A System based Approach to Construct a Kaposi Sarcoma-Associated Herpesvirus (KSHV) Specific Pathway Crosstalk Network[1]

Amy Wafa Daali, Yufei Huang, Mo Jamshidi
Department of Electrical and Computer Engineering
One UTSA Circle
San Antonio, TX, USA
wdaali@ieee.org

Abstract *- Kaposi's sarcoma-associated herpesvirus (KSHV) is one of the human cancer viruses associated with Kaposi's sarcoma, a dominant AIDS-related tumor of endothelial cells, and several B-cell malignancies. Although much advancements have been made in understanding the mechanism of KSHV, signaling pathways activated during early infection and their functions in virus entry are still unclear. From a system of systems point of view, a pathway can be viewed as a biosystem consisting of group of interacting molecules such as genes and proteins to carry out a specific biological function. We developed a method to determine clusters of functionally related pathway at different time conditions as well as identifying the significant subnetworks. This article is the first known to present a system based analysis to infer a biologically relevant functional network specific to KSHV. Such network may reveal discriminative subnetwork markers for the treatment of Kaposi's sarcoma.*

Keywords: complex systems, subnetworks, crosstalk

1 Introduction

To uncover the structural design of many complex systems from natural to technological networks, it is crucial to define the rules or interactions governing their behavior [1, 2]. This task is inherently challenging due to many difficulties related to the nature of these networks: structural and dynamical complexity, network evolution, connection and node diversity [3]. Most of biological networks are represented by gene regulatory networks and protein-protein interaction networks.

The availability of large biological data from high-throughput molecular profiling technologies has allowed scientists to infer gene regulatory networks through gene expression analysis; the monitoring of expression levels of thousands of genes. Gene regulatory networks are set of interactions between large numbers of genes and their regulators. Researchers have adopted different approaches to infer these networks. Some of the popular formalisms used include dynamic Bayesian networks (DBNs), probabilistic Boolean networks (PBNs) and ordinary differential equations (ODEs) [4]. DBNS are effective in inferring causal relationships between genes but are very computational expensive and fail to capture the temporal patterns of genes across time. PBNs incorporate the probabilistic behavior of genes; it assumes that genes are either in the "ON" state or "OFF" state and hence neglecting the intermediate expression levels. On the other hand, ODEs captures the mechanistic relationships between genes in great details but estimating large number of parameters proves to be a problem. To relieve some of the mentioned limitations, the genomic and proteomic research is moving toward a system or pathway based analysis which offers greater power of biological discovery [5-13].

System biology was first coined in 1968 by Mesarovic, a pioneer in the field of systems theory [14]. It is an interdisciplinary field that focuses on studying the mechanisms underlying complex biological components from systematic measurements [15, 16]. System biology serves as an integration platform for different types of genomics, transcriptomics and proteomics data by using network-based approaches. It has been shown that cellulars components are organized into functional modules or groups comprised of interacting nodes with common functions [17-19]. Pathway analysis has produced promising results in predicting these functions that otherswise would have been difficult to achieve when considering isolated components alone. By understanding the underlying relationship between groups of genes functioning in same pathways, it is possible to develop strategies to diagnose, treat and prevent complex diseases.

1 This work was supported, in part, by Lutcher Brown Endowed Chair and Qatar National Research Fund, Department of Electrical and Computer Engineering, the University of Texas, San Antonio

In this paper we adopt a system biology approach for pathway analysis to investigate interactions between significantly enriched pathways in the context of KSHV infection and to analyze discriminative subnetworks.

