nature genetics

Supplementary information

https://doi.org/10.1038/s41588-024-01853-0

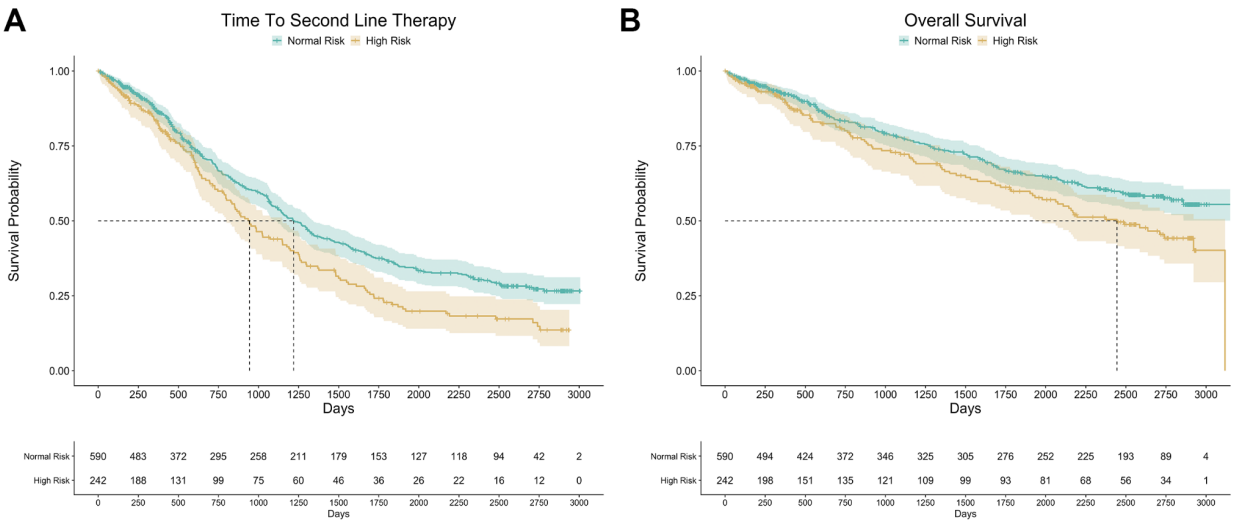
Comprehensive molecular profiling of multiple myeloma identifies refined copy number and expression subtypes

In the format provided by the authors and unedited

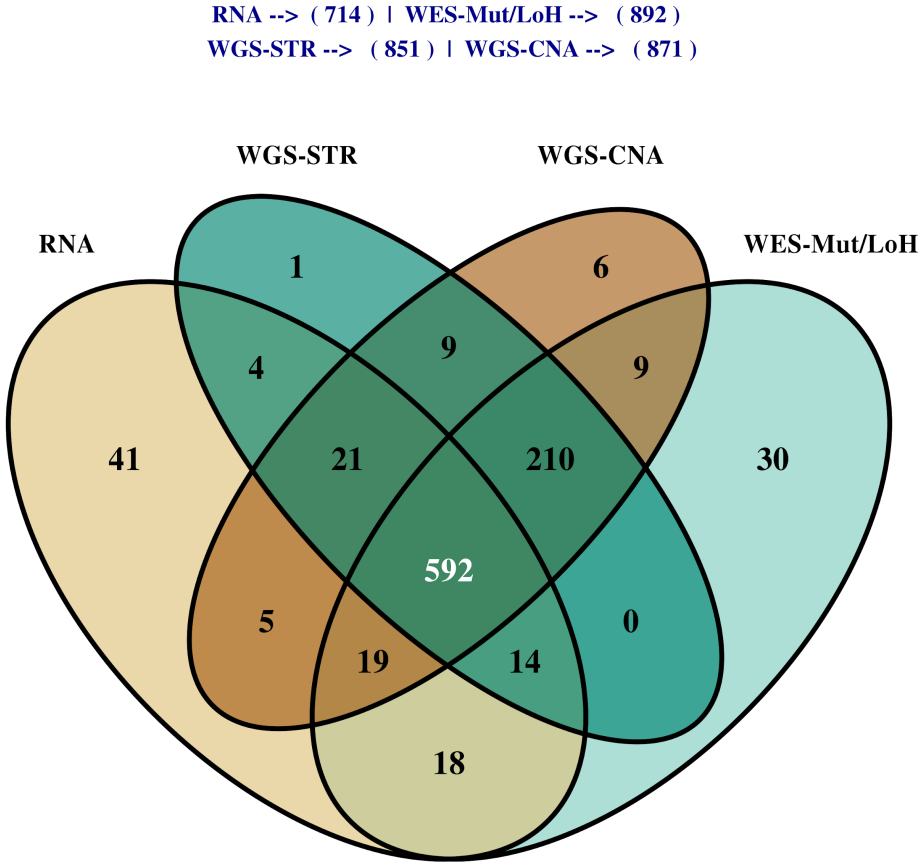
Supplementary Information

Table of Contents

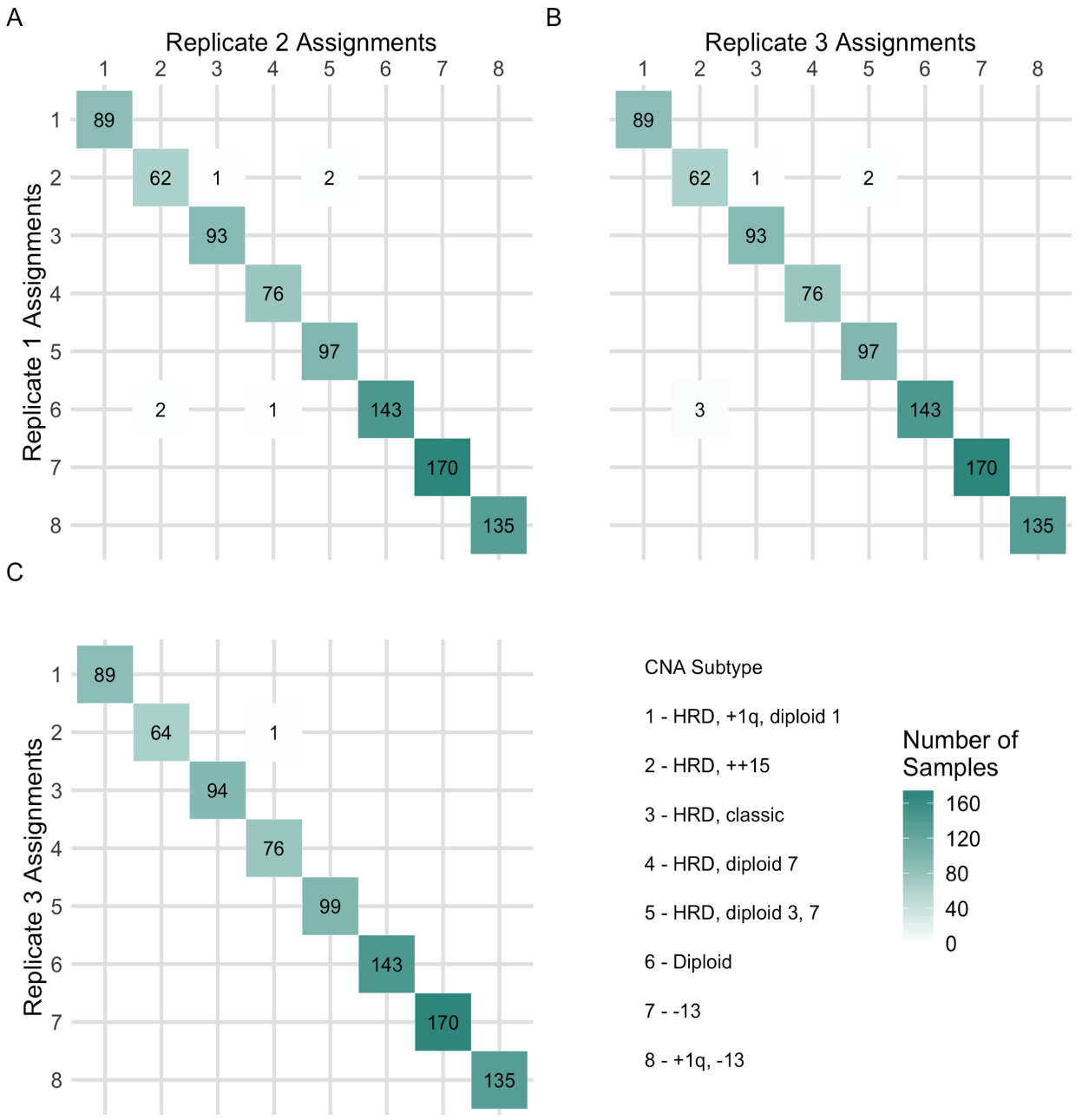
Supplementary Figures



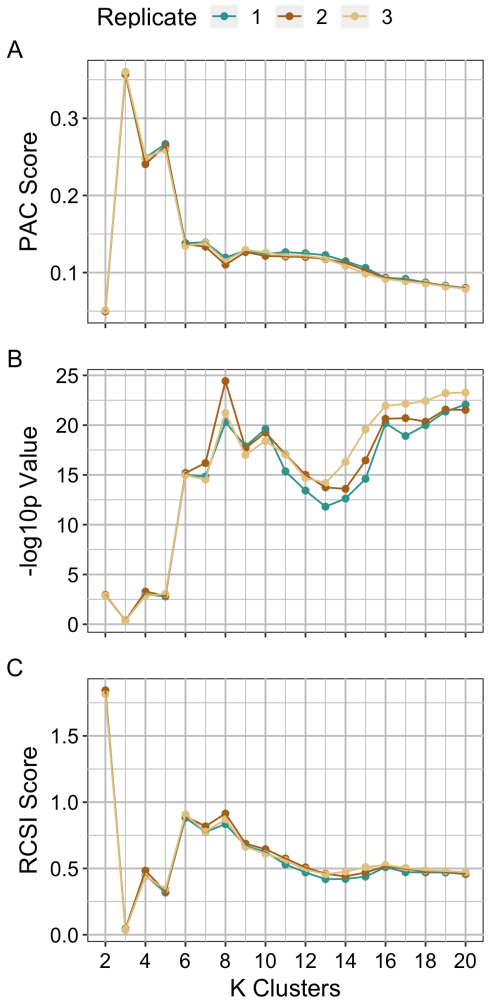
Supplementary Figure 1. Survival Outcomes of Patients with High-Risk Molecular Features. Pointwise survival estimates are shown by the respective dark lines while the 95% confidence interval is shown by the matching shaded bands and pairwise outcomes were compared by the log-rank test. Patients were assigned to the high risk group if they any of the following were detected: del17p13, t(14;16)-MAF, t(14;20)-MAFB, t(8;14)-MAFA, or t(4;14)-WHSC1/MMSET/NSD2 A) There is a significant difference in time to second line therapy (P = 0.00063) between normal-risk and high-risk patients were the medians were 40.6 months (95% CI = 36.9-44.6) and 31.4 months (95% CI = 26.5-38.3), respectively B) While the median OS for normal-risk patients was not met in the IA22 release there is a significant difference in outcome (P = 0.011) compared to high-risk patients were the median OS was 81.4 months (95% CI = 64.4 - NA).



