Impact of rare structural variant events in newly diagnosed multiple myeloma

Monika Chojnacka[1], **Benjamin Diamond**[1], **Bachisio Ziccheddu**[1], **Even Rustad**[2], **Kylee Maclachlan**[3], **Marios Papadimitriou**[1], **Eileen M. Boyle**[4], **Patrick Blaney**[4], **Saad Usmani**[3], **Gareth Morgan**[4], **Ola Landgren**[1,*], **Francesco Maura**[1,*]

[1.] Myeloma Division, Sylvester Comprehensive Cancer Center, University of Miami, Miami, FL, USA.

[2.] Institute for Cancer Research, Oslo University Hospital, Oslo, Norway.

[3.] Myeloma Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[4.] Myeloma Research Program, NYU Langone, Perlmutter Cancer Center, New York, NY, USA.

Abstract

Purpose: Whole genome sequencing (WGS) of newly diagnosed multiple myeloma patients (NDMM) has shown recurrent structural variant (SV) involvement in distinct regions of the genome (i.e. hotspots) and causing recurrent copy number alterations. Together with canonical immunoglobulin translocations, these SVs are recognized as "recurrent SVs". More than half SVs were not involved in recurrent events. The significance of these "rare SVs" has not been previously examined.

Patients and Methods: In this study, we utilize 752 WGS and 591 RNA-seq data from NDMM patients to determine the role of rare SVs in myeloma pathogenesis.

Results: 94% of patients harbored at least one rare SV event. Rare SVs showed an SV-class specific enrichment within genes and superenhancers associated with outlier gene expression. Furthermore, known myeloma driver genes recurrently impacted by point mutations were dysregulated by rare SVs.

CORRESPONDING AUTHORS: Francesco Maura, MD, Division of Myeloma, Sylvester Comprehensive Cancer Center, University of Miami, 1120 NW 14th Street, Clinical Research Building, Miami, FL 33136, Phone (305) 243-7687, fxm557@med.miami.edu.
*These authors contributed equally

Conclusions: Overall, we demonstrate the association of rare SVs with aberrant gene expression supporting a potential driver role in myeloma pathogenesis.

INTRODUCTION

Whole genome sequencing (WGS) has shown that the clinical behavior of various cancer types is driven by a complex genomic landscape characterized by multiple genomic drivers and mutational processes (1, 2). Among these somatic events, structural variants (SV) have emerged as key events causing the acquisition of copy number alterations (CNA), creating gene fusions, and dysregulating gene expression through superenhancer hijacking and the disruption of 3D genomic structure (1, 3–6). A single SV is comprised of 2 breakpoints and classified according to four main classes: duplication, deletion, inversion, and translocation. Each SV unit may exist as a singleton or as part of a complex SV event – such as chromothripsis, templated insertion, or chromoplexy, and occur at various rates across cancer types (1, 3, 7).

Multiple myeloma (MM) is a clinically and biologically heterogenous hematological malignancy characterized by an expansion of abnormal plasma cells in the bone marrow (6, 8). Historically, SV analysis in MM has been restricted to recurrent translocations between immunoglobulin genes (Ig) and key oncogenes such as *CCND1*, *MAF* and *NSD2*. These "Canonical Translocation" events are considered to be one of few initiating genomic myeloma drivers due to their clonal characteristic of conservation across several evolutionary phases of tumor development, their strong impact on gene expression and consequent impact on disease biology (9, 10).

Analysis of large datasets of clinically annotated WGS paired with RNA sequencing (RNA-seq) have allowed the interrogation of SVs within the MM genomic landscape. Multiple studies have shown a heavy SV enrichment in MM beyond the canonical SV events, which account for only a minor fraction of the full SV repertoire (11–14). Interestingly, the MM SV landscape shows a high prevalence of chromothripsis and chromoplexy, and has one of the highest prevalence of templated insertions described in cancer (12, 15). More than 60 recurrent SV genomic hotspot have been reported and linked to multiple driver genes and regulatory regions (i.e., "SV Hotspot") (12). In addition to these recurrent events, SVs outside of SV hotspots and canonical Ig events are often responsible for recurrent copy number alterations in 152 previously defined genomic regions, here named "GISTIC CNA" after the method of discovery (12). Despite this comprehensive characterization of SVs to date, more than half SV are not involved in a recurrent "Canonical Translocation", "SV Hotspot" or "GISTIC CNA" events. It is largely unknown if these unclassified SV events, here defined as "rare SV", play a potential driver or passenger role in MM disease biology and whether we can gain further insights into recurrently impacted pathways that could constitute novel targets for therapy. Overall, the relevance of rare SV events is not new in MM. In fact, several rare events have been reported and linked to distinct clinical behavior, like rare nonsynonymous mutations on *CRBN, IKZF3*, or *BCMA* deletions after exposure to immunomodulatory agents and CAR-T, respectively (16–18). In this work, to expand our understanding on the role of rare SVs in relation to clinical and biological heterogeneity, and potential implications on the development of individualized treatment strategies, we

utilize 752 WGS and 591 RNA-seq data in a newly diagnosed MM (NDMM) patient cohort, demonstrating that many rare SV events are not passengers but rather have a potential major impact on the gene expression and tumor biology in individual patients.

METHODS

Patient Cohort and Sample Processing

WGS and RNA-sequencing of bone marrow biopsies were utilized from NDMM patients enrolled in the CoMMpass study (NCT01454297; IA13). Processing of sequencing data was conducted as previously described in the preceding publication [9]. In brief, WGS samples underwent low-coverage long-insert sequencing (median 4–8X). Low-coverage sequencing prevented clonality analysis of genomic variants. Paired-end reads were aligned to human reference genome (USCS hg19) with the Burrow Wheeler Aligner (BWA; v0.7.8; RRID:SCR_010910). CNA were identified utilizing tCoNuT (https://github.com/tgen/MMRF_CoMMpass/tree/master/tCoNut_COMMPASS) and externally validated with controlFREEC as previously described (12). SV identification was performed by two SV callers, DELLY (v0.7.6; RRID:SCR_004603) and Manta (v1.5.0; RRID:SCR_022997). The final SV catalogue was obtained by quality filtering calls and validating by copy number alterations as previously described (12). Annotation of SVs as being involved in complex SV events or as stand-alone SV breakpoint pairs was conducted. SV considered to be causative of a copy number alteration were logged when the SV breakpoint occurs within 50 Kb of a copy number variable start or end position.

RNA-seq data with a target coverage depth of 100 million reads were utilized from 591 patients. Paired-end reads were aligned to the human reference genome (UCSC hg19) utilizing the STAR aligner (STAR; v2.3.1z; RRID:SCR_004463). Transcripts were normalized to TPM utilizing Salmon (v7.2; RRID:SCR_017036).

Defining Rare Structural Variants

Identification of recurrent SVs occurred in a step-wise fashion following an SV involvement hierarchy: 1) SVs involved in canonical myeloma translocations occurring between Ig loci (*IGH, IGK*, and *IGL*) or their partner genes (*CCND1, CCND2, CCND3, MAF, MAFA, MAFB,* and *NSD2*), 2) SVs occurring within or 100 kb from SV hotspot boundaries, and 3) SVs responsible for CNA in the patient cohort, as identified by the GISTIC algorithm (RRID:SCR_000151). SV hotspot and recurrent CNA region catalogues are reported with our previous publication(12). We considered the varying nature of recurrent CNA regions involving GISTIC peaks with SV breakpoints as well. CNA regions involving GISTIC peaks vary in size. Both focal (<3Mb) and large CNAs overlapping GISTIC peaks whose breakpoints were within 50kb of an SV were considered causal, involved, and recurrent. Any SV breakpoint involved in at least one of the above event categories defined the SV event as a recurrent event. Therefore, any remaining SV events were defined rare, not involving any known genomic region clinically defined in myeloma disease progression or targeted at significant rates across the patient cohort under investigation. Recurrent and rare SV identification in naïve and memory B-cells, as well as the precursor conditions, was conducted in a similar manner.

