

1 Genomic Basis of Multiple Myeloma Subtypes from the MMRF 2 CoMMpass Study

3
4 Sheri Skerget¹, Daniel Penaherrera¹, Ajai Chari², Sundar Jagannath², David S. Siegel³, Ravi Vij⁴,
5 Gregory Orloff⁵, Andrzej Jakubowiak⁶, Ruben Niesvizky⁷, Darla Liles⁸, Jesus Berdeja⁹, Moshe
6 Levy¹⁰, Jeffrey Wolf¹¹, Saad Z. Usmani¹², The MMRF CoMMpass Network¹³, Austin W.
7 Christofferson¹, Sara Nasser¹, Jessica L. Aldrich¹, Christophe Legendre¹, Brooks Benard¹, Chase
8 Miller¹, Bryce Turner¹, Ahmet Kurdoglu¹, Megan Washington¹, Venkata Yellapantula¹, Jonathan
9 R. Adkins¹, Lori Cuyugan¹, Martin Boateng¹, Adrienne Helland¹, Shari Kyman¹, Jackie McDonald¹,
10 Rebecca Reiman¹, Kristi Stephenson¹, Erica Tassone¹, Alex Blanski¹⁴, Brianne Docter¹⁴, Meghan
11 Kirchhoff¹⁴, Daniel C. Rohrer¹⁴, Mattia D'Agostino¹⁵, Manuela Gamella¹⁵, Kimberly Collison¹⁶,
12 Jennifer Stumph¹⁶, Pam Kidd¹⁶, Andrea Donnelly¹⁷, Barbara Zaugg¹⁷, Maureen Toone¹⁸, Kyle
13 McBride¹⁸, Mary DeRome¹³, Jennifer Yesil¹³, David Craig¹, Winnie Liang¹, Norma C. Gutierrez¹⁹,
14 Scott D. Jewell¹⁴, John Carpten¹, Kenneth C. Anderson²⁰, Hearn Jay Cho^{2,13}, Daniel Auclair¹³,
15 Sagar Lonial^{21,*}, Jonathan J. Keats^{1*}

16
17 ¹Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix, AZ,
18 85004, USA

19 ²Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA

20 ³Hackensack University Medical Center, Hackensack, NJ, 07601, USA

21 ⁴Division of Oncology, Washington University, St. Louis, MO, 63110, USA

22 ⁵Virginia Cancer Specialists, Fairfax, VA, 22031, USA

23 ⁶University of Chicago Medical Center, Chicago, IL, 60637, USA

24 ⁷Weill Cornell Medicine, New York, NY, 10065, USA

25 ⁸Division of Hematology/Oncology, East Carolina University, Greenville, NC, 27834, USA

26 ⁹Sarah Cannon Research Institute, Nashville, TN, 37203, USA

27 ¹⁰Baylor Scott & White Research Institute, Dallas, TX, 75246, USA

28 ¹¹Department of Medicine, UCSF Medical Center, San Francisco, CA, 94117, USA

29 ¹²Levine Cancer Institute, Charlotte, NC, 28204, USA

30 ¹³Multiple Myeloma Research Foundation, Norwalk, CT, 06851, USA

31 ¹⁴Van Andel Institute, Grand Rapids, MI, 49503, USA

32 ¹⁵Myeloma Unit, Division of Hematology, University of Torino, Azienda Ospedaliero-Universitaria
33 Città della Salute e della Scienza di Torino, Torino, Italy

34 ¹⁶Spectrum Health, Grand Rapids, MI, 49503, USA

35 ¹⁷Precision for Medicine, Flemington, NJ, 08822, USA

36 ¹⁸InStat Services, Chatham, NJ, 07928, USA

37 ¹⁹Department of Hematology, University Hospital of Salamanca, IBSAL, CIBERONC, Salamanca,
38 Spain

39 ²⁰Dana-Farber Cancer Institute, Harvard Cancer Center, Boston, MA, 02215, USA

40 ²¹Department of Hematology and Medical Oncology, Emory University School of Medicine,
41 Atlanta, GA, 30322, USA

42

43 * Correspondence: jkeats@tgen.org & sloni01@emory.edu

44 Abstract

45 Multiple myeloma is a treatable, but currently incurable, hematological malignancy of
46 plasma cells characterized by diverse and complex tumor genetics for which precision medicine
47 approaches to treatment are lacking. The MMRF CoMMpass study is a longitudinal, observational
48 clinical study of newly diagnosed multiple myeloma patients where tumor samples are
49 characterized using whole genome, exome, and RNA sequencing at diagnosis and progression,
50 and clinical data is collected every three months. Analyses of the baseline cohort identified genes
51 that are the target of recurrent gain- and loss-of-function events. Consensus clustering identified
52 8 and 12 unique copy number and expression subtypes of myeloma, respectively, identifying high-
53 risk genetic subtypes and elucidating many of the molecular underpinnings of these unique
54 biological groups. Analysis of serial samples showed 25.5% of patients transition to a high-risk
55 expression subtype at progression. We observed robust expression of immunotherapy targets in
56 this subtype, suggesting a potential therapeutic option.

57 Introduction

58 Multiple myeloma is a treatable, but currently incurable, hematological malignancy of
59 plasma cells. The incorporation of new treatment modalities over the last two decades vastly
60 improved overall survival of myeloma patients, however, patients still ultimately relapse and there
61 remains a subset of high-risk patients with poor outcomes. Despite significant efforts to
62 understand the molecular basis of the disease, predicting patient outcomes and identifying high-
63 risk patients remains a challenge.

64
65 Multiple myeloma is a genetically heterogeneous disease with two broad karyotypic
66 groups. A hyperdiploid (HRD) phenotype, with characteristic trisomies of chromosomes 3, 5, 7, 9,
67 11, 15, 19, and 21, is present in 50-60% of tumors¹⁻³. The remaining non-hyperdiploid (NHRD)
68 tumors, with pseudo-diploid karyotypes, typically have an immunoglobulin translocation
69 dysregulating NSD2/WHSC1/MMSET, MYC, CCND1, or MAF⁴⁻⁹. Tumors harbor many other
70 genetic aberrations including non-immunoglobulin structural abnormalities and mutations¹⁰⁻¹³.
71 Although previous genomic studies were instrumental in deconvoluting the genetic heterogeneity
72 of myeloma, they are mostly limited by small cohort sizes, the number and types of assays
73 performed, lack of longitudinal sampling, clinical follow up, and biased inclusion of heavily
74 pretreated patients, limiting our comprehensive understanding of the disease.

75
76 To better understand the impact of tumor genetic profile on patient outcomes and
77 treatment response, the Multiple Myeloma Research Foundation sponsored the Relating Clinical
78 Outcomes in Multiple Myeloma to Personal Assessment of Genetic Profile (CoMMpass) Study
79 (NCT01454297). CoMMpass is an ongoing, prospective, longitudinal, observational clinical study
80 which accrued 1143 newly diagnosed, previously untreated multiple myeloma patients from sites
81 throughout the United States, Canada, Spain, and Italy between 2011-2016. Comprehensive
82 molecular profiling is performed on bone marrow derived tumor samples collected at diagnosis
83 and each progression event using whole genome (WGS), whole exome (WES), and RNA
84 (RNAseq) sequencing. Clinical parameters are collected every three months through the eight
85 year observation period.

86
87 We present a molecular analysis of the complete baseline cohort, with a median follow up
88 of 3 years, identifying recurrent loss- and gain-of-function events and distinct copy number (CN)
89 and gene expression subtypes of myeloma. The comprehensive nature of this dataset and our

90 integrated analysis framework defines both the overall frequency of gene alterations in myeloma
91 and the genetic basis of a high-risk patient population that does not benefit from current therapies.
92

93 Results

94 Cohort Description

95 The CoMMpass cohort includes 1143 patients from 84 clinical sites located in the United
96 States, Canada, Spain, and Italy. The demographic and clinical parameters of the cohort at
97 diagnosis adhere to expected distributions (Tables 1 & S1). Median age at diagnosis was 63
98 (range 27-93) with an expected over representation of males (60.4%) versus females (39.6%).
99 This cohort is largely composed of patients from the United States and, unlike most clinical trials,
100 the distribution of self-reported ancestry reflects US Census Bureau statistics with 80.6%
101 Caucasian, 17.5% Black, and 1.9% Asian. Baseline prognostication of patients with the
102 international staging system (ISS) identified 35.1% ISSI, 35.1% ISSII, and 27.2% ISSIII¹⁴.

103
104 Due to the highly variable cytogenetic panels used at individual sites, we defined the
105 phenotype of each patient from WGS data. This identified 57.2% HRD and 42.8% NHRD patients
106 (based on the detection of a whole chromosome gain on at least two classic hyperdiploid
107 chromosomes: 3, 5, 7, 9, 11, 15, 19, and 21), 24.3% patients with del(1p22), 35.2% with
108 gain(1q21), 52.0% with del(13q14), and 12.5% with del(17p13). We identified translocations
109 involving common target genes from any of the three immunoglobulin loci occurring at the
110 following frequencies: 20.0% CCND1; 1.2% CCND2; 1.8% CCND3; 4.0% MAF; 0.7% MAFA;
111 1.3% MAFB; 14.3% MYC; and 12.8% WHSC1. Of these events, 83.0% involved the IgH locus
112 while 5.3% and 11.7% involved the kappa and lambda IgL loci, respectively.

113
114 Irrespective of treatment, the median progression-free survival (PFS) of the cohort was 36
115 months and median overall survival (OS) of the cohort was 74 months (Figure S1.1A-B). Median
116 OS for ISSIII patients was 54 months, while median OS for ISSI and ISSII patients could not be
117 confidently predicted (Figure S1.1C-D). Patients with at least one high-risk cytogenetic feature
118 had worse OS outcomes, even with uniform utilization of novel agents (Figure S1.2)¹⁵.

119 Integrated Analysis for Gain and Loss of Function Genes

120 To comprehensively identify loss-of-function (LOF) and gain-of-function (GOF) events in
121 myeloma patients, an integrated model was developed to overcome the limitations of analyzing
122 any one data type by combining measurements from WES, WGS, and RNAseq to assign a
123 functional state to each gene. In the LOF model, a single event in a gene was designated a partial
124 LOF, whereas genes with two or more events were designated a complete LOF. At diagnosis,
125 592 patients had all three sequencing assays performed and were included in the analysis (Figure
126 S1.3, Table S6). We identified at least one partial LOF in all patients, a complete LOF event in
127 92% of patients, and 70 genes where a complete LOF event was identified in ≥ 5 patients (Figure
128 1A). Complete LOF was observed in 12 genes in $>2\%$ of the cohort, including TRAF3 (10.1%),
129 DIS3 (6.9%), FAM46C (5.1%), CYLD (4.7%), TP53 (4.1%), MAX (3.5%), RB1 (3.2%), WWOX
130 (3.2%), HUWE1 (2.7%), PVT1 (2.5%), CDC42BPB (2.0%), and MAGEC1 (2.0%). However,
131 CDC42BPB is unlikely a tumor suppressor gene in myeloma as it is in a contiguous gene region
132 on chr14 with TRAF3, which was previously shown to be the target of bi-allelic loss in this region¹⁶.
133

134 The target gene(s) of chr13 loss continues to be controversial. WGS data detected 13q14
135 deletion in 52.0% of patients, while LOF analysis identified 26.5% had complete LOF of one or
136 more genes on chr13. The commonly assumed target, RB1, showed complete LOF in 3.2% while
137 DIS3 complete LOF was detected in 6.9%, however, a striking number of additional genes were
138 independently knocked-out in myeloma (Figure 1B). Two contiguous gene regions with complete
139 LOF were identified: the first comprising MPHOSPH8 (1.4%), PSPC1 (1.5%), ZMYM5 (1.4%),
140 ZMYM2 (1.0%); and the second comprising TGDS (1.9%) and GPR180 (0.8%) where the minimal
141 region of deletion and LOF frequency suggest the targets are PSPC1 and TGDS, respectively.
142 Additional complete LOF events were identified targeting LATS2 (1.4%), BRCA2 (1.2%), PARP4
143 (1.0%), MYCBP2 (1.0%), TPP2 (1%), CDK8 (0.8%), TSC22D1 (0.8%), and ARHGEF7 (0.8%).
144 These results highlight that monosomy 13 is associated with multiple independent gene
145 inactivation events.
146

147 The GOF analysis identified an event in 92% of patients at diagnosis and 27 genes where
148 a GOF event was identified in 5 or more patients (Figure 1C). There were 7 genes in which a
149 GOF event was identified in greater than 2% of the cohort, including KRAS (23.6%), NRAS
150 (21.6%), WHSC1 (10.3%), BRAF (7.1%), FGFR3 (4.9%), HIST1H1E (3.2%), and EGR1 (2.5%).
151 A number of patients had activating mutations described in other cancers including KRAS^{G12D}

152 (3.2%), KRAS^{G12V} (1.5%), NRAS^{Q61R} (7.4%), NRAS^{Q61K} (5.9%), NRAS^{Q13R} (2.4%), NRAS^{Q61H}
153 (2.0%), NRAS^{Q13D} (0.8%), and BRAF^{V600E} (3.9%).

