Am J Cancer Res 2021;11(11):5659-5679
www.ajcr.us /ISSN:2156-6976/ajcr0137108

*Original Article*

# Branching clonal evolution patterns predominate mutational landscape in multiple myeloma

Akanksha Farswan[1*], Lingaraja Jena[2*], Gurvinder Kaur[2*], Anubha Gupta[1], Ritu Gupta[2], Lata Rani[2], Atul Sharma[3], Lalit Kumar[3]

*[1]SBILab, Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology-Delhi (IIIT-D), Delhi 110020, India; [2]Laboratory Oncology Unit, Dr. B.R.A. IRCH, All India Institute of Medical Sciences (AIIMS), New Delhi 110029, India; [3]Department of Medical Oncology, Dr. B.R.A. IRCH, All India Institute of Medical Sciences (AIIMS), New Delhi 110029, India. [*]Equal contributors and co-first authors.*

Received July 7, 2021; Accepted September 27, 2021; Epub November 15, 2021; Published November 30, 2021

**Abstract:** Multiple Myeloma (MM) arises from malignant transformation and deregulated proliferation of clonal plasma cells (PCs) harbouring heterogeneous molecular anomalies. The effect of evolving mutations on clone fitness and their cellular prevalence shapes the progressing myeloma genome and impacts clinical outcomes. Although clonal heterogeneity in MM is well established, which subclonal mutations emerge/persist/perish with progression in MM and which of these can be targeted therapeutically remains an open question. In line with this, we have sequenced pairwise whole exomes of 62 MM patients collected at two time points, i.e., at diagnosis and on progression. Somatic variants were called using a novel ensemble approach where a consensus was deduced from four variant callers (Illumina's Dragen, Strelka2, SomaticSniper and SpeedSeq) and actionable/druggable gene targets were identified. A marked intraclonal heterogeneity was observed. Branching evolution was observed among 72.58% patients, of whom 64.51% had low TMBs (<10) and 61.29% had 2 or more founder clones. The hypermutator patients (with high TMB levels ≥10 to ≤100) showed a significant decrease in their TMBs from diagnosis (median TMB 77.11) to progression (median TMB 31.22). A distinct temporal fall in subclonal driver mutations was identified recurrently across diagnosis to progression e.g., in *PABPC1, BRAF, KRAS, CR1, DIS3* and *ATM* genes in 3 or more patients suggesting such patients could be treated early with target specific drugs like Vemurafenib/Cobimetinib. An analogous rise in driver mutations was observed in *KMT2C, FOXD4L1, SP140, NRAS* and other genes. A few drivers such as *FAT4, IGLL5* and *CDKN1A* retained consistent distribution patterns at two time points. These findings are clinically relevant and point at consideration of evaluating multi time point subclonal mutational landscapes for designing better risk stratification strategies and tailoring time to time risk adapted combination therapies in future.

**Keywords:** Multiple myeloma, NGS, exome sequencing, clonal evolution, driver genes, mutations, tumor mutation burden, progression

## Introduction

Multiple Myeloma (MM) is a malignancy of clonal plasma cells that tend to evolve and accumulate as disease progresses from precursor transition states of monoclonal gammopathy of undetermined significance (MGUS)/ Smoldering Multiple Myeloma (SMM) to active MM and ultimately Extramedullary disease/ Plasma cell leukemia (PCL). Reservoir founder clones may exist prior to MGUS [1], which may become detectable and dominant with progression and gradually evolve into heterogeneous subclones. The process of subclonal propagation of PCs during myelomagenesis is complex and is driven under the influence of selection pressures exerted by immune surveillance, microenvironment and therapeutic agents.

Molecular mechanisms that underlie early progression in newly diagnosed MM patients who fail to respond to existing treatments are not completely understood. MM shows heterogeneity in terms of clinical phenotypes, rates of disease progression, response to therapy and survival outcomes, all of which are influenced by the underlying genomic complexity of the

patient [2]. It is established that two types of primary oncogenic events are involved in initiation of myelomagenesis [3, 4]. These include IgH translocations (found in ~55% patients) and hyperdiploidy of odd numbered chromosomes (observed at a frequency of ~40%). These two kinds of aberrations may coexist in ~10% of cases. A gamut of secondary events (mutations in *RAS, NF-kB* pathway, overexpression of *MYC*, haploinsufficiency of *p53*, (1q) gain and (1p) loss) are known to occur that provide further growth advantage to evolving (sub)clones, promote drug resistance, genome instability and progression. Deletion 13q is commonly found among non-hyperdiploid MM as well as in MGUS which suggests its role as a primary event during early oncogenesis of MM [5-7].

Based on mutational complexity and subclonal architecture, different patterns of clonal evolution have been reported in MM. The branching type of clonal evolution analogous to Darwinian model is the most frequent one and is found in ≥50% MM patients whereas linear or stable evolution with no significant alteration in subclonal architecture have been observed in ≤30% cases [8-10]. Analysis of WES data obtained from MM patients on IMiDs from UK Myeloma XI phase 3 trial and the CoMMpass study has revealed that 20% MM patients experienced neutral tumor evolution associated with poor prognosis while remaining 80% encountered branching evolution [11]. Patients with branching evolution may respond well to IMiDs as these can reconfigure bone marrow stromal cum immune microenvironment and prolong survival [11]. Instead, patients with neutral clonal evolution with random genetic drift may benefit from combinations of PIs with high dose melphalan [11, 12].

Recent NGS studies conducted on pairwise myeloma genomes/exomes at two or more serial time points have reported presence of intraclonal heterogeneity during progression and relapse [1, 7, 8, 12-21]. A series of somatic mutations including substitutions, indels and copy number variations emerge during disease progression that contour the pattern of clonal evolution. Numerous driver mutations have been identified in myeloma genome [17] that may co-evolve mutually in cooperation or exclusively either in same or different (sub) clones and modulate their net impact on clinical outcomes.

Although clonal heterogeneity in MM is well established, subclonal remodelling of gains/losses and rewiring of functional pathways are not completely understood. There is currently a paucity of data available on longitudinal subclonal evolution profiles associated with progression in MM and a deeper understanding is required to assess mutations of clinical relevance that could potentially be targeted for treatment in future therapeutic approaches against MM and its precursor states [11, 22]. The progressing subclonal shifts are of paramount clinical significance as these could promote oncogenesis and lead to drug refractoriness. Estimation of their cellular prevalence could further predict likelihood of depth of response and a rationalized approach of combinatorial therapy. More and more longitudinal studies are needed to explore the progressing subclonal events and ultimately guide combinations of targeted therapy that can eradicate such subclonal populations and delay progression. Hence, we decided to conduct this study to capture subclonal mutational landscapes associated with progression of MM and identify potential actionable/druggable targets that can be treated with their corresponding drugs.

In this study, we have evaluated 186 pairwise whole exome sequences obtained from 62 MM patients at two time points representing tumor at diagnosis, tumor at progression and compared to their germline landscapes respectively using NGS. We have identified individual clonal genomic complexities, tumor mutation burdens (TMBs) and divergence of clusters of mutations in founder clones. This study has provided novel insights into recurrent subclonal shifts in drivers (DRV), oncogenes (ONC), tumor suppressor genes (TSGs) and the potential actionable targets (ACT) associated with progression of MM.

## Materials and methods

This study was approved by the Institute Ethics Committee and conducted as per ethical guidelines. Voluntary written informed consent was obtained from all the study individuals.

### *Whole exome sequencing*

Genomic DNA was isolated from CD138+ plasma cells enriched from bone marrow aspirates with MACS magnetic microbeads (Miltenyi Biotech, Germany), collected from 62 patients in-

**Table 1.** Baseline demographic, laboratory and clinical characteristics of multiple myeloma (MM) patients (n = 62)

| Parameter | No. of patients |
|---|---|
| Median Age (Range) In Years | 58 (31 to 72) |
| Gender | |
| Male | 38 |
| Female | 24 |
| Hemoglobin (g/dL) | |
| ≤10 | 39 |
| >10 | 23 |
| Platelet Count (/dL) | |
| <100 | 10 |
| ≥100 | 52 |
| Serum creatinine (mg/dL) | |
| ≤2 | 49 |
| >2 | 13 |
| Seruam albumin (g/dL) | |
| <3.5 | 30 |
| ≥3.5 | 32 |
| ISS 1/2/3 | 1/17/44 |
| RISS I/II/III/NA | 1/36/14/11 |
| MRS 1/2/3/NA | 7/33/21/1 |
| Serum calcium, mg/dL | |
| 0-11 | 54 |
| >11 | 8 |
| eGFR, mL/min | |
| <40 | 17 |
| ≥40 | 45 |
| IgG Isotype | |
| IgA | 14 |
| IgG | 37 |
| Light chain κ/λ | 11 |
| BM plasma cells, % | |
| ≤40 | 21 |
| >40 | 41 |
| Serum LDH (IU/L) | |
| ≤420 | 51 |
| >420 | 6 |
| NA | 5 |
| β2-microglobulin, mg/L | |
| <3.5 | 3 |
| ≥3.5 | 59 |

MRS = modified risk staging [52].

cluding 61 newly diagnosed treatment naïve MM patients and 1 MGUS (who later converted to MM at TP2) diagnosed as per IMWG guidelines (**Table 1**). Patients diagnosed and treated at our center from 2014 to 2019 in whom DNA samples were available prior to therapy and at the time of disease progression were included in this study. The patients were treated with triplet combination induction chemotherapy-VCD (bortezomib, cyclophosphamide and dexamethasone) or VTD (bortezomib, thalidomide, dexamethasone) or VRD (bortezomib, lenalidomide, dexamethasone) prior to time of progression. The median OS of the patient cohort was 152.5 weeks and median PFS was 87.21 weeks.

Whole exome sequencing (WES) was carried out on 186 DNA samples extracted from 62 MM patients collected at two time points- one prior to any therapy at diagnosis (Time Point 1 = TP1) and second at a follow up time point of disease progression (Time Point 2 = TP2). WES was also carried out on paired germline DNA obtained from peripheral blood mononuclear cells for all the patients.

For WES, DNA was extracted using Maxwell RSC cultured cells DNA kit (Promega, Wisconsin, USA) on automated nucleic acid extraction system (Promega, Wisconsin, USA). Prior to library construction, DNA was quantified fluorometrically with a DNA high sensitivity kit with Qubit (ThermoFisher Scientific, MA, USA). WES libraries were constructed from genomic DNA using the Nextera Exome kit (Illumina, San Diego, California, USA) which targets a genomic footprint of 62 Mb with >3,40,000, 95 mer probes. After quantification, the DNA was normalized to 10 ng/μl and a total of 50 ng DNA was tagmented with transposons. The tagmented DNA was purified from the transposome with sample purification beads. The purified tagmented DNA was subjected to a unique combination of dual index adapters and amplified with sequences required for cluster generation. After amplification, the DNA libraries were purified and the purified libraries containing unique indices were combined into a single pool using a 3-plex strategy. The target regions of interest in the purified libraries were hybridized with coding exome oligos and captured with streptavidin magnetic beads. The enriched libraries were eluted from the beads and subjected to a second round of hybridization with coding exome oligos. Final libraries were eluted and then quantified and evaluated for quality using DNA high sensitivity

Qubit kit (ThermoFisher Scientific, MA, USA) and DNA HS Kit (Agilent Technologies, Santa Clara, USA) on Agilent Bioanalyser respectively. The size range of generated libraries was 200-500 bp. The resultant captured libraries were pooled, normalized following standard normalization method and paired-end sequencing was carried out using the Illumina cBot system and HiSeq SBS kit V4-250 cycles on HiSeq 2500 (Illumina).

*Analysis of WES data*

The overall workflow of data analysis is shown in **Figure 1**. Raw sequencing reads were quality checked using FastQC software (v0.11.4, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The adapter sequences were removed using Trimmomatic software (v0.39, http://www.usadellab.org/cms/?page=trimmomatic). Illumina Dragen somatic pipeline (v3.5.7) was used to process the trimmed reads and aligned with human reference genome, hg19 available at UCSC (https://sapac.illumina.com/products/by-type/informatics-products/basespace-sequence-hub/apps/edico-genome-inc-dragen-somatic-pipeline.html).

The tumor and normal bam files obtained from Illumina Dragen somatic pipeline were used for variant calling using three additional variant callers, Strelka v2.9.10 [23]; SomaticSniper v1.0.5.0 [24] and SpeedSeq v0.1.2 (FreeBayes) [25] in order to validate the variants called by Dragen somatic pipeline. Only those variants called by all the four callers and passed filters of base quality (≥20), mapping quality (≥20), tumor reads (≥10) and normal reads (≥5) qualified as a consensus. These validated variants were further annotated using BaseSpace Variant Interpreter (https://variantinterpreter.informatics.illumina.com/home).

Further, COSMIC database was explored for assignment of variant pathogenicity (Pathogenic/Neutral/Unknown). Variants predicted as Deleterious/Damaging/Pathogenic by any of the three tools (SIFT/PolyPhen/FATHMM) were considered as Pathogenic. For identification of CNVs, the .bam files of tumor and normal samples obtained from Illumina Dragen (v3.5.7) somatic pipeline were analyzed using Sequenza (https://cran.r-project.org/web/packages/sequenza/) package along with human reference .fasta file from UCSC (ucsc.hg19.fasta).

Variants identified were compared with MMRF CoMMPass Study database (www.themmrf.org). The mutated genes were classified as driver genes, oncogenes and tumor suppressor genes based on publicly available resources listed at cBioPortal [26, 27] (https://www.cbioportal.org/); at intOgen (https://www.intogen.org/search) database [28]; OncoKB [29] (https://www.oncokb.org/) and as described by Walker *et al.,* 2014 [17].

Potentially actionable targets were identified in this study based on repository of FDA approved on label or off-label drugs or those experimentally druggable compiled and listed in literature [30, 31], at the TARGET (Tumor Alterations Relevant for Genomics driven Therapy) database of the Broad Institute (https://software.broadinstitute.org/cancer/cga/target) and the COSMIC actionability data v93 (https://cancer.sanger.ac.uk/cosmic). The TARGET database is a database of genes that when somatically altered in cancer, are directly linked to a clinical action. The tumor mutational burden (TMB) defined as the number of nonsynonymous mutations/Mb was calculated from average coverage with respect to total bases (313-7161264) in binary mode and with reference to human genome (hg19). Clonal evolution patterns were evaluated using QuantumClone [32] and the cellular prevalence values $\hat{\theta}$ were calculated as defined below (https://www.rdocumentation.org/packages/QuantumClone/versions/0.15.11).

$$\hat{\theta} = VAF \times \frac{N_{Ch} + N_{Ch(Normal)} \times \frac{1 - p}{p}}{NC}$$

where $N_{Ch}$ is the number of copies of the corresponding locus in cancer cells, $N_{Ch(Normal)}$ is the number of copies of the corresponding locus in the normal cells ($N_{Ch(Normal)}$ = 2 for autosomes) and *NC* is the number of chromosomal copies bearing the variant and *p* is the tumor purity.

The cellular prevalence values $\hat{\theta}$ of each cluster obtained from QuantumClone were subjected to fishplot R package for visualization [33]. Cellular prevalence values higher than 1 were set to 1 as suggested [32]. Clonal patterns were classified as branching or linear or stable as described [13]. In case of branching evolu-

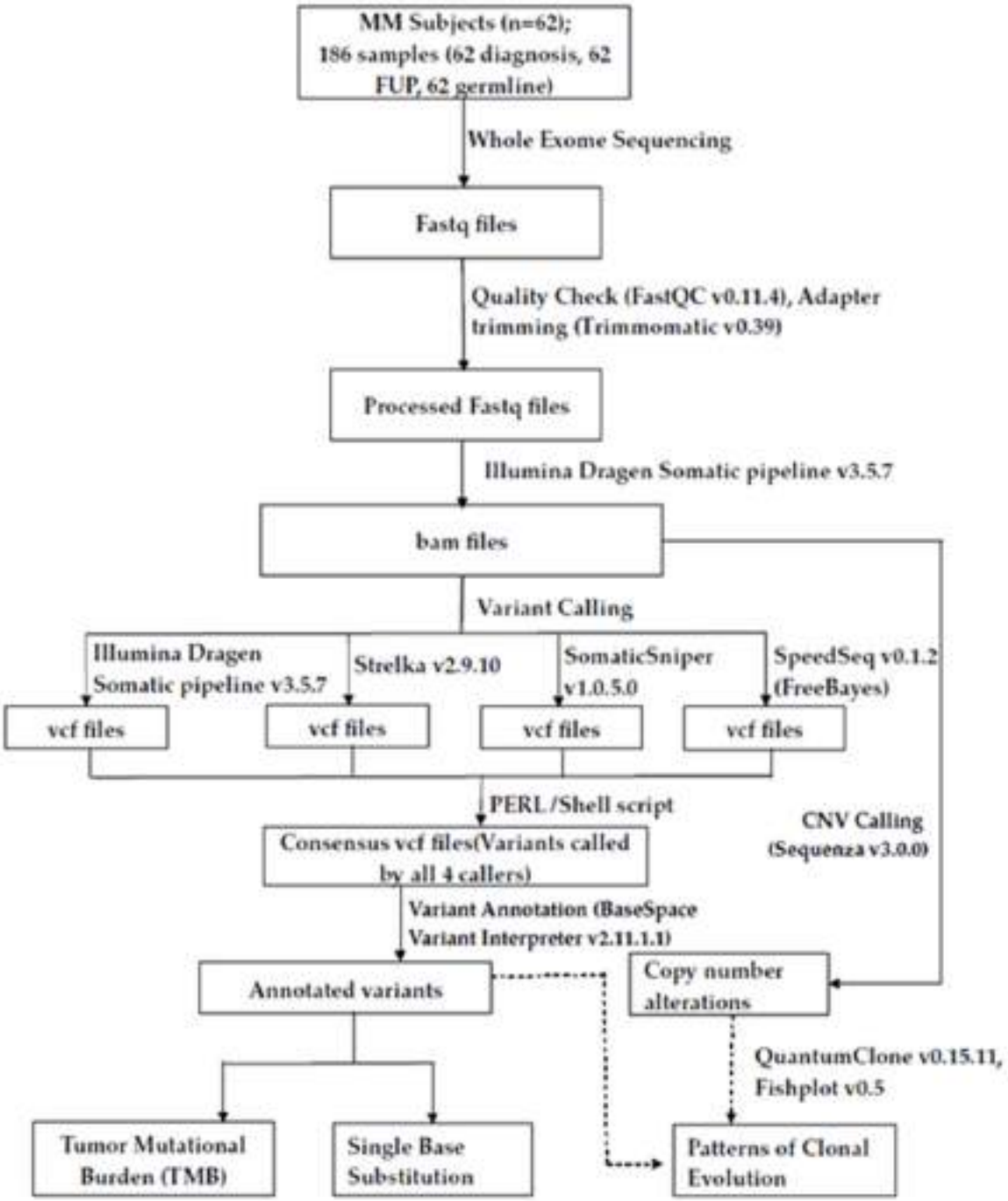



**Figure 1.** Workflow of study and data analysis. Analysis workflow of the WES study performed on 62 MM patients whose tumor PC samples were sequenced at diagnosis, at follow up and compared with their germline profiles. Fastq files were quality checked with FastQC, adaptors trimmed with Trimmomatic and processed further through Illumina Dragen Somatic pipeline for variant calling. Variants were validated with additional 3 variant callers (Strelka2, SomaticSniper and SpeedSeq), a consensus .vcf was derived and annotated with Variant Interpreter for deducing TMB and SBS. CNVs were identified with Sequenza and processed further with QuantumClone and Fishplot for interpretation of patterns of clonal evolution.

tion, both gain and loss of clones was observed. In case of linear evolution, there was gain of mutations but no clonal loss; while in stable progression, the clonal structure remained

**Table 2.** A comparison of number of nonsynonymous (NS) somatic mutations, tumor mutation burden (TMB) and single base substitutions (SBS) in MM at diagnosis and on progression

| Type of somatic mutations | Time point | |
|---|---|---|
| | TP1 (at Diagnosis) | TP2 (on progression) |
| IN ALL SAMPLES (n = 59) | | |
| Number of somatic mutations | 13951 | 11684 |
| Number of known pathogenic somatic mutations | 4410 | 3833 |
| Number of Missense somatic mutations | 10561 | 8996 |
| Number of Nonsense somatic mutations | 188 | 160 |
| Number of somatic mutations in 3'UTR | 1227 | 946 |
| Number of somatic mutations at Splicing sites | 1437 | 1207 |
| Number of somatic mutations in 5'UTR | 538 | 375 |
| MEANS PER SAMPLE | | |
| Average number of somatic mutations/sample | 236.45 | 198.03 |
| Average number of Missense somatic mutations/sample | 179 | 152.47 |
| Average number of Nonsense somatic mutations/sample | 3.19 | 2.71 |
| Average number of somatic mutations in 3'UTR/sample | 20.80 | 16.03 |
| Average number of somatic mutations at Splicing sites/sample | 24.36 | 20.46 |
| Average number of somatic mutations in 5'UTR/sample | 9.12 | 6.36 |
| MEDIAN number of NS somatic mutations | 32 | 34 |
| Tumor Mutation Burden (TMB) | | |
| MEDIAN TMB | 0.85 | 0.93 |
| AVERAGE SBS IN ALL SAMPLES | | |
| C>T | 128.88 | 101.86 |
| T>C | 85.02 | 64.66 |
| C>A | 34.34 | 27.56 |
| C>G | 28.64 | 21.92 |
| T>G | 21.98 | 16.47 |
| T>A | 15.90 | 12.25 |

preserved at two time points. Stable with loss pattern had predominantly conserved clonal structure but there was also evidence of clonal loss at a subsequent time point. The biological pathways relating to altered clonal mutational profiles were deduced by gene enrichment analysis using Enrichr (https://maayanlab.cloud/Enrichr/) as described [34].