Gene Expression Effects of Rare Structural Variants

Genes within SV deletion and duplication class were considered affected by the SV by default. Additionally, genes overlapping part of a duplication and deletion class SV, as well as those occurring up to 1 Mb away were considered in analysis. Genes up to 1 Mb away from inversions and translocations without copy number support were also considered. All SV events, single and complex, were removed from analysis if at least one breakpoint in the event occurred within 1 Mb of any Ig loci, to remove strong immunoglobulin associated enhancer effects. Genes involved in recurrent events were removed from consideration in rare SV analysis. Each rare SV breakpoint was considered individually, cataloguing breakpoints occurring within the gene body, 10–100 Kb, 100 Kb-1 Mb of genes with outlier expression. To examine the relationship between rare SV with outlier expression was analyzed. Extending on methods developed to test the penetrance of rare germline events, genes paired to rare SVs were considered affected if the gene expression was above a gene specific outlier z-score of $+/- 2$ (19). Gene outliers were defined as patient specific gene expression with a z-score of $+/- 2$. To determine if there is a class specific interaction with over-expressed and under-expressed gene expression outliers, a re-shuffle permutation was run on observed counts to model a random distribution of SV breakpoints. If the observed rates of class-specific SV to gene proximity was above the 97.5% CI defined by the random model, the SV class was defined as enriched, having a relationship to gene expression outliers. Cancer related gene data sets were pulled from Memorial Sloan Kettering's FDA recognized tumor mutation database OncoKB (RRID:SCR_014782) and COSMIC's Cancer Gene Census (RRID:SCR_002260), two publicly available repositories (20, 21). Myeloma driver gene list was pooled from previous publications on driver gene discovery in multiple myeloma (10, 22).

Superenhancers Involved by Rare Structural Variants

To investigate the relationship and possibility of rare SV influencing gene expression in an indirect manner, we analyzed the relationship between rare structural variants and superenhancers (23–25). We first observed SV breakpoint density, up to 10 Mb from the start and end of each superenhancer, considering the closest superenhancer to each breakpoint of the SV. For complex events where multiple breakpoints affect the same superenhancer, only 1 breakpoint per patient was counted. Because we are interested in SV event-to-enhancer relationships, not the number of breakpoints within each superenhancer, the analysis was performed this way to prevent artificially inflating SV involvement in superenhancers. To determine from which distance rare SVs target superenhancers, a re-shuffle permutation was performed randomly shifting the original breakpoint $+/- 10$–20 Mb, to model a random distribution of SV breakpoints. To determine SV classes that were enriched within 1 Mb of superenhancers, a Fisher Exact test was performed on observed vs modeled counts of class specific SVs within 1 Mb of superenhancers. Breast cancer and DLBCL specific superenhancers were obtained from publicly available data (26, 27).

DATA AND SOFTWARE AVAILABILITY

All the raw data used in the study are already publicly available (dbGap: phs000748.v1.p1). Analysis was carried out in R version 3.6.1. The full analytic workflow is reported in Supplementary Data 1. All other software tools used are publicly available.

RESULTS

Structural variant and complex event annotation

To characterize the patterns and biological impact of rare SVs in NDMM, we interrogated 752 WGS samples from patients enrolled in the CoMMpass study (NCT01454297; IA13), 591 of which had patient matched RNA-seq data available.

An SV (i.e., deletion, duplication, inversion and translocation) can occur alone as its own SV event, or be part of a complex SV event comprised of multiple basic SVs. We focused on 4 main complex events – chromothripsis, templated insertions, chromoplexy and finally bulk categorizing all remaining complex events as "complex". Annotation of each SV and SV event followed a hierarchical system, first annotating each SV, followed by the SV event. The hierarchical system prioritizes SVs that cause canonical translocations, followed by SVs whose breakpoint falls within SV hotspots, SVs whose breakpoint falls within recurrent copy number altered regions (GISTIC CNA), and finally rare SV events. An SV will first be annotated by the highest priority genomic region it involves. In the context of complex events, the highest priority basic SV will determine the event's annotation (Fig. 1A; Supplementary Fig. S1; Supplementary Tables 1–2; Methods) (12). Any SV with both breakpoints not included in any of the three recurrent groups was classified as rare. Reasonably, a complex event annotated as a recurrent SV can be composed of both recurrent and rare SVs (Fig. 1B–E). In contrast, a fully rare complex SV event will be composed of only rare SVs. Despite main drivers being involved by recurrent SVs, rare SVs can play an additive role, targeting genes with biological impact. In this work we investigated rare SVs in two contexts: 1) rare SVs within recurrent SV events, and 2) rare SVs within fully rare SV events.

Defining the landscape and role of rare SV and rare SV events in NDMM

At least one rare SV event was observed across 94% (705/752) of all patients (Fig. 2A, Supplementary Fig. S2A). Of the total 8,942 SV events across 752 patients, 269 (3%) were classified as Canonical Translocation SVs (128 *CCND1*, 8 *CCND2/CCND3*, 22 *MAF/A/B*, 87 *NSD2*, 24 other Ig translocations), 1,871 (21%) as SV hotspot, and 1,843 (20%) as GISTIC CNA SVs, leaving 4,959 (55%) fully rare SV events (Fig. 2A–B). Amongst the main MM cytogenetic groups, patients with canonical translocations involving *CCND1* showed significantly lower counts and proportions of rare SVs per event compared to *MAF/A/B, NSD2*, hyperdiploid and other translocation groups (p<0.001, using Wilcoxon test; Fig. 2A, Supplementary Fig. S2B, Supplementary Tables 3–4). Additionally, patients with *NSD2* translocations had considerably higher counts of rare SV per event compared to all other groups except the *CCND2/CCND3* group (p<0.01, Wilcoxon test; Supplementary Fig. S2B, Supplementary Table 3).

Overall, 42.7% (8,148/19,055) of basic SVs were directly involved in recurrent regions while 57.3% (10,907/19,055) basic SVs occurred in rare unclassified regions. Among rare SVs, 57.3% (6,256/10,907) occurred within fully rare SV events, while 4,651 (42.7%) occurred within recurrent SV events (i.e. complex events composed of both rare and recurrent SVs; Supplementary Fig. S2C). Across SV classes, 11% (63/547) of templated insertions, 12% (29/236) of chromothripsis, 18% (18/100) of chromoplexy, and 33% (159/471) of all other complex events were annotated as fully rare SV events. Among the single SV event classes, the ratio between rare and recurrent SV was lowest for translocations (31%; 334/1,069), followed by inversions (52%; 507/973) and duplications (53%; 886/1,661), and was highest among deletion (71%; 2,782/3,885) – with more rare deletions than in all recurrent class deletions combined (Fig. 2B, Supplementary Table 5). No SV class was found to be enriched between cytogenetic groups (p>0.05, Wilcoxon test). We were curious if naive and memory B-cells, as well as plasma cells in stable and progressive myeloma precursor conditions harbored similar rates of rare SV. Using WGS data from previous publications (28, 29) we found naive and memory B-cells and stable myeloma precursor conditions (i.e., monoclonal gammopathy of undetermined significance and smoldering myeloma) to have negligible rates of rare SVs, progressive precursor conditions harbored statistically more rare SVs than naïve and memory B-cells, with myeloma cells harboring significantly more than all other B-cell entities (Fig. 2C). Overall, these findings provide support for the potential driver and pathogenic role of a subset of these rare SV in MM.

