

1 **Gene interaction network analysis in multiple myeloma detects complex immune**
2 **dysregulation associated with shorter survival**

3
4
5 Anish K. Simhal^{1*}, Kylee H. Maclachlan^{2*}, Rena Elkin¹, Jiening Zhu³, Larry Norton², Joseph O.
6 Deasy¹, Jung Hun Oh¹, Saad Z. Usmani², Allen Tannenbaum^{1,4}

7
8 ¹ Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY

9 ² Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY

10 ³ Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY

11 ⁴ Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook

12 University, Stony Brook, NY

13

14 * Co-first authors; these authors contributed equally

15

16

17 ***Corresponding authors***

18 Kylee H Maclachlan, Myeloma Service, Memorial Sloan Kettering Cancer Center,

19 maclachk@mskcc.org

20 Allen Tannenbaum, Department of Computer Science, Stony Brook University,

21 arobertan@cs.stonybrook.edu

22

23 ***Running title*** Gene networks in myeloma

24

25 ***Keywords*** multiple myeloma, network analysis, RNA-sequencing, copy number aberration,
26 immune system, DNA damage.

27

28 ***Author contributions***

29 AKS: conceptualization, formal analysis, data curation, visualization, writing- original draft.

30 KHM: conceptualization, supervision, writing- original draft, writing- review and editing. RE:

31 methodology, writing – original draft. JZ: methodology, writing – original draft. SZU:

32 conceptualization, writing- review and editing. LN: conceptualization, supervision, writing-

33 review and editing. JOD: conceptualization, supervision, writing- review and editing. JHO:

34 conceptualization, supervision, writing- review and editing. AT: conceptualization, supervision,

35 writing- review and editing.

36

37

38 ***Acknowledgments***

39 The authors would like to thank Jonathan J. Keats at the Translational Genomics Research

40 Institute and the team at the Multiple Myeloma Research Foundation for their incredible work

41 and support with the CoMMpass dataset. KHM received support from the Multiple Myeloma

42 Research Foundation, the American Society of Hematology, and the Royal Australasian College

43 of Physicians. AKS, RE, JOD, JHO, and AT are supported by the Breast Cancer Research

44 Foundation. SZU is supported by both the Leukemia Lymphoma Society and International

45 Myeloma Society.

46

1 **Conflict of interests:**

2 SZU: Research funding: Amgen, BMS/Celgene, GSK, Janssen, Merck, Pharmacyclics, Sanofi,
3 Seattle Genetics, Takeda. Consulting/Advisory Board: Abbvie, Amgen, BMS, Celgene,
4 Genentech, Gilead, GSK, Janssen, Sanofi, Seattle Genetics, SecuraBio, SkylineDX, Takeda,
5 TeneoBio.

6

7 **Data availability**

8 The datasets used are available for download at <http://research.themmr.org>.

9

1 **ABSTRACT**

2

3 The plasma cell cancer multiple myeloma (MM) varies significantly in genomic characteristics,
4 response to therapy, and long-term prognosis. To investigate global interactions in MM, we
5 combined a known protein interaction network with a large clinically annotated MM dataset. We
6 hypothesized that an unbiased network analysis method based on large-scale similarities in gene
7 expression, copy number aberration, and protein interactions may provide novel biological
8 insights. Applying a novel measure of network robustness, Ollivier-Ricci Curvature, we examined
9 patterns in the RNA-Seq gene expression and CNA data and how they relate to clinical outcomes.
10 Hierarchical clustering using ORC differentiated high-risk subtypes with low progression free
11 survival. Differential gene expression analysis defined 118 genes with significantly aberrant
12 expression. These genes, while not previously associated with MM, were associated with DNA
13 repair, apoptosis, and the immune system. Univariate analysis identified 8/118 to be prognostic
14 genes; all associated with the immune system. A network topology analysis identified both hub
15 and bridge genes which connect known genes of biological significance of MM. Taken together,
16 gene interaction network analysis in MM uses a novel method of global assessment to demonstrate
17 complex immune dysregulation associated with shorter survival.

18

19 **STATEMENT OF SIGNIFICANCE**

20 Multiple myeloma has heterogenous clinical outcomes which are not well predicted by current
21 prognostic scoring systems. Global assessment of gene-protein interactions using Ollivier-Ricci
22 Curvature produces clusters of patients with defined prognostic significance, with high-risk groups
23 harboring complex gene dysregulation impacting immune function.

1 INTRODUCTION

2
3 The plasma cell cancer multiple myeloma (MM) has highly heterogenous clinical outcomes, with
4 a key determinant of response to treatment being genomic driver events. The most common
5 recurrent genomic events are hyperdiploidy, with a predominance of gains in chromosomes 3, 5,
6 7, 9, 11, 15, 19, and 21, and canonical chromosomal translocations affecting the immunoglobulin
7 heavy chain on chromosome 14 (1). MM harbors relatively few recurrent point mutations
8 compared with many other cancers, with only *NRAS*, *KRAS*, *TP53*, *FAM46C* and *DIS3* having a
9 prevalence above 10% (2).

10
11 Prognostic scoring updates have expanded the International Staging System (ISS) to incorporate
12 several chromosomal translocations [t(4;14), t(14;16)] and copy number aberrations (CNA;
13 deletion17p, gain/amplification1q), with each feature being considered as an individual event
14 (3,4). It has been recognized, however, that neither these features nor somatic mutations are
15 sufficient to define prognosis, with more extensive genomic assessments required to accurately
16 predict biological behavior.

17
18 Previous studies have described various genomic subtypes of MM using RNA-sequencing (RNA-
19 Seq) and/or CNA data (5–10). The subtypes identified by these methods tend to be dominated by
20 a single genomic event (i.e., hyperdiploidy, t(11;14), t(4;14), high proliferation index) or a
21 combination of previously described events (i.e., complex hyperdiploidy with gain1q and
22 monosomy 13) (9).

23

1 Here, we consider that integrating data from a comprehensive systems view, incorporating
2 complex interactions between multiple genes in a network, may define patterns of biological
3 behavior not captured by individual genomic events. Recently, a novel measure of network
4 robustness, Ollivier Ricci curvature (ORC), has been used to characterize breast and ovarian
5 cancers (11,12) and other pathological states (13). ORC measures the ability of a given connection
6 or interaction, between a pair of nodes — here being genes — to withstand perturbation,
7 considering both local and global connectivity in assessing the robustness of each pathway (see
8 **Methods** for a detailed description). In the context of cancer genomics, positive curvature implies
9 that there are multiple, robust active pathways for communication between the two genes. This
10 edge, or connection, can be described as “hub-like”. Negative curvature implies that if the
11 connection between two genes is impacted, the effect is relatively greater because of lack of direct
12 feedback controls; this edge can be considered “bridge-like”. Therefore, ORC analysis predicts the
13 effect of changes in gene expression within a wider network as opposed to just the individual gene.