*Statistical analysis*

Clinical and biological characteristics of the patients were analysed using Chi-squared or Fisher's exact test for discrete categorical variables as applicable. Nonparametric statistical analysis was carried out for continuous variables with Wilcoxon signed rank test. A *P*-value of <0.05 was considered statistically significant.

## Results

*Estimation of somatic mutations at two time points*

A total of 13951 and 11684 nonsynonymous (NS) somatic mutations were identified in myeloma pairwise whole exomes sequenced at diagnosis (TP1) and at progression (TP2) respectively (**Table 2**). Among these, 4410 somatic mutations in TP1 and 3833 in TP2 were classifiable as pathogenic. At diagnosis, 10561 somatic mutations were missense type, 1227 belonged to 3'UTR, 1437 were in splicing sites and 538 mapped in 5'UTR regions. On progression, these reduced to 8996, 946, 1207 and 375 somatic mutations representing missense, 3'UTR, splicing and 5'UTR mutations, respectively.

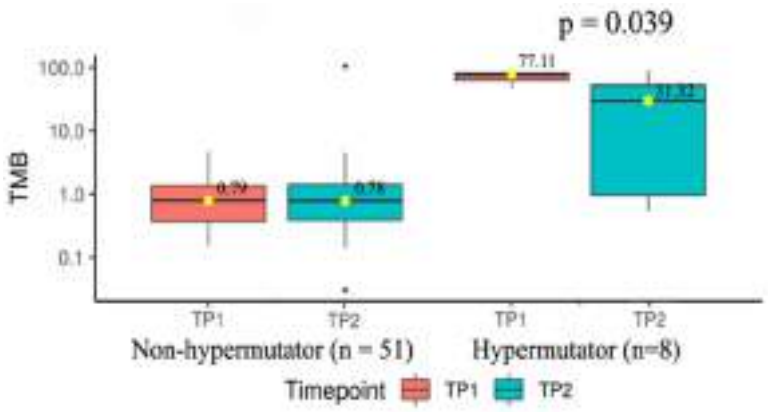



**Figure 2.** Changes in TMB at diagnosis and on progression. Comparison of median TMB across MM patients at TP1 and TP2 in non-hypermutator (n = 51) (TMB<10) versus hypermutator category (n = 8) (TMB between 10 and 100).

The average numbers of somatic mutations/sample at diagnosis totalled 236.45 at TP1 while 198.03 at TP2 (**Table 2**). At TP1, there were an average of 179 missense mutations/sample (152.47 at TP2), followed by 20.8 in 3'UTR (16.03 at TP2), 24.36 in splicing regions (20.46 at TP2), and 9.12 in 5'UTR region (6.36 at TP2). Patients with high somatic mutations may possess high neoantigen loads and may benefit from immunotherapies.

*Tumor mutation burden declines from diagnosis to progression in hypermutators*

Patients at diagnosis had an average tumor mutation burden (TMB) of 10.8 NS somatic mutations/Mb/sample (range 0.15 to 95) that reduced to 7.46 (range 0.03 to 105.47) on progression. The median TMB among patients at TP1 and TP2 were 0.85 and 0.93 respectively. The median TMB at two time points among patients with age at diagnosis ≤65 years (0.82 versus 0.76) and those with >65 years (1.62 versus 1.22) were comparable.

Patients were classified on the basis of their TMB levels at diagnosis as those with low TMB of ≤10 (n = 51) and high TMB levels ≥10 to ≤100 (n = 8) (i.e., hypermutators). Three patients (SM0007, SM0052 and SM0145) were outliers or super-hypermutators with ≥100 TMBs (134.43, 132.12 and 126.3 respectively) and were analyzed for clonal evolution exclusively. In particular, patients grouped into high TMB category (TMB levels ≥10 to ≤100) had median TMB levels at TP1 (77.11) that significantly reduced at TP2 (31.32; P = 0.039) (**Figure 2**). Hypermutators might sustain stable drug resistant clones and hence may benefit from combinations of IMiDs with novel therapeutics.

*Comparison of frequencies of driver genes mutated at diagnosis versus progression*

**Table 3** summarizes number of mutated genes and mutations that were encountered in MM in this study. Out of 8977 total mutated genes that got shortlisted, 8869 were found to be mutated in some form of cancer while 7107 genes were identified to be mutated in MM among which 6690 genes have been reported in MMRF CoMMPass dataset. A set of 131 mutated genes turned out to be known oncogenes, 176 were established tumor suppressors, 320 were known drivers across different cancers while 72 genes were found to be known driver genes in the context of MM. Of all these genes harbouring somatic mutations in MM, 100 genes got classified as COSMIC candidate actionable targets.

We screened the WES data for a total repertoire of 102 known driver genes for MM and found 72 driver genes to be mutated. We then analyzed which driver genes had subclonal gains or losses or remained stable with progression and arranged them in descending and ascending series (**Figure 3**). These drivers were further shortlisted to those that had topmost number of recurrent subclonal shifts and were observed in atleast 3 or more patients. **Figure 3A** shows topmost temporal falls in *PABPC1, BRAF, KRAS, CR1, DIS3, ATM* and other genes while **Figure 3C** shows topmost temporal increases that were observed in *KMT2C, FOXD4L1, SP140* and *NRAS*. Similarly, **Figure 3B** shows the most recurrent drivers like *FAT4* and *IGLL5* that remained stable on progression. Contrasting mutational landscapes at diagnosis and at progression highlight the importance of their immediate monitoring prior to tailoring therapy.

**Table 3.** Classification of genes harbouring NS somatic mutations and the variants observed in MM in this study

| Classification | Number of genes with mutations (n = 8977) | Number of mutations (n = 19022) |
|---|---|---|
| Known to be mutated in some cancer | 8869 | 18817 |
| Known to be mutated in MM | 7107 | 15864 |
| Mutated in MMRF CoMMPass study | 6690 | 15063 |
| Known oncogenes | 131 | 252 |
| Known tumor suppressor genes | 176 | 443 |
| Known to be driver genes in some cancer | 320 | 821 |
| Known to be driver genes in MM | 72 | 221 |
| Known as actionable (COSMIC) | 100 | 239 |
| Drivers with decreased frequencies on progression | 39 | 140 |
| Drivers with increased frequencies on progression | 12 | 36 |
| Drivers with constant frequencies both at diagnosis and on progression | 21 | 45 |

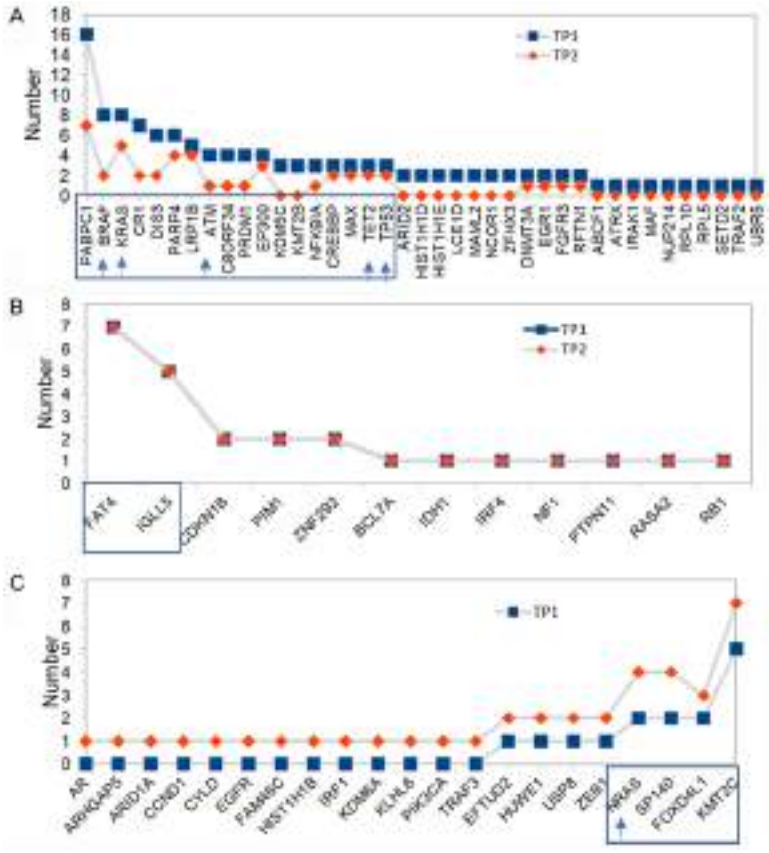



**Figure 3.** Temporal changes in distribution of driver genes on progression. Distribution of mutated driver genes in MM patients at TP1 and compared to TP2. (A) Falling mutated drivers whose frequencies decreased in TP2, (B) Drivers that are maintained at constant frequencies throughout the disease, and (C) Rising mutated drivers whose preponderance increased in patients at TP2. Driver mutation profiles observed in atleast 3 or more patients are shown inside boxed frames. Actionable genes are indicated by arrows on X axis.

*Distribution of mutated potential actionable target genes at diagnosis and progression*

As many as 19022 somatic mutations (**Table 3**) were observed at varying frequencies among 8977 genes in MM patients in this study. Of these, 18817 variants are known mutants in cancers of some kind, 15864 have been reported to be mutated in MM while 15063 have been described in MMRF dataset. These consisted of 821 mutations across drivers known to be associated with different cancers and 221 mutations in 72 driver genes (*BRAF, SP140, EP300, FAT4, PABPC1, CREBBP, FOXD4L1, PRDM1, KMT2C, C8ORF34, NRAS, KRAS, DIS3, NFKBIA, LRP1B, IGLL5, ZNF292, ATM, CR1, PTPN11, BCL7A, CDKN1B, PARP4, RB1, MAX, NF1, EFTUD2, TP53, DNMT3A, RASA2, RFTN1, TET2, EGR1, HIST1H1E, PIM1, ZEB1, FAM46C, LCE1D, CCND1, MAML2, ARID2, ARID1A, TRAF3, ARHGAP5, USP8, CY-*

*LD, ZFHX3, MAF, NCOR1, RPL5, KMT2B, IDH1, PIK3CA, KLHL6, SETD2, FGFR3, IRF1, HIST1H1D, HIST1H1B, ABCF1, IRF4, EGFR, UBR5, NUP214, TRAF2, IRAK1, RPL10, KDM6A, KDM5C, HUWE1, AR, ATRX*) known to be involved in MM. There were 252 somatic mutations in oncogenes, 443 in tumor suppressor genes and finally 239 variants were found across 100 potential actionable genes.

**Table 4** summarizes a list of variations in 22 actionable target genes that were found mutated in atleast 3 patients at either or both time points. These consisted of *BRAF, FANCM, MRE11, WRN, EXO1, FANCA, ALK, FANCD2, MSH3, NBN, NRAS, KRAS, FLT3 MAP2K1, PALB2, RAD51D, RAD51C, MERTK, KDR, RAD54B, FANCG, PTCH1.* The most common actionable mutation was Val600Glu in BRAF that was most abundant at the time of diagnosis. Identification of druggable targets at subclonal levels could aid in treating patients with genome defined target specific drugs.

*Comparison of single nucleotide substitutions at diagnosis and progression*

As shown in **Table 2**, six types of single base substitutions (SBS) were observed. The SBS C>T was the most predominant form of mutation found both at TP1 (128.88; 40.94%) and TP2 (101.86; 41.63%) followed by T>C (85.02; 27.017% at TP1, 64.66; 26.42% at TP2), C>A (34.34; 10.9% at TP1, 27.56; 11.26% at TP2), C>G (28.64, 9.09% at TP1; 21.92, 8.95% at TP2), T>G (21.98, 6.98% at TP1; 16.47, 6.73% at TP2), T>A (15.9, 5.05% at TP1; 12.25, 5% at TP2).

*Heterogeneity in clonal evolution*

Three types of clonal evolutionary patterns with 1 to 3 founder clones were observed in this study (**Figure 4**). The branching pattern of clonal evolution was observed in maximum number of patients (45; 72.58%) followed by Linear in 9 cases (14.51%) and Stable with loss of clone in 8 patients (12.90%) (**Figure 4A**). Distribution of founder clones in different subsets of patients with branching (n = 45) and non-branching (n = 17) evolution is shown in **Figure 4B**. One, two and three founder clones were detected in 18, 20 and 7 patients respectively out of 45 patients with branching patterns of clonal evolution. Patients with branching pattern of evolution had significantly higher number of founder clones (p = 0.0173, **Figure 4B**) than those with non-branching patterns. A significant number of patients with low TMB at TP1 developed branching clonal evolution (n = 40 out of 51) whereas those with high TMB had both branching (n = 5 out of 11) and non-branching evolutionary patterns (n = 6 out of 11) (P = 0.026) (**Figure 4C**).

Each case of MM was analyzed in depth by QuantumClone and their individual fish plots, clonal density and evolution plots were generated (Supplementary Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13). A median of 3 clones (range 2 to 9) was observed among 45 patients with branching clonal evolution. The number of clones was relatively lower among patients with non-branching evolution patterns-Linear (2 to 4) and Stable with loss of clone (2 to 3). **Figure 5A-C** shows a representative fish plot of each of the three types of clonal patterns of evolution (Branching, Linear and Stable with loss) observed in this study. The somatic mutational diversity in founder clones and their cellular prevalence was compared at two time points for each patient. A schematic representation of genes found to be mutated in founder clones including actionable/non-actionable genes and the significantly associated biological pathways predicted to be affected by such mutated genes in patients are shown in **Figures 6A**, **6B** and **7** respectively.

The heatmaps in **Figure 6A** and **6B** also depict falling/rising frequencies of actionable and non-actionable targets (including DRV/ONC/TSG/others) respectively. The topmost ten genes mutated in founder clones were *BAGE2* (37.28%) >*PABPC1* (30.5%) >*MUC17/NBPF1* (23.72%) >*DNAH14/FLG* (22.03%) >*FAT1/RHPN2/TPTE* (20.33%). The topmost frequently mutated actionable targets were *KRAS* (18.64%) >*BRAF/FANCM* (13.55%) >*FANCD2/WRN* (11.86%) >*FANCA/MLH1* (10.16%) >*NRAS/ATM* (8.47%) >*TET2/BRCA1* (6.77%) >*FGFR3/TP53* (5.08%), and others.

The cellular prevalence of topmost mutated tumor suppressor gene *KMT2C* showed an increase with progression in 6 out of 11 patients followed by *FAT1* (6 out of 12), *FANCA* (3 out of 6), *BRCA1* (3 out of 4), *TET2* (2 out of 4) and *NRAS* (4 out of 5) (**Figure 6**). On the

**Table 4.** Frequency of variations in actionable genes observed in atleast 3 or more multiple myeloma patients

| VARIANT | REF | ALT | EXON | HGVSC | HGVSP | CONSEQUENCE | GENE | Number of patients with mutation | Count at TP1 | Count at TP2 |
|---|---|---|---|---|---|---|---|---|---|---|
| 7:140453136:T | A | T | 15/18 | c.1799T>A | p.(Val600Glu) | missense_variant | BRAF | 7 | 5 | 2 |
| 14:45606287:T | C | T | 2/23 | c.524C>T | p.(Ser175Phe) | missense_variant | FANCM | 6 | 5 | 1 |
| 14:45650900:G | A | G | 16/23 | c.4378A>G | p.(Ile1460Val) | missense_variant | FANCM | 6 | 5 | 1 |
| 14:45665468:G | C | G | 21/23 | c.5434C>G | p.(Pro1812Ala) | missense_variant | FANCM | 6 | 4 | 2 |
| 11:94212048:T | C | T | | | c.403-6G>A | splice_region_intron_variant | MRE11 | 5 | 2 | 3 |
| 8:30999280:T | G | T | 26/35 | c.3222G>T | p.(Leu1074Phe) | missense_variant | WRN | 5 | 3 | 2 |
| 1:242042301:A | G | A | 13/16 | c.1765G>A | p.(Glu589Lys) | missense_variant | EXO1 | 4 | 3 | 1 |
| 16:89836323:T | C | T | 26/43 | c.2426G>A | p.(Gly809Asp) | missense_variant | FANCA | 4 | 2 | 2 |
| 16:89849480:T | C | T | 16/43 | c.1501G>A | p.(Gly501Ser) | missense_variant | FANCA | 4 | 2 | 2 |
| 2:29416366:C | G | C | 29/29 | c.4587C>G | p.(Asp1529Glu) | missense_variant | ALK | 4 | 3 | 1 |
| 3:10088266:T | G | T | 15/43 | c.1137G>T | c.1137G>T (p.(Val379=)) | splice_region,synonymous_variant | FANCD2 | 4 | 1 | 3 |
| 3:10140671:A | G | A | 43/43 | c.*37G>A | | 3_prime_UTR_variant | FANCD2 | 4 | 2 | 2 |
| 3:10140696:G | A | G | 43/43 | c.*62A>G | | 3_prime_UTR_variant | FANCD2 | 4 | 2 | 2 |
| 5:79960955:A | G | A | | | c.359-7G>A | splice_region_intron_variant | MSH3 | 4 | 3 | 1 |
| 8:30999123:A | G | A | | | c.3138+7G>A | splice_region,intron_variant | WRN | 4 | 3 | 1 |
| 8:90958530:C | T | C | | | c.1915-7A>G | splice_region_intron_variant | NBN | 4 | 2 | 2 |
| 8:90990479:G | C | G | 5/16 | c.553G>C | p.(Glu185Gln) | missense_variant | NBN | 4 | 2 | 2 |
| 1:115256529:C | T | C | 3/7 | c.182A>G | p.(Gln61Arg) | missense_variant | NRAS | 3 | 1 | 2 |
| 12:25362777:G | A | G | 6/6 | c.*73T>C | | 3_prime_UTR_variant | KRAS | 3 | 2 | 1 |
| 12:25380275:G | T | G | 3/6 | c.183A>C | p.(Gln61His) | missense_variant | KRAS | 3 | 2 | 1 |
| 13:28610183:G | A | G | | | c.1310-3T>C | splice_region_intron_variant | FLT3 | 3 | 1 | 2 |
| 15:66782048:T | C | T | | | c.1023-8C>T | splice_region,intron_variant | MAP2K1 | 3 | 1 | 2 |
| 16:23646191:C | T | C | 4/13 | c.1676A>G | p.(Gln559Arg) | missense_variant | PALB2 | 3 | 2 | 1 |
| 17:33433487:T | C | T | 6/10 | c.494G>A | p.(Arg165Gln) | missense_variant | RAD51D | 3 | 2 | 1 |
| 17:56811608:G | C | G | 9/9 | c.*25C>G | | 3_prime_UTR_variant | RAD51C | 3 | 2 | 1 |
| 2:112686988:A | G | A | 2/19 | c.353G>A | p.(Ser118Asn) | missense_variant | MERTK | 3 | 2 | 1 |
| 2:29416481:C | T | C | 29/29 | c.4472A>G | p.(Lys1491Arg) | missense_variant | ALK | 3 | 2 | 1 |
| 3:10106532:T | C | T | 23/43 | c.2141C>T | p.(Pro714Leu) | missense_variant | FANCD2 | 3 | 2 | 1 |
| 4:55972974:A | T | A | 11/30 | c.1416A>T | p.(Gln472His) | missense_variant | KDR | 3 | 2 | 1 |
| 5:80168937:A | G | A | 23/24 | c.3133G>A | p.(Ala1045Thr) | missense_variant,splice_region_variant | MSH3 | 3 | 2 | 1 |
| 8:95479680:C | G | C | 2/15 | c.88C>G | p.(Leu30Val) | missense_variant | RAD54B | 3 | 2 | 1 |
| 9:35074917:C | T | C | | | c.1636+7A>G | splice_region,intron variant | FANCG | 3 | 3 | 0 |
| 9:98239147:G | A | G | | | c.1504-8T>C | splice_region,intron_variant | PTCH1 | 3 | 1 | 2 |