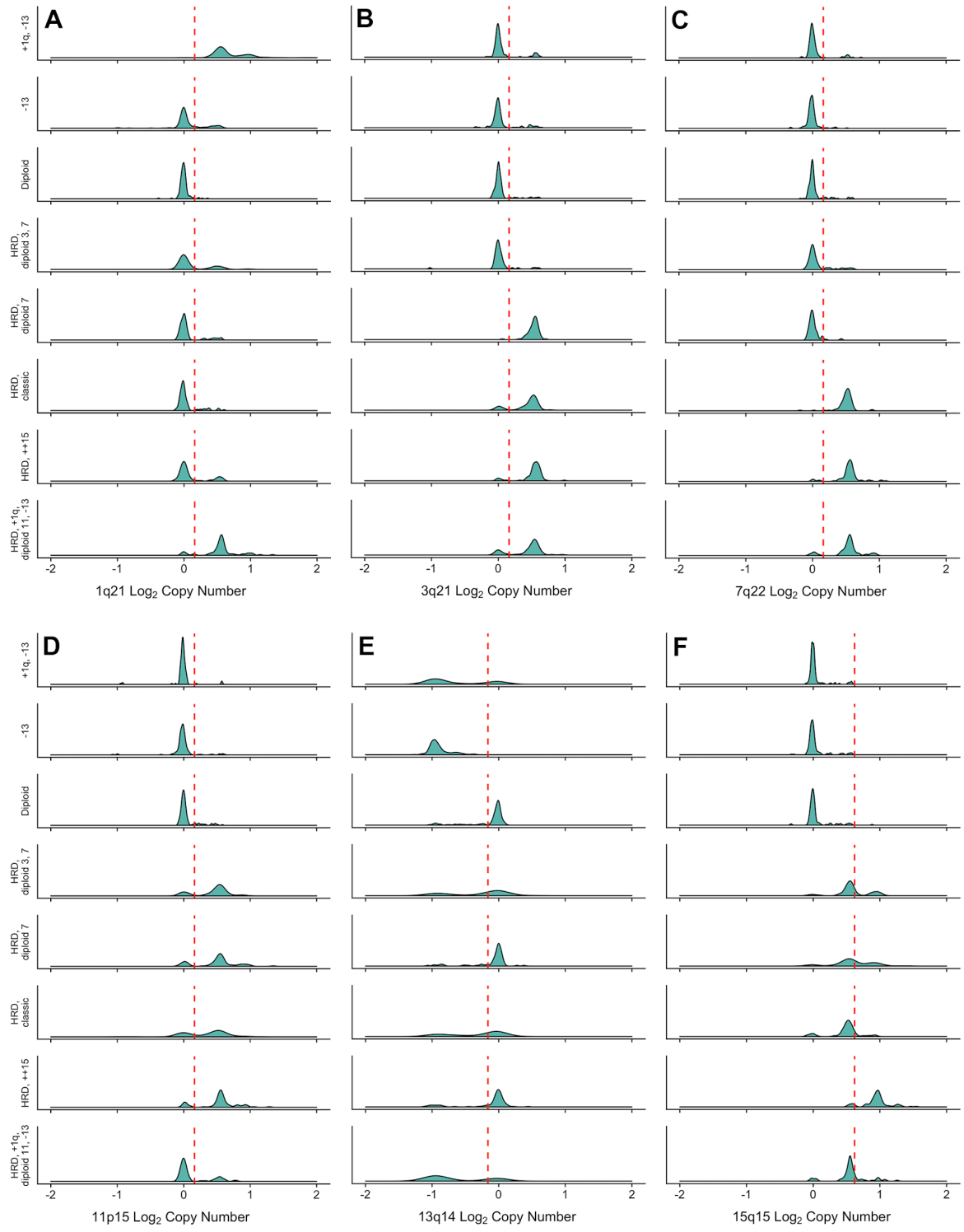
Supplementary Figure 2. Baseline Molecular Assay Intersection. Venn diagram showing the number of patients in the baseline CoMMpass cohort with available RNAseq, WGS structural (WGS-STR), WGS copy number (WGS-CNA), and WES mutation and loss-of-heterozygosity (WES-Mut/LoH) data for bone marrow (BM) derived tumor samples. There were 592 BM-derived tumor samples that were fully characterized with all data types available for analysis at diagnosis



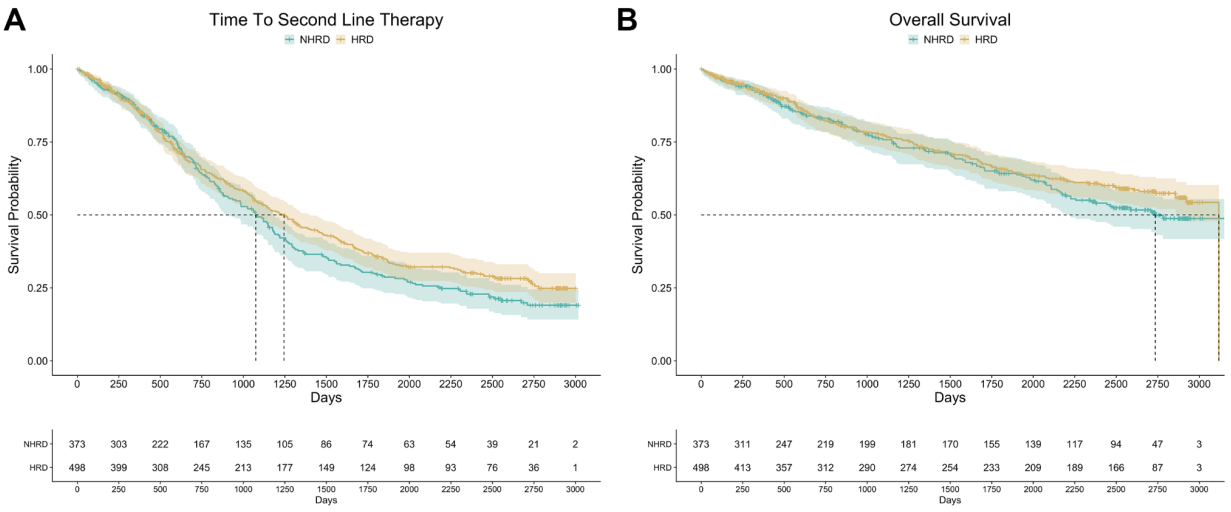
Supplementary Figure 3. Consistency of Subtype Assignments by CN Clustering. Confusion matrices showing the number of patients classified in each CN subtype across three replicates. CN subtype classifications were highly consistent across replicates with only 6/871 (0.7%) patients having different CN subtype classifications across replicates (A) 1 and 2 and (B) 1 and 3, and (C) only 1/871 (0.1%) patients having a different CN subtype classification across replicate 2 and 3.



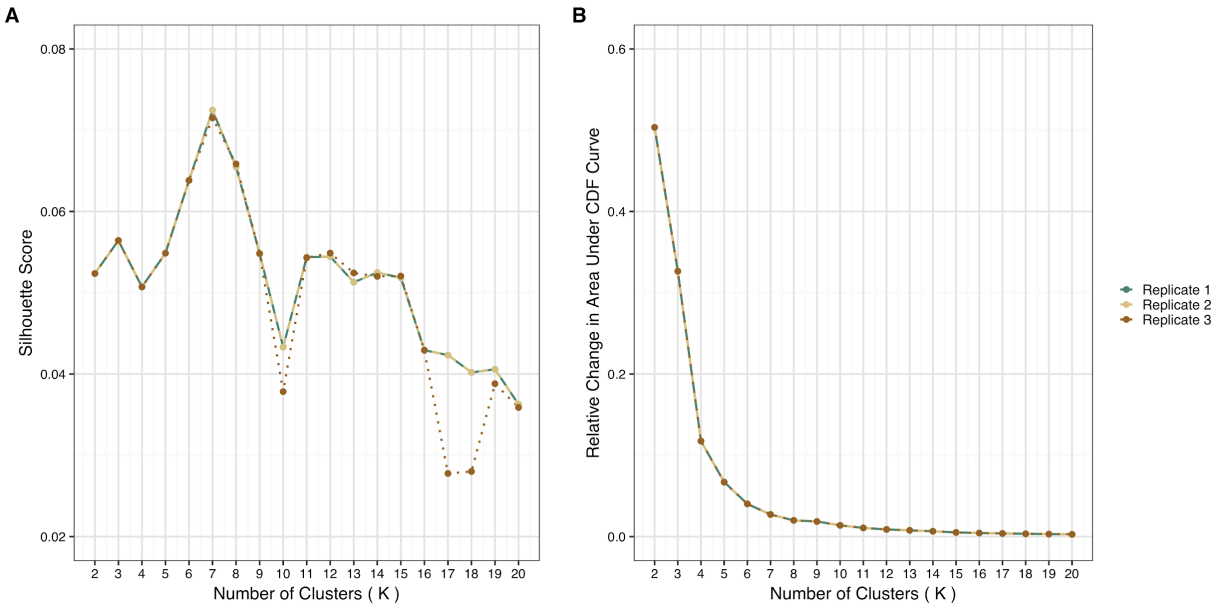
Supplementary Figure 4. Cluster Number Determination from CN Consensus Clustering. Consensus clustering was performed in triplicate using three different seeds and the optimum number of clusters was determined using M3C. M3C calculates (A) a PAC score, (B) -log10p value, and (C) RCSI score for K=2 to maxK. A lower PAC score, higher -log10p value, and higher RCSI score indicate an optimal number of clusters. An optimal class number of K=2 is supported by both the PAC score and the RCSI, with the lowest PAC score and the highest RCSI score across all three replicates, identifying the two broad HRD and NHRD genetic subtypes of myeloma. However, the p-value for K=2 is among the lowest when compared to the other clusters. For values of K greater than 12, the PAC score and p-score begin to overfit the data, indicated by the downward and upward trend respectively. Within the range of K=3-12, K=8 has the lowest PAC score, most significant p value, and a consistently high RCSI score across all three replicates, indicating K=8 is the optimal cluster number. The class assignments from replicate 2 were used for all further downstream analyses.



