Evaluate GSA extension on p53 data set

Zuguang Gu

August 23, 2024

The public p53 data set contains expression profile with 33 samples carrying mutated p53 and 17 samples classified as normal [4]. The aim is to find which processes are affected due to dysfunction of p53. We applied CePa and standard GSA method on NCI-Nature catalogue from PID database. For CePa, we used the absolute t-value as node-level statistic and mean as the pathway level transformation. For standard GSA, we also used the absolute t-value as the gene-level statistic and mean value as the pathway score. P-values are calculated from 1000 sample permutations and are adjusted (FDR) through BH method [1]. Cutoff of FDR was set to 0.1. Results from CePa and standard GSA are illustrated in figure 1.
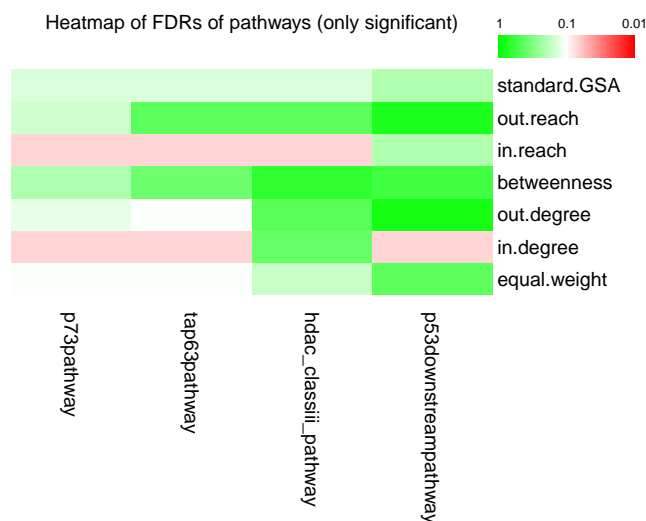

Figure 1: Heatmap of significance of pathways that are found by CePa and standard GSA.

In figure 1, FDRs are represented by continues colors in which red represents significant and green for insignificant. It is straightforward to see that standard GSA failed at identifying any significant pathways while CePa found four. The four pathways are all closely related to p53 functions in which p73 and p63 are p53 homologues [2] and histone deacetylase (HDAC) class III is essential for p53-mediated apoptosis [3].

Pathways are evaluated as significant according to choice of centrality options. P53downsteampathway is significant under in-degree centrality (FDR = 0.069, p-value = 0.001). Network graph for p53downsteampathway under in-degree is illustrated in figure 2. The differential expression for each node is colored by node's t-values in which red refers to up-regulation and green refers to down-regulation. In Figure 2, we only labeled nodes that have the top 5% largest centralities with symbols of member genes. It can be seen that nodes with large in-degree are highly down-regulated in the pathway (CDKN1A: t-value = -4.70, p-value = 0.00016; BAX: t-value = -4.67, p-value = 0.00016; BBC3: t-value = -2.59, p-value = 0.01604). The double contribution of differential expression and centrality improve the effect of important nodes and thus result in the significance of the pathway (it can be confirmed from null distribution of node-level score and pathway score in figure 3).
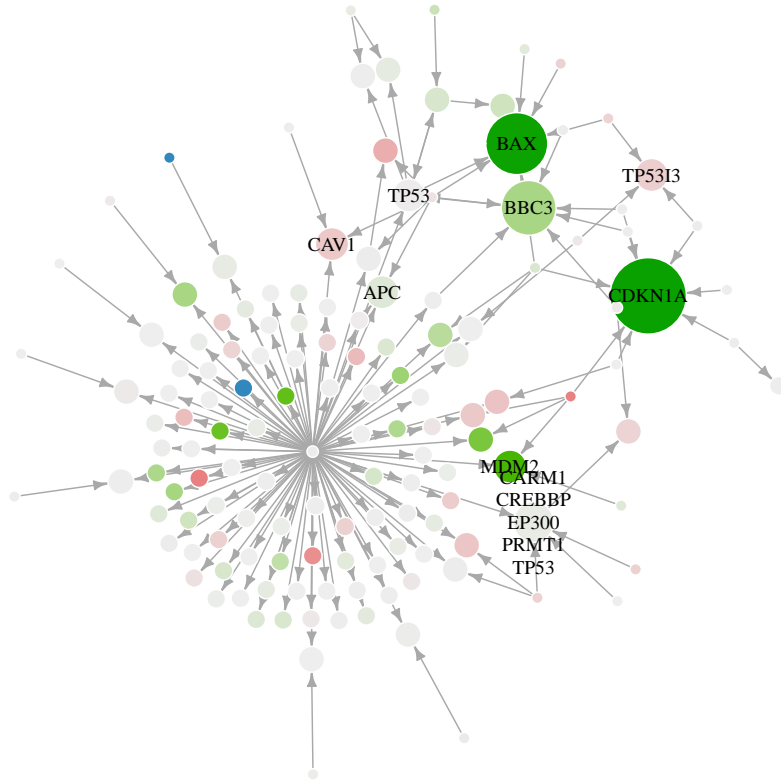