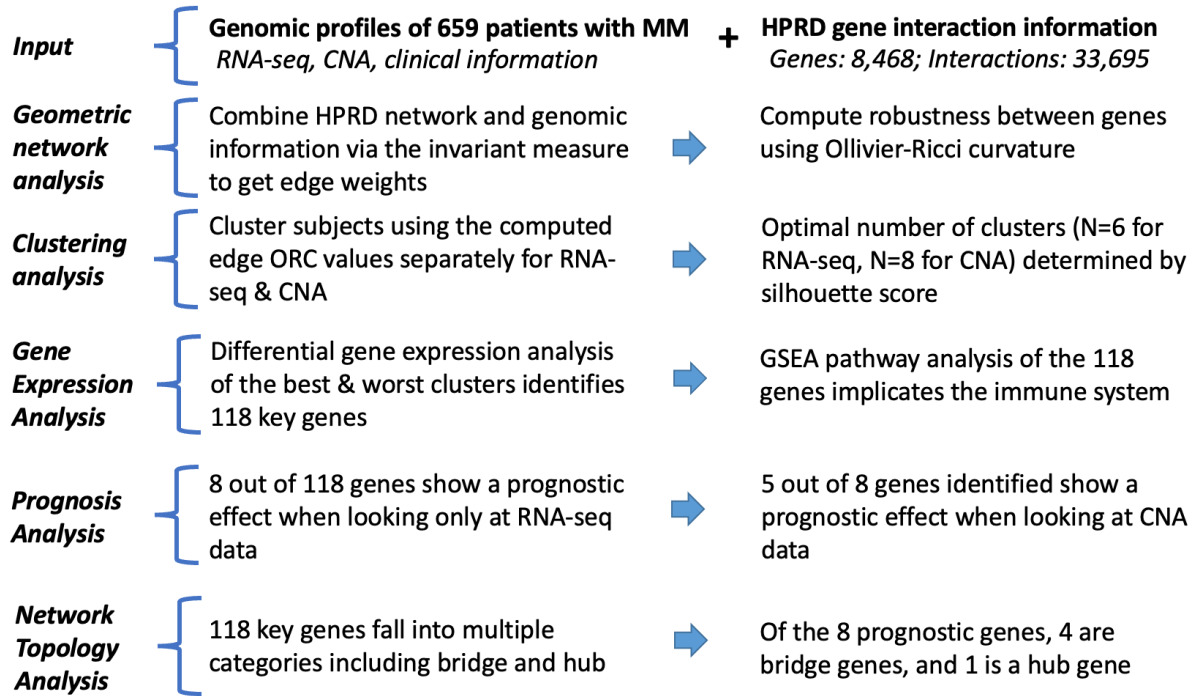
14
15 We utilize the ongoing Multiple Myeloma Research Foundation (MMRF) multi-site longitudinal
16 clinical registry study, which follows patients newly diagnosed with MM and collects both clinical
17 and genomic information periodically (9,14). The project, entitled CoMMpass (Relating Clinical
18 Outcomes in Multiple Myeloma to Personal Assessment of Genetic Profile), has over a thousand
19 patients enrolled in the latest interim analysis (IA19), and represents the largest publicly available
20 MM genomic data repository. The dataset includes clinical information, RNA sequencing (RNA-
21 Seq) information, copy number aberration (CNA), among others. To understand the relationship
22 between genes, we used a gene interactome derived from the Human Protein Reference Database
23 (HPRD) (15).

1
2 In this study, we apply an innovative geometric network analysis that integrates complex gene-
3 product interactions to characterize global patterns of MM biological behavior. Hierarchical
4 clustering defined groups of patients having different survival times, despite similar ISS
5 distributions. We identified 118 genes having significantly aberrant expression, most of which are
6 previously unassociated with MM, and 8 genes with prognostic capabilities which are part of the
7 immune system. These genes are not just hub genes, but bridge genes which help modulate
8 connections between two larger hub genes. Here, we demonstrate that protein-gene interaction
9 network analysis in MM demonstrates complex immune dysregulation which associates with
10 shorter survival.

11
12
13 **METHODS**

14
15 In this study, we perform a comprehensive geometric network analysis that integrates complex
16 gene-product interactions to characterize patterns in biological states. The methodology is
17 mathematically well-defined and has no fitting parameters, with an outline of the process
18 illustrated in **Figure 1**.

19



1
2 **Figure 1. Overview of the data processing pipeline.** This study uses a novel measure of network
3 robustness, Ollivier-Ricci curvature, to examine genes associated with shorter progression free
4 survival in multiple myeloma. RNA-Seq: RNA-sequencing; HPRD: Human Protein Reference
5 Database; CNA: copy number aberration; ORC: Ollivier-Ricci curvature; GSEA: gene set
6 enrichment analysis.

7
8 **Genomic data**

9 The MMRF CoMMpass dataset (release iteration: IA19), available to all researchers at
10 www.research.mmrf.org, includes clinical information, RNA-Seq gene expression, and CNA data
11 collected over time. Further information on the data collection and curation methods has
12 previously been published (9). For inclusion in this study, subjects must have RNA-Seq and CNA
13 data extracted from the bone marrow prior to the start of treatment and both demographic and
14 survival information. For the RNA-Seq data, the data provided by the MMRF was preprocessed
15 using the SALMON toolbox (16), included filtering unstranded immunoglobulin values, and was
16 normalized as transcripts per million (TPM) and log-transformed. For the CNA data, the data
17 provided by the MMRF was preprocessed using GATK (9).

1
2 Hyperdiploidy defined by more than 2 gains involving >60% of the chromosome affecting
3 chromosomes 3, 5, 7, 9, 11, 19 or 21. Mutational signatures were assessed using *mmsig*
4 (<https://github.com/UM-Myeloma-Genomics/mmsig>), a fitting algorithm designed specifically for
5 MM, to estimate the contribution of each mutational signature in each sample. APOBEC-
6 mutational activity was calculated by combining SBS2 and SBS13, with the top 10% being defined
7 as hyper-APOBEC (<https://cancer.sanger.ac.uk/signatures/sbs/>). The complex structural variant
8 chromothripsis was defined by manual curation according to previously published criteria (17).

9 10 ***Gene-product interaction data***

11 For network analysis on gene-product interactions, we used the curated network given by the
12 Human Protein Reference Database (HPRD) (15). The database consists of 9,600 genes and
13 notates 36,822 interaction pairs. We used the largest connected component of shared information
14 among the HPRD, RNA-Seq, and CNA data sets, which included 8427 of 9600 potential genes.

15 16 ***Ollivier Ricci curvature***

17 ORC integrates both local and global connectivity in assessing the robustness of each interaction
18 as characterized by the numerous feedback loops in a network modeled by a weighted graph or
19 Markov chain (18). Robustness, in this context, is defined as the ability of a system to return to its
20 original state following a perturbation. The ORC calculation is based on the ratio of an intrinsic
21 graph distance, capturing the metric properties of the network, to a distance defined via optimal
22 transport theory between the distributions of neighboring genomic values connected to a given
23 node. Capturing the sample-dependent pattern of curvature weighted edges provides a powerful

1 network-wide signature that integrates non-local information; illustrated in **Figure 2**, examples
2 zero, positive and negative curvature. ORC was calculated as per previous descriptions (11) and is
3 defined below.

4
5 Formally, ORC is defined as follows:
6

$$\kappa_{OR}(i, j) = 1 - \frac{W_1(\mu_i, \mu_j)}{d(i, j)}$$

7
8 where W_1 is the Wasserstein distance, also known as the Earth Mover's distance (EMD), between
9 the probability distributions, μ_i , μ_j . The probability distribution around a given node (gene), μ_i ,
10 is defined by the edge weights originating from the given node i to adjacent nodes as follows:

11

$$\mu_i(k) = \begin{cases} \frac{r_k}{\sum_{k \sim i} r_k} & k \sim i \\ 0 & k \not\sim i \end{cases}$$

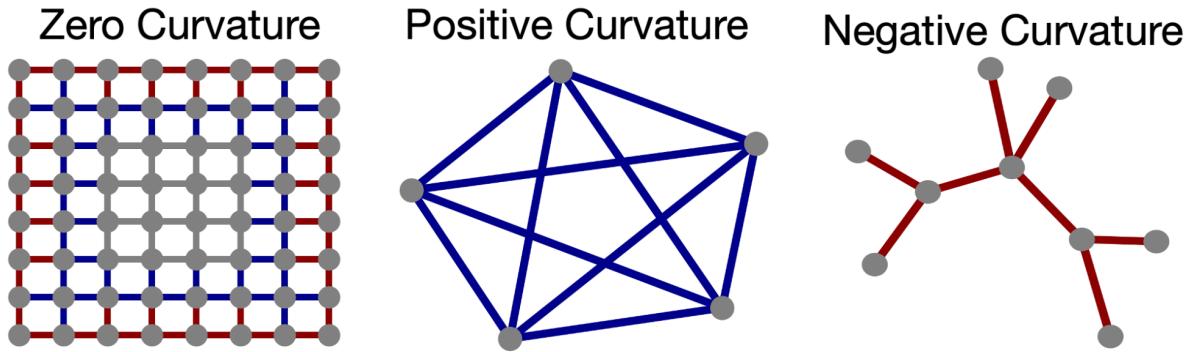
12
13
14

15 Where r_k indicates either RNA-Seq or CNA values in node k connected to node i . The
16 denominator $d(i, j)$ is the weighted shortest path between the two nodes, where the edge weights
17 of the weighted graph are derived from nodal values (RNA-Seq or CNA) quantifying the
18 information between two nodes and is formally defined below.