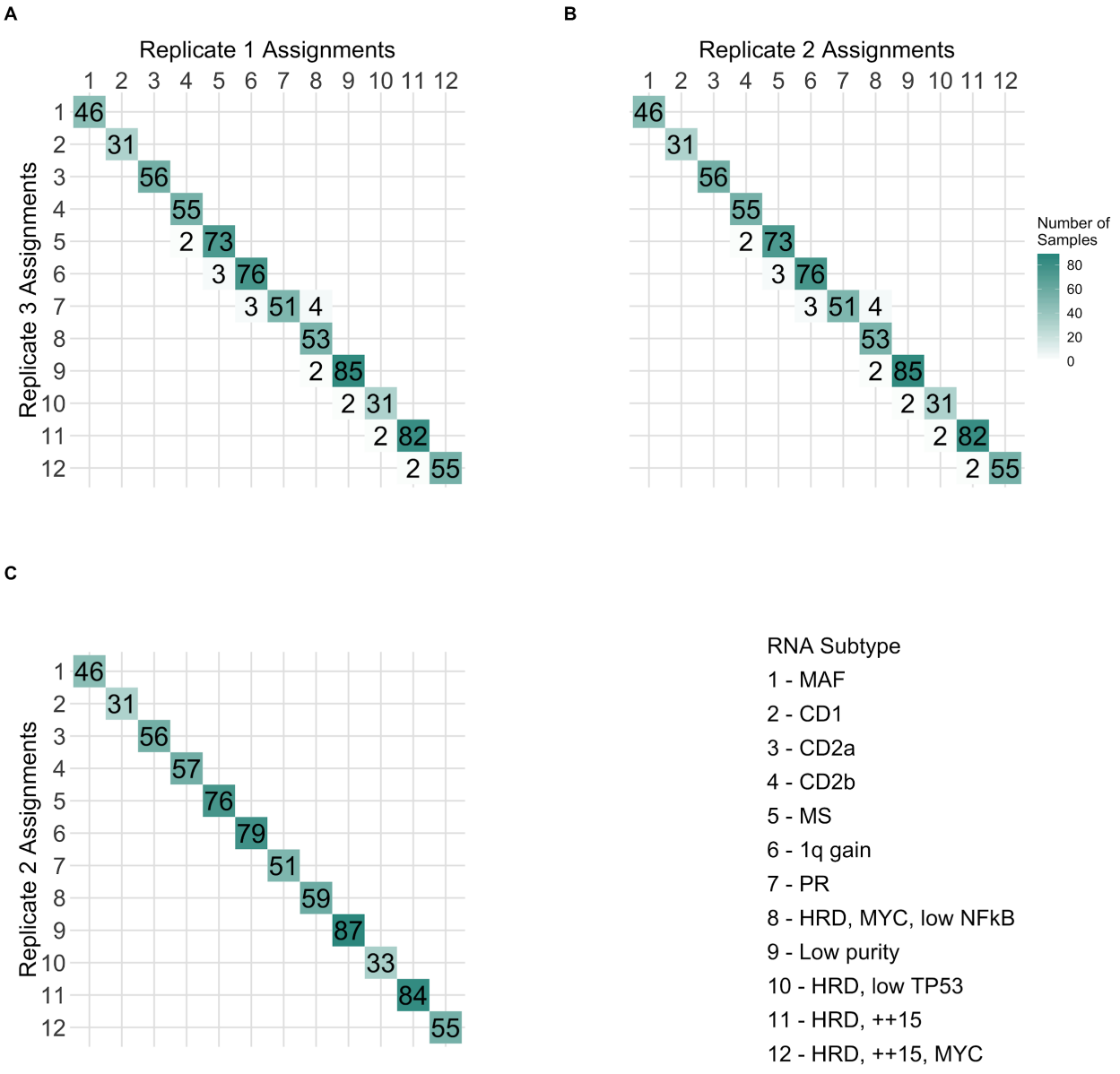
Supplementary Figure 5. CN Density Plots for Subtype Defining Events. Copy number (log2) density plot across CN subtypes for (A) 1q21, (B) 3q21, (C) 7q22, (D) 11p15, (E) 13q14, and (F) 15q15. The red dotted line in plots A-D indicates the 1 copy gain threshold, in plot E indicates the 1 copy loss threshold, and in plot F indicates the 2 copy gain threshold.



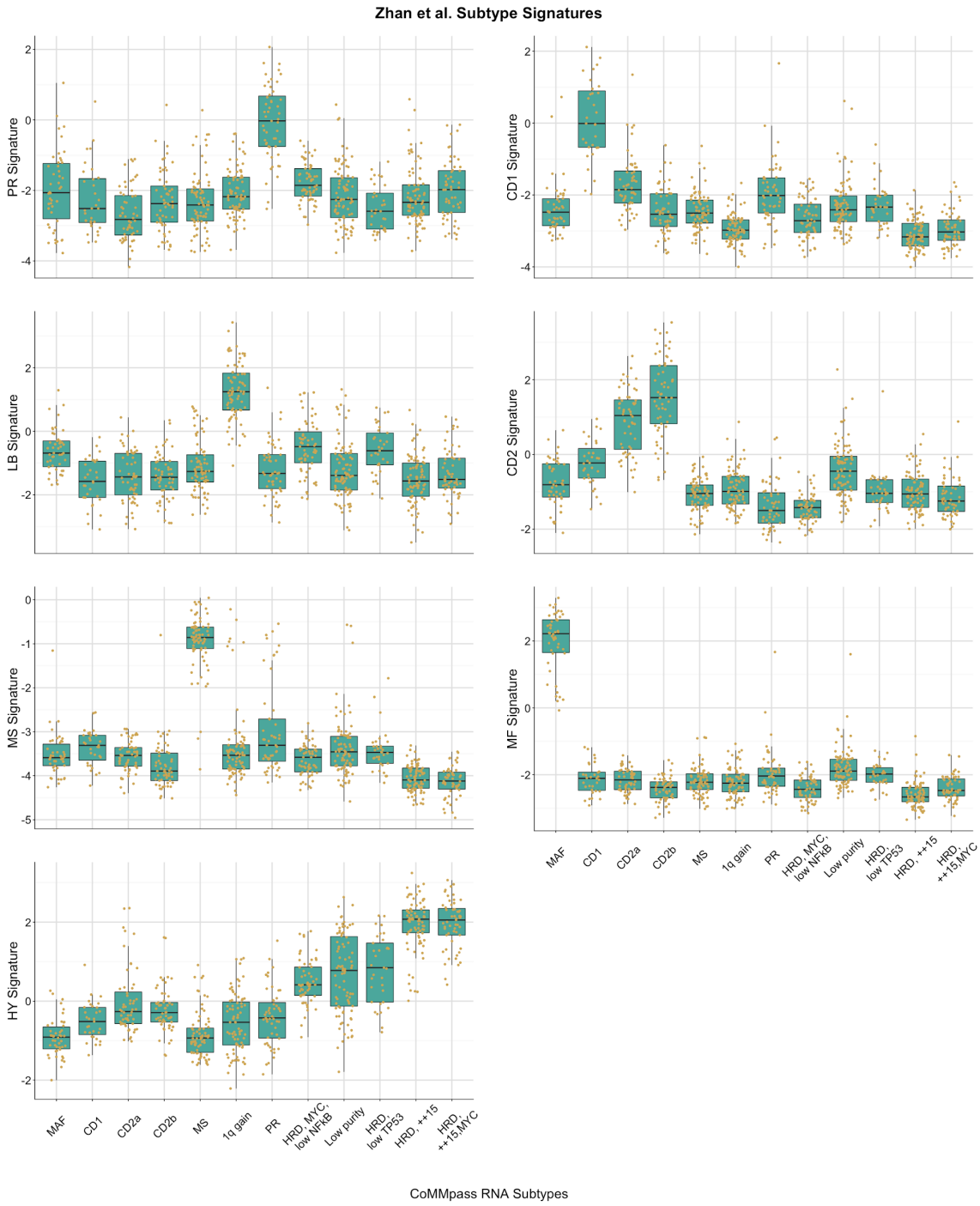
Supplementary Figure 6. Survival Outcomes for HRD and NHRD Patients. A) Time to second line therapy and B) OS outcomes for hyperdiploid (HRD) versus non-hyperdiploid (NHRD) CoMMpass patients. There was no significant difference in time to second line therapy for HRD (41.5 months, 95% CI = 35.5-46.9) versus NHRD (35.8 months, 95% CI = 30.2-39.4) patients, or OS for HRD (median 103.9 months, 95% CI = 97.4-not met) versus NHRD (median 91.6 months, 95% CI = 74.2-not met) patients.



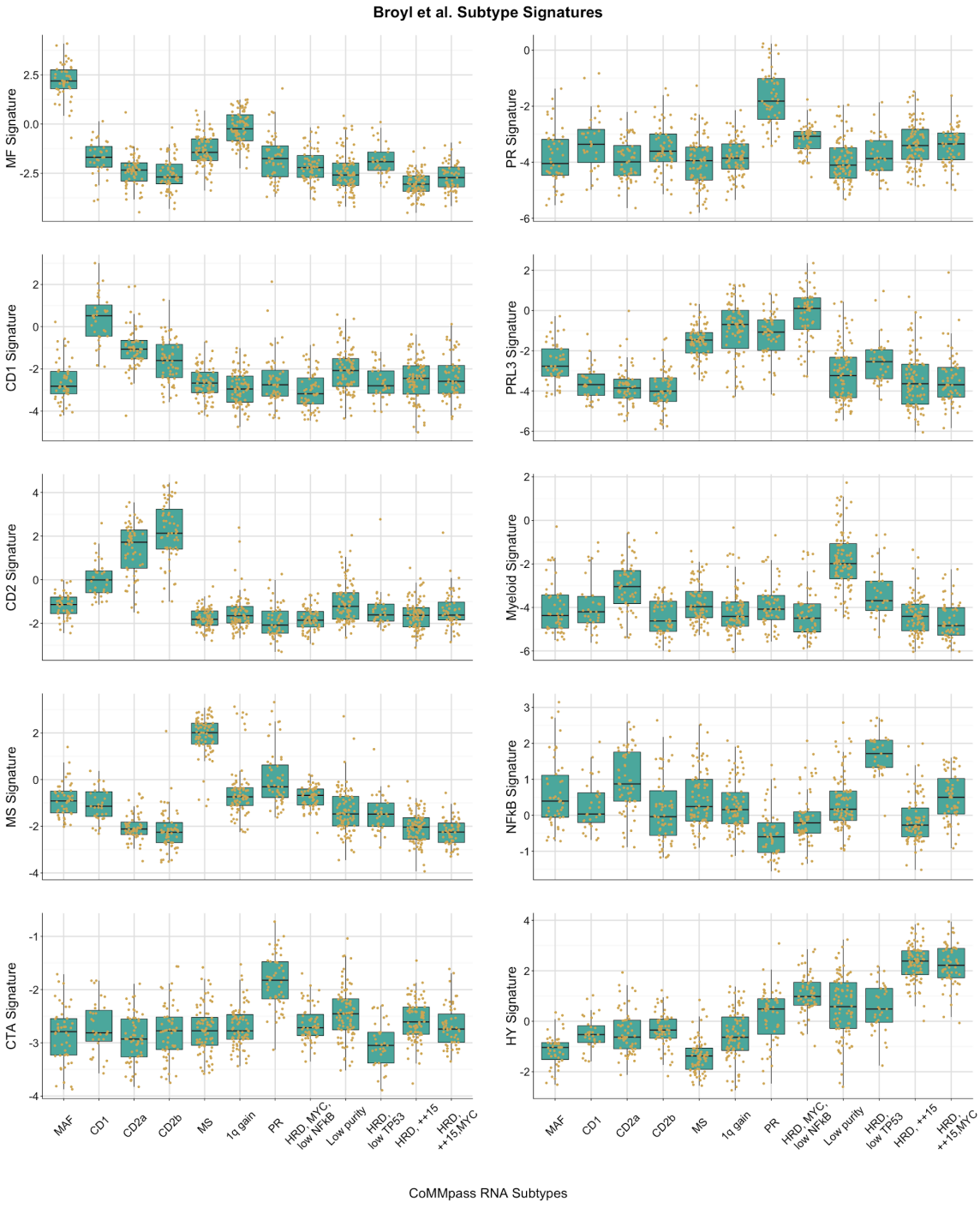
Supplementary Figure 7. Cluster Number Determination from Gene Expression Consensus Clustering. For each of the three replicates of consensus clustering, a (A) silhouette score and (B) relative change in area under the CDF curve was computed for the number of possible clusters tested (K, 2-20). (A) The silhouette score is defined as the average silhouette width, s(x), of all samples in the dataset, where s(x) is a measure of how appropriately grouped all samples are within a cluster, and the average of all silhouette widths reflects overall clustering quality. (B) The proportion increase in the CDF area as K increases, ΔK. Ideally, the optimal number of clusters (K) will correspond to the K that maximizes the silhouette score, s(x), while minimizing the ΔK. We evaluated the resulting grouping from K = 7-15 based on the aforementioned criteria (s(x) score and ΔK) in combination with CoMMpass WGS data and groups identified in previous studies to identify biologically relevant subtypes. Previous studies identified four common groups: MS, CD1, CD2, and PR. In our consensus clustering trials, these classes were not identified until K=11 or greater, in particular, the PR subtype which is the only group with a significant difference in outcome. We ultimately selected K=12, as it had the highest s(x) while minimizing ΔK when compared to other local maximums of s(x). K=9 was eliminated because one cluster only contained 2 patients, and similarly K=10 contained a cluster with only 1 patient, resulting in spurious groups that were not informative of myeloma biology.