2 Methodology

2.1 Time series microarray data analysis

We have applied our method to time series microarray data derived from Kaposi's Sarcoma-associated herpesvirus infection of human primary endothelial cells collected at different times t=[0,1,3,6,10,16,24,36,54,78] hours. Gene expression data obtained at t=0 hr corresponds to untreated control cell. The virus is introduced at t=1hr and maintains its latent state until t=54hr where it transitions to the lytic state (viral replication) [20]. In order to compare expression of a set of genes from an infected cell to the same set of genes from the control cell, we calculated the logarithm base 2 value of the fold change between these two samples. This value is calculated as follows for every time point with respect to the reference at t=0:

$$r_{ij} = log2\left(\frac{I_{ij}}{C_{0,j}}\right) \quad (1)$$

Where I_{ij} represents the gene expression value for gene j in the infected sample i for gene j, i=1,3,6,10,16,24,36,54,78 and $j = 1,2 \ldots . n$. Let $C_{0,j}$ represents the gene expression value for gene j in the control sample t=0hr. The fold change r_{ij} is calculated for all genes n. If $r_{ij} = 1$ then the expression of gene j in the infected sample is two-fold that of the control sample, whereas $r_{ij} = 0$ corresponds to zero change in the expressions, a positive and negative r_{ij} represent up- and down-regulation respectively.

2.2 Retrieving Active Biosystems: Enriched Pathways

To determine which biosystems are active upon viral infection, we performed rank-based gene set enrichment analysis (GSEA) where genes are ranked according to their fold change) [21]. Gene set enrichment analysis is a powerful computational method that tests whether expression of the gene set of interest is statistically significant between two different biological conditions. It is based on calculating the enrichment score Es and ranking genes by any user defined metric (difference, ratio). It computes the cumulative sum over ranked genes. This sum increases when a gene belongs to the set and decreases otherwise. The magnitude of increment is related to the correlation of a gene with the phenotype. This algorithm controls the proportion of false positives by calculating the False Discover Rate (FDR) value. We included only the significant pathways with FDR less than 0.25.

For this analysis, we used canonical pathways that consist of 880 curated gene sets collected from the Molecular Signatures Database (MSigDB) and computational studies. After running GSEA algorithm at t=1hr, we obtained 49 significant biological pathways that might be perturbed by the virus and interestingly enough none of the pathways were significantly down-regulated. Each pathway is assigned an enrichment score Es that reflects the degree to which a set *S* is overrepresented at the top or bottom of the total ranked list. This analysis was carried out on the remaining time points.

The reported up-regulated pathways were highly relevant to cell process triggered by the virus. Among these pathways, the interleukin-6 and interleukin-3 (IL-6, IL-3) pathways, IFN-gamma pathway, insulin like growth factor 1(IGF-1) pathway and PDGF-alpha pathway. Activation of the IL-6 pathway plays a critical role in immune response, inflammation and oncogenesis by controlling cell growth, gene activation, proliferation and survival [22]. IL-3 is produced mainly by activated T lymphocytes in response to virus infection. Interferon gamma is secreted from CD cells and activated natural killer cells with anti-microbial and anti-tumor properties. The Insulin like growth factor 1 signaling regulates growth in various different cell types and blocks apoptosis and has been associated with increased growth of existing cancer cells [23]. Finally, the Platelet Derived Growth Factor (PDGF) pathway which plays a central role in blood vessel formation (angiogenesis) commonly associated with kaposi sarcoma; the tumor caused by KSHV [24].

2.3 Hiearchical Clustering of Active Biosystems

After obtaining all active pathways induced by the viral infection, the enrichment scores are stored into a matrix P called the pathway matrix. P is a mxn matrix containing all pathway enrichment scores Es across all time points n=9 and m=354 representing total number of pathways used. In order to determine any relationships hidden in this large matrix, we performed hierarchical clustering analysis of values in P. We used the correlation distance as a similarity metric to compute the pairwise distances between rows. Figure 1 shows a hierarchical clustering heatmap of P_{mxn} where rows represent pathway names and columns represent time points:

