Genomic Evolution of Cancer Associated Viruses and Bacteria

Angelo Kontgas, Department of Chemistry and Biochemistry,
Utah State University, Logan, UT.

Research Mentor: Bino John, Department of Computational Biology,
University of Pittsburgh, Pittsburgh, PA

Specific Aims

The past several decades of research have linked several viruses and bacteria that predispose humans and animals to the development of cancer[1,2]. The molecular mechanisms which lead to the development of cancers that are linked to cancer-associated pathogens are not clearly understood. A common mechanism by which viruses evolve is known to play a central role in promoting cancer. The process involves the incorporation of genetic material from the DNA of the viral host genomes to form new viral genes that interfere with the normal functions of the host cells. Such viral genes are known to promote unrestricted cellular proliferation which leads to tumorigenesis[3]. Genes that are capable of promoting uncontrolled cellular growth upon mutation are called oncogenes. Viral oncogenes that are altered with respect to their host and are conserved between distantly related cancer-pathogens are promising candidates to understand the biological basis of cancer.

Recent studies have unexpectedly identified a large number of highly conserved genomic regions between distantly related species such as human, mouse, fish and flies[4]. Therefore, we hypothesize that similar approaches may unravel important evolutionarily conserved elements in several of the cancer-associated viral and bacterial genomes. Specifically, we propose to investigate the degree and extent of sequence conservation of genomic sequences between distantly related viruses, bacteria and the human genome. We expect that the proposed research will reveal novel evolutionary patterns between cancer-pathogens and humans.

Research Design & Methods

We have begun to curate the genomic sequences of several cancer-associated pathogens. We will analyze the evolutionary conservation of the Kaposi Sarcoma Herpes Virus (KSHV) genome with respect to the curated viral and bacterial genomes. The KSHV genome sequence will be used to scan the curated genome sequences using BLAST[5]. The BLAST output will be analyzed to identify sequences ("hits") that are conserved between KSHV and the scanned genomes. The BLAST E-value parameter that measures the statistical significance of a match between two given sequences will be set to a conservative threshold (E=0.0001). It is possible that the analyzed genomes may not share common sequence elements to the KSHV genome. If conserved sequence elements are not detected, we will apply the aforementioned approach to each genome in the curated list until novel sequence elements that are potential candidates for promoting the oncogenicity of the cancer-pathogens are identified. If common sequence elements are detected between genomic regions of several viruses, we will analyze the extent of conservation of these sequences by aligning the sequence hits using CLUSTALW[6]. We will also use MFOLD to fold the conserved sequences to investigate the presence of unusual secondary structures among the sequences[7]. The analysis of predicted secondary structures may help reveal structural constraints that may have helped preserve the sequence elements over millions of years of pathogen evolution. A similar approach in mammalian genomes has provided significant insights into the evolution of human genome[8].

Summary

The proposed study will help us to investigate whether the cancer-associated pathogens have common sequence signatures that may be identified at the level of their genome sequences. If novel molecular signatures are identified in this study, a broader investigation is planed to probe

the biological function of the conserved sequences and their evolution with respect to other animal genomes. It is likely that such commonly occurring sequences that are preserved over millions of years of viral, bacterial and animal evolution may hold key clues into the roles of cancer-associated pathogens in cancer. We expect that future investigations based on the current study will discover novel genes that are yet to be characterized.

Reference List

1. McCance, D. J. *Human tumor viruses*. American Society for Microbiology, Washington, D.C (1998).

2. Pellicano, R., Mladenova, I., Martinotti, R., Fagoonee, S. & Rizzetto, M. [Gastric cancer and Helicobacter pylori: an interdisciplinary point of view]. *Minerva Med.* **97**, 31-38 (2006).

3. Gatza, M. L., Chandhasin, C., Ducu, R. I. & Marriott, S. J. Impact of transforming viruses on cellular mutagenesis, genome stability, and cellular transformation. *Environ. Mol. Mutagen.* **45**, 304-325 (2005).

4. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034-1050 (2005).

5. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402 (1997).

6. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680 (1994).

7. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415 (2003).

8. Pedersen, J. S. *et al.* Identification and classification of conserved RNA secondary structures in the human genome. *PLoS. Comput. Biol.* **2**, e33 (2006).