushtari AN, Plitas G, Balachandran VP, Smith JJ, Crago AM, Long Roche KC, Kelly DW, Samstein RM, Rana S, Ganly I, Wong RJ, Hakimi AA, Berger MF, Zehir A, Solit DB, Ladanyi M, Riaz N, Chan TA, Seshan VE and Morris LGT. The association between tumor mutational burden and prognosis is dependent on treatment context. Nat Genet 2021; 53: 11-15.
[18] Sharma Y, Miladi M, Dukare S, Boulay K, Caudron-Herger M, Groß M, Backofen R and Diederichs S. A pan-cancer analysis of synonymous mutations. Nat Commun 2019; 10: 1-14.
[19] Goodman DB, Church GM and Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science 2013; 342: 475-479.
[20] Parmley JL, Chamary JV and Hurst LD. Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol 2006; 23: 301-309.
[21] Drummond DA and Wilke CO. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 2008; 134: 341-352.
[22] Supek F, Skunca N, Repar J, Vlahovicek K and Smuc T. Translational selection is ubiquitous in prokaryotes. PLoS Genet 2010; 6: e1001004.
[23] Savisaar R and Hurst LD. Exonic splice regulation imposes strong selection at synonymous sites. Genome Res 2018; 28: 1442-1454.
[24] Kimura M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 1977; 267: 275-276.
[25] Fan Y, Xi L, Hughes DS, Zhang J, Zhang J, Futreal PA, Wheeler DA and Wang W. MuSE: accounting for tumor heterogeneity using a sample-specific error model improves sensitivity and specificity in mutation calling from sequencing data. Genome Biol 2016; 17: 1-11.
[26] Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C and Lichtenstein L. Calling somatic SNVs and indels with Mutect2. BioRxiv 2019; 861054.
[27] Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L and Wilson RK. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012; 22: 568-576.
[28] Larson DE, Harris CC, Chen K, Koboldt DC, Abbott TE, Dooling DJ, Ley TJ, Mardis ER, Wilson RK and Ding L. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics 2012; 28: 311-317.
[29] Wang K and Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38: e164.
[30] Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR and Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 2018; 34: 511-513.
[31] Islam SA, Wu Y, Díaz-Gay M, Bergstrom EN, He Y, Barnes M, Vella M, Wang J, Teague JW, Clapham P and Moody S. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. BioRxiv 2021;2020-12.
[32] Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, Islam SMA, Lopez-Bigas N, Klimczak LJ, McPherson JR, Morganella S, Sabarinathan R, Wheeler DA and Mustonen V; PCAWG Mutational Signatures Working Group, Getz G, Rozen SG and Stratton MR; PCAWG Consortium. The repertoire of mutational signatures in human cancer. Nature 2020; 578: 94-101.
[33] Farswan A, Jena L, Kaur G, Gupta A, Gupta R, Rani L, Sharma A and Kumar L. Branching clonal evolution patterns predominate mutational landscape in multiple myeloma. Am J Cancer Res 2021; 11: 5659-5679.
[34] Eo SH, Kang HJ, Hong SM and Cho H. K-adaptive partitioning for survival data, with an application to cancer staging. arXiv preprint arXiv:1306.4615. 2013.
[35] Budczies J, Klauschen F, Sinn BV, Győrffy B, Schmitt WD, Darb-Esfahani S and Denkert C. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. PLoS One 2012; 7: e51862.

[36] Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, Dawson KJ, Iorio F, Nik-Zainal S, Bignell GR, Hinton JW, Li Y, Tubio JM, McLaren S, O'Meara S, Butler AP, Teague JW, Mudie L, Anderson E, Rashid N, Tai YT, Shammas MA, Sperling AS, Fulciniti M, Richardson PG, Parmigiani G, Magrangeas F, Minvielle S, Moreau P, Attal M, Facon T, Futreal PA, Anderson KC, Campbell PJ and Munshi NC. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. Nat Commun 2014; 5: 1-13.

[37] Yamazaki H, Shirakawa K, Matsumoto T, Hirabayashi S, Murakawa Y, Kobayashi M, Sarca AD, Kazuma Y, Matsui H, Maruyama W, Fukuda H, Shirakawa R, Shindo K, Ri M, Iida S and Takaori-Kondo A. Endogenous APOBEC3B overexpression constitutively generates DNA substitutions and deletions in myeloma cells. Sci Rep 2019; 9: 1-14.

[38] Law EK, Sieuwerts AM, LaPara K, Leonard B, Starrett GJ, Molan AM, Temiz NA, Vogel RI, Meijer-van Gelder ME, Sweep FC, Span PN, Foekens JA, Martens JW, Yee D and Harris RS. The DNA cytosine deaminase APOBEC3B promotes tamoxifen resistance in ER-positive breast cancer. Sci Adv 2016; 2: e1601737.

[39] Yan S, He F, Gao B, Wu H, Li M, Huang L, Liang J, Wu Q and Li Y. Increased APOBEC3B predicts worse outcomes in lung cancer: a comprehensive retrospective study. J Cancer 2016; 7:618-625.

[40] Du Y, Tao X, Wu J, Yu H, Yu Y and Zhao H. APOBEC3B up-regulation independently predicts ovarian cancer prognosis: a cohort study. Cancer Cell Int 2018; 18: 1-10.

