



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

의학박사 학위논문

Identification of Genomic Profiling of Multiple Myeloma Patients in Korea

한국인 다발골수종 환자의 유전체변이 분석

2021년 8월

서울대학교 대학원
의학과 검사의학 전공
이 누리

A thesis of the Degree of Doctor of Philosophy

한국인 다발골수종 환자의
유전체변이 분석

Identification of Genomic Profiling of Multiple Myeloma
Patients in Korea

August 2021

Department of Laboratory Medicine

Seoul National University

College of Medicine

Nuri Lee

한국인 다발골수종 환자의
유전체변이 분석

지도교수 이 동 순

이 논문을 의학박사 학위논문으로 제출함

2021 년 4 월

서울대학교 대학원

의학과 검사의학 전공

이 누 리

이누리의 박사학위논문을 인준함

2021 년 7 월

위 원 장 _____

부 위 원 장 _____

위 원 _____

위 원 _____

위 원 _____

ABSTRACT

Identification of Genomic Profiling of Multiple Myeloma Patients in Korea

Nuri Lee

Department of Medicine

The Graduate School

Seoul National University

To investigate the prognostic value of gene variants and copy number variations (CNVs) in patients with newly diagnosed multiple myeloma (NDMM), an integrative genomic analysis was performed using conventional cytogenetics, fluorescent in situ hybridization, and whole-exome sequencing. Sixty-seven patients with NDMM exhibiting more than 60% plasma cells in the bone marrow aspirate were enrolled in the study. Whole-exome sequencing was performed on bone marrow nucleated cells. Mutation and CNV analyses were performed using the CNVkit and Nexus Copy Number software. Eighty-three driver gene mutations were detected in 63 patients with NDMM. The median number of mutations per patient was 2.0 (95% confidence interval [CI] = 2.0–3.0, range = 0–8). *MAML2* and *BHLHE41* mutations were associated with decreased

survival. CNVs were detected in 56 patients (72.7%; 56/67). The median number of CNVs per patient was 6.0 (95% CI = 5.7–7.0; range = 0–16). Among the CNVs, 1q gain, 6p gain, 6q loss, 8p loss, and 13q loss were associated with decreased survival. Additionally, 1q gain and 6p gain were independent adverse prognostic factors. Increased numbers of CNVs and driver gene mutations were associated with poor clinical outcomes. Cluster analysis revealed that patients with the highest number of driver mutations along with 1q gain, 6p gain, and 13q loss exhibited the poorest prognosis. In addition to the known prognostic factors, the integrated analysis of genetic variations and CNVs could contribute to prognostic stratification of patients with NDMM.

.....

Keyword : multiple myeloma, copy number variation, somatic mutation

Student Number : 2018–39099

CONTENTS

Abstract.....	i
Contents.....	iii
List of tables and figures.....	iv
List of abbreviation.....	vii
1. Introduction.....	1
2. Materials and Methods.....	4
2.1. Patients.....	4
2.2. DNA extraction and exome sequencing.....	5
2.3. Bioinformatic evaluation of sequencing data.....	6
2.4. Copy number analysis.....	9
2.5. Statistical analysis.....	11
3. Results.....	12
3.1. Characteristics of the study population.....	12
3.2. Identification of CNVs in MM.....	14
3.3. Analysis of selected driver gene mutations.....	18
3.4. Correlation analysis of genomic variants.....	26
3.5. Prognostic impact of CNVs and somatic mutations.....	29
3.6. Factors affecting OS in patients with hyperdiploid NDMM.....	40
3.7. OS of clusters classified based on K-means analysis.....	42
4. Discussions.....	45
References.....	52
Abstract in Korean.....	61

LISTS OF TABLES

Table 1. List of driver genes including the study

Table 2. Demographic and clinical characteristics of the study population

Table 3. Frequency of copy number alterations in patients with newly diagnosed multiple myeloma

Table 4. Frequency of chromosome gain and loss including cytoband information

Table 5. Cox proportional hazards model for factors associated with overall survival.

Table 6. Results of copy number variations and rearrangements by WES, FISH and CG in patients with NDMM

LISTS OF FIGURES

Figure 1. Examples of copy number alteration in one patient.

Figure 2. Summary plot of chromosomal gain and loss in 67 patients with newly diagnosed multiple myeloma determined using whole-exome sequencing and copy number analysis.

Figure 3. Frequency and distribution of somatic mutations in patients with NDMM. Mutation burden per megabase (Mb) in tumor are represented (upper) and mutated genes are ranked by mutant frequency (lower).

Figure 4. Diagrams for mutations for most frequent genes in NDMM patients. (A)–(G) is diagrams for *BRAF*, *BHLHE41*, *KRAS*, *NRAS*, *IGLL5*, *ATM*, and *MAML2* respectively.

Figure 5. Pairwise associations between mutations and copy number variations.

Figure 6. Kaplan–Meier analysis of overall survival in patients with copy number variations.

Figure 7. Kaplan–Meier analysis of overall survival in patients with mutations in the driver genes.

Figure 8. Kaplan–Meier survival curves of patients with copy number variations identified using fluorescent in situ hybridization (FISH), conventional cytogenetics (CG), and whole-exome sequencing (WES) analysis.

Figure 9. Distribution of the number of mutated driver genes in each patient.

Figure 10. Kaplan–Meier curves showing the overall survival of patients with hyperdiploid multiple myeloma (MM).

Figure 11. Kaplan–Meier curves for overall survival (OS) with subdivided driver groups.

LISTS OF ABBREVIATIONS

BM	Bone marrow
BWA	Burrows–Wheeler aligner
CG	Conventional cytogenetics
CNVs	Copy number variations
FISH	Fluorescent in situ hybridization
HR	Hazard ratio
HRD	Hyperdiploid
IMWG	International myeloma working group
ISS	International staging system;
LDH	Lactate dehydrogenase
MM	Multiple myeloma
NDMM	Newly diagnosed multiple myeloma
OS	Overall survival
PCC	Pearson correlation coefficient
PCR	Polymerase chain reaction
PFS	Progression–free survival
qRT–PCR	Quantitative real–time PCR
WES	Whole exome sequencing

1. INTRODUCTION

Multiple myeloma (MM) is the second most common hematological malignancy in Korea and the USA [1–3]. The incidence of MM in Korea has increased tenfold in the past 20 years and doubling in the last ten years [1]. The treatment outcomes of patients with MM have markedly improved owing to the development of novel therapeutic agents and the technical advances in molecular diagnostics and detection of minimal residual diseases [4, 5]. There is a consensus on the risk factors that aid in stratifying patients with MM. The methods to detect genetic variations include conventional karyotyping and fluorescent in situ hybridization (FISH) [6, 7].

Recently, the genetic landscape of MM has expanded owing to the use of high throughput technologies, such as next generation sequencing for analyzing targeted genes and whole exome or genome [8–11]. Within the last five years, several next generation sequencing studies targeting large number of MM patients have been conducted. Niccolo B. et al. performed next generation sequencing and whole exome sequencing on 418 patients [9], and Brain A. W et al. also conducted whole exome sequencing on 1,273 patients [10]. Previous studies have revealed novel and frequently mutated genes, such as *KRAS*, *NRAS*, *BRAF*, and *FAM46C* in MM. Also, attempts have been made to find driver mutations among several MM related genes using various algorithms through frequency and/or functional based approaches. These studies found that

63 and 61 driver genes, respectively, would be significant in MM patients [8, 10]. Additionally, the chronological evolution of multiple driver events has been demonstrated using serial patient specimens [8].

However, the prognostic effect of most mutated genes, which have a low recurrence rate, has not been clearly identified [12]. There are limited studies on the comprehensive analysis of various cancer driver events and the correlation between structural variants and genomic events. Moreover, in most of the previous large cohort studies, the proportion of Asian patients was only about 2~4%. Such a low ethnical composition had limitations in analyzing the characteristics of genomic variations that reflected characteristics of Asians MM patients. For this reason, the differences between somatic mutations and copy number variations (CNVs) by ethnicity have not been clearly clarified. In particular, a few studies have investigated genomic features of Korean newly diagnosed multiple myeloma (NDMM). In case of CNVs, single nucleotide polymorphism array chip [13] and conventional karyotyping [14] analysis was utilized to investigate the frequency and prognosis of structural variants in Korean MM patients. However, researches through massive sequencing are limited to a small number of patients [15, 16], and more in depth studies are required to identify the genomic profiles of Korean MM patients.

Recent studies have performed clustering analysis that integrates various prognosis related genomic variations to classify patients with MM based on the characteristics of genomic alterations [8–10]. The heterogeneous nature of MM has warranted further studies on the multifaceted interpretation of

subgroups incorporating various genetic variations and their prognostic relevance. Several studies have utilized emerging technologies to examine the prognostic implications of genetic variations. These techniques include several clustering techniques to efficiently analyze the large amounts of data [17, 18]. In addition, various computational methods and tools to analyze structural variations based on NGS results have been developed [19, 20]. This analysis has an advantage of being able to simultaneously evaluate the somatic mutations and structural variations by detecting CNVs through strategies such as read depth, paired end mapping, and split read using SAM/BAM and FASTQ files obtained from the NGS process. Despite the fact that various analysis techniques have been explored so far, there are no consensus guidelines that include somatic variations and small structural variations.

In this study, the genomic profile of Korean patients with NDMM was examined using whole exome sequencing (WES) to investigate the characteristics of somatic mutations and CNVs in Korean. This study was intended to establish a Korean NDMM genome database to provide information in determining the composition of the NGS gene panel required for the diagnostic evaluation of MM, and development of biomarkers for predicting responses to treatments and prognosis. In addition, correlation analysis among various mutations was implemented, and new prognosis subgroups were identified through integrated analysis of CNVs and somatic variants.

2. MATERIALS AND METHODS

2.1. Patients

In this study, 67 patients with NDMM exhibiting more than 60% plasma cells in the bone marrow (BM) aspiration were recruited at the Seoul National University Hospital between July 2004 and January 2018. Patients were diagnosed to NDMM based on BM aspiration and biopsy according to the international myeloma working group (IMWG) 2016 guideline [21]. A total of three laboratory hematologists reviewed the slide for diagnosis and differential counting of plasma cells. Patients with >60% plasma cells were included in this study to maximize representativeness of variants for genetic changes of plasma cells. In particular, since the revised IWMG guideline designated plasma cell count over 60% as biomarkers of active myeloma, 60% was established as the selection criteria. The clinical characteristics, including the age of disease onset, sex, CRAB (hypercalcemia, renal impairment, anemia, and bone disease) symptoms, chemotherapy regimens, and survival, as well as the laboratory findings, including complete blood count, blood urea nitrogen/creatinine, albumin, lactate dehydrogenase levels, BM histological findings, and cytogenetic findings (FISH and conventional cytogenetics [CG]), of each patient were recorded. This study was approved by the Institutional Review Board of Seoul National University Hospital (IRB No. 1312-102-544). All study subjects provided their informed consent to participate in the study.

2.2. DNA extraction and exome sequencing

DNA isolated from the frozen BM mononuclear cells was used in the exome capture protocol. The SureSelectHuman All Exon V5+UTR probe set included 359,555 exons of 21,522 genes, and the size of the total targeted region was 75 Mb. To generate the standard exome capture libraries, the Agilent SureSelect Target Enrichment protocol was used for generating the Illumina paired-end sequencing library (ver. B.3, June 2015) with 3 μ g input genomic DNA (gDNA). The quantity and quality of DNA were examined using PicoGreen (Molecular Probes, OR, USA) and Nanodrop (ThermoFisher Scientific, MA, USA). DNA concentration \geq 50 ng/uL, purity \geq 1.7 (A260/A280), volume \geq 20 uL and total amount \geq 1 ug was passed for quality control criteria, and determined to acceptable for evaluation. The gDNA (1 μ g) was fragmented using adaptive focused acoustic technology (AFA; Covaris). The fragmented DNA was repaired with an adenine ligated to the 3' end, followed by ligation of the Agilent adapters. The adapter-ligated product was subjected to polymerase chain reaction (PCR). The final purified product was then quantified using quantitative real-time PCR (qRT-PCR) following the qPCR Quantification Protocol Guide and subjected to quality control using the Caliper LabChipHigh Sensitivity DNA (PerkinElmer, MA, USA). For exome capture, 250 ng of all exon capture libraries were mixed with hybridization buffers, blocking mixes, RNase block, and 5 μ L of SureSelect, following the standard Agilent SureSelect Target Enrichment protocol. Hybridization to the

capture baits was performed at 65° C using a heated thermal cycler with the lid temperature maintained at 105° C for 24 h in a PCR machine. The captured DNA was amplified and the final purified product was quantified using qRT-PCR according to the qPCR Quantification Protocol Guide. The amplified product was subjected to quality control using the TapeStationDNAscreentape (Agilent, CA, USA). The pooled DNA libraries were sequenced using the HiSeq™ 2500 platform (Illumina, CA, USA).

2.3. Bioinformatic evaluation of sequencing data

FASTQ files were aligned to the reference human genome (hg19; GRC37) using the Burrows–Wheeler aligner (BWA, v0.62) [22]. Duplicate PCR reads were removed using Picard 1.98. Variant calling was performed using the “HaplotypeCaller” in Genome Analysis Toolkit 2.7-2 [23]. To detect the candidate gene mutations, a filtering strategy was used. Low-quality variants with a low total depth (<20) and a low altered allele count (<10) were filtered out. Synonymous and noncoding variants were discarded. The variants with an allele frequency of more than 0.01 when compared with those in the 1000 Genomes Project (<http://browser.1000genomes.org/>), the Exome Aggregation Consortium (<http://exac.broadinstitute.org/>), and the NHLBI exome sequencing project (ESP6500, <http://evs.gs.washington.edu/EVS/>) databases were excluded. As a matched control sample was not included in this study, a stringent variant selection pipeline was applied to prioritize the high-confidence

set of somatic mutations. The driver genes were selected based on two previous studies [8, 10]. Genes that were identified as driver genes in at least one of these two studies were selected as driver genes for this study (Table 1). Additionally, the in silico prediction algorithms, SIFT [24, 25], CADD [26] and PolyPhen2 [27] were used, and the clinical significance was interpreted according to the ASCO/AMP/CAP guidelines [28].

