Revealing Heterogeneity in Gene Regulation through Network Edge Coloring: A Case Study in Pediatric Pulmonary Infections

Bryant Dang BS[1], Shyam Visweswaran MD PhD[2], Asuncion Mejias, MD PhD[3], Rohit Divekar MBBS PhD[4], Suresh K. Bhavnani PhD[1]

[1]Inst. for Translational Sciences, UTMB; [2]Dept. of BMI, Univ. of Pittsburgh, Pittsburgh, PA; [3]Div. of Pediatric Infectious Diseases, Ohio State University, Columbus, OH; [4]Div. of Allergic Diseases, Mayo Clinic, Rochester, MN

Abstract

Although bipartite network visualizations have been effective in revealing heterogeneity in diseases (e.g., through patient node clusters representing subphenotypes), the resulting layouts often have many intersecting edges which can conceal important patterns such as gene regulation. Here we demonstrate the utility of coloring edges in a bipartite network to represent fold change (with respect to the controls), to help reveal differences in the gene regulation patterns among subphenotypes. The results suggest that colored edges representing fold change can help domain experts detect complex regulation patterns in bipartite networks compared to just expression values.

Introduction

Bipartite networks (which can simultaneously represent both patients and their molecular information such as gene expression), have helped to reveal heterogeneity in disease. For example, we used a bipartite network where nodes represented children less than 2 years of age infected with either influenza or respiratory syncytial virus (RSV), matched controls, 18 genes that were significantly expressed in both types of infection, and edges that represented normalized gene expression[1]. The network (Figure 1) revealed 3 clusters of patients: *core cases* that had high gene expression of 14 genes at the top of the network suggesting hyper-responsiveness, *periphery cases* that had medium expression of all 18 genes suggesting medium responsiveness, and 4 *control-like* cases that had a gene expression signature that was similar to the controls at the bottom of the network suggesting normal responsiveness. However, the network was too dense to comprehend overall patterns of gene regulation. We therefore posed the question: *Can colored edges in a bipartite network help to reveal differences in gene regulation across subphenotypes.*

Method

We median-centered all gene expression values (by dividing the median gene expression of the controls across all samples, including the controls, for each gene), and mapped the values to colors using the following range: <=0.2 (green), 1 (black), >=4 (red). As shown by the edges of the network in Figure 1, this range was used to color the edges to represent fold change. The network was presented to a domain expert in immunology to detect and interpret novel patterns related to gene regulation, and the Mann Whitney *U* test was used to quantitatively verify the significance of the pattern.

Results and Conclusion

The domain expert inspected the network and identified a new pattern related to gene regulation that was not salient in the network without the colored edges that was previously analyzed. As shown in Figure 1, the *core cases* (blue triangles and diamonds) strongly upregulated (mostly red edges) the 14 genes at the top, and strongly downregulated (mostly green edges) the 4 genes at the bottom. In contrast, the *control-like cases* had medium regulation (mostly dark edges) of both sets of genes. Statistically, the median expression of the core cases (Median=4) across the 14 genes on the top was significantly (U=625, p<.001) higher compared to the median expression (Median=0.26) of the 4 genes at the bottom. In contrast, the median expression (Median=1.54) of the control-like cases across the 14 genes on the top was not significantly different (U=15, p<.06) compared to the median expression (Median=0.93) of the 4 genes at the bottom. The results therefore revealed an important difference in the amplitude of genomic perturbation across subphenotypes providing a testable translational application to clinical outcomes. Coloring edges in a bipartite network therefore appear to be an effective way to reveal gene regulation differences amongst subphenotypes.

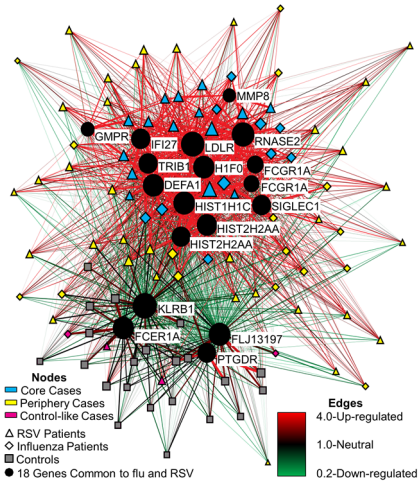

Figure 1. A bipartite network of patients with flu or RSV, 18 common genes, with edge length representing normalized gene expression, and edge color as fold change.

Acknowledgements: Supported by IHII, & NIH UL1TR000071 UTMB CTSA.

References

1. Bhavnani S.K., et al. Heterogeneity within and across Pediatric Pulmonary Infections: From Bipartite Networks to At-Risk Subphenotypes. *AMIA Summit on Translational Bioinformatics* (in review).