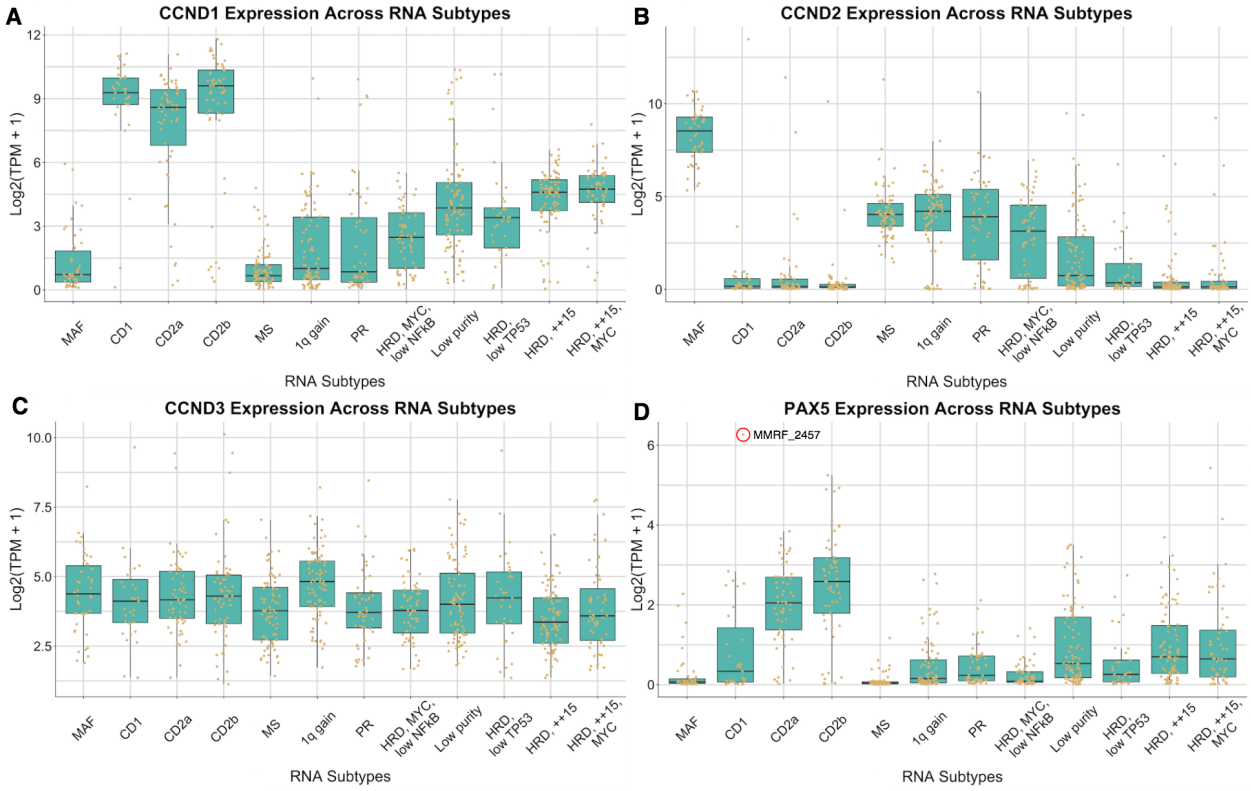
Supplementary Figure 8. Consistency of Subtype Assignments by RNA Expression Clustering Confusion matrices of the number of patients classified in each RNA subtype across three replicates. (C) Replicate 1 and 2 produced identical cluster assignments for all 714 samples. (A-B) 20 samples (2.8% of the total dataset) from replicate 3 were discordant from their assignments in replicates 1 and 2.



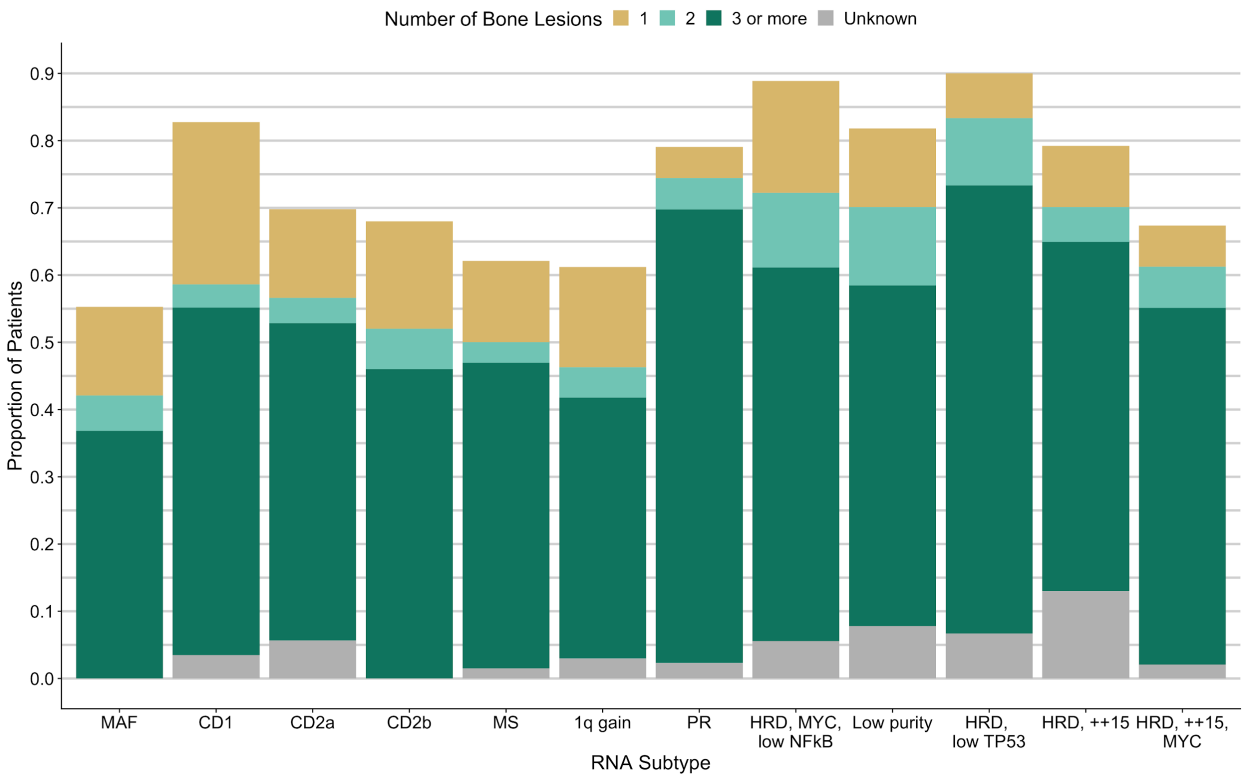
Supplementary Figure 9. Relationship between CoMMpass and Zhan et al. RNA Subtypes. Index values for each of the 7 subtypes defined by Zhan et al.[17] were calculated for each patient (n=714) and compared to the CoMMpass RNAseq native subtype assignments. The distribution of index values between the Zhan et al. subtypes and each identified CoMMpass subtype was used to identify related subtypes. The index ranges are individually shown as a boxplot with the upper and lower bounds of the box representing the 25th and 75th percentile while the center line indicates the median and whiskers represent the highest and lowest value within 1.5(IQR).