Considering their impact on the genome and clinical outcomes, chromothripsis and templated insertions were of particular interest (12, 28). A significantly higher proportion of rare SVs were found within chromothripsis events compared to both templated insertions involving two chromosomes and templated insertions involving more than 2 chromosomes, with median rare SV proportions of 64% (0–100% range), 0% (0–100% range), and 0% (0–100% range) of each complex SV event, respectively (p<0.001, using Wilcoxon test; Fig. 2D–E, Supplementary Fig. S2D–E). The differences observed between chromothripsis and templated insertions is expected, considering that the latter tends to be associated with focal amplification of recurrent driver genes and regulatory regions, while chromothripsis is often composed of multiple SVs affecting broad genomic segments often across multiple chromosomes. Chromoplexy and other complex events also displayed a large proportion of rare SVs within recurrent events, with median rare SV proportion of 55% (16–100% range) and 75% (12–100% range), each significantly higher than templated insertions (p<0.001; using Wilcoxon test; Supplementary Fig. S2D–E). Having established that rare SVs occur in most of the patients and across all SV classes, we first focused on rare SVs targeting myeloma genes of known impact. Given our strict criteria for calling SVs and manual validation of calls, we are confident in a low false-positive rate (Methods) (12).

Rare SV as a novel mechanism for dysregulation of known driver genes

MM has been reported to have oncogenic dependencies with recurrent mutations affecting 80 known driver genes (Supplementary Table 6) (10, 22). We therefore investigated the possible role of rare SVs both in recurrent and rare SV events as a novel mechanism for targeting and dysregulating known myeloma driver genes recurrently involved by single

nucleotide variation (SNV) and indels. To do so, we went to compare the expression impact of rare SVs, in comparison to patients with wild type and SNV/indel per gene of interest, within the 591 RNA-seq data at our disposal. Considering patients without recurrent SNVs in known driver genes, we found 45 rare SVs in 43 (6%) patients involving 21 of the 80 known MM drivers (Fig. 3A). Importantly, 14 of the driver genes impacted by rare SV had expression changes in line with their expected role in MM (i.e. duplication and deletions associated with expression within the 4th and 1st gene expression quartiles, respectively), with 18 patients displaying outlier expression in impacted genes (Fig. 3B; Supplementary Table 7). Among these events, we noted outlier gene expression mediated by rare focal tandem duplication or templated insertions on *KRAS* (n=1) and *PIM1* (n=1) (Fig. 3C–D), supporting the potential driver role of these SVs, and providing an alternate mechanism of dysregulating known driver genes. Importantly, in the absence of SNVs, these events would not be detected using whole exome sequencing and/or targeted panel approaches, highlighting the importance of WGS in order to comprehensively capture the genomic heterogeneity of each individual patient.

Chromothripsis with and without hypergains

Chromothripsis in MM can present with different genomic and copy number patterns. Overall, we identified two main profiles: 1) multiple deletions with copy number profile oscillating around either monoallelic loss or gain (67.5%, 75/111); or 2) focal segments amplified multiple times (32.5%, 36/111). The proportion of rare SV and fully rare events were not different between the 2 chromothripsis patterns (p>0.05, Wilcoxon test). Focusing on the latter, we investigated the frequency of focal (<3 Mb) copy number increases with 3 or more additional copies (i.e. total CN ≥ 5), which we define as hypergain events. With increasing extra CN, chromothripsis was increasingly responsible for most of the hypergains compared to other SV classes (226/268; 84%) (Fig. 4A). Focal hypergains associated with chromothripsis were studied in three possible SV patterns: recurrent SVs as part of recurrent chromothripsis events, rare SVs as part of a recurrent chromothripsis event, and rare SVs as part of fully rare chromothripsis events. The proportion of rare and recurrent SVs in recurrent chromothripsis events leading to hypergains was comparable (94/226; 41.6%) and was responsible for the majority of hypergains in rare SVs (94/97; 97%) (Fig. 4B). Testing the expression of genes involved by rare SVs, within recurrent chromothripsis events, showed clear overexpression of 180 genes, 6 of which do not have a defined driver role in MM including known cancer drivers *SOS1, NTRK1, and FCRL4* (z-score >2; Fig. 4C, Supplementary Fig. S3, Methods). Importantly, there is no difference in clinical outcomes between patients with hypergain and non-hypergain chromothripsis, with both groups displaying poorer outcomes compared to patients without chromothripsis (Fig. 4D). The poor overall outcomes in patients across both profiles of chromothripsis is likely due to the chromosomal instability that led to these chaotic chromosomal rearrangements targeting multiple oncodrivers at once (3).

Impact of fully rare events and implications on disease biology

To infer the biological impact of each rare SV and distinguish their role between a passenger or a potential driver, we investigated their impact on gene expression both in a direct manner, where an SV occurs within a gene body, and in an indirect manner

where they function through transcriptional deregulation via superenhancer hijacking. SVs were considered to involve a gene if either breakpoint intersected the gene region and/or occurred up to 1 Mb away, to take account of distal relationships within a topologically associated domain (TAD) (14, 24). Distinct duplications and deletions have been reported to recurrently dysregulate the interaction between gene-enhancers by remodeling TADs (4, 5, 30). To determine SV class-specific breakpoint enrichment in relation to distance from genes, a re-shuffle permutation was performed to create a random background model for each SV class and gene expression direction (i.e., up vs down regulation). Extending on methods developed to test the penetrance of rare germline events, genes paired with rare SV's were considered affected if the gene expression was outside a gene specific outlier z-score of +/− 2 (19). A more stringent criteria than quartile cutoffs, Z-scores allow for more confidence in determining the relationship between rare variants and their impact on non-driver gene expression. Across 591 patients, with available WGS and RNA-seq, we observed a significant association between rare SV events and outlier gene expression. Genes within recurrently affected areas (SV hotspots, GISTIC CNAs, and canonical Ig translocations), were excluded from this analysis, as by definition, a rare SV never involves a recurrently aberrant region.

Across all SV classes and SV to gene outlier distances, we saw a variable distribution of the number of gene outliers involved in each basic SV in fully rare SV events (Fig. 5A). Notably, rare templated insertion events affected the most gene outliers per event. Against a background model, rare templated insertions were enriched in overexpressing outliers within the body of genes and up to 1 Mb away (Fig. 5B). Rare duplication class SVs were enriched in overexpressing outliers within the gene body, in genes 100 kb of the breakpoint and up to 1 Mb away. Rare inversion class SVs were enriched in overexpressing outliers of genes 100kb and 1Mb away. Rare translocations were associated with overexpressing outliers 1 MB away. Rare complex SVs were enriched within the gene body of under expressing outliers, while deletion SVs were enriched in gene outliers with decreased expression within the gene body, 100 kb away, and up to 1 Mb away (Fig. 5B). A total of 201 (34%) patients had at least 1 rare SV event associated with gene expression outliers (Fig. 5C). Additionally, rare SVs in recurrent SV events were enriched within chromothripsis events associated with overexpressed outliers across all gene to SV distances and associated with under-expressing outliers in 100 Kb and 1 Mb relationships (Supplementary Fig. S4). Taking in aggregate all gene expression outliers with SV association up to 1 Mb away, gene ontology pathway analysis (Methods) showed involvement in key cell cycle pathways and proliferative Rac mediated signal transduction responsible for ERB2-mediated malignant transformation and metastasis, in addition to involving B-cell development genes such as *PRKDC*, and *POU2F2* (Fig. 5D, Supplementary Table 8) (31, 32). The enrichment of SV classes in their respective contexts to proximity to gene outliers against a random background model, suggests that a fraction of rare SV is actively involved in shaping the tumor genomic profile and therefore they are likely not passengers.

