

Telomere maintenance, cell cycle and cell-matrix adhesion genes are differentially expressed during progression of colon cancer from primary tumor growth to metastasis

Abstract

Gene expression in the SW480 primary tumor colon cancer cell line was compared to that in the SW620 metastatic colon cancer cell line using the Affymetrix® GeneChip® Human Genome U133A array. Using filtering criteria of a 1.5 or greater fold-change in expression, all “present calls” in at least one group and a false discovery rate of less than 5%, 1534 out of over 18,000 genes were differentially expressed. The expression of 1034 genes was increased in the metastatic cell line, while 520 showed decreased expression in the metastatic cells. The Gene Ontology and KEGG pathway terms associated with these genes were examined to identify biological themes associated with each set of genes. These analyses identified distinct biological themes associated with each group. The ontology term analysis identified an over-representation of genes involved in telomerase dependent telomere maintenance among the up-regulated genes and an over-representation of genes involved in cell matrix adhesion among the down regulated genes. The analysis of KEGG pathway terms identified a significant over representation of cell cycle genes among the up-regulated genes. The increased expression of genes involved in cell cycle and telomere maintenance, as well as the decreased expression of genes involved in cell matrix adhesion, may contribute to the progression to metastasis in colon cancer cells.

Introduction

In this study the GeneSifter® microarray data analysis system was used to analyze data generated from a comparison of the SW480 primary tumor colon cancer cell line with the SW620 metastatic colon cancer cell line. This analysis process used can be broken down into two discrete tasks: identification of differentially expressed genes and the determination of the biological significance of both individual genes and groups of genes. GeneSifter uses Gene Ontology (GO) reports, KEGG pathway reports and z-scores to summarize the biological processes represented in a particular gene list. Z-scores can then be used to identify GO terms that are significantly over- or under-represented in this list. This allows for rapid characterization of the broad biological themes affected in a particular experiment. This study outlines the use of these methods to identify biological themes associated with the progression of colon cancer from primary tumor growth to metastasis.

The Data

Gene expression profiles in the SW480 primary tumor colon cancer cell line were compared to that in the SW620 metastatic colon cancer cell line using the Affymetrix® GeneChip® Human Genome U133A array (GEO dataset accession GDS756). Three biological replicates were prepared for each cell line. CHP file data saved as text was obtained from the GEO website for each sample. The tab-delimited text files were loaded into GeneSifter and the MAS 5 derived signal and detection calls were used for analysis.

Identification of differentially expressed genes

The Affymetrix® GeneChip® Human Genome U133A array contains ~22,000 probe sets representing ~18,400 different transcripts. Using filtering criteria of a 1.5 or

greater fold-change in expression, all “present calls” in at least one group and a p-value of < 0.05 from a unpaired, two sample t-test produced a list of 1863 genes that were differentially expressed. Correction for multiple testing was then performed using the method of Benjamini and Hochberg (Reiner, et al. 2003) to derive a false discovery rate estimate from the raw p-values. A false discovery rate of 5% was used as a cutoff for statistical significance giving a final list of 1534 differentially expressed genes. 1034 of these genes showed higher expression in the metastatic cell line while 520 showed decreased expression. The parameters used for this filtering are summarized in table 1.

Parameter	Setting
Normalization	<i>Global Median</i>
Statistics	<i>t-test</i>
Quality	<i>1</i>
Threshold	<i>1.5</i>
Correction	<i>Benjamini and Hochberg</i>
Log Transformation	<i>Log transform</i>

Table 1: Parameters used to identify differentially expressed genes using GeneSifter.

Biological significance

The biological process ontologies and KEGG pathway terms associated with the differentially expressed genes were examined using a z-score report. The z-score report identifies ontologies or pathway terms that are significantly over-represented in a gene list. Among the biological processes ontologies were “telomerase-dependent telomere maintenance” which had a z-score of 4.16 for the up-regulated genes and “cell-matrix adhesion” which had a z-score of 2.22 for the down-regulated genes. The KEGG pathway term “cell cycle” had a z-score of 4.35 for the up-regulated genes. The

individual genes for are listed in Table 2 and Figure 1. See supplemental material to view a comprehensive list of ontologies and pathway terms identified.

A

Probe Set ID	Gene ID	Gene Name	Up
201174_s_at	TERF2IP	Telomeric repeat binding factor 2, interacting protein	3.50
203611_at	TERF2	Telomeric repeat binding factor 2	2.79
203448_s_at	TERF1	Telomeric repeat binding factor (NIMA-interacting) 1	2.46
203449_s_at	TERF1	Telomeric repeat binding factor (NIMA-interacting) 1	2.24
202561_at	TNKS	Tankyrase, TRF1-interacting ankyrin-related ADP-ribose polymerase	1.97
221481_x_at	HNRPD	Heterogeneous nuclear ribonucleoprotein D	1.80

B

Probe Set ID	Gene ID	Gene Name	Down
209835_x_at	CD44	CD44 antigen (homing function and Indian blood group system)	3.96
212014_x_at	CD44	CD44 antigen (homing function and Indian blood group system)	3.91
211819_s_at	SORBS1	Sorbin and SH3 domain containing 1	2.91
213867_s_at	CD47	CD47 antigen (Rh-related antigen)	1.72
204989_s_at	ITGB4	Integrin, beta 4	1.69
217007_s_at	ADAM15	A disintegrin and metalloproteinase domain 15 (metargidin)	1.53

Table 2: A) Telomerase maintenance genes up-regulated during progression to metastasis. B) Cell-matrix adhesion genes down-regulated during progression to metastasis.