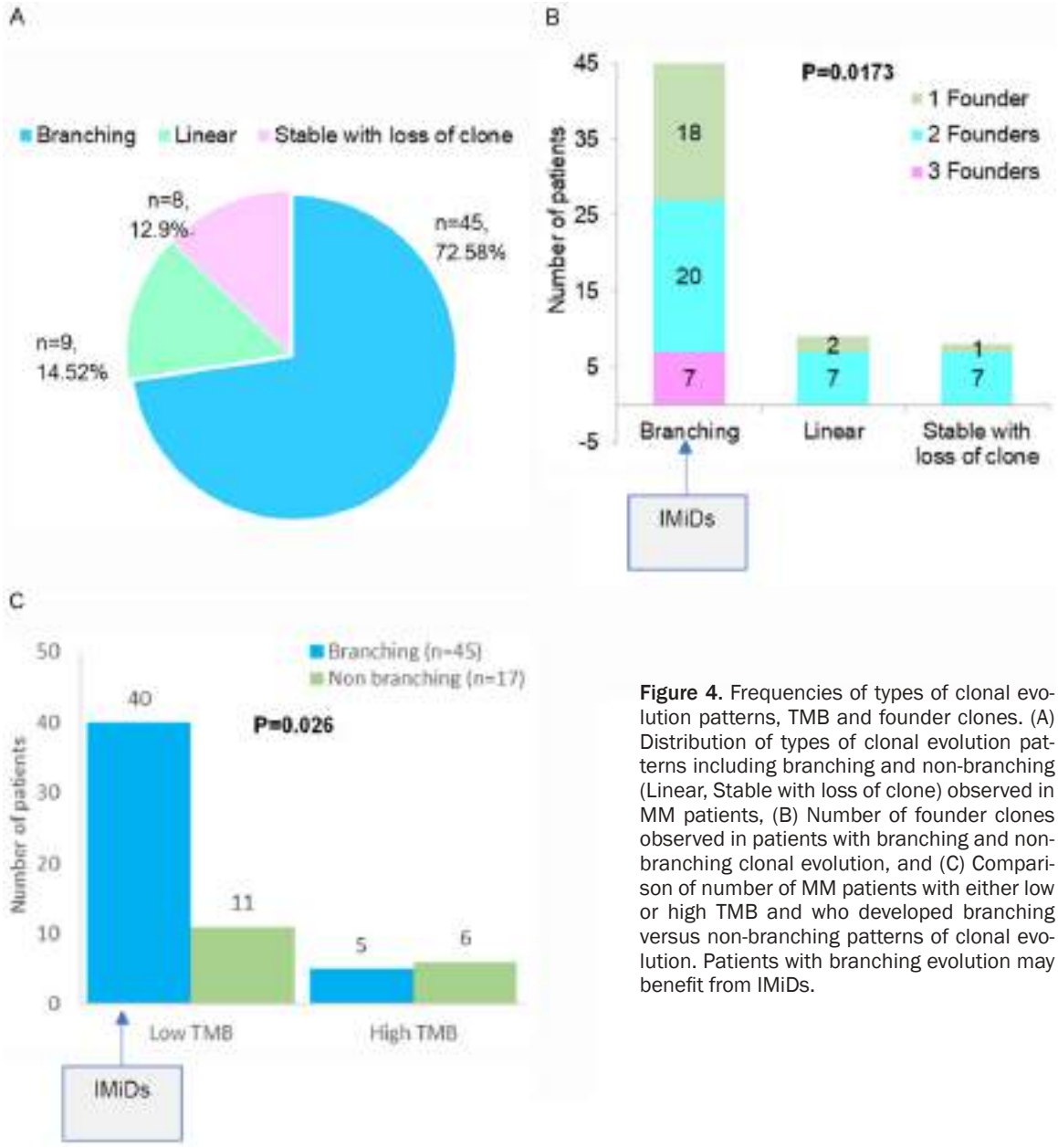



**Figure 4.** Frequencies of types of clonal evolution patterns, TMB and founder clones. (A) Distribution of types of clonal evolution patterns including branching and non-branching (Linear, Stable with loss of clone) observed in MM patients, (B) Number of founder clones observed in patients with branching and non-branching clonal evolution, and (C) Comparison of number of MM patients with either low or high TMB and who developed branching versus non-branching patterns of clonal evolution. Patients with branching evolution may benefit from IMiDs.

contrary, cellular prevalence of mutated driver *PABPC1* decreased with progression in 13 out of 18 patients, *KRAS* (8 out of 11), *BRAF* (6 out of 8), *ATM* (4 out of 5) and others (**Figure 6**).

*Prediction of biological pathways affected by somatic mutations*

A comprehensive gene enrichment analysis by Enrichr identified a network of biological pathways found to be significantly associated with somatic mutations on progression of MM (**Figure 7**). These included, notably, ECM-receptor interaction, Galactose metabolism, Protein digestion and absorption, Cholesterol metabolism, Antigen processing and presentation, Drug metabolism, RNA degradation, Starch and sucrose metabolism, Hematopoietic cell lineage, Base excision repair, MAPK signaling pathway, viral carcinogenesis, cell cycle, apoptosis, Th17 cell differentiation, Th1 and Th2 cell differentiation, beta-Alanine metabolism, cellular senescence and others.

Pathways that were affected by 2434 mutated genes found exclusively at diagnosis and those affected by new mutations in genes at TP2 are shown in **Figure 8**. Additional pathways (n =

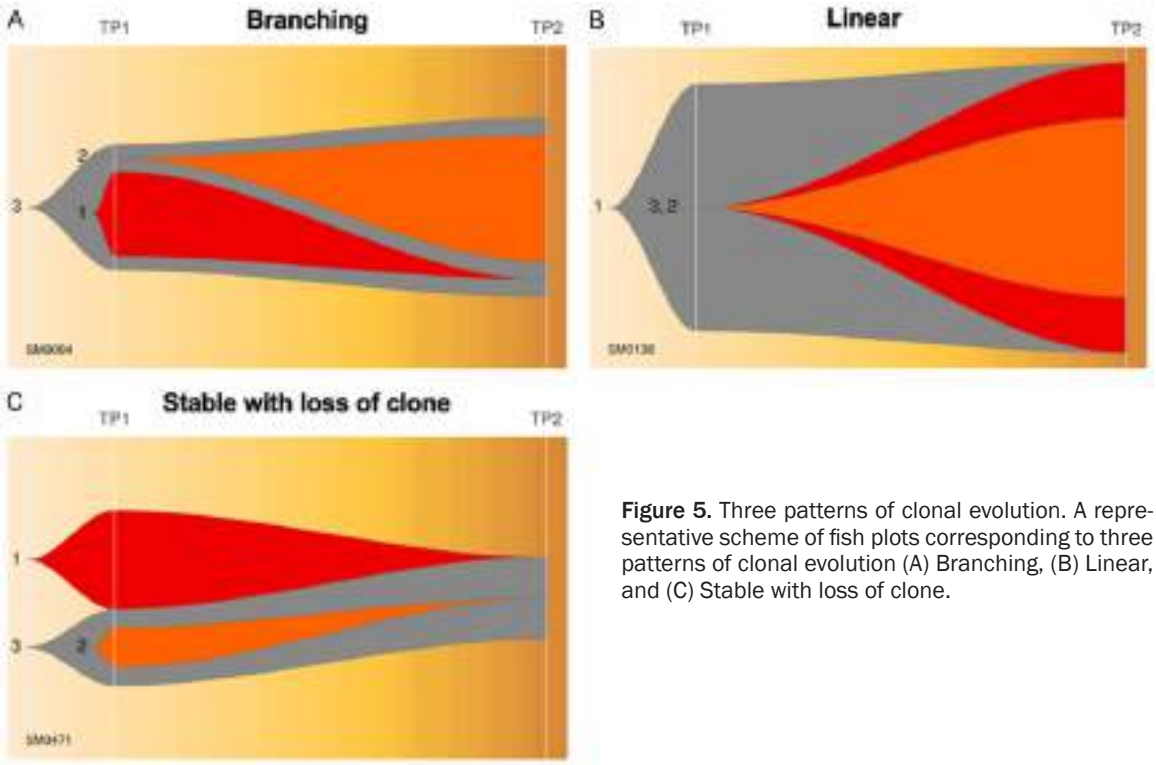



**Figure 5.** Three patterns of clonal evolution. A representative scheme of fish plots corresponding to three patterns of clonal evolution (A) Branching, (B) Linear, and (C) Stable with loss of clone.

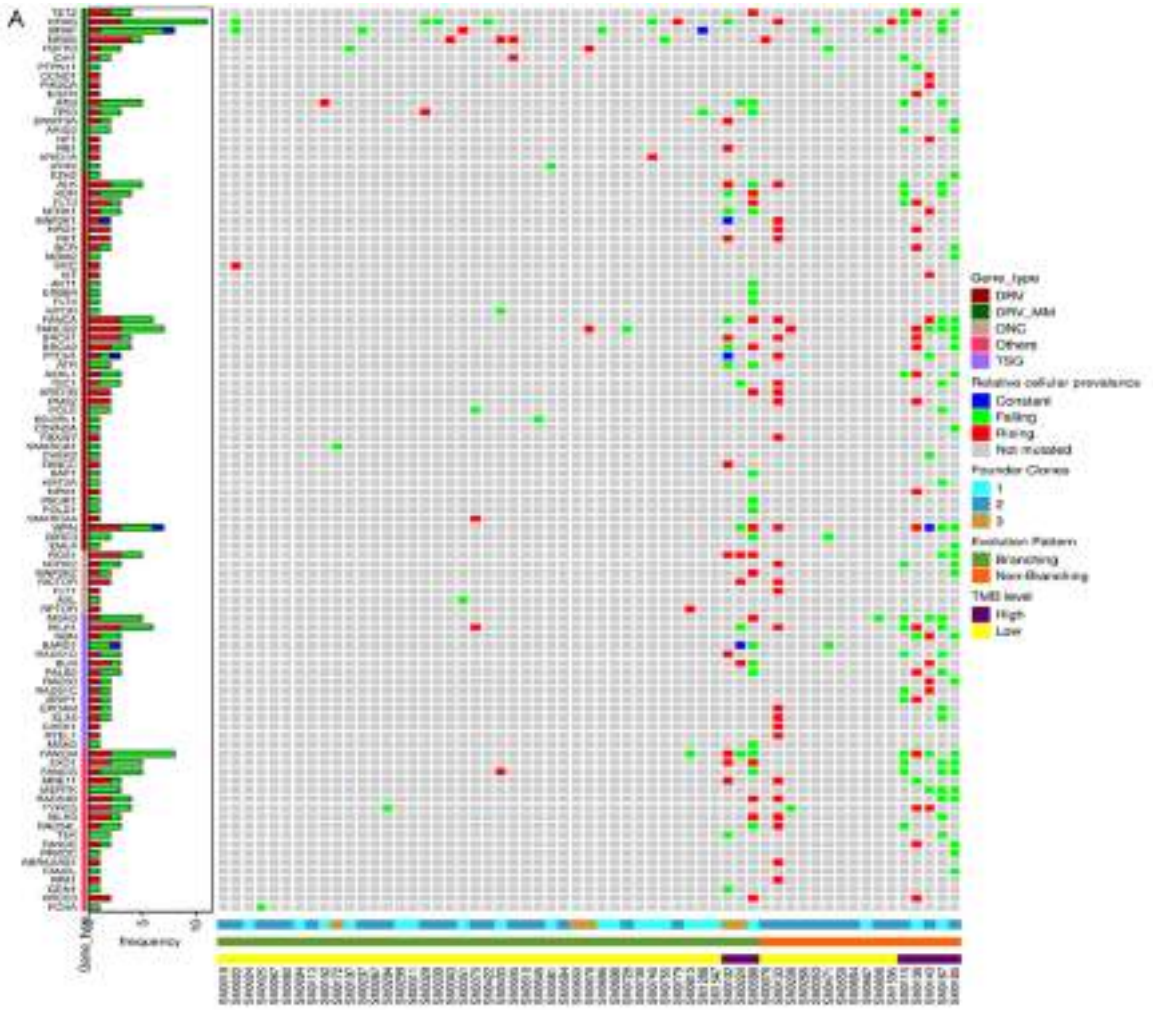



**Figure 6.** Comparison of potential actionable and non-actionable mutated genes in different samples grouped as with branching or non-branching clonal evolution patterns and low or high TMB levels. A. Heatmap depicting distribution of actionable targets including drivers, oncogenes and tumor suppressors with rising or falling frequency

trends across MM patients classified on the basis of branching/non-branching clonal evolutionary patterns, TMB levels and number of founder clones. B. Heatmap depicting distribution of non-actionable target genes drivers, oncogenes and tumor suppressors with rising or falling trends across MM patients classified on the basis of branching/non-branching clonal evolutionary patterns, TMB levels and number of founder clones.

13) found to be affected exclusively on progression included NK cell mediated cytotoxicity, chemical carcinogenesis, PI3K-Akt signaling, phototransduction, PPAR signaling, GnRH signaling and others. Likewise, 18 pathways were exclusively affected by mutations at TP1.

*Clonal divergence in individual cases*

**Figure 5A-C** Shows a representative fish plot of each of the three types of clonal patterns of evolution (Branching, Linear and Stable with loss) observed in this study. A case-wise description of subclones and their patterns of evolution are summarized in Supplementary Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 and in Supplementary Note 1.

## Discussion

Progression of MM is linked with a spatiotemporal shift in subclonal structure. The prime objective of this study was to explore subclonal evolution associated with progression of MM and identify potential actionable targets for each patient. In order to achieve this, we adopted a novel Ensemble algorithm approach for identification of mutations. As per our findings and as suggested by others [35, 36], there can be significant differences in the SNV outputs processed by different variant callers based on the properties of the caller used, their strengths and weaknesses. Since no somatic caller has the ultimate ability to perform, an ensemble approach that combines multiple callers has been reported to offer the best balance of both sensitivity and specificity [36-38]. Hence, we decided to call mutations through four common variant callers (Dragen, Strelka2, SomaticSniper and SpeedSeq) and generate a common consensus rather than depending on any single one. This innovative approach ensured that the clonal landscape of MM captured in our study was closest possible estimation to reality.

An important observation of this study is that we have been able to identify recurrent subclonal shifts in actionable/druggable targets of clinical importance such as *BRAF, KRAS, ATM, TET2* and *TP53* at diagnosis in multiple patients (in atleast 3 patients or more) (**Figure 3A**). A similar gain in subclonal *NRAS* mutations was observed at the time of progression (**Figure 3C**). The reduction in frequencies of driver genes with progression can be explained by their selective loss in response to therapy that may coincide with fulfillment of their initial functional role(s) needed in triggering myelomagenesis. On the other hand, an increase in another set of driver genes indicates an effect of evolutionary pressure that allows selection of topmost fit clones. These sweeping subclones may either be novel or may result from expansion of pre-existing mutations known to be present at low or undetectable frequencies at the time of diagnosis or earlier. The inability to detect low copy mutations is largely due to technical limitations of sequencing of bulk tumor tissue and recent advanced technologies of single cell sequencing may be able to resolve effect of evolving somatic mutations more lucidly.

Screening of actionable genetic mutations in these genes allows to match patients with future treatments that would be most beneficial, which is in coherence with the overall goal of the ongoing Multiple Myeloma Research Foundation (MMRF) MyDRUG (Myeloma-Developing Regimens Using Genomics) clinical trial (NCT03732703) [39]. The MyDRUG aims at enrolling patients with mutations in *BRAF, NRAS, KRAS, FGFR3, CDKN2C, IDH2* or t(11;14) and assign to appropriate targeted agent against that mutation. Patients with *BRAF* V600E or any *NRAS* or *KRAS* actionable mutations found in subclonal populations could thus benefit the most if treated early with BRAF inhibitor e.g. Vemurafenib or MEK inhibitor Cobimetinib respectively. Heat maps in **Figure 6A** show genomic signatures of actionable genes for each patient enrolled in this study that could be targeted specifically to select the right drug for the right patient based on the specificity of the mutation.

TMB is an emerging prognostic biomarker of response to immunotherapy, approximation of neoantigen load and overall survival especially in solid tumors [40, 41]. A high TMB is considered a biomarker of higher neoantigen load,

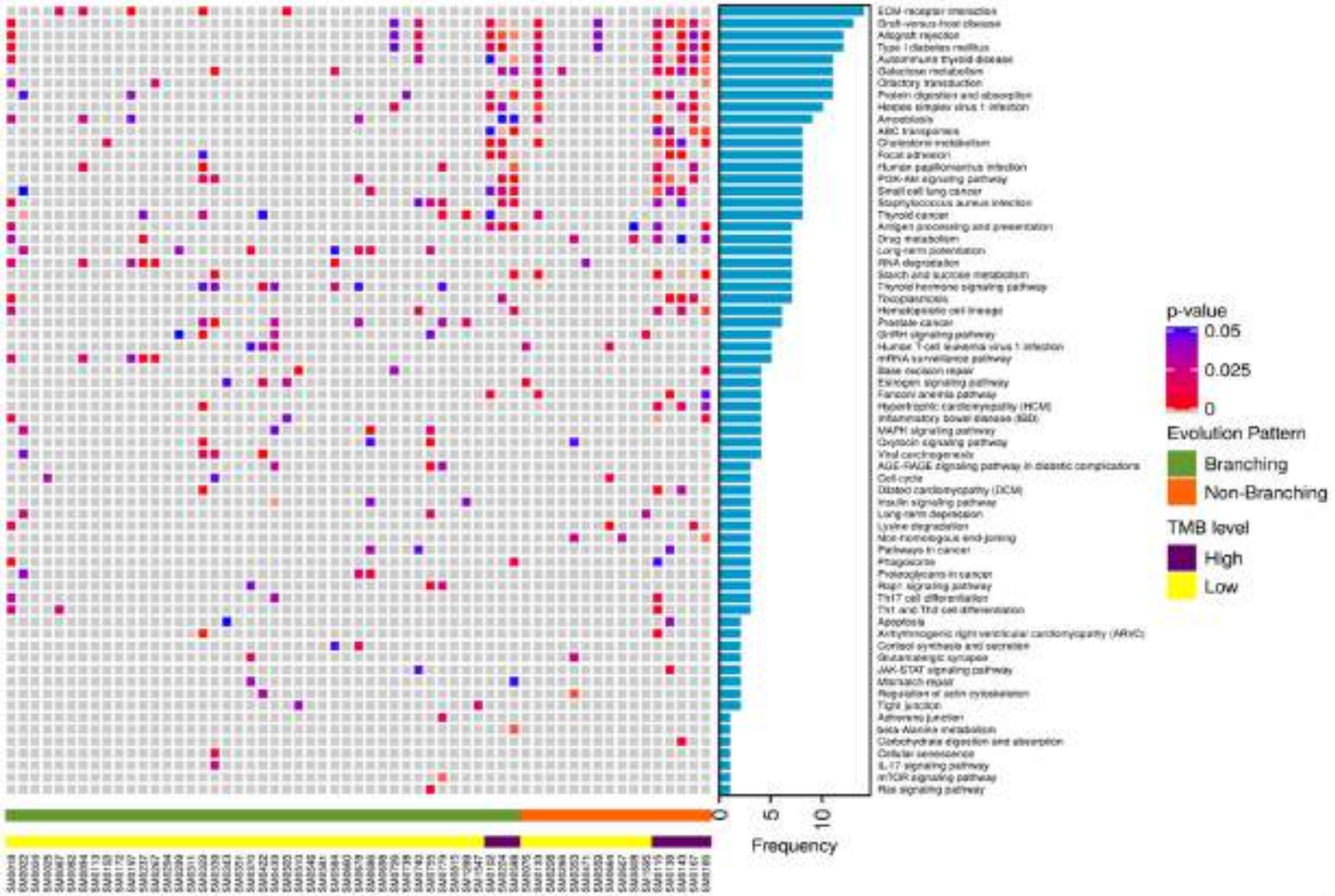



**Figure 7.** Predicted pathways affected by somatic mutations across samples. Heatmap depicting significantly affected biological pathways predicted to be altered by Enrichr across MM patients classified on the basis of branching/non-branching clonal evolutionary patterns and TMB levels.

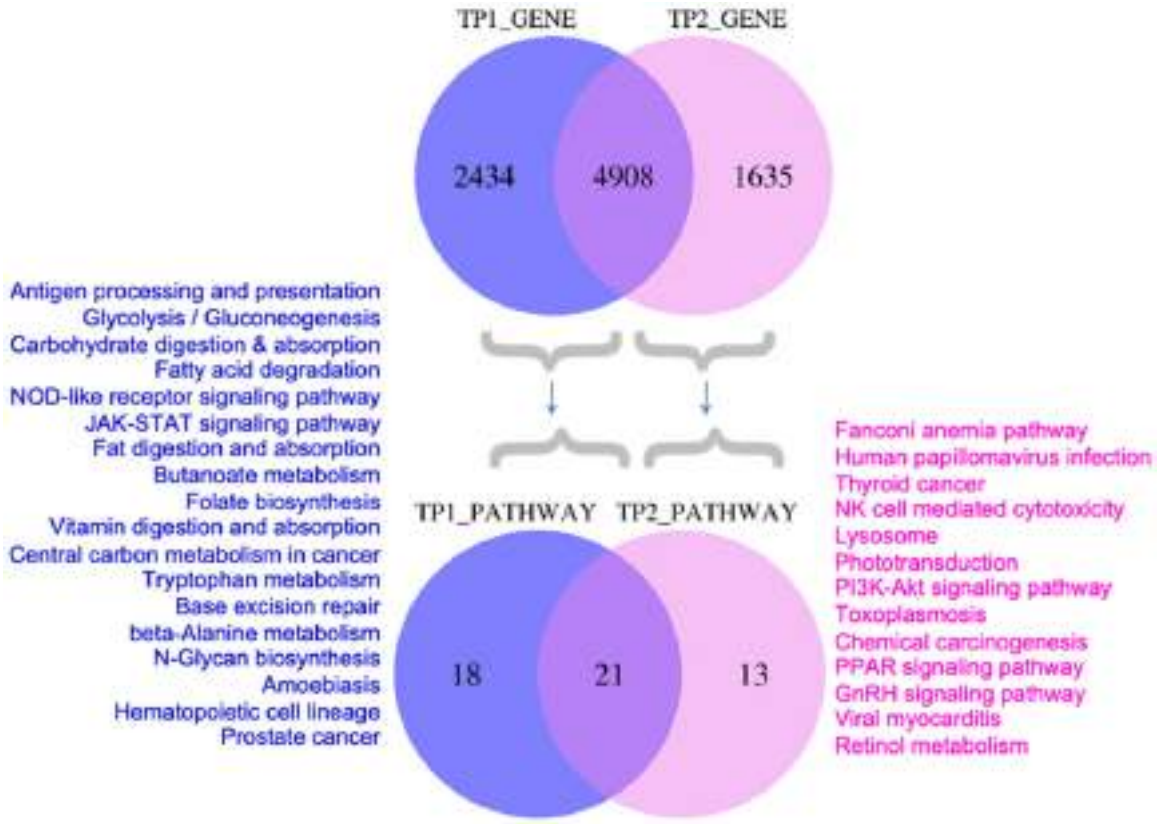



**Figure 8.** Comparison of mutated genes and associated pathways at diagnosis and at progression. Venn diagram representing number of mutated genes and the predicted biological pathways affected by mutations exclusively at diagnosis (TP1) or progression (TP2).

increased response rates to immunotherapy and better outcomes. High somatic mutation and neoantigen loads have been found to correlate with reduced PFS in MM [42]. Patients were classified in this study into those with low TMB between ≤1 to 10 or high TMB (≥10 or hypermutators). This study has shown a modest loss of TMB from diagnosis to progression but only in a subset of patients with hypermutator status (i.e. TMB≥10) (**Figure 2**). There could be a selective loss of less fit drug sensitive clones yet with persistence of drug resistant clones in such patients and hence combinations IMiDs with novel therapeutics could be used to treat such patients.