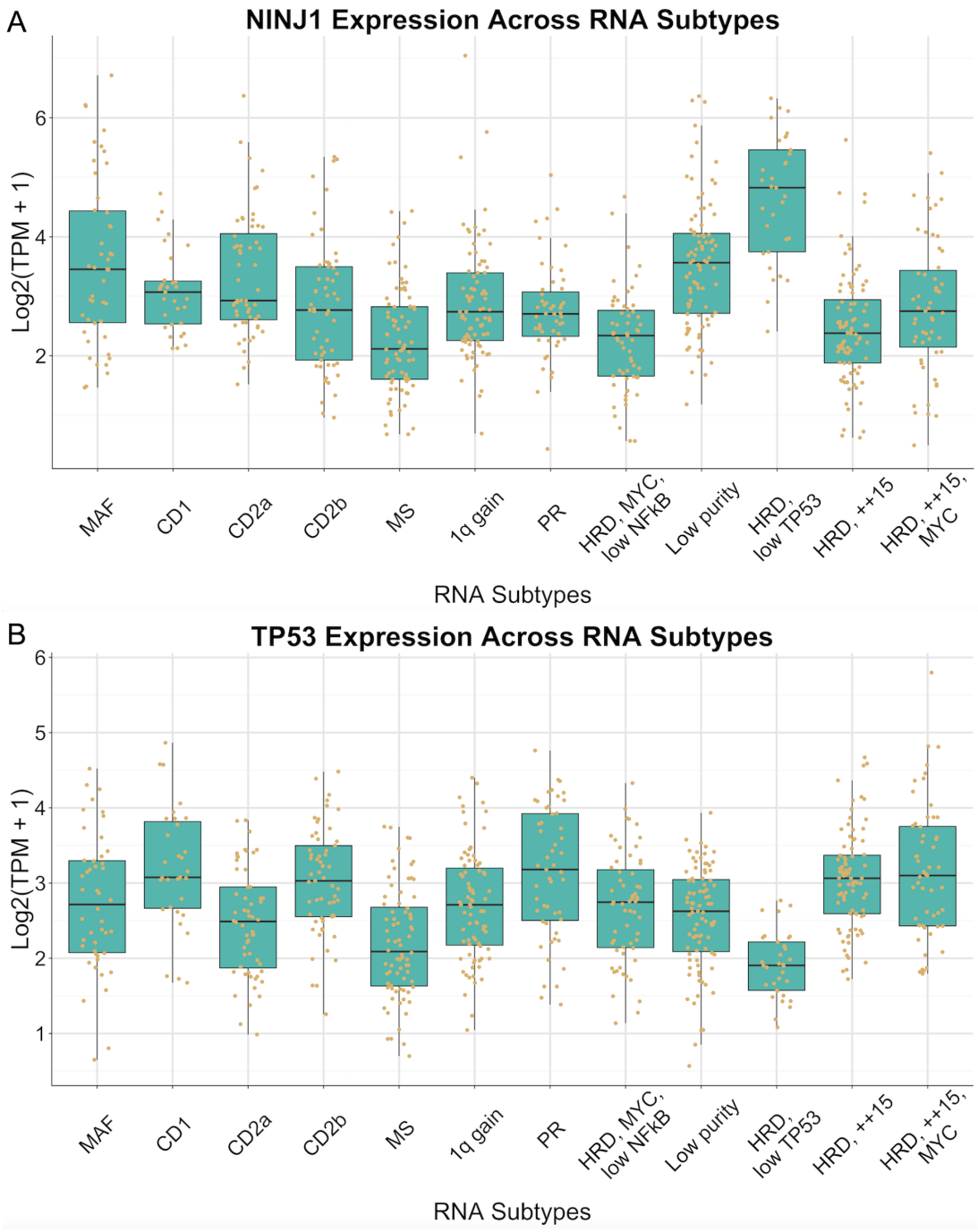
Supplementary Figure 10. Relationship between CoMMpass and Broyl et al. RNA Subtypes. Index values for each of the 10 subtypes defined by Broyl et al.[18] were calculated for each patient (n=714) and compared to the CoMMpass RNAseq native subtype assignments. The distribution of index values between the Broyl et al. subtypes and each identified CoMMpass subtype was used to identify related subtypes. The index ranges are individually shown as a boxplot with the upper and lower bounds of the box representing the 25th and 75th percentile while the center line indicates the median and whiskers represent the highest and lowest value within 1.5(IQR)



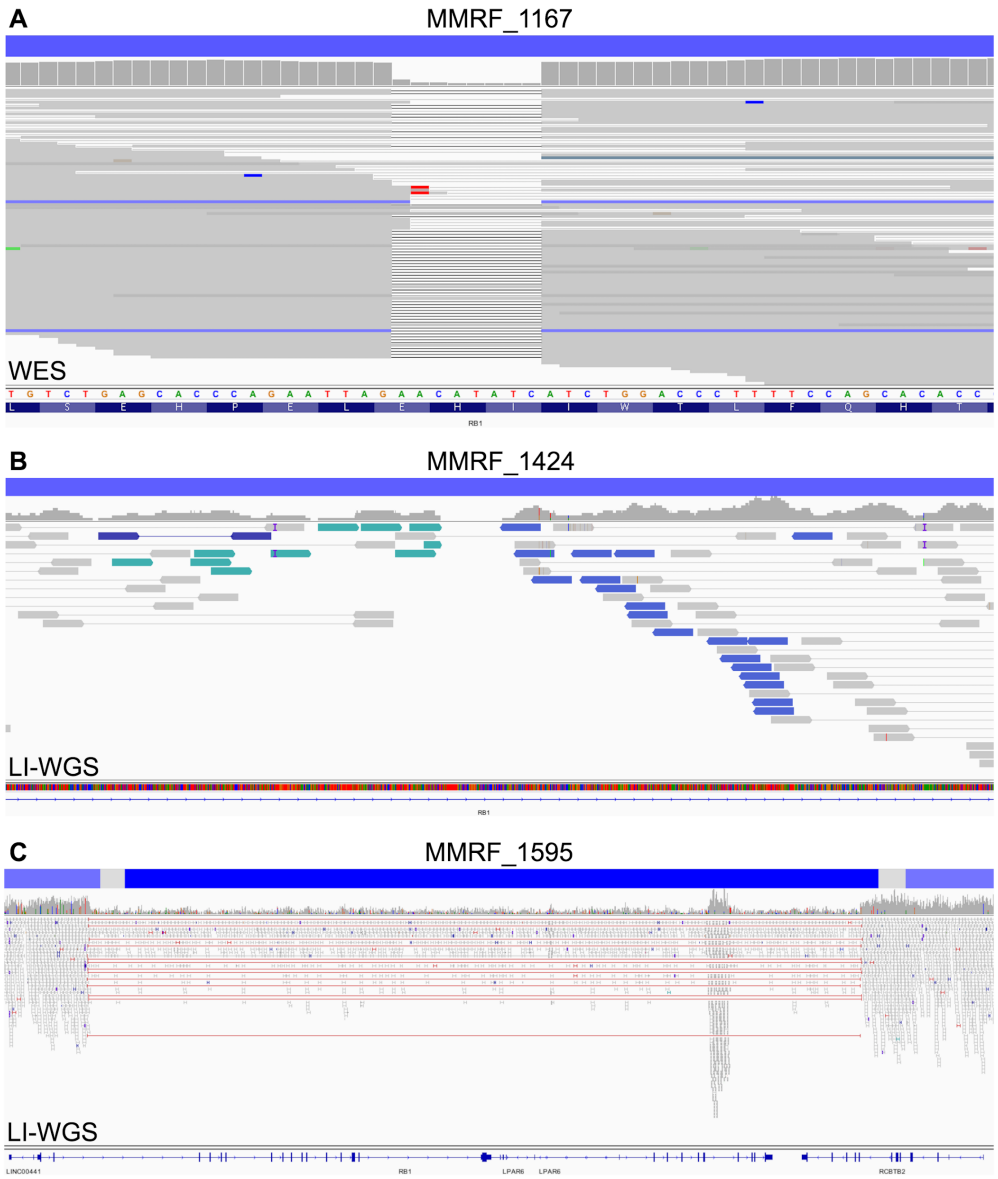
Supplementary Figure 11. CCND1/2/3 and PAX5 Expression Across RNA Subtypes (A) CCND1, (B) CCND2, (C) CCND3, and (D) PAX5 expression across RNA subtypes (n=714). Patients (brown dots) in the CD1, CD2a, and CD2b subtypes typically had overexpression of CCND1, CCND2, or CCND3 due to canonical immunoglobulin translocations targeting these genes. In MMRF_2457 (red circle), a translocation involving CCND1/2/3 was not identified, however, this patient had a t(9;14) resulting in overexpression of PAX5. Notably, the CD2a and CD2b subtypes had the highest median expression of PAX5 across RNA subtypes. In all plots the expression range is shown as a boxplot with the upper and lower bounds of the box representing the 25th and 75th percentile while the center line indicates the median and whiskers represent the highest and lowest value within 1.5(IQR)



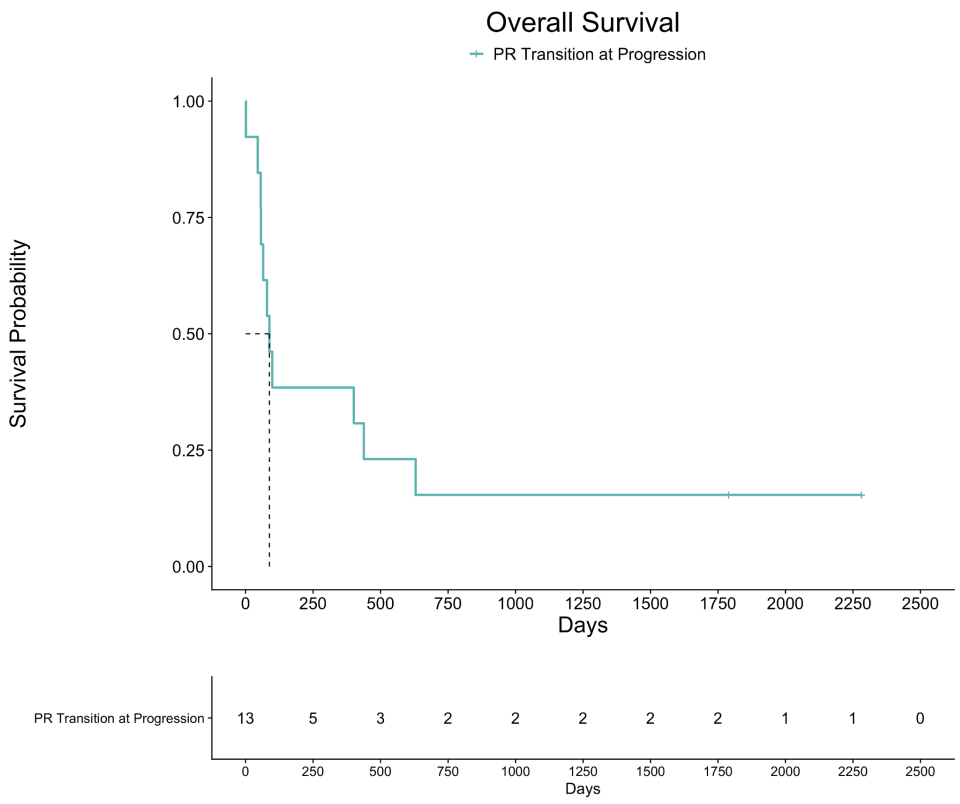
Supplementary Figure 12. Prevalence of Bone Disease Across RNA Subtypes. Proportion of patients in each RNA subtype with bone disease. The proportion of patients with 1, 2, or 3 or more bone lesions for each subtype is also shown. The Unknown (gray) category represents patients with bone disease for whom the number of lesions was not specified.



