Examining the Expression Variance of Non-Small Cell Lung Carcinoma Using a Protein-Protein Interaction Network

Kendyl Douglas, Ami Radunskaya (Advisor)
Systems Imagination Inc., Scottsdale AZ, Pitzer College, Department of Mathematics, Claremont, CA

Abstract

Using a data set of female non-smokers with non-small cell lung carcinoma (NSCLC) and a protein-protein interaction (PPI) network, we identified different trends in expression variance between the tumorous and normal states. We find significant differences between the expression variance in the protein interactions of normal and cancerous genes in the protein interaction network. Moreover, we find significant differences between the neighborhood expression variance of proteins in the normal and tumor states. From these two findings, we gain a better understanding of the NSCLC protein-protein interaction network.

Introduction

Higher expression variance (EV) can be interpreted as a change in the protein products of regulated genes. Lower EV can be interpreted as a change in the protein products of non-regulated genes. The higher EV genes have been known to be sensitive to extracellular conditions. The lower EV genes have been known to be stable in a large range of extracellular environments. When we say "stable," we mean the lower EV gene response is restricted when its pathway is activated. Physiologically, proteins that have high EV are known to be found on the edge of the cell. These are the extracellular receptor ligands and membrane receptors. The proteins that have low EV are known to be an integral member of signal transduction or cellular communication pathways. Some genes have little variation in response to a change in an extracellular conditions and others have a lot of variation under the same extracellular conditions.[2]

The PPI network uses vertices and edges to illustrate physical protein interactions. The vertices are represented as proteins, and the edges are represented as the pairwise protein-protein interactions between them.[4,5]

We hoped to find an association or trend within the group of genes in a PPI network with proteins of high EV and within the group of low EV. Furthermore, we want to make inferences about how proteins gather in a network and about protein behavior in general.

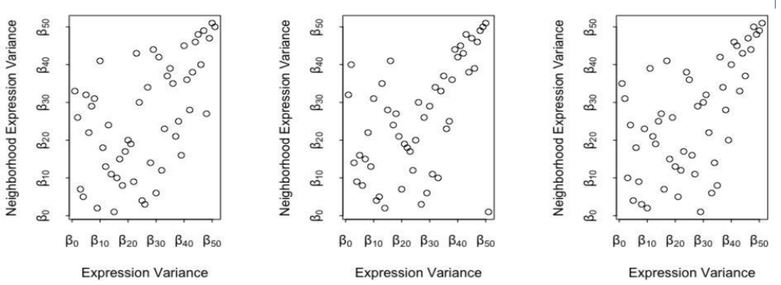

Figure 1. The neighborhood EV plotted against the EV of the normal (left), tumor (middle) and normal+tumor (right) groups in the dataset. The proteins with high EV are in the higher numbered bins. The proteins with low expression variance are in the lower numbered bins.

Materials and Methods

1. We obtained the quantile normalized genomic EV for the tumor, normal and the entire dataset in R
2. EV was assigned to the vertex representing each protein in the network. Then we obtained the mean EV of the neighborhood of individual proteins. The neighborhood EV of a protein v is defined as
$$\frac{\sum_{j=1}^{N+1} v_j}{|N+1|}$$
where v_j is the variance of protein *j* and *j* ranges over all neighbors of node i.
3. Sorted our proteins into 50 bins. Each bin, denoted β, contained proteins with similar EV values. Similarly, we sorted the neighborhood EV values into 50 bins.
4. The median EV of each bin was plotted against the neighborhood EV for the tumor, nontumor and the whole dataset including the non-tumor and tumor genes (Figure 1).
5. Then generated a 50×50 matrix of the interactions between proteins (Figure 2).

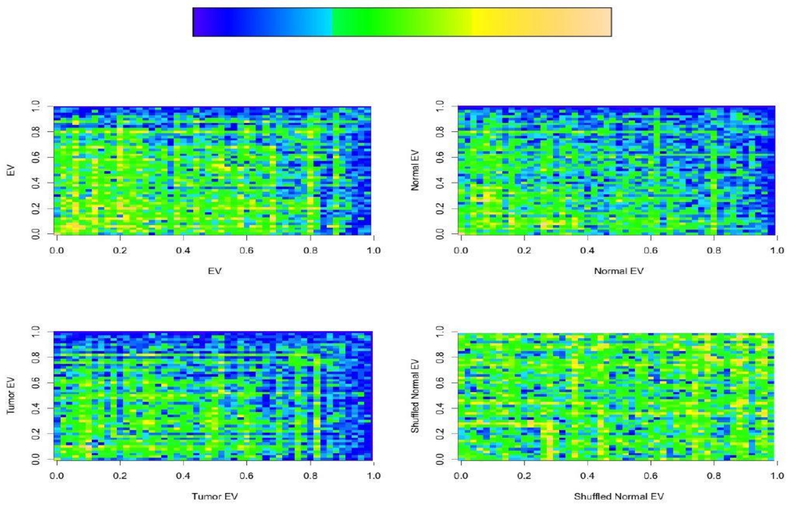

Figure 2. Heat map of the interactions between bins of the entire network (a), the normal proteins exclusively (b), the tumor proteins exclusively (c) and a randomized network where the interactions between proteins had been shuffled and then placed into bins (d).