This study has shown a predominance of branching pattern of clonal evolution in MM in concurrence to other studies [8, 13-19] (**Figure 4A**). An increase in DNA damage and a branching pattern of evolution are considered hallmarks of effectiveness of therapy and attainment of deep response [13]. Although the branching type of evolution reflects on the better response rates to therapy while tumor strives to mutate and acquire fitter clones to survive, it is also a prominent underlying mechanism of relapse. While mutations in founder clones are primarily involved in initiation of myelomagenesis, those in subclones may contribute significantly to relapse. The study has further shown that branching evolution is more predominant among patients with 2 or more founder clones (**Figure 4B**) and those with low tumor mutation burden (TMB<10) (**Figure 4C**). Since, this happens under the positive selection pressure of therapy and the microenvironment, such patients could perhaps benefit more from immunomodulatory drugs (IMiDs) such as thalidomide/lenalidomide and analogues [12].

Studies have shown that ongoing DNA damage intensifies from MGUS to MM and provides a mechanism by which chromosomal aberrations and heterogeneity are acquired by malignant plasma cells [43]. **Figures 7** and **8** show the functional pathways that were affected by genetic mutations on progression. These include

pathways in cancer, metabolism of galactose, cholesterol, drugs, cellular senescence, cell cycle, apoptosis, viral carcinogenesis, RNA degradation, base excision repair and several other crucial signalling pathways involved in pathogenesis of MM or immune surveillance. Deregulated DNA damage repair related pathways as also seen in our study have been associated with poor prognosis [44] since the tumor cells can withstand DNA damaging drugs and repopulate with therapy resistant cells on treatment. It has been suggested that a 'synthetic lethality' approach [45] may be more beneficial where co-treatment of patients with current drugs and those targeting DNA repair pathways [46] (e.g, Bortezomib with PARP1 inhibitor [47] or Spironolactone [48] or a novel compound DCZ3301 [49]) may reverse drug resistance in such patients [50, 51].

Studies like this have shown genomic plasticity of mutational landscapes and how relative preponderance of mutated drivers changes with disease progression. Supplementary Figures 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13 shows individual evolution patterns as FISH plots followed by summarized individual case reports on 62 newly diagnosed MM patients enrolled in this study. It provides a detailed genomic architecture and cellular prevalence of each and every subclone identified for every patient at diagnosis and at progression. **Table 4** summarizes the number of patients who had an actionable/druggable mutation and who could qualify for targeted treatments with target specific drugs. Comprehensive analysis of mutational subclonal landscapes of patients as observed in this study is a pre-requisite to infer the genomic mutations that can be treated in future in similar lines as in MyDRUG trial. An integration of such early genomic biomarkers with clinical biomarkers could help in risk estimation and identification of patients who could benefit more from a rationalized therapeutic approach at early stages. It is indeed not just the individual mutations but an extended treatment landscape that needs to be monitored preferably at multiple time points to tailor therapy. An early assessment of TMB along with mutations in drivers and actionable target genes during decision making, may therefore, allow most appropriate therapeutic personification in clinics.

In conclusion, a systematic analysis of evolving mutational landscapes, TMB and SBS signatures could help in better stratification of high risk MGUS/SMM/MM patients prior to subclonal expansion and therefore open the opportunities of early and personified cure for the disease.

## Acknowledgements

This work was supported by grants from Department of Biotechnology, Govt. of India [BT/MED/30/SP11006/2015] and Department of Science and Technology, Govt. of India [DST/ICPS/CPS-Individual/2018/279(G)]. Akanksha Farswan would like to thank University Grants Commission, Govt. of India for UGC-Senior Research Fellowship. Authors acknowledge Multiple Myeloma Research Foundation (MMRF) for providing the dataset. These data were generated as part of the Multiple Myeloma Research Foundation Personalised Medicine Initiative.

## Disclosure of conflict of interest

None.

**Address correspondence to:** Ritu Gupta, Laboratory Oncology Unit, Dr. B.R.A. IRCH, AIIMS, Ansari Nagar, New Delhi 110029, India. Tel: +91-9873433275; E-mail: drritugupta@gmail.com; drritu.laboncology@aiims.edu; Anubha Gupta, SBILab, Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology-Delhi (IIIT-D), Delhi 110020, India. Tel: +91-8826066166; E-mail: anubha@iiitd.ac.in

## References

[1] Maura F, Rustad EH, Boyle EM and Morgan GJ. Reconstructing the evolutionary history of multiple myeloma. Best Pract Res Clin Haematol 2020; 33: 101145.

[2] Kumar SK and Rajkumar SV. The multiple myelomas-current concepts in cytogenetic classification and therapy. Nat Rev Clin Oncol 2018; 15: 409-421.

[3] Landgren O and Morgan GJ. Biologic frontiers in multiple myeloma: from biomarker identification to clinical practice. Clin Cancer Res 2014; 20: 804-813.

[4] van Nieuwenhuijzen N, Spaan I, Raymakers R and Peperzak V. From MGUS to multiple myeloma, a paradigm for clonal evolution of premalignant cells. Cancer Res 2018; 78: 2449-2456.

[5] Kaufmann H, Ackermann J, Baldia C, Nösslinger T, Wieser R, Seidl S, Sagaster V, Gisslinger H, Jäger U, Pfeilstöcker M, Zielinski C and Drach J. Both IGH translocations and chromosome 13q deletions are early events in monoclonal gammopathy of undetermined significance and do not evolve during transition to multiple myeloma. Leukemia 2004; 18: 1879-1882.

[6] Rajan AM and Rajkumar SV. Interpretation of cytogenetic results in multiple myeloma for clinical practice. Blood Cancer J 2015; 5: e365.

[7] Manier S, Salem KZ, Park J, Landau DA, Getz G and Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. Nat Rev Clin Oncol 2016; 14: 100-113.

[8] 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 JMC, 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: 2997.

[9] Furukawa Y and Kikuchi J. Molecular pathogenesis of multiple myeloma. Int J Clin Oncol 2015; 20: 413-422.

[10] Bolli N, Maura F, Minvielle S, Gloznik D, Szalat R, Fullam A, Martincorena I, Dawson KJ, Samur MK, Zamora J, Tarpey P, Davies H, Fulciniti M, Shammas MA, Tai YT, Magrangeas F, Moreau P, Corradini P, Anderson K, Alexandrov L, Wedge DC, Avet-Loiseau H, Campbell P and Munshi N. Genomic patterns of progression in smoldering multiple myeloma. Nat Commun 2018; 9: 3363.

[11] Johnson DC, Lenive O, Mitchell J, Jackson G, Owen R, Drayson M, Cook G, Jones JR, Pawlyn C, Davies FE, Walker BA, Wardell C, Gregory WM, Cairns D, Morgan GJ, Houlston RS and Kaiser MF. Neutral tumor evolution in myeloma is associated with poor prognosis. Blood 2017; 130: 1639-1643.

[12] Furukawa Y and Kikuchi J. Molecular basis of clonal evolution in multiple myeloma. Int J Hematol 2020; 111: 496-511.

[13] Jones JR, Weinhold N, Ashby C, Walker BA, Wardell C, Pawlyn C, Rasche L, Melchor L, Cairns DA, Gregory WM, Johnson D, Begum DB, Ellis S, Sherborne AL, Cook G, Kaiser MF, Drayson MT, Owen RG, Jackson GH, Davies FE, Greaves M and Morgan GJ; NCRI Haemato-Oncology CSG. Clonal evolution in myeloma: the impact of maintenance lenalidomide and depth of response on the genetics and subclonal structure of relapsed disease in uniformly treated newly diagnosed patients. Haematologica 2019; 104: 1440-1450.

[14] Weinhold N, Ashby C, Rasche L, Chavan SS, Stein C, Stephens OW, Tytarenko R, Bauer MA, Meissner T, Deshpande S, Patel PH, Buzder T, Molnar G, Peterson EA, van Rhee F, Zangari M, Thanendrarajan S, Schinke C, Tian E, Epstein J, Barlogie B, Davies FE, Heuck CJ, Walker BA and Morgan GJ. Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. Blood 2016; 128: 1735-1744.

[15] Melchor L, Brioli A, Wardell CP, Murison A, Potter NE, Kaiser MF, Fryer RA, Johnson DC, Begum DB, Hulkki Wilson S, Vijayaraghavan G, Titley I, Cavo M, Davies FE, Walker BA and Morgan GJ. Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. Leukemia 2014; 28: 1705-1715.

[16] Lohr JG, Stojanov P, Carter SL, Cruz-Gordillo P, Lawrence MS, Auclair D, Sougnez C, Knoechel B, Gould J, Saksena G, Cibulskis K, McKenna A, Chapman MA, Straussman R, Levy J, Perkins LM, Keats JJ, Schumacher SE, Rosenberg M, Getz G and Golub TR. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell 2014; 25: 91-101.

[17] 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.

[18] 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.

[19] Maura F, Bolli N, Angelopoulos N, Dawson KJ, Leongamornlert D, Martincorena I, Mitchell TJ,

Fullam A, Gonzalez S, Szalat R, Abascal F, Rodriguez-Martin B, Samur MK, Glodzik D, Roncador M, Fulciniti M, Tai YT, Minvielle S, Magrangeas F, Moreau P, Corradini P, Anderson KC, Tubio JMC, Wedge DC, Gerstung M, Avet-Loiseau H, Munshi N and Campbell PJ. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. Nat Commun 2019; 10: 3835.

[20] Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K and Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CWM, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES and Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013; 499: 214-218.

[21] Samur MK, Aktas Samur A, Fulciniti M, Szalat R, Han T, Shammas M, Richardson P, Magrangeas F, Minvielle S, Corre J, Moreau P, Thakurta A, Anderson KC, Parmigiani G, Avet-Loiseau H and Munshi NC. Genome-wide somatic alterations in multiple myeloma reveal a superior outcome group. J Clin Oncol 2020; 38: 3107-3118.

[22] Keats JJ, Chesi M, Egan JB, Garbitt VM, Palmer SE, Braggio E, Van Wier S, Blackburn PR, Baker AS, Dispenzieri A, Kumar S, Rajkumar SV, Carpten JD, Barrett M, Fonseca R, Stewart AK and Bergsagel PL. Clonal competition with alternating dominance in multiple myeloma. Blood 2012; 120: 1067-1076.

[23] Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X, Kim Y, Beyter D, Krusche P and Saunders CT. Strelka2: fast and accurate calling of germline and somatic variants. Nat Methods 2018; 15: 591-594.

[24] Larson DE, Abbott TE and Wilson RK. Using somaticsniper to detect somatic single nucleotide variants. Curr Protoc Bioinforma 2014; 45: 15.5.1-8.

[25] Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, Marth GT, Quinlan AR and Hall IM. SpeedSeq: ultra-fast personal genome analysis and interpretation. Nat Methods 2015; 12: 966-968.

[26] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, Cerami E, Sander C and Schultz N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013; 6: pl1.

[27] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C and Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012; 2: 401-404.

[28] Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, Santos A and Lopez-Bigas N. IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods 2013; 10: 1081-1082.

[29] Chakravarty D, Gao J, Phillips SM, Kundra R, Zhang H, Wang J, Rudolph JE, Yaeger R, Soumerai T, Nissan MH, Chang MT, Chandarlapaty S, Traina TA, Paik PK, Ho AL, Hantash FM, Grupe A, Baxi SS, Callahan MK, Snyder A, Chi P, Danila D, Gounder M, Harding JJ, Hellmann MD, Iyer G, Janjigian Y, Kaley T, Levine DA, Lowery M, Omuro A, Postow MA, Rathkopf D, Shoushtari AN, Shukla N, Voss M, Paraiso E, Zehir A, Berger MF, Taylor BS, Saltz LB, Riely GJ, Ladanyi M, Hyman DM, Baselga J, Sabbatini P, Solit DB and Schultz N. OncoKB: a precision oncology knowledge base. JCO Precis Oncol 2017; 2017: PO.17.00011.

[30] Schuh A, Dreau H, Knight SJL, Ridout K, Mizani T, Vavoulis D, Colling R, Antoniou P, Kvikstad EM, Pentony MM, Hamblin A, Protheroe A, Parton M, Shah KA, Orosz Z, Athanasou N, Hassan B, Flanagan AM, Ahmed A, Winter S, Harris A, Tomlinson I, Popitsch N, Church D and Taylor JC. Clinically actionable mutation profiles in patients with cancer identified by whole-genome sequencing. Cold Spring Harb Mol Case Stud 2018; 4: a002279.

[31] Galanina N, Bejar R, Choi M, Goodman A, Wieduwilt M, Mulroney C, Kim L, Yeerna H, Tamayo P, Vergilio JA, Mughal TI, Miller V, Jamieson C and Kurzrock R. Comprehensive genomic profiling reveals diverse but actionable molecular portfolios across hematologic malignancies: implications for next generation clinical trials. Cancers (Basel) 2019; 11: 11.

[32] Deveau P, Colmet Daage L, Oldridge D, Bernard V, Bellini A, Chicard M, Clement N, Lapouble E, Combaret V, Boland A, Meyer V, Deleuze JF, Janoueix-Lerosey I, Barillot E, Delattre O, Maris JM, Schleiermacher G and Boeva V. QuantumClone: clonal assessment of functional mutations in cancer based on a genotype-aware method for clonal reconstruction. Bioinformatics 2018; 34: 1808-1816.

[33] Miller CA, McMichael J, Dang HX, Maher CA, Ding L, Ley TJ, Mardis ER and Wilson RK. Visu-

alizing tumor evolution with the fishplot package for R. BMC Genomics 2016; 17: 880.

[34] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW and Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016; 44: W90-97.

[35] Roberts ND, Kortschak RD, Parker WT, Schreiber AW, Branford S, Scott HS, Glonek G and Adelson DL. A comparative analysis of algorithms for somatic SNV detection in cancer. Bioinforma Oxf Engl 2013; 29: 2223-2230.

[36] Koboldt DC. Best practices for variant calling in clinical sequencing. Genome Med 2020; 12: 91.

[37] Fang LT, Afshar PT, Chhibber A, Mohiyuddin M, Fan Y, Mu JC, Gibeling G, Barr S, Asadi NB, Gerstein MB, Koboldt DC, Wang W, Wong WH and Lam HY. An ensemble approach to accurately detect somatic mutations using SomaticSeq. Genome Biol 2015; 16: 197.

[38] Anzar I, Sverchkova A, Stratford R and Clancy T. NeoMutate: an ensemble machine learning framework for the prediction of somatic mutations in cancer. BMC Med Genomics 2019; 12: 63.

[39] Auclair D, Anderson K, Avigan D, Biran N, Chaudhry M, Cho H, Furlong M, Hofmeister C, Kansagra A, Krishnan A, Larsen J, Orloff G, Vij R, Voorhees P, Yee A, Ye J, Zonder J, Lonial S and Kumar S. The myeloma-developing regimens using genomics (MyDRUG) master protocol. J Clin Oncol 2019; 37: TPS8057.

[40] McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT, Watkins TBK, Shafi S, Murugaesu N, Mitter R, Akarca AU, Linares J, Marafioti T, Henry JY, Van Allen EM, Miao D, Schilling B, Schadendorf D, Garraway LA, Makarov V, Rizvi NA, Snyder A, Hellmann MD, Merghoub T, Wolchok JD, Shukla SA, Wu CJ, Peggs KS, Chan TA, Hadrup SR, Quezada SA and Swanton C. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science 2016; 351: 1463-1469.

[41] Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, Ibrahim F, Bruggeman C, Gasmi B, Zappasodi R, Maeda Y, Sander C, Garon EB, Merghoub T, Wolchok JD, Schumacher TN and Chan TA. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 2015; 348: 124-128.

[42] Miller A, Asmann Y, Cattaneo L, Braggio E, Keats J, Auclair D, Lonial S, Russell SJ and Stewart AK. High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. Blood Cancer J 2017; 7: e612.

[43] Walters DK, Wu X, Tschumper RC, Arendt BK, Huddleston PM, Henderson KJ, Dispenzieri A and Jelinek DF. Evidence for ongoing DNA damage in multiple myeloma cells as revealed by constitutive phosphorylation of H2AX. Leukemia 2011; 25: 1344-1353.

[44] Kassambara A, Gourzones-Dmitriev C, Sahota S, Rème T, Moreaux J, Goldschmidt H, Constantinou A, Pasero P, Hose D and Klein B. A DNA repair pathway score predicts survival in human multiple myeloma: the potential for therapeutic strategy. Oncotarget 2014; 5: 2487-2498.

[45] Cottini F, Hideshima T, Suzuki R, Tai YT, Bianchini G, Richardson PG, Anderson KC and Tonon G. Synthetic lethal approaches exploiting DNA damage in aggressive myeloma. Cancer Discov 2015; 5: 972-987.

[46] Szalat R, Samur MK, Fulciniti M, Lopez M, Nanjappa P, Cleynen A, Wen K, Kumar S, Perini T, Calkins AS, Reznichenko E, Chauhan D, Tai YT, Shammas MA, Anderson KC, Fermand JP, Arnulf B, Avet-Loiseau H, Lazaro JB and Munshi NC. Nucleotide excision repair is a potential therapeutic target in multiple myeloma. Leukemia 2018; 32: 111-119.

[47] Gourzones C, Bret C and Moreaux J. Treatment may be harmful: mechanisms/prediction/prevention of drug-induced DNA damage and repair in multiple myeloma. Front Genet 2019; 10: 861.

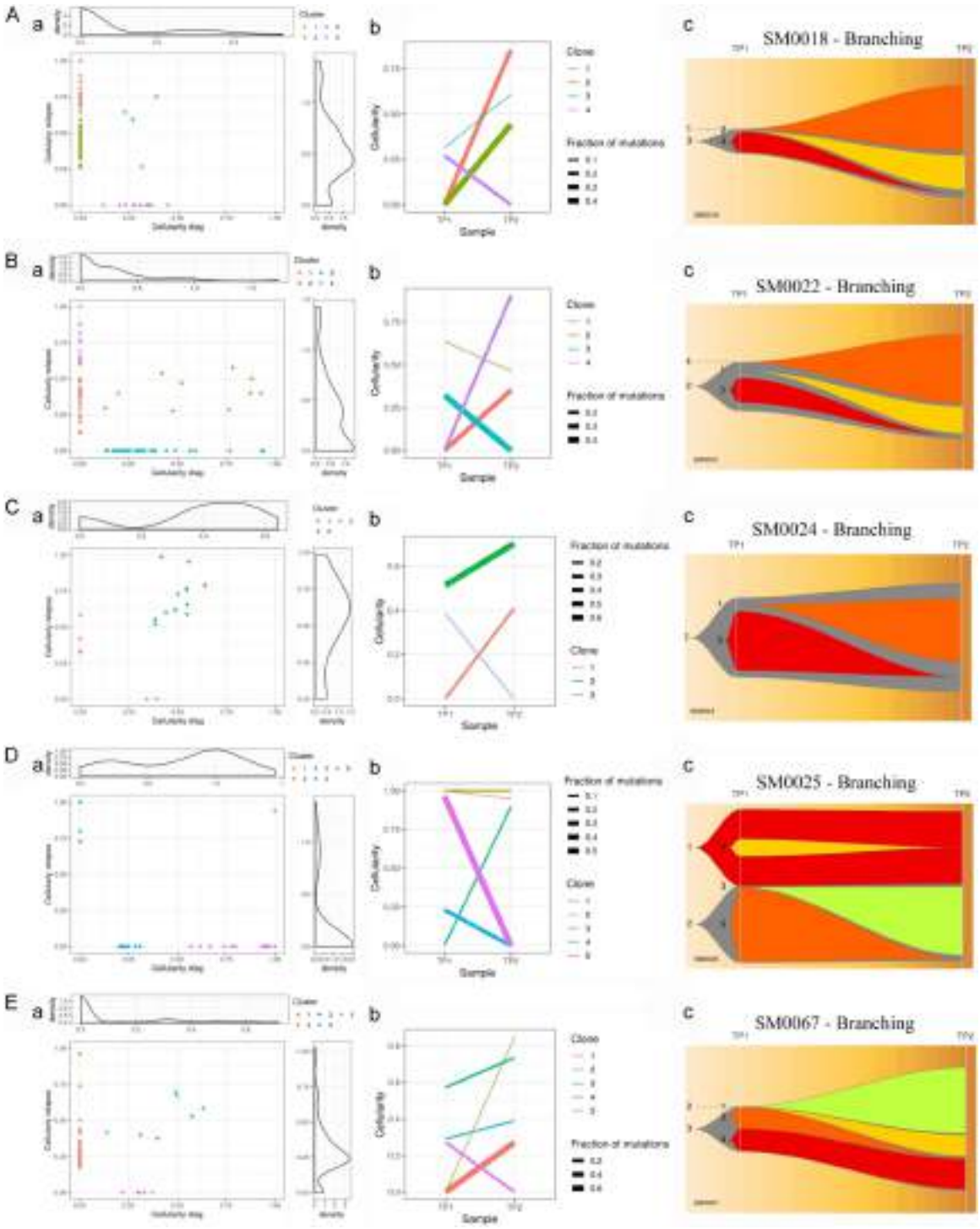
[48] Chauhan D, Ray A, Viktorsson K, Spira J, Paba-Prada C, Munshi N, Richardson P, Lewensohn R and Anderson KC. In vitro and in vivo antitumor activity of a novel alkylating agent, melphalan-flufenamide, against multiple myeloma cells. Clin Cancer Res 2013; 19: 3019-3031.