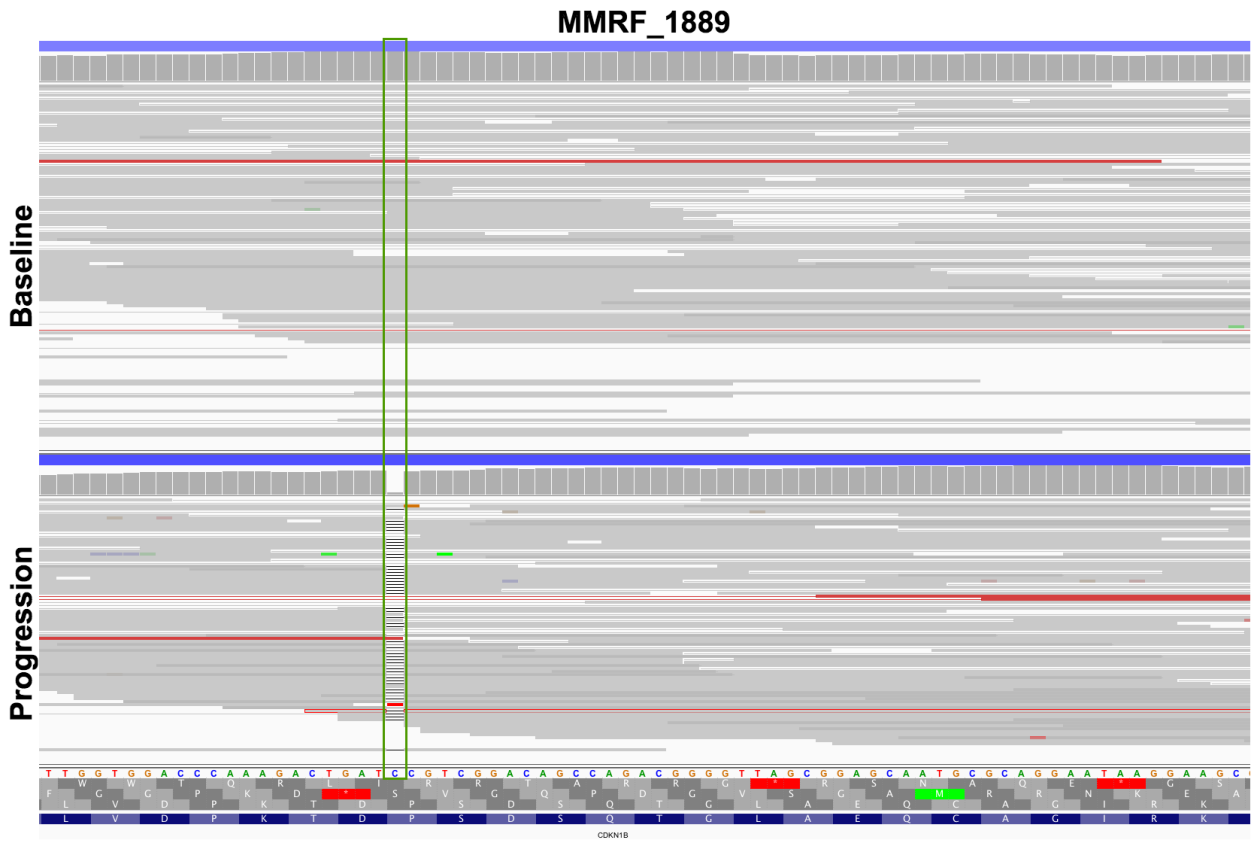
Supplementary Figure 13. *NINJ1* and *TP53* Expression Across RNA Subtypes The expression of (A) *NINJ1* and (B) *TP53* is shown for each CoMMpass RNA subtype (n=714). The expression range is shown as a boxplot with the upper and lower bounds of the box representing the 25th and 75th percentile while the center line indicates the median and whiskers represent the highest and lowest value within 1.5(IQR).



Supplementary Figure 14. Mechanisms of RB1 Complete Loss at Diagnosis. Different mechanisms of complete loss of RB1 are observed in tumors of the PR RNA subtype, but generally involve a one copy deletion of 13q coupled with a second molecular event. Panels A-C show CN segmentation (blue bars) and sequencing data (WES or LI-WGS) for three tumors of the PR RNA subtype at baseline. (A) Patient MMRF_1167 had complete loss of RB1 as a result of 13q copy loss (log2 CN = -1.004) and a small deletion (AR = 0.87). (B) Patient MMRF_1424 had complete loss of RB1 as a result of 13q copy loss (log2 CN = -0.9708) and an inversion where the breakpoint in the intronic region between exons 2 and 3 prevents splicing. (C) Patient MMRF_1595 had complete loss of RB1 as a result of 13q copy loss (log2 CN = -1.006) and a second interstitial deletion (log2 CN = -5.3759) including all RB1 exons except exon 1.



Supplementary Figure 15. Overall Survival of Patients After Transition to PR Subtype. OS outcome for the 13 patients that transition to the PR subtype at progression from any non-PR RNA subtype at baseline, excluding the low purity subtype. Days are landmarked to the date at which the progression visit bone marrow sample was obtained to the date of last follow up (no OS censor flag, 2 patients) or to the date of death (OS censor flag, 11 patients). Patients that transitioned to the PR subtype exhibited extremely poor survival outcomes, with median OS of 88 days (3 months) after the progression visit.



Supplementary Figure 16. Deletion of CDKN1B in a Patient that Transitioned to PR. One patient that transitioned to the PR subtype at progression acquired complete loss of function of CDKN1B due to copy loss and mutation. Panels A and B show WES data for patient MMRF_1889 at (A) baseline and (B) progression, when the patient transitioned to the PR subtype. (A) At baseline the patient had a one copy loss of CDKN1B due to an arm-level deletion of 12p. (B) At progression, the patient had complete loss of CDKN1B due to copy loss of 12p and a clonal frameshift mutation (AR = 0.98). There was no read evidence supporting the existence of a subclone with this mutation at diagnosis, suggesting that this mutation was acquired.

Detailed Methods

RNA Sequencing (RNAseq) Library Preparation

All RNA sequencing libraries were constructed using the Illumina TruSeq RNA library kit following the manufacturer's recommendations. Over the course of the project the primary target input changed from 2000 ng to 500 ng. When 500 ng of RNA was not available, samples were processed using a lower input of 250 ng or 150 ng. In all cases, the input quantity was based on nanodrop quantification. RNA quality was confirmed at TGen to have a required RIN ≥ 7 using an Agilent Bioanalyzer with the RNA 6000 Nano Kit (Agilent, 5067-1511) or eRIN ≥ 6 on an Agilent Tapestation using the RNA ScreenTape assay (Agilent, 5067-5576, 5067-5577). Libraries were prepared following the manufacturer's recommendation for the TruSeq RNA Library Prep Kit v2 (Illumina, RS-122-2001), unless otherwise noted. Eight cycles of PCR amplification for 2000 ng and 500 ng input, 9 cycles for 250 ng input, and 10 cycles for 150 ng input were performed. Final libraries were quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and assessed on the Agilent TapeStation using the High Sensitivity D1000 ScreenTape assay (Agilent, 5067-5584, 5067-5585).

Long Insert Whole Genome Sequencing Library Preparation

Long Insert whole genome sequencing (LI-WGS) libraries were initially generated using the Illumina TruSeqDNA Whole Genome kit (TSWGL). Fragmentation of 600-1100 ng of DNA was performed on a Covaris E210 Focused Ultrasonicator with the following parameters: duty cycle = 2, peak power = 6, cycles per burst = 200, and time = 20 sec. Libraries were prepared according to the standard Illumina protocol, however, the AMPure bead ratios were adjusted to either 1:1 or 1:0.8. After ligation clean up, samples were run on a 1.5% agarose gel and Xtracta gel extractors (USA Scientific, 5454-2500) were used to extract library molecules at approximately 1.3, 1.0, and 0.8 kb. Size-selected samples were processed with Freeze 'N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad, 732-6165). Purified 1 kb samples were concentrated using AMPure XP beads and amplified with 10 PCR cycles. Final libraries were run on Agilent Bioanalyzer DNA 12000 chips (Agilent, 5067-1508 & 6067-1509). If a patient's tumor and constitutional samples were not

within ~100 bp of each other, alternate excised samples (1.3kb or 0.8kb) were processed to generate a better match.

For samples processed using the KAPA HTP/LTP Library Preparation Kit (Roche, KK8234) and off-bead protocol (KAWGL), 200-1100 ng of DNA was fragmented on the Covaris E210 (see parameters above) or E220 with the following parameters: duty cycle = 4%, initial / peak power = 170 W, cycles per burst = 200, and time = 20 sec. Libraries were prepared according to the standard Kapa protocol, however, the AMPure bead ratio was adjusted to either 1:1 or 1:0.8. Libraries were amplified for two cycles before size selection to linearize the fragments and five cycles after size selection using KAPA HiFi HotStart ReadyMix (Roche, KK2602)

For samples processed using the KAPA HyperPrep kit (Roche, KK8505) (KHWGL), 200 ng of DNA was fragmented on the E210 or E220 Covaris with the parameters defined above. Libraries were prepared according to the manufacturer's protocol, however the post ligation AMPure bead ratio was adjusted to 1:0.8. Molecules between 950-1050 bp were either extracted automatically from a Sage Science Pippin Prep 1.5% gel (Sage Science, CSD1510) or hand punched from a 1.5% agarose gel. One cycle of PCR amplification pre size selection, followed by 6 cycles of amplification post size selection, was performed using KAPA HiFi HotStart ReadyMix.

Whole Exome Sequencing (WES) Library Preparation

Initially samples were processed using the Illumina TruSeq Exome Library Prep Kit (TSE61) until the product was discontinued. Genome libraries were created using 1100 ng of DNA as input for fragmentation on the E210 Covaris using the following settings: duty cycle = 10, intensity = 5, cycles per burst = 200, time = 120 seconds. Individual DNA samples were mixed with TElowE (10 mM Tris-HCL, 0.1 mM EDTA) to a final volume of 55 µl in a covaris microTUBE plate. Six libraries were pooled together before enrichment following the manufacturer’s protocol.

Subsequently, samples were processed using the KAPA HTP/LTP Library Preparation Kit using the off-bead protocol and 8-plex pooled enrichment was performed using Agilent

SureSelect XT2 Human All Exon V5+UTR baits (Agilent, G9661B) (KAS5U). Genome library preparation was performed according to the manufacturer's protocol with 500 ng of input DNA fragmented to an average target size of 160 bp using a Covaris E220 focused-ultrasonicator with the following settings: duty cycle 10%; peak power 175; cycles per burst 200; time 300; power mode “frequency sweeping”; bath temperature 7°C. Five cycles of PCR amplification was performed pre-capture using KAPA HiFi HotStart ReadyMix, and the resulting libraries were quantified using either the Agilent Bioanalyzer 1000 chip (Agilent, #5067-1504) or the Agilent Tapestation D1000 kit (Agilent, 5067-5582, 5067-5583). Eight libraries were pooled at 187.5 ng each before capture following the Agilent XT2 protocol. The enriched library pool was amplified for 8 cycles using KAPA HiFi HotStart ReadyMix. Final libraries were quantified using the Agilent Bioanalyzer High Sensitivity DNA Kit (Agilent, #5067-4626) or Agilent Tapestation HSD1000 (Agilent, #5067- 5584 & 5067- 5585) and Qubit High-Sensitivity assay (Invitrogen, #Q32854). Samples processed using the KAPA HTP/LTP Library Preparation Kit and the on-bead protocol followed by 8-plex pooled enrichment performed using Agilent SureSelect XT2 Human All Exon V5+UTR baits (KBS5U) were performed according to the manufacturer's protocol with 500 ng of input DNA as described above for the KAS5U protocol.

