PROTEIN NETWORK ANALYSIS UNCOVERS A POOR-SURVIVAL SUBTYPE IN MULTIPLE MYELOMA

Anish K. Simhal[1]*, Kylee H. Maclachlan[2]*, Rena Elkin[1], Jiening Zhu[3], Saad Z. Usmani[2], Jonathan J. Keats[4], Larry Norton[5], Joseph O. Deasy[1], Jung Hun Oh[1], Allen Tannenbaum[3,6]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY. [2]Myeloma Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.
[3]Department of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY. [4]Translational Genomics Research Institute (TGen), Phoenix, AZ.
[5]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY. [6]Department of Computer Science, Stony Brook University, Stony Brook, NY

* simhala@mskcc.org, maclachk@mskcc.org

INTRODUCTION

- It is well-known that multiple myeloma (MM) subtypes vary significantly in terms of prognosis, response to therapy, and genomic characteristics.
- To date, no study has integrated genomic information with known protein interaction information in an analysis to simultaneously integrate local and global network information.
- We hypothesized that novel, unbiased analyses based on large scale similarities of mRNA, copy number alteration (CNA) characteristics, and the resulting information processing alterations of MM might expose useful new subtype classifications as well as unappreciated therapeutic targets.

KEY RESULTS

- Ollivier-Ricci curvature based network analysis identified high-risk subtypes.
- Differential gene expression analysis defined 118 genes associated with poor outcomes.
- Univariate analysis identified 8/118 to be prognostic genes; all associated with the immune system.
- A network topology analysis identified both hub and bridge genes which connect known genes of biological significance of MM.

SAMPLE INFORMATION

- Multiple Myeloma Research Foundation's CoMMpass dataset, release IA19.
- 659 newly diagnosed MM patients with both baseline RNA-seq and CNA data were included.

	Patients (N=659)
Sex	Male: 393 Female: 266
Age (mean ± SD)	61.58 ± 13.09 years
ISS	Stage I: 229 Stage II: 232 Stage III: 179
Treatment class	Bortezomib-based: 138 Carfilzomib-based: 38 IMIDs-based: 32 Combined bortezomib/IMIDs-based: 317 Combined IMIDs/carfilzomib-based: 110 Other: 24

Subject characteristics at baseline

METHODS

Ollivier Ricci Curvature (ORC)

- ORC integrates both local and global connectivity in assessing the robustness of each pathway as characterized by a numerous feedback loops in a network.
- Robustness is the ability of a system to return to its original state following a perturbation.
- The higher the curvature value between two nodes (genes), the more likely to have more feedback loops connecting them.

Ollivier Ricci curvature on example networks. Blue edges indicate positive curvature between nodes. Black edges indicate zero curvature, and red edges indicate negative curvature between nodes.

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

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

Curvature formulation where k(i,j) is the curvature measure between nodes i and j. W_1 is the Wasserstein distance between the probability distributions.

Study summary

Input	**Genomic profiles of 659 patients with MM** *RNA-seq, CNA, clinical information*	+	**HPRD gene interaction information** *Genes: 8,468; Interactions: 33,695*
Geometric network analysis	Combine HPRD network and genomic information via the invariant measure to get edge weights	⇨	Compute robustness between genes using Ollivier-Ricci curvature
Clustering analysis	Cluster subjects using the computed edge ORC values separately for RNA-seq & CNA	⇨	Optimal number of clusters (N=6 for RNA-seq, N=8 for CNA) determined by silhouette score
Gene Expression Analysis	Differential gene expression analysis of the best & worst clusters identifies 118 key genes	⇨	GSEA pathway analysis of the 118 genes implicates the immune system
Prognosis Analysis	8 out of 118 genes show a prognostic effect when looking only at RNA-seq data	⇨	5 out of 8 genes identified show a prognostic effect when looking at CNA data
Network Topology Analysis	118 key genes fall into multiple categories including bridge and hub	⇨	Of the 8 prognostic genes, 4 are bridge genes, and 1 is a hub gene