Figure 2: Network graph of p53downsteampathway. In the graph, size of nodes corresponds to the value of in-degree. Colors correspond to the t-values of nodes in which red means up-regulation, green for down-regulation and light grey for no-regulation.
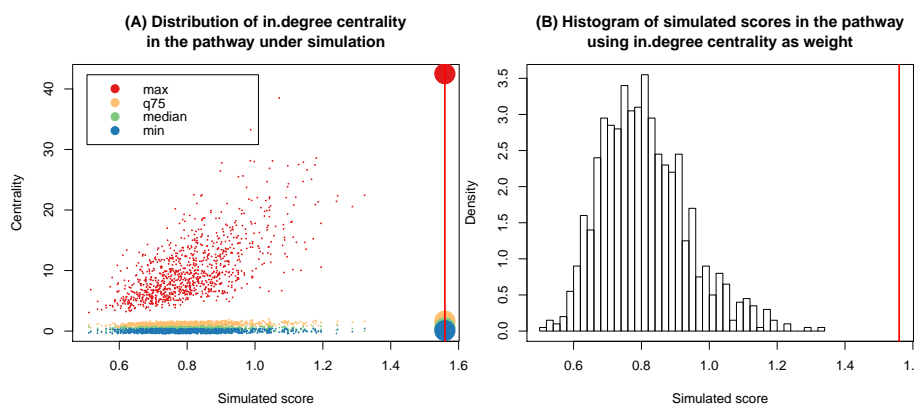
2

Figure 3: Null distribution of node-level score and pathway score. A) Distribution of node-level score under simulation. Because the pathway score is integrated through a list of node-level scores, we measure the distribution of node-level scores in a pathway by four values: maximum, the 75% quantile, median and minimum value. In figure A, pathway score for the real pathway is represented by a red vertical line and the four node scores are marked by large points. Through figure A, we can get how node-level scores contribute to the pathway scores and the significance. B) Distribution of pathway score under simulation. Pathway score for the real pathway is represented by a red vertical line.

There may be more information that can be inferred from the result. For p53downsteampathway, first the three most important nodes (BAX, BBC3 and CDKN1A) are with high in-degree. It means these nodes are controlled by many regulators, which may lead to a robust regulation. Second, these three nodes are with highly differential expression while their regulators (nodes that directly point to them) only show low differentiations. It can be explained that the regulation is additive and collaborative.

References

[1] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[2] E. R. Flores, K. Y. Tsai, D. Crowley, S. Sengupta, A. Yang, F. McKeon, and T. Jacks. p63 and p73 are required for p53-dependent apoptosis in response to dna damage. *Nature*, 416:560–564, 2002.

[3] T. Liu, P. Y. Liu, and G. M. Marshall. The critical role of the class III histone deacetylase SIRT1 in cancer. *Cancer research*, 69(5):1702–1705, Mar. 2009.

[4] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. 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, Oct. 2005.