[41] Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, Proszek PZ, Melchor L, Pawlyn C, Kaiser MF, Johnson DC, Qiang YW, Jones JR, Cairns DA, Gregory WM, Owen RG, Cook G, Drayson MT, Jackson GH, Davies FE and Morgan GJ. APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. Nat Commun 2015; 6: 1-11.

[42] Hoang PH, Cornish AJ, Dobbins SE, Kaiser M and Houlston RS. Mutational processes contributing to the development of multiple myeloma. Blood Cancer J 2019; 9: 1-11.

[43] Timurağaoğlu A, Demircin S, Dizlek S, Alanoğlu G and Kiriş E. Microsatellite instability is a common finding in multiple myeloma. Clin Lymphoma Myeloma 2009; 9: 371-374.

[44] Miyashita K, Fujii K, Suehiro Y, Taguchi K, Uike N, Yoshida MA and Oda S. Heterochronous occurrence of microsatellite instability in multiple myeloma - an implication for a role of defective DNA mismatch repair in myelomagenesis. Leuk Lymphoma 2018; 59: 2454-2459.

[45] Chang L, Chang M, Chang HM, Chang F. Microsatellite Instability: A Predictive Biomarker for Cancer Immunotherapy. Appl Immunohistochem Mol Morphol 2018; 26: e15-e21.

[46] Oh DY, Venook AP and Fong L. On the Verge: Immunotherapy for Colorectal Carcinoma. J Natl Compr Canc Netw 2015; 13: 970-978.

[47] Zhang Z, Zhou B, Gao Q, Wu Y, Zhang K, Pu Y, Song Y, Zhang L and Xi M. A polymorphism at miRNA-122-binding site in the IL-1α 3'UTR is associated with risk of epithelial ovarian cancer. Fam Cancer 2014; 13: 595-601.

[48] Chen X, Paranjape T, Stahlhut C, McVeigh T, Keane F, Nallur S, Miller N, Kerin M, Deng Y, Yao X, Zhao H, Weidhaas JB and Slack FJ. Targeted resequencing of the microRNAome and 3'UTRome reveals functional germline DNA variants with altered prevalence in epithelial ovarian cancer. Oncogene 2015; 34: 2125-2137.

[49] Melaiu O, Macauda A, Sainz J, Calvetti D, Facioni MS, Maccari G, Ter Horst R, Netea MG, Li Y, Grząśko N, Moreno V, Jurczyszyn A, Jerez A, Watek M, Varkonyi J, Garcia-Sanz R, Kruszewski M, Dudziński M, Kadar K, Jacobsen SEH, Mazur G, Andersen V, Rybicka M, Zawirska D, Raźny M, Zaucha JM, Ostrovsky O, Iskierka-Jazdzewska E, Reis RM, Stępień A, Beider K, Nagler A, Druzd-Sitek A, Marques H, Martìnez-Lopez J, Lesueur F, Avet-Loiseau H, Vangsted AJ, Krawczyk-Kulis M, Butrym A, Jamroziak K, Dumontet C, Vogel U, Rymko M, Pelosini M, Subocz E, Szombath G, Sarasquete ME, Silvestri R, Morani F, Landi S, Campa D, Canzian F and Gemignani F. Common gene variants within 3'-untranslated regions as modulators of multiple myeloma risk and survival. Int J Cancer 2021; 148: 1887-1894.

[50] Evans DGR, van Veen EM, Byers HJ, Wallace AJ, Ellingford JM, Beaman G, Santoyo-Lopez J, Aitman TJ, Eccles DM, Lalloo FI, Smith MJ and Newman WG. A dominantly inherited 5'UTR variant causing methylation-associated silencing of BRCA1 as a cause of breast and ovarian cancer. Am J Hum Genet 2018; 103: 213-220.

[51] Ozturk K, Onal MS, Efiloglu O, Nikerel E, Yildirim A and Telci D. Association of 5'UTR polymorphism of secretory phospholipase A2 group IIA (PLA2G2A) gene with prostate cancer metastasis. Gene 2020; 742: 144589.

[52] Halbert B and Einstein DJ. Hot or Not: Tumor mutational Burden (TMB) as a biomarker of immunotherapy response in genitourinary cancers. Urology 2021; 147: 119-126.

[53] Boumber Y. Tumor mutational burden (TMB) as a biomarker of response to immunotherapy in small cell lung cancer. J Thorac Dis 2018; 10: 4689-4693

[54] Miller A, Cattaneo L, Asmann YW, Braggio E, Keats JJ, Auclair D, Lonial S, Network TM, Russell SJ and Stewart AK. Correlation between somatic mutation burden, neoantigen load and progression free Survival in multiple myeloma: analysis of MMRF CoMMpass study. Blood 2016; 128: 193.

[55] Laubach JP, Paba Prada CE, Richardson PG and Longo DL. Daratumumab, elotuzumab, and the development of therapeutic monoclonal antibodies in multiple myeloma. Clin Pharmacol Ther 2017; 101: 81-88.

[56] Moreno L, Perez C, Zabaleta A, Manrique I, Alignani D, Ajona D, Blanco L, Lasa M, Maiso P, Rodriguez I, Garate S, Jelinek T, Segura V, Moreno C, Merino J, Rodriguez-Otero P, Panizo C, Prosper F, San-Miguel JF and Paiva B. The mechanism of action of the anti-CD38 monoclonal antibody isatuximab in multiple myeloma. Clin Cancer Res 2019; 25: 3176-3187.

