An intuitive network-based approach to investigate clinical features among breast cancer subtypes

Fonseca, AL[123], De Souza, SJ[13]

Universidade Federal do Rio Grande do Norte[1],
Pós-graduação em Bioinformática[2], Núcleo Interdisciplinar de Bioinformática (BioME)[3]

Breast cancer is a heterogeneous disease that covers a broad spectrum of pathologies with each subtype having particular characteristics, including: morphological, behavioral and molecular features. Furthermore, breast tumors have a variety of clinical outcomes, hormonal response and different therapeutic options. Nowadays, the 'omics' technologies have allowed the understanding of that heterogeneity and also their impact in patient prognosis. However, fundamental issues at clinical level remains unsolved, mostly regarding the association between molecular subtypes and treatment, such as the case of triple-negative breast cancer (TNBC). Furthermore, others issues associated with tumor aggressiveness and patient survival rate are not extensively investigated. To approach such problems, we create a network-based strategy for prioritization of genes and topological modules, called Triads. First of all, breast cancer data were collected from the TCGA consortium, which span four subtypes, Luminal A (440 samples), Luminal B (123), Her2 Enriched (37) and TNBC (113). In briefly, our approach is based in two major principles, i) genomics data integration, and; ii) reduction of network complexity. For the data integration, TCGA data were summarized using the S-score method, which calculates for each gene a specific score, reporting alteration rate and their type (loss or gain-function). Finally, we apply a statistical correction to reduce the topological properties (nodes or edges) in each subtype network. As result, the network complexity reduction reached 50% and 30%, among the nodes and edges, respectively. Moreover, the nodes selected by our method presented a high proportion of extremely altered genes in each subtype. Furthermore, our findings report differential topological properties among the subtypes networks, for instance, exclusive nodes and edges in each network. To evaluate the biological meaning of these datasets, we carried out a enrichment analysis with the exclusive nodes against the KEGG and Gene Ontology databases. Finally, we analyze the filtered edges and their association with the survival outcome, through a Kaplan-Meier method. In addition, the frequency of survival edges in our networks seems to be significant when compared against full human PPI network. In conclusion, the development of network-based approaches is a powerful tool to understand complex diseases, mostly due to the ability of these methods to integrate biological data. We presented here, a flexible and intuitive approach able to reduces the network complexity and highlight relevant clinical features of different breast cancer subtypes.