Network-Theoretic Mapping of Breast Cancer in India through Functional Cartography in Gene Regulatory and Protein-Protein Interactions

Extended Abstract

Breast cancer is the leading cancer among women in India, representing 13.5% of all cancer diagnoses and 10.6% of cancer-related deaths as of 2020. It was estimated that India had 118,000 new cases of breast cancer, with a total number of living cases at over 525,000. In this study, we identified transcriptional level changes in Indian breast cancer tumours versus normal tissue from datasets available in the NCBI Gene Expression Omnibus (GEO) online. This helped to identify differentially expressed genes (DEGs) in the Indian breast cancer dataset, as shown in Figure 1. However, getting potential drug targets from this long list of DEGs is complex. Traditional network theoretical analysis often captures hub genes or proteins based on degree and centrality measures, which misses important information about their topological connectivity in the network. To address this limitation, we aim to apply functional cartography of hub genes/proteins for further classification into functional roles [1]. Here, we present an integrated network-science framework for both gene regulatory networks (GRNs) and protein-protein interaction (PPI) networks to identify specific "connector hubs," the key molecular nodes that bridge distinct functional communities and may drive cross-talk between pathways.

Our pipeline begins with large-scale transcriptional profiles of breast cancer cohorts, where we construct GRNs using the Weighted Gene Co-expression Network Analysis (WGCNA) approach [2]. Parallel PPI networks are curated from high-confidence resources such as STRING and BioGRID [3]. For both gene regulatory and protein-protein interaction networks, we perform **community (module) detection** using complementary algorithms such as Louvain/Leiden for modularity maximisation and Infomap for flow-based clustering, implemented in `igraph` and `graph-tool` [4]. To achieve robust consensus partitions, we apply co-assignment-based clustering across repeated runs and algorithms. Network quality is assessed using modularity Q and normalised mutual information (NMI) to confirm stability and reproducibility.

Within each consensus community, we apply **functional cartography** to classify nodes by their intra- and inter-module connectivity [1]. The within-module degree z-score

$$z_i = \frac{k_i^{(s)} - \overline{k}^{(s)}}{\sigma_k^{(s)}}$$

quantifies how well node i is connected to other nodes in its own module s, while the participation coefficient

$$P_i = 1 - \sum_{s=1}^{M} \left(\frac{k_i^{(s)}}{k_i} \right)^2$$

measures the distribution of its links across all M modules. Nodes with high z_i and intermediate P_i emerge as **connector hubs**, bridging multiple functional communities.

This dual-network strategy allows us to pinpoint candidate drug targets that are not only differentially expressed but also topologically critical genes or proteins that coordinate inter-pathway communication and may underlie therapy resistance or disease progression. Early

analyses highlight well-known oncogenic drivers as well as less-characterised connectors that warrant experimental follow-up. By coupling community detection with functional cartography, we demonstrate a scalable methodology for uncovering hidden architecture in complex biological networks and provide a foundation for precision oncology interventions.

Ethical Considerations

All transcriptomic and interaction data are derived from publicly available repositories (e.g., TCGA, GEO, STRING), ensuring patient anonymity and compliance with open-data guidelines. No individual-level clinical identifiers are used, and all analyses follow FAIR data principles.

Figure

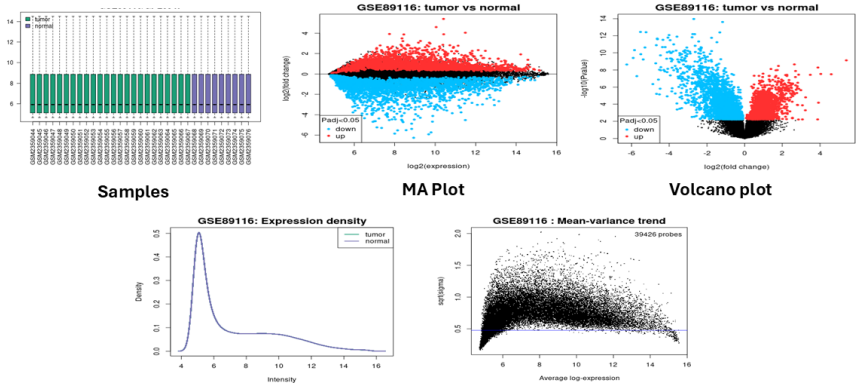

Figure: Gene expression analysis of an Indian Breast Cancer dataset comparing tumour and normal breast tissue samples. Visualisations (from left to right) include sample distribution, MA plot, volcano plot, expression density, and mean-variance trend to highlight differentially expressed genes.

References

1. Guimerà, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic networks. *Nature*, 433(7028), 895–900.

2. Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559.

3. Szklarczyk, D. *et al.* (2023). STRING v12: protein–protein association networks. *Nucleic Acids Research*, 51(D1), D638–D646.

4. Newman, M. E. J. (2006). Modularity and community structure in networks. *PNAS*, 103(23), 8577–8582.