154 Identification of Copy Number Subtypes of Multiple Myeloma

155 To discover potential underlying phenotypes of myeloma beyond the known dichotomy of
156 HRD and NHRD karyotypes, unsupervised consensus clustering was performed on CN data from
157 871 patients. Three independent replicates identified eight subtypes as the optimal solution
158 (Figure S2.1). These trials were highly consistent, with deviations between any two replicates
159 occurring in <1% of patients (Figures S2.2 & S2.3).

160

161 The CN subtypes consisted of five HRD and three NHRD subtypes and were annotated
162 based on defining features (Figure 2A). The HRD, classic subtype had gains of classic HRD
163 chromosomes, and the remaining HRD subtypes were annotated based on deviations from this
164 phenotype. The HRD, ++15 subtype exhibited tetrasomy of chr15 (Figure S2.4) while two
165 subtypes (HRD, diploid 7 and HRD, diploid 3, 7) were defined by the absence of chr7 and chr3
166 trisomies. Finally, the complex HRD, +1q, diploid 11, -13 subtype lacked chr11 trisomy but
167 harbored gain of chr1q and loss of chr13. Of the NHRD subtypes, the diploid subtype was mostly
168 devoid of CN events and highly associated with translocations targeting a D-type cyclin (71.3%).
169 The remaining two NHRD groups were strongly associated with canonical immunoglobulin
170 translocations (71.1%) and were defined by chr13 loss. The -13 subtype contained a
171 subpopulation of patients with chr14 loss, while the +1q, -13 subtype had gains of 1q.

172

173 In the cohort, there was no difference in outcomes between HRD and NHRD patients
174 (Figure S2.5). However, the HRD and NHRD subtypes with both 1q gain and chr13 loss had
175 inferior OS outcomes when compared to patients in other CN subtypes (Figure 2B), suggesting
176 that HRD patients should not universally be considered as a group with good outcomes.
177 Combining these two subtypes identified a group with inferior outcomes as compared to patients
178 with other genetic backgrounds (Figure 2C, HR=1.928, 95% CI=1.435-2.59, p<0.001). The
179 observation that NHRD patients in the +1q, -13 subtype exhibited poor OS outcomes as compared
180 to NHRD patients in the -13 subtype suggested 1q gain, rather than 13q loss, is the predictor of
181 poor outcome, which was confirmed in a cox proportional hazard model examining the
182 contribution of 13q14 and 1q21 CN on OS outcomes across the cohort (Figure S2.6).

183 RNA Subtypes of Multiple Myeloma

184 Consensus clustering was performed on RNA sequencing results from 714 baseline
185 samples to identify subtypes of myeloma defined by gene expression similarities. The number of
186 clusters that best represented the data was 12 across three independent replicates, for which two
187 were identical and a third had 20 (2.8%) patients assigned to different classes (Figure S3.1-3.3).
188 Many of the observed subtypes were associated with known immunoglobulin translocations and
189 CN states (Figures 3A & S3.4) and there were clear relationships with subtypes identified in
190 previous studies (Figures S3.5 & S3.6)^{17,18}. Four subtypes were identified across all studies,
191 including: MS, characterized by t(4;14) patients; MAF, characterized by t(14;16) patients; CD1
192 characterized by t(11;14) patients; and PR, characterized by patients with a high proliferation
193 index. To maintain consistency across studies we used subtype names from previous studies
194 when appropriate but otherwise assigned names based on common molecular features.

195

196 The MS subtype comprised 10.6% of patients for whom a t(4;14)-WHSC1 was detected
197 in 62/67 (92.5%) by WGS. In one patient, a t(2;4) involving the kappa locus was detected. Two of
198 the four remaining patients had detected fusion transcripts between WHSC1 and the highly
199 expressed genes FUT8 or CXCR4. The MAF subtype included 6.4% of patients in which we
200 detected immunoglobulin rearrangements in 38/41 (92.7%) of patients (27 t(14;16)-MAF, 4
201 t(8;14)-MAFA, 6 t(14;20)-MAFB and 1 t(20;22)-MAFB). All three patients with undetectable
202 immunoglobulin translocations had high expression of a MAF family gene and in two patients the
203 mechanism was identified. One had a t(1;16) juxtaposing the FAM46C super-enhancer with
204 MAF^{19,20}. Another had an atypical insertion of a class-switch circle telomeric of MAF. One patient
205 had both a t(14;16)-MAF and t(4;14)-WHSC1 yet strongly associated with the MAF subtype,
206 suggesting the MAF expression signature can overpower the MS signature. MAF immunoglobulin
207 translocations were associated with higher mutation load²¹ and in this cohort 8/10 patients with
208 high tumor mutation burden (>10 mut/mb) were in the MAF subtype, and could qualify to receive
209 a checkpoint inhibitor.

210

211 Three subtypes were highly associated with overexpression of a D-type cyclin caused by
212 t(11;14)-CCND1, t(12;14)-CCND2, or t(6;14)-CCND3 (Figure S3.7A-C). The CD1 subtype
213 included 4.3% of patients 24/25 (96%) had a detected D-type cyclin targeting translocation. The
214 remaining patient had a t(9;14) which linked the immunoglobulin heavy chain locus with the B-cell
215 master regulator PAX5 resulting in its overexpression (Figure S3.7D). Unlike previous studies that
216 identified a single CD2 subtype, we identified two related subtypes designated as CD2a and

217 CD2b. The CD2a subtype comprised 7.8% patients of whom 40/47 (85.1%) had a detected D-
218 type cyclin IgH translocation. The CD2b subtype included 8.0% patients of whom 51/56 (91.1%)
219 had a detected D-type cyclin targeting translocation. Both the CD2a and CD2b subtypes were
220 associated with cell surface expression of CD20, which is largely absent in other RNA subtypes,
221 including CD1.

222

223 The PR subtype contained 7.1% of patients with an admixture of classic genetic subtypes
224 and very poor clinical outcome, with a median OS of 21 months (Figure 3B-C). High proliferation
225 index scores were also concentrated in this subtype (Figure S3.8). Clearly, current treatment
226 regimens are ineffective for these patients. We compared the expression of current checkpoint
227 and immunotherapy targets in non-PR versus PR patients and observed that three of five
228 checkpoint targets showed no difference in expression between the two groups, whereas all
229 immunotherapy targets showed either no difference in expression (1 target), or had a higher
230 median expression (4 targets) in PR patients (Figure 3D-E).

231

232 A subtype representing 11.1% of patients most closely resembled the previously defined
233 low bone (LB) subtype¹⁷(Figure S3.5), however, there was no noticeable decrease in bone lesions
234 (Figure S3.9). This subtype comprised an admixture of 59.2% HRD and 40.8% NHRD patients,
235 but 74.0% had a gain of chr1q with 26.0% having ≥ 4 copies and was thus termed the 1q gain
236 subtype.

237

238 Four of the RNA subtypes were associated with a HRD karyotype (Figure S3.4) and either
239 did not uniquely associate with a subtype from a previous study, or the original name could not
240 be justified. Two of the HRD subtypes associated closely with the HY (hyperdiploid) subtype
241 identified previously but differed due to an enrichment of tetrasomy 15, observed in 58.7% and
242 60.8% of patients. Since structural events involving MYC are associated with HRD karyotypes²²,
243 we investigated the association between these two groups and MYC rearrangements. We
244 identified 37/49 (75.5%) patients versus 23/76 (30.3%) of patients had MYC rearrangements, and
245 thus these subtypes were named HRD ++15, MYC and HRD ++15 respectively. A third HRD
246 subtype comprising 8.3% of patients most closely associated with the PRL3 subtype¹⁸, however,
247 the signature was elevated in four subtypes (Figure S3.6). A MYC structural event was identified
248 in 35/49 (71.4%) of these patients and this group was also distinguished from all others except
249 PR in having a low NFkB index (Fig S3.10) and was thus named HRD MYC, low NFkB. The
250 smallest HRD group contained 4.6% of patients and was associated with the previously defined

251 NF- κ B subtype¹⁸, however, no clear association existed with the NF- κ B index used to define the
252 subtype (Figures S3.6 & S3.10). One of the predictors of this RNA subtype was overexpression
253 of NINJ1 (Table S7, Figure S3.11A), which inhibits translation of TP53²³. TP53 was also found to
254 be underexpressed, exhibiting the lowest median expression in this subtype as compared to all
255 other RNA subtypes (Figure S3.11B). Taken together, this subtype was termed HRD, low TP53.
256

257 The final subtype contained 12.2% of patients and strongly correlated with the previously
258 defined myeloid group¹⁸. Analysis of multiple data types indicated a bias to lower purity samples,
259 and was thus termed Low purity (Figures S3.6 & S3.12).

260 Clinical and Molecular Associations with RNA Subtypes

261 Some RNA subtypes were strongly associated with specific molecular events while others
262 seemed to be admixtures with a common transcriptional phenotype. To identify additional defining
263 features of each RNA subtype, we tested for significant associations between clinical and
264 molecular features, including complete LOF and GOF events. Overall, 21 genes with complete
265 LOF or GOF were identified to have a significant association with one or more RNA subtype
266 (Figure 4). As expected, GOF was detected in the translocation target genes associated with the
267 MAF, MS, and the CD subtypes. Although loss of one WWOX allele is expected in t(14;16) we
268 frequently detected complete LOF of WWOX ($p < 0.001$) supporting a possible role of WWOX in
269 myeloma. Both the MS and 1q gain subtypes were diminished for NRAS GOF, and the latter was
270 enriched for TRAF3 LOF. The CD2a subtype was enriched for GOF events in NRAS ($p < 0.005$)
271 and IRF4 ($p < 0.005$) while the CD2b subtype was enriched for GOF events in IRF4 ($p < 0.005$) and
272 EFTUD2 ($p < 0.01$) representing potential subtype-specific therapeutic targets. In general, the HRD
273 RNA subtypes were not enriched for any GOF or LOF events aside from the HRD, ++15, MYC
274 subtype, which was enriched for LOF events in FAM46C ($p < 0.001$).

275
276 The PR subtype was enriched for LOF of RB1 ($p < 0.001$) and MAX ($p < 0.01$), gain(1q21)
277 ($p < 0.001$), del(13q14) ($p < 0.001$), and ISSIII patients ($p < 0.001$). Interestingly, 50% of PR patients
278 were ISSIII while 22% and 28% were ISSI and ISSII, highlighting that ISS underestimates disease
279 severity in half of these high-risk patients. Different mechanisms of complete loss of RB1 were
280 observed but typically involved a one copy deletion of 13q coupled with a second molecular event
281 (Figure S4.1). Identifying LOF of RB1 and MAX provides the first defining genetic features of the
282 high-risk PR phenotype.

283 Transition to PR at Progression and Link with G1/S Checkpoint

284 To apply the RNA subtype classification to the serial samples, we developed a predictive
285 model that outputs the class probability associated with each of the 12 subtypes and assigned
286 each progression sample the subtype with the highest class probability. Overall 71 patients were
287 assigned a subtype at two or more timepoints, with 55 patients assigned a subtype other than
288 Low purity for at least two time points. At diagnosis, 5 serial patients were classified as Low purity,
289 however, at progression they all had a subtype other than Low purity, further supporting that this
290 phenotype was driven by relative sample purity rather than a distinct disease signature (Figure
291 S5.1). For each patient, subtype assignments were compared across visits. Although most
292 patients remained in the same subtype throughout their disease course 13/51 (25.5%) non-PR
293 patients transitioned into the PR subtype at progression (Figure 5A). Regardless of original
294 subtype, patients that transitioned to the PR subtype rapidly succumbed to their disease (Figure
295 5B), with a median OS after the detected progression of 88 days (Figure S5.2), and had inferior
296 outcomes compared to other patients who also progressed (Figure 5C).

297
298 To identify molecular events potentially driving the transition of patients to the PR subtype,
299 gene functional status was compared at the PR and prior non-PR time point. Molecular data was
300 available for comparison at both time points for 9/13 patients that transitioned to PR. Despite the
301 prevalence at baseline, none of the patients transitioning to PR acquired complete LOF of RB1.
302 However, three patients (33%) had complete LOF of a cyclin-dependent kinase inhibitor at
303 progression. Two patients had complete LOF of CDKN2C at progression due to homozygous
304 deletions (Figure S5.3). One patient acquired two independent deletions at progression while the
305 other had one clonal deletion at diagnosis and the second became clonal, increasing from 26%
306 to 100%, at progression, suggesting this aggressive clone existed even before any treatment was
307 received. One patient acquired complete loss of CDKN1B at progression from the combination of
308 a pre-existing deletion and a clonal frameshift mutation detected only at progression, suggesting
309 the mutation either existed at a frequency below our limit of detection, or was acquired (Figure
310 S5.4). Similar to the baseline observations, there are multiple genetic defects in G1/S checkpoint
311 genes that can result in the PR phenotype.

312 Discussion

313 The MMRF CoMMpass study represents the largest sequencing study of multiple
314 myeloma patients undertaken to date based on the number of enrolled patients and the total
315 number of sequencing assays performed. The cohort has facilitated the identification of distinct
316 CN and expression subtypes of myeloma, as well as both recurrent and rare molecular events
317 that occur at frequencies that would not be detected in smaller patient cohorts.

