# Revealing the impact of structural variants in multiple myeloma

Rustad et al.

**Supplementary Figures**

**Figure S1: Association between SV classes and genomic features.** Correlation matrix showing the breakpoint density of each SV class in the CoMMpass dataset (y-axis) and genomic features across the genome (x-axis), divided in 500 kb bins. Color and size of points are determined by the magnitude of positive (blue) and negative (red) spearman correlation coefficients.

Macintosh HD:Users:rustade:Documents:MSKCC:Projects:SV_project_clean:figures:Figure_S1:svCorrMatrix_trimmed.pdf

**Figure S2: Clinical and molecular features associated with adverse prognosis in the CoMMpass dataset.** Hazard ratios for PFS and OS estimated using multivariate Cox regression. Lines indicates 95% CI from multivariate cox regression models, statistically significant features indicated by asterisks (\* p<0.05; \*\* p < 0.01).

disk:Users:rustade:Dropbox:Forskning:Projects:SV_project_clean:figures:Figure_S2:cox_forest.pdf

**Figure S3: Breakpoint distribution within the *IGH* locus.** Density (y-axis) of IGH-locus breakpoints (position along the x-axis) for *MAP3K14* and canonical partners. The distribution of *MAP3K14* breakpoints is similar to that of other drivers, including the class-switch recombination (CSR) regions (i.e. IgA, IgG, IgM, IgD and IgE), particularly the IgM-switch region.

Macintosh HD:Users:rustade:Documents:MSKCC:Projects:SV_project_clean:figures:Figure_S3:igh_breakpoints.pdf

**Figure S4: Impact of copy number and SV involvement on gene expression after excluding genes after excluding genes with 4 or more copies**. Impact of copy number and SV involvement on normalized gene expression values (Z-scores) after excluding genes with 4 or more copies, estimated by multivariate linear regression. Estimates with 95% CI for each parameter are shown. Pooled analysis was performed for all expressed genes on autosomes across all patients (total copy number < 4), excluding structural events involving immunoglobulin loci. This analysis shows the same direction of gene expression effects as in the full analysis presented in **Figure 3D**, although the effect size is smaller for all SV classes.

disk:Users:rustade:Dropbox:Forskning:Projects:SV_project_clean:figures:Figure_3:Multivariate_expr_below4.pdf

**Figure S5: Recurrently translocated regions not identified as hotspots by the PCF algorithm**. Two recurrently translocated regions reported by Barwick et al(1) were not recapitulated in our hotspot analysis using the PCF algorithm: chromosome 20q12 (*MAFB;* upper panel) and chromosome 19p13.3 (bottom panel). Each panel displays, from top to bottom: SV density (black), enhancer density (brown), GISTIC peaks and cumulative copy number (gain = blue, red = loss). Vertical dashed gray lines mark the loci of putative driver genes in the SV density plots. In cumulative CNA plots, the number of patients with CNA is shown on the y-axis, with chromosomal coordinates (Mb) on the x-axis. *MAFB* is a known driver in multiple myeloma and clearly a false negative by the PCF algorithm due to the lack of a clear breakpoint cluster compared to the surrounding region. On chromosome 19p13.3, there appears to be a wide region enriched for SV breakpoints and copy number gains near the telomere, but again lacking a peak with high density of SV breakpoints. The potential driver role of SVs in this region remains unclear.

**Macintosh HD:Users:rustade:Documents:MSKCC:Projects:SV_project_clean:figures:Figure_S3:200524_Figure_S3.pdf**

**Figure S6: Hotspots with known and novel driver genes**. Three putative driver SV hotspots are shown as examples in the main manuscript (**Figure 4B-C**); the remaining are displayed here. Novel putative drivers have bold labels; the remaining hotspots contain known multiple myeloma driver genes. Each panel displays, from top to bottom: SV density (black), enhancer density (brown), GISTIC peaks and cumulative copy number (gain = blue, red = loss). Vertical dashed gray lines mark the loci of putative driver genes in the SV density plots. In cumulative CNA plots, the number of patients with CNA is shown on the y-axis, with chromosomal coordinates (Mb) on the x-axis.

Macintosh HD:Users:rustade:Documents:MSKCC:Projects:SV_project_clean:figures:Figure_S4:200524_figure_S4.pdf

**Figure S7: SV burden and the number of SV hotspots involved in each patient**. Each dot is a patient, with log10 SV burden along the x-axis and number of hotspots involved along the y-axis. **A)** Overall SV burden and SV hotspots in multiple myeloma. **B)** Rearrangement signature 1 (RS1) SV burden and RS1 hotspots in breast cancer. **C)** Rearrangement signature 3 (RS3) SV burden and RS3 hotspots in breast cancer. Breast cancer data shown in **B-C** were obtained from Glodzik et al(2).

Macintosh HD:Users:rustade:Documents:MSKCC:Projects:SV_project_clean:figures:S_Methods:200523_sv_hot_cor.pdf

**Supplementary Tables**

**Table S1: Cohort summary.** Baseline clinical variables andavailability of sequencing data from patients in the CoMMpass cohort (IA 13).

**Table S2: Recurrent CNAs in multiple myeloma**. Chromosomal coordinates used to define recurrent CNAs in multiple myeloma.

**Table S3: SV hotspots in multiple myeloma.** Summary table of all SV hotspots. Detailed explanations for each column are provided as a second sheet in the table excel file.

**Table S4: GISTIC amplification peaks.** Wide peaks of chromosomal gains as defined by the GISTIC algorithm (FDR < 0.1).

**Table S5: GISTIC deletion peaks.** Wide peaks of chromosomal loss as defined by the GISTIC algorithm (FDR < 0.1).

**Table S6: Expression analysis of genes in SV hotspots.** Linear regression analysis was run to assess the impact of putative driver SV involvement and copy number changes for each gene within 500 Kb of an SV hotspot. Results are shown here for SV involvement as well as CNAs present in more than five patients.

**Table S7: Summary of SV data for each patient.** For each patient (rows) we report the SV burden, presence or absence of chromothripsis, canonical primary events (i.e. *IGH* translocations and hyperdiploidy), any non-canonical putative driver events involving immunoglobulin loci, translocations involving *MAFB* (since this hotspot was missed) or *MYC*, and putative driver SVs in all 68 SV hotspots.

**Table S8: Hotspots of templated insertions.** Hotspot regions for templated insertions identified by the PCF algorithm when run exclusively with templated insertion breakpoints. Overlapping hotspots from the main hotspot analysis are annotated along with the putative driver genes involved.

**Table S9: Hotspots of chromothripsis.** Hotspot regions for chromothripsis identified by the PCF algorithm when run exclusively with chromothripsis breakpoints.

**Supplementary references**

1. Barwick BG, Neri P, Bahlis NJ, Nooka AK, Dhodapkar MV, Jaye DL*, et al.* Multiple myeloma immunoglobulin lambda translocations portend poor prognosis. Nature communications **2019**;10(1):1911 doi 10.1038/s41467-019-09555-6.

2. Glodzik D, Morganella S, Davies H, Simpson PT, Li Y, Zou X*, et al.* A somatic-mutational process recurrently duplicates germline susceptibility loci and tissue-specific super-enhancers in breast cancers. Nat Genet **2017**;49(3):341-8 doi 10.1038/ng.3771.