SUPPLEMENTARY APPENDIX

The spectrum of somatic mutations in monoclonal gammopathy of undetermined significance indicates a less complex genomic landscape than that in multiple myeloma

Aneta Mikulasova,[1,2,3,4] Christopher P. Wardell,[1] Alexander Murison,[5] Eileen M. Boyle,[5] Graham H. Jackson,[6] Jan Smetana,[2,3] Zuzana Kufova,[7,8] Ludek Pour,[9] Viera Sandecka,[9] Martina Almasi,[10] Pavla Vsianska,[10] Evzen Gregora,[11] Petr Kuglik,[2,3] Roman Hajek,[7,8] Faith E. Davies,[1] Gareth J. Morgan[1] and Brian A. Walker[1]

[1]Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA; [2]Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic; [3]Department of Medical Genetics, University Hospital Brno, Czech Republic; [4]Department of Pathological Physiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic; [5]Center for Myeloma Research, Division of Molecular Pathology, Institute of Cancer Research, London, UK; [6]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, UK; [7]Faculty of Medicine, University of Ostrava, Czech Republic; [8]Department of Hematooncology, University Hospital Ostrava, Czech Republic; [9]Department of Internal Medicine, Hematology and Oncology, University Hospital Brno, Czech Republic; [10]Department of Clinical Hematology, University Hospital Brno, Czech Republic and [11]Department of Internal Medicine and Hematology, University Hospital Kralovske Vinohrady, Prague, Czech Republic

 doi:10.3324/haematol.2017.163766

Received: January 10, 2017.

Accepted: May 16, 2017.

Pre-published: May 26, 2017.

Correspondence: bwalker2@uams.edu

Somatic mutation spectrum in monoclonal gammopathy of undetermined significance indicates a less complex genomic landscape compared to multiple myeloma

Running head: Somatic mutation spectrum in MGUS compared to myeloma

Authors

Aneta Mikulasova,[1,2,3,4] Christopher P. Wardell,[1] Alexander Murison,[5] Eileen M. Boyle,[5] Graham H. Jackson, [6] Jan Smetana,[2,3] Zuzana Kufova,[7,8] Ludek Pour,[9] Viera Sandecka,[9] Martina Almasi,[10] Pavla Vsianska,[10] Evzen Gregora,[11] Petr Kuglik,[2,3,4] Roman Hajek,[7,8] Faith E. Davies,[1] Gareth J. Morgan,[1] and Brian A. Walker[1]

Affiliations

[1]Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR, USA
[2]Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czech Republic
[3]Department of Medical Genetics, University Hospital Brno, Brno, Czech Republic
[4]Department of Pathological Physiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic
[5]Center for Myeloma Research, Division of Molecular Pathology, Institute of Cancer Research, London, United Kingdom
[6]Northern Institute for Cancer Research, Newcastle University, Newcastle upon Tyne, United Kingdom
[7]Faculty of Medicine, University of Ostrava, Ostrava, Czech Republic
[8]Department of Hematooncology, University Hospital Ostrava, Ostrava, Czech Republic
[9]Department of Internal Medicine, Hematology and Oncology, University Hospital Brno, Brno, Czech Republic
[10]Department of Clinical Hematology, University Hospital Brno, Brno, Czech Republic
[11]Department of Internal Medicine and Hematology, University Hospital Kralovske Vinohrady, Prague, Czech Republic

Supplementary Data

Supplementary Methods – Exome Sequencing

Data quality metrics and processing

FastQC (v0.10.0) was used for basic quality control of Illumina paired-end sequencing data. These files were aligned to the reference genome (GRCh37), using BWA (v.0.5.9) followed by Stampy (v.1.0.20) to improve gapped alignment. BAM files were recalibrated using GATK (v2.3.9) and deduplicated using Picard (v.1.85). Tumor and normal samples were realigned as pairs using the GATK indel realigner to improve indel call rates. Calls from a panel of 6668 highly variable SNPs within the exome capture were compared between the tumor and normal samples to confirm that they were correctly paired. Sequencing was performed to a mean depth of 67x for tumor samples and 64x for peripheral blood samples.

