INBIA: a boosting methodology for proteomic network inference

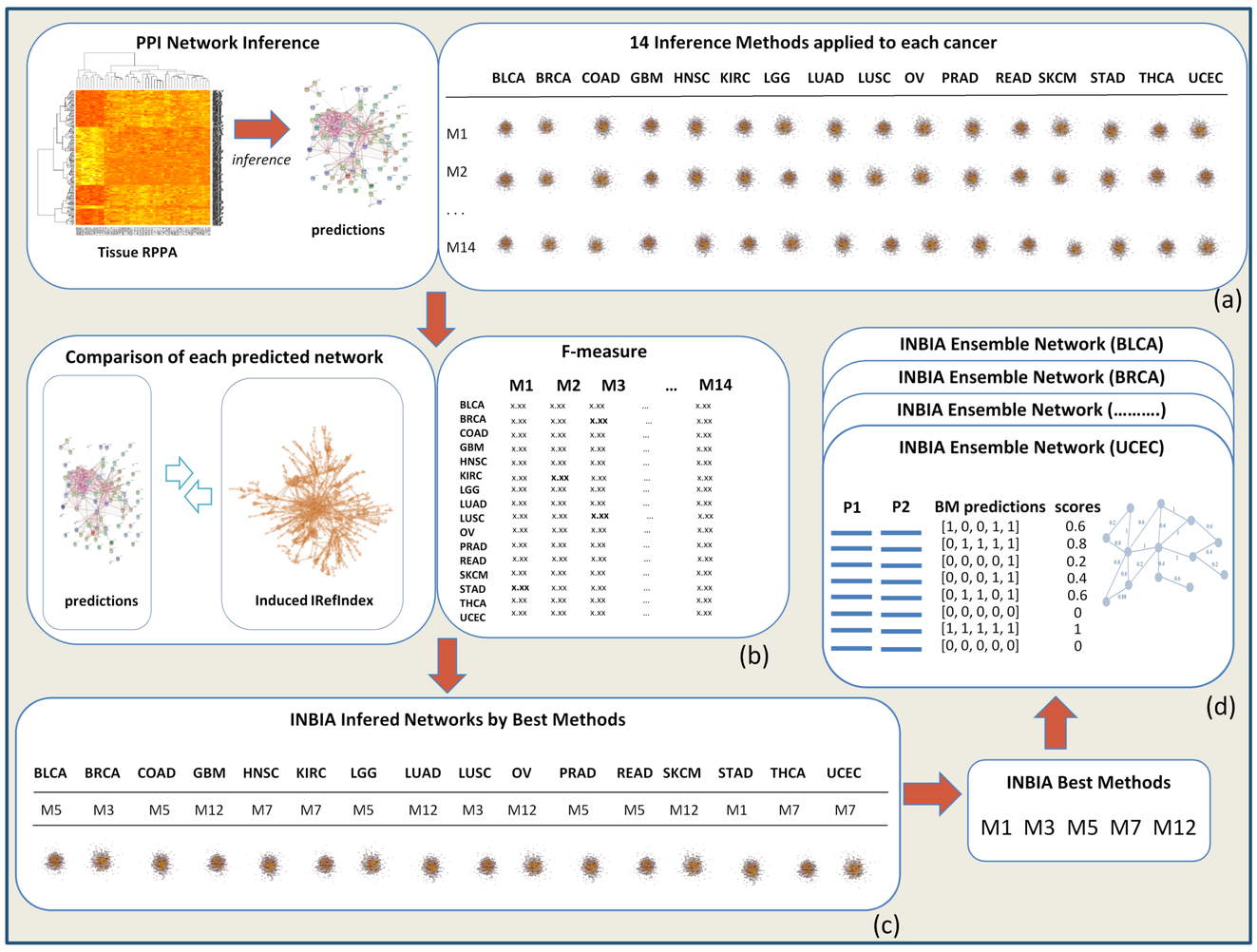
Davide Stefano Sardina, Giovanni Micale
Department of Mathematics and Computer Science, University of Catania, Viale Andrea Doria, 95125, Catania, Italy
dsardina@dmi.unict.it, gmicale@dmi.unict.it