Table 1. List of driver genes including the study

No	Driver gene selected for study	Driver genes detected in this study (among selected genes)
1	ABCF1	ABCF1
2	ACTG1	ARID1A
3	ARID1A	ATM
4	ARID2	ATRX
5	ATM	BHLHE41
6	ATRX	BRAF
7	BCL7A	C8orf34
8	BHLHE41	CREBBP
9	BRAF	CYLD
10	BTG1	DIS3
11	C8ORF34	DNMT3A
12	CCND1	DTX1
13	CDKN1B	DUSP2
14	CDKN2C	EGR1
15	CREBBP	EP300
16	CYLD	FAM46C
17	DIS3	FGFR3
18	DNMT3A	HIST1H1B
19	DTX1	HIST1H1D
20	DUSP2	HIST1H1E
21	EGR1	HUWE1
22	EP300	IDH1
23	FAM46C	IDH2

24	FGFR3	IGLL5
25	FUBP1	IRF1
26	HIST1H1B	IRF4
27	HIST1H1D	KDM6A
28	HIST1H1E	KLHL6
29	HIST1H2BK	KMT2C
30	HUWE1	KRAS
31	IDH1	LTB
32	IDH2	MAFB
33	IGLL5	MAML2
34	IRF1	MAN2C1
35	IRF4	NCOR1
36	KDM5C	NF1
37	KDM6A	NFKB2
38	KLHL6	NFKBIA
39	KMT2B	NRAS
40	KMT2C	PABPC1
41	KRAS	PIM1
42	LCE1D	POT1
43	LTB	PRDM1
44	MAF	PRKD2
45	MAFB	PTPN11
46	MAML2	RB1
47	MAN2C1	RFTN1
48	MAX	RPL10
49	NCOR1	SAMHD1
50	NF1	SETD2
51	NFKB	SF3B1
52	NFKB2	TBC1D29
53	NFKBIA	TET2
54	NRAS	TP53
55	PABPC1	TRAF2
56	PIK3CA	TRAF3
57	PIM1	UBR5
58	POT1	ZNF292
59	PRDM1	
60	PRKD2	
61	PTPN11	
62	RASA2	

63	RB1
64	RFTN1
65	RPL10
66	RPL5
67	RPRD1B
68	RPS3A
69	SAMHD1
70	SETD2
71	SF3B1
72	SP140
73	TBC1D29
74	TCL1A
75	TET2
76	TGDS
77	TP53
78	TRAF2
79	TRAF3
80	UBR5
81	XBP1
82	ZFP36L1
83	ZNF292

2.4. Copy number analysis

Copy number alterations were analyzed using a CNVkit [20] and Nexus software version 5.0 (Biodiscovery, CA, USA). The copy number in the NDMM dataset was called against an MM-negative karyotype FISH panel. Heatmap plots were drawn with the “heatmap” command in CNVkit. Data were loaded into Nexus 5.0 and the copy number calls were generated genome-wide for each sample based on the fixed thresholds for deletions and duplications specified in the settings. SNP-rank segmentation algorithm [29], a statistics-based algorithm similar to the circular binary segmentation, applies both the

copy number value and the B-allele frequency to the segmentation (Fig 1). GISTIC algorithms were used to determine the significance of focal somatic copy number alterations.

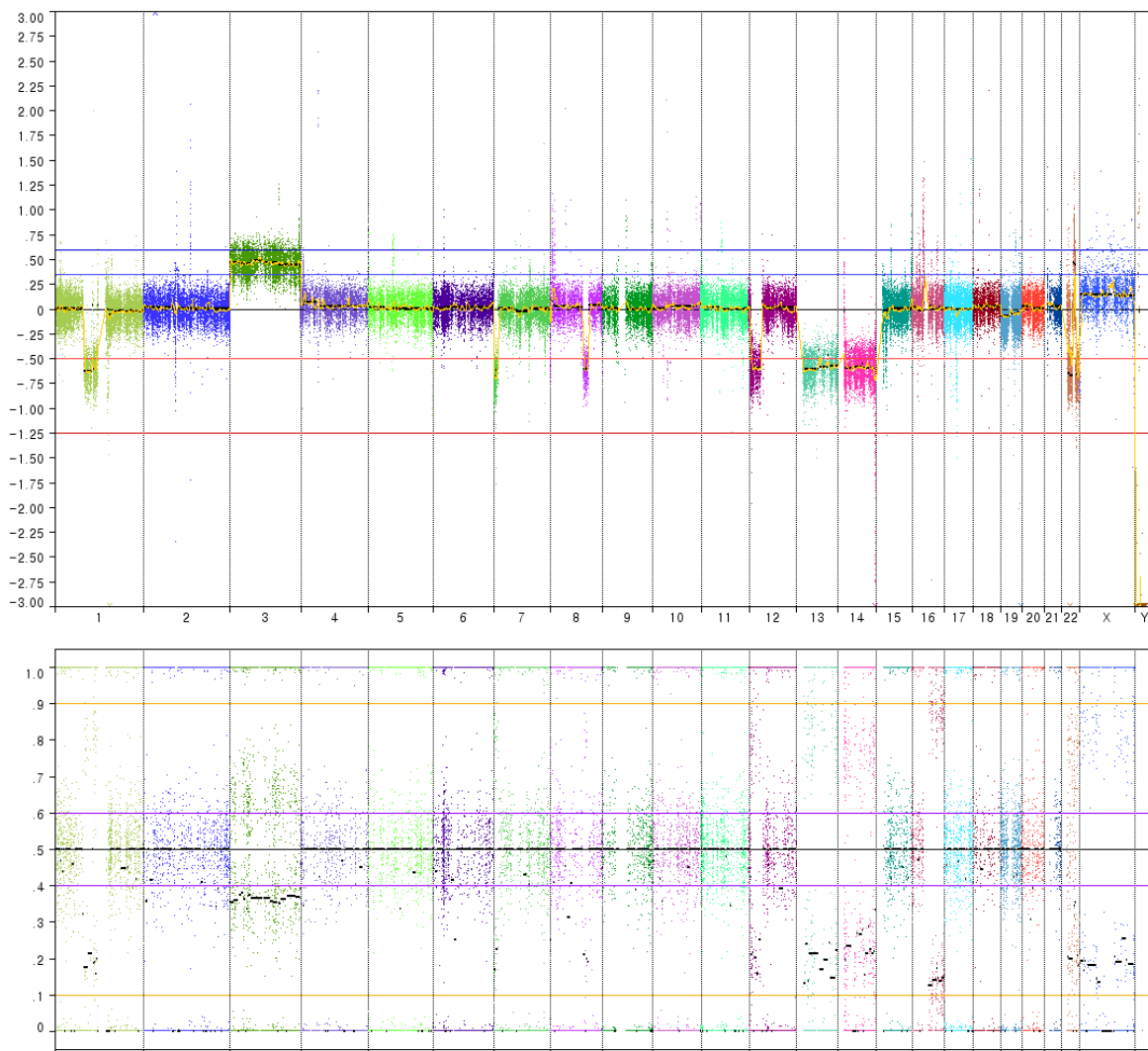


Fig 1. Examples of copy number alteration in one patient. Patient #1 showed hemizygous deletion of 1p, 8q, 12p, 13q, 14q, and 21q, and duplication of chromosome 3. Log R ratio (upper) and B allele frequency (lower) were described on each chromosome.

2.5. Statistical analysis

All statistical analyses were performed using the statistical R-project program (version 3.6.2), PASW statistics version 18 (SPSS Inc., IL, USA), and MedCalc version 12.0 (MedCalc Software, Mariakerke, Belgium). A pairwise correlation analysis between somatic mutations and CNVs was performed using the Fisher test. Cumulative overall survival (OS) curves for the groups with or without genomic variations were calculated using the Kaplan–Meier (KM) method and compared using the log–rank test and Breslow test. The Cox proportional hazards model was used to evaluate the prognostic impact of CNVs and mutated genes on OS. A multivariate analysis was performed on the full set of significant variables in the univariate analysis. The differences were considered significant at $P < 0.05$.

3. RESULTS

3.1. Characteristics of the study population

The demographic and clinical characteristics of the study population, including the laboratory tests performed on the day of MM diagnosis, are described in Table 1. The numbers of male and female patients were 34 and 33, respectively. The median age of the study population was 65 years. Among the total nucleated cells in the BM aspiration, the median percentage of BM plasma cells was 75.8%. The numbers of patients with ISS stage I, II, and III tumors were 7, 25, and 33, respectively (Table 2).

Table 2. Demographic and clinical characteristics of the study population

Variables	Baseline distribution of patients (N = 67)
Sex (Male/Female)	34 (50.7%) / 33 (49.3%)
Age (year)	65.0 (58.0–71.0)
Stage ISS (I/II/III/NA)	7 (10.4%) / 25 (37.3%) / 33 (49.3%) / 2 (3.0%)
Serum M protein (g/dL)	6.1 (4.3–9.1)
β 2-Microglobulin (mg/dL)	6.4 (3.7–15.3)
Hemoglobin (g/dL)	8.3 (7.3–9.3)
Calcium (mg/dL)	9.4 (8.6–10.0)
BUN (mg/dL)	17.0 (14.0–24.8)
Creatinine (mg/dL)	1.1 (0.9–1.9)
Albumin (mg/dL)	3.1 (2.7–3.5)
LDH (U/L)	174 (137.5–221)
Bone disease	45 (67.1%)
BM plasma cell (%)	75.8 (66.2–85.4)
Death (Yes/No)	31 (46.3%) / 36 (53.7%)
Treatment	
VAD	14 (21.9%)
VMP	13 (20.3%)
VTD	7 (10.9%)
TD	11 (17.2%)
MP (MPT)	8 (12.5%)
Others*	14 (20.9%)

Values are presented as median (Interquartile Range)

Abbreviations: NDMM, newly diagnosed multiple myeloma; ISS, International staging system; NA, Not available; LDH, Lactate dehydrogenase; BM, Bone marrow; VAD, vincristine/doxorubicin/dexamethasone; VMP, bortezomib/ melphalan/prednisolone; VTD, bortezomib/ thalidomide/dexamethasone; TD, thalidomide/dexamethasone; MPT, melphalan/prednisolone/thalidomide

*Other: four patients for bortezomib/dexamethasone, four patients for dexamethasone, two patients for ixzomib, one patients for lenalidomide, three patients are unknown for treatment regimen.

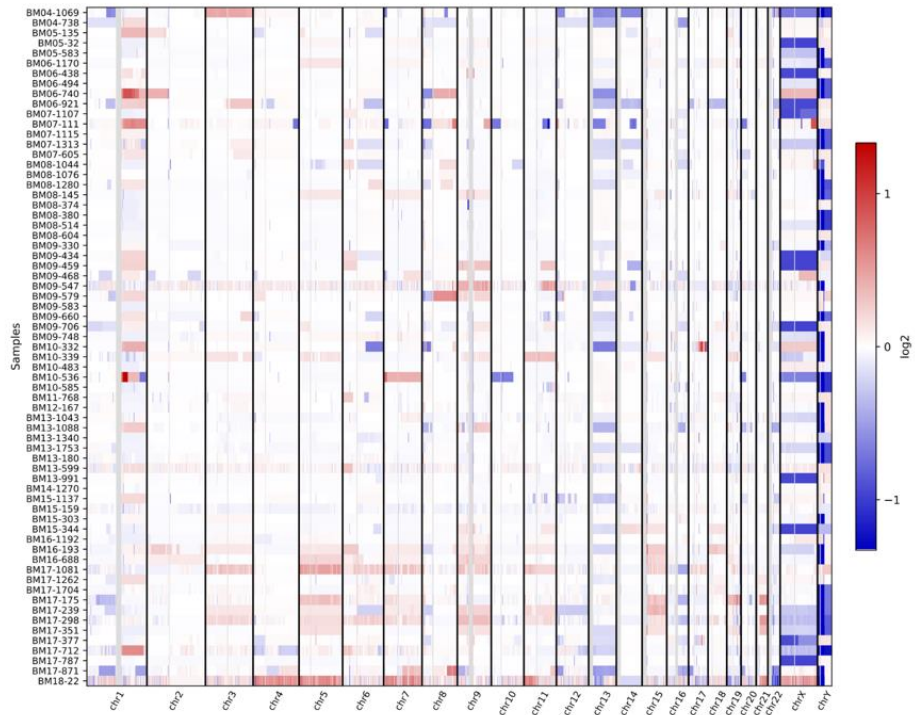
3.2. Identification of CNVs in MM

The gain and loss of the p arm, q arm, and both arms of chromosomes 1 to 22 were analyzed. The most common chromosomal gain was 1q. Among the odd-numbered chromosomes, the most frequent whole-arm gain was observed in chromosomes 5, 7, 9, 15, and 19. The most common chromosome loss was 13q loss, followed by the losses of 16q, 22q, 8p, 14q, 1p, and 6q (Table 3; Fig 2). The frequency of the 1q gain, 6p gain, and 13q loss with cytoband information is represented in Table 4.

Table 3. Frequency of copy number alterations in patients with newly diagnosed multiple myeloma

Chr	Gain			Total (%)	Loss			Total (%)
	p arm	q arm	whole arms		p arm	q arm	whole arms	
1	0	35	0	35 (17.5)	13	3	0	16 (8.1)
2	3	2	1	6 (3.0)	1	5	0	6 (3.0)
3	0	4	8	12 (6.0)	0	0	0	0 (0.0)
4	1	1	1	3 (1.5)	5	4	0	9 (4.5)
5	1	0	14	15 (7.5)	0	1	0	1 (0.5)
6	10	2	4	16 (8.0)	2	12	0	14 (7.1)
7	0	4	11	15 (7.5)	4	0	0	4 (2.0)
8	0	8	0	8 (4.0)	15	1	1	17 (8.1)
9	1	3	14	18 (9.5)	2	0	0	2 (1.0)
10	0	0	0	0 (0.0)	2	2	0	4 (2.0)
11	1	6	6	13 (6.5)	2	4	1	7 (3.5)
12	1	0	0	1 (0.5)	8	1	1	10 (5.1)
13	0	0	0	0 (0.0)	0	33	0	33 (16.7)
14	0	2	0	2 (1.0)	0	13	0	13 (6.6)
15	0	2	12	14 (7.0)	0	0	0	0 (0.0)
16	2	1	0	3 (1.5)	2	18	0	20 (10.1)
17	0	2	1	3 (1.5)	3	2	0	5 (2.5)
18	0	3	6	9 (4.5)	1	0	1	2 (1.0)
19	6	0	11	17 (8.5)	5	0	4	9 (4.5)
20	0	0	0	0 (0.0)	7	3	0	10 (5.1)
21	0	1	6	7 (3.5)	0	0	0	0 (0.0)
22	0	2	0	2 (1.0)	1	16	0	17 (8.6)
Total	26	78	95	199 (100)	73	118	8	199 (100)

(A)



(B)

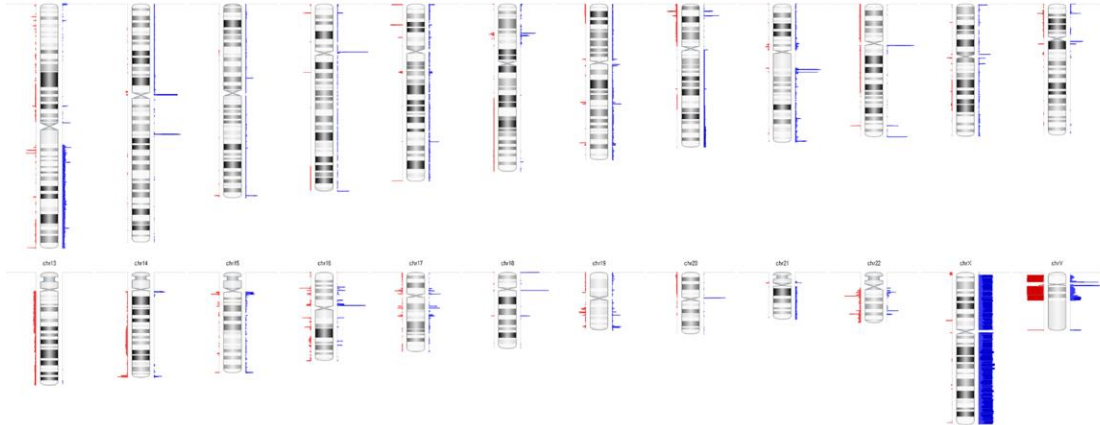


Fig 2. Summary plot of chromosomal gain and loss in 67 patients with newly diagnosed multiple myeloma determined using whole-exome sequencing and copy number analysis. (A) In the heatmap plot, each row represents one patient,

while each column represents chromosomes 1–23, X, and Y in order. (B) Ideogram with regions of chromosomal gain and loss. Copy number gains and deletions are marked as blue and red bars to the right and left side of the ideogram, respectively.