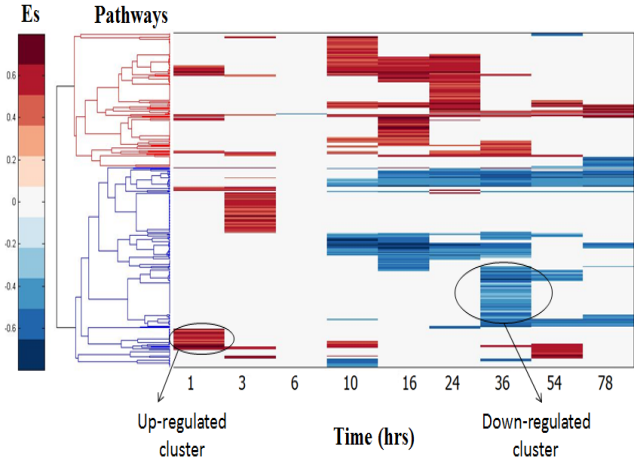

Figure 1. Heatmap of pathway matrix P_{mxn} reveals functional clustering of up- and down-regulated pathways

The strength of enrichment score of each pathway is indicated by different colors. Red denotes a positive Es corresponding to an up-regulated pathway, blue denotes a negative Es corresponding to a down-regulated pathway while white is equivalent to zero change in activity. As shown in Figure 1, it is interesting to note that first significant down-regulated pathway activities occur after t=10hr, a plausible explanation for this behavior might be due to the start of virus manipulation of host cells.

Pathways are grouped into functional blocks sharing similar up- or down-regulation patterns across different time points and therefore are more likely to share similar biological functions. It is beyond the scope of this paper to find functional relationships between all pathways. We have selected, however, the following three relevant pathways to include in our analysis: IL-3, IL-6 and Interferon gamma pathway.

3 Construction of KSHV specific pathway crosstalk network (KS-PCN)

The goal of this section is twofold: to construct a relevant functional network specific to KSHV based on pairwise correlation between genes and to analyse modular structures of the network that are functionally related. Given two vectors xi, yj representing gene expression values for two pathways, we measure the degree of similarity between genes by calculating the Pairwise Pearson Correlation (PPC) :

$$r_{ij} = \frac{\sum_{k=1}^{l} x_i(k) y_j(k)}{\sqrt{\sum_{k=1}^{l} x_i^2(k) \sum_{k=1}^{l} y_j^2(k)}} \quad (2)$$

The union of the three significantly enriched pathways resulted in 58 genes . All pairwise correlation coefficients r_{ij} between these genes are computed. By selecting gene pairs with PPC greater than 0.7, we built a KSHV specific pathway crosstalk network (KS-PCN) using Cytoscape 2.8.3 in Figure 2 [25].

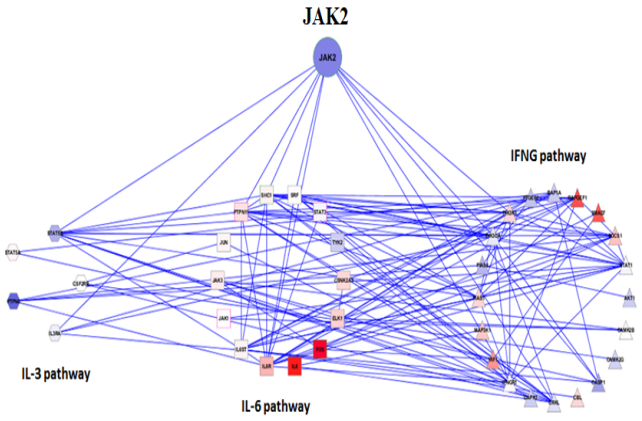

Figure 2. Visualization of KSHV specific pathway crosstalk network (KS-PCN)

Using the group attribute layout option in cytoscape, genes (nodes) belonging to a specific pathway were given a unique node shape. The crosstalk network includes 42 nodes with 125 interactions. Genes with high up-regulated activity are denoted by red, while down-regulated genes are represented by blue. The hexagon shaped genes belong to interleukin-3 pathway (IL-3); the square shaped genes belong to interleukin-6 (IL-6) pathway while the triangle shaped genes belong to interferon gamma pathway.

