POSTER

Network inference and clustering: mathematical tools to discover glioma biomarkers from gene expression patient data

Sofia Martins[a], Roberta Coletti[b], Marta B. Lopes[b,c,d]

[a] NOVA School of Science and Technology (NOVA SST), NOVA University of Lisbon, Caparica, Portugal
[b] Center for Mathematics and Applications (NOVA Math), NOVA SST, Caparica, Portugal
[c]Department of Mathematics, NOVA SST, Caparica, Portugal
[d] UNIDEMI, Department of Mechanical and Industrial Engineering, NOVA SST, Caparica, Portugal

E-mail: roberta.coletti@fct.unl.pt\qquadmarta.lopes@fct.unl.pt

Abstract

Gliomas are primary malignant brain tumors with high heterogeneity, often showing poor patient survival. Increasing the molecular understanding of different glioma types is essential to find novel targets for personalized therapies [1]. Given the huge amount of data from different molecular layers that nowadays are available, statistical methods and machine learning techniques can help to disclose relevant information.

In this work, we investigated the differences and similarities among the three main glioma types (astrocytoma, oligodendroglioma, and glioblastoma) arising from the gene expression data from TCGA glioma projects. After updating the sample diagnostic labels according to the 2016 glioma classification guidelines, we estimated a gene network for each glioma type by employing the Joint Graphical LASSO method [2]. The estimated sparse networks were analyzed to detect relations characteristic of the different tumor types, and a network-based variable selection was performed. Robust Sparse K-means Clustering [3] was applied in different cases of study, as a way to validate the ability of the selected features to distinguish classes of patients, as well as to disclose new patient groups and the relevant genes behind groups' separation.

The obtained results highlight a molecular distinction between glioblastoma and the other glioma subtypes, showing more similarities in astrocytoma and oligodendroglioma. Our findings, supported by a targeted literature review, identified some features as potential candidates as glioma biomarkers. To comprehend these genes' potential roles in diagnosis and the development of innovative therapeutics, a biological validation of these genes is recommended.

Keywords Network inference, Clustering, Glioma, Variable selection

References

[1] M. Weller, W. Wick, M.B. Ken Aldape *et al.*: *Glioma*. Nature Reviews Disease Primers **1 (15017)** (2015)

[2] P. Danaher, P. Wang and D.M. Witten: *The joint graphical lasso for inverse covariance estimation across multiple classes*. Journal of the Royal Statistical Society. Series B, Statistical methodology, **76 (2)**, 373–397 (2014).

[3] Y. Kondo, M. Salibian-Barrera and R. Zamar: *RSKC: An R Package for a Robust and Sparse K-Means Clustering Algorithm*. Journal of Statistical Software, **72 (5)**, 1–26 (2016)