Somatic mutation calling

Single nucleotide variants (SNVs) were called using MuTect (v1.1.4). They were further filtered using the following criteria: minimum of 10x depth at that site in both tumor and normal BAMs, a minimum of 1 non-reference base call in both directions, a mean Phred quality score of greater than 26 for that base in the tumor sample, a mean mapping quality score of equal to or greater than 50 across all reads at that site in the tumor sample, and the site must be uniquely alignable according to the CRG alignability tracks available from the UCSC genome browser and created using the GEM mappability tool.[1] C>A|G>T SNVs likely to be oxidation artefacts created during library preparation were removed.[2]

Short indels were called using the GATK Indelocator and further filtered according to the following criteria: a minimum of 1 non-reference base call in both directions, a mean Phred quality score of greater than 26 for that base in the tumor sample, a mean mapping quality score of equal to or greater than 50 across all reads at that site in the tumor sample, and the site must be uniquely alignable. No more than 2 reads covering that site in the normal sample could contain any indel. A window of 21 base pairs centered on the first base of the indel was taken and had to conform to the following rules which remove low-complexity sequences:

no homopolymers of greater than 6 base pairs and no dinucleotide may occur more than 5 times.

All somatic events were annotated using both SnpEff (v3.1) and Oncotator (v0.4.2.2) with SnpEff providing the most deleterious interpretation regardless of transcript and Oncotator annotating only the canonical transcript. Significantly mutated genes were detected by providing all SNV and short indels to the MutSigCV (v1.4) algorithm.[3] A q-value cut-off of 0.1 was used. Common frequently mutated genes (FLAGS) were filtered out from list of recurrent mutated genes.[4]

Translocations were called using Delly[5] and manually curated in the Integrative Genomics Viewer (IGV). Translocations were considered real when there were at least 10 supporting reads and if no translocations were also found in the peripheral blood sample.

Nonnegative matrix factorization (NMF)

Nonnegative matrix factorization (NMF) using the NMF package in R was used to factorize a matrix of frequency of trinucleotide mutation contexts per sample, in order to identify underlying mutation signatures among all mutations in the 33 samples.[6] The algorithm was run for between 1 and 18 underlying signatures, with 50 runs at each number.

The optimum number of signatures was determined to be two based on a number of quality control metrics. The signatures were compared to the 30 existing COSMIC signatures[6] using cosine similarity.

References

1. Derrien T, Estelle J, Marco Sola S, et al. Fast computation and applications of genome mappability. PLoS One. 2012;7(1):e30377.
2. Costello M, Pugh TJ, Fennell TJ, et al. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. Nucleic Acids Res. 2013;41(6):e67.
3. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-218.
4. Shyr C, Tarailo-Graovac M, Gottlieb M, Lee JJ, van Karnebeek C, Wasserman WW. FLAGS, frequently mutated genes in public exomes. BMC Med Genomics. 2014;7(1):64.
5. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-i339.
6. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature. 2013;500(7463):415-421.

Supplementary Tables

Supplementary Table S1: Clinical parameters of 33 MGUS patients.

Variable	MGUS cohort
Median age (range)	61 (35–86)
Sex ratio (M:F)	1.5:1
Median follow-up (months)	49 (0-188)
Isotype	
IgG	23 (69.7%)
IgA	6 (18.2%)
IgM	2 (6.1%)
Biclonal	1 (3.0%)
Light chain only	1 (3.0%)
Light chain type	
Kappa	16 (48.5%)
Lambda	17 (51.5%)
Serum M-Ig	
<15 g/l	29 (87.9%)
≥15 g/l	4 (12.1%)
% of PCs in bone marrow	
<5%	28 (84.8%)
≥5%	5 (15.2%)
Serum kappa/lambda FLC ratio	
Normal	17 (51.5%)
Abnormal	15 (45.5%)
Data missing	1 (3.0%)
Risk group[a]	
Low risk	14 (42.4%)
Low-intermediate risk	9 (27.3%)
High-intermediate risk	9 (27.3%)
High risk	0 (0.0%)
Data missing	1 (3.0%)

[a] Rajkumar SV, Kyle RA, Therneau TM, *et al.* Serum free light chain ratio is an independent risk factor for progression in monoclonal gammopathy of undetermined significance. Blood. 2005;106(3):812-817.