[57] Offidani M, Corvatta L, Morè S and Olivieri A. Belantamab mafodotin for the treatment of multiple myeloma: an overview of the clinical efficacy and safety. Drug Des Devel Ther 2021; 15: 2401-2415.

Supplementary Table 1. The table shows the cut-offs obtained for TMB_OTH via KAP

	Min	Median	Max	Cut-off (≤, >)	PFS	OS
TMB_OTH	0.1114	1.3742	193.673	1.84 (666, 166)	4.90E-06	9.16E-09

The same cut-off was obtained using PFS and OS. KM analysis was done using the proposed cut-offs. There was a significant difference (*p*-value <0.05) on the KM survival curves of the patients below and above the selected cut-offs of TMB_OTH.

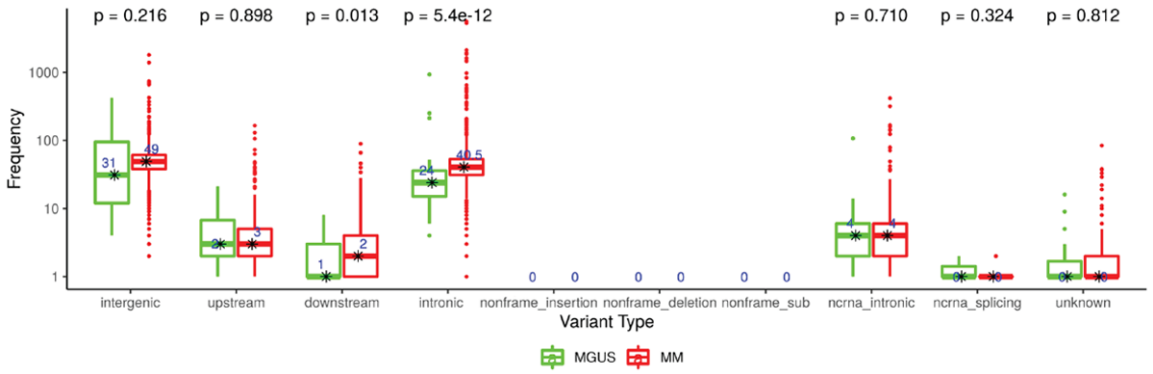
Supplementary Table 2. The table shows the cut-offs obtained for TMB_NS, TMB_SYN and TMB_OTH via Cutoff Finder

	Cut-off via PFS	*p*-value	Cut-off via OS	*p*-value
TMB_NS	0.6282	3.2E-07	0.6216	2.1E-08
TMB_SYN	0.5565	4.1E-05	0.5265	2.3E-09
TMB_OTH	1.84	4.9E-06	1.84	9.2E-09

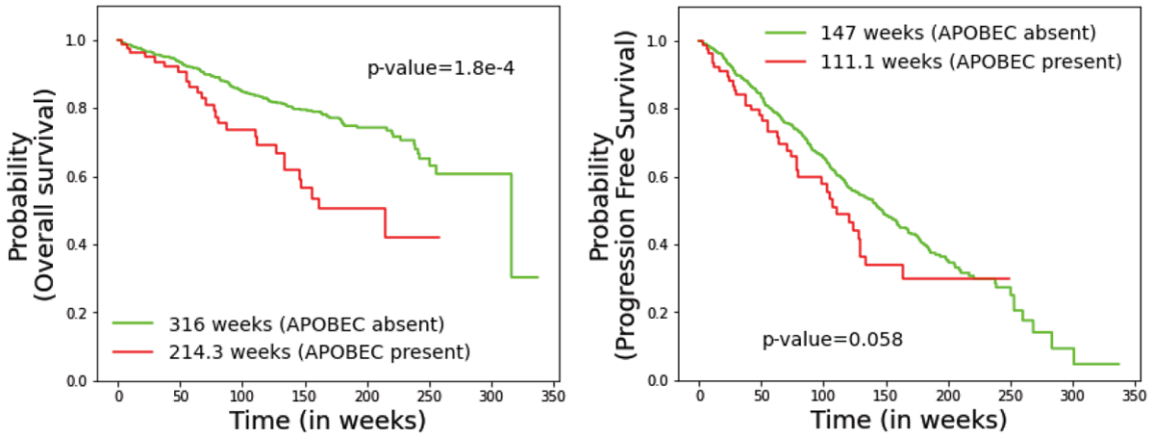
Similar cut-offs were obtained using PFS and OS. KM analysis revealed a significant difference (*p*-value <0.05) on the survival curves of the patients below and above the selected cut-offs.