[49] Hu L, Li B, Chen G, Song D, Xu Z, Gao L, Xi M, Zhou J, Li L, Zhang H, Feng Q, Wang Y, Lu K, Lu Y, Bu W, Wang H, Wu X, Zhu W and Shi J. A novel M phase blocker, DCZ3301 enhances the sensitivity of bortezomib in resistant multiple myeloma through DNA damage and mitotic catastrophe. J Exp Clin Cancer Res 2020; 39: 105.

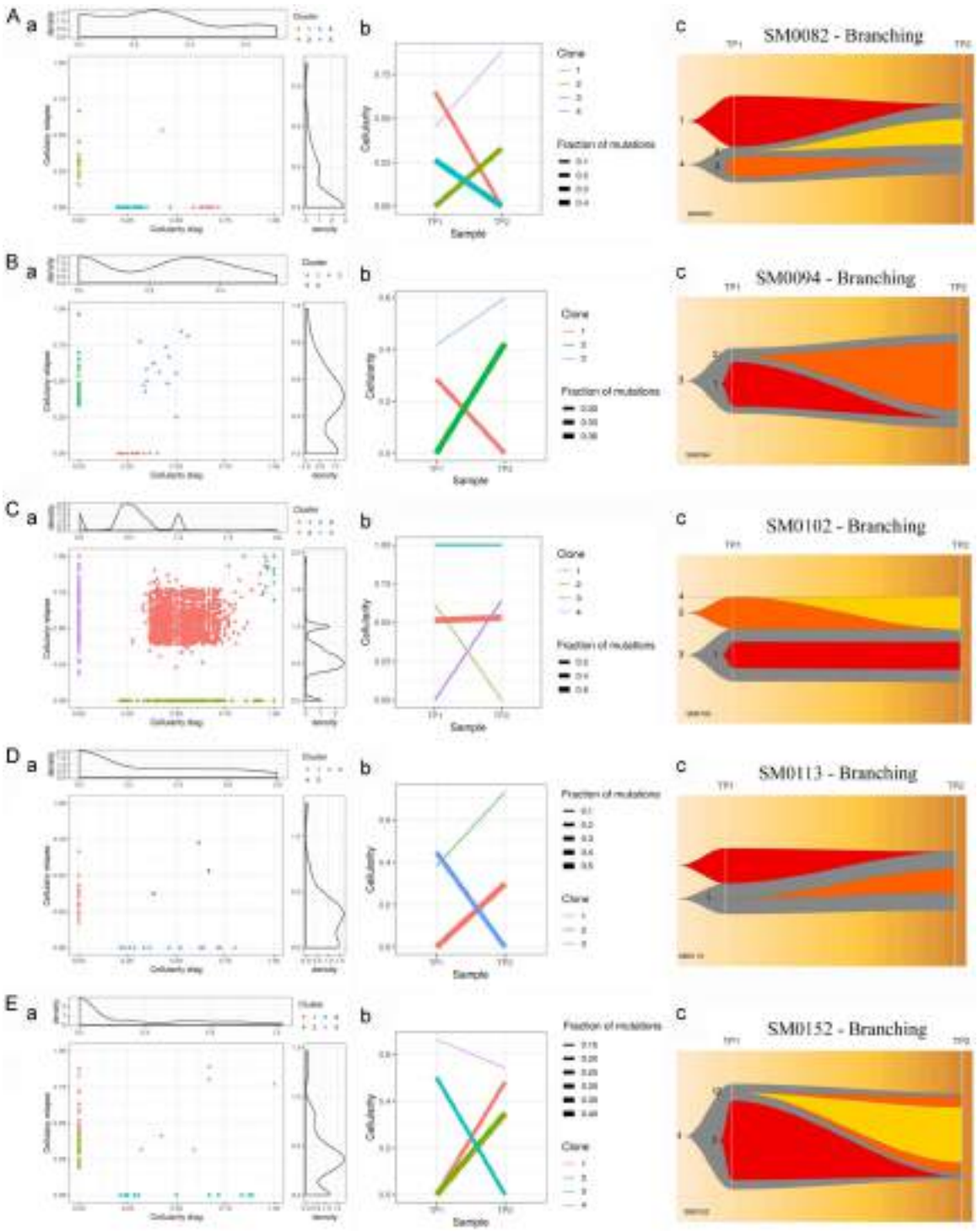
[50] Neri P, Ren L, Gratton K, Stebner E, Johnson J, Klimowicz A, Duggan P, Tassone P, Mansoor A, Stewart DA, Lonial S, Boise LH and Bahlis NJ. Bortezomib-induced "BRCAness" sensitizes multiple myeloma cells to PARP inhibitors. Blood 2011; 118: 6368-6379.

[51] Ino R, Saitoh T, Kitamura Y, Homma K, Takahashi N, Nanami G, Kasamatsu T, Shimizu H, Takizawa M, Matsumoto M, Sawamura M, Yokohama A, Tsukamoto N, Handa H and Murakami H. The role and therapeutic target of base excision repair genes in multiple myeloma (MM). Blood 2017; 130: 4403.

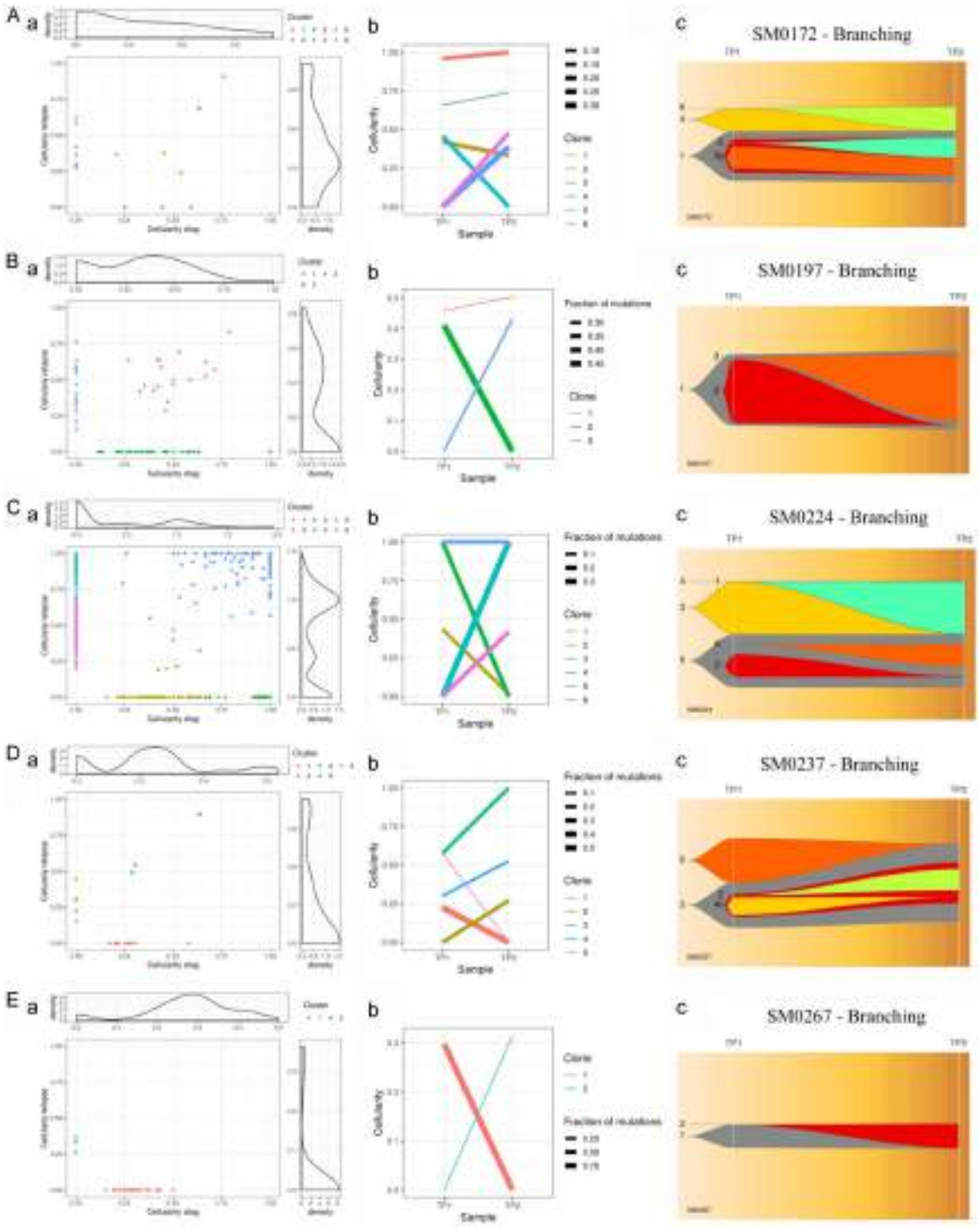
[52] Farswan A, Gupta A, Gupta R, Hazra S, Khan S, Kumar L and Sharma A. AI-supported modified risk staging for multiple myeloma cancer useful in real-world scenario. Transl Oncol 2021; 14: 101157.