Supplementary Table S2: Overview of CNAs found by CGH+SNP arrays in 33 MGUS patients.

Chromosome	Copy-number losses Cases (%)	Copy-number gains Cases (%)
1p	2 (6.1%)	–
1q	1 (3.0%)	9 (27.3%)
2p	–	–
2q	–	–
3p	–	6 (18.2%)
3q	–	6 (18.2%)
4p	–	4 (12.1%)
4q	1 (3.0%)	4 (12.1%)
5p	–	4 (12.1%)
5q	–	4 (12.1%)
6p	–	5 (15.2%)
6q	2 (6.1%)	4 (12.1%)
7p	1 (3.0%)	5 (15.2%)
7q	–	5 (15.2%)
8p	1 (3.0%)	1 (3.0%)
8q	–	2 (6.1%)
9p	1 (3.0%)	9 (27.3%)
9q	–	10 (30.3%)
10p	–	1 (3.0%)
10q	–	1 (3.0%)
11p	–	3 (9.1%)
11q	–	3 (9.1%)
12p	–	–
12q	–	–
13q	10 (30.3%)	–
14q	5 (15.2%)	1 (3.0%)
15q	–	5 (15.2%)
16p	–	1 (3.0%)
16q	2 (6.1%)	1 (3.0%)
17p	–	5 (15.2%)
17q	–	5 (15.2%)
18p	–	4 (12.1%)
18q	–	4 (12.1%)
19p	–	9 (27.3%)
19q	–	9 (27.3%)
20p	–	–
20q	–	–
21q	2 (6.1%)	3 (9.1%)
22q	2 (6.1%)	1 (3.0%)
Xp	6 (18.2%)	1 (3.0%)
Xq	6 (18.2%)	2 (6.1%)
Yp	4 (12.1%)	1 (3.0%)
Yq	4 (12.1%)	1 (3.0%)

– indicates not present.

Supplementary Table S3: Number of SNVs per case in MGUS compared to NDMM.

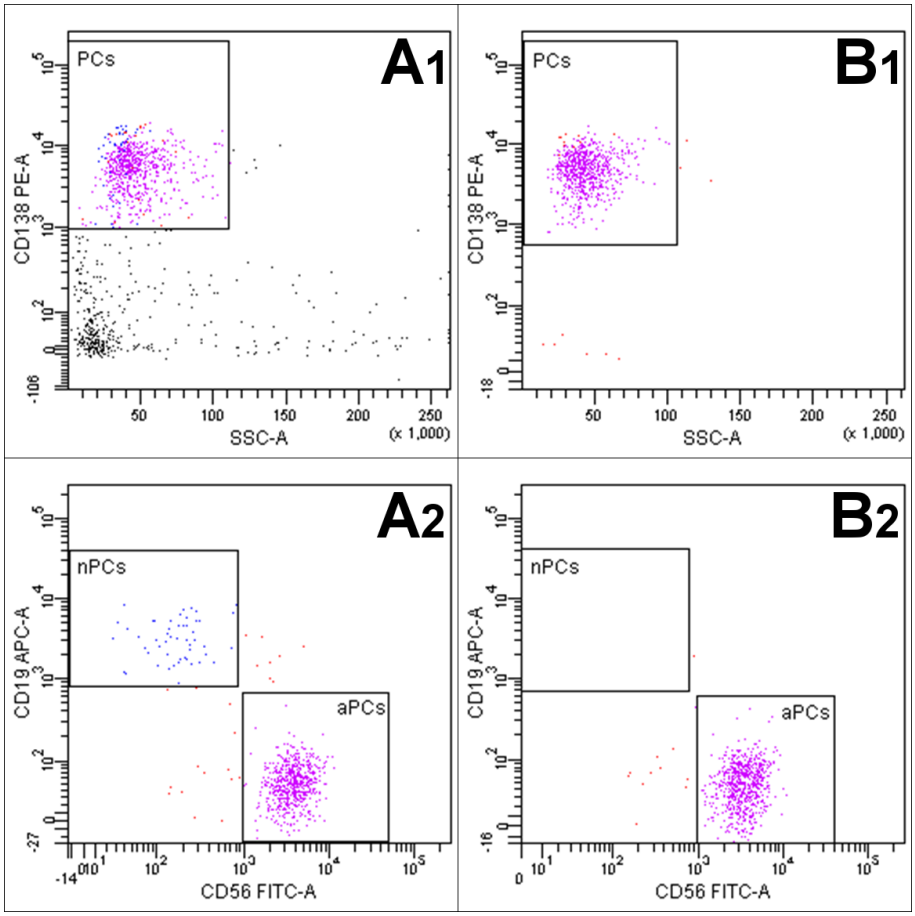
SNV category	MGUS (n = 33) Median (range)	NDMM (n = 463) Median (range)	*P*
Total SNVs	89 (9–315)	123 (1–897)	7.04×10^{-5}
Exonic SNVs	27 (2–111)	48 (0–599)	8.80×10^{-9}
NS-SNVs	19 (0–70)	38 (0–452)	2.49×10^{-9}
S-SNVs	6 (1–42)	12 (0–151)	5.86×10^{-5}