19

$$d(i, j) = \sum_{i \sim j} \frac{1}{\sqrt{w_{i,j}}}$$

20
21
22



1
2 **Figure 2. Ollivier Ricci curvature on example networks.** Gray edges indicate zero curvature
3 between nodes, blue edges indicate positive curvature, and red edges indicate negative curvature.
4 In the center image, there are multiple paths that can be traced out between any pair of nodes;
5 therefore, the curvature is positive. Conversely, the red edges in the right-most figure show
6 negative curvature values since the removal of any edge would bisect the graph.

7
8

9 ***Clustering analysis***

10 To explore the potential subtypes in the cohort, we used a hierarchical agglomerative clustering
11 method. For each data type, the RNA-Seq, CNA, and ORC matrices were separately clustered.
12 The number of clusters was determined by the silhouette score (19), a measure which takes into
13 account both the average intra-cluster distance and average nearest-cluster distance to determine
14 the optimal number of clusters. Survival analysis for progression free survival (PFS) was
15 performed using the Kaplan-Meier method and log-rank tests were used to determine statistical
16 significance. Multiple comparisons were corrected using the Benjamini Hochberg false discovery
17 rate (BH-FDR) (20).

18

19 ***Differential gene expression analysis***

20 To investigate biological differences between the identified subtypes, we conducted a differential
21 gene expression analysis between high and low-risk groups, as identified in prior steps, using RNA
22 sequencing read counts with DESeq2 (21). The p-values from this analysis were then BH-FDR

1 corrected. Genes with a corrected p-value less than 0.05 and an absolute log₂ fold change greater
2 than 3.5 were considered significant.

3

4 ***Pathway analysis***

5 Pathway analysis was performed using the Broad Institute's Gene Set Enrichment Analysis
6 (GSEA) tool (22,23). The utilized pathways are from the hallmark gene set collection from the
7 human molecular signatures database (MSigDB) (24). The fifty gene sets present different
8 biological states and processes identified using manual curation. Gene association with the
9 immune system was determined using ImmuneSigDB, an immune system pathways database
10 provided by GSEA (25).

11

12 ***Prognosis analysis***

13 To test whether or not an individual gene was prognostic, we used a Cox's proportional hazards
14 model (26) with the RNA-Seq data. The p-values from this analysis were corrected for multiple
15 hypothesis testing using BH-FDR. For genes that were significant with RNA-Seq, we repeated the
16 modeling analysis using CNA data.

17

18 ***Network topology analysis***

19 To understand how genes are connected to each other, a given gene's immediate neighbors are
20 visualized as a '1-hop plot.' Furthermore, a '2-hop plot' shows not only a gene's immediate
21 neighbors but also the nearest neighbors of the immediate neighbor genes, in order to contextualize
22 the relative portion of the overall network a given gene occupies. Bridge genes connect with

1 relatively few genes in the network, while hub genes form many connections relative to the rest of
2 the genes in the network.

3

4 ***Data and code availability***

5 The methods and instructions for how to use them are available for download at
6 www.github.com/aksimhal/mm-orc-subtypes. All data is available for download at
7 www.research.mmrf.org.

8

9

10

11 **RESULTS**

12

13 ***Patient cohort***

14

15 CoMMpass IA19 RNA-Seq and CNA data were available for 659 patients. The mean age in the
16 dataset was 62.5 ± 10.7 years; 60% were male, and the ISS distribution was 35% stage I, 35%
17 stage II, and 30% stage III. For the cohort, the five-year PFS rate was approximately 32%, with
18 the longest survival time listed at eight years. An overview is presented in **Supplementary Table**

19 **1.**

20

21 ***Hierarchical clustering using Ollivier-Ricci curvature differentiates subtypes with low*** 22 ***progression-free survival rates***

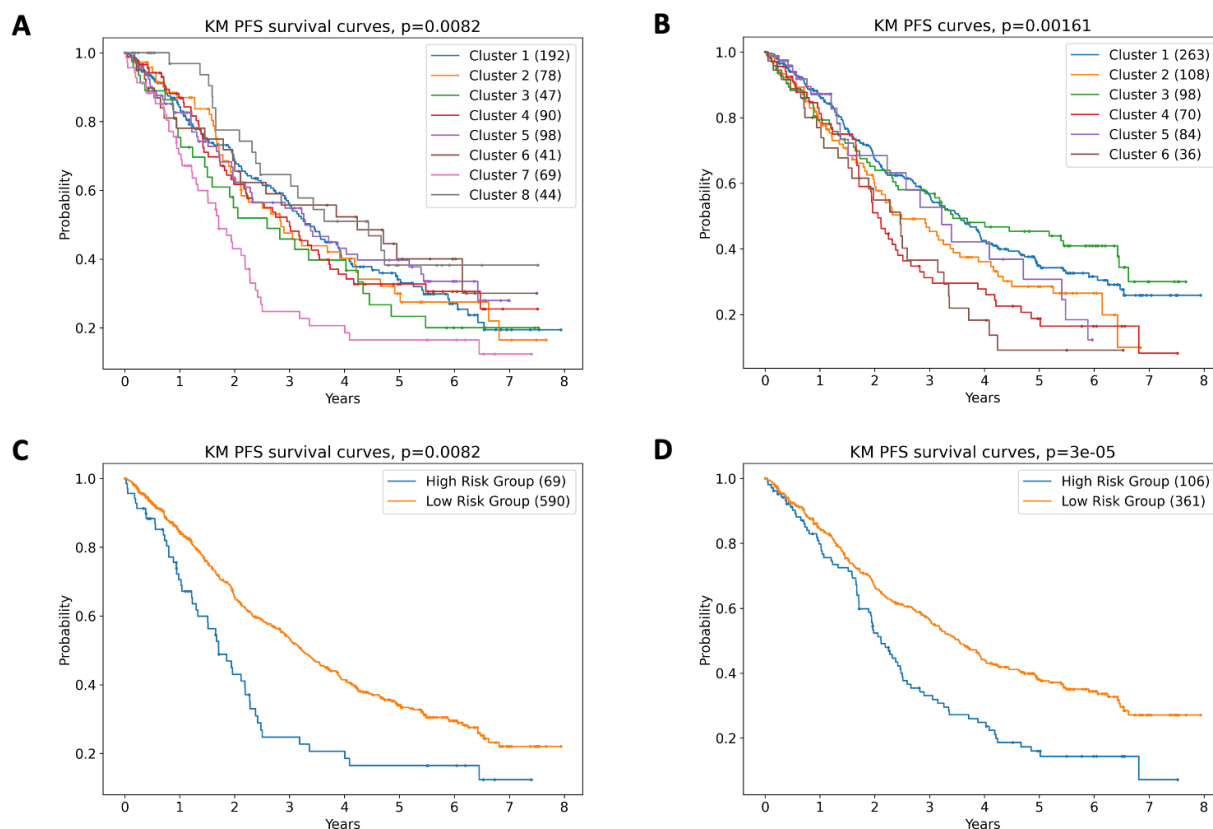
23

24 The largest connected network component from shared information between the HPRD, RNA-
25 Seq, and CNA data consisted of 8,468 nodes and 33,695 edges. ORC, a correlate for robustness of
26 strength between gene interaction pairs, was computed for each of the 33,695 interaction pairs in
27 each individual patient. Hierarchical clustering of the resultant ORC matrix together with CNA

1 data produced 8 clusters (**Supplementary Figure 1A, Figure 3A**), while clustering based on
2 RNA-Seq produced 6 clusters (**Supplementary Figure 1B, Figure 3B**); both methods being
3 significant for PFS (CNA; $p=0.0082$, RNA-Seq; $p=0.0016$, log-rank test). Interestingly, the
4 clustering appears to be defining biological differences not captured by the ISS prognostic score,
5 with a relatively even distribution of ISS stages in each cluster.