Table 4. Frequency of chromosome gain and loss including cytoband information

(A) 1q gain

Region	Number	Percentage
q21.1–q23.2	1	2.9%
q21.1–q23.3	1	2.9%
q21.1–q41	1	2.9%
q21.1–q44	23	65.7%
q21.2–q44	7	20.0%
q23.2–q41	1	2.9%
q31.3–q44	1	2.9%
Total	35	100.0%

(B) 6p gain

Region	Number	Percentage
Whole p arm	6	42.9%
Whole p, q arms	4	28.6%
p21.3	1	7.1%
p25.3–p12.3	1	7.1%
p25.3–p21.1	1	7.1%
p25.3–p21.2	1	7.1%
Total	14	100.0%

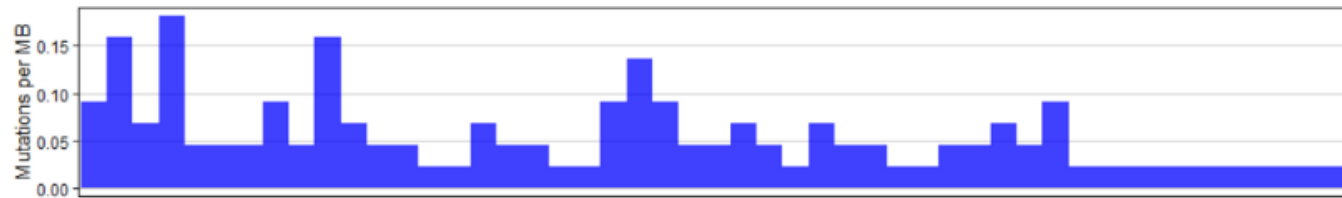
(C) 13q loss

Region	Number	Percentage
q11–q34	29	87.9
q12.1–21.3	1	3.0%
q12.1–q22.1	1	3.0%
q13.3–q14.2	1	3.0%
q33.1 – q34	1	3.0%
Total	33	100.0%

3.3. Analysis of selected driver gene mutations

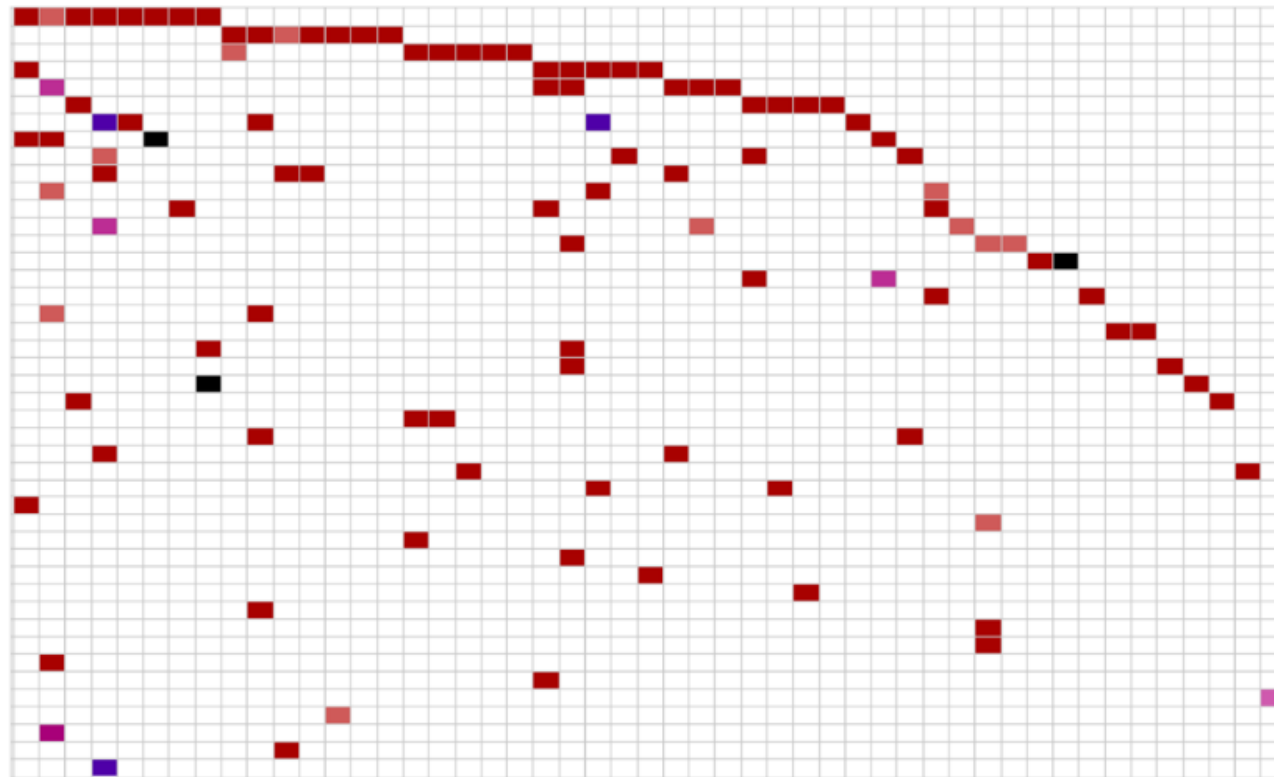
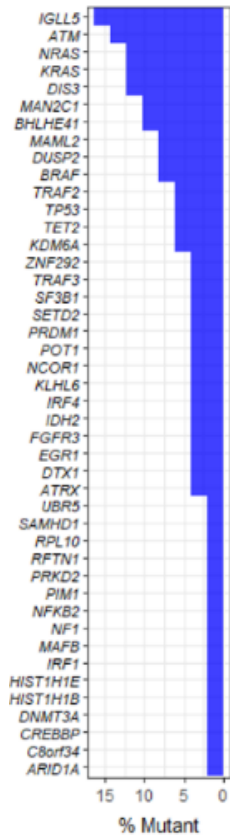
Driver gene mutations were detected in 58 of the 83 genes selected based on two previous studies [8, 10]. Variations classified as tier I, tier II, and tier III were separately selected based on the ASCO/AMP/CAP guidelines to analyze the frequency and mutation types of each gene (Fig 3). The most frequently mutated driver gene was *IGLL5*, which was detected in eight patients (11.9%). Additionally, seven patients (10.4%) had mutations in *ATM*, six patients (9.0%) had mutations in *NRAS*, *KRAS*, and *DIS3*, five patients (7.5%) had mutations in *MAN2C1* and *BHLHE41*, four patients (6.0%) had mutations in *MAML2*, *DUSP2*, and *BRAF*, and three patients (4.5%) had mutations in *TRAF2*, *TP53*, *TET2*, and *KDM6A*. The diagram including the variation and domain for each gene is shown in Fig 4. The p.V600E variation in *BRAF* was detected in three patients, while the p.Q61R(L) variation in *KRAS* was detected in three patients. In *NRAS*, the p.G12D(V) and p.Q61L(K) variations were detected in two patients.

Mutations, such as p.S403L, p.L581V, and p.Q584X, were detected in *ATM* (Fig 4).



Translational Effect

- Synonymous
- Non Synonymous



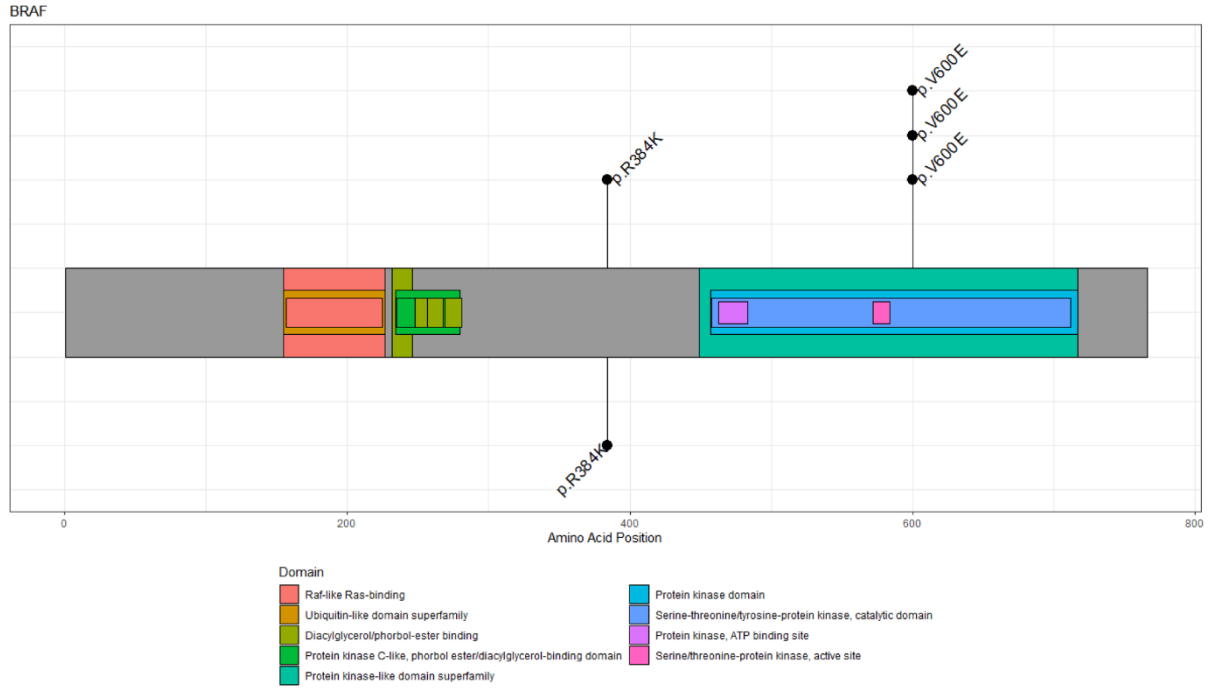
Mutation Type

- frameshift deletion
- nonsynonymous SNV
- stopgain
- nonframeshift deletion
- splicing
- frameshift insertion
- nonframeshift insertion

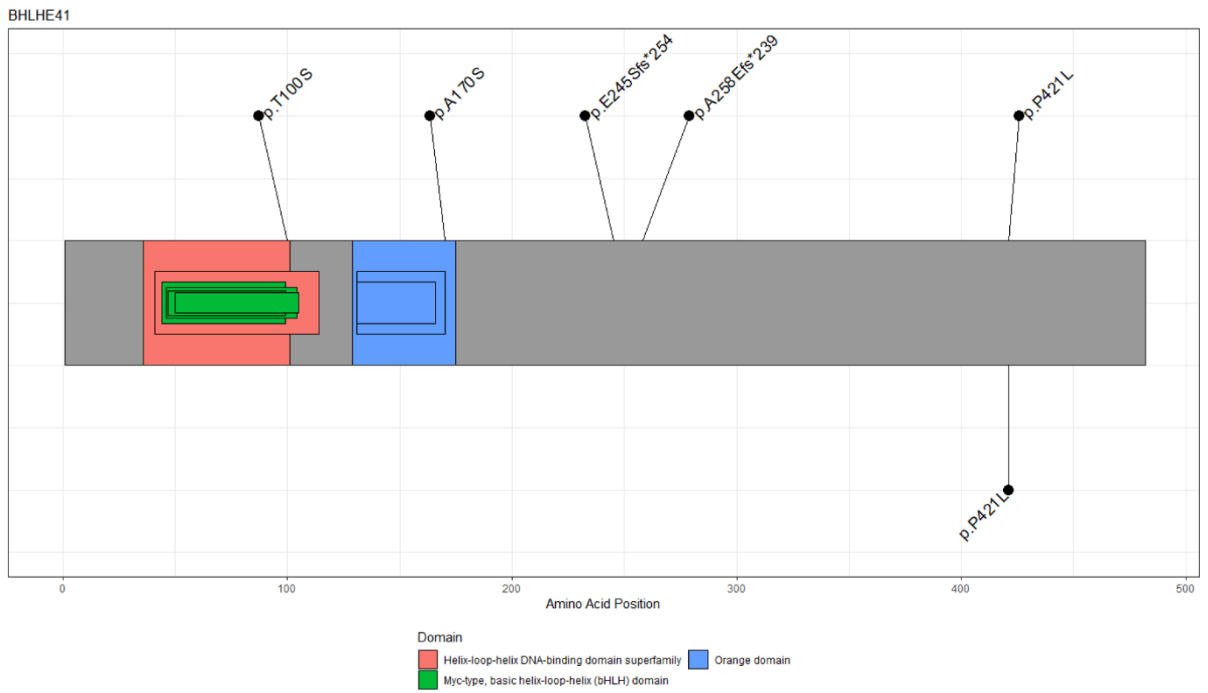
Sample (n=49)

Fig 3. Frequency and distribution of somatic mutations in patients with NDMM. Mutation burden per megabase (Mb) in tumor are represented (upper) and mutated genes are ranked by mutant frequency (lower). The heatmap showed individual mutations in patient samples, color-coded by type of mutation. Genes with significant somatic mutations (tier I, II, and II by ASCO/AMP/CAP guidelines) are represented.