318
319 A diverse array of genetic events can contribute to the development or progression of
320 cancer, with individual genes often being affected by multiple types of alterations, however, these
321 diverse processes are generally summarized in isolation. To accurately summarize the frequency
322 of these changes, we integrated seven different data formats extracted from WES, WGS, and
323 RNAseq data and identified 70 LOF and 27 GOF genes occurring in five or more patients. This
324 approach permitted the identification of genes that are often under-represented by a single
325 technology including complete LOF in RB1, CDKN2C, and WWOX, and allowed each gene to be
326 assigned a biologically relevant phenotype of functional, partial loss, or complete loss.
327 Differentiating between partial and complete LOF is pertinent for accurate identification of high-
328 risk patient populations. In TP53, solitary deletions or mutations have been associated with poor
329 prognosis, however, only patients with complete LOF of TP53 have poor outcomes, suggesting
330 only one third of patients with del(17p13) identified by clinical cytogenetic assays are true high-
331 risk patients^{12,24–26}. Finally, there is a long standing interest in determining the gene associated
332 with monosomy 13 in myeloma. Our analysis identified recurrent complete loss events in RB1
333 and DIS3, but also identified independent complete loss events in PSPC1, TGDS, LATS2,
334 BRCA2, PARP4, MYCBP2, TPP2, CDK8, TSC22D1, and ARHGEF7 suggesting multiple genes
335 on chr13 can independently contribute to myelomagenesis.

336
337 Few studies have performed CN clustering and, due to their limited sample size, have
338 barely resolved distinct CN subtypes beyond the classic HRD and NHRD phenotypes³. We
339 identified eight distinct CN subtypes, including five HRD, and three NHRD subtypes. Although
340 previous studies have shown HRD patients have favorable outcomes compared to NHRD
341 patients, in CoMMpass there was no difference in OS or PFS outcomes. This large cohort analysis
342 revealed a number of seemingly inter-related events such as 1q gains and monosomy 13; in HRD
343 patients lacking trisomy 3, a concomitant absence of trisomy 7; and groups with a classic HRD
344 phenotype defined by trisomy or tetrasomy 15. Interestingly, there were independent HRD and

345 NHRD subtypes with 1q gain and chr13 loss with the HRD group lacking the classic trisomy 11,
346 suggesting the combination of these events can phenocopy the benefits of trisomy 11. Although
347 patients with 1q gain and 13 loss represent poor outcome CN subtypes, the median OS of these
348 patients was just under 5 years, comparable to the 4.5 year OS associated with R-ISSIII which
349 includes patients with high-risk clinical and cytogenetic features¹⁵. Taken together, this highlights
350 that CN features alone are insufficient to distinguish the subset of ultra high-risk myeloma patients
351 with early OS events.

352

353 Previous studies clustering myeloma gene expression data have identified 8 to 10 unique
354 subtypes, many of which were consistent among studies including the MS, MAF, CD1, CD2, and
355 PR subtypes^{17,18}. Our consensus clustering of RNAseq data from CoMMpass identified 12 unique
356 RNA subtypes. In this study, some previously identified subtypes were further broken down, while
357 others were renamed to better reflect the underlying biology, which was aided by incorporating
358 the multitude of data types in this study. For example, previous studies identified a single CD2
359 group, however, we identified two CD2 groups, designated CD2a and CD2b, which were
360 associated with IRF4 mutations, PAX5 expression, and CD20 surface expression. Several studies
361 have sought to identify treatment strategies for CD20 positive patients, but based on the
362 distribution within CD2a and CD2b it may be pertinent to consider that these patients originate
363 from two unique populations. In the context of precision medicine, the strong link between t(11;14)
364 and the CD subtypes leads to question whether one of the CD subtypes better predicts response
365 to venetoclax than t(11;14)²⁷.

366

367 The PR RNA subtype defined a group of patients with extremely poor OS, high proliferative
368 index scores, and nearly ubiquitous 1q gains. This group has remained controversial because of
369 the competing supervised patient segregation TC classification models^{28,29}, that argue to group
370 patients by defined genetic features, however, given the mixture of genetic backgrounds observed
371 in this subtype, it would be difficult to identify these high-risk patients based on genomic features
372 alone. Through our integrated analysis we identified LOF of RB1 or MAX as common genetic
373 events in baseline PR patients, which provides the first genomic link to this gene expression
374 phenotype. Loss of MAX was recently associated with transformation and increased proliferation
375 in small cell lung cancer³⁰ and thus, LOF of RB1 or MAX likely contributes to the highly proliferative
376 phenotype observed in PR patients. PR patients exhibited similar or lower median expression of
377 checkpoint targets but higher expression of most immunotherapy targets when compared to non-
378 PR patients suggesting that immunotherapies may represent a viable therapeutic option in these

379 high-risk patients, and highlighting the importance of identifying these patients in future clinical
380 studies.

381
382 There was also a strong tendency for patients to transition to the PR subtype at
383 progression, with 25.5% of serial patients in a non-PR subtype at diagnosis transitioning to PR.
384 Patients that transitioned to the PR subtype had extremely poor outcomes after the transition,
385 with a median survival of less than three months after their progression visit. An acquired complete
386 loss of a cyclin-dependent kinase inhibitor, such as CDKN2C or CDKN1C, was observed in 33%
387 of patients transitioning to the PR subtype at progression suggesting that transition to the PR
388 phenotype at progression is highly associated with genetic events that further disrupt cell cycle
389 control.

390
391 These findings demonstrate that advanced molecular diagnostics such as WGS and
392 RNAseq are better predictors of disease behavior than current staging systems based on clinical
393 laboratory, conventional cytogenetic, and FISH data. These assays also identify therapeutic
394 targets including RAS, BRAF^{V600E}, and FGFR3 mutations, that are actionable with agents already
395 approved for other cancer indications. In fact CoMMpass findings drove the development of a
396 number of clinical trials, notably, MMRC-085 Myeloma-Developing Regimens Using Genomics
397 (MyDRUG, NCT03732703), an umbrella trial using targeted exome and transcriptome
398 sequencing to stratify subjects into sub-protocols with approved targeted agents. These findings
399 favor adoption of advanced molecular diagnostics into the routine care of myeloma, especially as
400 the breadth of clinical impact broadens and costs decrease.

401
402 Comprehensive molecular analyses of the baseline CoMMpass cohort has permitted a
403 more thorough understanding of the genetic diversity and subtypes of the disease. This approach
404 clearly defined the primary molecular features driving different subtypes of multiple myeloma and
405 identified high-risk patients at both diagnosis and progression. Innovative clinical trials targeting
406 this high-risk population are needed given the current poor outcomes with therapies that are
407 otherwise highly effective in other subtypes. Given that patients frequently transition to the high-
408 risk PR subtype at progression, it will be important to know the percentage of PR patients in
409 clinical trial populations, particularly in the relapse/refractory setting, to understand if the arms are
410 balanced and if there is a difference in response between these groups. The identification of
411 unique subtypes and the frequency of target gene dysregulation, via our integrated analysis,

412 provides a solid foundation to prioritize targets for precision medicine approaches in multiple
413 myeloma.

414 Methods

415 Sample Collection and Biobanking

416 All samples and clinical data analyzed in this study were collected as of interim analysis
417 14 from patients who consented to participate in the Relating Clinical Outcomes in Multiple
418 Myeloma to Personal Assessment of Genetic Profile (CoMMpass) Study (NCT01454297),
419 sponsored by the Multiple Myeloma Research Foundation (MMRF). The MMRF CoMMpass study
420 accrued patients from clinical sites in Canada, Italy, Spain and the United States. All patient
421 samples were shipped to one of three biobanking operations: Van Andel Research Institute
422 (VARI) in Grand Rapids, Michigan for all samples collected in Canada or the United States;
423 University Hospital of Salamanca for samples collected in Spain; or University of Torino for
424 samples collected in Italy. In North America and Spain, K2-EDTA tubes were used for collection
425 of peripheral blood (PB) and sodium heparin tubes were used for the collection of bone marrow
426 (BM) aspirates. These samples were shipped to their respective biobank using CoMMpass study
427 kits that maintained samples at 7-12°C. In Italy, clinical sites participating in the FORTE clinical
428 trial (NCT02203643) collected BM and PB sample aspirates in sodium citrate vacutainers.
429 Samples collected at sites in Italy were shipped at ambient temperature to the biorepository site.

430
431 At VARI, the received BM and PB specimens were first quality controlled by flow cytometry
432 to determine the percentage of plasma cells (PCs) in the PB and BM specimens. Patients were
433 only included in the study when the submitted BM contained at least 1% PCs. If the PB showed
434 less than 1% circulating PCs, white blood cells were used as the constitutional DNA source,
435 however, when >1% circulating PCs were observed enriched CD3 positive T-cells were used.
436 Whole BM PC enrichment, or PB PC enrichment when >5% circulating PCs were detected, was
437 performed using the Miltenyi autoMACS Pro Separator using anti-CD138 microbeads. The purity
438 of the enriched samples was assessed using a 3-color slide-based immunofluorescence assay
439 that identified cells with DAPI and the presence or absence of kappa or lambda light chains.
440 Clinically eligible baseline patients with tumor samples with greater than 250,000 cells recovered
441 after CD138 enrichment and monoclonal purity greater than or equal to 80% moved forward for
442 nucleic acid extraction. For progression samples the cell recovery requirement was 200,000 cells.
443 To minimize nucleic acid isolation failures, the first 750,000 cells were used exclusively for DNA
444 isolation. When more than 750,000 cells were recovered, the sample was split 50/50 for DNA and
445 RNA isolations. When more than 4 million cells were recovered, multiple aliquots were stored for

446 future use. Cells destined for DNA isolation were stored as snap frozen pellets, while samples for
447 RNA extraction were lysed in QIAzol before long-term storage at -80°C.

448

449 Samples from clinical sites in Spain were collected and shipped using the provided
450 CoMMpass collection kits. Red blood cells were removed from the PB and BM specimens using
451 a red cell lysis buffer and, following a PBS wash, the remaining white blood cells were counted.
452 After red cell lysis, the isolated cells were quality controlled using flow cytometry to determine the
453 percentage of PCs in the PB and BM specimens. If the PB showed less than 1% circulating PCs,
454 white blood cells were used as the constitutional DNA source; however, when >1% circulating
455 PCs were observed, enriched CD3 positive T-cells were used. The isolated PB cells (1-5 million
456 cells) were snap frozen as dry pellets for constitutional DNA isolation. The isolated BM cells were
457 stained with anti-CD138 microbeads and PCs were enriched using a Miltenyi autoMACS Pro
458 Separator. The enriched CD138⁺ cells were stored as snap frozen dry pellets and, when possible,
459 a separate aliquot was lysed with QIAzol and stored at -80°C until shipped on dry ice to VARI for
460 isolation. The purity of the CD138 enriched PC fractions was determined using flow cytometry
461 with antibodies against CD38, CD138, and CD45.

462

463 Samples collected in Italy were treated with red cell lysis buffer. After washing the
464 remaining WBC with PBS, the cells were counted. The isolated PB cells (1-5 million cells) were
465 snap frozen as dry pellets for constitutional DNA isolation. The BM WBC were stained with anti-
466 CD38 magnetic beads and PCs were enriched using a Miltenyi autoMACS Pro Separator. After
467 sorting, the purity of the enriched specimens was assessed by flow cytometry using a fluorescent
468 anti-CD38 antibody. Cells destined for DNA isolation were stored as snap frozen pellets, while
469 samples for RNA extraction were lysed in QIAzol before long-term storage at -80°C.

470 Flow Cytometry Phenotyping and Quality Control Process

471 All samples received at VARI were tested by flow cytometry to phenotype and quality
472 control the received specimens. The antigens and corresponding commercial antibodies used in
473 the flow cytometry assays are as follows: CD38 (BD Biosciences, 340677), CD45/PTPRC (BD
474 Biosciences, 340665), CD138/SDC1 (BD Biosciences, 347205), CD319/SLAMF7
475 (Invitrogen/eBioscience, 12-2229-42), CD13/ANPEP (BD Biosciences, 340686), CD19 (BD
476 Biosciences, 340720), CD20/MS4A1 (BD Biosciences, 346581), CD27 (BD Biosciences,
477 654665), CD28 (BD Biosciences, 348047), CD33 (BD Biosciences, 340679), CD52 (Life
478 Technologies, MHCD5204), CD56/NCAM1 (BD Biosciences, 340724), CD117/KIT (BD

479 Biosciences, 340867), FGFR3/CD333 (R&D Systems, FAB766P), Kappa (BD Biosciences,
480 643774), and Lambda (Life Technologies, MH10614). Flow panels performed included CD38 x
481 CD45 x CD138 x CD56 (initial BM & PB screening panel 1); CD38 x CD45 x CD138 x CD319
482 (updated screening panel 1 after the introduction of daratumumab); CD38 x CD45 x cytoplasmic
483 Kappa x cytoplasmic lambda (BM & PB screening panel 2); CD38 x CD45 x CD138 x (either
484 CD13, CD19, CD20, CD27, CD28, CD33, CD52, CD117, FGFR3); and propidium iodide stained
485 nuclei to determine the DNA content of each tumor.