Supplementary Table 3. The table shows the cut-offs obtained for the six different types of substitutions via Cutoff Finder

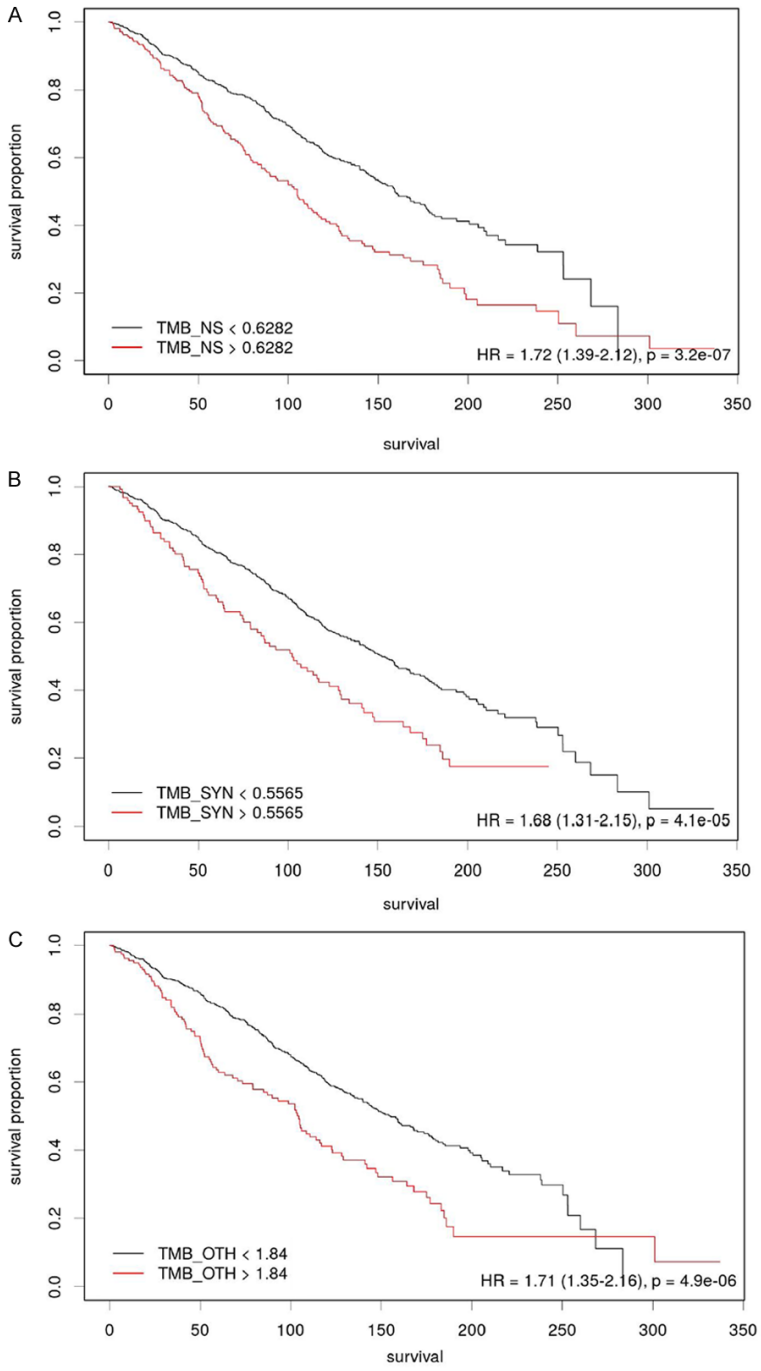
SBS	Min	Median	Max	PFS cutoff	OS cutoff	PFS *p*-value	OS *p*-value
C>A	0	17	1251	26.5	28.5	1.9E-5	5.1E-06
C>G	0	21	1575	3.5	34.5	0.027	8.6E-05
C>T	1	59	7315	79.5	110	0.00055	5.8E-07
T>A	0	17	684	5.5	2.5	0.0046	0.0024
T>C	0	35	4498	12.5	11.5	0.00019	0.0027
T>G	0	19	915	6.5	6.5	0.0013	0.0029



