

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Tumor proteomics

ArticleInfo		
ArticleID	:	3929
ArticleDOI	:	10.1186/gb-2001-2-8-reports0022
ArticleCitationID	:	reports0022
ArticleSequenceNumber	:	20
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2001-6-6 Received : 2001-6-6 OnlineDate : 2001-7-20
ArticleCopyright	:	BioMed Central Ltd2001
ArticleGrants	:	

Steven G Gray

## Abstract

A novel protein upregulated in certain cancers has been identified and characterized using an integrated proteomics, genomics, bioinformatics and standard molecular biology approach.

## Significance and context

From the wealth of sequence information generated by genomics techniques, bioinformatics data mining has revealed the presence of many putative genes of unknown function. The next logical development has been the production of methods for studying and identifying the proteins encoded by such genes, to discover how they fit in to the cellular environment. In a twist on this proteomics approach, Choong *et al.* have directly detected a novel protein rather than its gene. They first identified and analyzed a hitherto unknown protein that is upregulated in a hepatocellular carcinoma proteome. By sequencing small fragments of this protein they were then able to use bioinformatics to piece together the gene from sequences in the databases. Having identified the putative gene, they then characterized it extensively using standard molecular biology techniques.

## Key results

Choong *et al.* used a database of expressed proteins for a hepatocellular carcinoma-derived cell line (HCC-M), separated by two-dimensional gel electrophoresis, that they had developed previously. The separated proteins were fingerprinted by using peptide mass spectrometric techniques. From these data they could identify novel proteins that were not present in the SWISS-PROT or National Center for Biotechnology Information (NCBI) databases. One such protein was then taken for further analysis and was subjected to in-gel trypsin digestion and the products were sequenced using mass spectrometry. The sequences were then used to search an expressed sequence tag (EST) database and 40 DNA sequences obtained were assembled into a putative open reading frame. On the basis of this consensus DNA sequence, the authors named the protein HCC-1, and confirmed and extended the sequence using RACE (rapid amplification of cDNA ends).

A battery of bioinformatics tools was used to analyze the predicted protein, which was found to have no transmembrane segment (by PredictProtein), mitochondrial targeting signals (PSORT) or secretory signal (SignalP). A predominantly  $\alpha$ -helical structure (PHDsec) and three putative domains were

predicted. Of these, only the first domain, consisting of 42 amino acids, was found to have any homology to known proteins, with homology to a SAP domain (RPS-BLAST), a putative DNA-binding motif implicated in chromatin organization and transcriptional repression. A polyclonal antibody to HCC-1 was developed, and was used to show that the protein was localized predominantly in the nucleus. Using PSI-BLAST, the same domain was found to have similarities to heterogeneous nuclear ribonucleoproteins (hnRNPs). Analysis of the putative upstream promoter region revealed transcriptional elements associated with eukaryotic polymerase II promoters (ProScan). This region was confirmed as a *bona fide* promoter by cloning the fragment into a specialized vector and assaying for promoter activity. It was found that an enhancer sequence was needed for full transcriptional activity. The authors could find no introns by chromosome walking into the *HCC-1* genomic region and suggest that *HCC-1* is in fact a processed gene, or retrogene. Using radiation hybrid mapping, Choong *et al.* mapped the gene to chromosome 7q22.1, a region frequently altered in human tumors. Using a PCR screening method, they showed that *HCC-1* was expressed in various human tissues and that its expression was altered in liver and pancreatic tumors.

## Links

The tools used to analyze the protein data were [PredictProtein](#) , [SignalP](#) , [PSORT](#) , [RPS-BLAST](#) and [ProScan](#).

## Conclusions

Choong *et al.* suggest that *HCC-1* is a functional retrogene whose expression is altered in pancreatic adenocarcinomas and hepatocellular carcinoma. The protein product is a nuclear protein with potential DNA-binding capabilities and may have important cellular roles and/or may contribute to tumorigenesis.

## Reporter's comments

This paper illustrates the power of modern high-throughput techniques in the discovery of novel genes and proteins. The integrated approach using proteomics, genomics and bioinformatics allows researchers to identify new genes and proteins quickly, and to predict their structures and functions. Predictions can then be tested quickly and accurately using standard molecular techniques. One possible note of discord is the fact that the isolated gene appears to be a retrogene. Thus, the results of the PCR screening for gene expression could theoretically be argued to be from the original parental gene and not from the *HCC-1* retrogene. The authors did not examine this possibility. A method to test for the presence of other copies of the gene would be to screen a P1 phage artificial chromosome (PAC) library and isolate several positive clones which could be used for fluorescence *in situ* hybridization (FISH) experiments.

# Table of links

[FEBS Letters](#)

[PredictProtein](#)

[SignalP](#)

[PSORT](#)

[RPS-BLAST](#)

[ProScan](#)

## References

1. Choong ML, Tan LK, Lo SL, Ren E-C, Ou K, Ong S-E, Liang RCMY, Seow TK, Chung MCM: An integrated approach in the discovery and characterization of a novel nuclear protein over-expressed in liver and pancreatic tumors. *FEBS Lett.* 2001, 496: 109-116. 25030