486 Nucleic Acid Isolation

487 All nucleic acid isolations were performed at VARI. DNA was extracted from the dry cell
488 pellets with the Qiagen Genra Puregene Tissue Kit (Qiagen, 158667) with isolated DNA
489 suspended in Qiagen buffer ATE, and stored at -20°C. DNA was extracted from PB samples using
490 the Qiagen QIA Symphony, which uses magnetic beads for automated sample processing. Blood
491 tubes were either processed immediately upon receipt, or frozen at -20°C and processed in
492 batches. QIA Symphony extractions were performed using the DSP DNA Midi Kit (Qiagen,
493 937255). DNA was eluted in Qiagen buffer ATE and stored at -20°C. DNA was quantified by
494 Nanodrop spectrophotometric analysis, as well as by fluorescence using the Qubit 2.0 to
495 determine dsDNA content. Sample quality was determined by agarose gel or Agilent TapeStation
496 Genomic ScreenTape. Samples with at least 250 ng of dsDNA were submitted to TGen for
497 analysis.

498
499 Tumor cells designated for RNA extraction were dissolved in QIAzol Lysis Reagent
500 (Qiagen, 79306), stored at -80°C, and extracted with the Qiagen RNeasy Plus Universal Mini Kit
501 (Qiagen, 73404). RNA was eluted in nuclease-free water and stored at -80°C. RNA was quantified
502 by Nanodrop spectrophotometric analysis and RNA quality was evaluated using the Agilent
503 Bioanalyzer 2100. Samples with a RIN ≥ 6.0 and at least 200 ng of RNA were submitted to TGen
504 for analysis.

505 RNA Sequencing (RNAseq) Library Preparation

506 All RNA sequencing libraries were constructed using the Illumina TruSeq RNA library kit
507 following the manufacturer's recommendations. Over the course of the project the primary target
508 input changed from 2000 ng to 500 ng. When 500 ng of RNA was not available, samples were
509 processed using a lower input of 250 ng or 150 ng. In all cases, the input quantity was based on

510 nanodrop quantification. RNA quality was confirmed at TGen to have a required RIN ≥ 7 using an
511 Agilent Bioanalyzer with the RNA 6000 Nano Kit (Agilent, 5067-1511) or eRIN ≥ 6 on an Agilent
512 TapeStation using the RNA ScreenTape assay (Agilent, 5067-5576, 5067-5577). Libraries were
513 prepared following the manufacturer's recommendation for the TruSeq RNA Library Prep Kit v2
514 (Illumina, RS-122-2001), unless otherwise noted. Eight cycles of PCR amplification for 2000 ng
515 and 500 ng input, 9 cycles for 250 ng input, and 10 cycles for 150 ng input were performed. Final
516 libraries were quantified using the Qubit dsDNA HS Assay Kit (Invitrogen, Q32854) and assessed
517 on the Agilent TapeStation using the High Sensitivity D1000 ScreenTape assay (Agilent, 5067-
518 5584, 5067-5585).

519 Long Insert Whole Genome Sequencing (LI-WGS) Library Preparation

520 LI-WGS libraries were initially generated using the Illumina TruSeqDNA Whole Genome
521 kit (TSWGL). Fragmentation of 600-1100 ng of DNA was performed on a Covaris E210 Focused
522 Ultrasonicator with the following parameters: duty cycle = 2, peak power = 6, cycles per burst =
523 200, and time = 20 sec. Libraries were prepared according to the standard Illumina protocol,
524 however, the AMPure bead ratios were adjusted to either 1:1 or 1:0.8. After ligation clean up,
525 samples were run on a 1.5% agarose gel and Xtracta gel extractors (USA Scientific, 5454-2500)
526 were used to extract library molecules at approximately 1.3, 1.0, and 0.8 kb. Size-selected
527 samples were processed with Freeze 'N Squeeze DNA Gel Extraction Spin Columns (Bio-Rad,
528 732-6165). Purified 1 kb samples were concentrated using AMPure XP beads and amplified with
529 10 PCR cycles. Final libraries were run on Agilent Bioanalyzer DNA 12000 chips (Agilent, 5067-
530 1508 & 6067-1509). If a patient's tumor and constitutional samples were not within ~100 bp of
531 each other, alternate excised samples (1.3kb or 0.8kb) were processed to generate a better
532 match.

533
534 For samples processed using the KAPA HTP/LTP Library Preparation Kit (Roche,
535 KK8234) and off-bead protocol (KAWGL), 200-1100 ng of DNA was fragmented on the Covaris
536 E210 (see parameters above) or E220 with the following parameters: duty cycle = 4%, initial /
537 peak power = 170 W, cycles per burst = 200, and time = 20 sec. Libraries were prepared according
538 to the standard Kapa protocol, however, the AMPure bead ratio was adjusted to either 1:1 or
539 1:0.8. Libraries were amplified for two cycles before size selection to linearize the fragments and
540 five cycles after size selection using KAPA HiFi HotStart ReadyMix (Roche, KK2602)

541

542 For samples processed using the KAPA HyperPrep kit (Roche, KK8505) (KHWGL), 200
543 ng of DNA was fragmented on the E210 or E220 Covaris with the parameters defined above.
544 Libraries were prepared according to the manufacturer's protocol, however the post ligation
545 AMPure bead ratio was adjusted to 1:0.8. Molecules between 950-1050 bp were either extracted
546 automatically from a Sage Science Pippin Prep 1.5% gel (Sage Science, CSD1510) or hand
547 punched from a 1.5% agarose gel. One cycle of PCR amplification pre size selection, followed by
548 6 cycles of amplification post size selection, was performed using KAPA HiFi HotStart ReadyMix.

549 Whole Exome Sequencing (WES) Library Preparation

550 Initially samples were processed using the Illumina TruSeq Exome Library Prep Kit
551 (TSE61) until the product was discontinued. Genome libraries were created using 1100 ng of DNA
552 as input for fragmentation on the E210 Covaris using the following settings: duty cycle = 10,
553 intensity = 5, cycles per burst = 200, time = 120 seconds. Individual DNA samples were mixed
554 with TElowE (10 mM Tris-HCL, 0.1 mM EDTA) to a final volume of 55 µl in a covaris microTUBE
555 plate. Six libraries were pooled together before enrichment following the manufacturer's protocol.
556

557 Subsequently, samples were processed using the KAPA HTP/LTP Library Preparation Kit
558 using the off-bead protocol and 8-plex pooled enrichment was performed using Agilent SureSelect
559 XT2 Human All Exon V5+UTR baits (Agilent, G9661B) (KAS5U). Genome library preparation was
560 performed according to the manufacturer's protocol with 500 ng of input DNA fragmented to an
561 average target size of 160 bp using a Covaris E220 focused-ultrasonicator with the following
562 settings: duty cycle 10%; peak power 175; cycles per burst 200; time 300; power mode "frequency
563 sweeping"; bath temperature 7°C. Five cycles of PCR amplification was performed pre-capture
564 using KAPA HiFi HotStart ReadyMix, and the resulting libraries were quantified using either the
565 Agilent Bioanalyzer 1000 chip (Agilent, #5067-1504) or the Agilent Tapestation D1000 kit
566 (Agilent, 5067-5582, 5067-5583). Eight libraries were pooled at 187.5 ng each before capture
567 following the Agilent XT2 protocol. The enriched library pool was amplified for 8 cycles using
568 KAPA HiFi HotStart ReadyMix. Final libraries were quantified using the Agilent Bioanalyzer High
569 Sensitivity DNA Kit (Agilent, #5067-4626) or Agilent Tapestation HSD1000 (Agilent, #5067- 5584
570 & 5067- 5585) and Qubit High-Sensitivity assay (Invitrogen, #Q32854). Samples processed using
571 the KAPA HTP/LTP Library Preparation Kit and the on-bead protocol followed by 8-plex pooled
572 enrichment performed using Agilent SureSelect XT2 Human All Exon V5+UTR baits (KBS5U)
573 were performed according to the manufacturer's protocol with 500 ng of input DNA as described
574 above for the KAS5U protocol.

575

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

584 Illumina Sequencing

585 Sequencing was performed on an Illumina HiSeq2000 or HiSeq2500 at TGen using
586 Illumina HiSeq v3 or v4 chemistry. Diluted library pools with 1% PhiX control libraries were
587 clustered on Illumina cBOT instruments as recommended by the manufacturer. In all cases,
588 sequencing assays used a paired-end sequencing format with at least 82x82 nucleotide reads.
589 The majority of RNA sequencing libraries, which all had 6 bp sample indexes, were sequenced
590 using paired-end 83x83 reads. Exome libraries with 6 bp sample indexes were sequenced using
591 paired-end 83x83 reads, while those with 8 bp sample indexes were sequenced using 82x82
592 reads. Whole genome long-inserts were typically clustered on individual lanes, allowing a paired-
593 end 86x86 format. In all cases, raw sequencing data was extracted from the BCL files in the
594 resulting Illumina run folders using BCL2FASTQ v1.8.4 or BCL2FASTQ v2.17.1 to generate
595 industry standard FASTQ files.

596 Sequencing Data Analysis

597 Analysis of all sequencing data was performed at TGen on a high-performance computing
598 system using an internally developed analysis pipeline (Medusa Subversion,
599 <https://github.com/tgen/medusaPipe>) and the MMRF CoMMpass specific TGen05 recipe. This
600 recipe is based on the hs37d5 version of the GRCh37 reference genome used by the 1000
601 genomes project, with gene and transcript models from Ensembl v74. Additional automated
602 CoMMpass specific primary processing was also performed
603 (https://github.com/tgen/Post_Medusa_Processing). Code for the creation of the reference
604 genome and gene models used, as well as secondary and tertiary analysis methods, are available
605 on GitHub (https://github.com/tgen/MMRF_CoMMpass).

606 The paired-end fastq files generated in the LI-WGS and WES assays from each
607 sequencing lane were aligned with bwa (v0.7.8-r455). The output SAM file was converted to a
608 BAM file and sorted using SAMtools (v0.1.19-44428cd), after which base recalibration was
609 performed using GATK (3.1-1-g07a4bf8). When multiple lanes existed they were merged into a
610 single BAM file and duplicate reads were marked using Picard (v1.111(1901)), and joint indel
611 realignment was performed using GATK to produce the final BAM files used for analysis. The
612 quality of each assay was determined using multiple Picard and Samtools quality control metrics.
613 To be included in the analysis, both the tumor and constitutional sample needed to meet the
614 following criteria for genomes: physical coverage $\geq 25x$, chimera read rate $< 3\%$, and dlrs ≤ 0.2 ;
615 and for exomes: $> 90\%$ target bases at 20x coverage, and chimera read rate $< 3\%$. Somatic
616 mutations were identified using Seurat (v2.6, <https://github.com/tgen/seurat>), Strelka (v1.0.13),
617 and MuTect (v2.2-25-g2a68eab), and their outputs were combined to identify somatic events
618 called by at least two callers. The coding effect of each mutation was determined using snpEFF
619 (v4.2 (build 2015-12-05)), and additional annotations were added using snpSIFT. Somatic
620 structural abnormalities were detected using Delly (v0.7.6) to which additional filtering fields were
621 added to ensure informative read pairs spanned at least a 100 bp window on both breakends.
622 Somatic copy number (CN) abnormalities were identified with a CoMMpass specific
623 implementation of tCoNut (https://github.com/tgen/MMRF_CoMMpass).

624
625 Paired-end fastq files from the RNA sequencing assays were aligned using STAR (v2.3.1z
626 01/24/2013) and the output SAM file was converted to a BAM file and sorted using samTools
627 followed by duplicate marking with Picard. For RNAseq to be included in the analyses, we required
628 at least 50 million read-pairs (100 million reads) generated from each library, a 5' bias ratio ≥ 0.5 ,
629 and a 5'/3' bias ratio ≥ 0.75 . Gene expression estimates were determined using multiple tools.
630 Counts were extracted from the unsorted SAM file using HtSeq (v0.6.0). TPMs were estimated
631 with Salmon (0.7.2) using the fastq reads as input for quasi-alignment to a transcriptome defined
632 by the GTF gene model. To correct for the variable level of immunoglobulin transcription between
633 samples that compress the TPM values we removed plasma-cell specific transcripts including
634 immunoglobulin, mitochondrial, rRNA, and chrY genes from the final TPM calculation
635 (https://github.com/tgen/Post_Medusa_Processing). Gene fusion events were identified using the
636 TopHat-Fusion workflow in TopHat2 (2.0.8b) followed by independent cross-validation that an
637 associated genomic event existed in the matched LI-WGS assay.

638

639 The genotypes of each result file were compared using SNPsitter (v5) to ensure the
640 constitutional DNA, tumor DNA, and tumor RNA were from the same individual and that each
641 patient was uniquely represented. To ensure accurate alignment with clinical data, molecular
642 predicted gender was required to match the clinically recorded gender and, when available, the
643 clinical immunoglobulin isotype was confirmed to match the isotype defined by flow cytometry and
644 RNA sequencing. All constitutional DNA samples were manually reviewed to ensure they
645 represented a diploid genome. Potential low level cross contamination of the tumor DNA
646 specimens was determined by comparing the number of high confidence mutations detected
647 versus the percentage of those mutations which exist in dbSNP.

648 Survival Analyses

649 Survival curves were computed using the Kaplan-Meier method as implemented in R by
650 the survfit function from the survival (v3.1-8) package and plotted using the ggsurvplot function
651 from the survminer (v0.4.6) package. Pairwise comparisons of survival curves were calculated
652 using pairwise_survdiff from the survminer package. Progression-free survival estimates were
653 computed using the PFS_Censor_Flag (inv_censpfs) and Time_To_Censored_PFS (inv_ttcpfs)
654 fields (Table S1). Overall survival estimates were computed using OS_Censor_Flag (censos) and
655 Time_To_Censored_OS (ttcos) fields. Univariate and multivariate Cox proportional hazards
656 models were calculated using the coxph function from the survival package.