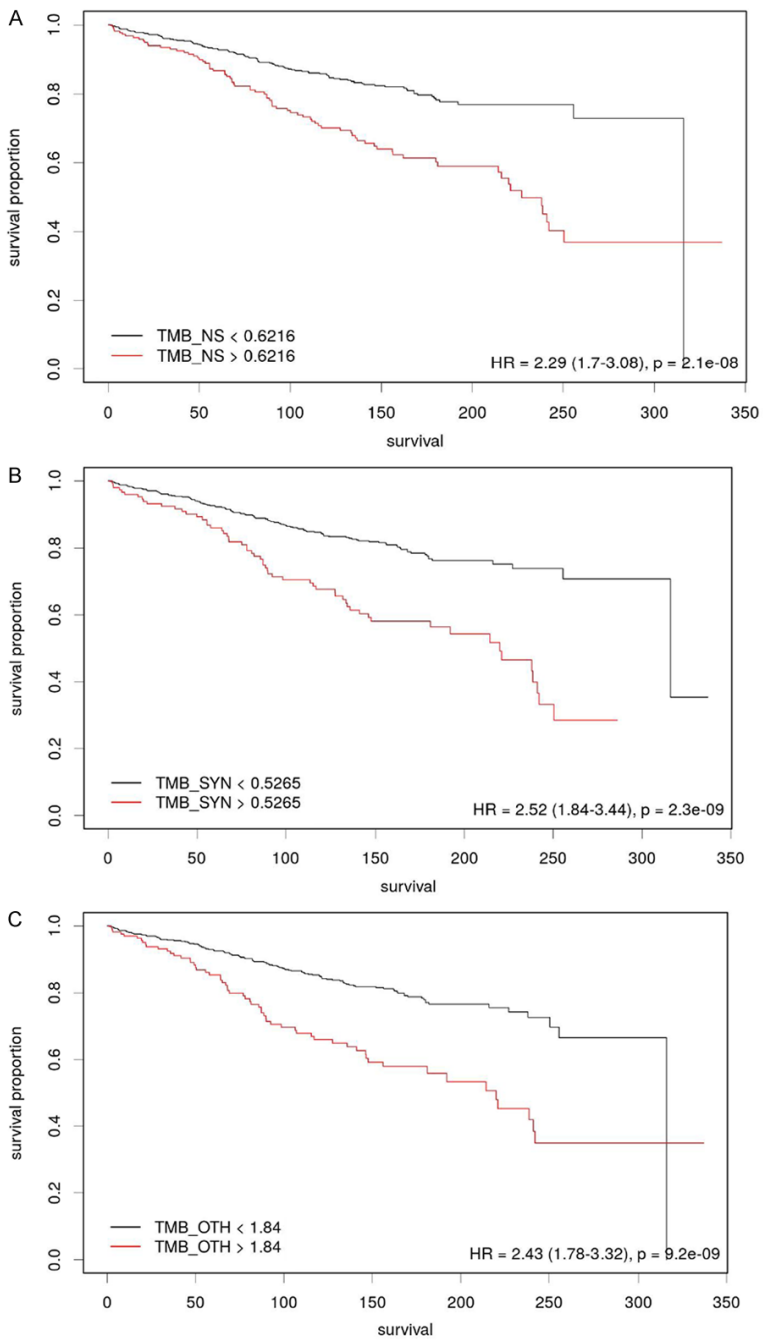
Supplementary Figure 1. Boxplot showing the variation in the frequency of the variants under the other variants category. There was a statistically significant rise in the frequency of intronic, intergenic, and downstream variants with *p*-values less than 0.05.



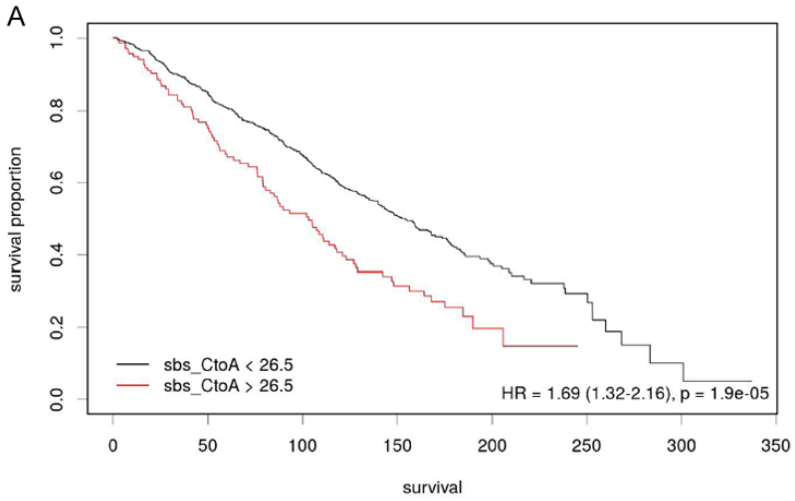
Supplementary Figure 2. KM curves reveal that APOBEC activity is associated with poor overall survival in NDMM patients. The difference in the overall survival probability between low and high TMB_NS is statistically significant with *p*-values 1.8e-4. However, there is no statistically significant difference between progression-free survival and APOBEC activity.

