ELSEVIER

Blood 142 (2023) 1960–1963

The 65th ASH Annual Meeting Abstracts

POSTER ABSTRACTS

652.MULTIPLE MYELOMA: CLINICAL AND EPIDEMIOLOGICAL

A Novel Biosignature for Potential Stratification and Elucidation of Newly Diagnosed Multiple Myeloma Patients at-Risk

Muhammad Kashif, PhD[1,2,3], *Yijie Zhou*[3,4,2], *Vincent Luong, MD*[4,2,5], *Katarina Uttervall, MDPhD*[4,2,5], *Gösta Gahrton, MDPhD*[4,2], *Hareth Nahi, MD PhD*[4,2,6], *Evren Alici, MDPhD*[4,7], *Johan Lund, MDPhD*[4,2,5]

[1] Department of Medicine, Karolinska Institutet, Huddinge, Sweden
[2] Center of Hematology and Regenerative Medicine (HERM), Karolinska Institutet, Stockholm, Sweden
[3] Department of Information Technology, Uppsala University, Uppsala, Sweden
[4] Department of Medicine, Karolinska Institutet, Stockholm, Sweden
[5] Department of Hematology, Karolinska University Hospital, Stockholm, Sweden
[6] Department of Medicine, Institution for biomedicine and clinical science, Linköping, Sweden
[7] Center for Hematology and Regenerative Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden

Multiple myeloma (MM) patients have highly variable overallsurvival (OS) ranging from few weeks to more than ten years. Discovering an early biosignature to stratify short-term from long-term survivors offers the prospect of treating at-risk patients. Machine learning (ML) algorithms are currently being tested to discover biosignatures, but they consist of several features that make their implementation in healthcare an arduous task.

Here, we have developed an algorithm called AlgoOS to stratify newly diagnosed MM (NDMM) patients by integrating a NetRank algorithm, a variation of the Google PageRank algorithm, and ML algorithms-the first of its kind in MM. Also, a dataset of NDMM patients (*n=31*) was built consisting of transcriptomic (features=28256), clinical (features=13), biochemical (features=12), and fluorescent in situ hybridization (FISH) (features=3) data. A cut-off OS of 46 months was used to group short and long-term survivors based on domain knowledge. Finally, AlgoOS was implemented on this dataset, and a biosignature predictive of NDMM patient stratification was extracted. The prediction model's performance was evaluated by accuracy, precision, and F1-score, and 5-fold cross-validations were performed. R was used to build a transcription-factor-gene-regulatory network, while all other analyses, including ML, were performed using Python.

During 1 st step of AlgoOS, all transcriptomic features were ranked by NetRank score, and the top 20 were selected for further processing. This ranking was similar to web page ranking done by the Google PageRank algorithm, except that the NetRank algorithm also takes into account the correlation of features to OS. In detail, we calculated NetRank scores of transcriptomic features by building a transcription-factor-gene-regulatory network using the JASPAR-v2022 database. Each transcription factor motif in JASPAR was matched to the putative promoter region (upstream 1000 base pairs) of genes in the hg38 human reference genome. Furthermore, correlations of transcriptomic features to OS were computed, and the NetRank score was calculated by iteratively optimizing a damping factor parameter *d*, see Figure 1. We trained the support vector machine (SVM) on top NetRanked 10-20000 transcriptomic features, trained it separately on the same number of randomly selected features and calculated their performance scores. The models were chosen by the criteria 1) precision and accuracy $\geq$ 80%, 2) kernel = non-linear, and 3) $C < 1$, where C is a regularization parameter. It was found that the top 20 NetRanked features were the best predictive of patient stratification. The results were better than random feature selection (accuracy= 90 vs. 62, precision= 100 vs. 44, F1-score= 91 vs. 51). The results were also validated by another classifier randomforest (RF) (accuracy= 70 vs. 44, precision= 71 vs. 32, F1-score= 77 vs. 44).

During the 2 nd AlgoOS step, the importance of the 20 features in the RF model was calculated to find that leaving out 2 of them did not affect the performance. Therefore, 18 transcriptomic features were integrated with clinical, biochemical, and FISH data, and a final RF was trained. It was found that data integration improved the performance of the final RF compared to the previous RF model that was trained only on transcriptomic data (accuracy= 90 vs. 70, precision= 100 vs. 71, F1-score= 91 vs. 77). Finally, we found a novel biosignature to stratify short-term from long-term survivors. The biosignature consisted of twenty nine features only, including eighteen transcriptomic (*TMEM62, BUB1, CXorf21, PLEKHM1, PRPF18, ULK2, VMP1, LONRF1, USP15, TBC1D5, NUP93, CASP4, NDFIP1, TPM1, NRXN2, ALG9, C1QC and SOX13*), seven biochemical (LDH, albu-

min, creatinine, calcium, M-proteins, Hb light chains, Hb heavy chains), three clinical (age, gender, ISS stage) and one FISH (17P13) features. Moreover, LDH, *TMEM62* and *BUB1* were the three most important features, see Table1.
Bioinformatics analysis of the 18 transcriptomic features in the biosignature showed an enrichment to autophagy (autophagic cell death, p-value=9.8×10^{-3}; autophagosome, p-value= 1.5×10^{-2}) that is a known mechanism of resistant to drugs and cell death in MM.
Our findings indicate that creating a biosignature using transcriptomic features in addition to previously used prognostic factors may improve prediction of outcome in MM. The model has to be tested in larger clinical material.