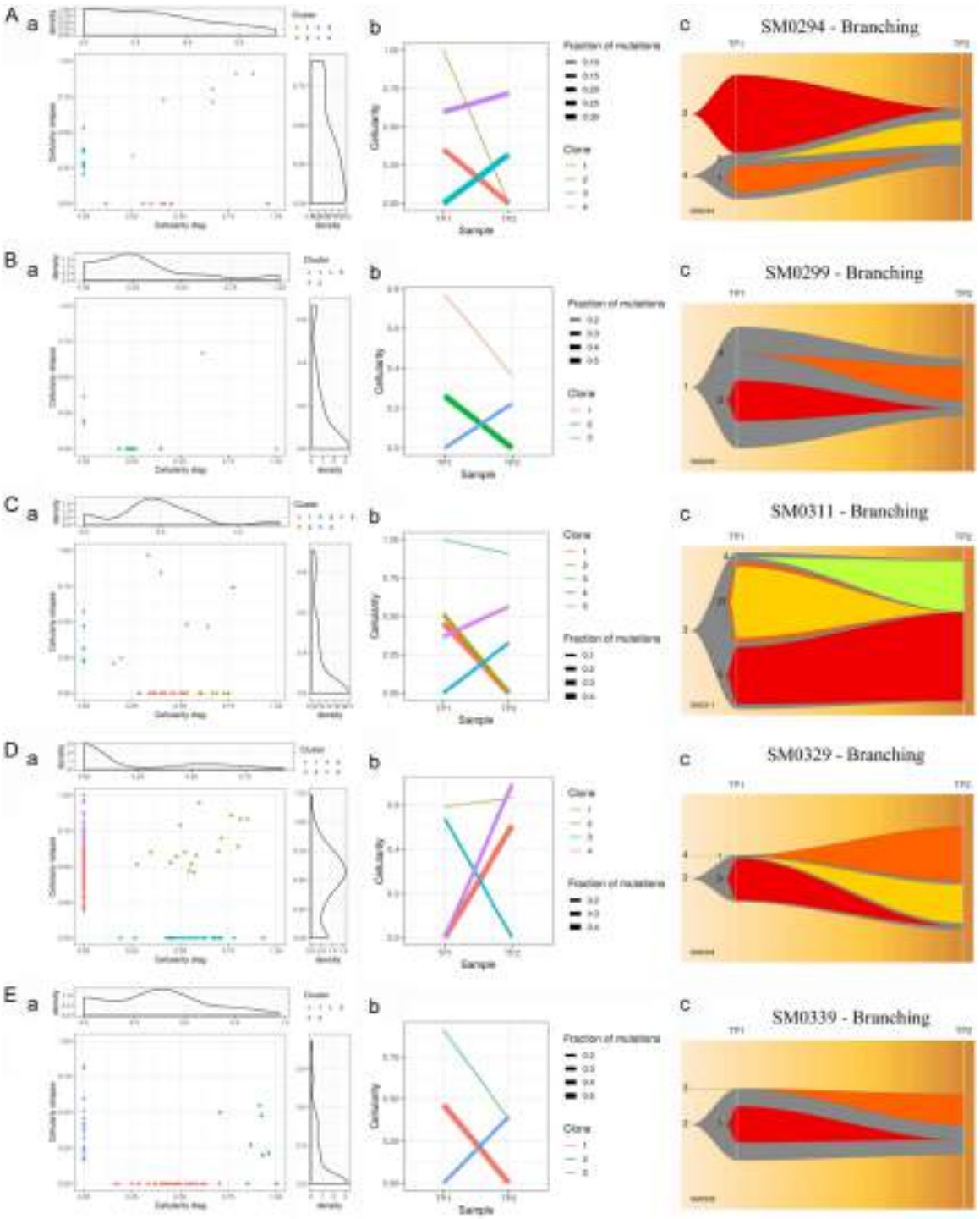

**Supplementary Figure 1.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



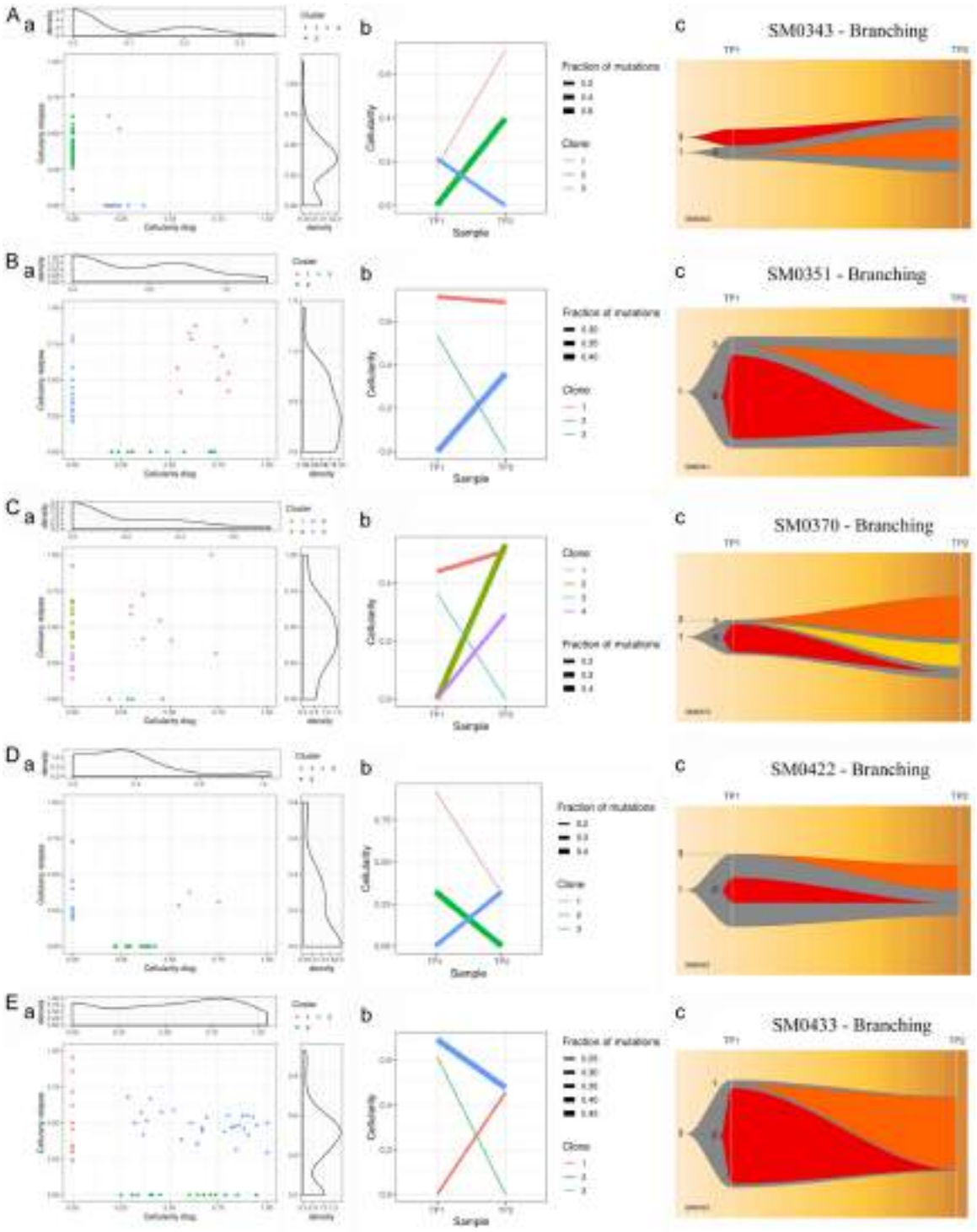

**Supplementary Figure 2.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



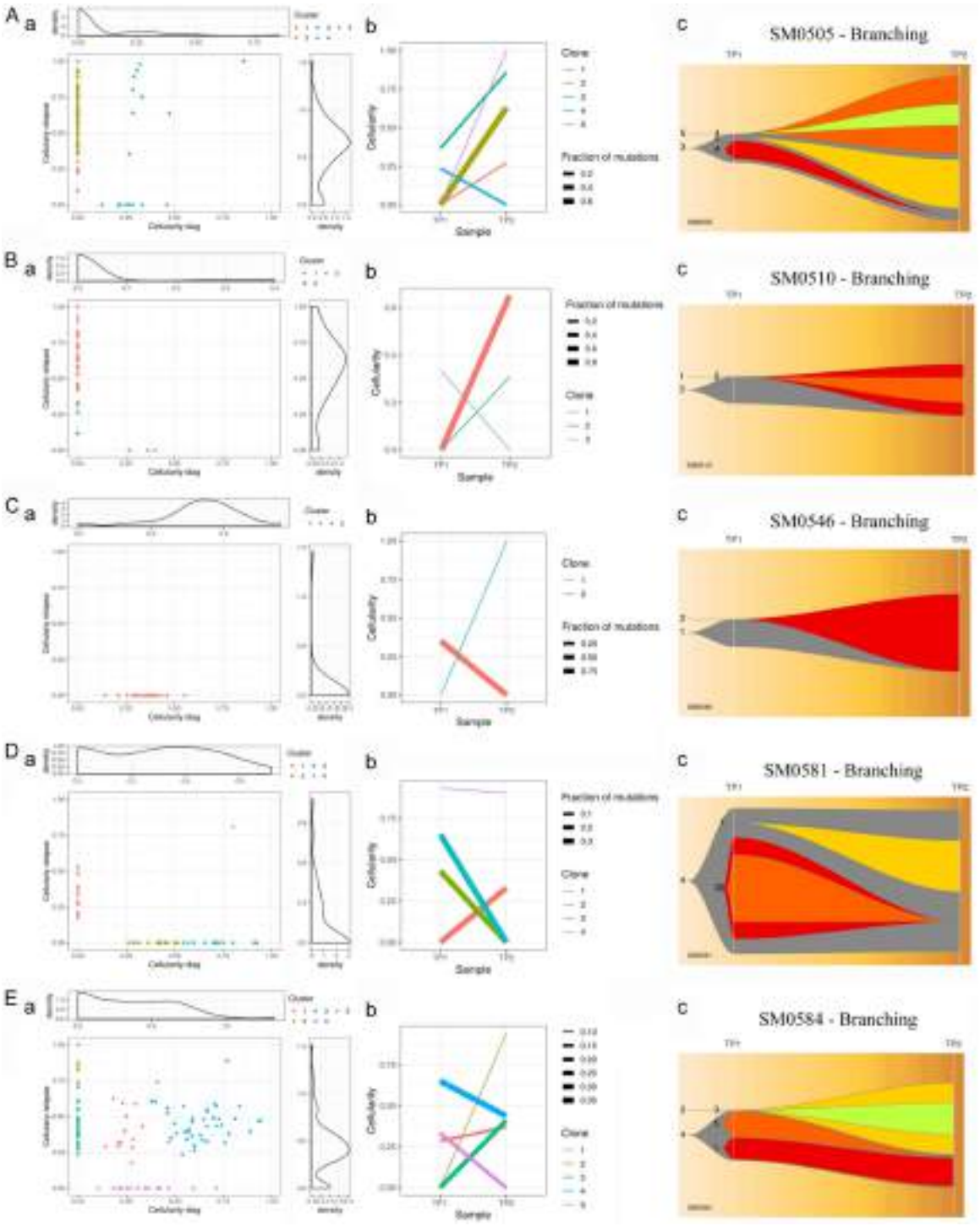

**Supplementary Figure 3.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



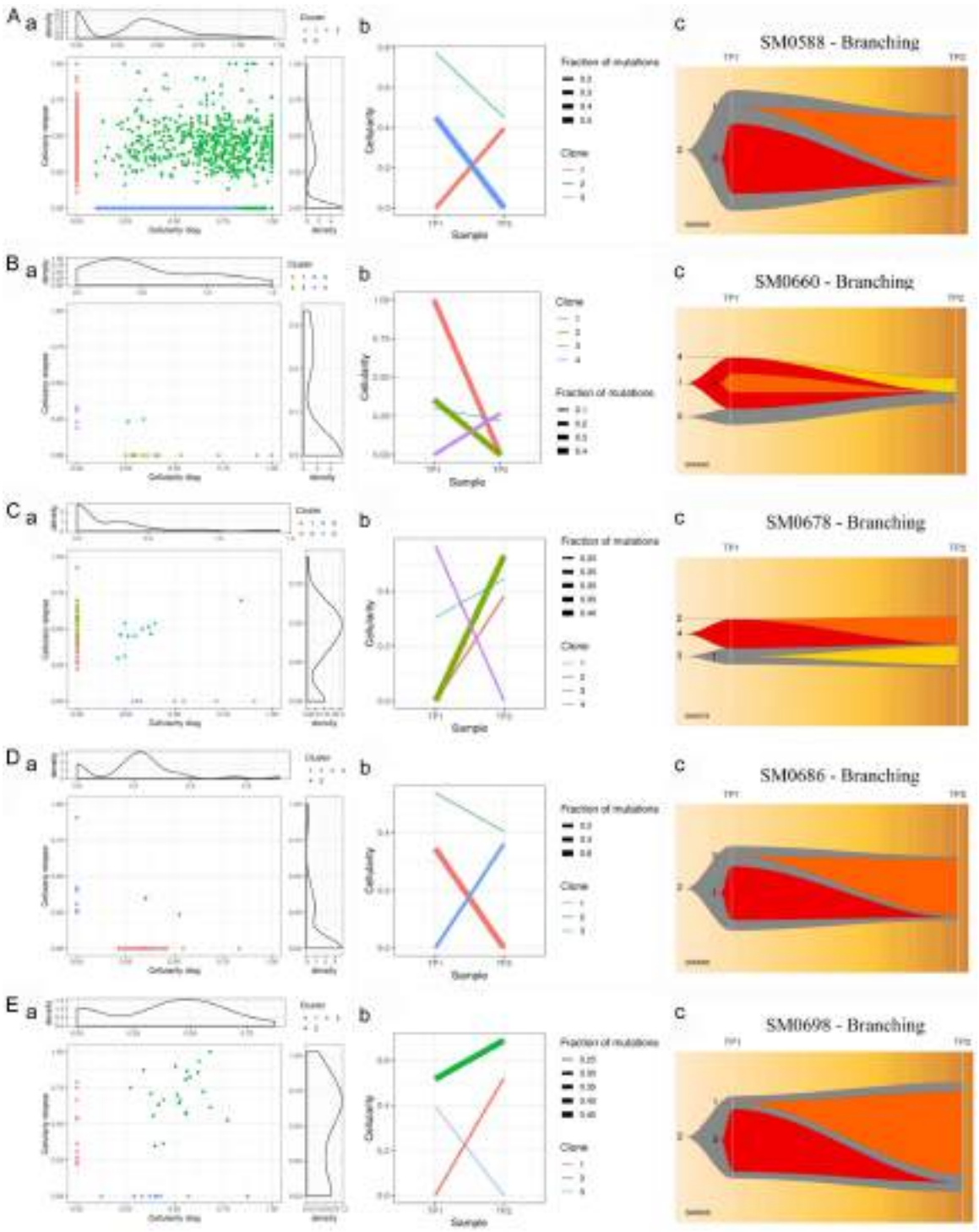

**Supplementary Figure 4.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



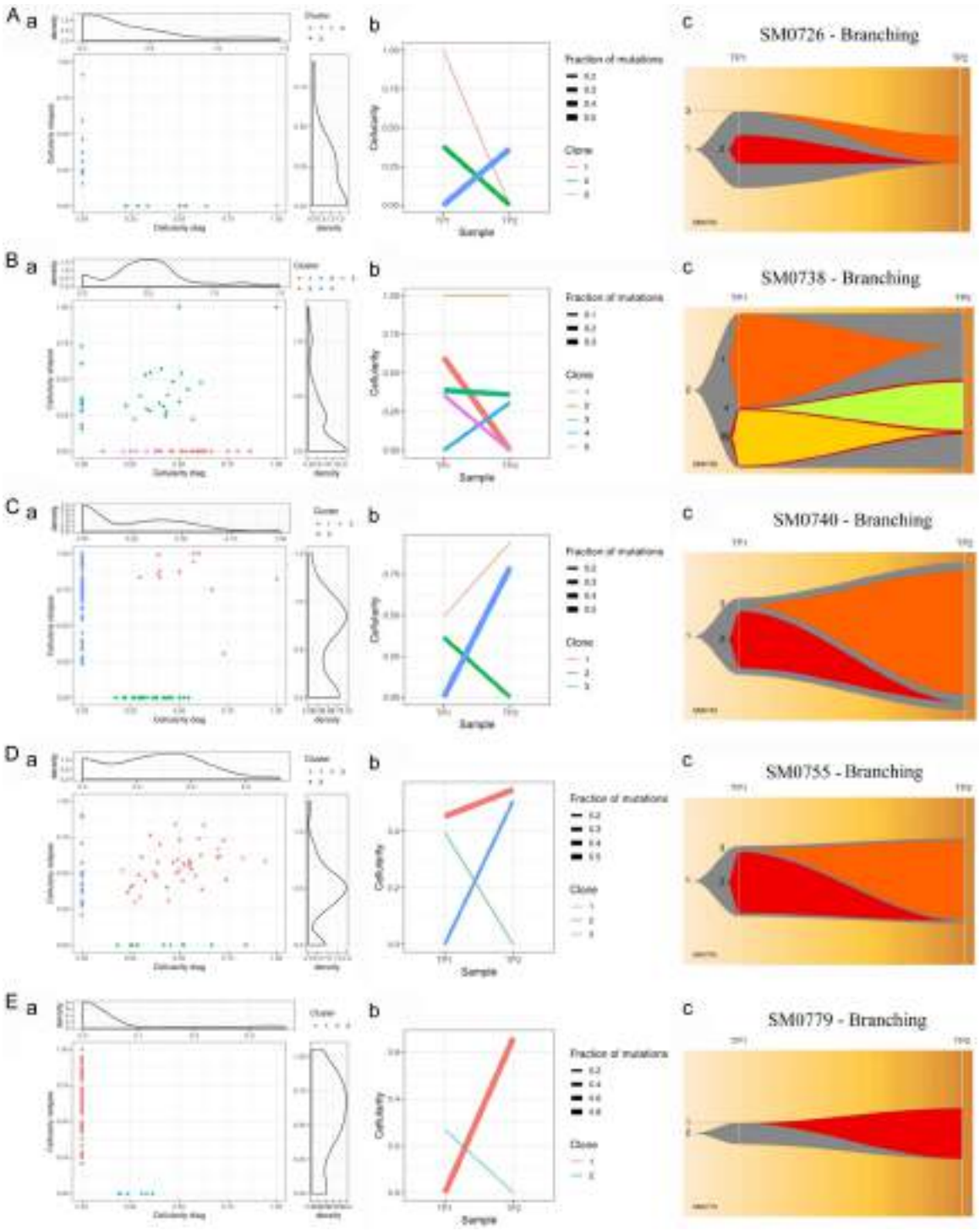

**Supplementary Figure 5.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



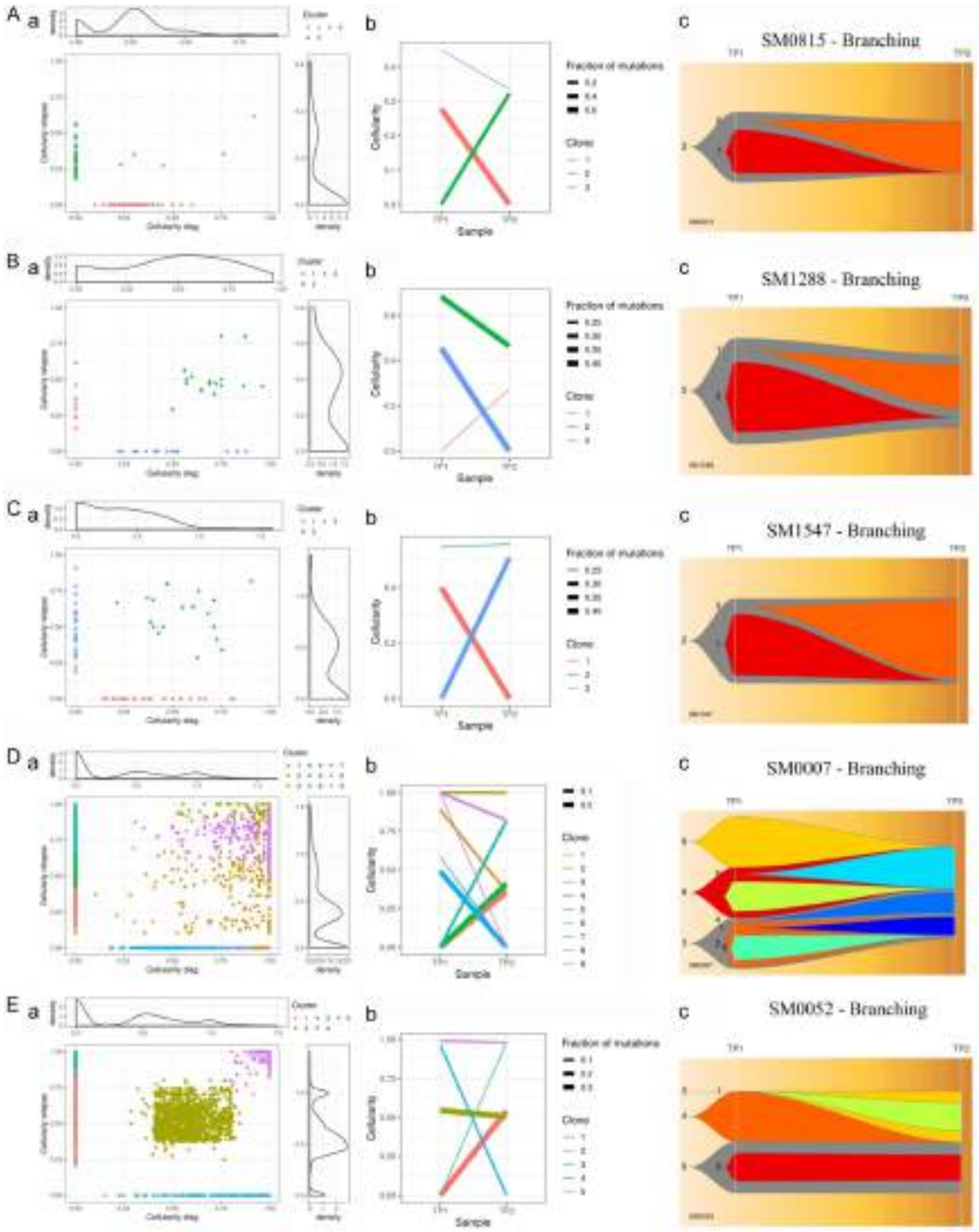

**Supplementary Figure 6.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



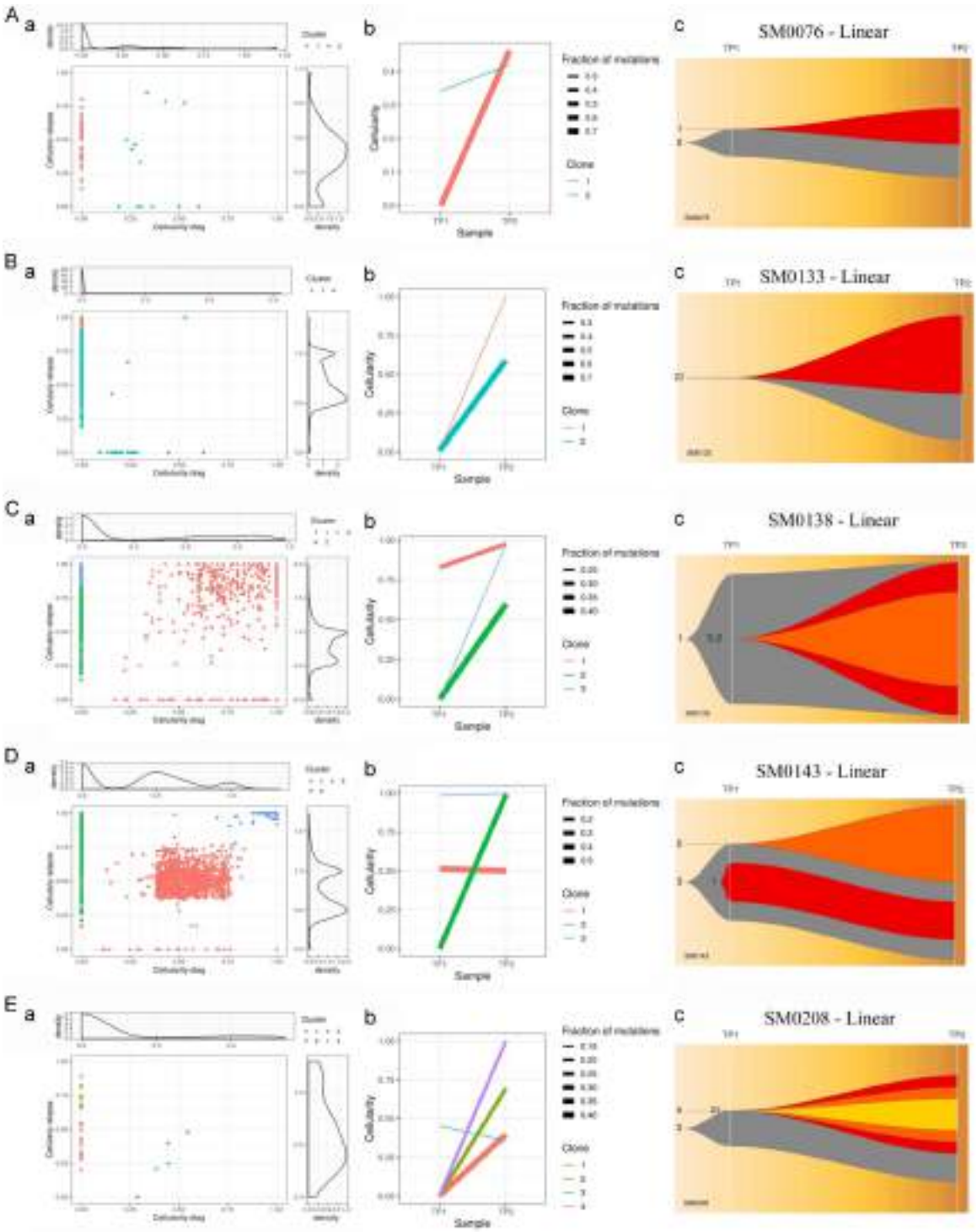

**Supplementary Figure 7.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



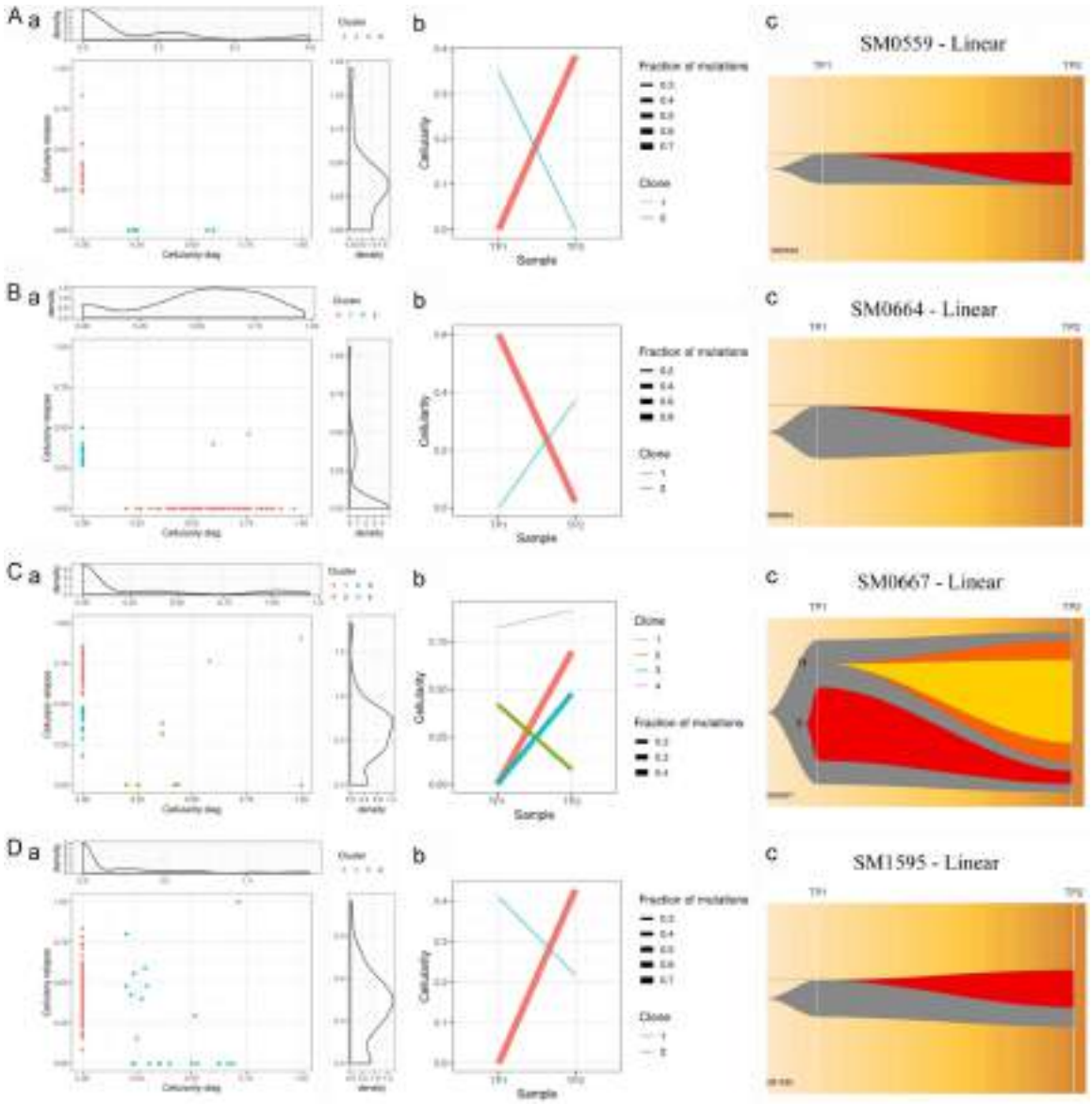

**Supplementary Figure 8.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



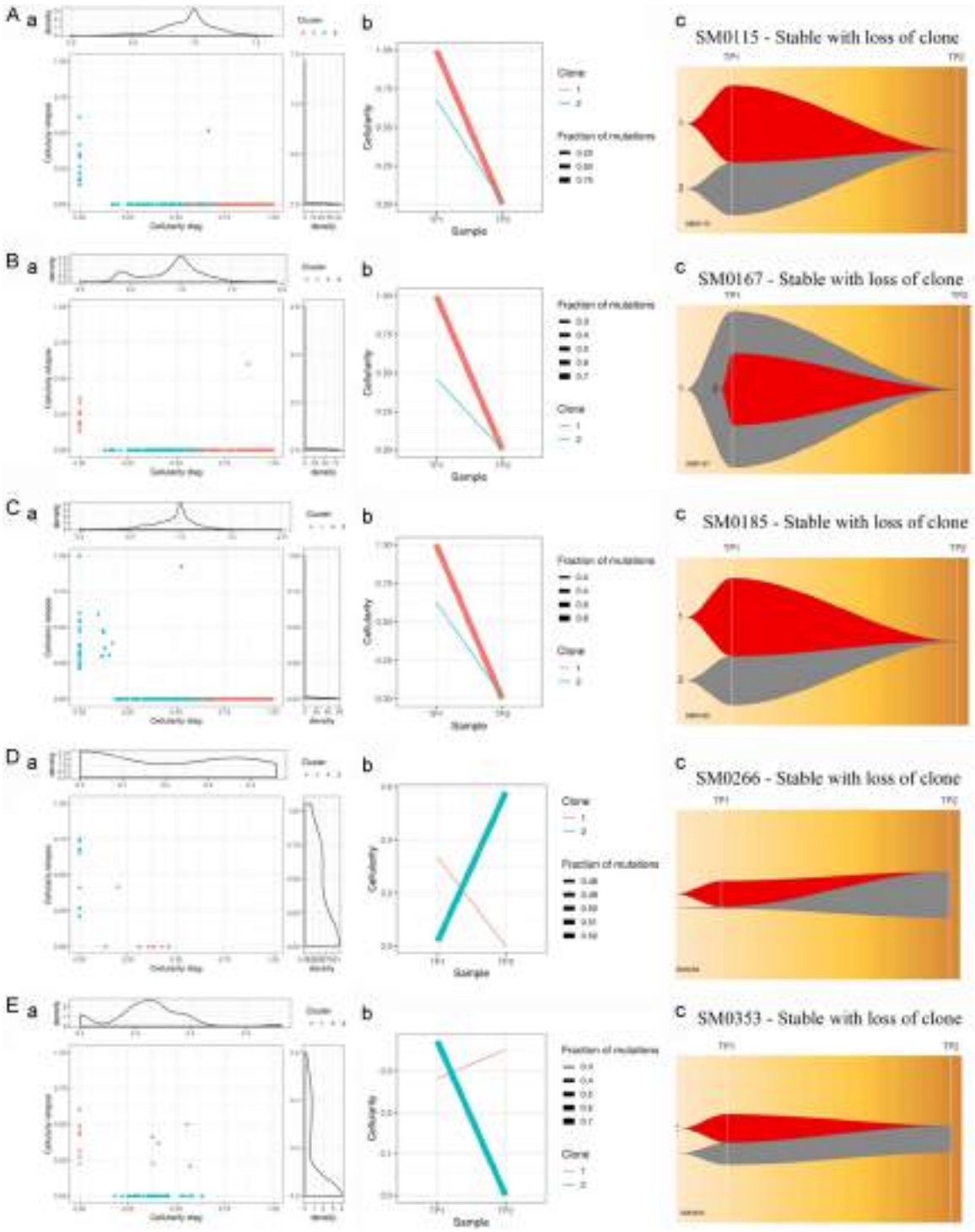

**Supplementary Figure 9.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).



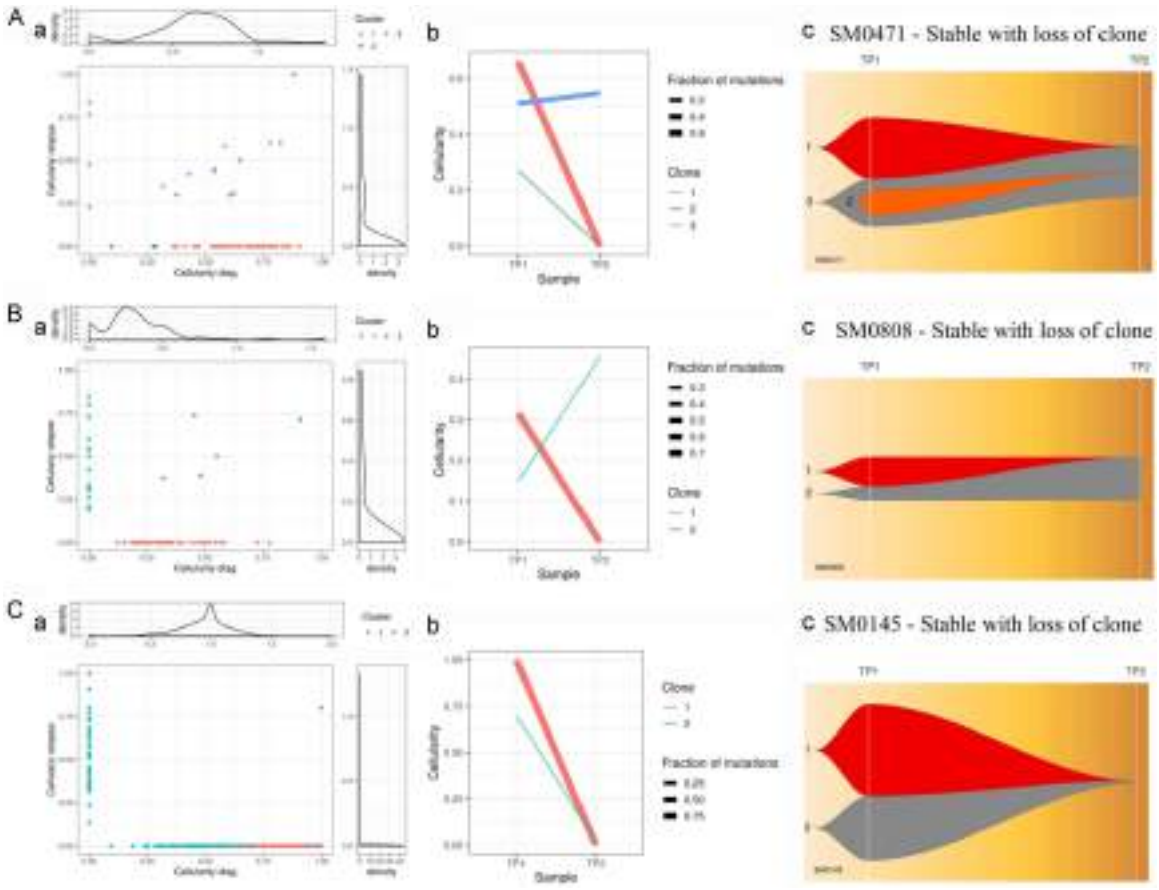

**Supplementary Figure 10.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).




**Supplementary Figure 11.** (A-D) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).




**Supplementary Figure 12.** (A-E) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).




**Supplementary Figure 13.** (A-C) Clonal evolution in each case of MM. Representation of clonal evolution through (a) Density, (b) Evolution and (c) Fish plots across individual MM patients with branching, linear and stable with loss of clone patterns of clonal evolution. Supplementary Note 1 (Casewise clonal evolution).

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

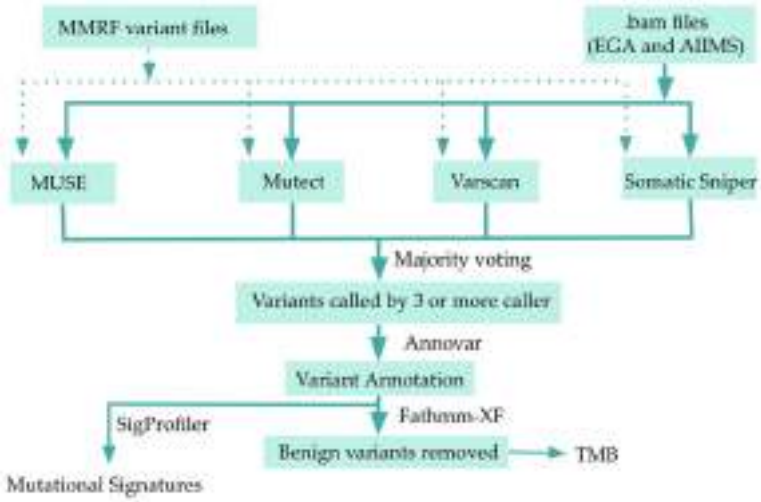



**Figure 1.** Workflow of the study and data analysis. Four different variant callers were used to identify variants in the MM and MGUS patients. Variants were finalized using the majority voting scheme. Variants were then annotated with Annovar for deducing TMB. Mutational signatures were inferred using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

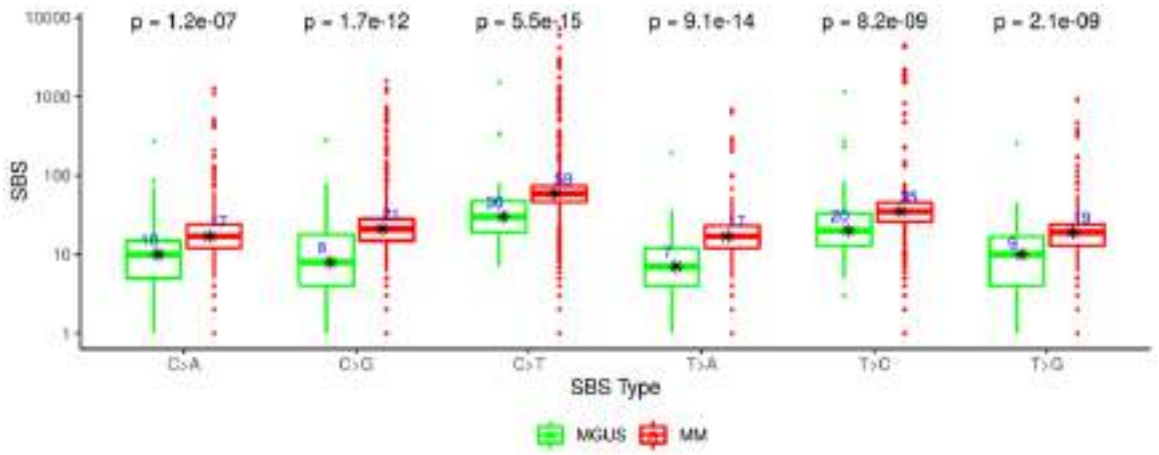