657 Integrated Analysis for Gain and Loss of Function Genes

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

667
668 The gene model from Ensembl v74 defines 63,677 distinct genes or gene elements,
669 however, only 57,997 of these map to a contig defined by the hs37d5 reference genome used in

670 this study, and only 57,736 of these map to a primary contig (chromosomes 1-22, X, and Y). The
671 list of genes was reduced to a final list of 23,221 analyzed genes by excluding immunoglobulin
672 elements, excluding genes whose gene names were missing or contained “.” or “-”, and requiring
673 the genes were annotated with the following biotypes: lincRNA, miRNA, processed_transcript,
674 protein_coding, snoRNA, snRNA, TR_C_gene, TR_D_gene, TR_J_gene, TR_J_pseudogene,
675 TR_V_gene, TR_V_pseudogene.

676 Loss-of-Function (LOF)

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

685
686 In the LOF analysis, a gene was defined as being functional, or exhibiting complete or
687 partial LOF. Complete LOF implied that all functional alleles of the gene are lost, whereas partial
688 LOF implied that there is evidence of inactivation of one allele, and a potential haploinsufficiency.
689 For example, a somatic deletion, NS mutation, structural rearrangement within a gene body, copy
690 neutral LOH, or a constitutional LOF mutation would result in partial LOF of a gene, whereas two
691 or more of these events impacting a gene would result in complete LOF. We used a heuristic
692 approach to determine the functional state of a gene where each position of an 11 bitset identified
693 the presence or absence of a specific genetic event category. The bits and their corresponding
694 definitions in order are: Cl, 1-copy loss; Cd, 2-copy loss; Cn, copy neutral (2-4 copies); M2, two
695 or more NS mutations; M1, one NS mutation; Lh, loss-of-heterozygosity; Ne, gene not expressed;
696 sD, structural deletion; sT, structural translocation; sI, structural inversion; Gm, constitutional
697 SNP. These 11 bits were summarized to assign each gene a status of complete loss, partial loss,
698 or functional.

699
700 The segmented CN data was transformed into a sample by gene matrix where each entry
701 represented the lowest log2 fold change observed overlapping the gene body (Table S5). These
702 entries were translated into three flags: ‘Cd’ for homozygous deletions or 2-copy loss; ‘Cl’ for 1-

703 copy or heterozygous loss; and ‘*Cn*’ designating 2-4 gene copies. We used a log₂ threshold of ≤-
704 2.32 to identify homozygous deletions, which represents the theoretical value if a homozygous
705 deletion exists in 80% of cells, while the remaining 20% of cells are in the normal copy state. A
706 threshold of ≤-0.1613 was used to identify single copy heterozygous losses, which represents the
707 theoretical log₂ value if a single copy deletion exists in ~21.1% of cells.

708
709 The only somatic mutations included in the analysis were those annotated as NS mutations
710 with the most damaging effect of the mutation to any transcript being the source of the effect used
711 in the LOF model (Table S2). The count of these mutations per gene was used to define the ‘M1’
712 bit, which is set when a single mutation is observed, and the ‘M2’ bit when multiple NS mutations
713 were detected. A single NS mutation in a gene, ‘M1’=1, was considered a partial loss and could
714 contribute to complete loss if an additional event from any other mechanism (ie. CN loss) was
715 also observed. When two independent NS mutations were detected in a gene, ‘M2’=1, we
716 assumed they exist in trans, which would represent a bi-allelic loss and thus could be interpreted
717 as a complete LOF based on mutation data alone.

718
719 To identify regions of copy neutral LOH we integrated the observed copy number and B-
720 allele frequency (BAF). To set the ‘*Lh*’ bit, we extracted the largest overlapping segment from the
721 absolute BAF segmentation data, and set ‘*Lh*’=1 when ‘*Cn*’=1 and the absolute BAF value was
722 between 0.45 and the max value of 0.5 (Table S5).

723
724 To capture loss of gene expression through potential secondary mechanisms, such as
725 epigenetic repression, we identified genes with outlier loss of expression by integrating MAD
726 analysis and the absolute expression value. We used the MAD approach utilized for Cancer
727 Outlier Profile Analysis (COPA) to normalize the gene expression TPM values³². Using this
728 approach, the TPM for each gene was median centered and scaled to their median absolute
729 deviation (MAD). Given the TPM of a gene, the MAD value was computed as:

$$734 \quad MAD_{sample} = [\log_2(TPM_{sample}) - median(\log_2(TPM_{cohort}))] / mad(\log_2(TPM_{cohort}))$$

730 A gene was defined as not expressed if the MAD value was less than the 25th percentile value
731 minus 1.5 times the interquartile range (IQR) and the observed corrected TPM was less than 0.1.
732 When these criteria exist the ‘*Ne*’ bit was set to 1 to denote a potential epigenetic loss of RNA
733 expression.

735

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

744
745 Individuals can also inherit a defective copy of a gene contributing to a LOF event when
746 the remaining functional allele is lost. To identify inherited LOF events we performed joint variant
747 calling on the constitutional exomes from the entire MMRF cohort using GATK (v3.5) best practice
748 guidelines. Individual gVCFs were created using HaplotypeCaller with -L to limit the analysis to
749 the regions targeted by the exome kit. The patient specific gVCF files were combined in batches
750 of 100 using CombineGVCFs and then joint SNV and indel calling was performed across the
751 batched gVCFs using GenotypeGVCFs. Variant quality recalibration (VQSR) was applied
752 independently on the SNP and INDEL detected to reduce false positives. While applying the
753 recalibration, any SNP below the 99.0 tranche and INDEL below the 95.0 tranche were marked
754 as a PASS. These variants were annotated using GATK VariantAnnotator to flag when a variant
755 existed in one of the following databases: dbSNP (v147), ExAC (v3.0), ClinVar (v20170530),
756 NHLBI (v2 ssa137), 1000 Genomes (Phase 3), and COSMIC (v78). To identify highly damaging
757 LOF variants we annotated the variants independently using snpEFF (v4.2) and
758 VEP(v87)/LOFTEE following established rules³³. Expected LOF variants were extracted from the
759 annotated files when a heterozygous variant was marked as LOF by snpEFF or a high-confidence
760 LOF by LOFTEE and the allele frequency in the CoMMpass cohort was less than or equal to 0.05,
761 and the variant was marked as PASS or the variant was known in dbSNP or ExAC. Variants
762 annotated as LOF by both tools were then annotated with the observed allele ratio in the matching
763 tumor exome and tumor RNAseq alignments. Using the available data, the 'Gm' bit was set to
764 indicate if the inherited LOF variant was detectable in the tumor at an expected allele frequency.
765 In the absence of other somatic events, the bit was set when the AR was at least 0.4. When a CN
766 loss was detected we required the AR to be above 0.6 to confirm the LOF impacted the remaining
767 allele, and when copy neutral LOH was detected we required the AR to exceed 0.85.

768

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

775 Gain-of-Function (GOF)

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

783

784 To identify high level amplifications, the 'Cg' bit was set to 1 when the CN segment for a
785 gene was estimated to exceed 6 total copies ($\log_2 \text{CN} \geq 1.5$), the full gene was contained within
786 the amplification, and the gene was detectably expressed ($\text{TPM} \geq 1$). The expression filter
787 prevents non-expressed genes within an amplified segment from being nominated as a potential
788 oncogene.

789

790 To flag NS SNV/INDEL that likely result in a GOF events, we identified events that fell into
791 one of three categories based on the altered amino acid position. To identify recurrent mutations,
792 the 'Rm' bit is set for a gene in any patient where the observed mutation altered the same amino
793 acid in at least two independent patients. When a gene was flagged as having recurrent mutations
794 in the cohort, we also flagged clustered mutations in the gene to set the 'Cm' bit. Mutations were
795 considered to be clustered when at least 5 different patients had NS mutations within 10% of the
796 CDS space of a gene. Finally, when a gene was flagged as having recurrent mutations, we
797 nominated any NS mutation in a gene by setting the 'Nm' bit (even including those without the
798 'Rm' or 'Cm' bits set) to capture potential GOF events by rare activating mutations.

799

800 To detect overexpression of a gene, which is common in myeloma through events like
801 immunoglobulin translocations, we identified outlier RNA expression using the same MAD

802 approach described in the LOF analysis. A gene was defined as overexpressed if the MAD value
803 was greater than the 75th percentile value plus 1.5 times the interquartile range (IQR) and the
804 observed corrected TPM ≥ 1 . To limit false positive overexpression calls, we only set the 'Oe' bit
805 when an accompanying DNA structural event was detected in the gene or its upstream or
806 downstream region.

807
808 When an inframe fusion transcript was detected in the RNAseq data we set the 'Hf' bit for
809 both genes involved in the fusion. The fusion events included were independently validated in the
810 matching WGS assay, and the expression of both genes needed to be ≥ 1 TPM.

811
812 A gene was defined as gained in a sample if any of the 'Cg', 'Rm', 'Cm', 'Oe', or 'Hf' bits
813 were set to 1. Samples with only the 'Nm' bit set for a gene were considered to have a potential
814 GOF event and were only included in the count of samples with GOF events in a particular gene
815 when at least 5 unique samples had one of the other bits set. In addition to the individual
816 mechanism of gain, we also required the gene to be expressed (≥ 1 TPM) in the sample for it to
817 be considered a GOF gene.

818 Curation of LOF and GOF Genes

819 The integrated analysis focused on identifying genes with potential recurrent LOF and
820 GOF events, however, some genes were nominated in both the GOF and LOF analysis (e.g.
821 NRAS and KRAS). For genes that were identified as the target of both LOF and GOF in 5 or more
822 patients, we relied on COSMIC Cancer Gene Census annotations (accessed August 2017) and
823 reports of gene function from published literature. Known oncogenes removed from the LOF but
824 retained on the GOF list were NRAS, KRAS, and WHSC1, while known tumor suppressor genes
825 removed from the GOF list but retained on the LOF list were CDKN1B, DIS3, EGR1, FAM46C,
826 MAX, NOTCH2, PABPC1, PARP4, RPL10, SDHA, TP53, and TRAF3. Genes that were
827 exclusively nominated to the GOF list by mutational events were further curated based on their
828 reported function in the literature and were removed if there was strong evidence supporting that
829 the gene functions as a tumor suppressor rather than an oncogene. The genes removed through
830 this process were DUSP2 and KLHL6, based on evidence from the literature, and BTG1, which
831 is listed in the Cancer Gene Consensus as a tumor suppressor gene.

832
833 In patients where a gene had complete LOF as a result of two mutations, visual validation
834 of mutations within close proximity was performed to determine if the mutations were in cis or

835 trans. In order to perform visual validation, we required that sequencing reads or read-pairs
836 spanned both of the mutation positions. If the mutations occurred in trans, no action was taken,
837 however, if the mutations occurred in cis the gene was downgraded from complete to partial LOF
838 for that particular sample. CDKN1B, EGR1, LRP6, MAGEC3, PABPC1, PRR14L, SYNE1, and
839 UBR5 had samples whose variants were phased in cis, and after downgrading these events from
840 complete to partial LOF, were removed from the list of recurrent complete LOF events occurring
841 in 5 or more patients.

842
843 To further curate the LOF list, visual validation was performed to remove double calling
844 events that can arise such as when a copy number deletion breakpoint overlaps a translocation
845 breakpoint. Visual verification was limited to samples with contiguous genes identified as having
846 a LOF involving a structural event. From the LOF list, CDKN2C, PSPC1, RB1, RCBTB2,
847 CDC42BPB, CYLD, SNX20 had samples with LOF events that were manually modified after
848 visual inspection. In addition, MAGI1 was removed from the GOF list because four of the five
849 patients were flagged due to a translocation event in SLC25A26 which is believed to be a false-
850 positive arising due to poor read mapping in this area.

851 Identification of Copy Number Subtypes in Multiple Myeloma

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

861 Identification of RNA Subtypes in Multiple Myeloma

862 Prior to unsupervised mRNA expression clustering, the $\log_2(\text{TPM}+1)$ transformed
863 expression data of 56,430 genes and 714 BM baseline tumor samples were adjusted for
864 confounding biobank effects by applying ComBat from the sva R package (v3.26.0)³⁴. The original
865 list of genes were then ranked by their corresponding geometric coefficient of variation (GCV).

866 Larger GCV values are associated with greater variability in random variables that follow a log-
867 normal distribution and therefore genes with a GCV less than 1 were removed from the dataset.
868 The resulting 4811 gene by 714 sample expression matrix was mean centered as a method for
869 standardizing the data. Clustering analysis was performed by running three separate trials of
870 ConsensusClusterPlus (v1.42.0) with the PAM algorithm and Pearson distance metric³⁵. Each
871 trial computed sample class assignments for an increasing number of clusters (K), from 2 to 20,
872 using the average linkage option and 80% subsampling with the number of repetitions set to
873 10,000. The optimum K was chosen based on the silhouette score for each cluster across all
874 trials. All three trials with different initial seed values (3000094, 2862787, 8275806) converged to
875 the solution $K^* = 12$ by silhouette score.

