GEOMETRIC NETWORK ANALYSIS DEFINES POOR-PROGNOSIS SUBTYPES 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

INTRODUCTION

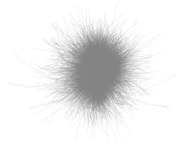
It is well-known that multiple myeloma (MM) subtypes vary significantly in terms of prognosis, response to therapy, and genomic characteristics. We hypothesized that novel, unbiased analyses based on large scale similarities of mRNA, copy number alteration (CNA) characteristics, and the resulting information processing alterations of MM might expose useful new subtype classifications as well as unappreciated therapeutic targets.

AIM

- Identify high-risk MM subtypes by synthesizing protein interaction information with RNA-seq data.
- Use a novel method of graph robustness, Ollivier-Ricci curvature, to identify weak points in the graph.
- Contextualize network differences in the context of hallmark pathways.

INTERACTOME

- To understand the interactions between genes, we used the Human Protein Reference Database (HPRD), a manually curated database.
- After filtering for genes in both the HPRD and CoMMpass dataset, the resulting interactome had 8,488 genes with 33,695 interactions.
- Average degree centrality: 8.4 ± 14.9

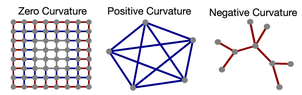


Cytoscape based visualization of the interactome was used. The un-interpretability of this figure highlights the need for comprehensive network analysis techniques.

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.



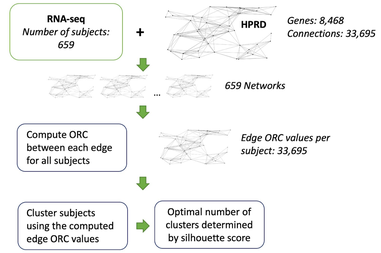
Ollivier Ricci curvature on example networks. Blue edges indicate positive curvature between nodes. Black edges indicate zero curvature, and red edges indicate negative curvature between nodes.

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

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

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.

ORC-based clustering analysis pipeline



SAMPLE INFORMATION

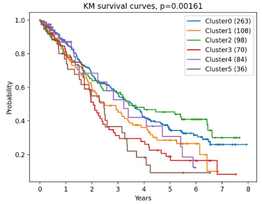
- Multiple Myeloma Research Foundation's CoMMpass dataset, release IA19.
- 659 subjects had both baseline RNA-seq and CNA data.

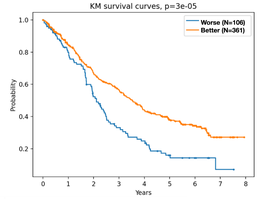
	Patients (N=659)
Sex	Male: 393 Female: 266
Age (mean ± SD)	61.58 ± 13.09 years
ISS	Stage I: 229 Stage II: 232 Stage III: 179
Treatment class	Bortezomib-based: 138 Carfilzomib-based: 38 IMIDs-based: 32 Combined bortezomib/IMIDs-based: 317 Combined IMIDs/carfilzomib-based: 110 Other: 24

Subject characteristics at baseline

RNA-SEQ RESULTS

- RNA-seq based clustering analysis revealed six unique clusters





Progression free survival curves of the RNA-seq clustering results. When combining the high-risk versus low-risk groups, the five-year survival rates diverged significantly.

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.

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

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

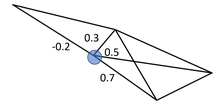
Differential gene expression pathway analysis

- Evaluated using DESEQ2. 118 genes remained significant after FDR correction and had an absolute log2 fold change > 3.5.
- 19 genes were under expressed in the high-risk group compared to the low-risk group.
- Pathway analysis was performed using the Broad Institute's Gene Set Enrichment Analysis tool.
- Pathways used are from the hallmark gene set collection from the human molecular signatures database (MSigDB).

SCALAR ORC RESULTS

Scalar curvature analysis

- Scalar curvature is defined as the summation of edge curvature values originating in the node (gene) of interest.
- The difference in the average scalar curvature between the poor-prognosis group versus the good-prognosis group reflects changes in robustness irrespective of the protein interaction network.
- Genes were ranked based on the difference in scalar curvature and a pathway analysis was done for the 100 genes which became the most robust and 100 genes which became the most fragile.



