Tumor marker-encoding genes: a bunch of mysterious diamonds in the pile of evolutionary compost

To be presented by Ancha BARANOVA

[1]*Molecular and Microbiology Department, College of Science, George Mason University, Fairfax, VA 22003 USA*

abaranov@gmu.edu

The search for human genes specifically expressed in different tumours is one of the major challenges for modern tumour biology. A number of experimental methods are designed for tumour-specific gene search. Most of them are based on time-consuming and expensive experimental protocols (numerous modifications of the diffrential display approach, cDNA microarrays, serial analysis of gene expression). Growing number of ESTs in publicly available databases provides a strong basis for the development of computer-based procedures for the subtraction of different EST pools instead of traditional experimental approaches used to compare expression profiles. A computer-based differential display (CDD) is an analogue of experimental differential displays that compares expression patterns in a particular tissue versus any other tissue source.

In our previous works we pioneered global approach to CDD by subtracting all available tumour cDNA libraries against all available normal libraries instead of widely described pairwise comparisons. We developed software HSAnalyst for EST sorting on the basis of the cDNA tissue source. For complete CDD analysis we used the LibraryRegistry database consisting of about 4000 library descriptions that we manually verified and curated. An algorithm executed by the program consists of two major steps: (1) for each cluster the number of ESTs is retrieved from the cluster description and (2) the number of ESTs from the `tumour' cDNA libraries is counted according to the LibraryRegistry database. The whole range of possible EST numbers is dissected into sub-ranges. HSAnalyst makes it possible to arrange sub-ranges exponentially (sub-ranges with exponents 1^2, 3^4, 5^8, 9^16, etc.) or linearly (sub-ranges with factors 1^10, 11^20, 21^30, etc.). Simultaneously, the ratio cancer ESTs/all ESTs is calculated for each cluster, and those which exceed the user-defined bottom threshold value are listed in the output file. To be sure that we found `true' tumour-specific clusters not generated by chance among the total number of the EST clusters (more than 90 000 units) we calculated the theoretical number of `tumour' clusters for every sub-range. The underlying model was the binomial distribution with the mean value of `cancer/all' ratio that could be defined by a user (0^100%). The number of clusters that exceeded threshold value was calculated. It enabled us to automatically select a set of genes highly expressed in different tumours but not in normal tissues. The products of such genes could serve as novel tumour markers and potential targets for anti-tumour therapy.

Most of the tumor-specific EST clusters reveled by global CDD were derived from the plurality of the tumor types originated in tissues of both ectodermal and mesodermal origin. Among a list of various genes highly represented in tumor-derived libraries and rarely observed in normal tissue-derived libraries was the Unigene cluster Hs.389457, which contains the complete mRNA sequence of human Brachyury, a

homologue of the mouse Brachyury gene. The product of the Brachyury gene is a member of the T-box family of transcription factors, characterized by a highly conserved DNA-binding domain designated as T-domain and is a key player in mesoderm specification during embryonic development. Reverse transcription-PCR analysis validated the in silico predictions of the expression of the human Brachyury gene. Brachyury mRNA was found in tumors of the small intestine, stomach, kidney, bladder, uterus, ovary, and testis, as well as in cell lines derived from lung, colon, and prostate carcinomas, but not in the vast majority of the normal tissues tested. We showed that an HLA-A0201 epitope of human Brachyury able to expand CTLs in vitro. The lung carcinoma cell line H441 was effectively lysed in the presence of Brachyury specific CTLs even at a low ratio of effector T cells to targets in an antigen-specific and MHC-restricted manner. Furthermore, we were also able to show that Brachyury–T-p2–specific CTLs can be expanded in vitro from PBMCs of a colorectal cancer patient and an ovarian carcinoma patient, satisfying a critical prerequisite for the use of Brachyury as a therapeutic target for cancer vaccine regimens.

Interestingly, many of tumor-specific ESTs clusters reveled by global CDD do not contain long open reading frames and cannot be classified as protein-encoding genes. These mRNA may or may not possess some regulatory or other meaningful functions, most likely remaining a mere consequence rather than the cause of the tumorigenesis. There are ways, though, to translate RNAs even without long open reading frames, for example, these RNAs might be edited, thus, allowing low-level translation. In this connection, it is remarkable that mRNA for the RNA-editing protein APOBECI was identified among top 20 tumor specific sequences revealed by CDD.

To experimentally assess patterns of expression of these EST clusters, we studied four of them in PCR experiments on Clontech MTC panels. The experimental data confirm the results obtained by in silico screening, i.e. tumor specificity of their expression. We suggest that a significant increase in the expression of non-coding RNA is a fundamental feature of cancer cells. It is well known that tumors are characterized by an imbalance in the expression of the multiple transcribed units, most of which are consequences but not the causes of the tumor development. Deregulated environment with tumor cell may provide conditions for the expression of evolutionarily new and/or sleeping genes. Thus, tumors could be considered as an evolutionary proving ground for an expression of mutated or newly evolved transcripts. If proven to be correct, this concept may substantially increase our capabilities in the diagnosis and treatment of cancer.

