Network geometry analysis reveals unique multiple myeloma genomic subtypes and potential new therapeutic targets

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

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center; [2]Myeloma Service, Department of Medicine, Memorial Sloan Kettering Cancer Center; [3]Department of Applied Mathematics & Statistics, Stony Brook University; [4]Department of Computer Science, Stony Brook University; *Co-first authors

*simhala@mskcc.org; maclachk@mskcc.org

Challenge

- A range of genomic aberrations predicts prognosis in multiple myeloma (MM), including copy number alteration (CNA), mutational processes, and structural variation.
- Recent studies have subtyped MM by genomic information, but no study has explored these abnormalities together with a global assessment of gene interactions.
- Network topology analysis integrates complex interactions to define patterns of biological behavior not captured by individual genomic events.
- We hypothesize that incorporating gene interaction networks will delineate biologically relevant MM subtypes and potential new therapeutic targets.

Study Sample

MMRF CoMMpass dataset (IA19)

- RNA-seq, CNA, and clinical data
- Number of subjects: 659

Category	Alive (N=295)	Dead (N=364)	P-value
Age	59.8 ± 9.5	64.8 ± 11.0	1.6e-9
Sex	M:164, F:131	M:229, F:135	0.07
ISS stage	(I) 133, (II) 94, (III) 62	(I) 96, (II) 138, (III) 117	1.7e-12

Table 1. Demographic information about the study sample.

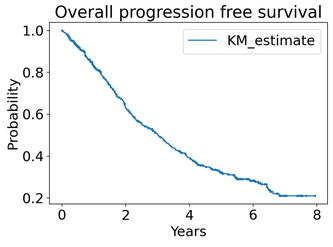


Figure 1. Kaplan Meier curve showing progression free survival for the 659 subjects.

Interactome — HPRD

- To understand the interactions between genes, we used the Human Protein Reference Database (HPRD).
- Interactome had 8,488 genes; 33,695 interactions.

Ollivier Ricci Curvature

- ORC considers both local and global connectivity in assessing the robustness of each pathway characterized by feedback loops in a network.

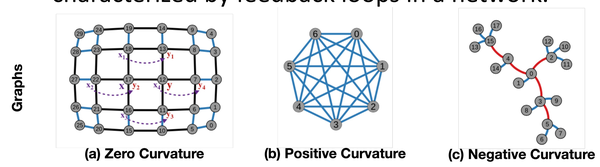
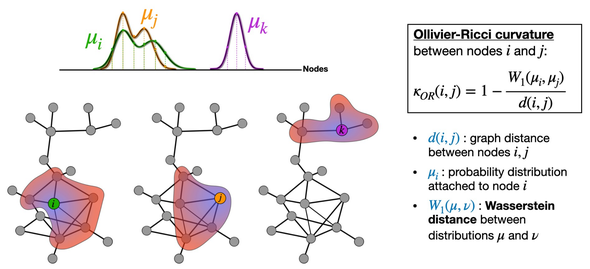


Figure 2. Ollivier Ricci curvature on example networks. Blue edges indicate positive curvature between nodes. Black edges indicate zero curvature, and red edges indicate negative curvature between nodes.



Ollivier-Ricci curvature between nodes i and j:

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

- $d(i,j)$: graph distance between nodes i, j
- μ_i : probability distribution attached to node i
- $W_1(\mu,\nu)$: **Wasserstein distance** between distributions μ and ν

Figure 3. Overview of ORC formulation. (Left) Example of how to form a probability distribution around a node. (Right) ORC formulation with definitions.

- Positive curvature -> multiple active pathways for information communication.
- Negative curvature -> limited active pathways for information communication.

Clustering Analysis

- Data clustered using a hierarchical agglomerative clustering approach.
- Optimal number of clusters determined via silhouette score.
- Survival analysis for progression-free survival (PFS) was performed using the Kaplan-Meier method.

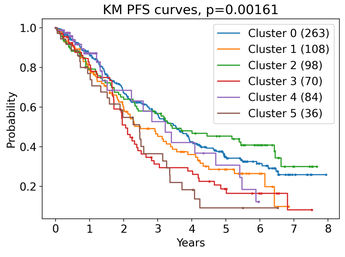


Figure 4. RNA-seq based Kaplan Meier curves showing PFS for each cluster identified by ORC cluster analysis.