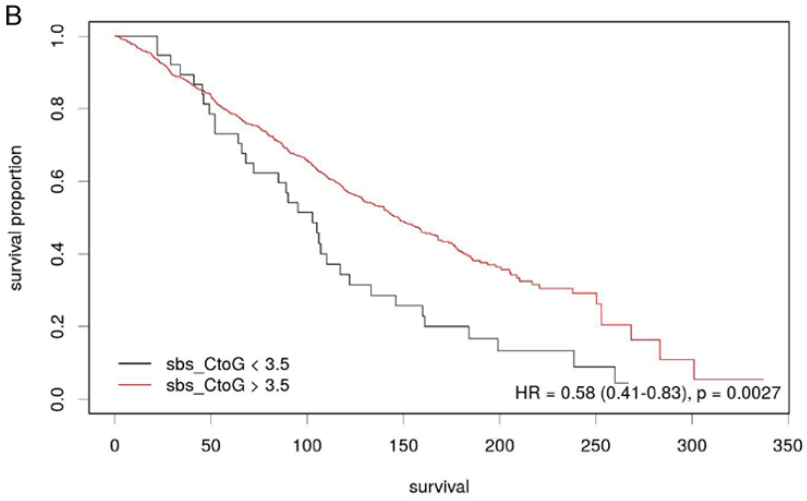


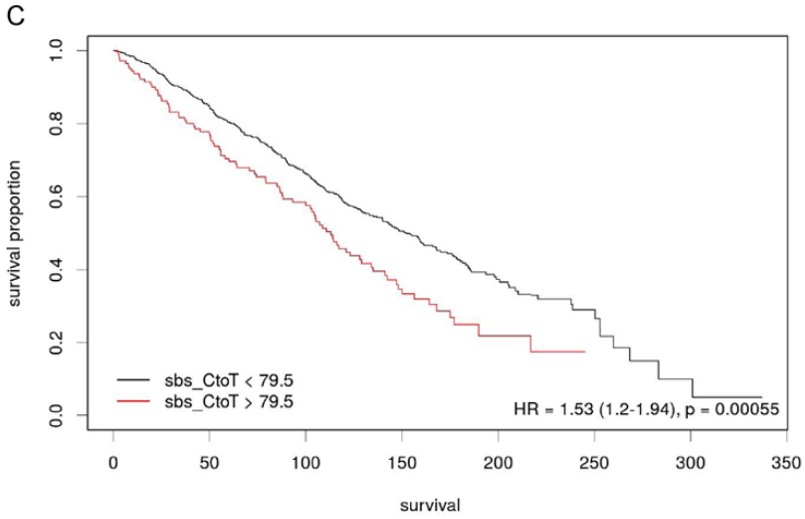
Supplementary Figure 3. KM curves reveal significant differences in the PFS survival patterns of (A) TMB_NS, (B) TMB_SYN and (C) TMB_OTH at the thresholds obtained via Cutoff Finder.

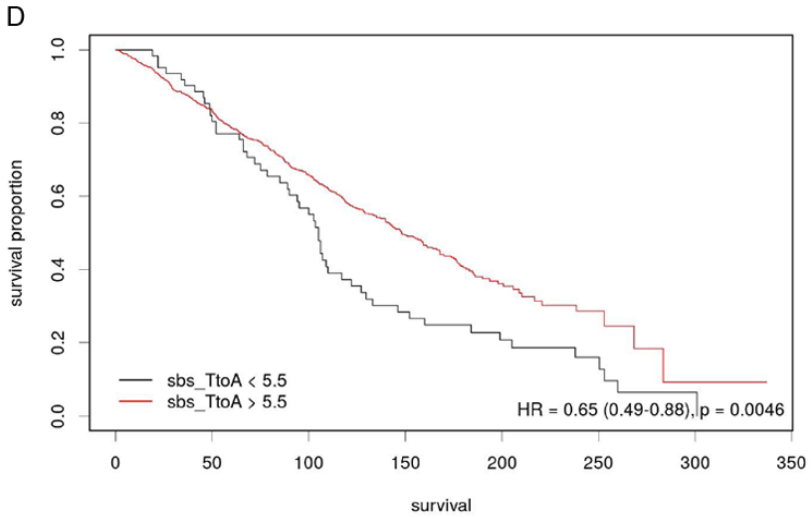


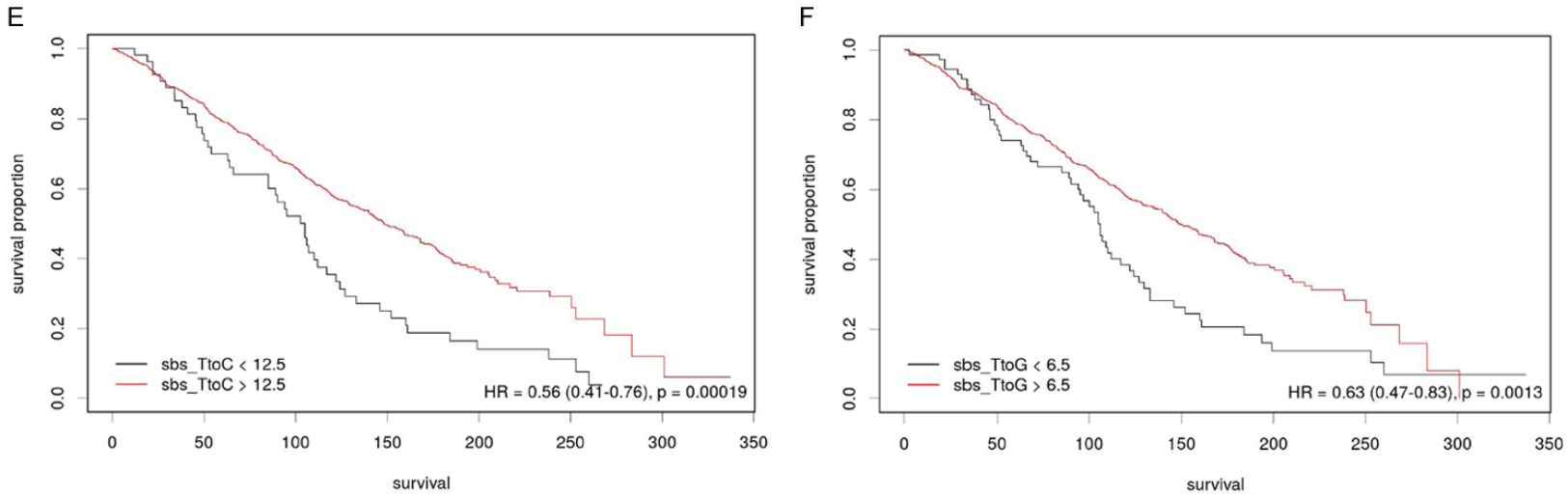
Supplementary Figure 4. KM curves reveal significant differences in the OS survival patterns of (A) TMB_NS, (B) TMB_SYN and (C) TMB_OTH at the thresholds obtained via Cutoff Finder.