6
7 Considering the dominant impact of hyperdiploidy on CNA analyses, we repeated hierarchical
8 clustering on the non-hyperdiploid samples and found PFS prediction remained significant
9 ($p=0.0002$, log-rank test). Of note, analyzing CNA via ORC produced a cluster representing 10%
10 of patients with a markedly inferior PFS when compared to the remaining clusters (**Figures 3A,**
11 **3C**); median PFS was 1.7 years, despite only 35% of patients being ISS III. When assessing
12 previously described copy number risk factors (**Supplementary Table 2**), patients in this cluster
13 almost universally contain aberration in chr1q (gain; 57%, amplification; 29%, diploid 3%), while
14 also harboring the highest proportion of the complex structural variant chromothripsis (43% of
15 patients, $p<0.0001$ compared with the remaining clusters, Fisher's exact test). This finding is
16 congruent with previously published data demonstrating chromothripsis to be an independent
17 prognostic factor in MM (17), and with an increasing body of knowledge demonstrating that
18 multiple genomic insults compound to worse survival (17,27).

19



1
2 **Figure 3. Hierarchical clustering using Ollivier Ricci Curvature (ORC) predicts progression-**
3 **free survival (PFS) in multiple myeloma.** Kaplan-Meier analysis of PFS based on ORC
4 according to (A) copy number aberration, and (B) RNA sequencing. To better understand the
5 differences between the high risk and low risk cohorts, clusters with similar outcomes were
6 grouped. C) For CNA based clustering, clusters 1-6 and 8 were combined into the low-risk group.
7 Cluster 7 was the high-risk group. D) For RNA-sequencing data, clusters 4 and 6 were combined
8 into a high-risk group. Clusters 1 and 3 were combined into a low-risk group.

9
10 Clustering of the ORC matrix with RNA-Seq data produced more variation in PFS between
11 clusters (**Figure 3B**). Of note, clusters 2 and 3 contain the majority of t(11;14) patients
12 (**Supplementary Table 3**). Considering the dominant role of *CCND1* in MM pathophysiology,
13 we repeated hierarchical clustering in the non-t(11;14) samples, which remained significant for
14 PFS-prediction ($p=0.0002$, log-rank test). When clustering with all patients; 98% of those in cluster
15 4 harbor t(4;14), and 81% of those in cluster 6 have a translocation affecting *MAF*, *MAFA* or
16 *MAFB*, with 72% having increased APOBEC-mutational activity. Clusters 1 and 5 are more
17 heterogenous, with a combination of hyperdiploidy, canonical translocations, gain/ampl1q, *TP53*

1 aberration and chromothripsis. While a high proportion of patients in the 2 clusters with the
2 shortest PFS (4 and 6) carry a previously described genomic risk factor, the other clusters (1 and
3 3) demonstrate a longer PFS despite 29.2% being ISS III, and 34% harboring a risk factor included
4 in R-ISS / R2-ISS. Given that clustering with ORC using RNA-Seq demonstrated better
5 discrimination of PFS compared with CNA, we have elected to focus on RNA-Seq for the
6 remainder of the current study. We hypothesized that expanding on the ORC analysis with gene
7 set enrichment analysis (GSEA), prognostic modeling, and network topology analysis will provide
8 further biological insights.

9

10 *Expression analysis using ORC-based risk groups demonstrates differential DNA damage and*
11 *immune system signaling*

12

13 Differential gene expression analysis was conducted comparing high-risk (clusters 4 and 6) and
14 low-risk (clusters 1 and 3) as defined by ORC analysis of RNA-Seq data. Gene sets enriched in
15 the high-risk group includes inflammatory response, IL-6/JAK/STAT3 signaling and DNA
16 damage response (DDR) signaling (P53 pathway, DNA repair and apoptosis, **Table 1**). Of note,
17 there was no significant difference between the groups in p53 function by traditional methods
18 (*TP53* mutations and del17p), therefore our methods are capturing more global dysregulation in
19 DNA damage signaling than is evident by standard mutation and copy number analysis.

20

21

1 **Table 1. Differential gene expression analysis according to ORC-based risk groups.**

2 Directionality indicates the gene-set expression in the high-risk group compared with the low-
 3 risk group, with risk being defined by ORC of RNA-Seq data.

<i>Pathway</i>	<i>Genes</i>	<i>Q-value</i>	<i>Directionality</i>
Mitotic spindle	<i>BINI, GEMIN4, LATS1</i>	5.15e-3	Underexpressed
DNA repair	<i>ADA, CCNO, ERCC4, GTF2H5, NFX1, DCTN4</i>	9.16e-5	Overexpressed
IL6 JAK STAT3 signaling	<i>CCL7, JUN, IFNGR1, IL2RA</i>	1.52e-3	Overexpressed
Inflammatory response	<i>CCL7, KIF1B, MEPIA, PDPN, KCNJ2</i>	1.62e-3	Overexpressed
P53 pathway	<i>ADA, JUN, SAT1, PLK2, NOL8</i>	1.62e-3	Overexpressed
Apoptosis	<i>JUN, IFNGR1, SAT1, PAK1</i>	6.44e-3	Overexpressed

4
 5 Within these differentially expressed pathways, 118 genes were selected for further pathway
 6 analysis (having absolute log fold change > 3.5 and corrected p-value < 0.05, **Supplementary**
 7 **Table 4**). Of these 118 genes, 19 were under-expressed and 99 were overexpressed in the short
 8 survival group compared to the longer survival group in the poor survival group. In univariate
 9 analysis, 8/118 genes were predictors of PFS (*BUB1, MCMI, NOSTRIN, PAM, RNF115, SNCAIP,*
 10 *SPRR2A* and *WEE1*, **Table 2**), with 5 of these also being significant when analyzing based on
 11 CNA (*NOSTRIN, PAM, RNF115, SNCAIP* and *SPRR2A*). Interestingly, none of these genes feature
 12 in previously described lists of MM driver genes (27,28), suggesting that we are capturing novel
 13 aspects of MM biology. In addition to differential expression in the inflammatory response and
 14 IL-6/JAK/STAT3 signaling gene sets, interrogation of the ImmuneSigDB database demonstrated
 15 110 /118 genes to overlap with ImmuneSigDB pathways, including all 8 of the independently
 16 prognostic genes (**Table 2**). Taken together, these findings suggest that global assessment of gene
 17 interactions can detect complex immune dysregulation.

18

1 **Table 2. Gene expression in 8 novel immune-network genes associate with survival.**
 2 Coefficients less than 1 indicate a protective effect — associated with longer PFS. Coefficients
 3 greater than 1 indicate a detrimental effect — associated with a shorter PFS.

<i>Gene</i>	<i>Coefficient</i>	<i>95% - 105% Range</i>	<i>Q-value</i>	<i>Gene description</i>	<i>Number of ImmunoSigDB gene sets</i>
<i>BUB1</i>	1.36 ± 0.05	1.22-1.51	1.71e-8	BUB1 mitotic checkpoint serine/threonine kinase	5
<i>MCM6</i>	1.45 ± 0.07	1.27-1.66	6.19e-8	Minichromosome maintenance complex component 6	4
<i>NOSTRIN</i>	1.58 ± 0.11	1.27-1.98	4.49e-5	Nitric oxide synthase trafficking	1
<i>PAM</i>	0.72 ± 0.08	0.62-0.83	1.34e-5	Peptidylglycine alpha-amidating monooxygenase	7
<i>RNF115</i>	1.42 ± 0.11	1.14-1.77	1.72e-3	Ring finger protein 115	6
<i>SNCAIP</i>	1.40 ± 0.09	1.17-1.67	2.03e-4	Synuclein alpha interacting protein	1
<i>SPRR2A</i>	1.34 ± 0.05	1.22-1.46	1.43e-10	Small proline rich protein 2A	3
<i>WEE1</i>	1.32 ± 0.04	1.23-1.41	6.19e-15	WEE1 G2 checkpoint kinase	9

4
 5
 6 ***Local neighborhood 1-hop and 2-hop gene networks demonstrate differential DNA damage***
 7 ***and immune system signaling***

8
 9 A key feature of gene network analysis is the ability to capture a wide range of gene-pair
 10 interactions, above and beyond the expression levels of a single gene. While this analysis may be
 11 difficult to interpret in the context of highly connected genes, it can detect complex patterns (i.e.,
 12 an overall increase or decrease in network robustness) or specific individual interactions (i.e., a
 13 gene-pair demonstrating an increase in robustness while all other local connections become more
 14 fragile).