To experimentally examine our prediction that at least some tumor-related sequences are evolutionarily new, we performed Southern hybridization of [α-32P]-labeled newly described tumor-related fragments with genomic DNA from different animal species. Hybridization signals were detected only with human and orangutan DNA. Comparative genomics analysis have shown that the tumor-related transcripts under consideration have orthologs in mammal genomes only and not in those of fishes, amphibia, and birds, with the single exception of a short sequence in the chicken genome with low homology for AI952931.

According to our observations, the majority of the previously described tumor markers also possess relatively low conservativeness. A typical example of the known tumor marker and evolutionary new sequence tumor-associated testis-specific antigen

PRAME that was identified as a melanoma antigen recognized by cytotoxic T cells (CTLs) and expressed in a variety of cancer cells including leukaemic cells (Matshushita M. et al., 2003). Human PRAME is BCR/ABL-inducible gene (Watari K. et al., 2000) that is located on 22q11.22, consists of 6 exons and encodes 509 aa protein with unclear function. Human genome also contains 21 low-copy repeats that roughly correspond to exons 4 and 5 of the human PRAME that are not included in the any gene models supported by UNIGENE expression data. Alignment of the human PRAME nucleotide sequence with mouse genome did not result in any matches. On the other hand, mouse tissues express multiple transcripts that are characterized by very low abundance (1 to 3 transcripts per corresponding murine UNIGENE cluster) and encode proteins with significant level of homology to human PRAME ranging form 32.78% for Pramel3 to 40.86% for Pramel1. Mouse Pramel1 is located on chromosome 4E1 and consists of 4 exons. Mouse Pramel3 consists of 5 exons and locates on chromosome X, where is present in 5 tightly clustered copies. Other murine PRAME-like sequences are also located on chromosome X often being clustered together with partial paralogous sequences resembling internal exons of PRAME-like genes. For example, murine Pramel3 represents a group of low-copy repeats retaining 70-80% of identical nucleotide positions. Both Pramel1 and Pramel3 are weakly homologous to human PRAME at the protein level, but not to each other. Moreover, nucleotide alignment of Pramel1 and Pramel3 indicates divergence of these sequences to the level below detection with standard cut-off parameters of BLAST2 (word size 11, open gap and extension gap penalties are 5 and 2, correspondingly). On the other hand, pairwise BLAST2 comparisons of human PRAME with murine PRAME-like sequences revealed 80-88% of nucleotide homology between PRAME and Pramel1 that correspond to four consequent but not contiguous DNA regions. Taken together, these finding indicates that both in the mouse and in the human genomes PRAME-like sequences independently undergo partial duplications accompanied by rapid, but mostly synonymous changes in its nucleotide sequences. While staying almost silent in the mouse genome, in the setting of the human cell PRAME sequence become an actively transcribed gene, preferentially expressed in tumors.

Tumor marker encoding genes tend to be located close to points of the chromosomal rearrangements at the borders of the syntheny between two mammalian genomes. Often these rearrangements lead to the perturbations in the structure of the tumor marker encoding gene and possible change of the mechanism of its regulation. We hypothesize that the fraction of the genes that discordantly regulated in two compared genomes is proportional to the number of the chromosomal rearrangements that accompanied evolution of these genomes. As the rate of rearrangement' accumulation in course of the evolution of the murid rodents was much higher than in hominides and other mammalian lineages, tumor markers have to be studied in alternative, non-rodent animal models. Comparative genomic analysis of the known tumor markers PRAME, SEACAM6, SKALP, SCGB2A2, DPCR1 and MAGE family confirmed our prediction.

Taken together, data obtained by PCR amplification, Southern hybridization and computational analysis support the prediction that newly evolved sequences may be expressed in tumor cells. The fundamental concept of the tumors as the reservoir for the background expression of the novel transcripts, if proven to be correct, may dramatically increase our capabilities in diagnostics and treatment of cancer.

References:

1. Kozlov AP: Evolution of Living Organisms as a Multilevel Process. J Theor Biol 1979, 81:1.
2. Kozlov AP: Gene Competition and the Possible Evolutionary Role of Tumours. Medical Hypotheses 1996, 46:81.
3. Baranova AV, Lobashev AV, Ivanov DV, Krukovskaya LL, Yankovsky NK, Kozlov A: In silico screening for tumour-specific expressed sequences in human genome. FEBS Lett 2001, 508:143.
4. Krukovskaja LL, Baranova AV, Tyezelova T, Polev D, Kozlov AP: Experimental study of human expressed sequences newly identified in silico as tumour specific. Tumour Biol 2005, 26:17.
5. Palena C, Polev DE, Tsang KY, Fernando RI, Litzinger M, Krukovskaya LL, Baranova AV, Kozlov AP, Schlom J. The human T-box mesodermal transcription factor Brachyury is a candidate target for T-cell-mediated cancer immunotherapy. Clin Cancer Res. 2007 Apr 15;13(8):2471-8.
6. Kozlov AP, Galachyants YP, Dukhovlinov IV, Samusik NA, Baranova AV, Polev DE, Krukovskaya LL. Evolutionarily new sequences expressed in tumors. Infect Agent Cancer. 2006 Dec 25;1:8.