The scalar curvature value around the highlighted node is the sum of edges originating from it: 1.4

Pathway	*Genes*	*Q-value*
Apoptosis	CAV1, NEDD9, CASP3, MMP2, PTK2, CCND1, RELA, GNA15, BCL2L1, CASP8, PAK1	1.26e-11
UV downregulated response	CAV1, DLG1, MYC, MET, COL1A1, PRKCA, ATXN1, KIT	6.72e-8
PI3K AKT MTOR signaling	GRB2, PTPN11, PLCG1, AKT1, MAPK1, FGF6, FGF17	1.29e-7
Mitotic spindle	NEDD9, DLG1, LMNB1, ABL1, BCAR1, AKAP13, CTTN, BCR	3.52e-7
Complement pathway	CASP3, GRB2, SRC, JAK2, FN1, C1QA, CASP10, CBLB	3.52e-7
IL6 JAK STAT3 signaling	GRB2, PTPN11, A2M, CSF2RB, CBL, ITGA4	6.47e-7
Apical junction	MMP2, PTK2, DLG1, PLCG1, SRC, ACTA1, ARHGEF6	4.6e-6
Estrogen early response	CCND1, MYC, JAK2, KRT8, BCL2, GAB2	5.64e-5
Inflammatory response	RELA, GNA15, MYC, MET, ABCA1, AXL	5.64e-5
Coagulation pathway	MMP2, FN1, C1QA, A2M, F12	1.17e-4

*Top ten hallmark pathways associated with the top 100 genes which became **more robust** in the high-risk group defined by the RNA-seq data. Pathways becoming more robust would generally be expected to exert increased effect.*

Pathway	*Genes*	*Q-value*
Apoptosis	BMP2, FU, CLU, CD44, FAS, CCND2, ATF3, BRCA1, KRT18, APP, CDK2, CREBBP	3.44e-13
TGF BETA signaling	BMP2, ID2, SMAD3, NCOR2, SKIL, SMURF1, ACVR1	1.67e-9
Complement pathway	F2, CLU, PLAUR, CFH, CCL5, CEBPB, LYN, FYN, HSPA1A	1.41e-8
Epithelial mesenchymal transition	CD44, FAS, ID2, PLAUR, IGFBP3, FLNA, ITGB1, FGF2, APLP1	1.41e-8
KRAS up signaling	BMP2, CCND2, ID2, PLAUR, CFH, IGFBP3, JUP, CPE, GADD45G	1.41e-8
TNFA signaling via NFKB	BMP2, CD44, ATF3, ID2, SMAD3, PLAUR, CCL5, CEBPB, EFNA1	1.41e-8
Allograft rejection	F2, FAS, CCND2, BRCA1, CCL5, LYN, FLNA, HIF1A	1.6e-7
Apical junction	ITGB1, JUP, ACTC1, GNAI1, IRS1, SHC1, ACTB, ADAM15	1.6e-7
Estrogen early response	CD44, KRT18, NCOR2, ABLIM1, IGF1R, NRIP1, KRT15, AR	1.6e-7
Myogenesis	CLU, APP, IGFBP3, ITGB1, FGF2, ACTC1, ABLIM1, DTNA	1.6e-7

*Top ten hallmark pathways associated with the top 100 genes which became **more fragile** in the high-risk group defined by the RNA-seq data. Pathways becoming more fragile would be expected to exert less normal control.*

CONCLUSIONS

- In this study, we applied the geometric network analysis tool ORC to multi-omics data in MM represented as biological networks to identify individuals at high risk of short PFS and relevant biological correlates.
- For the RNA-seq data, patient clusters were associated with known predictors of PFS, indicating that our methods are valid, as well as providing new biological insights.

ACKNOWLEDGEMENTS

The work was supported by the Breast Cancer Research Foundation. KM is supported by an MMRF Research Scholar Award, an ASH Research Restart Award, and Royal Australasian College of Physicians.