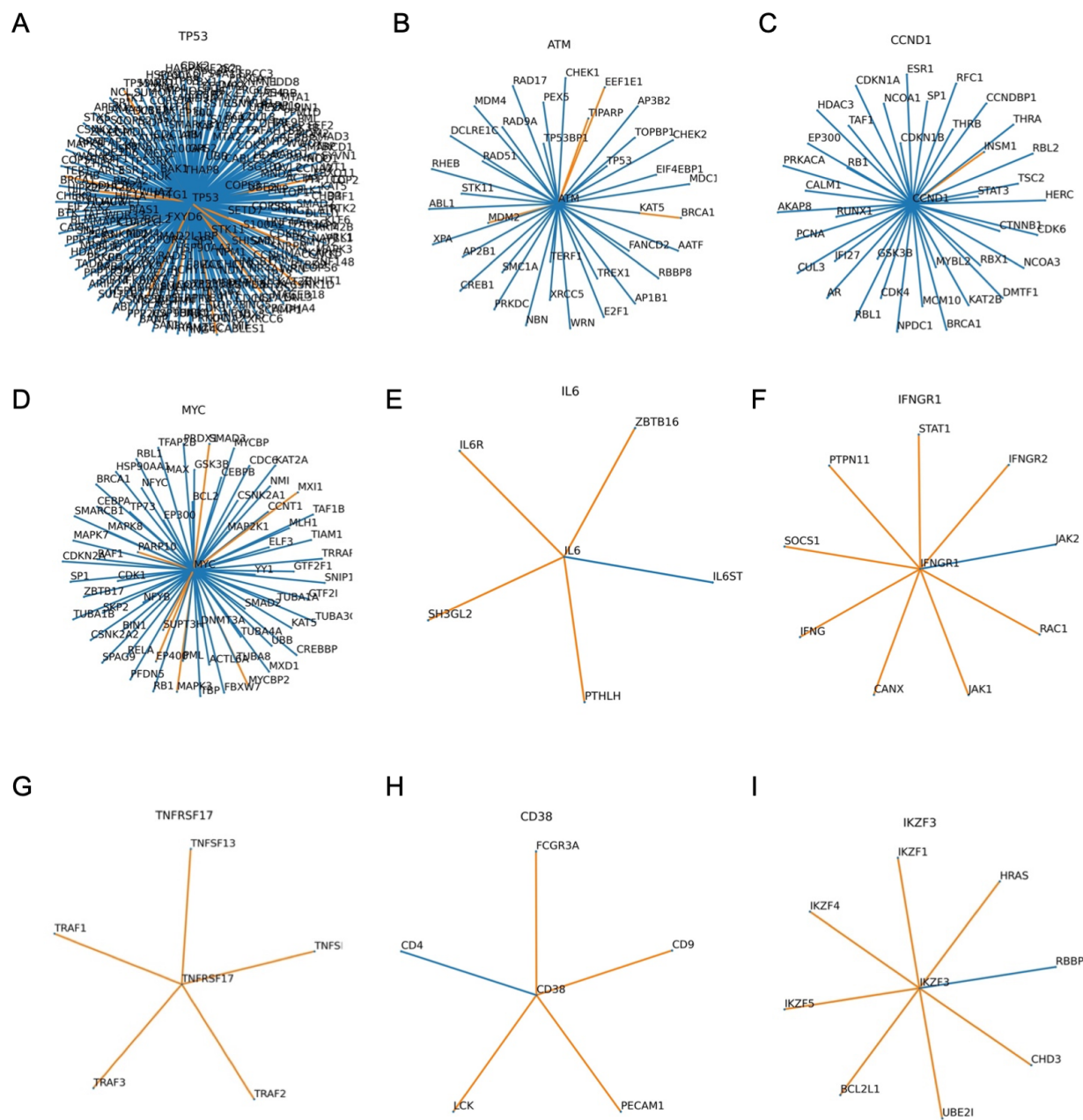
1
2 Comparing high-risk and low-risk clusters as defined by ORC analysis of RNA-Seq data, we note
3 several interesting network expression patterns. Within DDR-signaling, *TP53* and *ATM* signaling
4 pathways overwhelming become more robust in the high-risk group (**Figures 4A, 4B**), with more
5 robust pathways generally expected to exert increased effects. While we typically associate loss of
6 p53 function with poor prognosis in cancer, global network analysis is detecting global changes in
7 expression that may not fully capture functional protein levels. The same analysis performed on
8 the basis of CNA demonstrates a mixture of *TP53* connections becoming more robust and more
9 fragile, possibly reflecting the impact of del17p (**Supplementary Figure 2A**).

10
11 In addition to DDR-signaling, networks centered on *CCND1* and *MYC* become more robust overall
12 (**Figures 4C, 4D**), which suggests these signaling and transcriptional hubs remain dominant in the
13 context of high-risk disease. In contrast to the above networks showing a clear signal of robustness,
14 the effect on *RAF / RAS / MAPK* and *NFKB* signaling are more heterogenous (**Supplementary**
15 **Figure 2B-D**), suggesting that some parts of this network may play an oversized role in MM
16 biology compared with the other interactions.

17
18 Considering the immune dysregulation observed on GSEA analysis, signaling through some
19 cytokines and receptors become more fragile (i.e., IL-6, IFN γ ; **Figures 4E, 4F**), while others
20 demonstrate a more heterogenous effect (i.e., TNF, IFN α ; **Supplementary Figures 2F, 2G**). In
21 this context, pathways becoming more fragile would be expected to exert less than normal control.
22 Interestingly, multiple networks involving therapeutic targets for MM immune-based therapies
23 become more fragile, suggesting potential therapeutic vulnerabilities. This included *TNFRSF17*

1 (encoding for BCMA, a cellular-therapy target), *CD38* (the target of monoclonal antibody
2 daratumumab), *IZKF3* (a target of immunomodulatory agent lenalidomide) and *SLAMF7* (the
3 target of monoclonal antibody elotuzumab) (**Figures 4G-I, Supplementary Figures 2H, 2I**).

