Transcriptional Regulatory Elements and Factors of Kaposi's Sarcoma Human Virus

Manjula Kasoji, Department of Bioinformatics Georgia Institute of Technology

Bino John, Mentor, Department of Computational Biology, University of Pittsburgh School of Medicine

Introduction

Kaposi's sarcoma human virus (KSHV) is a tumor developing agent that belongs to the human herpes virus 8 family. Kaposi's sarcoma is now the fourth most common cancer caused by an infectious agent worldwide following gastric, cervical, and hepatic cancers. It is a cancerous disease that causes abnormal tissue growth under the skin and leads to lesions. KSHV is usually not life threatening unless it reaches to the internal organs such as lungs, liver, and gastrointestinal tract. Functional studies of KSHV indicate that its genes interfere with established tumor suppressor pathways, modify the host cellular environment, and thus contribute to the pathogenesis of KSHV-associated disorders.

The goal of this experiment is to identify the transcriptional regulatory sites and factors that are involved in the transcription of the viral genome. Finding out more about transcriptional regulation will allow us to help decrease the spread of KSHV by disrupting its function. We believe that the identification of regulatory elements and factors is an attainable goal because KSHV uses human transcription factors to carry out its function.

We will attempt to identify regulation sites or motifs by using a computational method called K-factor. K-Factor predicts regulatory motifs in a set of functionally related sequences [Li]. KSHV is not evolutionary conserved therefore we are using K-factor because it does not rely on evolutionary conservation, and it is not based on existing motifs from a database as opposed to other motif prediction programs. The transcription factors will also be identified by using other computational programs that predict human transcription factors. KSHV does not have transcription factors of its own. It uses human factors to regulate its transcription. This is why we are searching for associated human transcription factors.

Research Design and Methods

This experiment involves two major steps to reach our goals, the identification of transcription factors and motifs, and the validation of these findings. Identification involves obtaining the KSHV genome, identifying all of the protein encoding genes, finding the upstream regions of up to one thousand nucleotides, or 1 kb of the protein coding genes, and using programs to identify the factors and motifs that are involved in these upstream regions. We are searching for sequences of at least 1kb in length upstream of protein coding genes because these are the regulatory areas. The KSHV genome and its protein coding genes will be identified by via a search in the National Center for Biotechnology Information (NCBI) website. Using these locations, the upstream sequences of all of the genes will be obtained.

For validation, first a set of random sequences will be generated by a simple Perl script. These random sequences will act as reference or control sequences to which the upstream sequences will be compared against. Both the upstream region sequences and the reference sequences will be entered into the K-factor program which will subsequently predict transcription factor binding motifs. The results from the random sequences will allow us to validate the out put of the prediction programs. If the programs are giving the same results for both the random and upstream sequences then we will know that the program is not reporting correctly; the results should be different since the random sequences do not contain any legitimate data. These predicted motifs will be matched with predicted human transcription factors that will be obtained via programs obtained from the World Wide Web.

Expected Results

The results expected from this experiment is the identification of regulatory elements and transcription factors that bind to those elements of KSHV protein coding genes. The KSHV genome is large so we expect to find numerous elements and factors. These elements and factors are expected to be correlated to upstream human protein coding genes.

There are three main results that may be expected from this experiment. KSHV is a virus that causes cancer by using human transcription factors. Therefore, the transcription factors that are identified to be associated with KSHV may already be associated with cancer causing genes in humans. Also, transcription factors of all cell types in the human body are recognized by KSHV since it may infect any cell type. This leads to the notion that the transcription factors associated with KSHV might be ubiquitously expressed. Lastly, compared to the human genome, the KSHV genome is relatively small. So, in order to use the human genome KSHV must have highly condensed information content; KSHV may have more regulatory motifs per kilo base of sequences than humans, and the KSHV genes may have evolved to be recognized by the human genome.

Reference List

Bulyk, M. L. "Computational prediction of transcription-factor binding site locations." Genome Biol. 5.1 (2003): 201.

Fukao, T., et al. "An evolutionarily conserved mechanism for microRNA-223 expression revealed by microRNA gene profiling." Cell 129.3 (2007): 617-31.

Li, Ji., et al. “Regulatory Circuit of Human MicroRNA Biogenesis.” PLoS Computational Biol. 3 (2007): 0721-0732.