Supplementary Table S4: List of 5 MGUS patients with myeloma-significantly mutated genes found in 463 NDMM dataset.

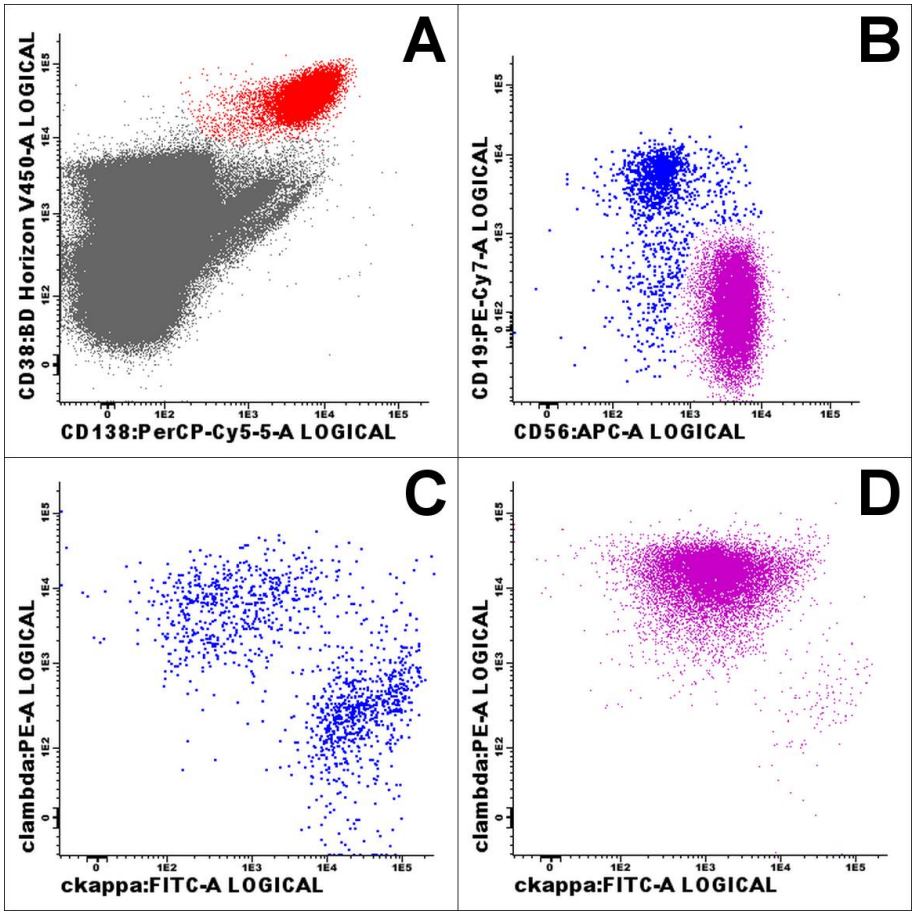
MGUS case	Gene (Mutation)	CCF	Clonal/Subclonal
1	*EGR1* (p.M29L)	0.61	Clonal
	LTB (p.T56M)	0.80	Clonal
2	*KRAS* (p.Q61L)	0.25	Subclonal
	NRAS (p.G13R)	0.16	Subclonal
3	*HIST1H1E* (p.V57L)	1.00	Clonal
	HIST1H1E (p.S89T)	0.77	Clonal
4	*KRAS* (p.A146T)	0.72	Clonal
5	*HIST1H1E* (p.A65P)	0.78	Clonal
	DIS3 (p.D488N)	0.18	Subclonal