Results

- Weak positive correlation for the non-cancerous proteins (Spearman's, ρ = 0:54, n = 8207, P = 4.1× 10^{-5}) (Figure 1).
- Weak positive correlation between the EV against the neighborhood EV after isolating the tumor genes from the non-cancerous genes (Spearman's, ρ = 0:52; n = 8207; P=7.5 × 10^{-5}) (Figure 1).
- Moderate positive correlation between an individual protein's EV and the neighborhood EV of the entire dataset (Figure 1) (Spearman's, ρ = 0:58; n = 8207; P = 9.17 × 10^{-6}).

All heat maps depicted a high density of interactions among the proteins in the low EV bins (Figure 2a, b, c).

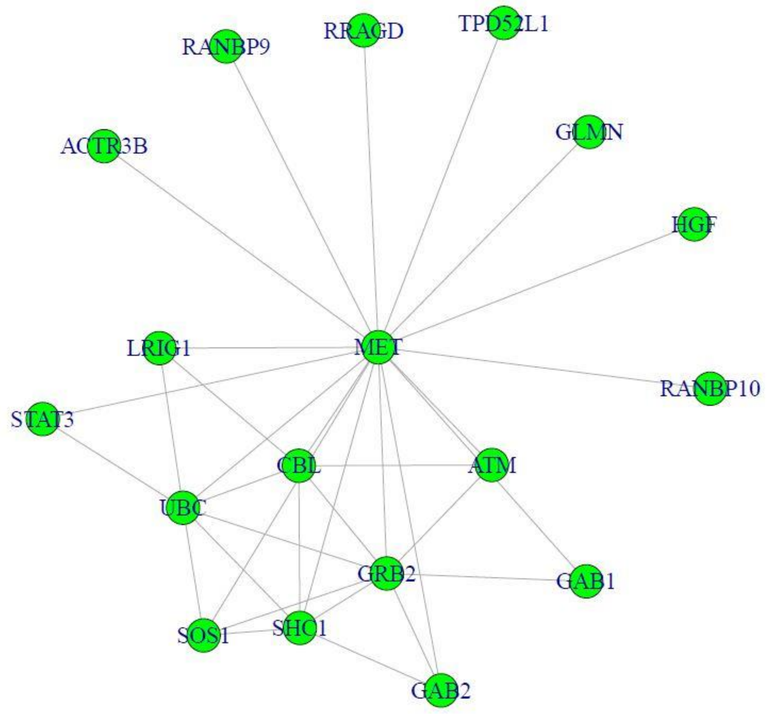

Figure 3. The gene MET and its neighborhood. The MET gene has been identified as an oncogene for NSCLC.

Discussion

High and low variance genes have been known to conglomerate with other genes of the same, or similar, EV-value to form functionally specific network neighborhoods.[2,3] However, in this study genes of average EV had virtually no interactions. Since the high variance nodes have high EV (Figure 1) and low EV nodes are connected to low EV nodes (Figure 2), it would seem that we would likely observe functionally specific network neighborhoods given further research. It would be interesting to further pinpoint the discrepancies that exist between our data and other data.[2,3]

High node degrees influenced the number of interactions between bins (Figure 2). In order to test how gene connectivity is correlated with variation in expression, we could obtain the distribution between high and low variance genes.

In a *Saccharomyces* dataset, the relationship between EV and number of interactions was more apparent when looking at proteins of high node degrees, otherwise known as hubs.[2] Hubs are defined as proteins with 7 or more interactions.[1,3] Examining these hubs would allow us to get a better sense of how proteins are gathering in this NSCLC network with respect to their EV. Hubs, or functionally-specific neighborhoods, typically provide robustness to the cell, and if there is an absence or less functionally-specific network neighborhoods, we could gain a larger grasp of this network.

Contact

Kendyl Douglas
Systems Imagination Inc.
Email: kendyldouglas@gmail.com
Phone: (310) 318 4968

References

1. Han,J.D.,N.Bertin,D.S.Goldberg,G.F.Berriz,L.V.Zhang,D.Dupuy, and A.J. Walhout. "Evidence for Dynamically Organized Modularity in the Yeast Protein-Protein Interaction Network." N.p., 1 July 2014.
2. Komurov, Kakajan, and Prahlad T Ram."Patterns of Human Gene Expression Variance Show Strong Associations with Signaling Network Hierarchy." BMC Systems Biology 4 (2010): 154. 31 Mar. 2016.
3. Komurov K., White M. "Revealing static and dynamic modular architecture of the eukaryotic protein interaction network". Mol Syst Biol. (2007); 3:110.
4. Lehne,B.,andT.Schlitt."Protein-ProteinInteractionDatabases: Keeping up with Growing Interactomes." Human Genomics, 3 Apr. 2009. In: Human Genomics, Vol. 3, No. 3, 04.2009, p. 291 - 297.
5. Yook, Soon-Hyung, Zoltn N. Oltvai, and Albert-Lszl Barabsi. "Functional and Topological Characterization of Protein Interaction Networks." Proteomics PROTEOMICS 4.4 (2004): 928-42.