**Figure 2.** Boxplot shows the difference in the frequency of the single base substitutions between MGUS and MM patients. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not. For all the substitutions, there is significant variation in the frequency with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

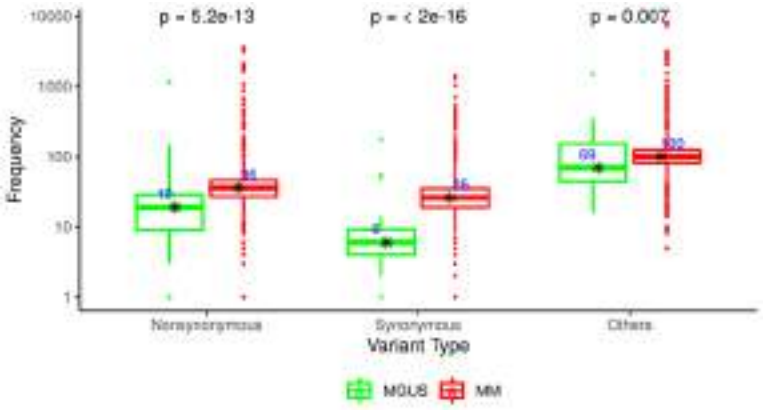



**Figure 3.** Boxplot showing the variation in the frequency of the three different categories of variants-Nonsynonymous, Synonymous, and Others between MGUS and MM. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not. There was a statistically significant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

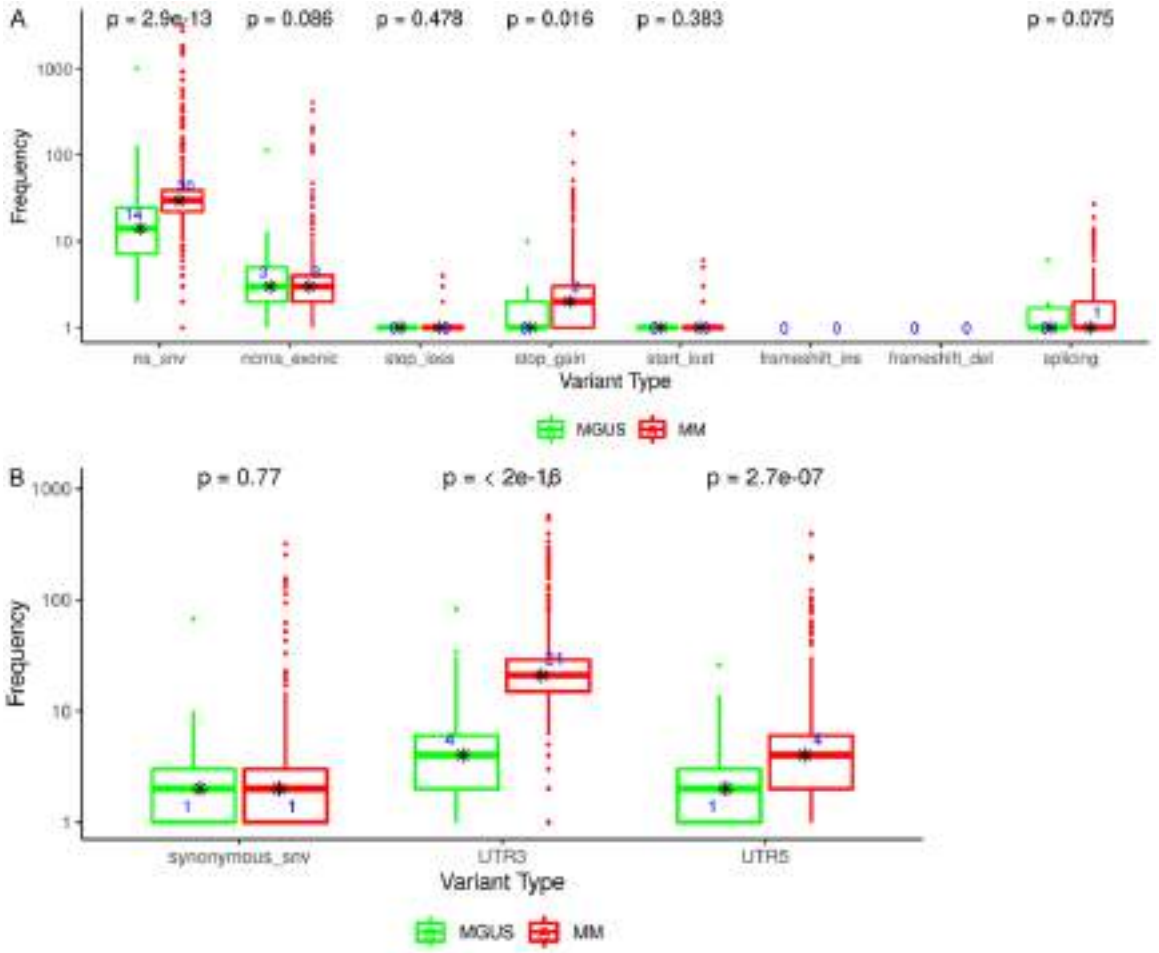



**Figure 4.** A. Boxplot showing the variation in the frequency of the variants under the nonsynonymous category. There was a statistically significant variation in the frequency of nonsynonymous_snv and stop_gain variants with *p*-values less than 0.05. B. Boxplot showing the variation in the frequency 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.

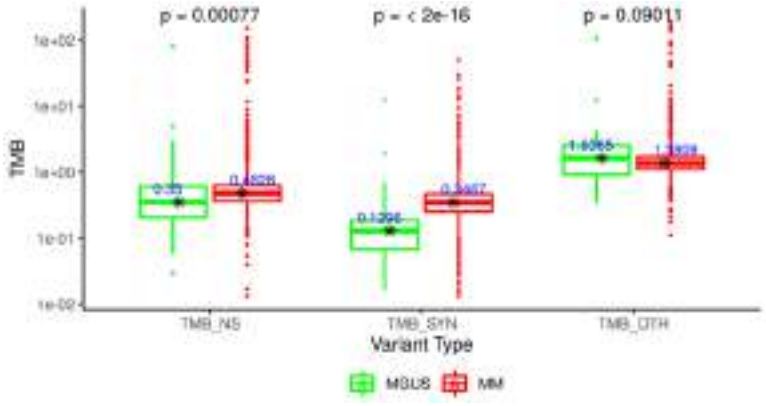



**Figure 5.** Boxplot reveals that the difference in the low TMB and high TMB groups is statistically significant with *p*-values less than 0.05 for TMB_NS and TMB_SYN. Wilcoxon rank-sum test was applied to determine if the change is statistically significant or not.

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.

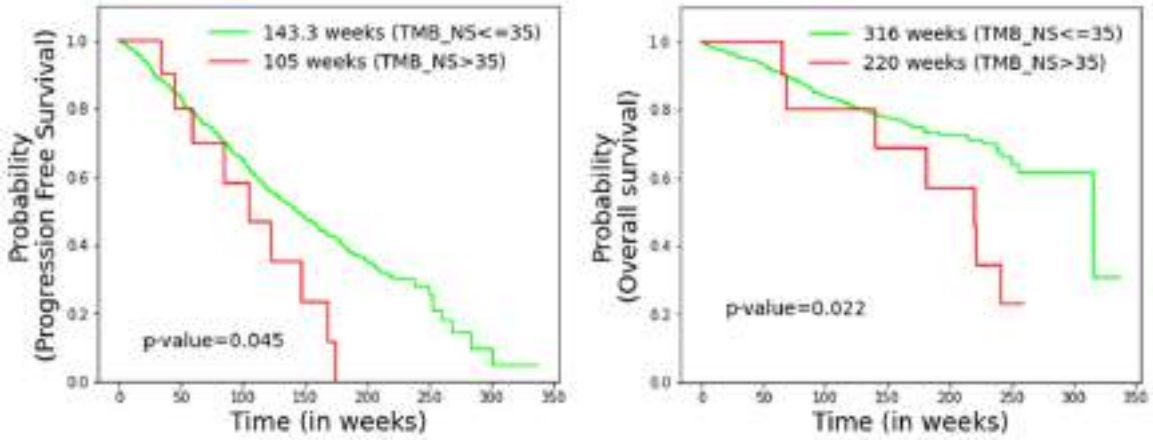



**Figure 6.** High TMB is associated with poor overall survival in NDMM patients. The difference in the overall survival probability between low and high TMB_NS is statistically significant with *p*-values 0.045 and 0.022 for PFS and OS respectively.

**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.

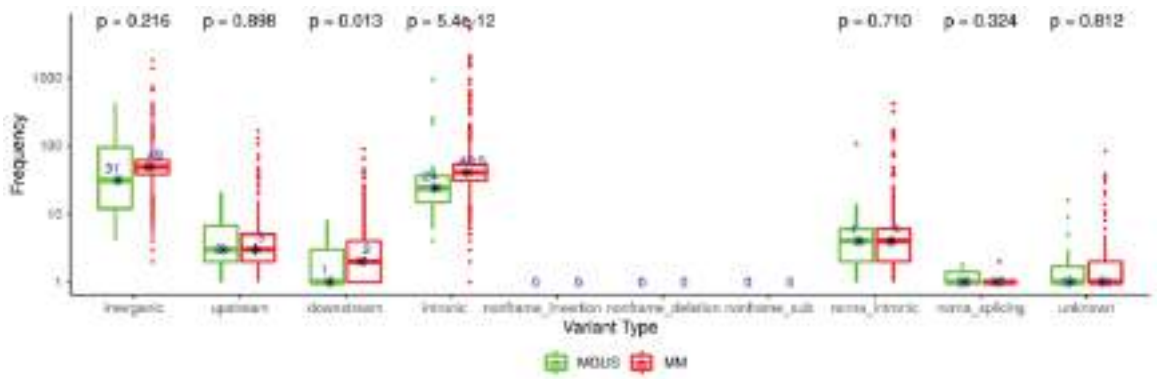
**Supplementary Table 2.** The table shows the cut-offs obtained for TMB_NS, TMB_SYN and TMB_OTH via Cutoff Finder

| | 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 |

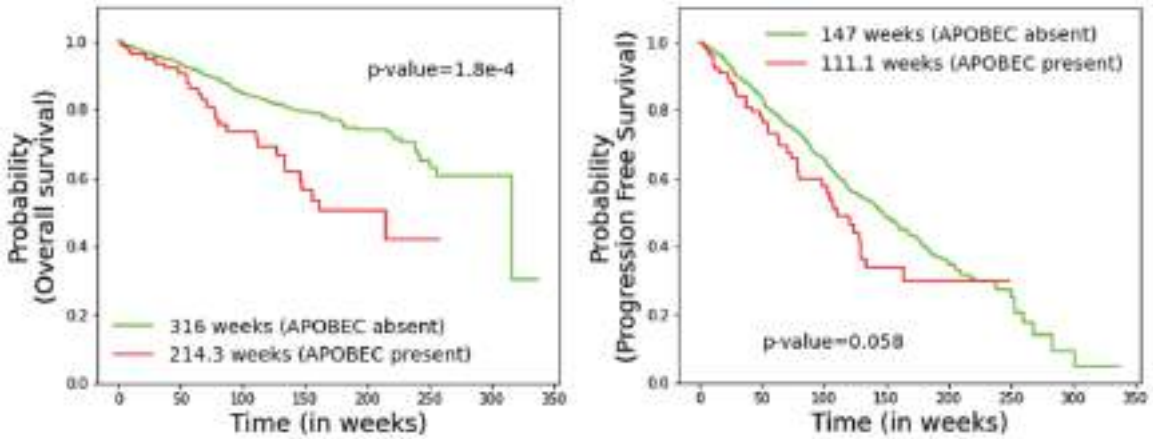
Similar cut-offs were obtained using PFS and OS. KM analysis revealed a significant difference (*p*-value <0.05) on the survival curves of the patients below and above the selected cut-offs.