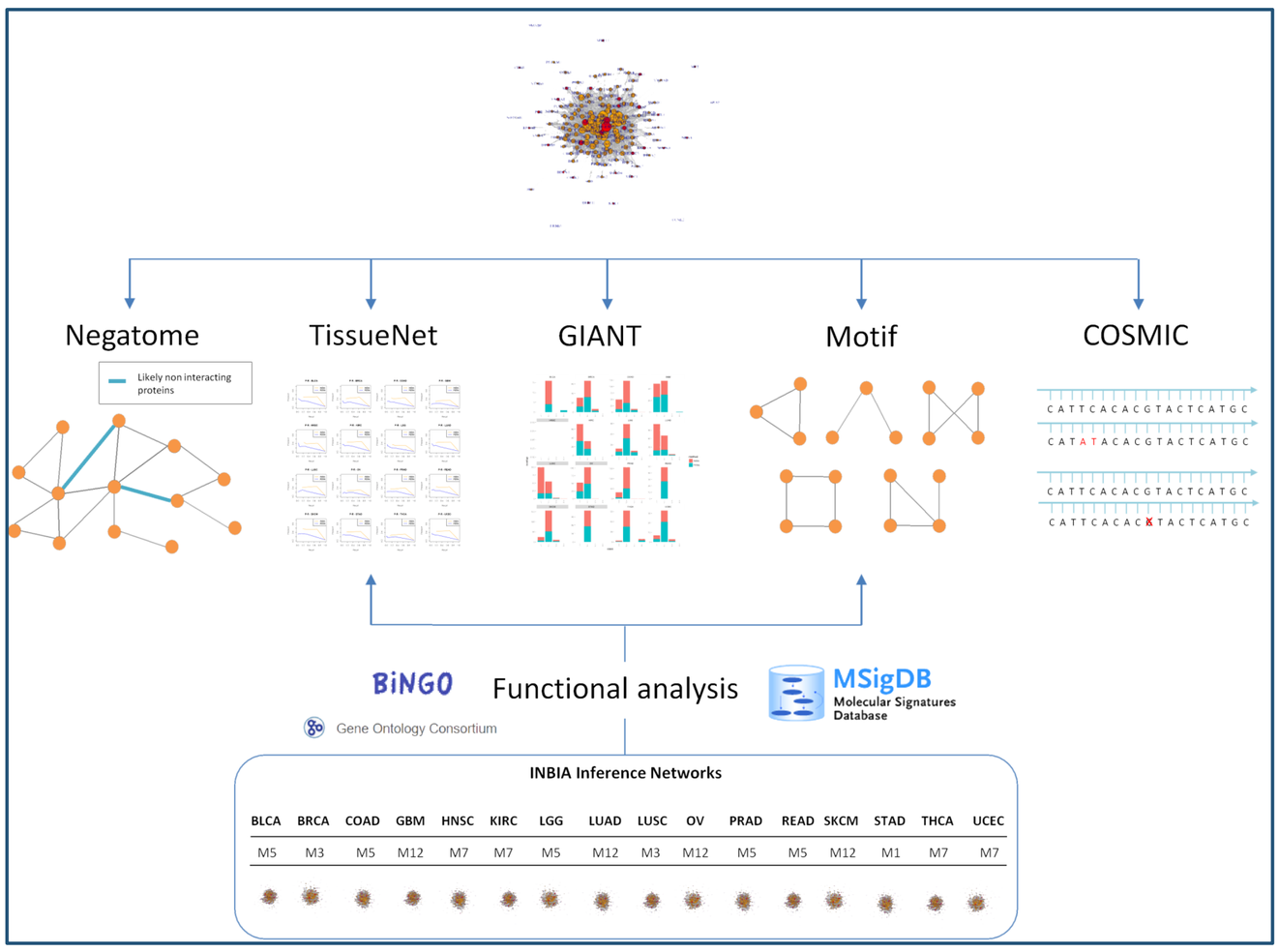
Background

The analysis of tissue-specific protein interaction networks and their functional enrichment in pathological and normal tissues provides insights on the etiology of diseases. The release of The Cancer Proteome Atlas (TCPA) has provided proteomic expression data for 190 proteins in 16 cancer types using reverse-phase protein arrays (RPPA) technology (Li et al., 2013). However, established protocols to infer interaction networks from protein expressions are still missing.

Materials and methods

We have developed a methodology called Inference Network Based on iRefIndex Analysis (INBIA) to accurately correlate proteomic inferred relations to protein-protein interaction (PPI) networks. INBIA makes use of 14 network inference methods (Sardina et al., 2016) on protein expressions related to 16 cancer types from TCPA (Li et al., 2013) and it uses as reference model the iRefIndex (Razick et al., 2008) human PPI network. For each inferred network, INBIA measures the goodness of predictions starting from the computation of true positive and false positive rates with respect to the gold standard and finally indentifies the best methods by measuring F-measure (Sardina et al., 2016). Predictions are then validated through non-interacting and tissue specific PPI networks resources. The first, Negatome (Blohm et al., 2014), takes into account likely non-interacting proteins while TissueNet (Basha et al., 2016) and GIANT (Greene et al., 2015), report experimentally verified PPIs in more than 50 human tissues. The reliability of the proposed methodology is assessed by comparing INBIA with PERA (Şenbabaoğlu et al., 2016), a tool which infers protein interaction networks from Pathway Commons, by both functional and topological analysis.





