In Silico Analysis of the CST6 Tumor Suppressor Gene

Athanasia Pavlopoulou, Department of Biotechnology, Bioinformatics and Medical Informatics Team, Biomedical Research Foundation of the Academy of Athens, Athens, Greece

Georgios Tsaramirsis, Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

ABSTRACT

The gene encoding cystatin E/M, CST6, is a Class II tumor suppressor. Using bioinformatics tools for database mining and virtual gene expression profiling, the authors showed that CST6 is differentially expressed in various types of cancer. Moreover, epigenetic silencing mediated by hypermethylation of the CpG island located at the CST6 promoter was found to be conserved in mammalian species. Comprehensive analyses of animal genomes led to the identification of novel CST6 transcript orthologs and splice variants that enabled us to trace the evolutionary origin of CST6. Moreover, eight novel and potentially regulatory SNPs were identified in CST6 gene. Conserved cancer-relevant regulatory elements were identified in the CST6 gene promoter. In addition, miRNAs that are differentially expressed in human cancers were identified as putative posttranscriptional regulators of CST6. Collectively, the authors suggest that expression of CST6 in normal and cancer cells is coordinately regulated by genomic, transcriptional and post-transcriptional mechanisms.

Keywords: Breast Cancer, Cystatin E/M (CST6), Epigenetic Silencing, In Silico Analysis, Micro Ribonucleic Acid (miRNAs), Single-Nucleotide Polymorphisms (SNPs), Transcript Variants

INTRODUCTION

The gene CST6 was cloned by differential display based on its highly downregulated expression in 21MT-1 metastatic breast cancer cell line when compared to 21PT cell line isolated from the primary tumor of the same patient, while it was found inactivated in the majority of metastatic breast cancers (Sotiropoulou, Anisowicz, & Sager, 1997). The corresponding gene (CST6) was mapped to chromosomal locus 11q13 (Stenman et al., 1997), which is believed to harbor tumor suppressor genes because loss of heterozygosity (LOH) has been frequently observed in several types of cancer. Later the same gene was identified by expressed sequence tag (EST) sequencing in amniotic cells and fetal skin epithelial cells and was named cystatin E (Ni et al., 1997). Cystatin M, re-named cystatin E/M, is an endogenous inhibitor of asparaginyl endopeptidase/legumain and of the lysosomal cysteine proteases cathepsins B and L (Shridhar et al., 2004; J. Zhang et al., 2004). Increased proteolytic activities of lysosomal proteases are known to promote cancer progression, invasion and metastasis via degradation of the basement membrane and extracellular matrix (ECM) (Mohamed & Sloane, 2006). In this respect, it was shown that cystatin E/M is a tumor suppressor for breast cancer (Shridhar...
et al., 2004; J. Zhang et al., 2004). Further, cystatin E/M was shown to inhibit adhesion of breast cancer cells to endothelial cells, an important step for cancer metastasis (Shridhar et al., 2004), as well as invasion of human melanoma by suppressing legumain activity (Briggs et al., 2010). In a recent study, CST6 was also shown to suppress the proliferation, colony formation, migration and invasion of breast cancer cells (Jin et al., 2012).

Several studies have shown that aberrant hypermethylation of a CpG island located in the proximal promoter and first exon is associated with epigenetic silencing of CST6 expression in breast cancer cell lines and primary breast carcinomas, as well as in other types of cancer (Ai et al., 2006; Kioulafa et al., 2009; Qiu et al., 2008; Veena et al., 2008). In cancer cells, promoter CpG islands are prone to de novo methylation that prevents binding of regulatory transcription factors (TFs) to their cognate DNA sequences or TF-binding sites (TFBS) (Jones & Baylin, 2002). On the other hand, microRNAs (miRNAs) are small non-coding RNAs which negatively regulate gene expression at the post-transcriptional level acting by sequence-specific base-pairing at the 3’-UTR of the target mRNA that induces either mRNA degradation or translation inhibition by masking the ribosome-binding sequence. While it is now well established that miRNAs represent a distinct level of epigenetic gene regulation, an expanding body of evidence suggests a strong link between miRNAs and human diseases, including most types of cancer (Sotiropoulou, Pampalakis, Lianidou, & Mourelatos, 2009). Certain miRNAs are aberrantly expressed in cancer cells and were shown to contribute to tumor growth and/or progression functioning either as oncogenes (oncomirs) or tumor suppressors (Sotiropoulou et al., 2009).

Therefore, to this end, we searched for a potential association between miRNAs and CST6 silencing in cancer cells. In order to identify cancer-associated transcriptional and post-transcriptional regulators of CST6, we employed bioinformatics tools to search both the CST6 promoter and 3’-UTR for TFBS and miRNA targets, respectively. Moreover, the association between hypermethylation of the CpG island located in the CST6 promoter and to epigenetic silencing of cystatin E/M tumor suppressor in breast cancer cells was examined. Finally, we identified novel SNPs, as well as novel transcripts and cystatin E/M orthologous proteins by comprehensive phylogenetic analysis.

**METHODS**

**Identification of Single Nucleotide Polymorphisms (SNPs)**

SNPs represent the most common genetic variation. The human CST6 gene (GenBank accession no: NC_000011.9) was searched for SNPs using the NCBI dbSNP (Saccone et al., 2011) and CGAP SNPpipeline (Clifford et al., 2000) publicly available databases.

**Identification of CST6 Transcripts**

The cystatin E/M full amino acid sequence was searched against EST databases of various organisms using tBLASTn (Altschul, Gish, Miller, Myers, & Lipman, 1990) to identify CST6 orthologous transcripts. The retrieved coding sequences were subsequently aligned with the corresponding genomic sequences using the sim4 program (Florea, Hartzell, Zhang, Rubin, & Miller, 1998) in order to detect intron-exon junctions. The alignments were then checked for splicing signs, specifically for gaps at least 32 nucleotide long with ends matching the GT/AG intron consensus. All identified CST6 transcripts were analyzed for translational start sites and a polyadenylation signal in the 3’ untranslated region.

**Microarray Gene Expression Data Mining**

Microarray gene expression data (transcriptome) was mined using the publicly available database NCBI GEO (Barrett et al., 2013).
Computer-Aided Drug Design and Biological Evaluation of Novel Anti-Greek Goat Encephalitis Agents

Predicting Aging-Genes in Drosophila Melanogaster by Integrating Network Topological Features and Functional Categories
Yan-Hui Li, Jian-Hui Li, Xin Song, Kai Feng and Yuan-Chun Zhou (2012). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 16-26).
[www.igi-global.com/article/predicting-aging-genes-drosophila-melanogaster/77928?camid=4v1a](www.igi-global.com/article/predicting-aging-genes-drosophila-melanogaster/77928?camid=4v1a)