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

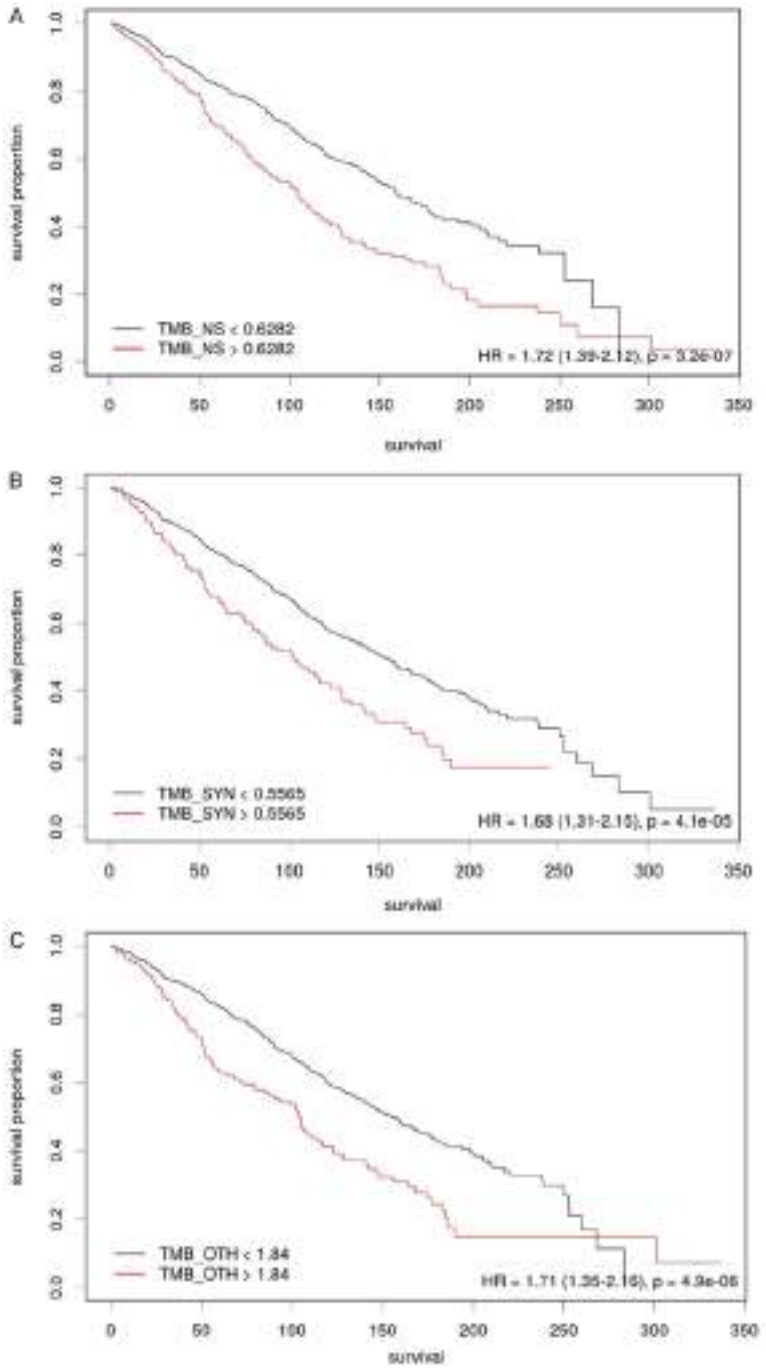
| SBS | Min | Median | Max | PFS cutoff | OS cutoff | PFS *p*-value | OS *p*-value |
|---|---|---|---|---|---|---|---|
| C>A | 0 | 17 | 1251 | 26.5 | 28.5 | 1.9E-5 | 5.1E-06 |
| C>G | 0 | 21 | 1575 | 3.5 | 34.5 | 0.027 | 8.6E-05 |
| C>T | 1 | 59 | 7315 | 79.5 | 110 | 0.00055 | 5.8E-07 |
| T>A | 0 | 17 | 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 |



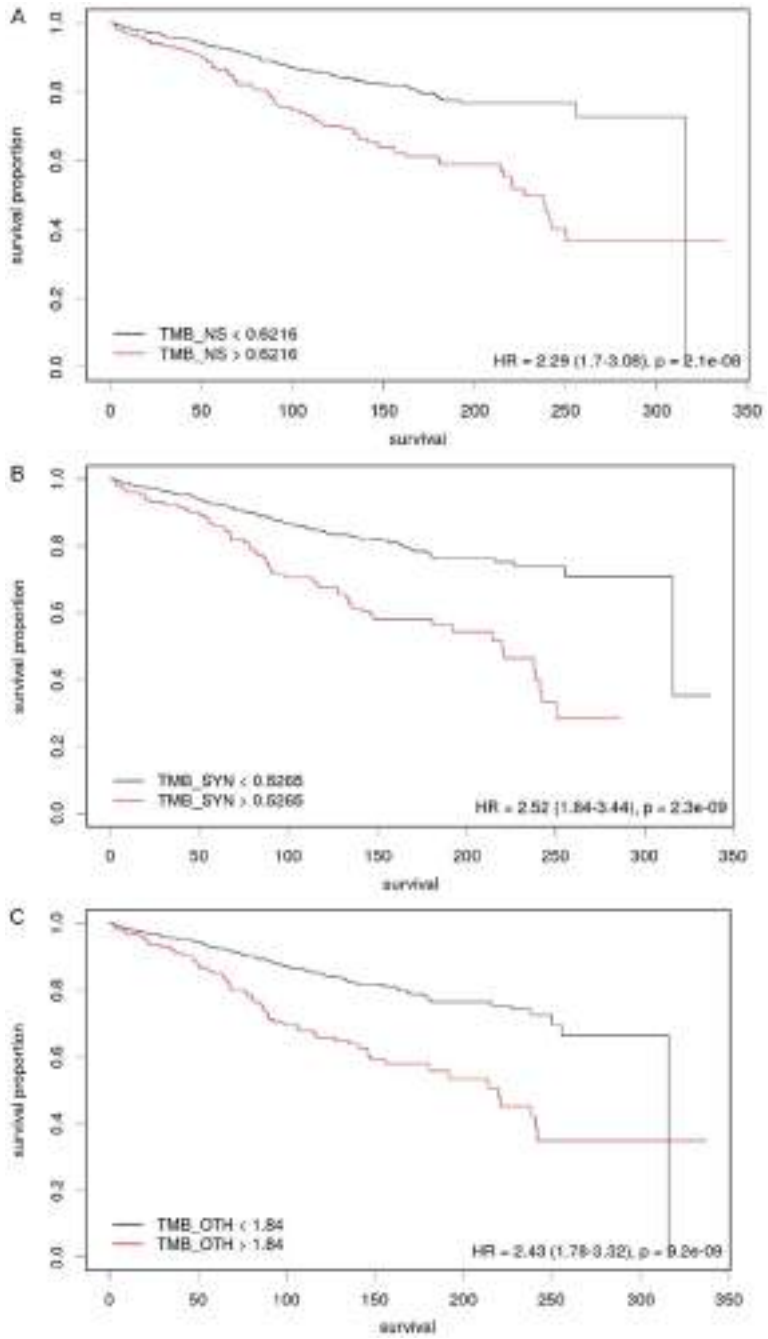

**Supplementary Figure 1.** Boxplot showing the variation in the frequency of the variants under the other variants category. There was a statistically significant rise in the frequency of intronic, intergenic, and downstream variants with *p*-values less than 0.05.



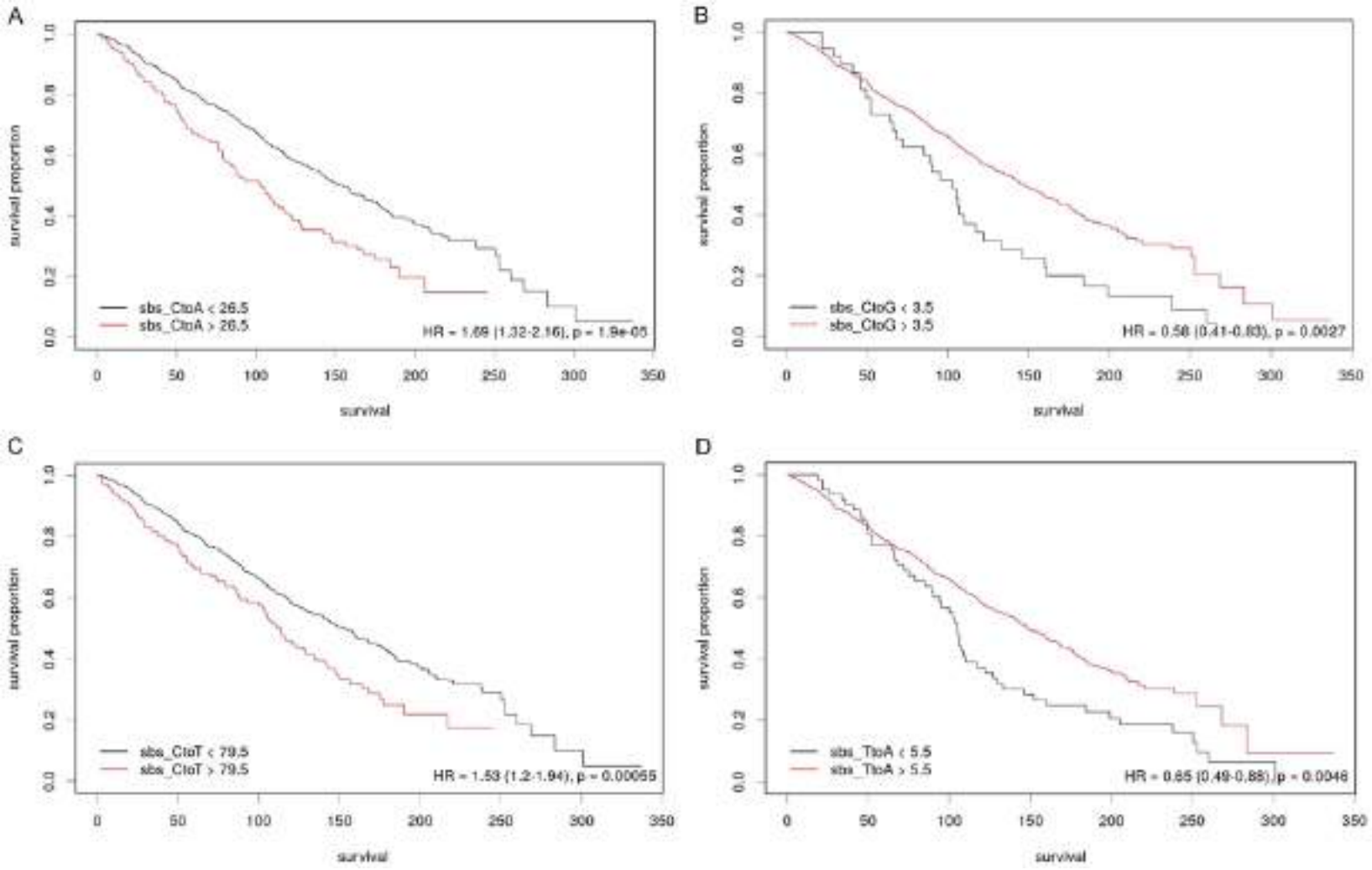

**Supplementary Figure 2.** KM curves reveal that APOBEC activity is associated with poor overall survival in NDMM patients. The difference in the overall survival probability between low and high TMB_NS is statistically significant with *p*-values 1.8e-4. However, there is no statistically significant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

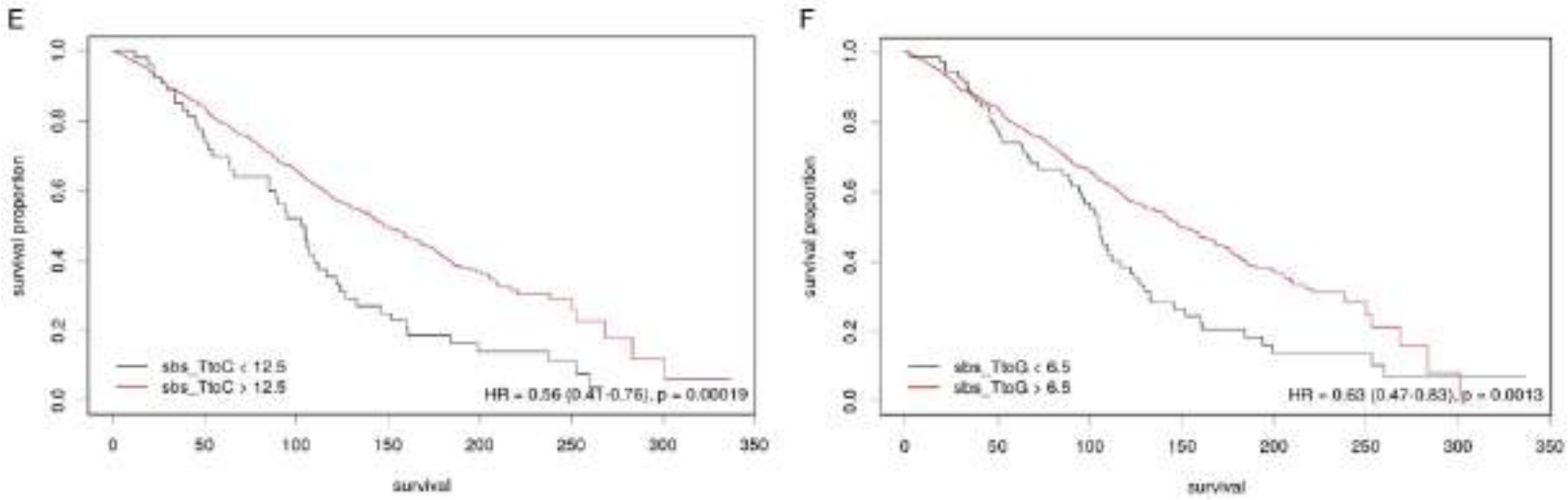



**Supplementary Figure 3.** KM curves reveal significant differences in the PFS survival patterns of (A) TMB_NS, (B) TMB_SYN and (C) TMB_OTH at the thresholds obtained via Cutoff Finder.

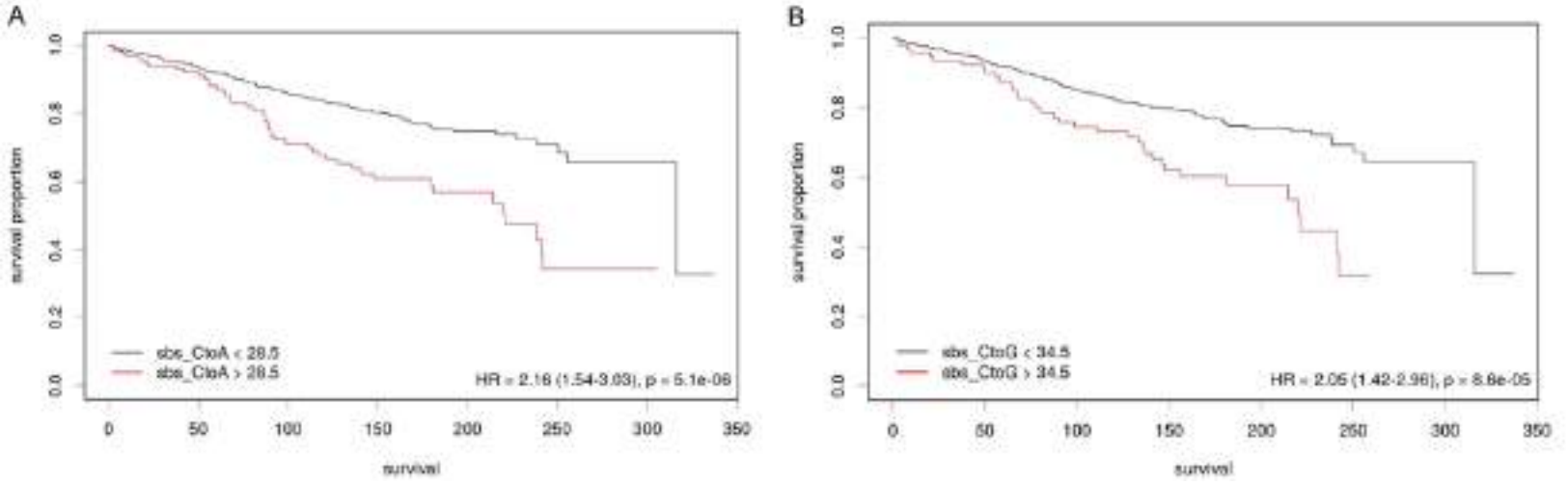


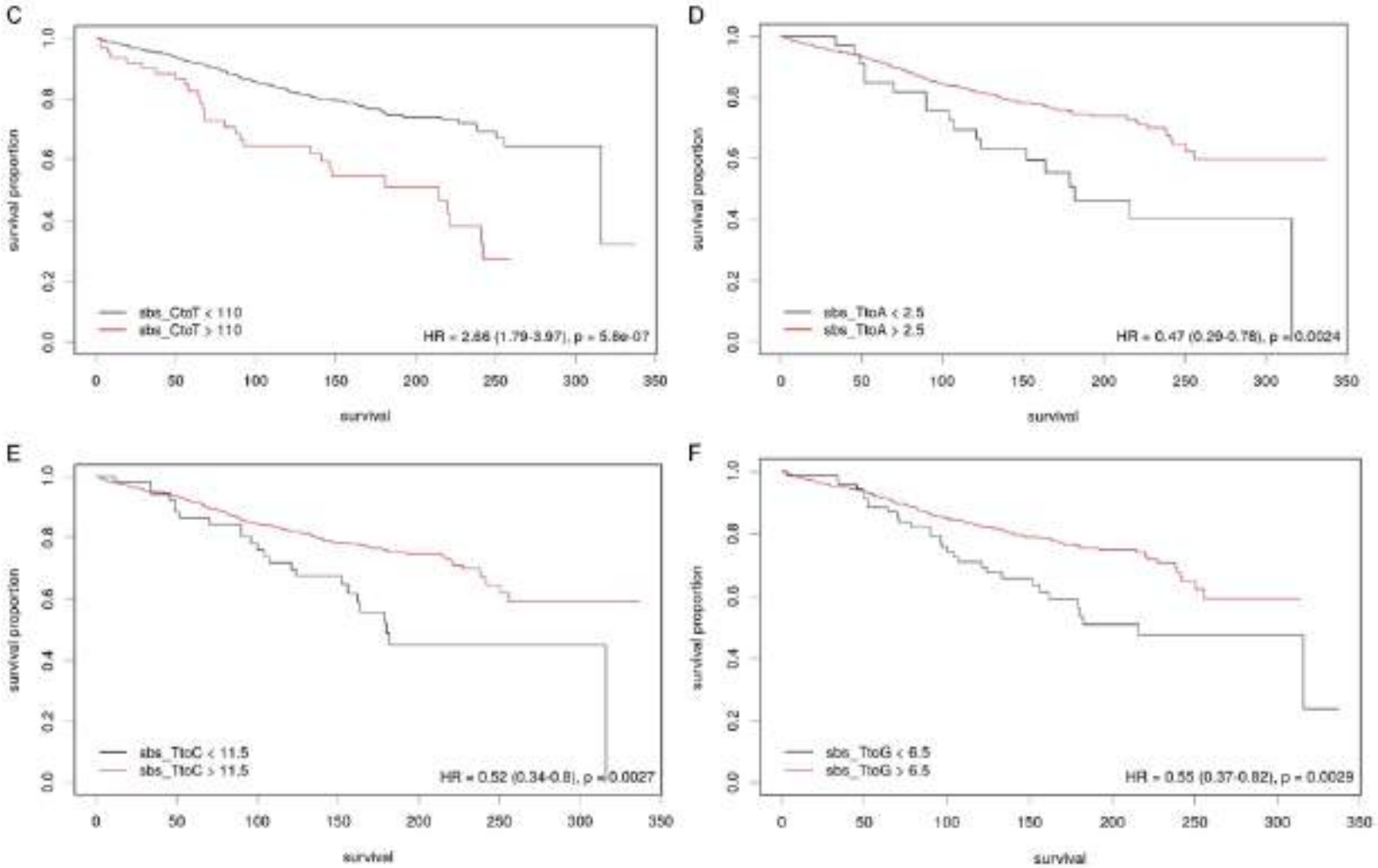

**Supplementary Figure 4.** KM curves reveal significant differences in the OS survival patterns of (A) TMB_NS, (B) TMB_SYN and (C) TMB_OTH at the thresholds obtained via Cutoff Finder.


A
survival proportion
survival
sbs_CtoA < 26.5
sbs_CtoA > 26.5
HR = 1.69 (1.32-2.16), p = 1.9e-05
B
survival proportion
survival
sbs_CtoG < 3.5
sbs_CtoG > 3.5
HR = 0.58 (0.41-0.83), p = 0.0027
C
survival proportion
survival
sbs_CtoT < 79.5
sbs_CtoT > 79.5
HR = 1.53 (1.2-1.94), p = 0.00055
D
survival proportion
survival
sbs_TtoA < 5.5
sbs_TtoA > 5.5
HR = 0.65 (0.49-0.88), p = 0.0046




**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.