4
5
6

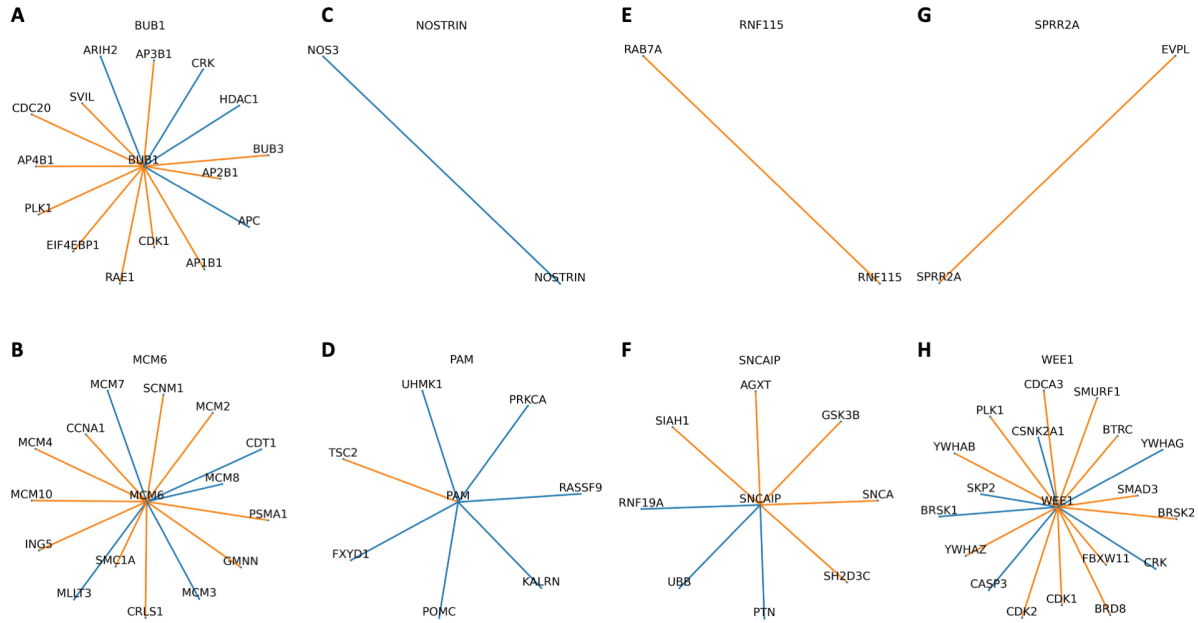


1
2 **Figure 4. Local neighborhood of selected genes relevant to MM biology and the immune**
3 **system.** Each line or edge represents the interaction between a gene-pair in a network, comparing
4 the median interactions observed in the high-risk group compared with those in the low-risk group.
5 Blue edges indicate that the connections are more robust in the high-risk group, while orange edges
6 are more fragile, risk being defined by the RNA-Seq-based clustering analysis. Higher resolution
7 images are available at www.github.com/aksimhal/mm-orc-subtypes.
8
9
10
11
12
13

1
2 From the list of 8 novel genes having expression associated with PFS in MM, all have a recognized
3 role in immune regulation (**Table 2**). In contrast with the other genes, only *WEE1*, (encoding for
4 a tyrosine kinase which affects G2-M transition), has been previously implicated in MM biology
5 (29). In the HPRD, *WEE1* acts as a hub gene, forming an above average number of connections
6 with its immediate neighbors (18 versus 8.4 for the whole graph). Interestingly, within the 8
7 prognostic genes, *BUB1* and *WEE1* connect to each other in a 2-hop analysis via *PLK1*, *CDK1*,
8 and *CRK*. From the genes with significantly different expression between risk groups, 24/118
9 (20.3%) connect to the 8 prognostic genes within the two-hop analysis.

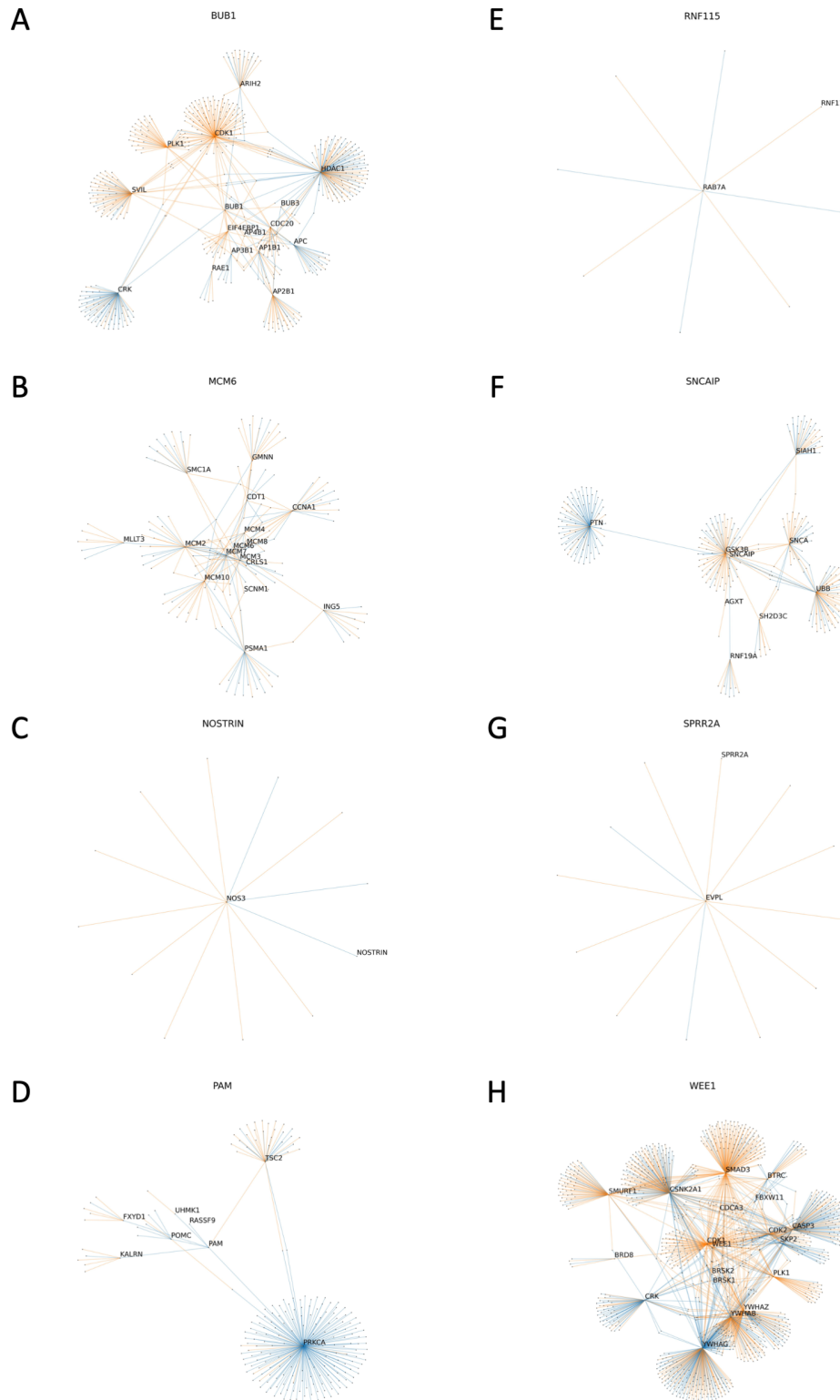
10
11 The 8 genes identified play different roles in their local neighborhoods (**Figure 5**); *NOSTRIN*, (a
12 nitric oxide synthase trafficker), *RNF115*, (an E3 ubiquitin ligase), and *SPRR2A* (induced by type-
13 2 cytokines in response to infection) form bridge-like connections to a single other gene. *NOSTRIN*
14 connects to another nitric oxide gene, *NOS3*, *RNF115* to the RAS oncogene family member
15 *RAB7A*, while *SPRR2A* connects with *EVPL* (associated with squamous cell cancer and
16 autoimmune disease). Four genes act as bridges for their local neighborhood: *BUB1*, *MCM6*, *PAM*,
17 and *SNCAIP* (**Figures 5, 6**). While these genes are not hub genes per se, they connect to multiple
18 hub genes and could therefore play a modulating role.

19



1
 2 **Figure 5. Local neighborhood of the eight genes identified as being predictive of PFS.** Each
 3 line or edge represents the interaction between a gene-pair in a network, comparing the median
 4 interactions observed in the high-risk group compared with those in the low-risk group. Blue edges
 5 indicate that the connections are more robust in the high-risk group, while orange edges are more
 6 fragile, risk being defined by the RNA-Seq-based clustering analysis.

7
 8



1
2 **Figure 6. ‘Two-hop’ neighborhood of the eight genes identified as being predictive of PFS.**
3 Each line or edge represents the interaction between a gene-pair in a network, comparing the
4 median interactions observed in the high-risk group compared with those in the low-risk group.

1 Blue edges indicate that the connections are more robust in the high-risk group, while orange edges
2 are more fragile, risk being defined by the RNA-Seq-based clustering analysis. Higher resolution
3 images are available at www.github.com/aksimhal/mm-orc-subtypes.