(A)

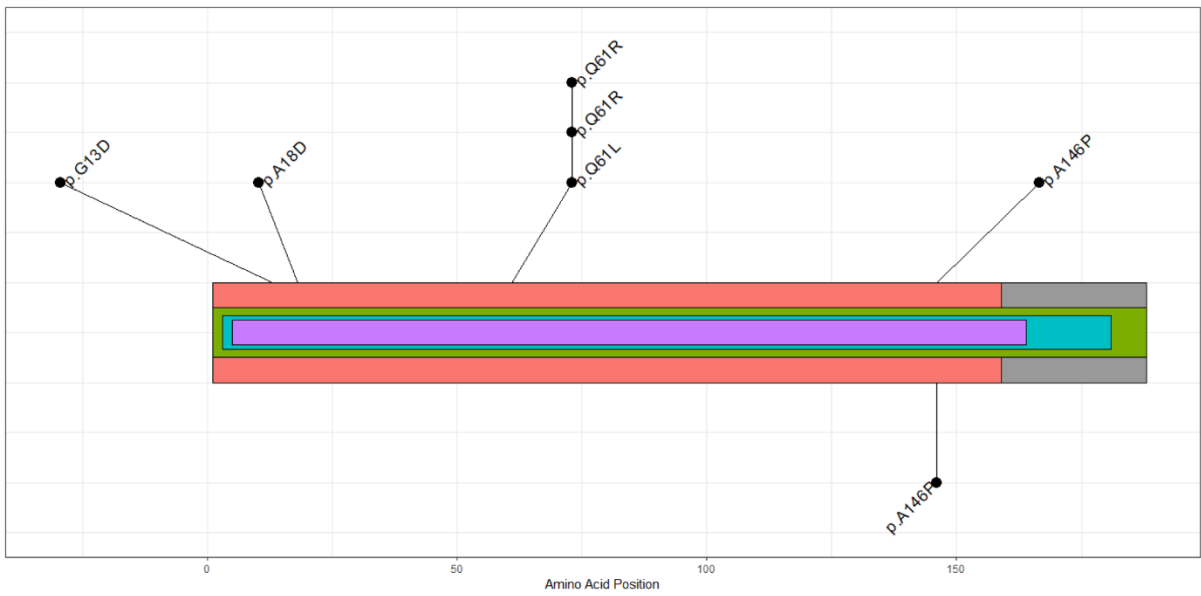


(B)



(C)

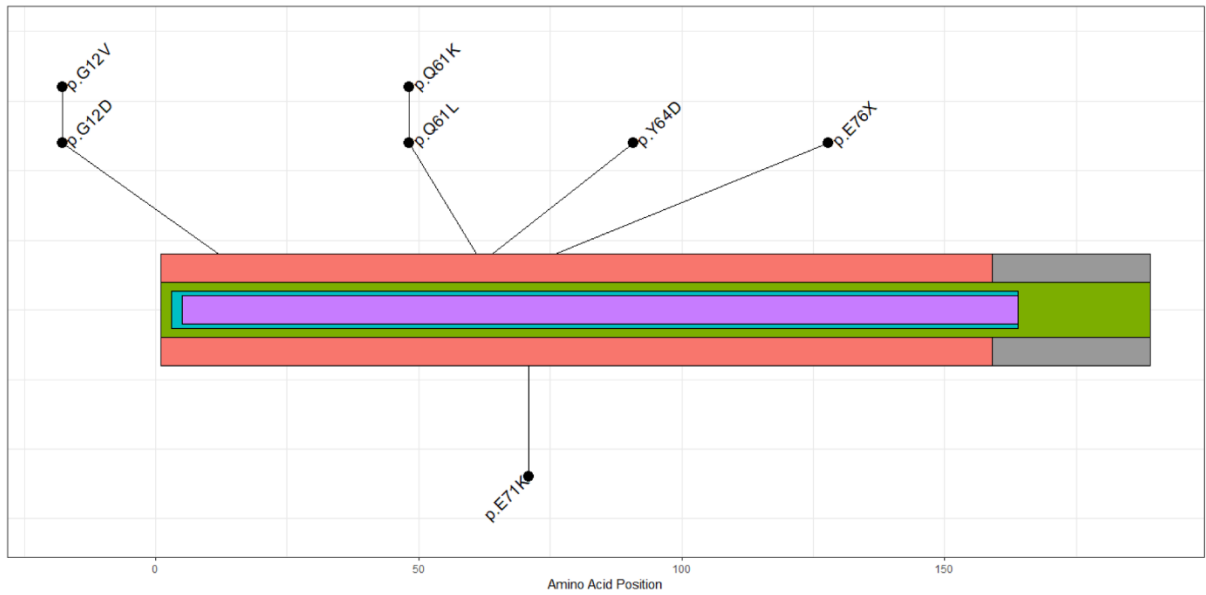
KRAS



Domain
Small GTP-binding protein domain P-loop containing nucleoside triphosphate hydrolase
Small GTPase superfamily, Ras-type Small GTPase

(D)

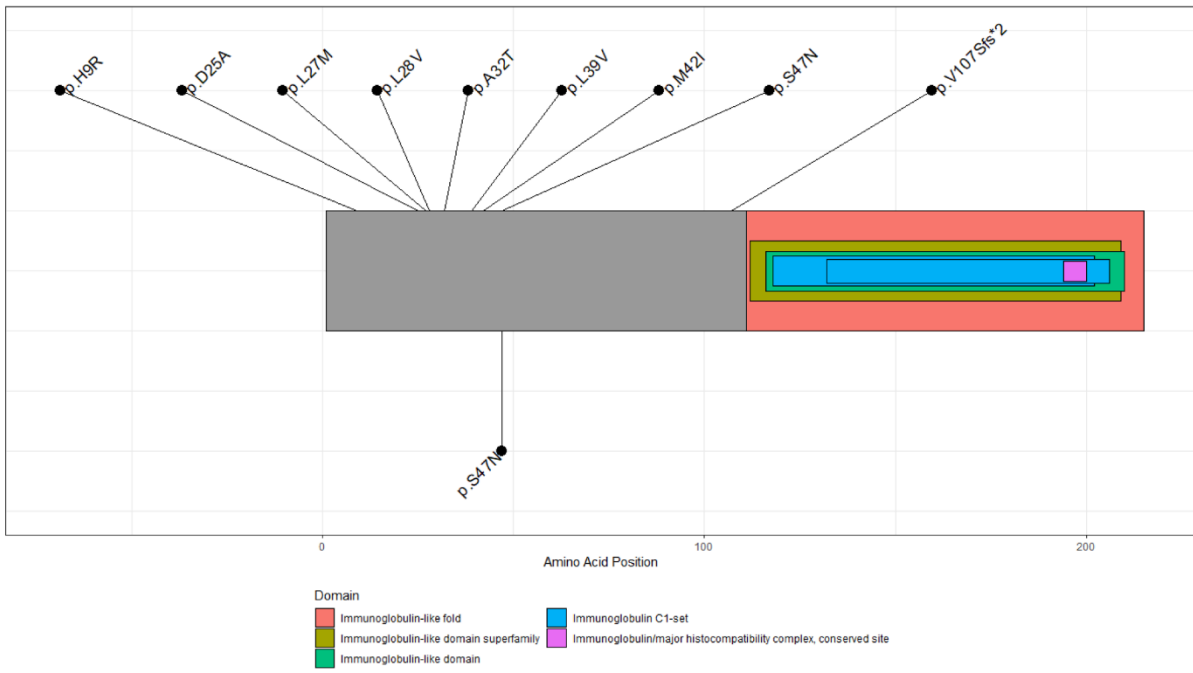
NRAS



Domain
Small GTP-binding protein domain P-loop containing nucleoside triphosphate hydrolase
Small GTPase superfamily, Ras-type Small GTPase

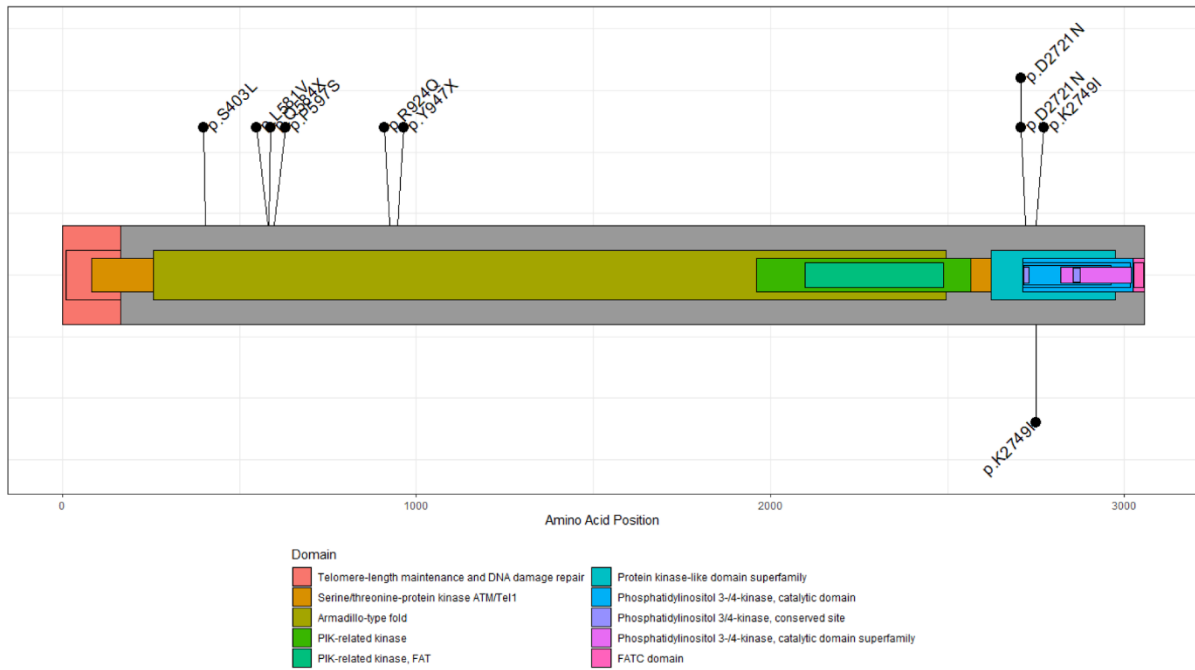
(E)

IGLL5



(F)

ATM



(G)

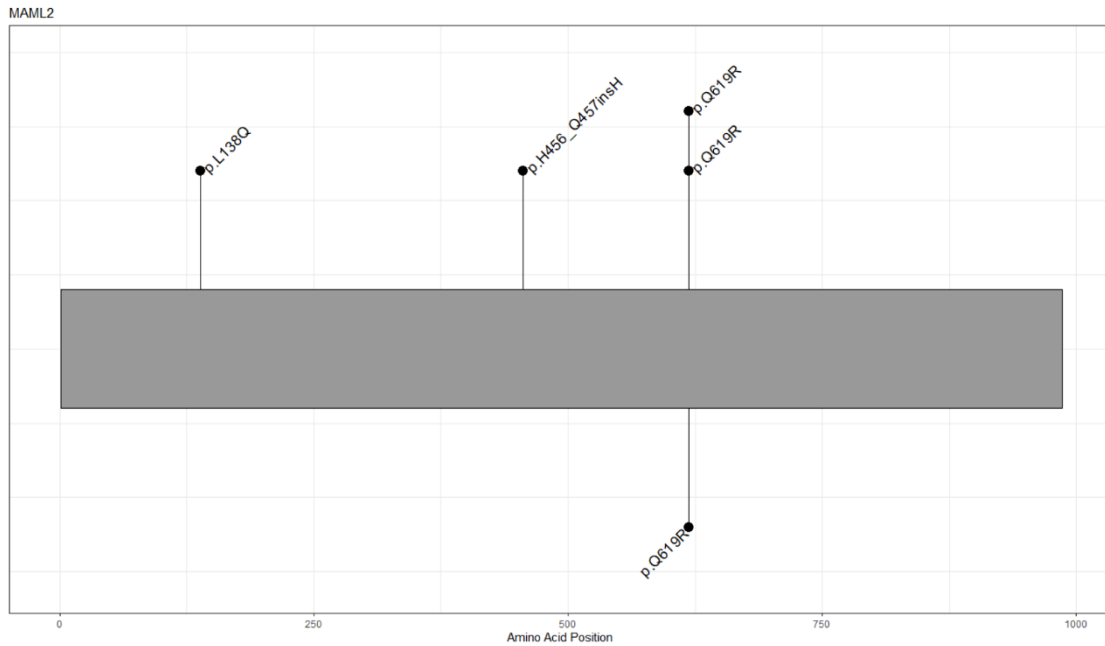


Fig 4. Diagrams for mutations for most frequent genes in NDMM patients.

(A)–(G) is diagrams for *BRAF*, *BHLHE41*, *KRAS*, *NRAS*, *IGLL5*, *ATM*, and *MAML2* respectively. In the case of the *BRAF* gene, p.V600E occurred in 3 patients, and *KRAS* showed p.Q61R(L) in 3 patients. In the case of *NRAS*, p.G12D(V) and p.Q61L(K) were confirmed in each of 2 patients. In the *ATM*, mutations such as p.S403L, p.L581V, and p.Q584X occurred.

3.4. Correlation analysis of genomic variants

CNVs were frequently detected with moderate or high correlation between each other. In patients with hyperdiploid MM, CNVs in odd-numbered chromosomes were highly correlated with each other. The CNVs between the following chromosome pairs that occurred simultaneously were highly correlated: chromosomes 5 and 9 (Pearson correlation coefficient [PCC] = 0.745; $P < 0.001$), chromosomes 5 and 15 (PCC = 0.791; $P < 0.001$), chromosomes 5 and 19 (PCC = 0.791; $P < 0.001$), chromosomes 9 and 15 (PCC = 0.763; $P < 0.001$), and chromosomes 15 and 19 (PCC = 0.782; $P < 0.001$) (Fig 5A). Among patients with non-hyperdiploid MM, there was a high correlation between 1q gain and 13q loss (PCC = 0.734, $P < 0.001$), and a moderate correlation between the following CNV pairs: 20 loss and 22q loss (PCC = 0.577, $P < 0.001$); 6p gain and 6q loss (PCC = 0.553 and $P < 0.001$). The analysis of correlations between genes with significant prognostic value among driver gene mutations revealed a significant correlation between *BRAF* and *EGR1* mutations (PCC = 0.564, $P < 0.001$). Additionally, the analysis of the correlation between CNVs and driver gene mutations revealed a significant correlation of *MAML2* mutation with 17q loss (PCC = 0.485, $P < 0.001$), 22q loss (PCC = 0.419, $P = 0.004$), 6q loss (PCC = 0.423, $P = 0.004$), and 6p gain (PCC = 0.373, $P = 0.001$). Additionally, *EGR1* mutation was significantly correlated with 12p loss (PCC = 0.426, $P = 0.004$) and 1p loss (PCC = 0.324,

$P = 0.03$). The other driver gene mutations were not significantly correlated with CNVs (Fig 5B).

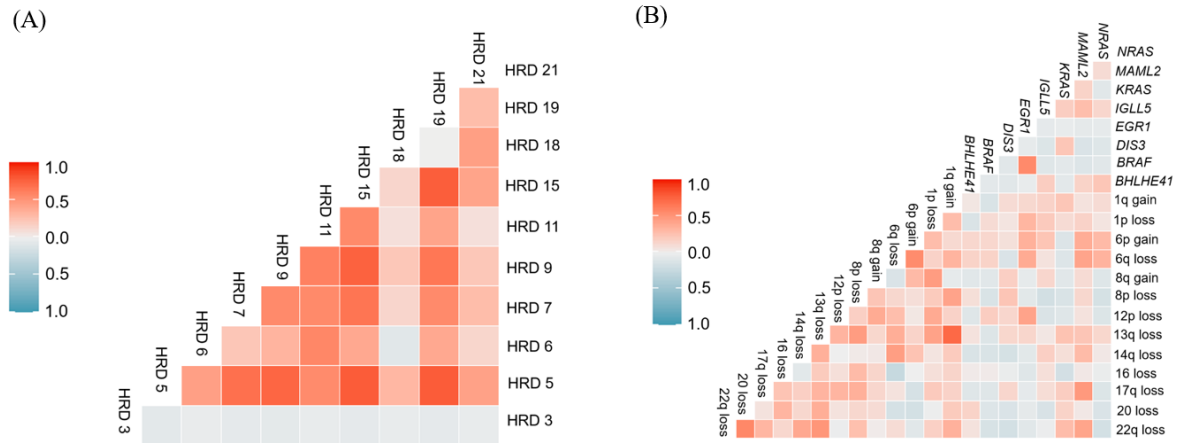


Fig 5. Pairwise associations between mutations and copy number variations. (A) In patients with hyperdiploid multiple myeloma, chromosome 5 variations were correlated with variations in chromosomes 9, 15, and 19, while chromosome 15 variations were correlated with variations in chromosomes 9 and 19. (B) Among patients without hyperdiploid multiple myeloma, 1q gain was correlated with 13q loss, *BRAF* mutation was correlated with *EGR1* mutation, and *MAML2* mutation was correlated with 17q loss.

