Statstical Methods for the Study of Extracellular Vesicles Content and their Potential as Biomarkers of Multiple Myeloma Agressiveness

AUTHORS: Carolina Pestana | Lisete Sousa, Emilie Carneiro, Filipa Barahona, Joana Caetano, Raquel Lopes, Bruna Ferreira, Cristina João

R&D UNIT: CEAUL

CONTACT: lmsousa@fc.ul.pt

CEAUL
Centro de Estatística e Aplicações
Universidade de Lisboa

INTRODUCTION & OBJECTIVES

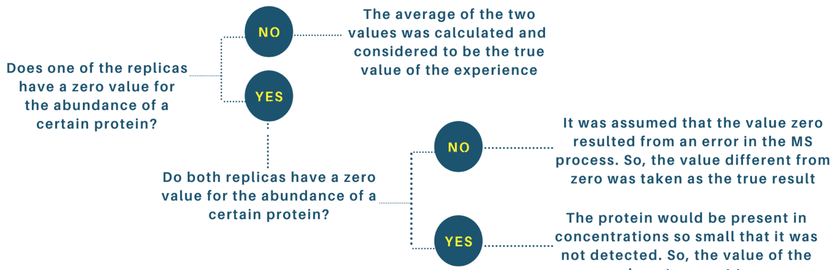
Multiple Myeloma (MM) is an aggressive and incurable haematologic malignancy preceded by the precursor stage of monoclonal gammopathy of undetermined significance (MGUS) [1]. Tools such as the disease aggressiveness score (R-ISS) [2] are currently used to estimate prognosis but they fail to accurately predict which patients will relapse earlier and die prematurely [3], therefore new biomarkers for disease aggressiveness are needed. Extracellular vesicles (EV) are emerging players in cancer, currently being explored as liquid biopsies. The present work aims at performing a statistical analysis that could allow the detection of differentially expressed proteins to identify the most promising EV proteins biomarkers in a group of MM patients.

MATERIALS & METHODS

A mass spectrometry (MS) analysis was performed on EV isolated from peripheral blood' samples in order to obtain the quantitative protein abundance for each sample. Then, the Rank Product (RP, RankProd R package) statistical method was applied to detect differentially expressed proteins in each one of the two interest groups: (A) patients' diagnosis and (B) R-ISS score. Although this method was originally designed to identify differentially expressed genes in a single experiment [4], it received widespread acceptance and is now used in several domains of *-omics*, such as transcriptomics and proteomics.

PRE-PROCESSING OF MS DATA

The mass spectrometry (MS) analysis allowed the identification of 1067 proteins. Since MS data was already normalized and filtered, the following procedure was applied to merge the two technical replicas of each sample into a single one:



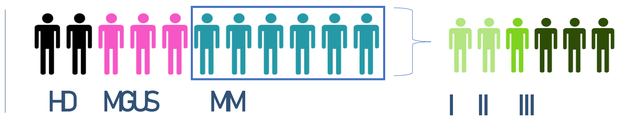
BUT WHY DID WE USE THE RANK PRODUCT METHOD?

- Is a non-parametric statistical method that makes it possible to detect variables consistently up-regulated or down-regulated in a number of replicate experiments [5]
- This method allowed us to rank the expression levels for all proteins from all replicate samples, being the proteins with the smallest RP value the ones with the most biological interest [7]
- Since the RP does not depend on estimations of the measurement variance for each single variable, it is very useful when this estimate becomes unreliable in result to a low number of replicates [4]
- Some of the advantages of this method are the fewer assumptions under the model and the strong performance with noisy data or a low number of biological replicates [6]

A False Discovery Rate (FDR) equals to 0,2 was used as a cut-off to define the differentially expressed proteins and the level of significance for protein selection was set at 0,05.

References
[1] Landgren, O. et.al., *Blood* (2009); [2] Palumbo, A. et al., *Journal Of Clinical Oncology* (2015).; [3] Manier, S. et. al., *Blood* (2017).; [4] Breitling, R. et. al., *FEBS Letters* (2004).; [5] Del Carratore, F. et al., *Bioinformatics* (2017); [6] Hong, F. et al., *Bioinformatics* (2006).; [7] Koziol, J. A., *FEBS letters* (2010).