4
5
6 For example, in the 2-hop analysis, the mitotic checkpoint kinase *BUB1* connects to *HDAC1*
7 (**Figure 6A**), a histone deacetylase commonly upregulated in MM cells with a well-defined impact
8 on prognosis (30). We note multiple network connections between *BUB1* and *HDAC1*, as well as
9 connections between *BUB1* and each of *CDK1* (cell-cycle transition regulator) and *APC* (a tumor-
10 suppressor protein within the Wnt signaling pathway). *PAM*, encoding for a protein with multiple
11 functions described, connects to *PRKCA*, a protein kinase involved in regulation of proliferation,
12 tumorigenesis, and inflammation. Interestingly, the network connections around *PRKCA* are
13 predominantly more robust in the high-risk group. *SNCAIP*, (which inhibits ubiquitin ligase
14 activity), connects with *PTN* (**Figure 6F**; a hub gene encoding for a protein having a role in cell
15 survival, angiogenesis and tumorigenesis), previously noted to be elevated in MM patients (31).
16 Our analysis finds that the connection between *SNCAIP* and *PTN* becomes more robust in the high-
17 risk group. Interestingly, when comparing the 1- and 2-hop networks between RNA-Seq and CNA
18 data, several gene networks were highly analogous between the two methods (Supplementary
19 Figure 3).

20
21 Overall, the complex gene interactions captured through ORC analysis have the capacity to
22 significantly improve our understanding of biological differences between patients have short and
23 long survival, extending on what we understand from traditional mutation and copy number
24 analysis.

25

26

1 **DISCUSSION**

2
3 In order to investigate global gene-protein interaction networks in MM and their impact on
4 prognosis, we combined a known protein interaction network, HPRD, with a large MM dataset;
5 CoMMpass. We applied a novel measure of network robustness, ORC, to examine patterns in the
6 RNA-Seq gene expression and CNA data and how they relate to clinical outcomes. Hierarchical
7 clustering using ORC produced 6 clusters based on RNA-Seq and 8 clusters based on CNA data,
8 with both data sources predictive of PFS. Previously published genomic classifications in MM
9 based on RNA-Seq and/or CNA data have defined between four to twelve clusters, depending on
10 the data and analytical approach (5–10). To date, no study has integrated genomic information
11 with known protein interaction information in an analysis able to simultaneously integrate local
12 and global network information. By using techniques previously shown to uncover differences in
13 network strength in other domains, such as ovarian cancer and autism spectrum disorders (12,13),
14 we were able to demonstrate a new way of characterizing MM genomic data.

15
16 Our results demonstrate fidelity with known genomic risk factors (i.e., t(4;14), gain 1q, *TP53*
17 aberration) as well as emerging factors not yet in clinical use (i.e., APOBEC mutational activity
18 and the complex structural variant chromothripsis (17,32,33). While some genomic subgroups
19 were defined by a single event (i.e., 98% of RNA-Seq cluster 4 harboring t(4;14), the network
20 analysis approach produced other groups not previously described, with a combination of genomic
21 events defining prognostically significant clusters. It is notable that the cluster having the shortest
22 PFS was defined not by ISS, R-ISS, hyperdiploidy or IgH translocations but associated with the
23 combination of gain/amp 1q and chromothripsis. This finding supports the hypothesis that more
24 comprehensive, global genomic characterization is able to better define MM prognosis.

1
2 As ORC measures relative robustness between genes, GSEA analysis comparing high-risk and
3 low-risk groups as identified by ORC analysis of RNA-Seq data allowed exploration of gene-pair
4 interaction changes in robustness associated with survival differences between groups. GSEA
5 located 118 differentially expressed genes associated with six key biological pathways, five of
6 which were overexpressed in the group with the poor survival. The underexpressed pathway,
7 mitotic spindle assembly, has previously been reported to be associated with poor prognosis in
8 MM (34), while the overexpressed pathways were all associated with DNA damage response
9 (DDR) and acute phase inflammation / immune response. While del 17p is included in the R-ISS
10 prognostic score, and genomic complexity and instability are recognized features of high-risk MM
11 biology (35–38), there is not currently any immune component to routine prognostication of
12 NDMM patients. Furthermore, there is likely a biological link between the pathways we describe,
13 with an inflammatory hypoxic microenvironment potentially contributing to aberrant DDR (39),
14 and functional high-risk patients who relapse within 12 months described to harbor both mutations
15 affecting the IL-6/JAK/STAT pathway and abnormal gene expression associated with mitosis /
16 DDR (40).

17
18 Univariate analysis of the 118 differentially expressed genes identified 8 prognostic genes which
19 are all associated with immune function according to ImmunoSigDB. Network topology analysis
20 identified most of these 8 to be bridge genes, connecting to genes known to have biological impact
21 in MM (i.e., *HDAC1*, *CDK1*, *PRKCA* and *PTN*). The near-neighbor and 2-hop gene topology
22 networks capture more global gene dysregulation, potentially missed in single-gene expression

1 analysis. Our results may also suggest a new set of therapeutic targets to further investigate in
2 high-risk MM patients.

3

4 Considering possible limitations; while CoMMpass represents the largest multi-site, international
5 genomic MM dataset compiled to date, it does contain patients who received drug regimens no
6 longer in common usage, and a low proportion of patients receiving the most potent modern
7 regimens. Ideally our methods would be applied to datasets including daratumumab- based
8 induction therapy. Considering possible extension of our analytical methods: while the choice of
9 using the HPRD as the protein interaction network is common in literature (41), other networks,
10 such as STRING (42), may provide complementary results. Finally, no network analysis method
11 represents the ‘gold standard’, and it is plausible that other clustering and network analysis
12 methods may provide alternative results. Future studies may consider whether or not the 118 genes
13 associated with high-risk individuals are dysregulated at precursor MM stages, and how the
14 expression of these genes is altered in response to treatment.

15

16

1 **References**

- 2
- 3 1. Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. *Nat Rev*
- 4 *Cancer*. 2012 Apr 12;12(5):335–48.
- 5 2. Hu Y, Chen W, Wang J. Progress in the identification of gene mutations involved in
- 6 multiple myeloma. *Onco Targets Ther*. 2019 May 24;12:4075–80.
- 7 3. Palumbo A, Avet-Loiseau H, Oliva S, Lokhorst HM, Goldschmidt H, Rosinol L, et al.
- 8 Revised International Staging System for multiple myeloma: A report from international
- 9 myeloma working group. *J Clin Oncol*. 2015 Sep 10;33(26):2863–9.
- 10 4. D’Agostino M, Cairns DA, Lahuerta JJ, Wester R, Bertsch U, Waage A, et al. Second
- 11 Revision of the International Staging System (R2-ISS) for overall survival in multiple
- 12 myeloma: A European Myeloma Network (EMN) report within the HARMONY project. *J*
- 13 *Clin Oncol*. 2022 Oct 10;40(29):3406–18.
- 14 5. Zhan F, Huang Y, Colla S, Stewart JP, Hanamura I, Gupta S, et al. The molecular
- 15 classification of multiple myeloma. *Blood*. 2006 Sep 15;108(6):2020–8.
- 16 6. Chng WJ, Kumar S, Vanwier S, Ahmann G, Price-Troska T, Henderson K, et al. Molecular
- 17 dissection of hyperdiploid multiple myeloma by gene expression profiling. *Cancer Res*.
- 18 2007 Apr 1;67(7):2982–9.
- 19 7. Broyl A, Hose D, Lokhorst H, de Knecht Y, Peeters J, Jauch A, et al. Gene expression
- 20 profiling for molecular classification of multiple myeloma in newly diagnosed patients.
- 21 *Blood*. 2010 Oct 7;116(14):2543–53.
- 22 8. Jang JS, Li Y, Mitra AK, Bi L, Abyzov A, van Wijnen AJ, et al. Molecular signatures of
- 23 multiple myeloma progression through single cell RNA-Seq. *Blood Cancer J*. 2019 Jan
- 24 3;9(1):2.
- 25 9. Skerget S, Penaherrera D, Chari A, Jagannath S, Siegel DS, Vij R, et al. Genomic Basis of
- 26 Multiple Myeloma Subtypes from the MMRF CoMMpass Study [Internet]. *bioRxiv*.
- 27 medRxiv; 2021. Available from:
- 28 <http://medrxiv.org/lookup/doi/10.1101/2021.08.02.21261211>
- 29 10. Bustoros M, Anand S, Sklavenitis-Pistofidis R, Redd R, Boyle EM, Zhitomirsky B, et al.
- 30 Genetic subtypes of smoldering multiple myeloma are associated with distinct pathogenic
- 31 phenotypes and clinical outcomes. *Nat Commun*. 2022 Jun 15;13(1):3449.
- 32 11. Sandhu RS, Georgiou TT, Reznik E, Zhu L, Kolesov I, Senbabaoglu Y, et al. Graph
- 33 curvature for differentiating cancer networks. *Sci Rep*. 2015;5:12323.
- 34 12. Elkin R, Oh JH, Liu YL, Selenica P, Weigelt B, Reis-Filho JS, et al. Geometric network
- 35 analysis provides prognostic information in patients with high grade serous carcinoma of
- 36 the ovary treated with immune checkpoint inhibitors. *NPJ Genom Med*. 2021 Nov
- 37 24;6(1):99.