A
survival proportion
1.0
0.8
0.6
0.4
0.2
0.0
sbs_CtoA < 26.5
sbs_CtoA > 26.5
HR = 1.69 (1.32-2.16), p = 1.9e-05
0
50
100
150
200
250
300
350
survival

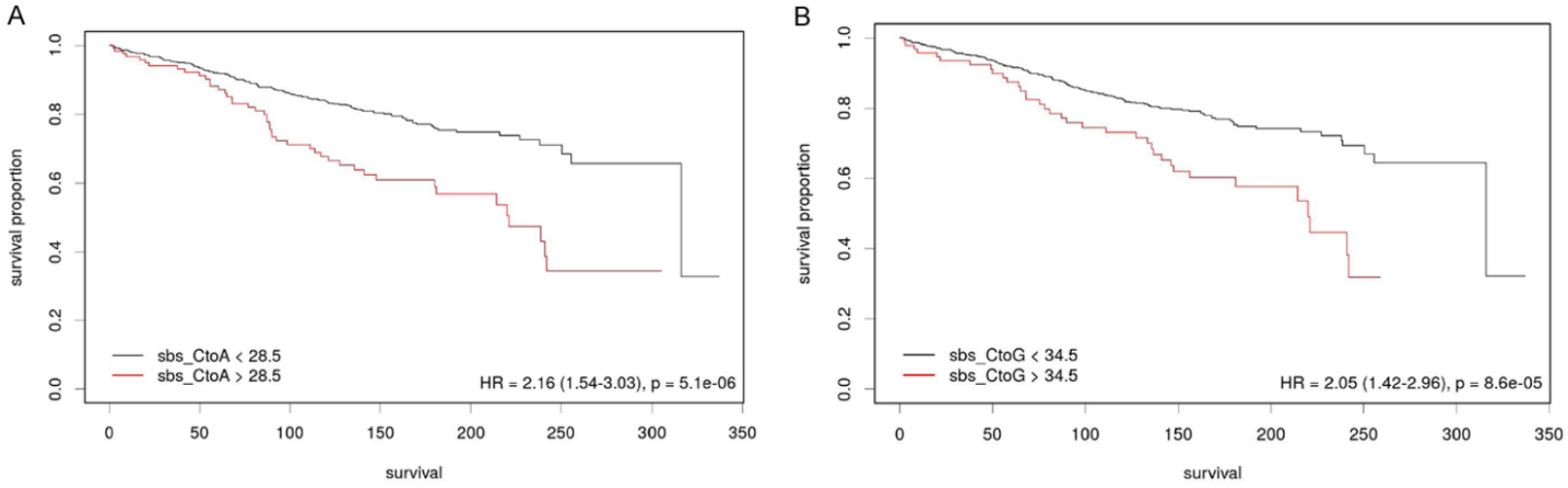

B
survival proportion
1.0
0.8
0.6
0.4
0.2
0.0
sbs_CtoG < 3.5
sbs_CtoG > 3.5
HR = 0.58 (0.41-0.83), p = 0.0027
0
50
100
150
200
250
300
350
survival