To further demonstrate the link between rare SV events and the associated outliers, we investigated all potential fusion genes caused by rare SV genes. A total 87 rare SV-associated fusions were observed across 69 patients, resulting in 18 overexpressing gene outliers (Supplementary Table 9).

Rare structural variants and superenhancers

Of the enriched SV to gene outlier relationships, translocation, duplication and templated insertion events were enriched in gene outliers up to 1 Mb distance were of particular interest for their possible involvement in expression regulation through distal regulatory elements. Translocation class SVs have been known to be involved in superenhancer hijacking, leading to up-regulation of distinct oncogenes in MM (33). This known mechanism is particularly enriched among templated insertions and recurrently affect important oncodrivers such as *CCND1*, *MYC, TENT5C*, and *MCL1* (34). While these associations have been extensively explored across recurrent SV, the interaction between rare SV and superenhancers has never been investigated. To do so we focused on rare SVs within fully rare SV events, modelling SV breakpoint density to the nearest known MM superenhancer up to 10 Mb (24). To determine if SV breakpoints occurred in a clustered or random manner with regard to superenhancer proximity, rare SV breakpoints were re-shuffled, adding a random length between 10–20 Mb to the original position, and proximity to nearest superenhancer was re-calculated to develop a random density distribution (Methods).

Overall, rare duplications, translocations and templated insertions were significantly enriched within and/or up to 1 Mb from superenhancer regions, when comparing all other SV classes. Compared to the random density distribution model, only templated insertions emerged as significantly enriched within 1 Mb of superenhancers ($p<0.001$, Fisher Exact, Fig. 6A). To further validate rare SV breakpoint involvement and enrichment with myeloma-specific superenhancers, we modeled the breakpoint density of myeloma rare SVs to the nearest known breast cancer-specific and diffuse large B-cell lymphoma-specific superenhancers, and found no significant enrichment (Supplementary Fig. S5A–B) (26, 27). Conversely, when rerunning the breakpoint density distribution to myeloma superenhancers not shared with breast cancer cells against the background model, the rare templated insertion enrichment was recapitulated (Supplementary Fig. S5C). These data further support the link between rare templated insertion and distinct MM super enhancer. Considering genes that were directly involved with enriched rare SVs involved SE regions directly and up to 1 Mb distance, and genes within 1 Mb of affected superenhancers, we found 303 gene expression outliers (z-score >2 or <−2). Overall, 26% (127/494) of rare duplications associated with superenhancers were also associated with gene expression outliers, 207 genes being over-expressed (e.g. *IDH2*) and 57 under-expressed, with a median of 3 genes affected per rare duplication event (1–9 gene count range). 36% (65/179) of rare translocation events were associated with outlier genes, with 123 outliers being overexpressed and 33 under-expressed, with a median of 2 genes affected per rare translocation event (1–12 gene count range). 27% (22/82) of rare templated insertion events associated with superenhancers were associated with outlier genes, with 162 being overexpressed (e.g. *IRF6*) and 13 under-expressed, with a median of 21 genes affected per templated insertion event (1–30 gene count range) (Fig. 6B). The association between rare single and complex SV with superenhancers and gene over expression support the idea that a fraction of these events are not passengers but rather play a potential driver role.

DISCUSSION

In recent years, DNA sequencing and expression studies have allowed the description of several new myeloma genomic and clinical subgroups. However, clinically, patients within groups harboring similar or even identical recurrent genomic features present with significant clinical heterogeneity. This clinical reality suggests that our understanding of the MM biology is incomplete, and this might be partially due to unknown genomic drivers contributing to significant disease heterogeneity in each individual patient. WGS, being the only method for genome wide SV characterization, has allowed us to expand our catalogue of recurrent genomic drivers, in particular, the characterization of CNA and SVs not involving Ig loci. Nevertheless, several somatic SV events remain unexplained and not associated with any recurrent gene or biological mechanisms known to play a driver role in MM pathogenesis. In this study, we have analyzed a large series of NDMM to decipher the role of rare SV events in NDMM in particular by their impact on gene expression. We focused on rare SVs specifically for two main reasons: 1) the frequency of rare SV events across the genome is low but they have the potential for significant biological impact compared to other genomic features such as SNV and indels, where the great majority are neutral and phenotypically inconsequential, and 2) unlike SNV on driver genes, somatic SV and CNA have rarely been found in normal cells across different tissues, suggesting that SV and CNA are essential to gene deregulation, selection and tumor progression (19, 29, 35, 36). These two findings suggest that the fraction of passenger SV is likely to be lower than what is seen with SNV and indels. Based on this hypothesis, we applied an analytical workflow that is similar to what has been previously performed to define the impact of rare genetic events (19). Three rare SV categories were identified. The first group was characterized by fully rare SV events involving genes without a clear and known driver role in MM. Across the entire cohort, 264 patients (35%) had at least one fully rare SV linked to gene expression outliers of z-scores outside +/− 2. In addition, we found a significant enrichment of duplications, translocations and templated insertions at known MM superenhancer regions supporting the link between rare SV and gene expression outliers. The second group is composed of genes without any known driver role in either MM or other cancers that were affected by multiple focal amplifications. The majority of these events were caused by chromothripsis (226/268; 84%) with 87% (180/205) of genes overexpressed as a consequence of multiple gains associated with rare SVs in recurrent chromothripsis events. This finding demonstrates that rare genes can be involved by distinct and complex events actively impacting individual tumor gene expression. The third group comprised of rare SVs involving known MM driver genes that have been previously shown to be involved by mutations, but not CNA or SV. Interestingly, 43 patients had at least one of these events, and in 42% (18/43) of patients these events impacted gene expression resulting in expression within the 1st or 4th quartile. This is an important finding, as some of these genes are potentially targetable and therefore expand the number of patients who might benefit from future targeted therapies. Taken together these data suggest that genomic and transcriptomic profiles in a significant fraction of patients are affected by rare SV events. The pathogenetic role of rare SVs in myelomagenesis is further supported by their low prevalence or absence in normal B-cells and stable myeloma precursor conditions, in contrast to the observed frequency in MM. This suggests that rare SVs may play a crucial

role in the progression from myeloma precursor conditions to MM. However, larger studies will be necessary to fully assess the impact of rare SVs on the progression of myeloma precursor conditions into MM. Importantly, WGS is the only method that can provide whole genome SV resolution and a comprehensive picture of therapeutically targetable alterations in patients.