3.5. Prognostic impact of CNVs and somatic mutations

3.5.1. Prognostic impact of each of the CNVs and somatic mutations

The univariate analysis revealed that 1q gain (hazard ratio [HR] = 2.54; 95% confidence interval [CI] = 1.06–6.04; $P = 0.036$), 6p gain (HR = 3.50; 95% CI = 1.45–8.46; $P = 0.005$), 6q loss (HR = 3.33; 95% CI = 1.29–8.65; $P = 0.013$), 8p loss (HR = 2.48; 95% CI = 1.06–5.78; $P = 0.036$), and 13q loss (HR = 2.25; 95% CI = 0.99–5.08; $P = 0.052$) were associated with poor OS in patients with non-hyperdiploid MM. The multivariate analysis of these CNVs revealed that 1q gain and 6p gain were significant independent prognostic markers (Table 5). Additionally, KM analysis revealed that patients with 1q gain ($P = 0.03$; Fig 6A), 6p gain ($P = 0.003$; Fig 6B), 6q loss ($P = 0.009$; Fig 6C), 8p loss ($P = 0.03$; Fig 6D), and 13q loss ($P = 0.046$; Fig 6E) exhibited poor prognosis. In patients with driver gene mutations, univariate and multivariate analyses revealed that mutations in *MAML2* (HR = 3.32; 95% CI = 1.34–8.27; $P = 0.010$) and *BHLHE41* (HR = 5.16; 95% CI = 1.91–13.9; $P = 0.001$) were significantly associated with poor OS. The KM plot of patients with driver gene mutations is shown in Fig 7. Patients with *BHLHE41* ($P < 0.001$; Fig 7A) and *MAML2* ($P = 0.016$; Fig 7B) mutations exhibited a significantly poor prognosis in the log-rank test, while those with *BRAF* mutations exhibited a significantly poor prognosis in the Breslow test ($P = 0.036$; Fig 7C). Additionally, *CREBBP* ($P < 0.001$), *HIST1H1D* ($P = 0.044$), *PRDM1* ($P = 0.025$), *KLHL6* ($P < 0.001$),

TRAF3 ($P = 0.049$), and *UBR5* ($P = 0.003$) mutations, which were detected only in one or two patients, were significantly associated with poor prognosis.

Table 5. Cox proportional hazards model for factors associated with overall survival.

Variables		N	Univariable		Multivariable	
			HR (95% CI)	<i>P</i>	HR (95% CI)	<i>P</i>
CNV (Non- hyperdiploid)	1q gain	26	2.54 (1.06–6.04)	0.036	2.46 (1.02–5.92)	0.045
	6p gain	7	3.50 (1.45–8.46)	0.005	3.40 (1.38–8.38)	0.008
	6q loss	6	3.33 (1.29–8.65)	0.013		
	8p loss	9	2.48 (1.06–5.78)	0.036		
	13q loss	24	2.25 (0.99–5.08)	0.052		
Mutated driver genes	<i>MAML2</i>	4	3.41 (1.19–9.82)	0.023	3.96 (1.35–11.67)	0.013
	<i>BHLHE41</i>	5	5.16 (1.91–13.90)	0.001	6.93 (2.21–21.73)	0.001

Abbreviations: CNV, copy number variation; HR, hazard ratio; CI, confidential interval

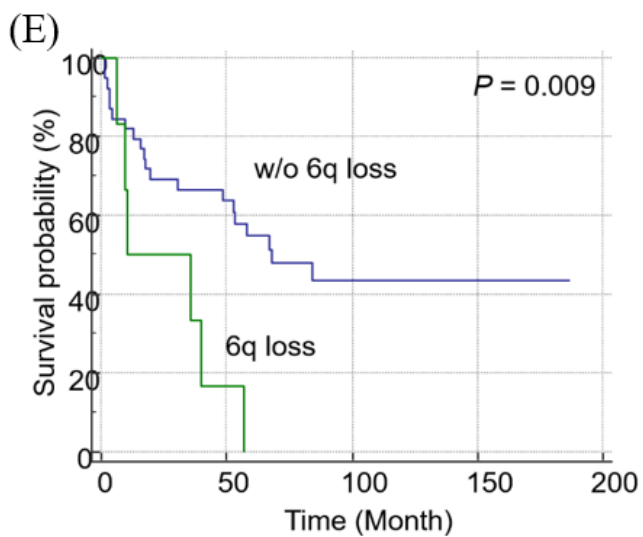
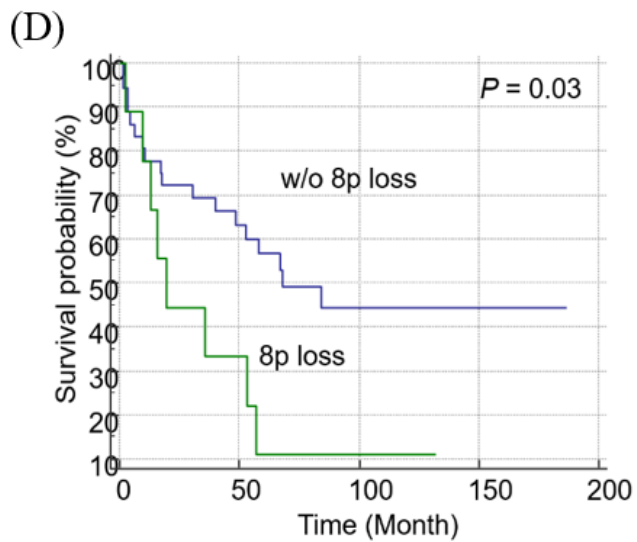
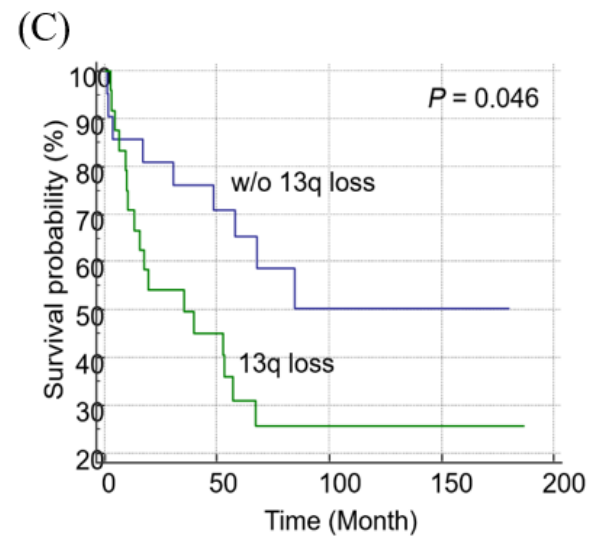
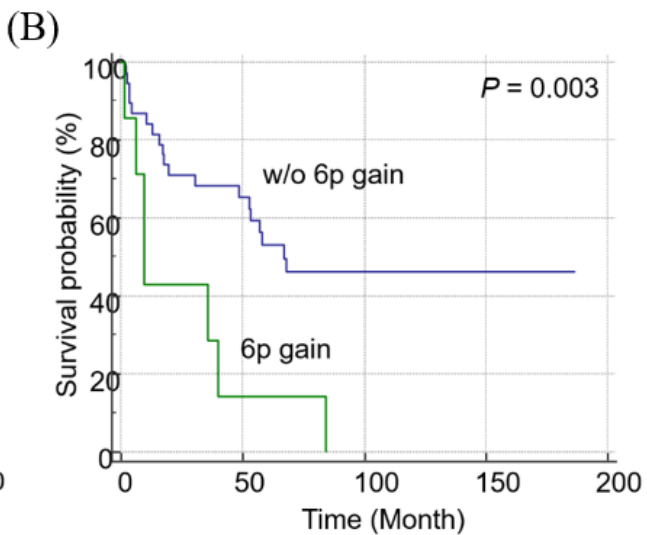
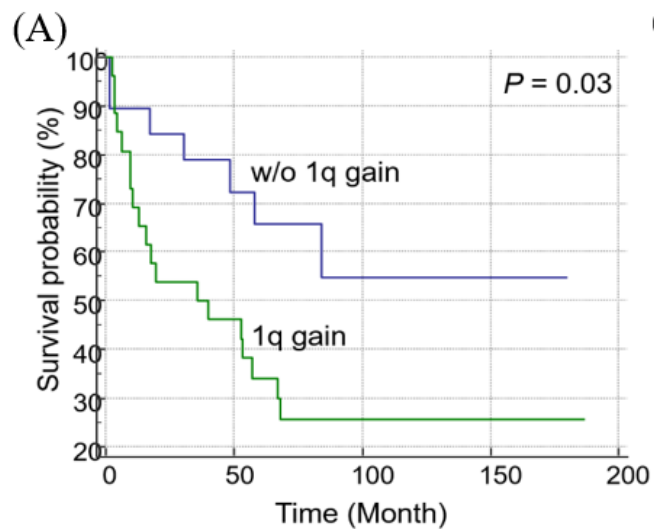


Fig 6. Kaplan–Meier analysis of overall survival in patients with copy number variations. Patients with (A) 1q gain ($P = 0.03$), (B) 6p gain ($P = 0.003$), (C) 6q loss ($P = 0.009$), (D) 8p loss ($P = 0.03$), and (E) 13q loss ($P = 0.046$) were significantly correlated with poor prognosis in the log–rank test.

Abbreviations: w/o; without

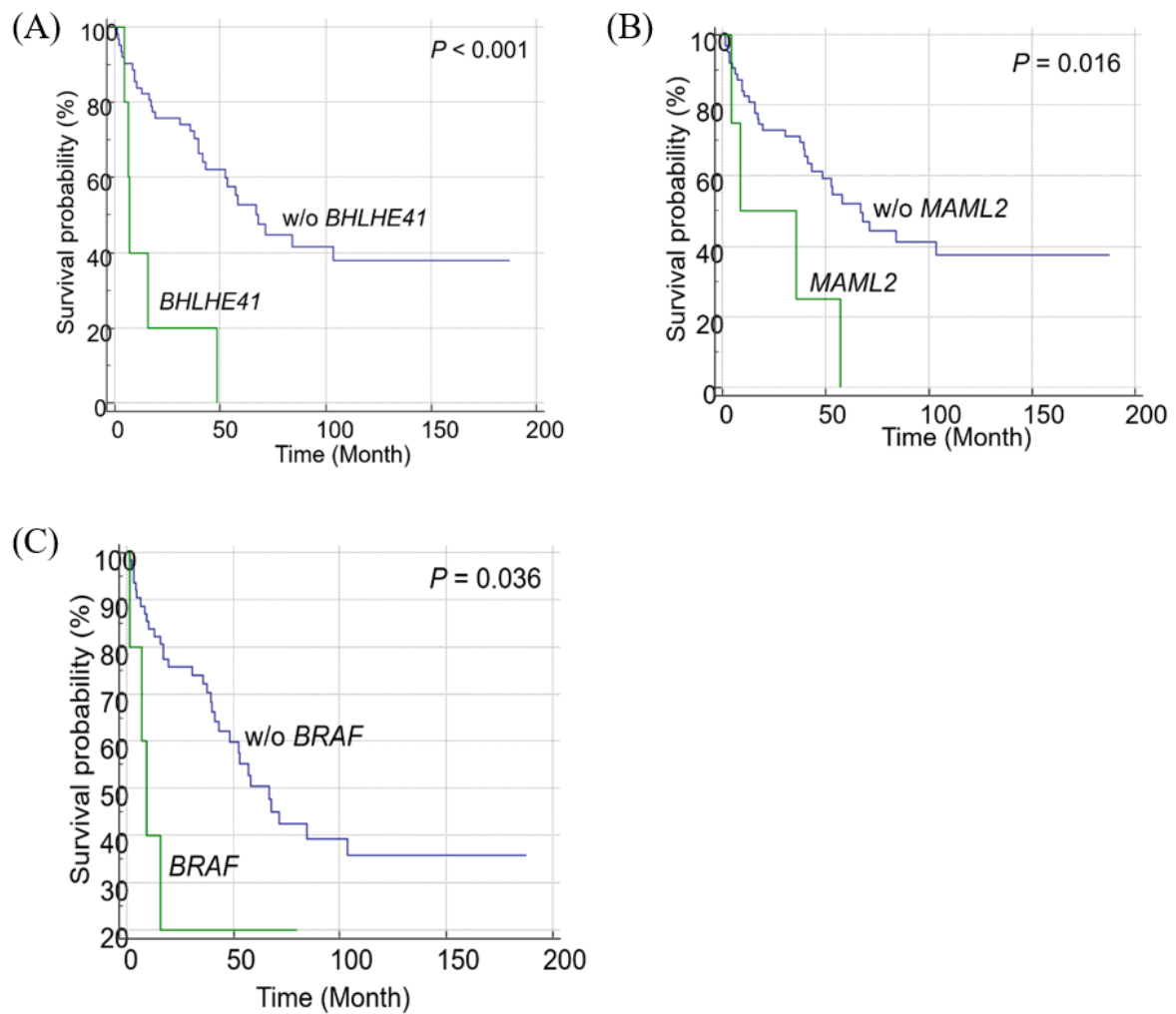


Fig 7. Kaplan–Meier analysis of overall survival in patients with mutations in the driver genes. (A) *BHLHE41* ($P < 0.001$) and (B) *MAML2* ($P = 0.016$) mutations were significantly correlated with poor prognosis in the log–rank test, while (C) *BRAF* mutations ($P = 0.036$) were significantly correlated with poor prognosis only in the Breslow test. Abbreviations: w/o; without

3.5.2. Prognostic impact of CNVs identified through FISH, CG, and WES analyses

Quantitative results of FISH and CG are shown in Table 6. Compared with the CG analysis of metaphase chromosomes, FISH and WES analyses revealed a higher frequency of abnormalities. CNV and FISH analyses identified 1q gain in 52.2% and 60.0% of the patients, respectively. Moreover, CNV and FISH analyses identified 13q deletion in 49.3% and 48.1% of the patients, respectively. Survival analysis of patients with or without copy number alterations identified through FISH, CG, and WES analyses was performed (Fig 6). KM analysis revealed that 1q gain identified through FISH was not significantly correlated with OS ($P = 0.831$; Fig 6A). Similarly, 1q gain identified through FISH and CG analyses (FISH+CG) was not significantly associated with OS ($P = 0.174$; Fig 6B). Meanwhile, patients with 1q gain identified through WES ($P = 0.03$; Fig 4A) and FISH+CG+WES ($P = 0.022$; Fig 6C) analyses exhibited a significantly adverse OS. Furthermore, patients with 13q deletion identified through CNV exhibited poor OS ($P = 0.046$; Fig 4E). Patients with 13q deletion identified through FISH+CG+WES also exhibited poor OS ($P = 0.013$; Fig 6F). However, the genetic variations identified through FISH ($P = 0.086$; Fig 6D) and FISH+CG ($P = 0.073$; Fig 6E) were not significantly correlated with OS.