876

877 Specific gene markers for each group were then determined using multinomial logistic
878 regression with L1 regularization trained to predict class assignment³⁶. For training, the model
879 took the $\log_2(\text{TPM}+1)$ mean centered expression matrix of the 4811 genes and 714 samples as
880 input and class assignments from the unsupervised clustering as targets. The value of the
881 regularization parameter that minimized the model's objective function was determined during
882 training via 20 fold cross-validation. The resulting non-zero values for the model coefficients were
883 indicative of genes that were most predictive for a given class; 607 genes in total (Table S7).

884

885 Further analysis of the resulting RNA clusters included cross-examination with: a gene
886 expression profiling index, used to assess cellular proliferation; an NFkB index; and previously
887 identified RNA subtypes^{17,18,28,37,38}. The proliferation index developed by Bergsagel et al. was
888 calculated for each sample by measuring the average $\log_2(\text{TPM}+1)$ transformed expression of 12
889 genes: TYMS, TK1, CCNB1, MKI67, KIAA101, KIAA0186, CKS1B, TOP2A, UBE2C, ZWINT,
890 TRIP13, KIF11. In addition, the NFkB(11) index for each sample was determined by calculating
891 the geometric mean expression of 11 genes: BIRC3, TNFAIP3, NFkB2, IL2RG, NFkB1, RELB,
892 NFkBIA, CD74, PLEK, MALT1, and WNT10A. Each subtype specific expression signature is
893 composed of a list of up-regulated and down-regulated genes. In both Broyl and Zhan et al., these
894 genes were identified using Affymetrix U133Plus2.0 microarray data. In order to apply these gene
895 lists to RNAseq data, we identified the corresponding Ensembl v74 ENSG ID for each Affymetrix
896 probe set using bioMartR, however, not all probesets had unique mappings to ENSG IDs. Many
897 probesets were associated with multiple ENSG while other probesets were not associated with
898 an ENSG. Only genes with unique probeset to ENSG mappings were used to calculate subtype
899 specific expression signatures. Once the unique mappings were identified, each subtype specific

900 signature was calculated for each sample using: $mean(\log_2(x_{up-regulated}) -$
901 $\log_2(x_{down-regulated}))$. Integration of the genes predictive of each CoMMpass subtype,
902 canonical structural events in MM, common CN events, previous subtype signatures, and
903 expression index data was used to name the 12 subtypes.

904

905 To assign longitudinal samples with RNAseq data to each of these classes we developed
906 a trained model classifier, which has the benefit of outputting class probabilities that can be used
907 to evaluate transitions in time. The trained classifier uses the corrected TPM values for the 4811
908 genes used for clustering as input. These values are transformed, $\log_2(TPM+1)$, followed by
909 mean centering using the pre-computed mean expression values from the training set. Running
910 this input through the trained classifier resulted in classifications for each of the 107 samples as
911 well as the corresponding class probabilities associated with each of the 12 classes.

912

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

918 Clinical and Molecular Associations with RNA subtypes

919 We used the non-parametric Fisher's Exact test to determine if a clinical feature, LOF, or
920 GOF event was enriched in the identified CN and RNA subtypes. The test works on the null
921 hypothesis that no nonrandom associations exist between two categorical variables, against the
922 alternative hypothesis that there is a nonrandom association between variables. The enrichment
923 analysis was performed using the function *fishertest* in Matlab2019a. The inputs for the
924 enrichment analyses were the LOF and GOF file filtered for 5 or more patients, the patient features
925 table (Table S1), and the gene model. The two-tailed test was used to check for enrichment or
926 depletion of each measurement within each subtype. An odds ratio was computed that indicated
927 significant enrichment or depletion of the measurement in each subtype such that an odds ratio
928 >1 indicated enrichment and <1 indicated depletion. In addition, for a condition to be depleted in
929 a group it was required to be present in at least 20% of the cohort. The Benjamini-Hochberg test
930 correction was used for multiple testing, which computed a pFDR value.

931 Acknowledgements

932 The CoMMpass study was funded by the Multiple Myeloma Research Foundation. We
933 would like to thank all of the MMRF CoMMpass study participants and their families for making
934 this research possible. We would also like to thank all of the laboratory staff at the Spectrum
935 Health Advanced Technology Laboratory who assisted with sample processing and flow
936 cytometry analysis.

937 Data Availability

938 DNA and RNA sequencing data are available from dbGAP, phs000748, and the Genomic
939 Data Commons. Summarized somatic and clinical data are also available on the MMRF
940 researcher gateway (<https://research.themmr.org/>). Clinical data is updated and new molecular
941 data is added with each bi-annual interim release.

942 Author Contributions

943 Conceptualization, S.S., D.A., J.C., S.L., J.K.; Methodology, S.S., D.P., A.W.C., S.N., M.B., S.K.,
944 K.S., D.C., W.L., J.K.; Software, S.S., D.P., A.W.C., S.N., J.L.A., C.L., B.B., C.M., B.T., A.K.,
945 M.W., V.Y., M.T., K.M., D.C., J.K.; Validation, S.S., A.W.C., M.T., K.M., J.K.; Formal Analysis,
946 S.S., D.P., A.W.C., S.N., J.L.A., C.L., B.B., C.M., B.T., A.K., M.W., V.Y., J.K.; Investigation, J.R.A.,
947 L.C., M.B., A.H., S.K., J.M., R.R., K.S., E.T., A.B., B.D., M.K., M.D.A., M.G., J.S., P.K., A.D., B.Z.,
948 M.T., K.M., H.J.C., A.C., S.J., D.S.S., R.V., G.O., A.J., R.N., D.L., M.L., J.B., S.Z.U., D.C., S.D.J.,
949 S.L., J.K.; Resources, H.J.C., S.J., D.S.S., R.V., G.O., A.J., R.N., D.L., M.L., J.B., S.L.; Data
950 Curation, S.S., D.P., A.W.C., S.N., J.L.A., C.L., D.C.R., M.D.A., M.G., A.D., B.Z., M.T., K.M., M.D.,
951 H.J.C., J.Y., D.A., S.D.J., J.K.; Writing - Original Draft, S.S., J.K.; Writing - Review & Editing S.S.,
952 B.B., D.S.S., N.C.G., A.C., S.Z.U., K.C.A., H.J.C., J.K.; Visualization, S.S., D.P., A.W.C., S.N.,
953 J.L.A.; Supervision, S.S., D.C., J.Y., D.A., S.D.J., J.C., S.L., J.K.; Project Administration, D.C.R.,
954 D.C., A.D., B.Z., K.M., M.D., J.Y., D.A., N.C.G., S.D.J., J.C., S.L., J.K.; Funding Acquisition, P.K.,
955 S.D.J., J.C., J.K.

956 Competing Interests

957 The authors declare no competing interests.

958

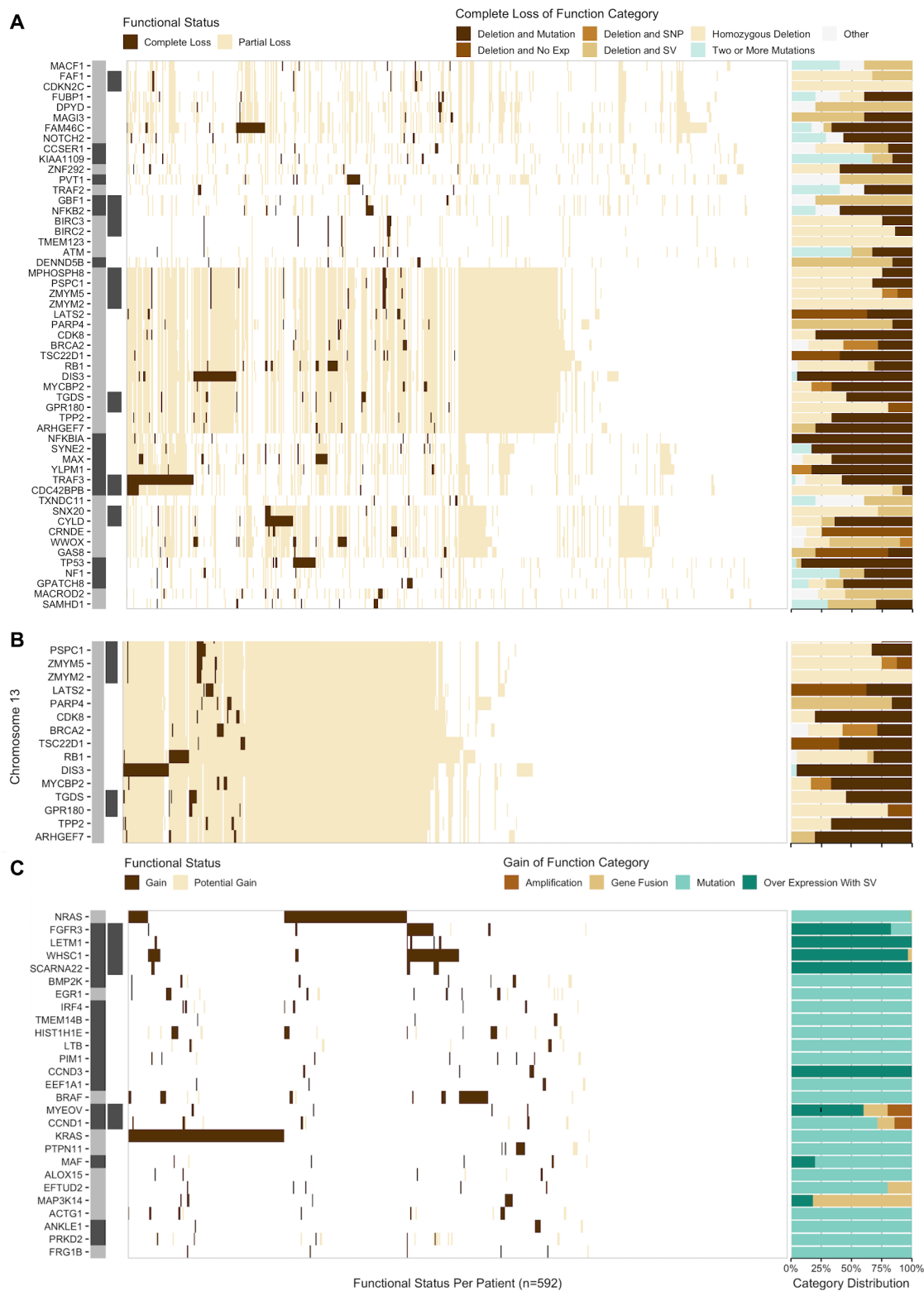
959 References

- 960 1. Lewis, J. P. & MacKenzie, M. R. Non-random chromosomal aberrations associated with
961 multiple myeloma. *Hematol. Oncol.* 2, 307–317 (1984).
- 962 2. Avet-Loiseau, H. et al. Prognostic significance of copy-number alterations in multiple
963 myeloma. *J. Clin. Oncol.* 27, 4585–4590 (2009).
- 964 3. Carrasco, D. R. et al. High-resolution genomic profiles define distinct clinico-pathogenetic
965 subgroups of multiple myeloma patients. *Cancer Cell* 9, 313–325 (2006).
- 966 4. Venti, G., Mecucci, C., Donti, E. & Tabilio, A. Translocation t(11;14) and trisomy 11q13---
967 qter in multiple myeloma. *Ann. Genet.* 27, 53–55 (1984).
- 968 5. Bergsagel, P. L. et al. Promiscuous translocations into immunoglobulin heavy chain switch
969 regions in multiple myeloma. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13931–13936 (1996).
- 970 6. Shaughnessy, J., Jr et al. Cyclin D3 at 6p21 is dysregulated by recurrent chromosomal
971 translocations to immunoglobulin loci in multiple myeloma. *Blood* 98, 217–223 (2001).
- 972 7. Avet-Loiseau, H. et al. Oncogenesis of multiple myeloma: 14q32 and 13q chromosomal
973 abnormalities are not randomly distributed, but correlate with natural history, immunological
974 features, and clinical presentation. *Blood* 99, 2185–2191 (2002).
- 975 8. Boersma-Vreugdenhil, G. R. et al. The recurrent translocation t(14;20)(q32;q12) in multiple
976 myeloma results in aberrant expression of MAFB: a molecular and genetic analysis of the
977 chromosomal breakpoint. *Br. J. Haematol.* 126, 355–363 (2004).
- 978 9. Fonseca, R. et al. The recurrent IgH translocations are highly associated with
979 nonhyperdiploid variant multiple myeloma. *Blood* 102, 2562–2567 (2003).
- 980 10. Bolli, N. et al. Heterogeneity of genomic evolution and mutational profiles in multiple
981 myeloma. *Nature Communications* vol. 5 (2014).
- 982 11. Lohr, J. G. et al. Widespread genetic heterogeneity in multiple myeloma: implications for
983 targeted therapy. *Cancer Cell* 25, 91–101 (2014).
- 984 12. Walker, B. A. et al. Mutational Spectrum, Copy Number Changes, and Outcome: Results of
985 a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J. Clin. Oncol.* 33, 3911–
986 3920 (2015).
- 987 13. Chapman, M. A. et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*
988 471, 467–472 (2011).
- 989 14. Greipp, P. R. et al. International staging system for multiple myeloma. *J. Clin. Oncol.* 23,
990 3412–3420 (2005).
- 991 15. Palumbo, A. et al. Revised International Staging System for Multiple Myeloma: A Report

