

Tumour-specific Causal Inference Discovers Distinct Disease Mechanisms Underlying Cancer Subtypes

Yifan Xue¹, Gregory Cooper¹, Chunhui Cai¹, Songjian Lu¹, Baoli Hu^{2,3,4}, Xiaojun Ma¹, Xinghua Lu^{1,*}

¹Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, 15260, United States

²Department of Neurological Surgery, University of Pittsburgh School of Medicine, Pittsburgh, 15260, United States

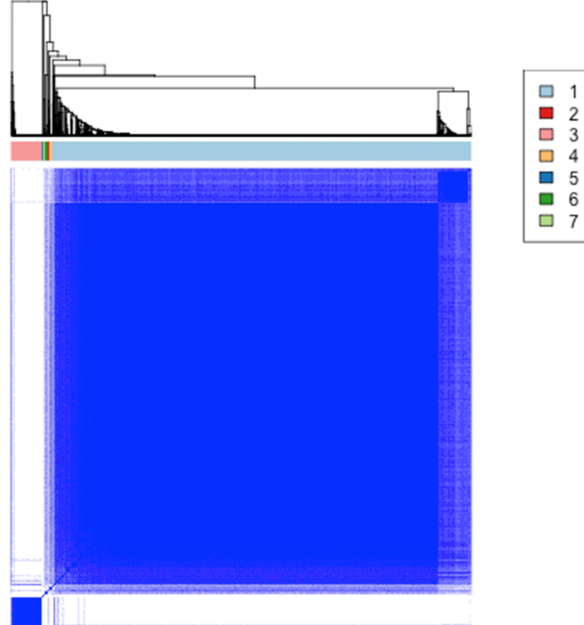
³Paediatric Neurosurgery, UPMC Children's Hospital of Pittsburgh, Pittsburgh, 15213, United States

⁴Molecular and Cellular Cancer Biology Program, UPMC Hillman Cancer Centre, Pittsburgh, 15232, United States

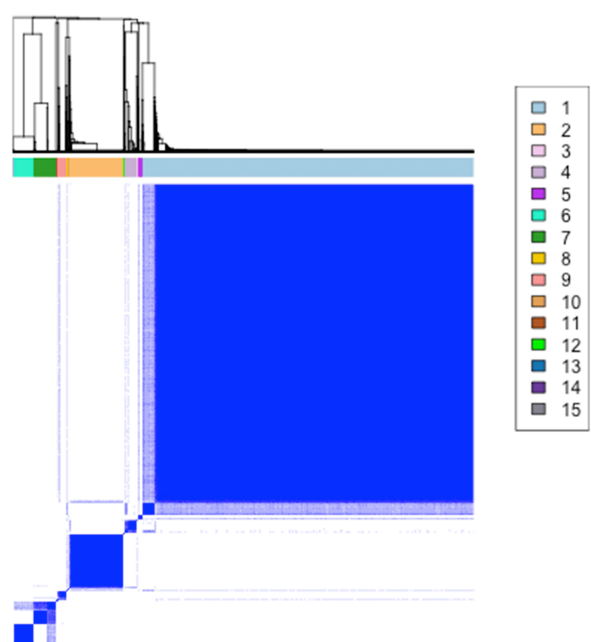
*xinghua@pitt.edu

Supplementary figure file.

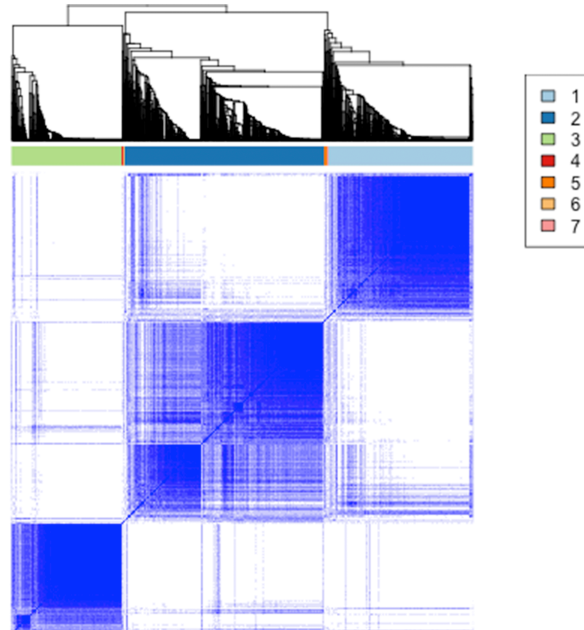
a. BRCA hierarchical clustering using affinity matrix



b. GBM hierarchical clustering using affinity matrix



c. BRCA hierarchical clustering using expression profile



d. GBM hierarchical clustering using expression profile

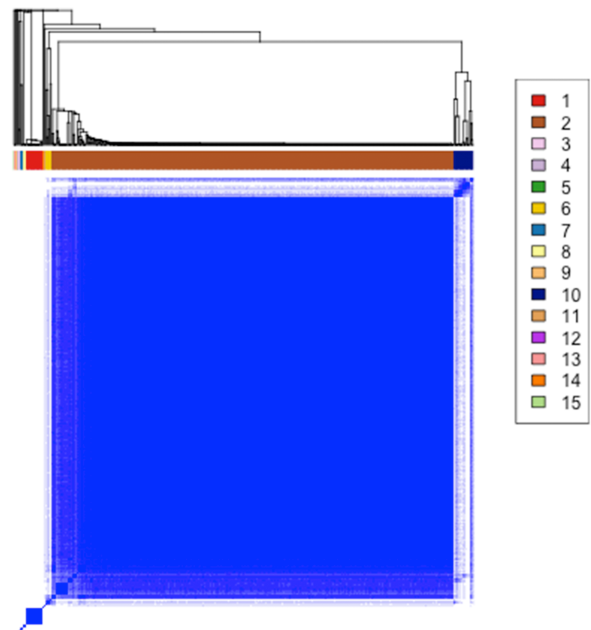


Figure S1. The consensus matrices of different clustering methods for identifying DEG modules, including hierarchical clustering using affinity matrix (a and b), and hierarchical clustering using expression profile with $1 - \text{Pearson correlation}$ as distance (c and d). The higher the frequency two DEGs are clustered into the same module, the darker blue the corresponding spot on the matrix. Each block sitting on the diagonal corresponds to a DEG module. The single dominant module composed of the majority of genes in a, b and d, and the overlapping across modules in c suggest that hierarchical clustering was unable to untangle the correlations between DEGs to identify robust modules, no matter whether the correlations were measured as the co-regulation frequency (affinity) or expression profile distance.