According to Figure 2, JAK2 acts as a major gene hub linking between IL-3, IL6 and IFNG pathway which agrees with the already existing biological knowledge [26]. Our analysis confirms the fact that JAK2 plays an essential role in activating IFNG pathway. Interferon-gamma is a cytokine with antiviral, immunoregulatory, anti-tumor characteristics and a major microphage activator. IFNGR2 binds to kinase JAk2 resulting in auto-phosphorylation and activation of this protein which in turn regulates the phosphorylation and activation of STAT1 [27]. We also confirmed that JAk2 is connected to IL-6 pathway through IL6R part of IL6-mediated signaling events [26].

In addition to verifying some of the already known biological interactions, we revealed several novel interactions between JAK2 and the three previously mentioned pathways. As shown in Figure 2, JAK2 interacts with IFN pathway through CRKL, PIK3CA, CASP1, and PIK3R1. JAK2 is related to IL-6 pathway through SRF and to IL-3 pathway through IL3RA and STAT5B.

We next extracted the important modular structures of the network using AllegroMCODE algorithm which identifies tightly connected clusters within the network [28]. It uses the weights nodes by finding local neighbourhood density to extract densely connected regions. After running the algorithm, we obtained the following largest subnetwork in Figure 3.

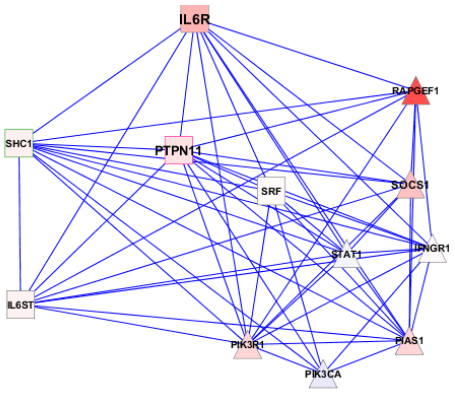

Figure 3- IL6ST subnetwork

The interleukin 6 signal transducer (IL6ST) subnetwork ranked first in the analysis with a score of 4.7 , 12 nodes and 56 edges. It is a signal transducer molecule essential for many cytokines. IL6ST is activated when cytokines bind to their receptors. Surprisingly, it has been reported that vIL6, viral protein related to IL6 encoded by the Kaposi sarcoma-associated herpesvirus can directly activate this signal transducer IL6ST. Furthermore, this network is characterized by the overexpression of the following genes IL6R and RAPGEF1 (denoted in red). The majority of KSHV infected B cells express elevated levels of IL-6R but the exact mechanism of how it effects the cells is still unclear [29]. The highly up-regulated RAPGEF1 activates several elements of the Ras family; Ras is the most common oncogene in human cancer found in 20-25% of all human tumors [30]. This discriminative subnetwork could be used as a target for therapy by possibly inhibiting IL6R or RAPGEF1 and possibly controlling the activity of RAS pathway.

Figure 4 shows the structure of the other two modular subnetworks which clearly have a closed-loop network topology. In comparison to the previous IL6ST subnetwork, they differ dramatically in size, function and number of node connections .