- 992 From International Myeloma Working Group. *J. Clin. Oncol.* 33, 2863–2869 (2015).
- 993 16. Keats, J. J. et al. Promiscuous Mutations Activate the Noncanonical NF- κ B Pathway in
994 Multiple Myeloma. *Cancer Cell* vol. 12 131–144 (2007).
- 995 17. Zhan, F. et al. The molecular classification of multiple myeloma. *Blood* 108, 2020–2028
996 (2006).
- 997 18. Broyl, A. et al. Gene expression profiling for molecular classification of multiple myeloma in
998 newly diagnosed patients. *Blood* 116, 2543–2553 (2010).
- 999 19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human
1000 genome. *Nature* 489, 57–74 (2012).
- 1001 20. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update.
1002 *Nucleic Acids Res.* 46, D794–D801 (2018).
- 1003 21. Walker, B. A. et al. APOBEC family mutational signatures are associated with poor
1004 prognosis translocations in multiple myeloma. *Nat. Commun.* 6, 6997 (2015).
- 1005 22. Misund, K. et al. MYC dysregulation in the progression of multiple myeloma. *Leukemia* 34,
1006 322–326 (2020).
- 1007 23. Cho, S.-J. et al. Ninjurin1, a target of p53, regulates p53 expression and p53-dependent
1008 cell survival, senescence, and radiation-induced mortality. *Proc. Natl. Acad. Sci. U. S. A.*
1009 110, 9362–9367 (2013).
- 1010 24. Chng, W. J. et al. Clinical significance of TP53 mutation in myeloma. *Leukemia* 21, 582–
1011 584 (2007).
- 1012 25. Boyd, K. D. et al. A novel prognostic model in myeloma based on co-segregating adverse
1013 FISH lesions and the ISS: analysis of patients treated in the MRC Myeloma IX trial.
1014 *Leukemia* 26, 349–355 (2012).
- 1015 26. Walker, B. A. et al. A high-risk, Double-Hit, group of newly diagnosed myeloma identified by
1016 genomic analysis. *Leukemia* 33, 159–170 (2019).
- 1017 27. Kumar, S. et al. Efficacy of venetoclax as targeted therapy for relapsed/refractory t(11;14)
1018 multiple myeloma. *Blood* vol. 130 2401–2409 (2017).
- 1019 28. Bergsagel, P. L. et al. Cyclin D dysregulation: an early and unifying pathogenic event in
1020 multiple myeloma. *Blood* 106, 296–303 (2005).
- 1021 29. Stein, C. K. et al. The varied distribution and impact of RAS codon and other key DNA
1022 alterations across the translocation cyclin D subgroups in multiple myeloma. *Oncotarget* 8,
1023 27854–27867 (2017).
- 1024 30. Augert, A. et al. MAX Functions as a Tumor Suppressor and Rewires Metabolism in Small
1025 Cell Lung Cancer. *Cancer Cell* (2020) doi:10.1016/j.ccell.2020.04.016.

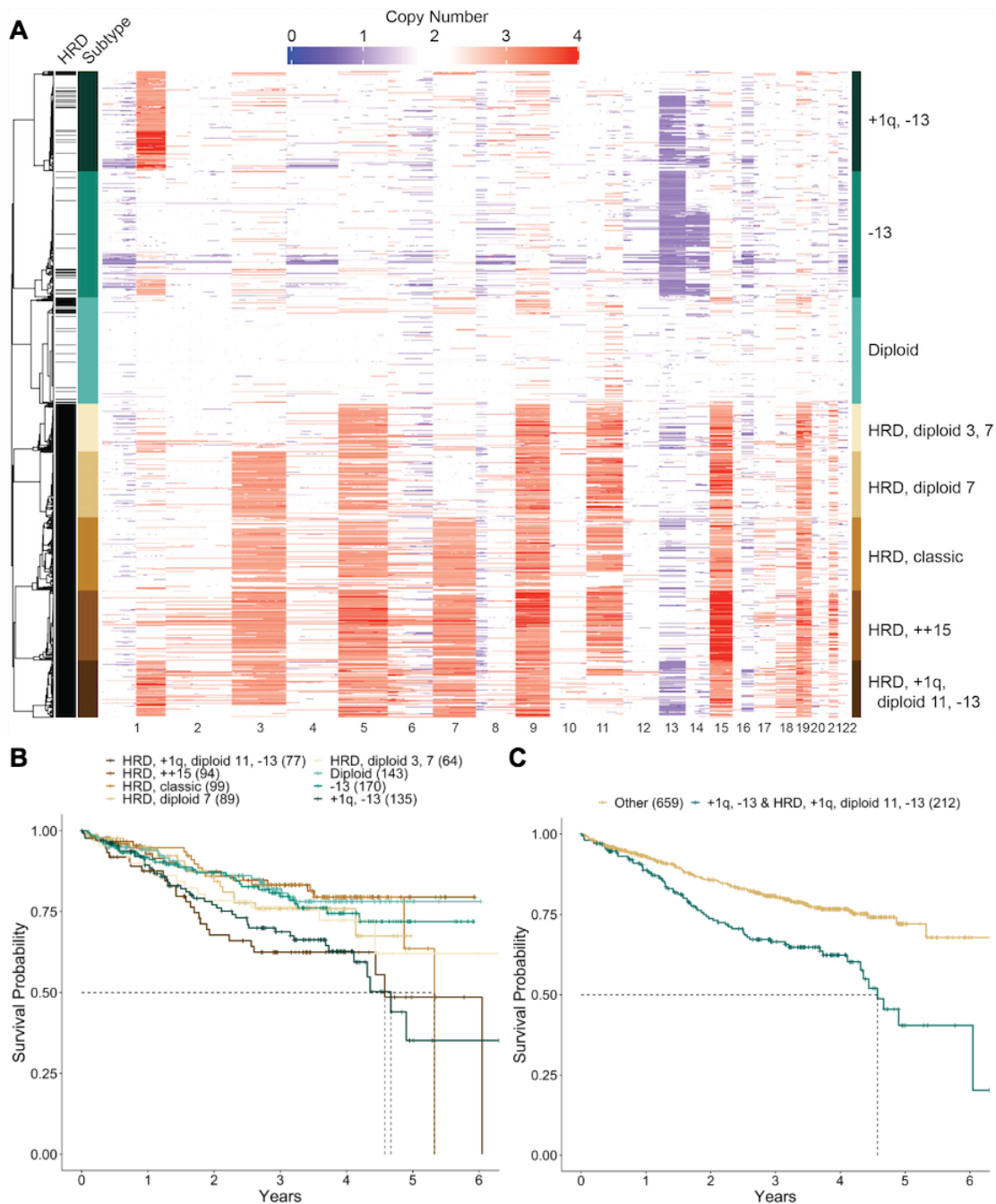
- 1026 31. Oken, M. M. et al. Toxicity and response criteria of the Eastern Cooperative Oncology
1027 Group. AMERICAN JOURNAL OF CLINICAL ONCOLOGY vol. 5 649–656 (1982).
- 1028 32. Tomlins, S. A. Recurrent Fusion of TMPRSS2 and ETS Transcription Factor Genes in
1029 Prostate Cancer. Science vol. 310 644–648 (2005).
- 1030 33. MacArthur, D. G. et al. A systematic survey of loss-of-function variants in human protein-
1031 coding genes. Science 335, 823–828 (2012).
- 1032 34. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for
1033 removing batch effects and other unwanted variation in high-throughput experiments.
1034 Bioinformatics 28, 882–883 (2012).
- 1035 35. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with
1036 confidence assessments and item tracking. Bioinformatics 26, 1572–1573 (2010).
- 1037 36. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear
1038 Models via Coordinate Descent. J. Stat. Softw. 33, 1–22 (2010).
- 1039 37. Annunziata, C. M. et al. Frequent engagement of the classical and alternative NF-kappaB
1040 pathways by diverse genetic abnormalities in multiple myeloma. Cancer Cell 12, 115–130
1041 (2007).
- 1042 38. Demchenko, Y. N. et al. Classical and/or alternative NF-κB pathway activation in multiple
1043 myeloma. Blood vol. 115 3541–3552 (2010).

1044 Figures



1045

1046 **Figure 1. Recurrent LOF and GOF events occurring in at least five patients at diagnosis**
1047 **ordered by event frequency.** The location and proximity of individual genes is shown next to
1048 each gene with the alternating gray and black bar illustrating when the chromosomal location
1049 changes, while black bars directly to the right denote contiguous genes. (A) Complete LOF was
1050 observed in 53 autosomal located genes. (B) Genes on chromosome 13q that were the target of
1051 complete LOF events in at least five patients in the baseline cohort. (C) GOF events were detected
1052 in 27 autosomal genes.