Results

INBIA's ensemble set is made by 4 methods (CLR, GLASSO, PLS, and MRNET). Comparing with Negatome, we found that there was, in all cases and for both methods, a very small set of interactions in common, meaning that both methodologies predicted few validated false negative interactions.
We computed the predicted PPIs in common among all inferred networks. INBIA unravels a total of 83 PPIs that includes EIF4EBP1, AKT1, GSK3A, RPS6, MAPK1, SRC proteins as most central in the network based on computed collective influence centrality measure. INBIA's best networks produce specific gene sets in MSigDB (Subramanian et al, 2005), e.g. HALLMARK MTORC1 SIGNALING.
We ran FlashMotif algorithm (Micale et al., 2017) to find all possible non-induced colored motifs of 3 and 4 nodes, where a motif represents a subgraph in which each node is 'colored' with a specific GO term. It found 959 colored motifs with 3 nodes and 9,006 motifs with 4 nodes in INBIA networks. It found 7 over-represented motifs with 4 nodes in the INBIA network.

Conclusions

The results show that our approach is capable to recover better PPI interaction networks, in terms of precision-recall, with respect to those retrieved by PERA when compared on TissueNet. Furthermore, a comparison of the inferred networks using GIANT tissue/edges classification shows that INBIA networks contain more interactions among tissue specific genes positively co-expressed in the tissue. The comparison clearly highlights that the selection of a proper reference database is crucial to establish the actual soundness of inference network models. In conclusion, INBIA is a valuable approach to predict proteomic interactions in pathological conditions starting from the current knowledge of human protein interactions.

References

1. Basha, O., Ruth, B., Sharon, M., Lerman, E., Kirson, B., Hekselman, I., Yeger-Lotem, E.: The tissuenet v.2 database: A quantitative view of protein-protein interactions across human tissues. Nucleic Acids Research 45, 427-431 (2016)
2. Blohm, P., Frishman, G., Smialowski, P., Goebels, F., Wachinger, B., Ruepp, A., Frishman, D.: Negatome 2.0: A database of non-interacting proteins derived by literature mining, manual annotation and protein structure analysis. Nucleic Acids Research 42, 396-400 (2014)
3. Greene, C.S., Krishnan, A., Wong, A.K., Ricciotti, E., Zelaya, R.A., Himmelstein, D.S., Zhang, R., Hartmann, B.M., Zaslavsky, E., Sealfon, S.C., et al.: Understanding multicellular function and disease with human tissue-specific networks. Nature genetics 47(6), 569-576 (2015)
4. Li, J., Lu, Y., Akbani, R., Ju, Z., Roebuck, P.L., Liu, W., Yang, J.-Y., Broom, B.M., Verhaak, R.G., Kane, D.W et al.: Tcpa: a resource for cancer functional proteomics data. Nature methods 10(11), 1046-1047 (2013)
5. Micale, G., Giugno, R., Ferro, A., Mongiovì, M., Shasha, D., Pulvirenti, A.: Fast analytical methods for finding signicant colored graph motifs. To be published in Data Mining and Knowledge Discovery (2017)
6. Razick, S., Magklaras, G., Donaldson, I.: IRefIndex: A consolidated protein interaction database with provenance. BMC Bioinformatics 9(1), 405 (2008)
7. Sardina, D.S., Micale, G., Ferro, A., Pulvirenti, A., Giugno, R.: Correlation between proteomic network inference and protein-protein interaction networks. In: Proc. 13th International Conferenceon on Computational Intelligence Methods for Bioinformatics and Biostatistics., pp. 39-44 (2016)
8. Şenbabaoğlu, Y., Sümer, S., Sánchez-Vega et, F., Bemis, D., Ciriello, G., Schultz, N., Sander, C.: A multi-method approach for proteomic network inference in 11 human cancers. PLOS Computational Biology 12, 1-31 (2016)
9. Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., Paulovich, A., Pomeroy, S., Golub, T., Lander, E., Mesirov, J.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America 102(43), 15545{15550 (2005)