Supplementary Figures

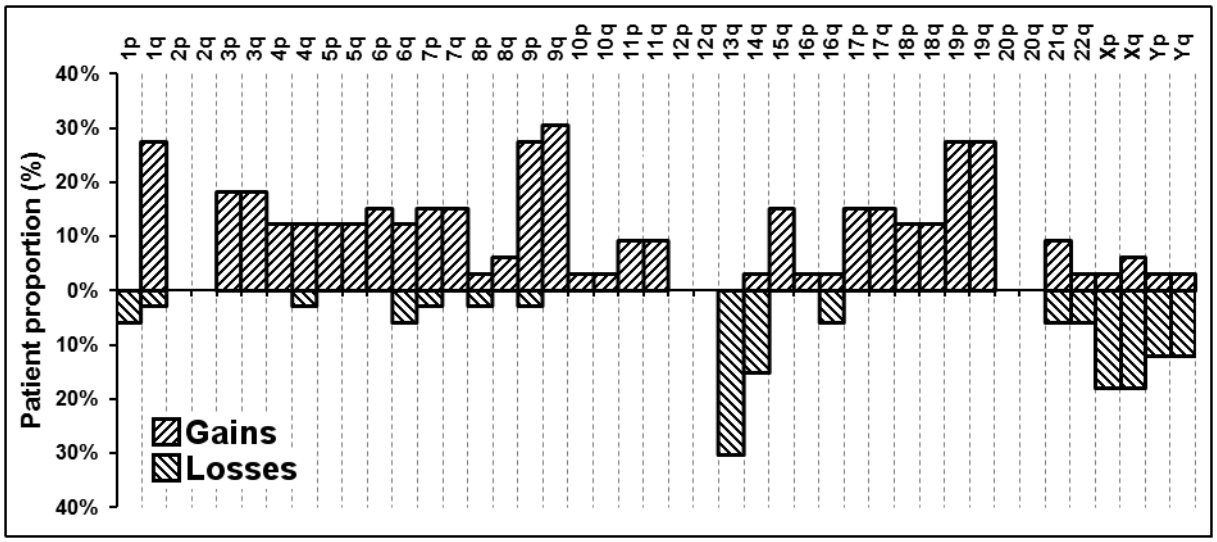
Supplementary Figure S1: Illustration of PCs sorting from mononuclear cell fraction (A) and purity evaluation of post-sorting cell suspension (B) in MGUS. PCs were identified according to CD138 and side scatter (SSC) (**A1**, **B1**) and divided according to expression of CD19 and CD56 into normal PCs (nPCs) and abnormal PCs (aPC) (**A2**, **B2**). Cell sorting as well as purity evaluation were made by BD FACSAria using acquisition software BD FACSDiva Software 6.1.3 (both BD Biosciences, San Jose, CA, USA).