1053

1054

1055

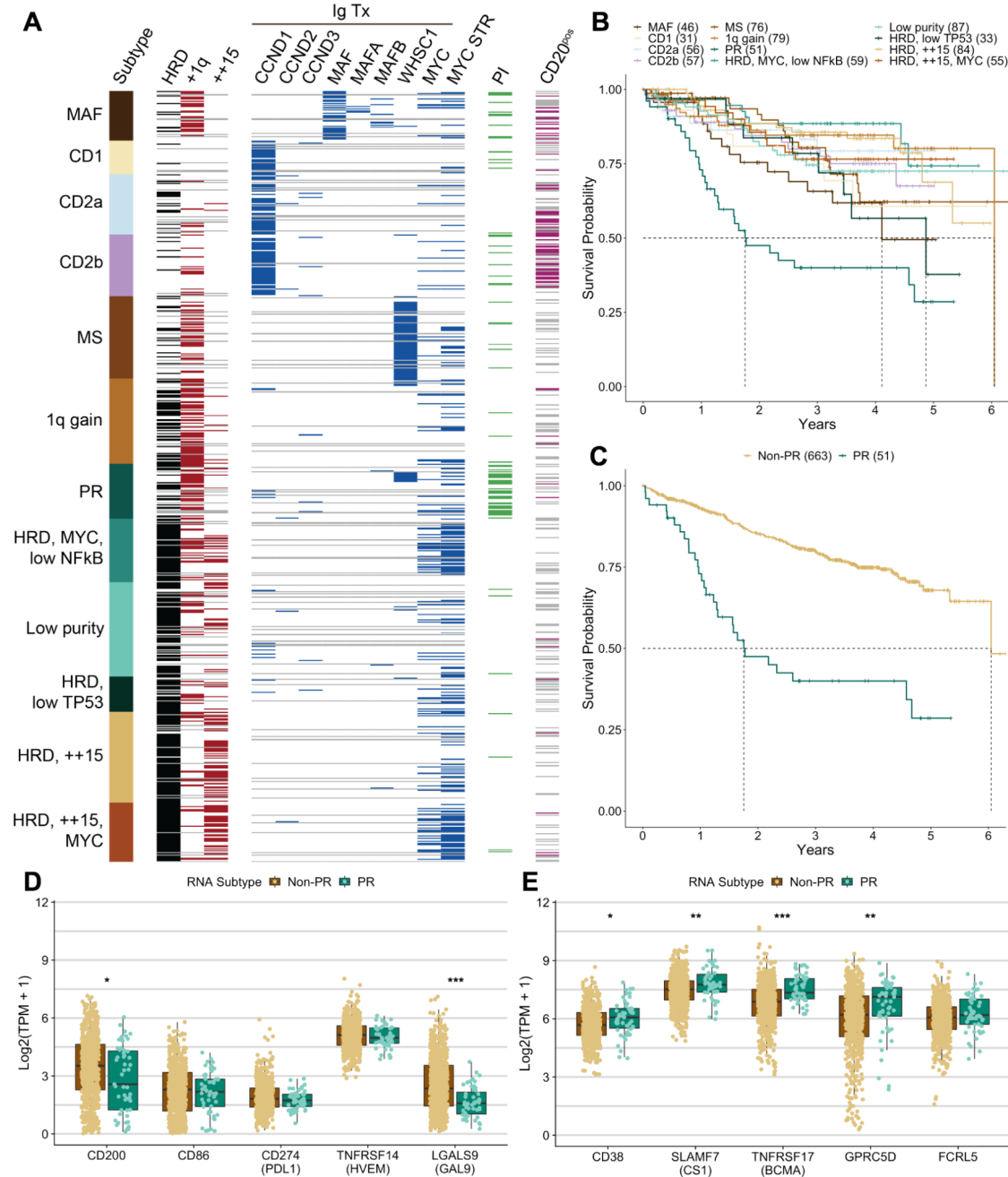
1056

1057

1058

1059

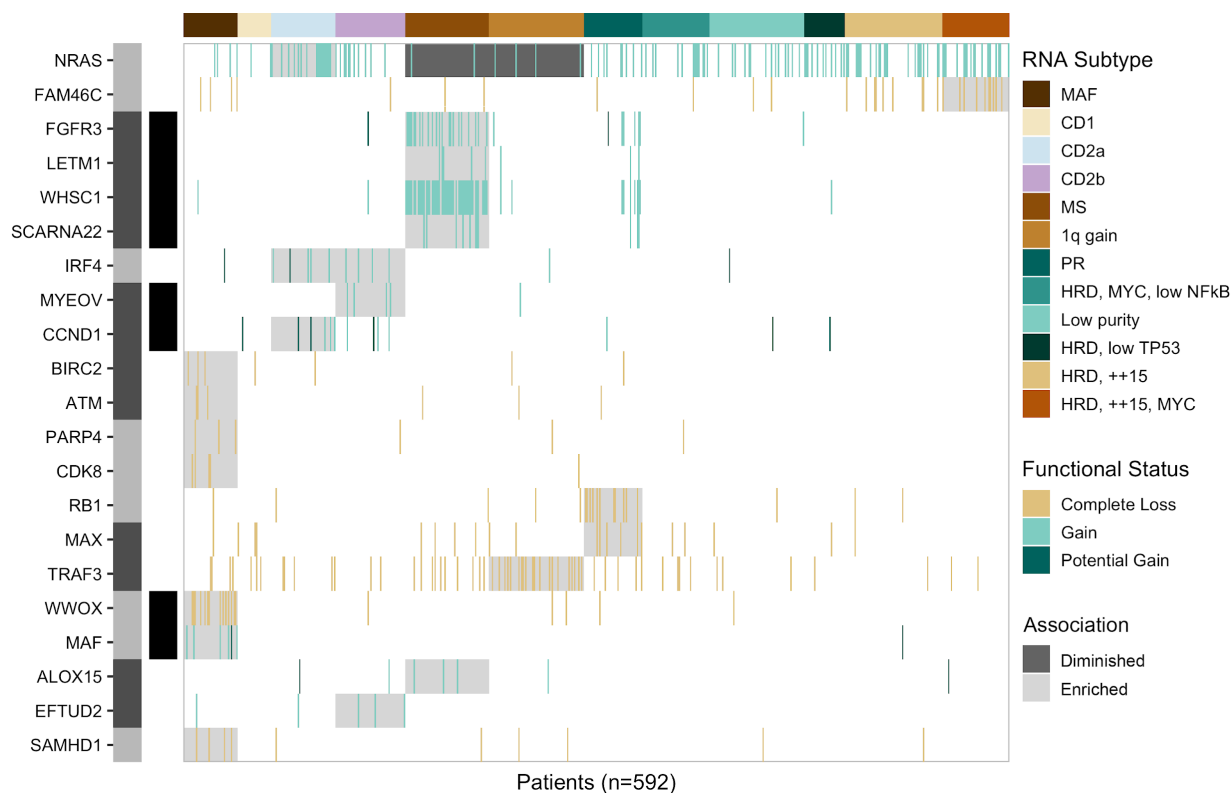
Figure 2. Copy number subtypes of multiple myeloma. (A) Consensus clustering of WGS CN data identified eight unique CN subtypes, comprising five HRD and three NHRD clusters that were annotated based on common CN features. (B) OS outcomes by CN subtype. Median OS was met for the HRD, +1q, diploid 11, -13 (56 months), +1q, -13 (57 months), and HRD, classic (65 months) subtypes. (C) OS outcomes of patients in the +1q, -13 and HRD, +1q, diploid 11, -13 groups (median = 56 months) versus patients in other CN subtypes ($p < 0.001$).



1060

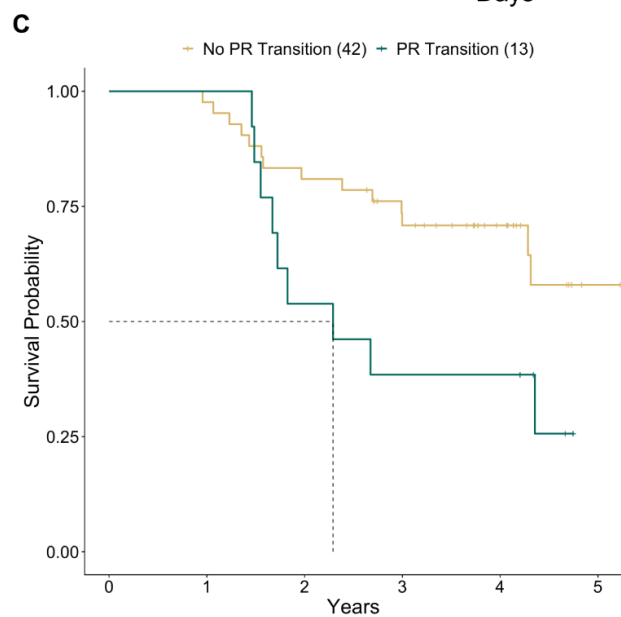
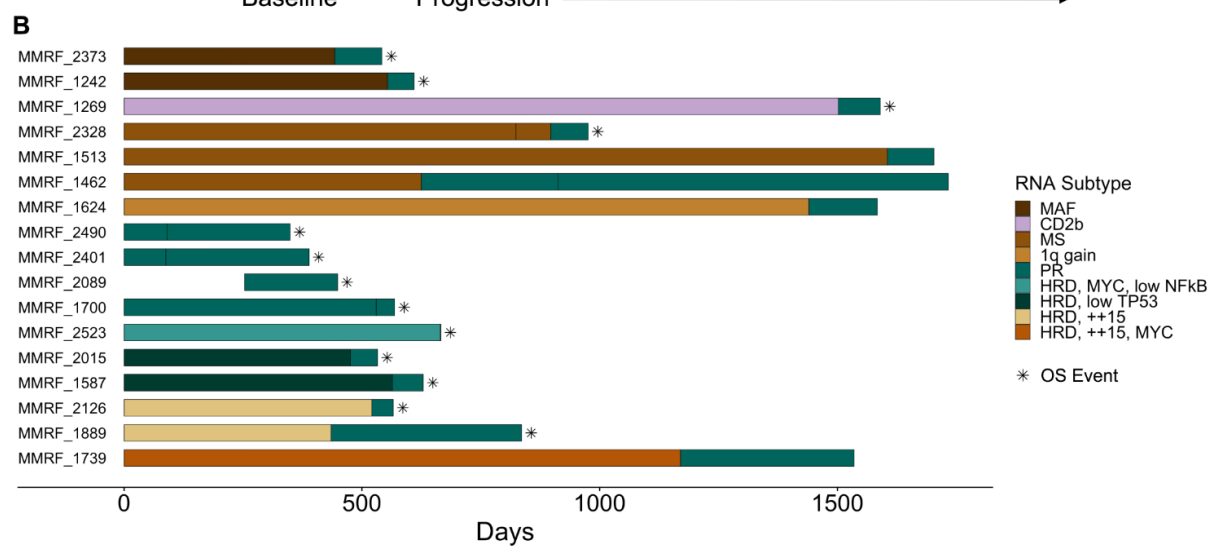
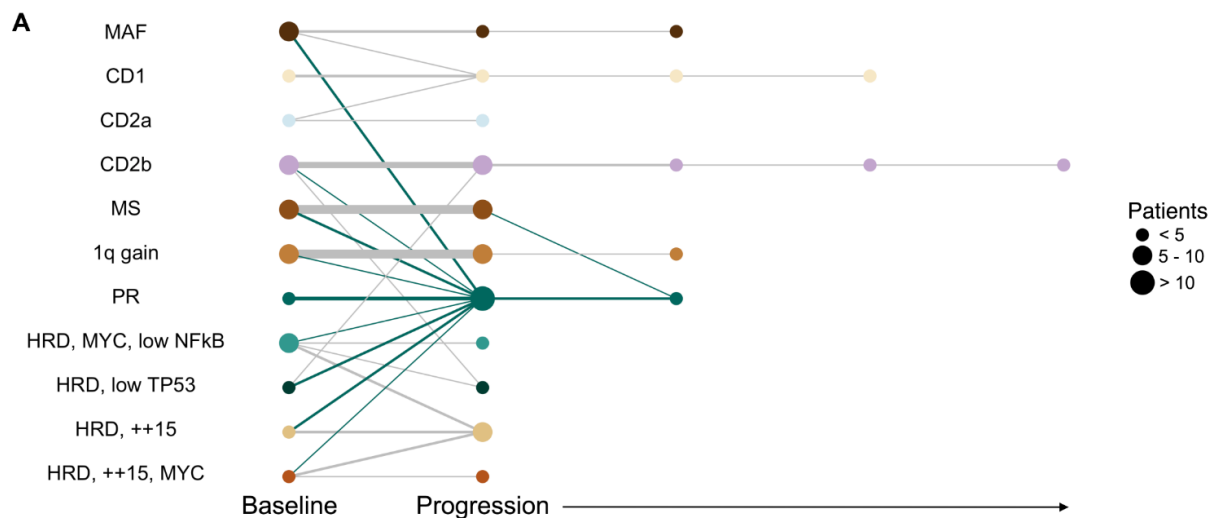
1061 **Figure 3. RNA subtypes of multiple myeloma and associated characteristics.** (A) Consensus
 1062 clustering of RNAseq data revealed 12 RNA subtypes of multiple myeloma. (B) OS outcomes for
 1063 patients by RNA subtype. Median OS was reached for the PR (21 months); MAF (50 months);
 1064 HRD, low TP53 (59 months); and 1q gain (74 months) subtypes. (C) OS outcomes of patients in
 1065 the PR versus non-PR (median = 74 months) subtype at diagnosis ($p < 0.0001$, HR = 3.73 [95%
 1066 CI 2.49-5.58]). Expressed (TPM > 1 in at least one group) checkpoint inhibitor (D) and
 1067 immunotherapy (E) targets in non-PR versus PR patients. A significant difference in median
 1068 expression between the two groups is designated (** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$).

1069



1070

1071 **Figure 4. RNA subtypes and significantly associated LOF and GOF events (pFDR≤0.1,**
1072 **p<0.05).** The location and proximity of individual genes is shown next to each gene with the
1073 alternating gray and black bar illustrating when the chromosomal location changes, while black
1074 bars directly to the right denote contiguous genes.



1076 **Figure 5. RNA subtype of serial patients at baseline and progression.** (A) Node size reflects
1077 the relative number of patients in each RNA subtype at each timepoint while edge width reflects
1078 the relative number of patients remaining in, or transitioning to, a particular RNA subtype, with the
1079 thinnest line and thickest line representing 1 and 7 patients respectively. (B) Swimmers plot of
1080 patients in the PR subtype at either baseline or progression. Vertical breaks indicate visits with
1081 available RNA sequencing data for RNA subtype prediction. Fill color indicates RNA subtype
1082 between visits. Asterisks denote OS events. (C) OS outcomes for serial patients that transition to
1083 the PR subtype at progression (median = 28 months) versus those that do not (median not met,
1084 $p < 0.05$).

1085 **Tables**

1086 **Table 1. Characteristics of the baseline CoMMpass cohort**

Characteristic	n = 1143
Age at Diagnosis – yr	
Median (Range)	63 (27 – 93)
Distribution - no. (%)	
<55 yr	239 (20.9)
55 - 64 yr	400 (35.0)
65 – 74 yr	358 (31.3)
≥ 75 yr	146 (12.8)
Sex – no. (%)	
Male	690 (60.4)
Female	453 (39.6)
Ethnicity – no. (%)	
Caucasian	742 (64.9)
Black	161 (14.1)
Asian	18 (1.6)
Other / Unknown	222 (19.4)
International Staging System (ISS) – no. (%)	
I	401 (35.1)
II	401 (35.1)
III	311 (27.2)
Unknown	30 (2.6)

Type of Myeloma – no. (%)	n = 971
Heavy Chain	
IgG	568 (58.5)
IgA	169 (17.4)
IgM	2 (0.2)
Bi-clonal	11 (1.1)
Negative	123 (12.7)
Unknown	98 (10.1)
Light Chain	
IgK	547 (56.3)
IgL	322 (33.2)
Bi-clonal	15 (1.5)
Negative	25 (2.6)
Unknown	62 (6.4)
Ploidy Status* - no. (%)	n = 871
Hyperdiploid	498 (57.2)
Non-hyperdiploid	373 (42.8)
Immunoglobulin Translocations* - no (%), IgH, IgK, IgL	n = 851
CCND1	170 (20.0), 168, 2, 0
CCND2	10 (1.2), 5, 1, 4
CCND3	15 (1.8), 13, 0, 2
MAF	34 (4.0), 33, 1, 0
MAFA	6 (0.7), 5, 1, 0
MAFB	11 (1.3), 9, 1, 1
MYC	122 (14.3), 55, 18, 49
WHSC1/MMSET/NSD2	109 (12.8), 108, 1, 0
Common Copy Number Alterations* - no. %	n = 871
del(1p22)	212 (24.3)
gain(1q21)	307 (35.2)
del(13q14)	453 (52.0)

del(17p13)	109 (12.5)
ECOG Performance** - no. %	n = 844
0	295 (35.0)
1	406 (48.1)
2	99 (11.7)
3	38 (4.5)
4	6 (0.7)
Cytogenetic Risk Profile***	n = 832
Standard risk	585 (70.3)
High risk	247 (29.7)

1087 *Ploidy Status, immunoglobulin translocation and copy number event data was extracted from
 1088 WGS data
 1089 **ECOG Performance, 0=Fully active, 1=Restricted in physically strenuous activity, 2=Ambulatory
 1090 and capable of all selfcare, 3=Capable of only limited selfcare, 4=Completely disabled³¹
 1091 ***High risk defined as those patients with one or more high-risk events: del17p13, t(14;16) [MAF],
 1092 t(14;20) [MAFB], t(8;14) [MAFA], t(4;14) [NSD2/WHSC1/MMSET])