- 1 13. Simhal AK, Carpenter KLH, Kurtzberg J, Song A, Tannenbaum A, Zhang L, et al. Changes
2 in the geometry and robustness of diffusion tensor imaging networks: Secondary analysis
3 from a randomized controlled trial of young autistic children receiving an umbilical cord
4 blood infusion. *Front Psychiatry* [Internet]. 2022;13. Available from:
5 <https://www.frontiersin.org/articles/10.3389/fpsy.2022.1026279>
- 6 14. Keats JJ, Craig DW, Liang W, Venkata Y, Kurdoglu A, Aldrich J, et al. Interim analysis of
7 the mmrf CoMMpass trial, a longitudinal study in multiple myeloma relating clinical
8 outcomes to genomic and immunophenotypic profiles. *Blood*. 2013 Nov 15;122(21):532–
9 532.
- 10 15. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, Muthusamy B, et al.
11 Human protein reference database as a discovery resource for proteomics. *Nucleic Acids*
12 *Res*. 2004 Jan 1;32(Database issue):D497-501.
- 13 16. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-
14 aware quantification of transcript expression. *Nat Methods*. 2017 Apr;14(4):417–9.
- 15 17. Rustad EH, Yellapantula VD, Glodzik D, Maclachlan KH, Diamond B, Boyle EM, et al.
16 Revealing the impact of structural variants in multiple myeloma. *Blood Cancer Discov*.
17 2020 Nov;1(3):258–73.
- 18 18. Ollivier Y. Ricci curvature of metric spaces. *C R Math*. 2007;345(11):643–6.
- 19 19. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster
20 analysis. *J Comput Appl Math*. 1987 Nov 1;20:53–65.
- 21 20. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful
22 approach to multiple testing. *J R Stat Soc* [Internet]. 1995; Available from:
23 <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x>
- 24 21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for
25 RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
- 26 22. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene
27 set enrichment analysis: a knowledge-based approach for interpreting genome-wide
28 expression profiles. *Proc Natl Acad Sci U S A*. 2005 Oct 25;102(43):15545–50.
- 29 23. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-
30 1alpha-responsive genes involved in oxidative phosphorylation are coordinately
31 downregulated in human diabetes. *Nat Genet*. 2003 Jul;34(3):267–73.
- 32 24. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
33 Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015 Dec
34 23;1(6):417–25.
- 35 25. Godec J, Tan Y, Liberzon A, Tamayo P, Bhattacharya S, Butte AJ, et al. Compendium of
36 immune signatures identifies conserved and species-specific biology in response to

- 1 inflammation. *Immunity*. 2016 Jan 19;44(1):194–206.
- 2 26. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds
3 models for censored survival data, with application to prognostic modelling and estimation
4 of treatment effects. *Stat Med*. 2002 Aug 15;21(15):2175–97.
- 5 27. Walker BA, Mavrommatis K, Wardell CP, Ashby TC, Bauer M, Davies F, et al. A high-
6 risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis.
7 *Leukemia*. 2019 Jan;33(1):159–70.
- 8 28. Maura F, Bolli N, Angelopoulos N, Dawson KJ, Leongamornlert D, Martincorena I, et al.
9 Genomic landscape and chronological reconstruction of driver events in multiple myeloma.
10 *Nat Commun*. 2019 Aug 23;10(1):3835.
- 11 29. Barbosa RSS, Dantonio PM, Guimarães T, de Oliveira MB, Fook Alves VL, Sandes AF, et
12 al. Sequential combination of bortezomib and WEE1 inhibitor, MK-1775, induced
13 apoptosis in multiple myeloma cell lines. *Biochem Biophys Res Commun*. 2019 Nov
14 12;519(3):597–604.
- 15 30. Mithraprabhu S, Kalff A, Chow A, Khong T, Spencer A. Dysregulated Class I histone
16 deacetylases are indicators of poor prognosis in multiple myeloma. *Epigenetics*. 2014
17 Nov;9(11):1511–20.
- 18 31. Yeh HS, Chen H, Manyak SJ, Swift RA, Campbell RA, Wang C, et al. Serum pleiotrophin
19 levels are elevated in multiple myeloma patients and correlate with disease status. *Br J*
20 *Haematol*. 2006 Jun;133(5):526–9.
- 21 32. Walker BA, Wardell CP, Murison A, Boyle EM, Begum DB, Dahir NM, et al. APOBEC
22 family mutational signatures are associated with poor prognosis translocations in multiple
23 myeloma. *Nat Commun*. 2015 Apr 23;6:6997.
- 24 33. Maura F, Petljak M, Lionetti M, Cifola I, Liang W, Pinatel E, et al. Biological and
25 prognostic impact of APOBEC-induced mutations in the spectrum of plasma cell dyscrasias
26 and multiple myeloma cell lines. *Leukemia*. 2018 Apr;32(4):1044–8.
- 27 34. Tao Y, Yang G, Yang H, Song D, Hu L, Xie B, et al. TRIP13 impairs mitotic checkpoint
28 surveillance and is associated with poor prognosis in multiple myeloma. *Oncotarget*. 2017
29 Apr 18;8(16):26718–31.
- 30 35. Kassambara A, Gourzones-Dmitriev C, Sahota S, Rème T, Moreaux J, Goldschmidt H, et
31 al. A DNA repair pathway score predicts survival in human multiple myeloma: the potential
32 for therapeutic strategy. *Oncotarget*. 2014 May 15;5(9):2487–98.
- 33 36. Ali JYH, Fitieh AM, Ismail IH. The Role of DNA Repair in Genomic Instability of Multiple
34 Myeloma. *Int J Mol Sci [Internet]*. 2022 May 19;23(10). Available from:
35 <http://dx.doi.org/10.3390/ijms23105688>
- 36 37. Giesen N, Paramasivam N, Toprak UH, Huebschmann D, Xu J, Uhrig S, et al.

- 1 Comprehensive genomic analysis of refractory multiple myeloma reveals a complex
2 mutational landscape associated with drug resistance and novel therapeutic vulnerabilities.
3 *Haematologica*. 2022 Aug 1;107(8):1891–901.
- 4 38. Maura F, Boyle EM, Rustad EH, Ashby C, Kaminetzky D, Bruno B, et al. Chromothripsis
5 as a pathogenic driver of multiple myeloma. *Semin Cell Dev Biol*. 2022 Mar;123:115–23.
- 6 39. Saitoh T, Oda T. DNA damage response in multiple myeloma: The role of the tumor
7 microenvironment. *Cancers (Basel)*. 2021 Jan 28;13(3):504.
- 8 40. Soekojo CY, Chung T-H, Furqan MS, Chng WJ. Genomic characterization of functional
9 high-risk multiple myeloma patients. *Blood Cancer J*. 2022 Jan 31;12(1):24.
- 10 41. Wu J, Vallenius T, Ovaska K, Westermarck J, Mäkelä TP, Hautaniemi S. Integrated
11 network analysis platform for protein-protein interactions. *Nat Methods*. 2009 Jan;6(1):75–
12 7.
- 13 42. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of
14 predicted functional associations between proteins. *Nucleic Acids Res*. 2003 Jan
15 1;31(1):258–61.
- 16