Table 6. Results of copy number variations and rearrangements by WES, FISH and CG in patients with NDMM

Anomaly	WES CNV*	FISH*	CG*	FISH+CG*†	FISH+CG+WES*†
1q (1q25) gain	52.2% (35/67)	60.0% (30/50)	27.0% (17/63)	46.3% (31/67)	67.2% (45/67)
13q (13q14, RB1) deletion	49.3% (33/67)	48.1% (25/52)	34.9% (22/63)	50.7% (34/67)	58.2% (39/67)
17p (17p13, p53) deletion	7.5% (5/67)	12.5% (3/24)	1.6% (1/63)	6.5% (4/62)	9.0% (6/67)
<i>IGH/MAF</i> rearrangement	NA	8.0% (2/25)	3.2% (2/63)	4.8% (3/62)	NA
<i>IGH/FGFR3</i> rearrangement	NA	18.5% (5/27)	1.6% (1/63)	9.7% (6/62)	NA

*In parentheses, number of patients with positive results/number of patients with total MM cases tested.

†Patients with positive on at least one of the tests were counted.

Abbreviations: WES, whole-exome sequencing; CNV, copy number variation; FISH, fluorescence in situ hybridization; CG, conventional cytogenetics; NDMM, newly diagnosed multiple myeloma

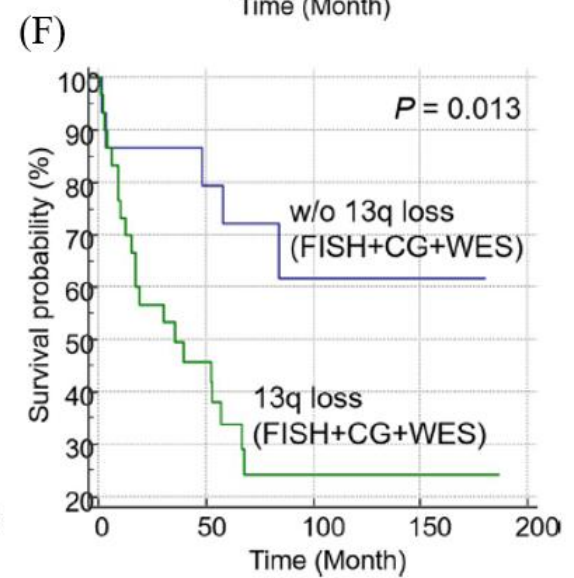
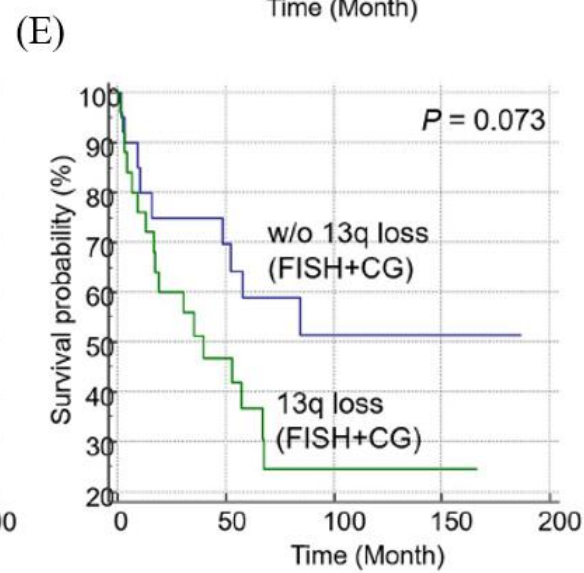
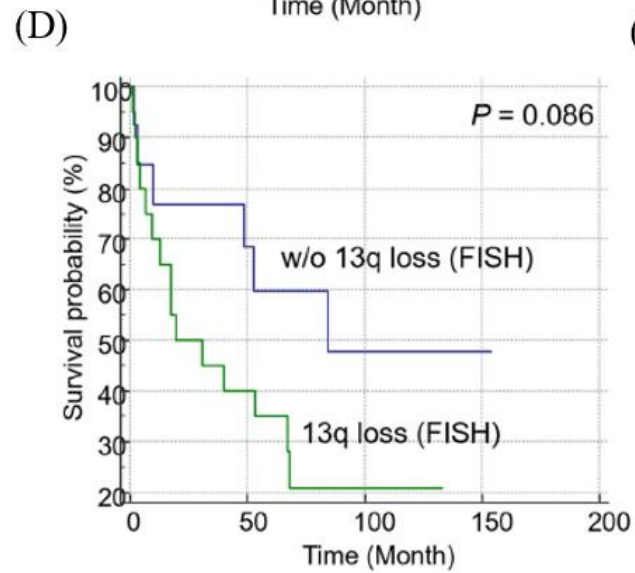
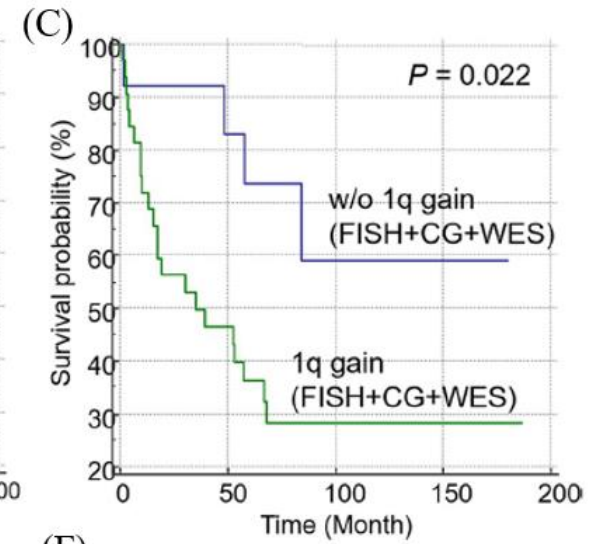
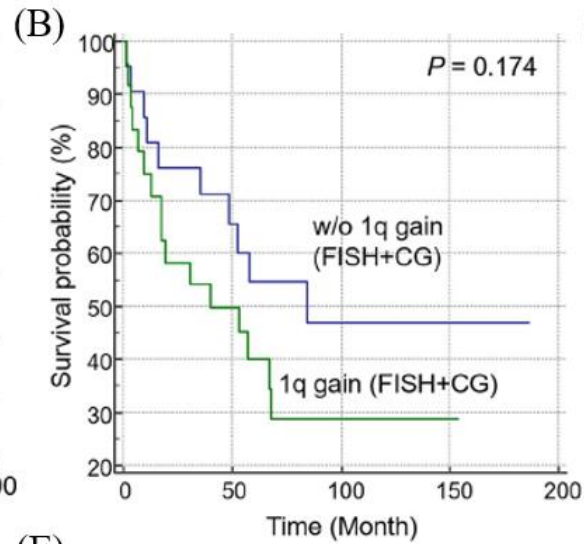
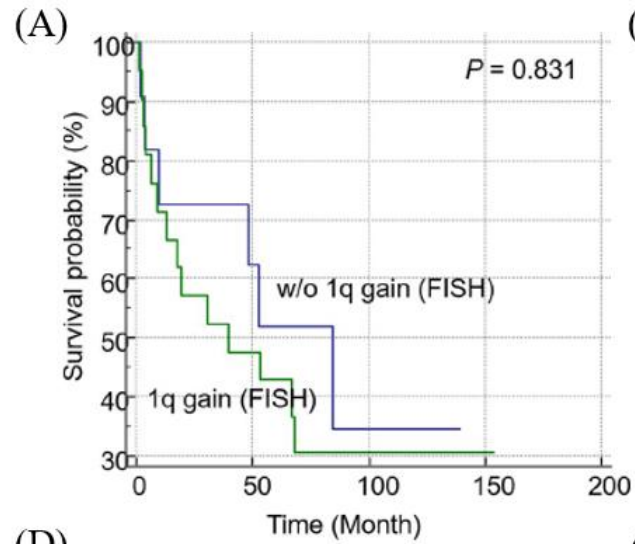


Fig 8. Kaplan–Meier survival curves of patients with copy number variations identified using fluorescent in situ hybridization (FISH), conventional cytogenetics (CG), and whole–exome sequencing (WES) analysis. Overall survival (OS) of patients with 1q gain identified through FISH (A) or FISH+CG (B) was not significantly different. Whereas, 1q gain identified through FISH+CG+WES analysis was significantly poor (C). Similarly, the OS of patients with 13q loss identified through FISH (D) and FISH+CG (E) was not significantly different. However, patients with 13q loss identified through FISH+CG+WES analysis exhibited significantly poor prognosis (F).

3.5.3. Mutational burden (CNVs and driver gene mutations) as a prognostic factor

The distribution of the number of mutated driver genes in each patient is shown in Fig. 9A. One or more mutations were detected in 63 patients (94%). The median number of total mutations was 2.0 (95% CI = 2.0–3.0; range = 0–8). The distribution plot of only tier I, II, and III mutations is shown in Fig. 9B. The median number of mutations was 1.0 (95% CI = 1.0–2.0; range = 0–8). The CNVs were detected in 56 patients (83.6%). The median total CNV number in each patient was 6.0 (95% CI = 5.7–7.0; range 0–16; Fig 9C).

KM analysis revealed that both CNVs and driver gene mutations were associated with poor prognosis in the group with a high mutational burden. Patients with ≥ 4 gene mutations exhibited lower OS than those with < 4 gene

mutations ($P = 0.035$; Fig 9D). Additionally, patients with ≥ 3 mutant driver genes (tier I, II, and III) exhibited poorer prognosis than patients with < 3 mutant driver genes ($P = 0.033$; Breslow test; Fig 9E). The survival rate of patients with ≥ 4 CNVs was lower than that of patients with < 4 CNVs ($P = 0.035$; Fig 9F).

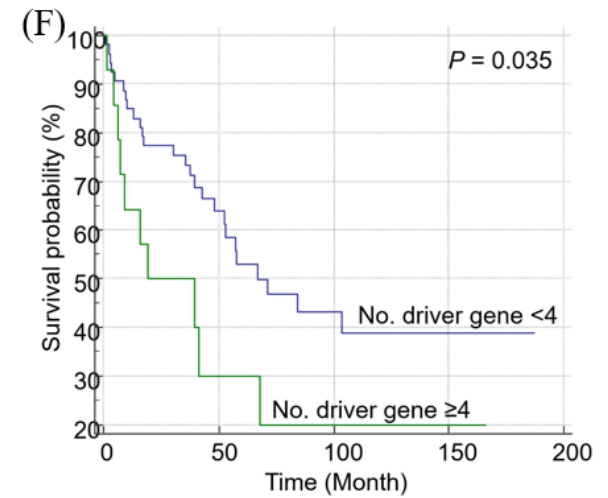
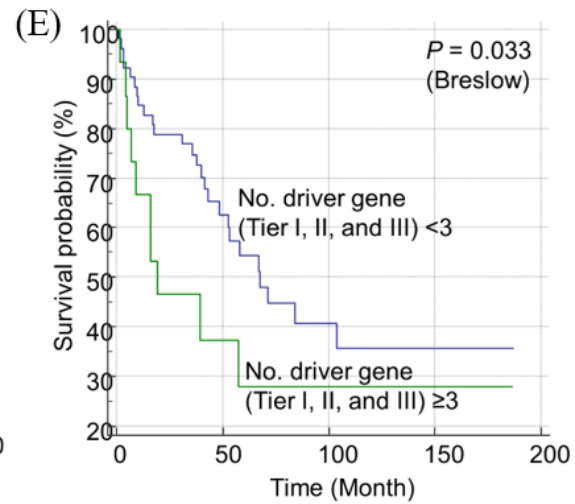
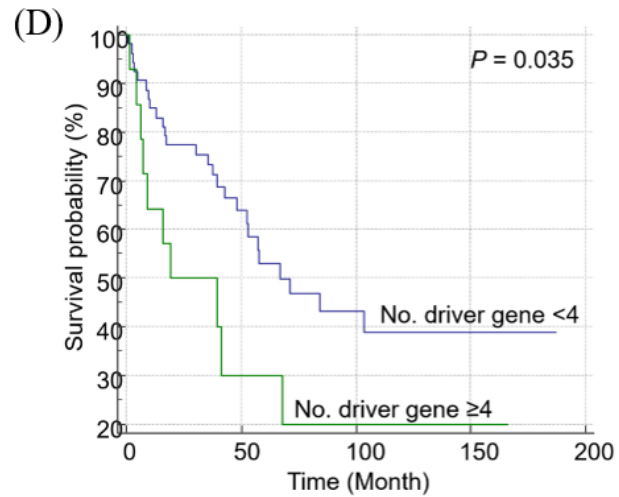
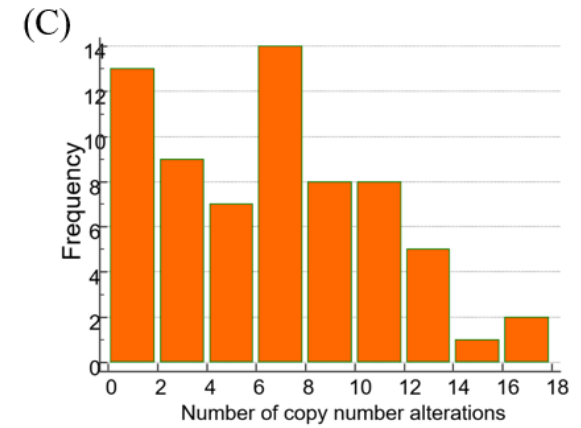
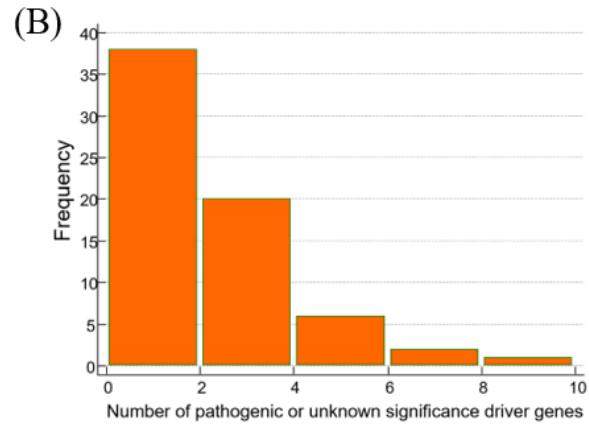
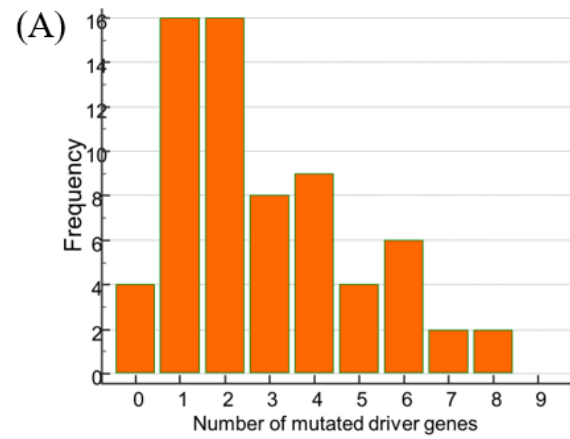


Fig 9. Distribution of the number of mutated driver genes in each patient. Distribution plots for the total number of (A) driver mutations, (B) selected mutations with tier I, II, and III according to the ASCO/AMP/CAP guidelines, and (C) copy number variations in each patient. Kaplan–Meier plots revealed that patients with (D) > 5 gene variants, (E) > 4 selected gene mutations, and (F) > 5 copy number variations exhibited poor overall survival.