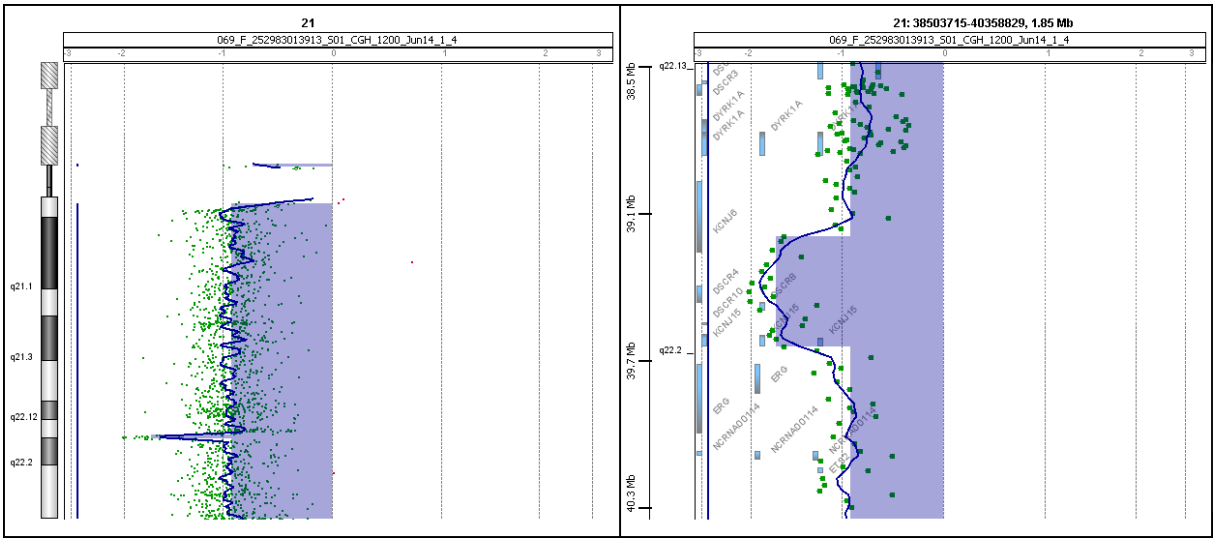
Supplementary Figure S2: Illustration of flow-cytometry analysis of whole bone marrow PCs clonality in MGUS. Total PCs were identified according to CD38 and CD138 expression (**A**). Identification of normal PCs (blue dots) and abnormal PCs (purple dots) was done based on surface expression of CD19 and CD56 (**B**) and clonality was confirmed according to cytoplasmic immunoglobulin κ and λ expression – normal PCs with normal ratio κ/λ (**C**) vs. λ+ restricted abnormal PCs (**D**). Analyses were made by flow cytometr BD FACSCantoII using acquisition software BD FACSDiva Software 6.1.3 (both BD Biosciences, San Jose, CA, USA) and analysis software Infinicyt 1.6.0 (Cytognos S.L., Salamanca, Spain).



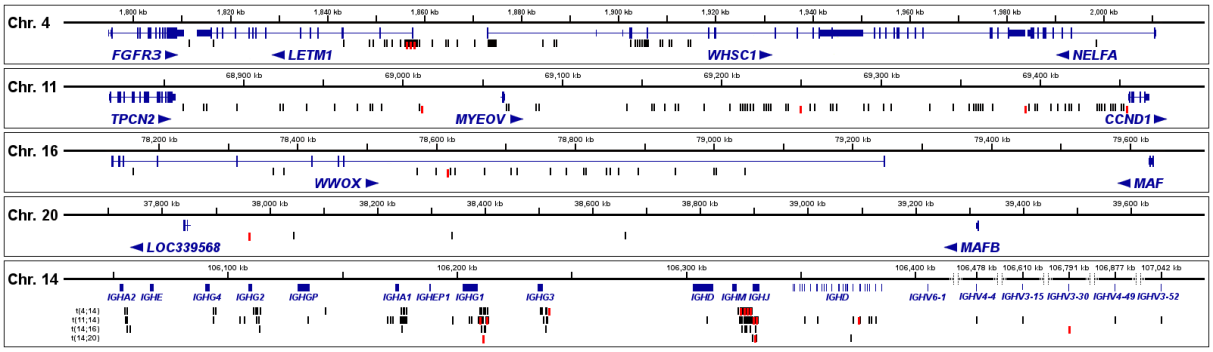
Supplementary Figure S3: Summary plot of CNAs at the chromosome-arm level in 33 MGUS patients. Chromosome arms are along the horizontal axis and frequencies of abnormalities are along the vertical axis.