While larger cohorts of WGS are needed to identify new hotspot and driver genes significantly enriched for SV, the picture that emerges from these data is that the heterogeneity between individual patients is also potentially driven by rare SV events, not easily identifiable using standardized statistical approaches based on prevalence and frequency. Furthermore, some of the observed rare SV events may in fact be deemed recurrent in larger datasets with better statistical power. The definition and characterization of these events is not only important for deciphering the individual patient heterogeneity, but might have significant implications for the development of individualized treatment strategies and in our understanding of mechanisms of resistance to novel targeted therapies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

1. Consortium I.T.P.-C.A.o.W.G., Pan-cancer analysis of whole genomes. Nature, 2020. 578(7793): p. 82–93. [PubMed: 32025007]

2. Dentro SC, et al. , Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes. Cell, 2021. 184(8): p. 2239–2254 e39. [PubMed: 33831375]

3. Li Y, et al. , Patterns of somatic structural variation in human cancer genomes. Nature, 2020. 578(7793): p. 112–121. [PubMed: 32025012]

4. Shanta O, et al. , The effects of common structural variants on 3D chromatin structure. BMC Genomics, 2020. 21(1): p. 95. [PubMed: 32000688]

5. Akdemir KC, et al. , Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat Genet, 2020. 52(3): p. 294–305. [PubMed: 32024999]

6. Manier S, et al. , Genomic complexity of multiple myeloma and its clinical implications. Nat Rev Clin Oncol, 2017. 14(2): p. 100–113. [PubMed: 27531699]

7. Cortes-Ciriano I, et al. , Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. Nat Genet, 2020. 52(3): p. 331–341. [PubMed: 32025003]

8. Rajkumar SV, Multiple myeloma: 2020 update on diagnosis, risk-stratification and management. Am J Hematol, 2020. 95(5): p. 548–567. [PubMed: 32212178]

9. Misund K, et al. , Clonal evolution after treatment pressure in multiple myeloma: heterogenous genomic aberrations and transcriptomic convergence. Leukemia, 2022. 36(7): p. 1887–1897. [PubMed: 35643867]

10. Maura F, et al. , Genomic landscape and chronological reconstruction of driver events in multiple myeloma. Nat Commun, 2019. 10(1): p. 3835. [PubMed: 31444325]

11. Ansari-Pour N, et al. , Whole-genome analysis identifies novel drivers and high-risk double-hit events in relapsed/refractory myeloma. Blood, 2022: p. blood.2022017010.

12. Rustad EH, et al. , Revealing the impact of structural variants in multiple myeloma. Blood Cancer Discov, 2020. 1(3): p. 258–273.

13. Walker BA, et al. , MYC Translocations In Multiple Myeloma Involve Recruitment Of Enhancer Elements Resulting In Over-Expression and Decreased Overall Survival. Blood, 2013. 122(21): p. 274–274.

14. Barwick BG, et al. , Chromatin Accessibility Identifies Regulatory Elements Predictive of Gene Expression and Disease Outcome in Multiple Myeloma. Clin Cancer Res, 2021. 27(11): p. 3178–3189. [PubMed: 33731366]

15. Maura F, et al. , Chromothripsis as a pathogenic driver of multiple myeloma. Semin Cell Dev Biol, 2022. 123: p. 115–123. [PubMed: 33958284]

16. Gooding S, et al. , Multiple cereblon genetic changes are associated with acquired resistance to lenalidomide or pomalidomide in multiple myeloma. Blood, 2021. 137(2): p. 232–237. [PubMed: 33443552]

17. Samur MK, et al. , Biallelic loss of BCMA as a resistance mechanism to CAR T cell therapy in a patient with multiple myeloma. Nat Commun, 2021. 12(1): p. 868. [PubMed: 33558511]

18. Lee H, et al. , Mechanisms of antigen escape from BCMA- or GPRC5D-targeted immunotherapies in multiple myeloma. Nat Med, 2023. 29(9): p. 2295–2306. [PubMed: 37653344]

19. Chiang C, et al. , The impact of structural variation on human gene expression. Nat Genet, 2017. 49(5): p. 692–699. [PubMed: 28369037]

20. Chakravarty D, et al. , OncoKB: A Precision Oncology Knowledge Base. JCO Precision Oncology, 2017(1): p. 1–16.

21. Tate JG, et al. , COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res, 2019. 47(D1): p. D941–D947. [PubMed: 30371878]

22. Walker BA, et al. , Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. Blood, 2018. 132(6): p. 587–597. [PubMed: 29884741]

23. Abdallah N, et al. , Implications of MYC Rearrangements in Newly Diagnosed Multiple Myeloma. Clin Cancer Res, 2020. 26(24): p. 6581–6588. [PubMed: 33008815]

24. Jia Y, et al. , Myeloma-specific superenhancers affect genes of biological and clinical relevance in myeloma. Blood Cancer J, 2021. 11(2): p. 32. [PubMed: 33579893]

25. Jin Y, et al. , Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. Blood, 2018. 131(19): p. 2138–2150. [PubMed: 29519805]

26. Bal E, et al. , Super-enhancer hypermutation alters oncogene expression in B cell lymphoma. Nature, 2022. 607(7920): p. 808–815. [PubMed: 35794478]

27. Chapuy B, et al. , Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. Cancer Cell, 2013. 24(6): p. 777–90. [PubMed: 24332044]

28. Oben B, et al. , Whole-genome sequencing reveals progressive versus stable myeloma precursor conditions as two distinct entities. Nat Commun, 2021. 12(1): p. 1861. [PubMed: 33767199]

29. Machado HE, et al. , Diverse mutational landscapes in human lymphocytes. Nature, 2022. 608(7924): p. 724–732. [PubMed: 35948631]

30. Loven J, et al. , Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell, 2013. 153(2): p. 320–34. [PubMed: 23582323]

31. Arias-Romero LE, et al. , A Rac-Pak signaling pathway is essential for ErbB2-mediated transformation of human breast epithelial cancer cells. Oncogene, 2010. 29(43): p. 5839–49. [PubMed: 20711231]

32. Cooke M, Baker MJ, and Kazanietz MG, Rac-GEF/Rac Signaling and Metastatic Dissemination in Lung Cancer. Front Cell Dev Biol, 2020. 8: p. 118. [PubMed: 32158759]

33. Affer M, et al. , Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. Leukemia, 2014. 28(8): p. 1725–1735. [PubMed: 24518206]

34. Mikulasova A, et al. , Microhomology-mediated end joining drives complex rearrangements and overexpression of MYC and PVT1 in multiple myeloma. Haematologica, 2020. 105(4): p. 1055–1066. [PubMed: 31221783]

35. Abascal F, et al. , Somatic mutation landscapes at single-molecule resolution. Nature, 2021. 593(7859): p. 405–410. [PubMed: 33911282]

36. Moore L, et al. , The mutational landscape of human somatic and germline cells. Nature, 2021. 597(7876): p. 381–386. [PubMed: 34433962]