RNA-seq Results

- RNA-seq clustering analysis identified 6 clusters.
- The clusters predicting the best and worst PFS were associated with specific genomic features including CCND1 and MAF/MAFB translocations, respectively (each p<0.05, Fisher's exact test).

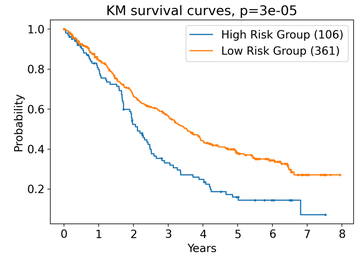


Figure 5. RNA-seq KM plot highlighting the different outcomes of high risk and low risk groups.

	Low risk (N=361)	High risk (N=106)
Hyperdiploidy	206/311	17/93
t(4;14)	0/332	66/97
t(11;14)	80/332	0/97
MAF/MAFB translocation	1/332	28/97
MYC translocation	65/332	11/97
Chromothripsis	69/332	40/97
APOBEC activity	6/330	24/97
TP53 inactivation (1 monoallelic, 2 biallelic)	(0) 238; (1) 29; (2) 9	(0) 71; (1) 11; (2) 5
Gain 1q21 (1 gain, 2 amplification)	(0) 226; (1) 76; (2) 9	(0) 45; (1) 30; (2) 18

Table 2. Differences in key biomarkers between high risk and low risk groups.

- Differential gene expression analysis carried out using raw read counts with DESeq2.
- 118 genes derived with FDR<0.05 and fold change > or < 3.5 was input into MetaCore to identify biological correlates associated with the risk of PFS.
- Key implicated biological processes included cell cycle progression, inflammation and immune response.

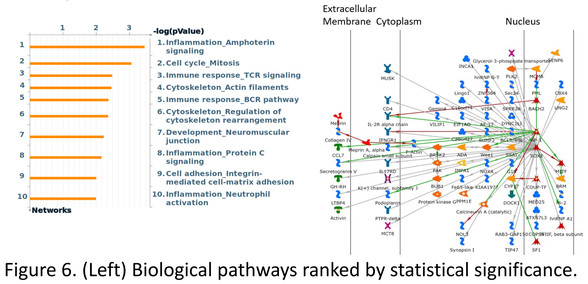


Figure 6. (Left) Biological pathways ranked by statistical significance. (Right) The largest connected network using top differential genes associated with progression free survival.

CNA Results

- CNA clustering analysis identified 8 clusters.
- Genomic features associated with the clusters included hyperdiploidy, CCND1 (longest PFS) and MAF-translocations, and APOBEC-activity (shortest PFS) (each p<0.05, Fisher's).
- The worst PFS cluster had a higher prevalence of chromothripsis (p<0.0001, Fisher's), known to be a strong predictor for MM survival.

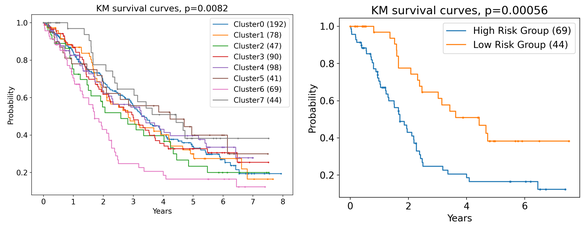


Figure 7. CNA based Kaplan Meier curves. (Left) KM plot highlighting PFS for each cluster identified by CNA ORC cluster analysis. (Right) KM plot highlighting the different outcomes of high risk and low risk groups.

	Low risk (N=44)	High risk (N=69)
Hyperdiploidy	26/37	7/52
t(4;14)	8/38	24/58
t(11;14)	1/38	13/58
MAF/MAFB translocation	2/38	15/58
MYC translocation	4/38	11/58
Chromothripsis	15/38	25/58
APOBEC activity	3/38	11/58
TP53 inactivation (1 monoallelic, 2 biallelic)	(0) 24; (1) 8; (2) 2	(0) 43; (1) 5; (2) 1
Gain 1q21 (1 gain, 2 amplification)	(0) 18; (1) 17; (2) 2	(0) 2; (1) 33; (2) 17

Table 3. Differences in key biomarkers between high risk and low risk groups.

Summary

- 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 both RNAseq and CNA-based 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 and an ASH Research Restart Award