Summary

1534 differentially expressed genes were identified in this comparison of the SW480 primary tumor colon cancer cell line with the SW620 metastatic colon cancer cell line. Analysis of the Gene Ontology and KEGG pathway terms

associated with these genes identified several gene families associated with tumor growth and metastasis that were enriched among these differentially expressed genes. There was an enrichment of genes involved in telomerase-dependent telomere maintenance and the cell cycle among the up-regulated genes, while cell-matrix genes were enriched among the down-regulated genes. This analysis has identified a set of genes that may contribute to the progression to metastasis in colon cancer cells.

Supplemental Material

Raw data is available from Gene Expression Omnibus (GEO), dataset GDS756.

Data can be viewed and analyzed using the GeneSifter Data Center (www.genesifter.net/web/DC).

References

- Reiner, et al. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19(3):368-375
- Doniger, et al. 2003. MAPPFinder: using Gene Ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biology* 4:R7

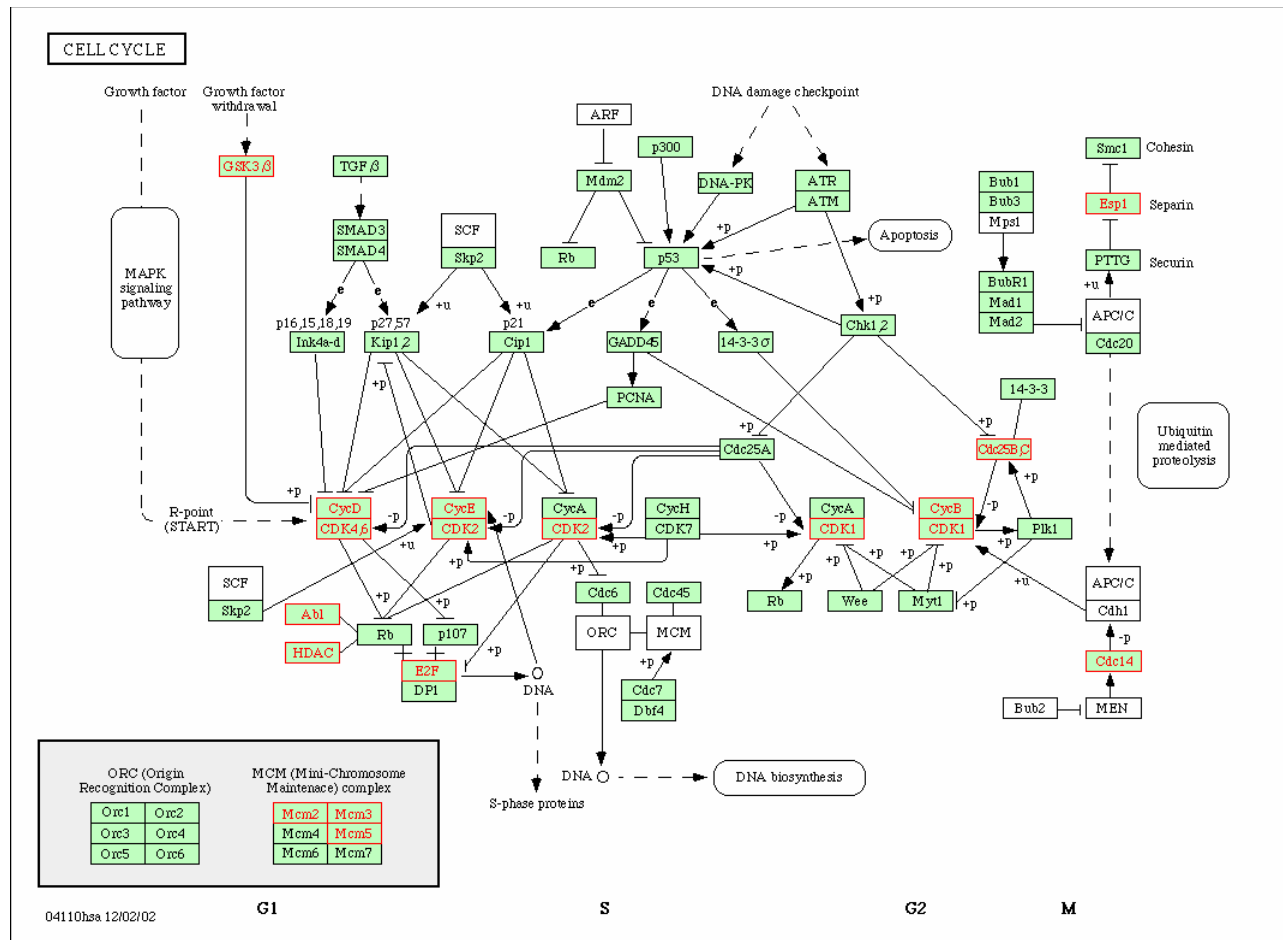


Figure 1: KEGG pathway diagram showing cell cycle genes up-regulated during progression to metastasis. The genes that were up-regulated in the metastatic cell line are highlighted in red.