Samples processed using KAPA HyperPrep Kits were prepared on an Agilent Bravo liquid handler followed by single-plex Agilent SureSelect XT enrichment using V5+UTR baits (KHS5U). Genome libraries were prepared from 200ng, 100ng, or 50ng of dsDNA. Each sample was fragmented to an average target size of 160 bp using a Covaris E220 with the following settings: duty cycle = 10%; peak power = 175; cycles per burst = 200; time = 300; power mode = “frequency sweeping”; bath temperature = 7°C. Pre-capture PCR using 7, 8, or 9 cycles of amplification was performed for the 200, 100, and 50 ng inputs, respectively, followed by 8 cycles of post-capture PCR.

Illumina Sequencing

Sequencing was performed on an Illumina HiSeq2000 or HiSeq2500 at TGen using Illumina HiSeq v3 or v4 chemistry. Diluted library pools with 1% PhiX control libraries were clustered on

Illumina cBOT instruments as recommended by the manufacturer. In all cases, sequencing assays used a paired-end sequencing format with at least 82x82 nucleotide reads. The majority of RNA sequencing libraries, which all had 6 bp sample indexes, were sequenced using paired-end 83x83 reads. Exome libraries with 6 bp sample indexes were sequenced using paired-end 83x83 reads, while those with 8 bp sample indexes were sequenced using 82x82 reads. Whole genome long-inserts were typically clustered on individual lanes, allowing a paired-end 86x86 format. In all cases, raw sequencing data was extracted from the BCL files in the resulting Illumina run folders using BCL2FASTQ v1.8.4 or BCL2FASTQ v2.17.1 to generate industry standard FASTQ files.

Integrated Analysis for Gain and Loss of Function Genes

To predict the functional status of each gene in each sample, we performed an analysis integrating multiple measurements from different sequencing assays for each sample. The functional state of each gene was predicted independently to capture loss-of-function (LOF) and gain-of-function (GOF) events. This analysis leveraged 7 outputs from WGS, WES, and RNAseq data for samples with all three data types (592 baseline samples). We sourced the somatic gene level CN and structural abnormalities (deletion, inversion, duplication, translocation) from WGS; the somatic non-synonymous (NS) mutations, B-allele frequency (BAF) and constitutional loss of function mutations from WES; and the gene expression (TPM), cohort normalized median absolute deviation (MAD) expression, and inframe fusion transcripts from RNAseq.

The gene model from Ensembl v74 defines 63,677 distinct genes or gene elements, however, only 57,997 of these map to a contig defined by the hs37d5 reference genome used in this study, and only 57,736 of these map to a primary contig (chromosomes 1-22, X, and Y). The list of genes was reduced to a final list of 23,221 analyzed genes by excluding immunoglobulin elements, excluding genes whose gene names were missing or contained “.” or “-”, and requiring the genes were annotated with the following biotypes: lincRNA, miRNA, processed_transcript, protein_coding, snoRNA, snRNA, TR_C_gene, TR_D_gene, TR_J_gene, TR_J_pseudogene, TR_V_gene, TR_V_pseudogene.

Loss-of-Function (LOF)

For the LOF analysis we limited the genes to potential tumor suppressor genes (TSG) based on the assumption that these genes would be detectably expressed in the majority of the cohort. To perform this filtering step, 71 known TSGs were obtained from the Cosmic Cancer Gene Census (v75) (http://cancer.sanger.ac.uk/census) that existed in our gene model. For a gene to be included in the LOF analysis, the median expression of the gene in the baseline CoMMpass cohort had to be greater than the median of the 10^{th} percentile of the gene expression of all 71 TSGs. A total of 10,577 genes were included in the LOF analysis after applying the above filter.

In the LOF analysis, a gene was defined as being functional, or exhibiting complete or partial LOF. Complete LOF implied that all functional alleles of the gene are lost, whereas partial LOF implied that there is evidence of inactivation of one allele, and a potential haploinsufficiency. For example, a somatic deletion, NS mutation, structural rearrangement within a gene body, copy neutral LOH, or a constitutional LOF mutation would result in partial LOF of a gene, whereas two or more of these events impacting a gene would result in complete LOF. We used a heuristic approach to determine the functional state of a gene where each position of an 11 bitset identified the presence or absence of a specific genetic event category. The bits and their corresponding definitions in order are: Cl, 1-copy loss; Cd, 2-copy loss; Cn, copy neutral (2-4 copies); M2, two or more NS mutations; M1, one NS mutation; Lh, loss-of-heterozygosity; Ne, gene not expressed; sD, structural deletion; sT, structural translocation; sI, structural inversion; Gm, constitutional SNP. These 11 bits were summarized to assign each gene a status of complete loss, partial loss, or functional.

The segmented CN data was transformed into a sample by gene matrix where each entry represented the lowest log2 fold change observed overlapping the gene body (Table S5). These entries were translated into three flags: '*Cd*' for homozygous deletions or 2-copy loss; '*Cl*' for 1-copy or heterozygous loss; and '*Cn*' designating 2-4 gene copies. We used a log2 threshold of ≤-2.32 to identify homozygous deletions, which represents the theoretical value if a homozygous deletion exists in 80% of cells, while the remaining 20% of cells are in the normal copy state. A

threshold of ≤-0.1613 was used to identify single copy heterozygous losses, which represents the theoretical log2 value if a single copy deletion exists in ~21.1% of cells.

The only somatic mutations included in the analysis were those annotated as NS mutations with the most damaging effect of the mutation to any transcript being the source of the effect used in the LOF model (Table S2). The count of these mutations per gene was used to define the 'M1' bit, which is set when a single mutation is observed, and the 'M2' bit when multiple NS mutations were detected. A single NS mutation in a gene, 'M1'=1, was considered a partial loss and could contribute to complete loss if an additional event from any other mechanism (ie. CN loss) was also observed. When two independent NS mutations were detected in a gene, 'M2'=1, we assumed they exist in trans, which would represent a bi-allelic loss and thus could be interpreted as a complete LOF based on mutation data alone.

To identify regions of copy neutral LOH we integrated the observed copy number and B-allele frequency (BAF). To set the '*Lh'* bit, we extracted the largest overlapping segment from the absolute BAF segmentation data, and set '*Lh'=1* when '*Cn'=1* and the absolute BAF value was between 0.45 and the max value of 0.5 (Table S5).

To capture loss of gene expression through potential secondary mechanisms, such as epigenetic repression, we identified genes with outlier loss of expression by integrating MAD analysis and the absolute expression value. We used the MAD approach utilized for Cancer Outlier Profile Analysis (COPA) to normalize the gene expression TPM values[33]. Using this approach, the TPM for each gene was median centered and scaled to their median absolute deviation (MAD). Given the TPM of a gene, the MAD value was computed as: $MAD_{sample} = [\log_2(TPM_{sample}) - median(\log_2(TPM_{cohort})] / mad(log2(TPM_{cohort})$. A gene was defined as not expressed if the MAD value was less than the 25th percentile value minus 1.5 times the interquartile range (IQR) and the observed corrected TPM was less than 0.1. When these criteria exist the *'Ne'* bit was set to 1 to denote a potential epigenetic loss of RNA expression.

To identify when an allele of a gene was inactivated by a structural event we annotated each breakend of structural abnormalities detected by Delly (deletions, inversions, duplications, translocations) independently with an intersecting gene if the breakend occurred between the start and end positions of the gene. To minimize potential double counting of deletions detected by copy number analysis and structural analysis, we only included deletions smaller than 15kb, as the copy number analysis rarely detected deletions below 20kb. When an intersecting gene was annotated on an individual translocation, deletion, or inversion breakend we set the 'sT', 'sD' and 'sI' bits, respectively.

Individuals can also inherit a defective copy of a gene contributing to a LOF event when the remaining functional allele is lost. To identify inherited LOF events we performed joint variant calling on the constitutional exomes from the entire MMRF cohort using GATK (v3.5) best practice guidelines. Individual gVCFs were created using HaplotypeCaller with -L to limit the analysis to the regions targeted by the exome kit. The patient specific gVCF files were combined in batches of 100 using CombineGVCFs and then joint SNV and indel calling was performed across the batched gVCFs using GenotypeGVCFs. Variant quality recalibration (VQSR) was applied independently on the SNP and INDEL detected to reduce false positives. While applying the recalibration, any SNP below the 99.0 tranche and INDEL below the 95.0 tranche were marked as a PASS. These variants were annotated using GATK VariantAnnotator to flag when a variant existed in one of the following databases: dbSNP (v147), ExAC (v3.0), ClinVar (v20170530), NHLBI (v2 ssa137), 1000 Genomes (Phase 3), and COSMIC (v78). To identify highly damaging LOF variants we annotated the variants independently using snpEFF (v4.2) and VEP(v87)/LOFTEE following established rules[34]. Expected LOF variants were extracted from the annotated files when a heterozygous variant was marked as LOF by snpEFF or a high-confidence LOF by LOFTEE and the allele frequency in the CoMMpass cohort was less than or equal to 0.05, and the variant was marked as PASS or the variant was known in dbSNP or ExAC. Variants annotated as LOF by both tools were then annotated with the observed allele ratio in the matching tumor exome and tumor RNAseq alignments. Using the available data, the *'Gm'* bit was set to indicate if the inherited LOF variant was detectable in the tumor at an expected allele frequency. In the absence of other

somatic events, the bit was set when the AR was at least 0.4. When a CN loss was detected we required the AR to be above 0.6 to confirm the LOF impacted the remaining allele, and when copy neutral LOH was detected we required the AR to exceed 0.85.

After setting each bit for each gene in every sample we used a simple heuristics approach to combine and score these events to assign each gene an interpreted functional state. Each bit was assigned a specific weight when set (Cl=0.5, Cd=1.0, Cn=0, M2=1.0, M1=0.5, Lh=0.5, Ne=0.5, sD=0.5, sT=0.5, sI=0.5, Gm=0.5) and these weights are summed to get a final LOF score where values ≥ 1 denote complete LOF, values = 0.5 denote a partial LOF (haplo-insufficiency), and values = 0 denote genes that are expected to be fully functional.

Gain-of-Function (GOF)

To identify potential oncogenes, we performed a GOF analysis designed to detect common oncogene activation mechanisms including activating mutations, gene amplification, RNA over-expression with concomitant structural events, and inframe gene fusions. The status of each gene was summarized into a 6 bitset which identified the presence or absence of a specific genetic event. The bits and their corresponding definition in order are: Cg, copy gain; Rm, recurrent mutation; Nm, nominated mutation; Cm, clustered mutation; Oe, overexpression; and Hf, gene fusion.