Disclosures Gahrton: *Fujimoto Pharmaceutical Corporation, Japan:* Consultancy. **Alici:** *Vycellix, Inc:* Current equity holder in private company, Other: Co-founder.

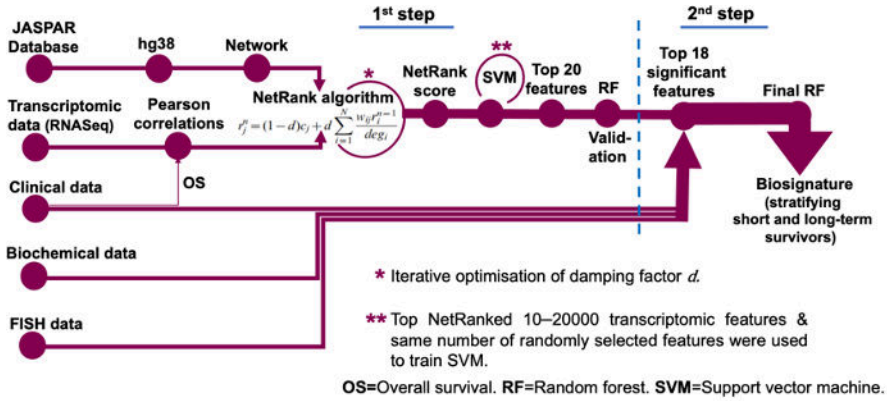

Figure 1: AlgoOS: An algorithm to discover a biosignature stratifying NDMM patients at-risk using integrated transcriptomic, clinical, biochemical, and FISH data.
During **1st step** of the algorithm, a transcription-factor-gene-regulatory network was built using the JASPAR database, and Pearson correlations of transcriptomic features to OS were calculated. A NetRank algorithm was used to calculate the NetRank score through iterative optimization of a damping factor parameter *d*. All transcriptomic features were sorted by NetRank score, and top 10–20000 NetRanked features were used to train SVMs to stratify NDMM patients into short and long-term survivors. Also, SVMs with the same numbers but randomly selected features were trained ten times, and average performance scores were calculated. It was found that the performance of SVM trained on NetRanked sorted features was better than SVM trained on randomly selected features. We found that the top 20 NetRanked features were the best predictor of patient stratification. This result was validated using a RF. During **the 2nd AlgoOS step,** removing 2 of these 20 features did not affect the model's performance. Therefore, 18 transcriptomic, along with clinical, biochemical, and FISH data, were used to train final RF, and finally, a novel biosignature to stratify NDMM patients was discovered.

Table 1: A novel biosignature to stratify NDMM at-risk patients. The biosignature consisted of eighteen transcriptomic, seven biochemical, three clinical, and one FISH features. In this table, all features were sorted based on their importance in the final RF.

Rank by importance in the final RF model	Features	Category	NetRank Score (Calculated only for transcriptomic features) (max= 0.63, min= 0)	Pearson correlation to OS (Calculated only for transcriptomic features)
1	Lactate dehydrogenase (LDH)	Biochemical		
2	*TMEM62*	Transcriptomic	0.54	0.30
3	*BUB1*	Transcriptomic	0.51	-0.35
4	*PRPF18*	Transcriptomic	0.49	0.62
5	Albumin	Biochemical		
6	*CXorf21*	Transcriptomic	0.60	-0.56
7	*PLEKHM1*	Transcriptomic	0.63	0.52
8	Creatine	Biochemical		
9	Calcium	Biochemical		
10	*USP15*	Transcriptomic	0.49	0.23
11	*ULK2*	Transcriptomic	0.52	0.43
12	*LONRF1*	Transcriptomic	0.56	0.51
13	*ALG9*	Transcriptomic	0.51	0.11
14	*NUP93*	Transcriptomic	0.58	-0.25
15	M-proteins	Biochemical		
16	*VMP1*	Transcriptomic	0.51	0.47
17	*NDFIP1*	Transcriptomic	0.485	0.52
18	Age	Clinical		
19	*TPM1*	Transcriptomic	0.53	0.35
20	*CASP4*	Transcriptomic	0.50	0.41
21	*C1QC*	Transcriptomic	0.52	0.36
22	Type of Hb light chains	Biochemical		
23	*TBC1D5*	Transcriptomic	0.48	0.33
24	*NRXN2*	Transcriptomic	0.49	0.33
25	Type of Hb heavy chains	Biochemical		
26	Sex	Clinical		
27	*SOX13*	Transcriptomic	0.58	0.47
28	ISS stage I/II-III	Clinical		
29	17p13	FISH		

Figure 1

https://doi.org/10.1182/blood-2023-179999