3.6. Factors affecting OS in patients with hyperdiploid NDMM

Among patients with hyperdiploid NDMM, patients with ≥ 5 trisomies exhibited better prognosis than those with < 5 trisomies ($P = 0.037$; Fig 8A). Patients with < 4 driver gene mutations exhibited a more favorable prognosis than those with ≥ 4 driver gene mutations ($P = 0.004$; Fig 8B). The survival analysis of patients with hyperdiploid NDMM with or without driver gene mutations revealed that patients without *BRAF* mutations exhibited a better prognosis than those with *BRAF* mutations ($P < 0.001$; Fig 8C).

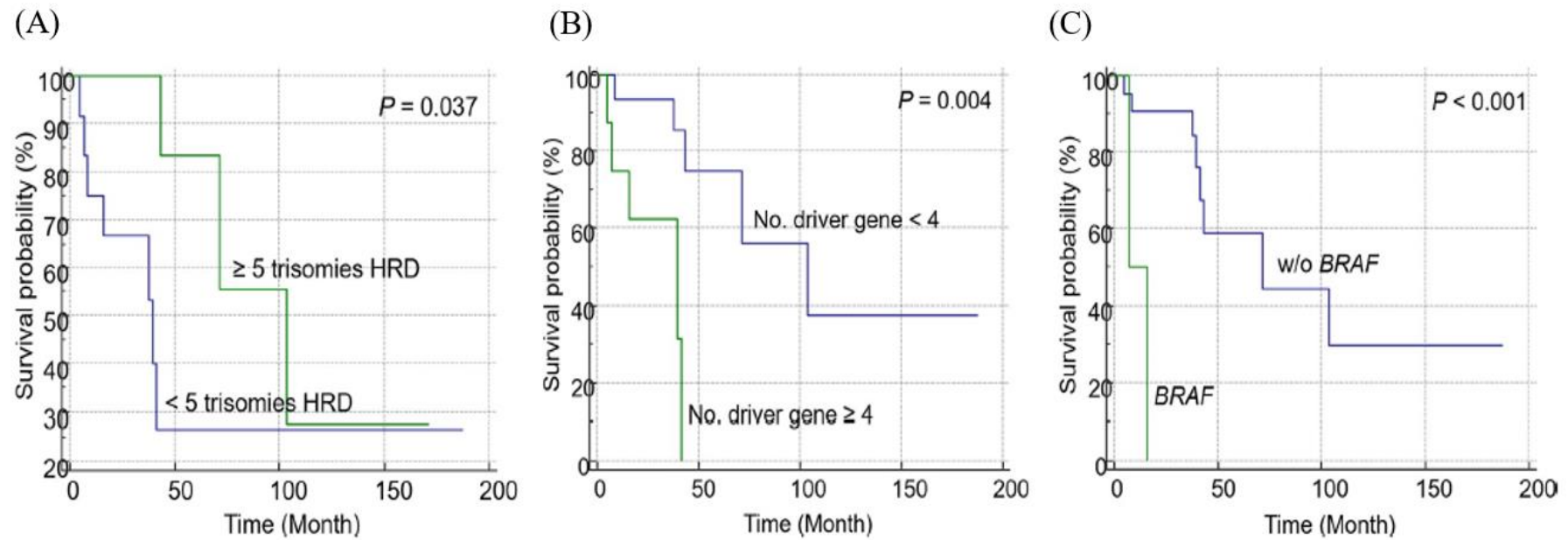
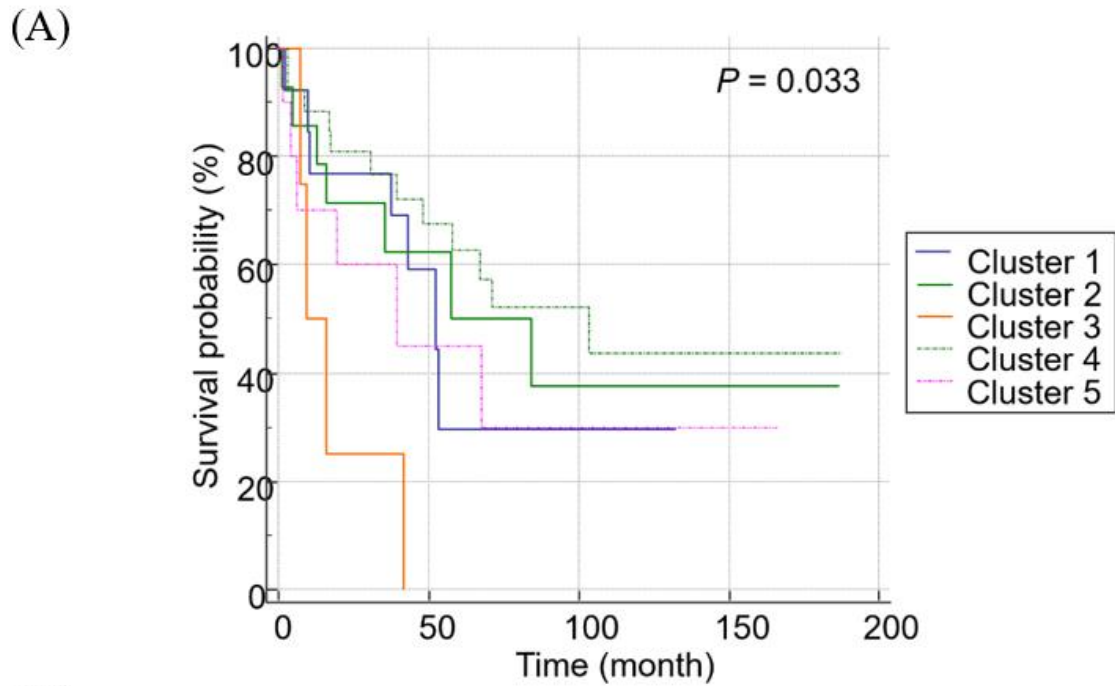


Fig 10. Kaplan–Meier curves showing the overall survival of patients with hyperdiploid multiple myeloma (MM). Patients with hyperdiploid (HRD) MM were subdivided based on the (A) number of trisomies (≥ 5 vs. < 5) ($P = 0.037$), (B) number of mutated driver genes (≥ 3 vs. < 3) ($P < 0.001$), and (C) the presence of *BRAF* mutation ($P < 0.001$).

3.7. OS of clusters classified based on K-means analysis

Several factors affecting the survival rate of patients with NDMM identified in this study were classified according to the K-means clustering method. The following factors were included in the analysis: True-hyperdiploid (T-HRD) (>5 trisomy), 1q gain, 6q loss, 6p gain, 8p loss, 13q loss, 17 loss, t(4:14), t(14:16), and the total number of driver gene mutations for each patient. Upon classification into five clusters, 1q gain, 6p gain, 8p loss, 13q loss, and the number of driver gene mutations were identified as the significant clustering factors. Among them, the cluster 3 group with 1q gain, 6p gain, 13q loss, and driver gene mutation number of 7.5 as the clustering center was associated with poor prognosis. In contrast, the cluster 4 group with no CNVs and a low centering value with a driver gene mutation number of 1.8 exhibited the best prognosis. The cluster 5 group with driver gene mutation number of 5.6, 1q gain, and 13q loss exhibited significantly poorer OS than the cluster 4 group ($P = 0.033$; Fig 11).



(B)

Factors	Cluster					P^*
	1 (N = 13)	2 (N = 14)	3 (N = 4)	4 (N = 26)	5 (N = 10)	
T-HRD	0.385	0.143	0.000	0.154	0.000	0.117
1q gain	0.692	0.714	0.750	0.269	0.600	0.020
6q loss	0.154	0.214	0.500	0.038	0.200	0.135
6p gain	0.308	0.500	0.250	0.077	0.200	0.049
8p loss	0.462	0.214	0.250	0.000	0.100	0.005
13q loss	0.769	0.500	0.750	0.269	0.600	0.027
17 loss	0.077	0.143	0.000	0.000	0.200	0.236
t(4:14)	0.231	0.143	0.000	0.038	0.000	0.216
t(14:16)	0.077	0.071	0.000	0.038	0.000	0.884
No. Driver genes	0.692	3.643	7.500	1.846	5.600	<0.001

Fig 11. Kaplan–Meier curves for overall survival (OS) with subdivided driver groups. (A) The cluster 3 group exhibited significantly lower OS than clusters 1, 2, 4, and 5 ($P = 0.019$). (B) Final cluster centers for variables, including those used for K–means analysis, are classified based on each cluster. The

significant clustering factors were 1q gain, 6p gain, 8p loss, 13q loss, and the number of driver gene mutations ($P < 0.05$). For example, the cluster 3 group, which exhibited the worst prognosis, had a clustering center with 1q gain, 6p gain, 13q loss, and driver gene mutation number of 7.5.

*P-value was calculated using analysis of variance.

4. Discussion

In this study, the CNV and somatic variant profiles of Korean patients with NDMM were identified using WES. We identified that 94.0% of enrolled patients had one or more driver gene mutations, and found frequently mutated genes including *IGLL5*, *ATM*, *NRAS*, *KRAS*, *DIS3*, *MAN2C1*, *BHLHE41*, *MAML2*, *DUSP2*, and *BRAF*. A comprehensive analysis revealed that the gain of 1q and 6p chromosome arms and the loss of 6q, 8q, and 13q chromosome arms were associated with poor OS. Additionally, *MAML2* and *BHLHE41* mutations were shown to have an adverse prognostic impact on OS. Patients with a high frequency of CNVs or a high number of mutations exhibited poor prognosis. Cluster analysis revealed that patients with the highest number of total driver gene mutations along with 1q gain, 6p gain, and 13q loss were associated with the poorest prognosis. These findings suggested the utility of integrated analysis of CNVs and somatic mutations in predicting prognosis.

The differences in the genomic profile between different ethnic groups were examined. The genomic profile of Korean patients with NDMM was compared with that reported previously [30–33]. Among the CNVs, 1q gain (52.2%) and 13q loss (49.3%) were most frequently detected, followed by losses of 8p, 14q, 1p, and 6q. The prevalence of 6p gain and 1q gain (26–45%) in this study was higher than that reported in previous studies. The frequency

of 1q gain detection using FISH was high in Korean patients with myelodysplastic syndrome [34]. This indicated that 1q gain is a candidate CNV that is specific for Korean patients with myeloma. Among the somatic variants, the frequency of mutations in the following genes was higher than 6%: *IGLL5*, *ATM*, *NRAS*, *KRAS*, *DIS3*, *MAN2C1*, *BHLHE41*, *MAML2*, and *BRAF*. The frequency of *IGLL5* mutation was high (18%), which is reported to occur exclusively with *KRAS/NRAS* mutation [35]. In contrast, the frequency of mutated genes in the MAPK pathway, such as *KRAS*, *NRAS*, and *BRAF*, was lower than that reported in previous studies on the Caucasian population (20–36%) [9, 32, 36, 37]. This low frequency might be owing to the characteristics of the enrolled patients. Previously, Cifola et al. reported that the frequency of *KRAS/NRAS* mutations was low in plasma cell leukemia, which is the aggressive and high-risk form of plasma cell dyscrasia [38]. As this study performed integrative analysis on patients with $\geq 60\%$ plasma cells in the BM aspirate, the characteristics of patients with advanced disease could contribute to the low frequency of *KRAS/NRAS* mutations in this study. Meanwhile, the incidence of other mutated genes in patients with NDMM varied in different studies. This variability might result from the heterogeneous composition of patients with myeloma rather than the ethnic difference.

CNVs are considered to be one of the most important drivers of cancer development and progression [12–14, 31, 32, 37, 39, 40]. Recently, various methods have been developed for analyzing CNV through WES, which has

enabled the integrated analysis of CNVs and somatic mutations [41] in addition to overcoming the limitations of CNV detection through CG and/or FISH. In the case of NDMM, the prognostic relevance was not clearly revealed except for the well-known variants such as t(4;14) and del(17). The effects of these abnormalities, especially for 1q gain and 13q loss on prognosis are controversial, and only a few studies have included them for risk stratification criteria [42–44]. The prognostic discrepancy could be caused by method sensitivity in addition to several unmeasuring confounding factors. A previous study showed that 5–53% of patients who were normal in FISH and cytogenetics had abnormality in chromosomal microarray test [45]. CNV detection through WES can overcome the limitations associated with low proliferative activity of cells or a low number of neoplastic plasma cells in the BM when compared with CNV detection through CG. Moreover, the detection of CNVs through single nucleotide polymorphism array or WES, which can encompass the entire chromosome, would be more helpful in predicting prognosis than FISH, which only detects small target chromosomal areas. In this study, 1q gain and 13q loss identified through WES analysis adversely affected OS, whereas 1q gain and 13q loss identified through FISH and/or CG methods did not have prognostic relevance. The deletion of whole arm of 13q (N = 29; 87.9%) and the gain of whole arm of 1q (N = 30; 85.7%) were observed in a majority of patients with 13q deletion and 1q gain in this study. Previously, Binder et al. analyzed the prognosis of patients with monosomy 13 and del(13q) separately in 1,181

NDMM patients, and reported that only patients with monosomy 13 had adverse prognosis [46]. Both of these studies suggest that the 13q loss, which had previously shown neutral or favorable results, may have included many patients with interstitial deletion rather than whole chromosome loss. Furthermore, it provides additional rationale for enabling the integration of WES and FISH tests to provide enhanced genetic information for risk stratification and improved prediction of outcomes.

Additionally, WES could explore the regions that are not covered by routine FISH probes. Recently, CNVs including 6p gain, 6q loss, and 8p loss have been reported to emerging recurrent alterations in multiple myeloma [45]. As CNV detection technologies have been developed, interests in the clinical implications of these new and emerging marker have raised. In this study, patients with 6p gain, 6q loss, and 8p loss were statistically significantly associated with poor prognosis. Further studies are needed to validate these CNVs as adverse prognostic markers in a large cohort.

In this study, *BHLHE41* and *MAML2* somatic mutations were associated with poor prognosis. *BHLHE41*, which is located at 12p12.1, is known as a helix–loop–helix superfamily domain that is involved in various cellular functions, such as proliferation, differentiation, tumorigenesis, and circadian rhythms [47]. The expression of *BHLHE41* is reported to be upregulated in patients with Waldenstrom macroglobulinemia [48], and might be associated with poor prognosis in NDMM. *MAML2* is located at 11q21. The prognostic

relevance of MECT1–MAML2 and CRTC1–MAML2 fusion oncogenes has been reported in mucoepidermoid carcinomas [49, 50]. In this study, *MAML2* mutations exhibited positive correlation with various CNVs including 17q loss and 22q loss, and these variations may aid in predicting the prognosis. *TP53* mutations are a well-known adverse prognostic factor in myeloma. Niccolo et al. demonstrated that *TP53* mutations were associated with adverse progression-free survival (PFS) and OS and that a rare mutation in *DNAH11* affected OS [9]. Another study reported that *TP53*, *ATR*, *ATM*, and *ZFHX4* mutations are associated with poor PFS or OS [37]. Although *TP53* did not show statistical significance in OS in this study (P = 0.613), it is presumed to be due to too small number of mutation-positive patients (N = 3). However, it could be owing to the differences in inclusion criteria, such as selecting only patients with >60% PC or targeting the Korean population. Most previous studies report the heterogeneous mutational landscape of MM. Some patients exhibited redundancy in gene mutations as two or more mutations were detected in genes involved in the same pathway [12, 51, 52]. Consistent with these findings, the current study suggested the heterogeneity of genomic variants in MM. We suggest that the prediction power of mutational burden in structural and somatic variants is higher than that of a single variation.