To identify high level amplifications, the '*Cg*' bit was set to 1 when the CN segment for a gene was estimated to exceed 6 total copies (log2 CN ≥1.5), the full gene was contained within the amplification, and the gene was detectably expressed (TPM ≥1). The expression filter prevents non-expressed genes within an amplified segment from being nominated as a potential oncogene.

To flag NS SNV/INDEL that likely result in a GOF events, we identified events that fell into one of three categories based on the altered amino acid position. To identify recurrent mutations, the 'Rm' bit is set for a gene in any patient where the observed mutation altered the same amino

acid in at least two independent patients. When a gene was flagged as having recurrent mutations in the cohort, we also flagged clustered mutations in the gene to set the '*Cm*' bit. Mutations were considered to be clustered when at least 5 different patients had NS mutations within 10% of the CDS space of a gene. Finally, when a gene was flagged as having recurrent mutations, we nominated any NS mutation in a gene by setting the '*Nm*' bit (even including those without the '*Rm*' or '*Cm*' bits set) to capture potential GOF events by rare activating mutations.

To detect overexpression of a gene, which is common in myeloma through events like immunoglobulin translocations, we identified outlier RNA expression using the same MAD approach described in the LOF analysis. A gene was defined as overexpressed if the MAD value was greater than the 75th percentile value plus 1.5 times the interquartile range (IQR) and the observed corrected TPM ≥1. To limit false positive overexpression calls, we only set the '*Oe*' bit when an accompanying DNA structural event was detected in the gene or immediately upstream or downstream before the start or stop of the next coding gene in chromosomal order.

When an inframe fusion transcript was detected in the RNAseq data we set the '*Hf*' bit for both genes involved in the fusion. The fusion events included were independently validated in the matching WGS assay, and the expression of both genes needed to be ≥1 TPM.

A gene was defined as gained in a sample if any of the '*Cg*', '*Rm*', '*Cm*', '*Oe*', or '*Hf*' bits were set to 1. Samples with only the '*Nm*' bit set for a gene were considered to have a potential GOF event and were only included in the count of samples with GOF events in a particular gene when at least 5 unique samples had one of the other bits set. In addition to the individual mechanism of gain, we also required the gene to be expressed (≥1 TPM) in the sample for it to be considered a GOF gene.

Curation of LOF and GOF Genes

The integrated analysis focused on identifying genes with potential recurrent LOF and GOF events, however, some genes were nominated in both the GOF and LOF analysis (e.g. NRAS and

KRAS). For genes that were identified as the target of both LOF and GOF in 5 or more patients, we relied on COSMIC Cancer Gene Census annotations (accessed August 2017) and reports of gene function from published literature. Known oncogenes removed from the LOF but retained on the GOF list were *NRAS*, *KRAS*, and *WHSC1*, while known tumor suppressor genes removed from the GOF list but retained on the LOF list were *CDKN1B*, *DIS3*, *EGR1*, *FAM46C*, *MAX*, *NOTCH2*, *PABPC1*, *PARP4*, *RPL10*, *SDHA*, *TP53*, and *TRAF3*. Genes that were exclusively nominated to the GOF list by mutational events were further curated based on their reported function in the literature and were removed if there was strong evidence supporting that the gene functions as a tumor suppressor rather than an oncogene. The genes removed through this process were *DUSP2* and *KLHL6*, based on evidence from the literature, and *BTG1*, which is listed in the Cancer Gene Consensus as a tumor suppressor gene.

In patients where a gene had complete LOF as a result of two mutations, visual validation of mutations within close proximity was performed to determine if the mutations were in cis or trans. In order to perform visual validation, we required that sequencing reads or read-pairs spanned both of the mutation positions. If the mutations occurred in trans, no action was taken, however, if the mutations occurred in cis the gene was downgraded from complete to partial LOF for that particular sample. *CDKN1B*, *EGR1*, *LRP6*, *MAGEC3*, *PABPC1*, *PRR14L*, *SYNE1*, and *UBR5* had samples whose variants were phased in cis, and after downgrading these events from complete to partial LOF, were removed from the list of recurrent complete LOF events occurring in 5 or more patients.

To further curate the LOF list, visual validation was performed to remove double calling events that can arise such as when a copy number deletion breakend overlaps a translocation breakpoint. Visual verification was limited to samples with contiguous genes identified as having a LOF involving a structural event. From the LOF list, *CDKN2C*, *PSPC1*, *RB1*, *RCBTB2*, *CDC42BPB*, *CYLD*, *SNX20* had samples with LOF events that were manually modified after visual inspection. In addition, *MAGI1* was removed from the GOF list because four of the five patients were flagged

due to a translocation event in *SLC25A26* which is believed to be a false-positive arising due to poor read mapping in this area.

Identification of Copy Number Subtypes in Multiple Myeloma

CN subtype discovery and membership of the BM baseline samples was determined using the Monte Carlo reference-based consensus clustering tool M3C (v3.9) using the PAM clustering algorithm, Euclidean distance, and a max K of 20. In total, WGS data from 871 BM baseline tumor-normal pairs met quality control requirements for use in CN analysis. A uniform data matrix was created by extracting the largest overlapping CN state value for 26,771 non-overlapping 100 kb intervals that excluded centromere, immunoglobulin, and HLA regions along with sex chromosomes to remove biological and sex bias. To confirm the accuracy of the clustering results we ran three replicates, each with 200 iterations of consensus clustering using M3C, with the seed changing between each run.

Identification of RNA Subtypes in Multiple Myeloma

Prior to unsupervised mRNA expression clustering, the log2(TPM+1) transformed expression data of 56,430 genes and 714 BM baseline tumor samples were adjusted for confounding biobank effects by applying ComBat from the sva R package (v3.26.0)[35]. The original list of genes were then ranked by their corresponding geometric coefficient of variation (GCV). Larger GCV values are associated with greater variability in random variables that follow a log-normal distribution and therefore genes with a GCV less than 1 were removed from the dataset. The resulting 4811 gene by 714 sample expression matrix was mean centered as a method for standardizing the data. Clustering analysis was performed by running three separate trials of ConsensusClusterPlus (v1.42.0) with the PAM algorithm and Pearson distance metric in R (v3.5.2)[36]. Each trial computed sample class assignments for an increasing number of clusters (K), from 2 to 20, using the average linkage option and 80% subsampling with the number of repetitions set to 10,000. The optimum K was chosen based on the silhouette score for each cluster across all trials. All three trials with different initial seed values (3000094, 2862787, 8275806) converged to the solution K* = 12 by silhouette score.

Specific gene markers for each group were then determined using multinomial logistic regression with L1 regularization trained to predict class assignment[37]. For training, the model took the log2(TPM+1) mean centered expression matrix of the 4811 genes and 714 samples as input and class assignments from the unsupervised clustering as targets. The value of the regularization parameter that minimized the model's objective function was determined during training via 20 fold cross-validation. The resulting non-zero values for the model coefficients were indicative of genes that were most predictive for a given class; 607 genes in total (Table S7).

Further analysis of the resulting RNA clusters included cross-examination with: a gene expression profiling index, used to assess cellular proliferation; an NFkB index; and previously identified RNA subtypes[17,18,29,38,39]. The proliferation index developed by Bergsagel et al. was calculated for each sample by measuring the average $\log_2$(TPM+1) transformed expression of 12 genes: *TYMS*, *TK1*, *CCNB1*, *MKI67*, *KIAA101*, *KIAA0186*, *CKS1B*, *TOP2A*, *UBE2C*, *ZWINT*, *TRIP13*, *KIF11*. The Mann-Whitney test was used to compare proliferation index scores for the CD2a versus CD2b subtypes. In addition, the NFkB(11) index for each sample was determined by calculating the geometric mean expression of 11 genes: *BIRC3*, *TNFAIP3*, *NFKB2*, *IL2RG*, *NFKB1*, *RELB*, *NFKBIA*, *CD74*, *PLEK*, *MALT1*, and *WNT10A*. Each subtype specific expression signature is composed of a list of up-regulated and down-regulated genes. In both Broyl and Zhan et al., these genes were identified using Affymetrix U133Plus2.0 microarray data. In order to apply these gene lists to RNAseq data, we identified the corresponding Ensembl v74 ENSG ID for each Affymetrix probe set using bioMartR, however, not all probesets had unique mappings to ENSG IDs. Many probesets were associated with multiple ENSG while other probesets were not associated with an ENSG. Only genes with unique probeset to ENSG mappings were used to calculate subtype specific expression signatures. Once the unique mappings were identified, each subtype specific signature was calculated for each sample using: $mean(\, log_2(\, x_{up-regulated}) - log_2(\, x_{down-regulated}\,)\,)$. Integration of the genes predictive of each CoMMpass subtype, canonical structural events in MM, common CN events, previous subtype signatures, and expression index data was used to name the 12 subtypes.

To assign longitudinal samples with RNAseq data to each of these classes we developed a trained model classifier, which has the benefit of outputting class probabilities that can be used to evaluate transitions in time. The trained classifier uses the corrected TPM values for the 4811 genes used for clustering as input. These values are transformed, log2(TPM+1), followed by mean centering using the pre-computed mean expression values from the training set. Running this input through the trained classifier resulted in classifications for each of the 107 samples as well as the corresponding class probabilities associated with each of the 12 classes.

To determine whether checkpoint inhibitor and immunotherapy targets were expressed in PR patients, we compared the expression of each target in 51 PR patients and 663 non-PR patients at baseline. Targets for which the median expression in PR and non-PR patients was <1 TPM were excluded. An unpaired, two-sided Wilcoxon (Mann-Whitney) test was used to test the null hypothesis of equal median expression in both groups.

Clinical and Molecular Associations with RNA subtypes

We used the non-parametric Fisher's Exact test to determine if a clinical feature, LOF, or GOF event was enriched in the identified CN and RNA subtypes. The test works on the null hypothesis that no nonrandom associations exist between two categorical variables, against the alternative hypothesis that there is a nonrandom association between variables. The enrichment analysis was performed using the function *fishertest* in Matlab2019a. The inputs for the enrichment analyses were the LOF and GOF file filtered for 5 or more patients, the patient features table (Table S1), and the gene model. The two-tailed test was used to check for enrichment or depletion of each measurement within each subtype. An odds ratio was computed that indicated significant enrichment or depletion of the measurement in each subtype such that an odds ratio >1 indicated enrichment and <1 indicated depletion. In addition, for a condition to be depleted in a group it was required to be present in at least 20% of the cohort. The Benjamini-Hochberg test correction was used for multiple testing, which computed a pFDR value.

References

33. Tomlins, S. A. *et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644–648 (2005).

34. Macarthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).

35. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

36. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).

37. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).

38. Annunziata, C. M. *et al.* Frequent engagement of the classical and alternative NF-kappaB pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell* **12**, 115–130 (2007).

39. Demchenko, Y. N. *et al.* Classical and/or alternative NF-kappaB pathway activation in multiple myeloma. *Blood* **115**, 3541–3552 (2010).