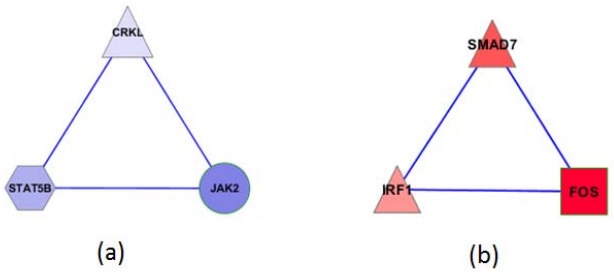

Figure 4 : (a) CRKL down-regulated subnetwork
(b) SMAD7 up-regulated subnetwork

As shown in Figure 4 (a), CRKL, a v-crk sarcoma virus CT10 oncogene homolog like , activates the RAS and JUN kinase signaling pathways [22]. Association between the transcription factor STAT5B and Crkl was not observed but our analysis suggests a functional relationship between them. In Figure 4 (b), the up-regulated subnetwork consists of SMAD7 forming a closed loop with the following transcription factors FOS and IRF1. It has been shown that SMAD7 interacts with TGF-beta receptor type-1 causing the degradation to TGFBR1. Dysregulation of this gene results in an increased susceptibility to colorectal cancer and based on our observation there is a possibility that it contributes to kaposi sarcoma infection [22]. Furthermore, we infer that FOS interacts with IFN gamma pathway through SMAD7 and IRF1.

In order to draw more powerful conclusions from these sub-graphs, algorithms such as mutual information network inference methods could be used to infer edge directionally [31]. Most three-node networks in biology take the form of a feed-forward loop (FFL) loop as discussed by Uri Alon et al. [32]. The feed-forward loop is a three gene pattern consisting of two input transcription factors both regulating same target gene. It was proved that biological networks use these type of loops as a mechanism for speeding response time in transcription networks [32].

4 Conclusion

Kaposi sarcoma (KS) is an incurable and devastating malignancy associated with HIV in AIDS patients. A better understanding of the underlying KSHV mechanism is very crucial to the development of effective KS treatment. In this paper, we verified the functional relationship between three significantly enriched pathways taken from the same up-regulated cluster: interlukin-3, interlukin6 and IFNG pathways. Using gene set enrichment analysis, we uncovered the functional clustering of up- and down-regulated pathways across different time conditions. As a result, we were able to construct a KSHV specific pathway crosstalk network in early stage of infection by calculating pairwise Pearson correlation between every gene pair. To

validate the predicted novel interactions between these pathways, biological experiments are needed to prove these dependencies. The system level approach proved helpful in understanding complex biological systems within the context of KSHV infection.

Acknowledgement

The authors acknowledge fund support from Lutcher Brown Endowed Chair and Qatar National Research 09-874-3-235, Department of Electrical and Computer Engineering, the University of Texas, San Antonio

References

[1] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: simple building blocks of complex networks," *Science Signaling,* vol. 298, p. 824, 2002.

[2] S. Draghici, P. Khatri, A. L. Tarca, K. Amin, A. Done, C. Voichita*, et al.*, "A systems biology approach for pathway level analysis," *Genome research,* vol. 17, pp. 1537-1545, 2007.

[3] S. H. Strogatz, "Exploring complex networks," *Nature,* vol. 410, pp. 268-276, 2001.

[4] H. De Jong, "Modeling and simulation of genetic regulatory systems: a literature review," *Journal of computational biology,* vol. 9, pp. 67-103, 2002.

[5] P. Khatri, M. Sirota, and A. J. Butte, "Ten years of pathway analysis: current approaches and outstanding challenges," *PLoS computational biology,* vol. 8, p. e1002375, 2012.

[6] V. Varadan, P. Mittal, C. J. Vaske, and S. C. Benz, "The Integration of Biological Pathway Knowledge in Cancer Genomics: A review of existing computational approaches," *Signal Processing Magazine, IEEE,* vol. 29, pp. 35-50, 2012.

[7] V. K. Ramanan, L. Shen, J. H. Moore, and A. J. Saykin, "Pathway analysis of genomic data: concepts, methods, and prospects for future development," *Trends in Genetics,* vol. 28, pp. 323-332, 7// 2012.

[8] Y. Li, P. Agarwal, and D. Rajagopalan, "A global pathway crosstalk network," *Bioinformatics,* vol. 24, pp. 1442-1447, 2008.

[9] T. Ideker and N. J. Krogan, "Differential network biology," *Molecular systems biology,* vol. 8, 2012.

[10] K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A.-L. Barabasi, "The human disease network," *Proceedings of the National Academy of Sciences,* vol. 104, pp. 8685-8690, 2007.