In this study, we presented novel prognostic subgroups in patients with NDMM through K-means clustering analysis. The clusters were divided according to CNVs and mutations that affect OS. The “Cluster 3” exhibited

the highest number of driver gene mutations and CNVs, including 1q gain, 13q loss, and 6q loss. Although efforts to stratify potential subgroups according to OS were attempted previously [9, 10], the characteristics of each cluster group have not been distinguished. However, this study revealed that prognosis-related CNVs, such as 1q gain, 13q loss, and 6q loss have clustering relevance, whereas somatic gene mutations have no clustering importance. The overall number of mutations in the driver gene panel played a significant role in stratifying each group.

Recently, prognosis-related classifications are being reconsidered for patients with hyperdiploid MM [53, 54]. In this study, patients with hyperdiploid MM and more than five trisomies exhibited favorable OS, which was based on a review of hyperdiploid variations that affected prognosis. Moreover, patients with hyperdiploid MM and an increased number of driver gene mutations exhibited poor prognosis. These results suggest that the integrated analysis of mutation burden (both CNVs and somatic mutations) using WES could aid in precisely predicting the clinical outcomes of patients with MM.

This study has several limitations: CD138 purification and sorting steps have not been implemented. Instead, only patients with >60% plasma cells in the BM aspirate were selected for the study. Most patients showed high numbers of plasma cells in the BM. Additionally, there were no control samples to remove the germline background. To overcome this limitation, the gene variants observed in healthy Korean individuals (n = 2,000) were removed. And,

we focused on the driver genes selected by multiple algorithms in several previous studies. Based on the ASCO/AMP/CAP guidelines, variants with tier I, II, and III were separately analyzed. Furthermore, CNVs under 2,500 kb were excluded in CNV analysis. However, this study demonstrated that large CNVs with size $> 2,500$ kb were poor prognostic factors. As CNV analysis cannot detect chromosomal translocations, the results of translocations, such as t(4;14) and t(14;16) were referenced for clustering integrated analysis.

In conclusion, this study comprehensively analyzed the somatic mutations and CNVs of patients with NDMM using WES. Additionally, this study proposed a new method for classifying patient groups with poor prognosis and predicting OS. MM is a heterogeneous disease comprising several subclones, and multiple driver gene mutations are detected in one patient. Therefore, an integrated analysis through the application of WES in the future would aid in predicting the prognosis in a clinical setting.

References

1. Hong J, Lee JH. Recent advances in multiple myeloma: a Korean perspective. *Korean J Intern Med.* 2016;31(5):820–34.
2. Palumbo A, Anderson K. Multiple myeloma. *The New England journal of medicine.* 2011;364(11):1046–60.
3. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. *Nature reviews Cancer.* 2012;12(5):335–48.
4. Kumar SK, Rajkumar SV, Dispenzieri A, Lacy MQ, Hayman SR, Buadi FK, et al. Improved survival in multiple myeloma and the impact of novel therapies. *Blood.* 2008;111(5):2516–20.
5. Kumar SK, Dispenzieri A, Lacy MQ, Gertz MA, Buadi FK, Pandey S, et al. Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia.* 2014;28(5):1122–8.
6. Avet-Loiseau H, Durie BG, Cavo M, Attal M, Gutierrez N, Haessler J, et al. Combining fluorescent in situ hybridization data with ISS staging improves risk assessment in myeloma: an International Myeloma Working Group collaborative project. *Leukemia.* 2013;27(3):711–7.
7. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al. Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *Journal of clinical oncology :*

official journal of the American Society of Clinical Oncology. 2015;33(26):2863–9.

8. Maura F, Bolli N, Angelopoulos N, Dawson KJ, Leongamornlert D, Martincorena I, et al. Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nature communications*. 2019;10(1):3835.

9. Bolli N, Biancon G, Moarii M, Gimondi S, Li Y, de Philippis C, et al. Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia*. 2018;32(12):2604–16.

10. Walker BA, Mavrommatis K, Wardell CP, Ashby TC, Bauer M, Davies FE, et al. Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. 2018;132(6):587–97.

11. Walker BA. Whole Exome Sequencing in Multiple Myeloma to Identify Somatic Single Nucleotide Variants and Key Translocations Involving Immunoglobulin Loci and MYC. *Methods in molecular biology* (Clifton, NJ). 2018;1792:71–95.

12. Robiou du Pont S, Cleynen A, Fontan C, Attal M, Munshi N, Corre J, et al. Genomics of Multiple Myeloma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2017;35(9):963–7.

13. Kim M, Lee SH, Kim J, Lee SE, Kim YJ, Min CK. Copy number variations could predict the outcome of bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. *Genes, chromosomes & cancer*. 2015;54(1):20–7.

14. Kim M, Ju YS, Lee EJ, Kang HJ, Kim HS, Cho HC, et al. Abnormalities in Chromosomes 1q and 13 Independently Correlate With Factors of Poor Prognosis in Multiple Myeloma. *Annals of laboratory medicine*. 2016;36(6):573–82.
15. Koh Y, Kim D, Jung WJ, Ahn KS, Yoon SS. Revealing Genomic Profile That Underlies Tropism of Myeloma Cells Using Whole Exome Sequencing. *International journal of genomics*. 2015;2015:675379.
16. Ryu D, Kim HJ, Joung JG, Lee HO, Bae JS, Kim SJ, et al. Comprehensive genomic profiling of IgM multiple myeloma identifies IRF4 as a prognostic marker. *Oncotarget*. 2016;7(30):47127–33.
17. Aldhyani THH, Alshebami AS, Alzahrani MY. Soft Clustering for Enhancing the Diagnosis of Chronic Diseases over Machine Learning Algorithms. *J Healthc Eng*. 2020;2020:4984967–.
18. Ambigavathi M, Sridharan D, editors. *Analysis of Clustering Algorithms in Machine Learning for Healthcare Data*. International Conference on Advances in Computing and Data Sciences; 2020: Springer.
19. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. *BMC bioinformatics*. 2013;14 Suppl 11(Suppl 11):S1.
20. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS computational biology*. 2016;12(4):e1004873.

21. Rajkumar SV, Dimopoulos MA, Palumbo A, Blade J, Merlini G, Mateos MV, et al. International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *The Lancet Oncology*. 2014;15(12):e538–48.
22. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)*. 2009;25(14):1754–60.
23. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next–generation DNA sequencing data. *Genome research*. 2010;20(9):1297–303.
24. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research*. 2003;31(13):3812–4.
25. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non–synonymous variants on protein function using the SIFT algorithm. *Nature protocols*. 2009;4(7):1073–81.
26. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics*. 2014;46(3):310–5.
27. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nature methods*. 2010;7(4):248–9.
28. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for

Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *The Journal of molecular diagnostics* : JMD. 2017;19(1):4–23.

29. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* (Oxford, England). 2004;5(4):557–72.

30. Walker BA, Leone PE, Jenner MW, Li C, Gonzalez D, Johnson DC, et al. Integration of global SNP-based mapping and expression arrays reveals key regions, mechanisms, and genes important in the pathogenesis of multiple myeloma. *Blood*. 2006;108(5):1733–43.

31. Avet-Loiseau H, Li C, Magrangeas F, Gouraud W, Charbonnel C, Harousseau JL, et al. Prognostic significance of copy-number alterations in multiple myeloma. *Journal of clinical oncology* : official journal of the American Society of Clinical Oncology. 2009;27(27):4585–90.

32. Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nature reviews Clinical oncology*. 2017;14(2):100–13.

33. Shah V, Sherborne AL, Walker BA, Johnson DC, Boyle EM, Ellis S, et al. Prediction of outcome in newly diagnosed myeloma: a meta-analysis of the molecular profiles of 1905 trial patients. *Leukemia*. 2018;32(1):102–10.

34. Lee DS, Kim SH, Seo EJ, Park CJ, Chi HS, Ko EK, et al. Predominance of trisomy 1q in myelodysplastic syndromes in Korea: is there an ethnic difference?

A 3-year multi-center study. *Cancer genetics and cytogenetics*. 2002;132(2):97–101.

35. White BS, Lanc I, O'Neal J, Gupta H, Fulton RS, Schmidt H, et al. A multiple myeloma-specific capture sequencing platform discovers novel translocations and frequent, risk-associated point mutations in IGLL5. *Blood cancer journal*. 2018;8(3):35.

36. Hu Y, Chen W, Wang J. Progress in the identification of gene mutations involved in multiple myeloma. *OncoTargets and therapy*. 2019;12:4075–80.

37. Walker BA, Boyle EM, Wardell CP, Murison A, Begum DB, Dahir NM, et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2015;33(33):3911–20.

38. Cifola I, Lionetti M, Pinatel E, Todoerti K, Mangano E, Pietrelli A, et al. Whole-exome sequencing of primary plasma cell leukemia discloses heterogeneous mutational patterns. *Oncotarget*. 2015;6(19):17543–58.

39. Prideaux SM, Conway O'Brien E, Chevassut TJ. The Genetic Architecture of Multiple Myeloma. *Advances in Hematology*. 2014;2014:864058.

40. Walker BA, Leone PE, Chiecchio L, Dickens NJ, Jenner MW, Boyd KD, et al. A compendium of myeloma-associated chromosomal copy number abnormalities and their prognostic value. *Blood*. 2010;116(15):e56–65.

41. Marchuk DS, Crooks K, Strande N, Kaiser–Rogers K, Milko LV, Brandt A, et al. Increasing the diagnostic yield of exome sequencing by copy number variant analysis. *PloS one*. 2018;13(12):e0209185.
42. Mikhael JR, Dingli D, Roy V, Reeder CB, Buadi FK, Hayman SR, et al. Management of newly diagnosed symptomatic multiple myeloma: updated Mayo Stratification of Myeloma and Risk–Adapted Therapy (mSMART) consensus guidelines 2013. *Mayo Clinic proceedings*. 2013;88(4):360–76.
43. Sonneveld P, Avet–Loiseau H, Lonial S, Usmani S, Siegel D, Anderson KC, et al. Treatment of multiple myeloma with high–risk cytogenetics: a consensus of the International Myeloma Working Group. *Blood*. 2016;127(24):2955–62.
44. Munshi NC, Anderson KC, Bergsagel PL, Shaughnessy J, Palumbo A, Durie B, et al. Consensus recommendations for risk stratification in multiple myeloma: report of the International Myeloma Workshop Consensus Panel 2. *Blood*. 2011;117(18):4696–700.
45. Pugh TJ, Fink JM, Lu X, Mathew S, Murata–Collins J, Willem P, et al. Assessing genome–wide copy number aberrations and copy–neutral loss–of–heterozygosity as best practice: An evidence–based review from the Cancer Genomics Consortium working group for plasma cell disorders. *Cancer genetics*. 2018;228–229:184–96.
46. Binder M, Rajkumar SV, Ketterling RP, Greipp PT, Dispenzieri A, Lacy MQ, et al. Prognostic implications of abnormalities of chromosome 13 and the

presence of multiple cytogenetic high-risk abnormalities in newly diagnosed multiple myeloma. *Blood cancer journal*. 2017;7(9):e600.

47. Shen Z, Zhu L, Zhang C, Cui X, Lu J. Overexpression of BHLHE41, correlated with DNA hypomethylation in 3'UTR region, promotes the growth of human clear cell renal cell carcinoma. *Oncology reports*. 2019;41(4):2137–47.

48. Trojani A, Greco A, Tedeschi A, Lodola M, Di Camillo B, Ricci F, et al. Microarray demonstrates different gene expression profiling signatures between Waldenström macroglobulinemia and IgM monoclonal gammopathy of undetermined significance. *Clinical lymphoma, myeloma & leukemia*. 2013;13(2):208–10.

49. Behboudi A, Enlund F, Winnes M, Andrén Y, Nordkvist A, Leivo I, et al. Molecular classification of mucoepidermoid carcinomas—prognostic significance of the MECT1–MAML2 fusion oncogene. *Genes, chromosomes & cancer*. 2006;45(5):470–81.

50. Anzick SL, Chen WD, Park Y, Meltzer P, Bell D, El-Naggar AK, et al. Unfavorable prognosis of CRTC1–MAML2 positive mucoepidermoid tumors with CDKN2A deletions. *Genes, chromosomes & cancer*. 2010;49(1):59–69.

51. Leich E, Weissbach S, Klein HU, Grieb T, Pischmarov J, Stuhmer T, et al. Multiple myeloma is affected by multiple and heterogeneous somatic mutations in adhesion- and receptor tyrosine kinase signaling molecules. *Blood cancer journal*. 2013;3:e102.

52. Bolli N, Avet-Loiseau H, Wedge DC, Van Loo P, Alexandrov LB, Martincorena I, et al. Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications*. 2014;5:2997.
53. Chretien M-L, Corre J, Lauwers-Cances V, Magrangeas F, Cleyne A, Yon E, et al. Understanding the role of hyperdiploidy in myeloma prognosis: which trisomies really matter? *Blood*. 2015;126(25):2713-9.
54. Barilà G, Bonaldi L, Grassi A, Martines A, Liço A, Macri N, et al. Identification of the true hyperdiploid multiple myeloma subset by combining conventional karyotyping and FISH analysis. *Blood cancer journal*. 2020;10(2):18.

국문 초록

이번 연구는 처음 진단된 다발골수종 환자에서 유전자 변이 및 구조적 변이의 예후를 분석하기 위해 기존의 세포유전학, 형광접합보인자법 및 전체엑솜염기서열분석을 이용하여 통합적인 유전체 분석을 수행하였다. 골수흡인액에서 60% 이상의 형질 세포를 보이는 다발골수종 환자 67 명을 대상으로 골수 유핵 세포에서 전체엑솜염기서열분석을 수행하였다. 체세포 돌연변이 및 구조적변이 분석에는 두 가지의 응용 프로그램(CNVkit 및 Nexus Copy Number)이 이용되었다. 63 명의 다발골수종 환자에서 83 개의 유발 유전자 돌연변이가 발견되었다. 환자당 체세포 돌연변이 수의 중앙값은 2.0 이었다(95 % 신뢰 구간 = 2.0-3.0, 범위 = 0-8). *MAML2* 및 *BHLHE41* 돌연변이는 짧은 생존률과 관련이 있었다. 구조적 변이는 56 명의 환자 (72.7%; 56/67)에서 발견되었다. 환자 당 구조적 변이 수의 중앙값은 6.0 (95 % CI = 5.7 - 7.0, 범위 = 0 - 16)이었다. 구조적 변이 중 1q 증폭, 6p 증폭, 6q 결실, 8p 결실 및 13q 결실은 생존 감소와 관련이 있었다. 또한 1q 증폭과 6p 증폭은 독립적인 불량한 예후 인자였다. 한 환자에서 보이는 높은 빈도의 구조적 변이와 유발 유전자 돌연변이는 불량한 임상 예후와 상관관계가 높았다. 클러스터링 분석을 통해 1q 증폭, 6p 증폭 및 13q 결실과 함께 가장 높은 빈도의 유발 유전자 돌연변이를 가진 환자가 가장 나쁜 예후를 나타냄을 알 수 있었다. 다발골수종 기존의 알려진 예후인자와 함께, 체세포 돌연변이와 구조적변이의 통합 분석은 환자의 보다 정확한 예후 예측 및 분류에 있어 의미가 높을 것으로 기대된다.