TRANSLATIONAL RELEVANCE

Characterization of multiple myeloma genome revealed that more than half structural variants are not involved in recurrent events. Here, we demonstrate that these rare SVs hold potential for myeloma pathogenesis through their gene expression impact. Rare SVs contribute to MM heterogeneity and have implications for development of individualized treatment.
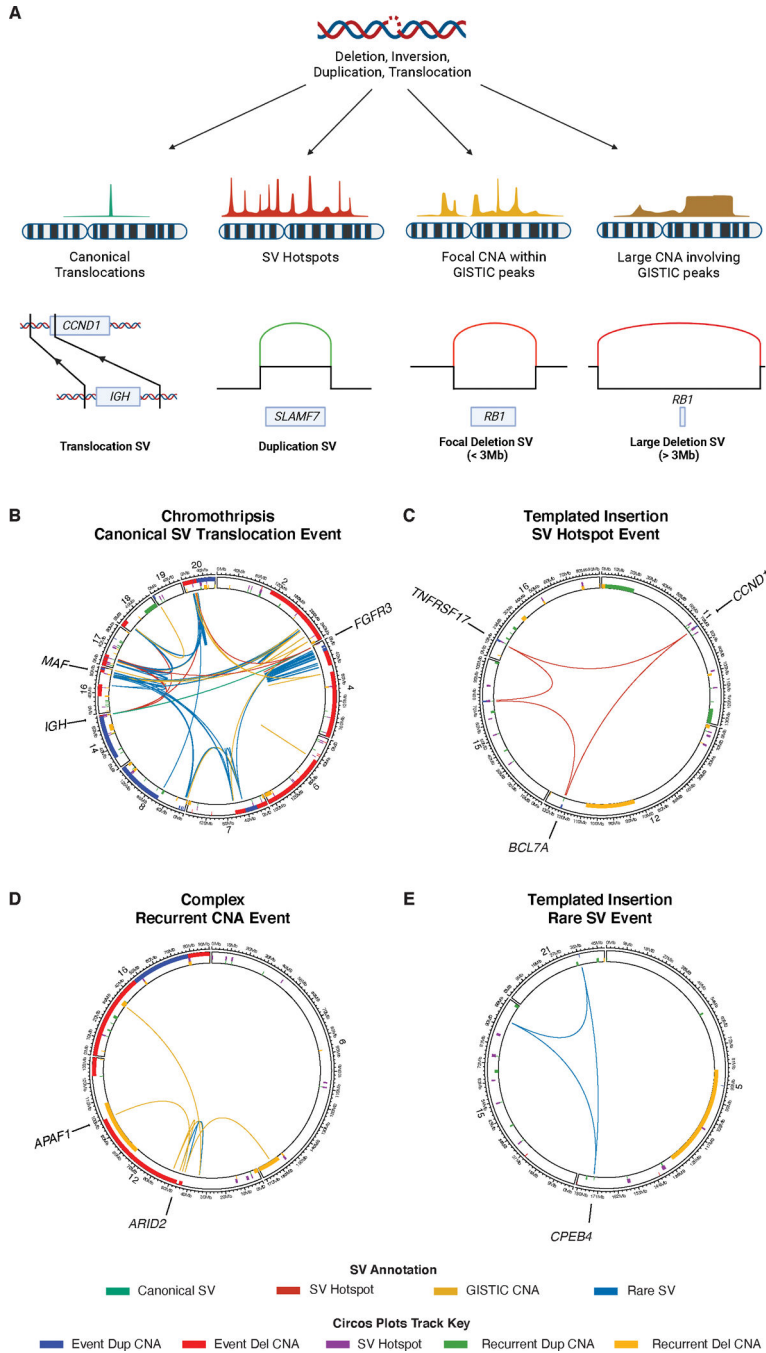
Figure 1: Defining recurrent and rare SV events.

A) Any SV events which causes a canonical myeloma translocation, or SV breakpoint that falls within the previously defined recurrent regions was identified as a recurrent SV. Examples depict SV breakpoints falling within SV hotspots and causing focal and large copy umber gains or losses occurring within previously defined recurrent GISTIC CNA regions of the myeloma genome. **B)** Chromothripsis class SV annotated as "Canonical Translocation" event based on the highest priority SV involving a translocation between *IGH* and *FGFR3*/*NSD2* (green) comprising of 123 basic SVs, of which 1 is a canonical