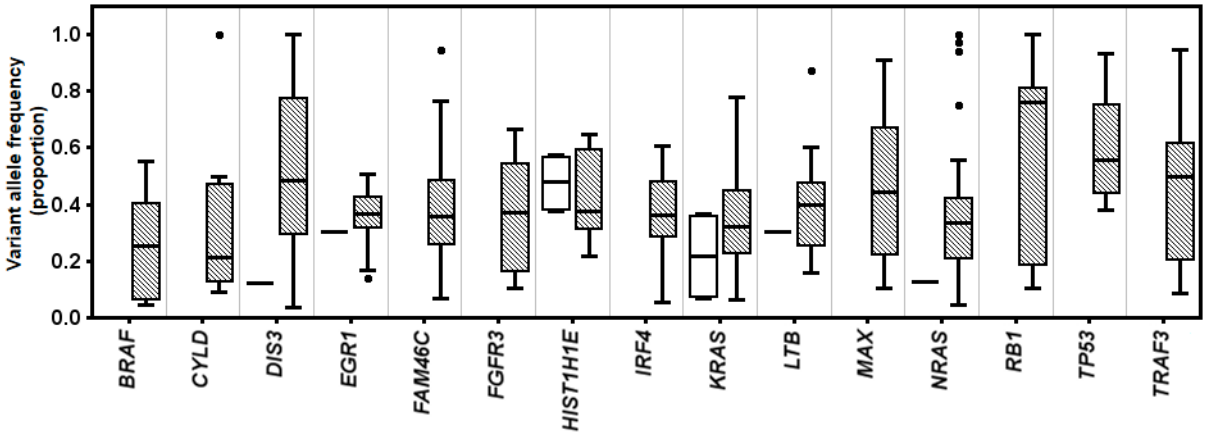
Supplementary Figure S4: Homozygous deletion affecting *KCNJ6*, *DSCR4*, *DSCR8*, *DSCR10*, *KCNJ15* genes at 21q22.13 in MGUS patient.



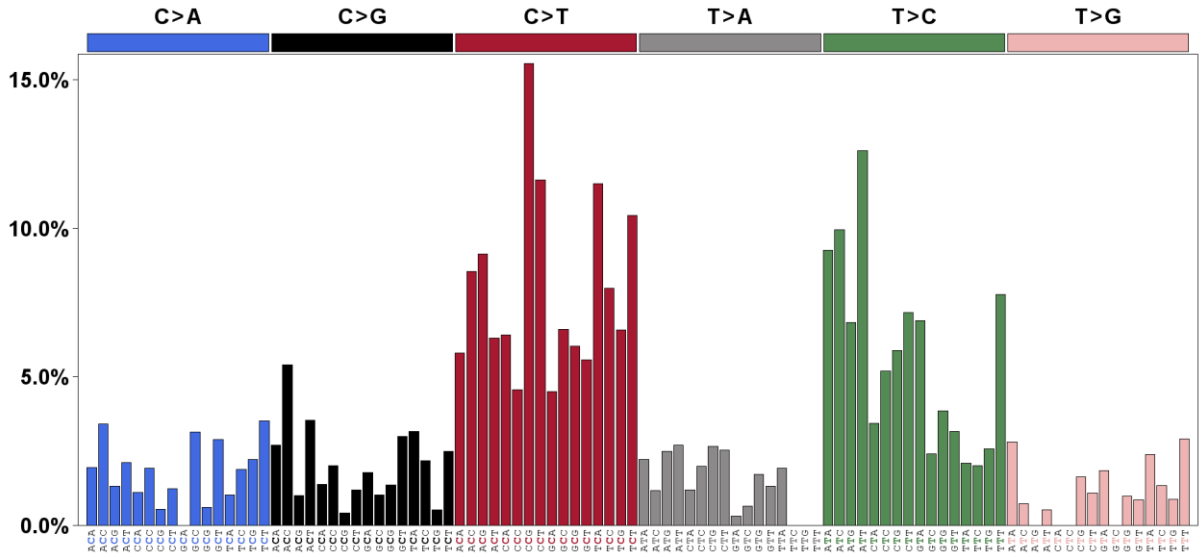
Supplementary Figure S5: Translocation breakpoints in 9 cases of MGUS compared to 166 cases of NDMM with *IGH* translocations as type t(4;14), t(11;14), t(14;16) and t(14;20). Genes are showed by blue color with arrows as transcription direction. Breakpoints found in MGUS and NDMM cases are marked by red and black vertical lines, respectively.