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

HIGH RISK DIFFERENCES

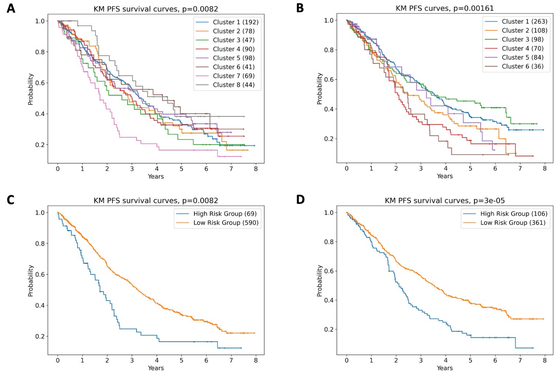
Biomarker	*Low-risk (N=361)*	*High-risk (N=106)*	*P-value*
Age (years, mean ± SD)	62.8 ± 10.5	61.1 ± 10.7	0.156
Sex	M: 214, F: 147	M: 63; F: 43	1.0
ISS	1: 138, 2: 114, 3: 100	1: 26, 2: 47, 3: 28	0.0153
Hyperdiploidy	206/311	17/93	1.32e-16
t(4;14)	0/332	66/97	4.14e-54
t(11;14)	80/332	0/97	1.41e-10
MAF/MAFB translocation	1/332	28/97	6.54e-19
MYC translocation	65/332	11/97	6.97e-2
Chromothripsis	69/332	40/97	1.01e-4
APOBEC activity	6/330	24/97	5.76e-12
TP53 inactivation (1 monoallelic, 2 biallelic)	(0) 238; (1) 29; (2) 9	(0) 71; (1) 11; (2) 5	0.472
Gain 1q21 (1 gain, 2 amplification)	(0) 226; (1) 76; (2) 9	(0) 45; (1) 30; (2) 18	9.29e-9

Differences in multiple myeloma biomarkers between the better and worse PFS outcomes as clustered with RNA-seq data.

Outcomes are influenced by genetic factors, rather than clinical factors.

Hyperdiploidy and t(11;14) have protective effects; t(4;14), MAF, APOBEC activity, and chromothripsis have detrimental effects.

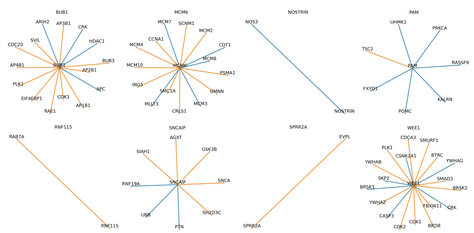
NETWORK RESULTS



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

Differential gene expression analysis according to ORC-based risk groups. *Directionality indicates the gene-set expression in the high-risk group compared with the low-risk group, with risk being defined by ORC of RNA-Seq data.*

Pathway	Genes	Q-value	Directionality
Mitotic spindle	BIN1, GEMIN4, LATS1	5.15e-3	Underexpressed
DNA repair	ADA, CCNO, ERCC4, GTF2H5, NFX1, DCTN4	9.16e-5	Overexpressed
IL6 JAK STAT3 signaling	CCL7, JUN, IFNGR1, IL2RA	1.52e-3	Overexpressed
Inflammatory response	CCL7, KIF1B, MEP1A, PDPN, KCNJ2	1.62e-3	Overexpressed
P53 pathway	ADA, JUN, SAT1, PLK2, NOL8	1.62e-3	Overexpressed
Apoptosis	JUN, IFNGR1, SAT1, PAK1	6.44e-3	Overexpressed



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

CONCLUSIONS

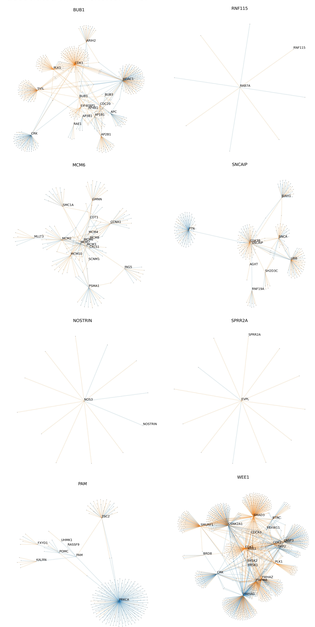
- The high-risk groups were defined not by ISS, R-ISS, hyperdiploidy or IgH translocations but associated with the combination of gain/amp 1q and chromothripsis. This finding supports the hypothesis that more comprehensive, global genomic characterization can better define MM prognosis.
- Overall, the complex gene interactions captured through ORC analysis have the capacity to significantly improve our understanding of biological differences between patients have short and long survival, extending on what we understand from traditional mutation and copy number analysis.

Want to learn more? Scan this QR code for a link to our manuscript on biorxiv.org.

PROGNOSTIC GENES

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

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



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

In contrast with the other genes, only WEE1, (encoding for a tyrosine kinase which affects G2-M transition), has been previously implicated in MM biology. In the HPRD, WEE1 acts as a hub gene, forming an above average number of connections with its immediate neighbors (18 versus 8.4 for the whole graph). Interestingly, within the 8 prognostic genes, BUB1 and WEE1 connect to each other in a 2-hop analysis via PLK1, CDK1, and CRK.

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

ACKNOWLEDGEMENTS

The authors would like to thank the team at the Multiple Myeloma Research Foundation for their incredible work and support with the CoMMpass dataset. KHM received support from the Multiple Myeloma Research Foundation, the American Society of Hematology, and the Royal Australasian College of Physicians. AKS, RE, JOD, JHO, and AT are supported by the Breast Cancer Research Foundation. SZU is supported by both the Leukemia Lymphoma Society and International Myeloma Society.