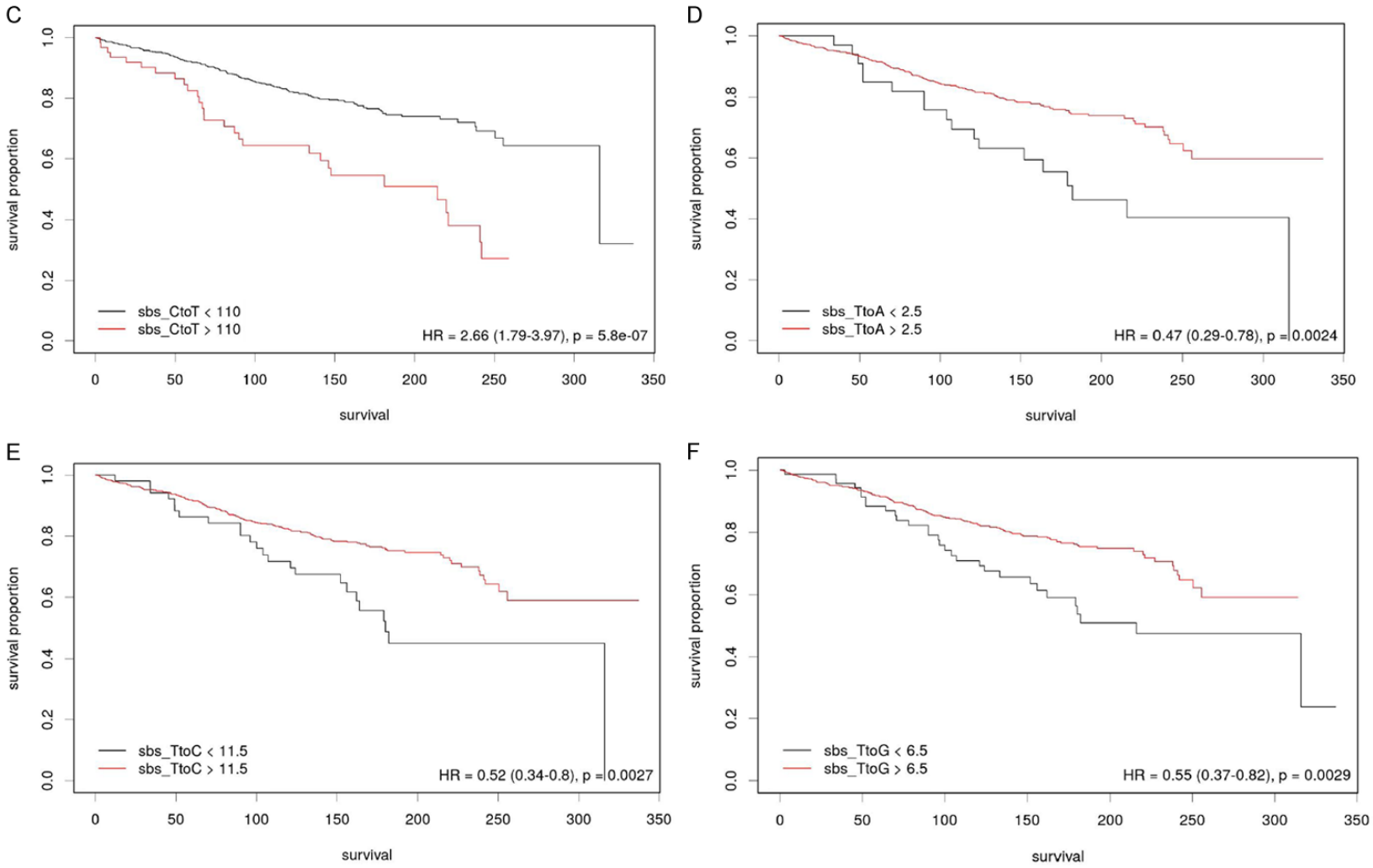

C
survival proportion
1.0
0.8
0.6
0.4
0.2
0.0
sbs_CtoT < 79.5
sbs_CtoT > 79.5
HR = 1.53 (1.2-1.94), p = 0.00055
0
50
100
150
200
250
300
350
survival


D
survival proportion
1.0
0.8
0.6
0.4
0.2
0.0
sbs_TtoA < 5.5
sbs_TtoA > 5.5
HR = 0.65 (0.49-0.88), p = 0.0046
0
50
100
150
200
250
300
350
survival



Supplementary Figure 5. KM curves reveal differences in the PFS survival patterns of substitutions (A) C>A, (B) C>G, (C) C>T, (D) T>A, (E) T>C and (F) T>G at the thresholds obtained via Cutoff Finder. Separation in the survival curves is significant if *p*-values <0.05.





Supplementary Figure 6. KM curves reveal differences in the OS survival patterns of substitutions (A) C>A, (B) C>G, (C) C>T, (D) T>A, (E) T>C and (F) T>G at the thresholds obtained via Cutoff Finder. Separation in the survival curves is significant if *p*-values <0.05.

Supplementary Table 4. The table shows the univariate hazard analysis and multivariate hazard analysis obtained on TMB_NS, TMB_SYN and TMB_OTH

	pfs				os			
	HR	CI	*p*-value	C-index	HR	CI	*p*-value	C-index
Univariate								
TMB_NS	1.71	1.39-2.12	<0.005	0.56	2.26	1.68-3.05	<0.005	0.58
TMB_SYN	1.68	1.31-2.15	<0.005	0.54	2.46	1.78-3.40	<0.005	0.56
TMB_OTH	1.71	1.35-2.16	<0.005	0.55	2.43	1.78-3.32	<0.005	0.58
Multivariate								
TMB_NS	1.45	1.11-1.90	0.01	0.57	1.55	1.04-2.31	0.03	0.6
TMB_SYN	1.13	0.81-1.58	0.48		1.41	0.89-2.24	0.14	
TMB_OTH	1.26	0.92-1.74	0.16		1.48	0.94-2.34	0.09	

Supplementary Table 5. The table shows the univariate hazard analysis and multivariate hazard analysis on the six different substitutions

	PFS				OS			
	HR	CI	*p*-value	C-index	HR	CI	*p*-value	C-index
Univariate								
C>A	1.63	1.26-2.11	<0.005	0.54	2.16	1.54-3.03	<0.005	0.55
C>G	1.46	1.04-2.04	0.03	0.52	2.11	1.41-3.16	<0.005	0.53
C>T	1.65	1.22-2.24	<0.005	0.53	2.36	1.61-3.45	<0.005	0.55
T>A	1.61	1.11-2.32	0.01	0.51	1.93	1.20-3.11	0.01	0.52
T>C	1.47	0.83-2.61	0.19	0.50	2.27	1.19-4.32	0.01	0.51
T>G	1.73	1.09-2.75	0.02	0.51	2.14	1.21-3.77	0.01	0.51
Multivariate								
C>A	1.43	1.02-1.99	0.04	0.55	1.67	1.06-2.63	0.03	0.58
C>G	0.84	0.49-1.43	0.52		0.97	0.49-1.93	0.93	
C>T	1.38	0.86-2.22	0.18		1.71	0.91-3.22	0.10	
T>A	1.19	0.71-1.97	0.65		1.22	0.61-2.44	0.58	
T>G	1.03	0.52-2.05	0.94		0.82	0.34-1.99	0.66	

T>C was removed from multivariate analysis as it was not significant for PFS in univariate analysis.