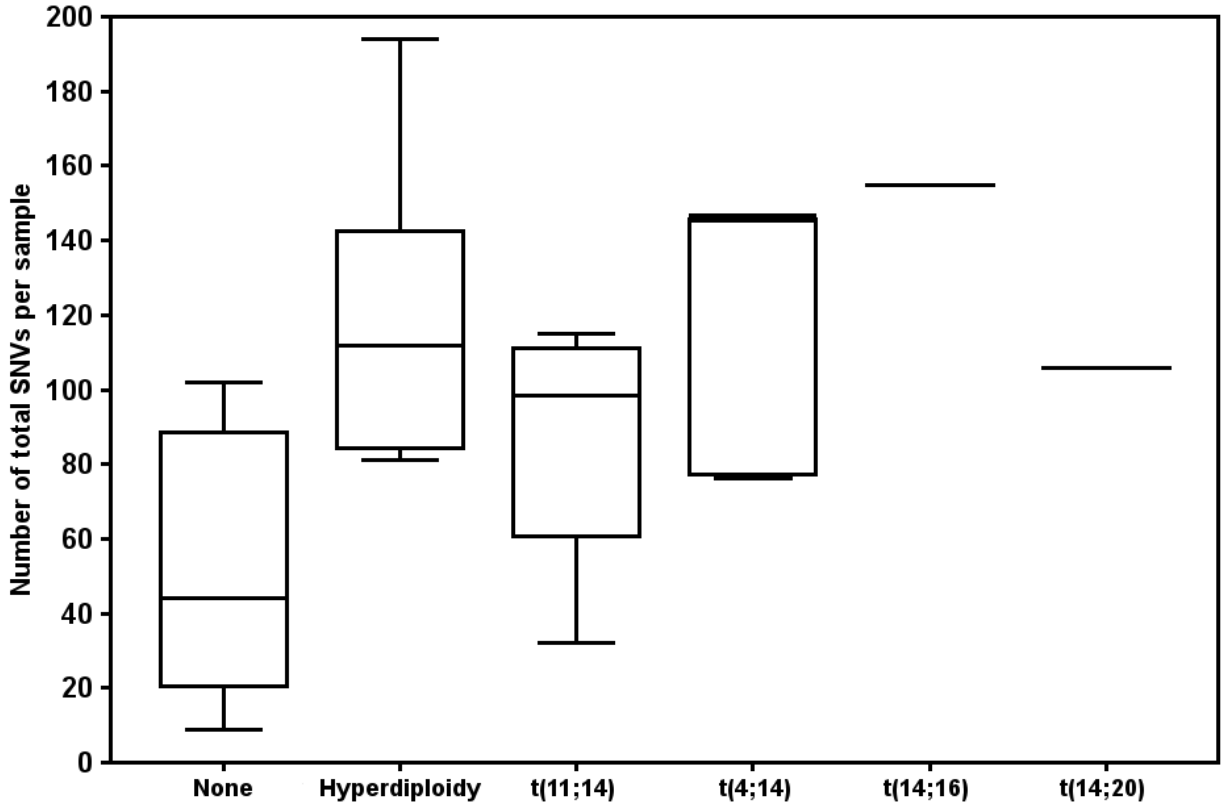
Supplementary Figure S6: Variant allele frequency of 15 myeloma-significantly mutated genes found in 463 NDMM cohort. Empty and filled boxes show cohorts of 33 MGUS and 463 NDMM patients, respectively.



Supplementary Figure S7. Mutational signature in the cohort of 33 MGUS patients. Mutation type are along the horizontal axis and percentage of mutations are on the vertical axes. Different colors show specific substitutions class. Cosine similarity between this signature and COSMIC signature 5 of unknown etiology was 0.89, suggesting that they are the same. All other detected signatures had cosine similarities of <0.5 to other COSMIC signatures and therefore seem unlikely to be genuine signals.



Supplementary Figure S8. Number of total SNVs per MGUS sample in specific sub-groups defined by presence and type of *IGH* translocation.



Supplementary Figure S9. Association between SNVs, chromosome abnormalities and clinical parameters in the cohort of 33 MGUS patients.