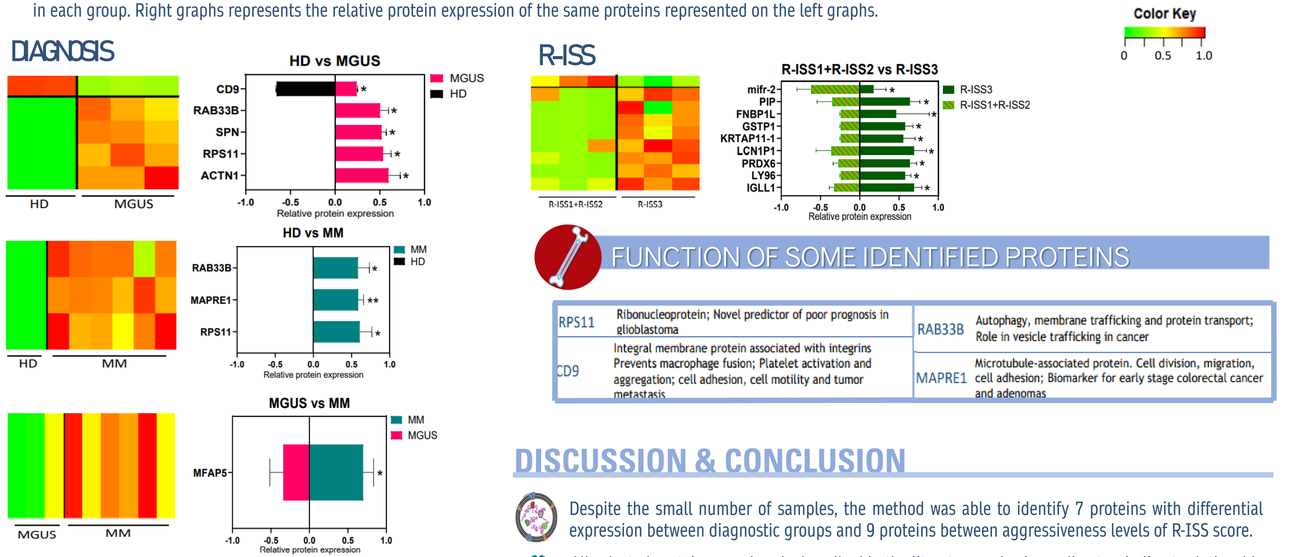
COHORT'S DESCRIPTION



RESULTS

EV PROTEINS WITH DIFFERENTIAL EXPRESSION LEVEL

The heat maps' (left graphs) colour scheme represents the lower or higher abundance (on a green to red scale, respectively) of the selected proteins with differential expression in each group. Right graphs represents the relative protein expression of the same proteins represented on the left graphs.



FUNCTION OF SOME IDENTIFIED PROTEINS

RPS11	Ribonucleoprotein; Novel predictor of poor prognosis in glioblastoma	RAB33B	Autophagy, membrane trafficking and protein transport; Role in vesicle trafficking in cancer
CD9	Integral membrane protein associated with integrins Prevents macrophage fusion; Platelet activation and aggregation; cell adhesion, cell motility and tumor metastasis	MAPRE1	Microtubule-associated protein. Cell division, migration, cell adhesion; Biomarker for early stage colorectal cancer and adenomas

DISCUSSION & CONCLUSION

Despite the small number of samples, the method was able to identify 7 proteins with differential expression between diagnostic groups and 9 proteins between aggressiveness levels of R-ISS score.

All selected proteins are already described in the literature as having a direct or indirect relationship with cancer. Those preliminary results require confirmation with a larger cohort of patients but yet, exosomal content analysis seems a promising approach for the identification of informative and non-invasive biomarkers for MM.

More than 100 patients' samples are currently undergoing analysis by MS. This new set of data will be analysed with the Limma package, implemented in the R/Bioconductor. This is a parametric method, which fits a linear model to each protein, using a moderated t-statistics from the empirical Bayes procedure. This empirical approach makes use of the parallel structure between the *-omics* data, making a reliable estimation.

With this new data set we aim estimating the probabilities of disease progression, relapse and response to treatment through multi-state models and survival analysis, associated with protein expression and other clinically relevant variables.

This work is partially financed by national funds through FCT – Fundação para a Ciência e a Tecnologia under the project UIDB/00006/2020.

FCT Fundação para a Ciência e a Tecnologia | Ciências ULisboa | Champalimaud Foundation