[11] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics,* vol. 12, pp. 56-68, 2011.

[12] H.-Y. Chuang, E. Lee, Y.-T. Liu, D. Lee, and T. Ideker, "Network-based classification of breast cancer metastasis," *Molecular systems biology,* vol. 3, 2007.

[13] X. Wu, R. Jiang, M. Q. Zhang, and S. Li, "Network-based global inference of human disease genes," *Molecular Systems Biology,* vol. 4, 2008.

[14] M. D. Mesarović, *Systems theory and biology: proceedings*: Springer-Verlag, 1968.

[15] H.-Y. Chuang, M. Hofree, and T. Ideker, "A decade of systems biology," *Annual review of cell and developmental biology,* vol. 26, p. 721, 2010.

[16] M. A. Fischbach and N. J. Krogan, "The next frontier of systems biology: higher-order and interspecies interactions," *Genome Biol,* vol. 11, p. 208, 2010.

[17] B. Alberts, "The cell as a collection overview of protein machines: Preparing the next generation of molecular biologists," *Cell,* vol. 92, pp. 291-294, 1998.

[18] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature,* vol. 402, pp. C47-C52, 1999.

[19] I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria*, et al.*, "Dynamic modularity in protein interaction networks predicts breast cancer outcome," *Nature biotechnology,* vol. 27, pp. 199-204, 2009.

[20] S.-J. Gao, J.-H. Deng, and F.-C. Zhou, "Productive lytic replication of a recombinant Kaposi's sarcoma-associated herpesvirus in efficient primary infection of primary human endothelial cells," *Journal of virology,* vol. 77, pp. 9738-9749, 2003.

[21] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette*, et al.*, "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,* vol. 102, pp. 15545-15550, 2005.

[22] National Center for Biotechnology Information Database [Online]. Available: http://www.ncbi.nlm.nih.gov/gene/

[23] F. I. Arnaldez and L. J. Helman, "Targeting the insulin growth factor receptor 1," *Hematology/oncology clinics of North America,* vol. 26, p. 527, 2012.

[24] M. Hannink and D. J. Donoghue, "Structure and function of platelet-derived growth factor (PDGF) and related proteins," *Biochimica et biophysica acta,* vol. 989, pp. 1-10, 1989.

[25] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage*, et al.*, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome research,* vol. 13, pp. 2498-2504, 2003.

[26] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez*, et al.*, "The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored," *Nucleic acids research,* vol. 39, pp. D561-D568, 2011.

[27] K. B. Nguyen, L. P. Cousens, L. A. Doughty, G. C. Pien, J. E. Durbin, and C. A. Biron, "Interferon α/β-mediated inhibition and promotion of interferon γ: STAT1 resolves a paradox," *Nature immunology,* vol. 1, pp. 70-76, 2000.

[28] G. D. Bader and C. W. Hogue, "An automated method for finding molecular complexes in large protein interaction networks," *BMC bioinformatics,* vol. 4, p. 2, 2003.

[29] M.-Q. Du, H. Liu, T. C. Diss, H. Ye, R. A. Hamoudi, N. Dupin*, et al.*, "Kaposi sarcoma–associated herpesvirus infects monotypic (IgMλ) but polyclonal naive B cells in Castleman disease and associated lymphoproliferative disorders," *Blood,* vol. 97, pp. 2130-2136, 2001.

[30] J. Downward, "Targeting RAS signalling pathways in cancer therapy," *Nature Reviews Cancer,* vol. 3, pp. 11-22, 2003.

[31] A. Noor, E. Serpedin, M. Nounou, H. Nounou, N. Mohamed, and L. Chouchane, "Information theoretic methods for modeling of gene regulatory networks," in *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012 IEEE Symposium on*, 2012, pp. 418-423.

[32] U. Alon, *An introduction to systems biology :design principles of biological circuits*. Boca Raton, FL: Chapman & Hall/CRC, 2007.