translocation between *IGH* and *FGFR3/NSD2*, 17 SVs within hotspot regions, 12 SVs within recurrent GISTIC CNA regions, and 91 are rare SVs. **C)** Templated insertion class SV annotated as an "SV Hotspot" event, based on the highest priority SV (red) involving a hotspot region (purple). **D)** Complex SV annotated as a "Recurrent CNA" event based on the highest priority SV involving a recurrent GISTIC deletion CNA region (yellow). **E)** Templated insertion class SV annotated as a "Rare SV" event, based on the exclusion of any recurrent regions affected by the event's SVs (blue).
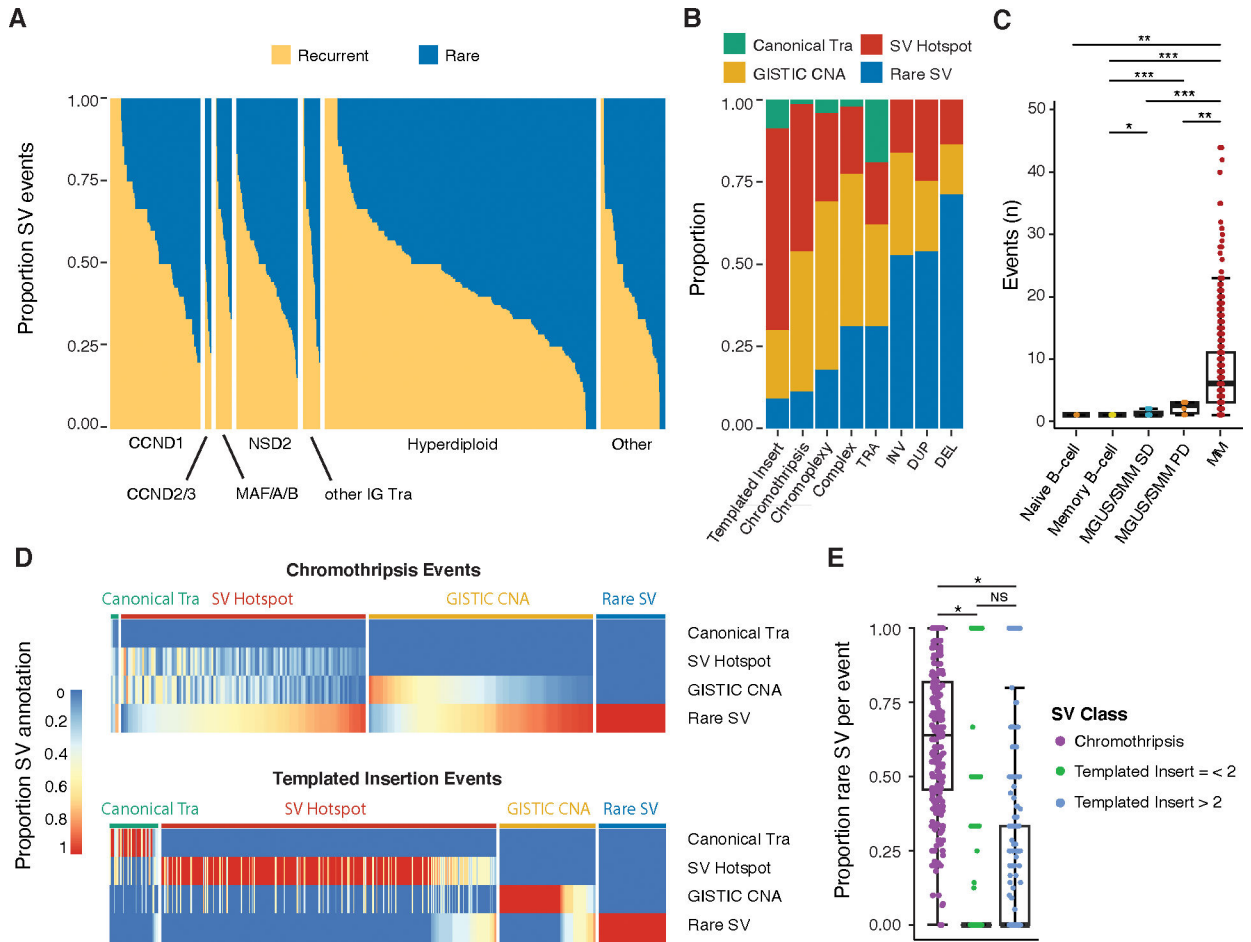
Figure 2: Rare SV landscape of newly diagnosed multiple myeloma patients.
A) Proportion of rare and recurrent SV events per patient, subdivided into classical clinical cytogenetic groups. *CCND1, CCND2/3, MAF/A/B*, and *NSD2* denote canonical translocations between immunoglobulin (Ig) loci. Group *MAF/A/B* includes genes *MAF, MAFA*, and *MAFB*. "other IG Tra" group represents the cytogenetic group with Ig translocations between *Ig* and non-canonical partners. "Other" cytogenetic group denotes patients without canonical translocations, translocations involving the Ig locus, or a hyperdiploid genome. **B)** Proportion of SV events in each classification category based on annotation hierarchy as described, grouped by SV class. **C)** Frequency of rare SV events in naïve and memory B-cells, stable (SD) and progressive (PD) myeloma precursor conditions (i.e., monoclonal gammopathy of undetermined significance and smoldering myeloma), and MM. **D)** Breakdown of SV annotation within each SV event, in chromothripsis and templated insertion complex SV events. **E)** Statistical comparison of proportion of rare SVs per event in all chromothripsis events compared to templated insertion events involving up to 2 chromosomes and, templated insertion involving more than 2 chromosomes (p< 0.001, using Wilcoxon test). * = p.value < 0.05, ** = p.value < 0.001, *** =p.value < 0.0001.
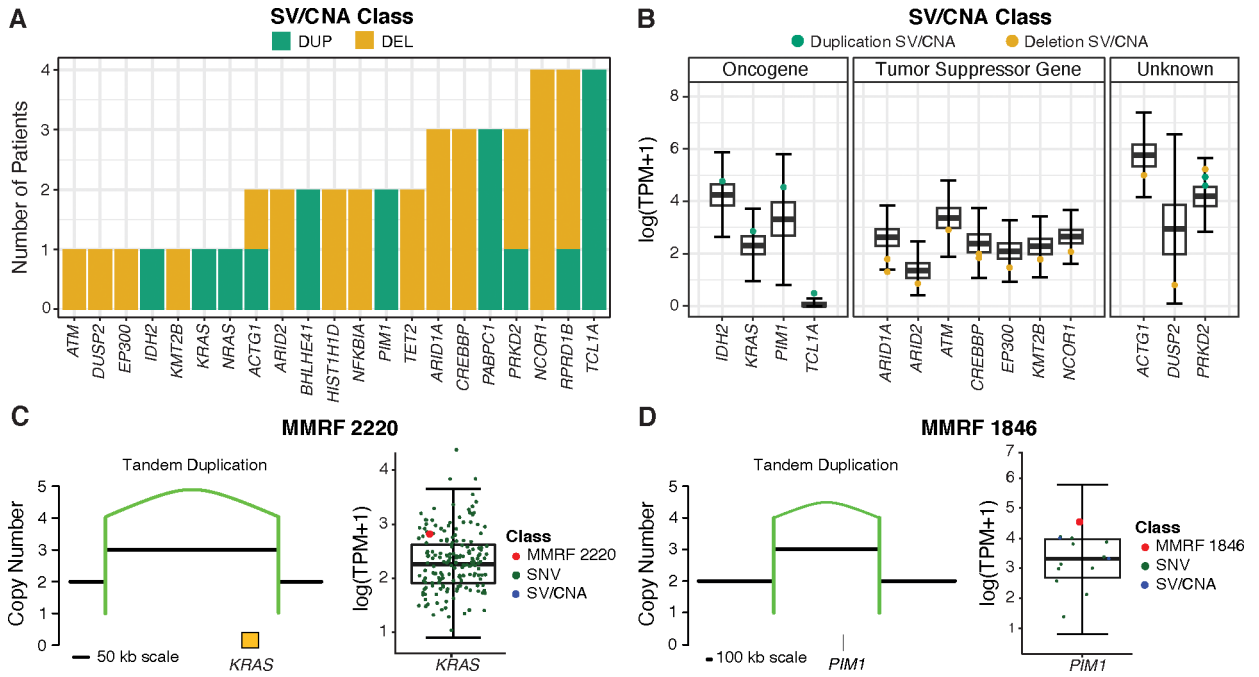
Figure 3: SVs as alternate mechanisms for targeting known multiple myeloma driver genes.
A) Panel of driver genes classically dysregulated by SNV point mutations focally duplicated or deleted by SVs and CNAs in patient cohort, showing oncogenes affected by duplications, tumor suppressor genes by deletions, and both duplications and deletions impacting genes with undetermined impact. **B)** Expression of genes impacted by rare SVs, denoted by colored points and CNA impact. Only SV events associated with gene expression in either the 1st or 4th quartile were reported. **C-D)** Examples of rare SV targeting known multiple myeloma driver genes. **C)** Tandem duplication SV increasing copy number of *KRAS*, with associated increase in gene expression measured in the 4th quartile (z-score >2). **D)** Tandem duplication SV increasing *PIM1* copy number, with associated increase in gene expression in the 4th quartile (z-score >2). In **C)** and **D)** the minor allele was retained. All expression analyses were conducted in context of 591 RNA-seq samples.
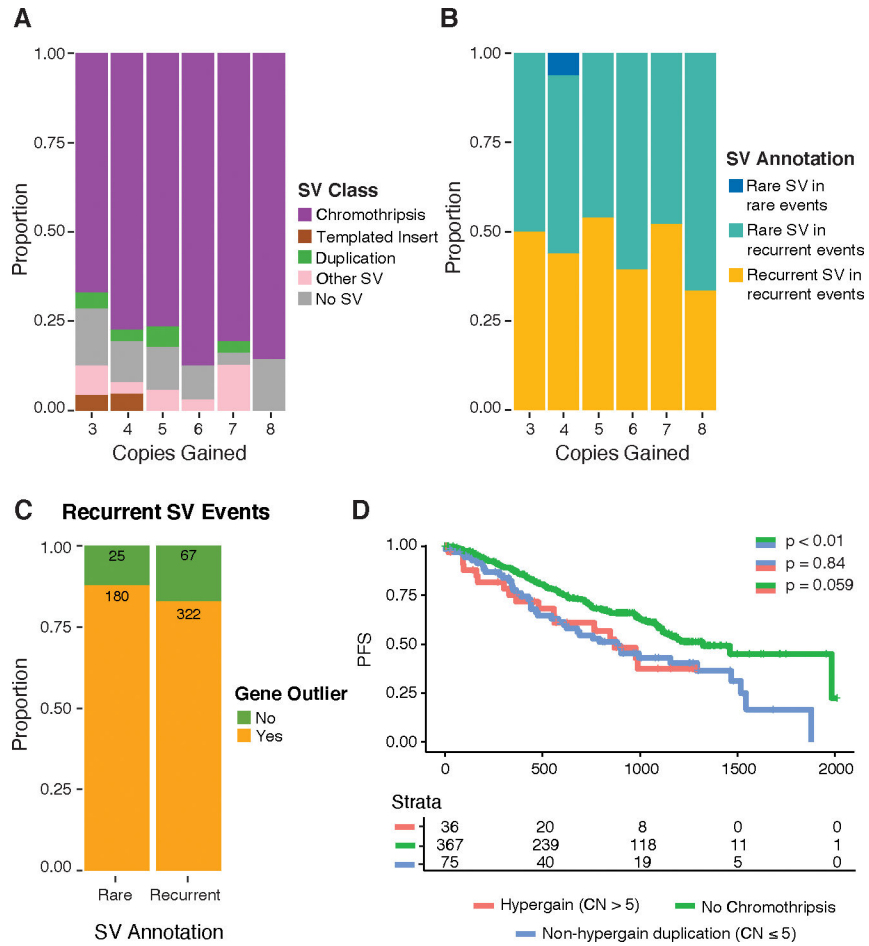
Figure 4: Chromothripsis can cause multiple focal copy number gains associated with increased gene expression.

A) Chromothripsis dominates as the primary SV class to cause focal copy number increases beyond 3 additional copies – here defined as hypergains. **B)** Rare SVs within recurrent chromothripsis events cause the majority of hypergains. **C)** Of all genes within hyepergain regions affected by rare chromothripsis SVs, the majority exhibit gene expression within the 4th quartile and beyond. **D)** Patients with hypergain chromothripsis display a shorter progression free survival (PFS) compared to patients without chromothripsis. PFS is not different between patients with non-hypergain duplication (CN   5) and hypergain chromothripsis events. All expression analyses were conducted in context of 591 RNA-seq samples.
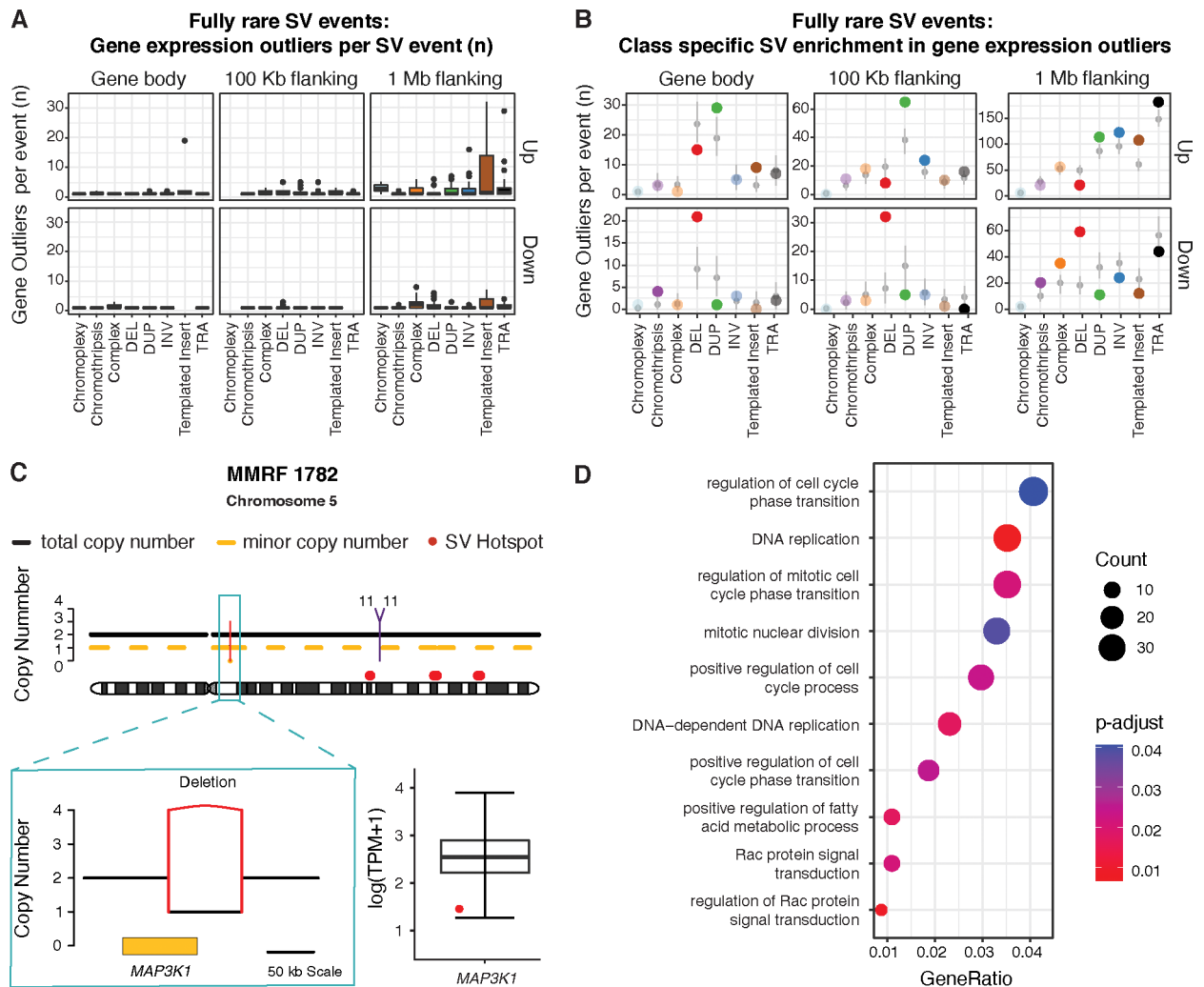
Figure 5: Impact of fully rare SV events on gene expression.
A) Number of gene outliers (z-score >2 or < −2) in rare SV events subdivided into SV classes and proximity from SV to gene outlier. **B)** Class specific rare SV enrichment in relation to gene proximity, subdivided into over-expressing outliers (z-score >2) and under-expressing outlier (z-score < −2). Solid points represent significant enrichment against background premutation rates. Translucent points represent non-significant relationship against background premutation rates. **C)** Example of a rare SV event involving a tumor suppressor gene. Single focal rare deletion SV bisecting *MAP3K1* gene with associated low expression in the 1st quartile (z-score < −2), in patient MMRF 1782. **D)** Gene ontology pathway analysis of all gene outliers with relationship to fully rare SV events across all classes and SV to gene outlier proximity. All expression analyses were conducted in context of 591 RNA-seq samples.
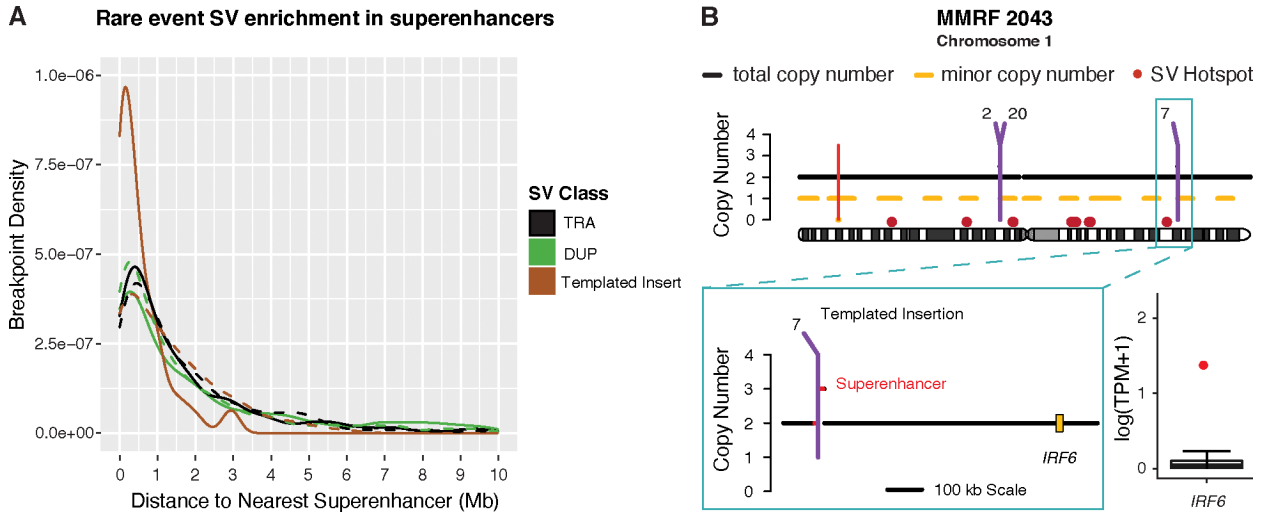
Figure 6: Rare SVs involving myeloma superenhancers.
A) Fully rare templated insertion class SV events were enriched within 1 Mb of known myeloma superenhancer regions (SE). Solid line depicts observed SV breakpoint density. Dashed line represents breakpoint density expected by permutations background rates. **B)** Example of rare templated insertion SV bisecting and duplicated an SE region. Within 1 Mb downstream of the affected superenhancer, *IRF6* displayed high outlier expression in the 4th quartile (z-score >2). All expression analyses were conducted in context of 591 RNA-seq samples.