S1-3 *[2000 Oct. 19 (Thur), 11:00~11:35/Hall A]*

Data mining: Finding organ specific genes using the EST database

Byungkook Lee

Molecular Modeling Section, Laboratory of Molecular Biology, Division of Basic Sciences, National Cancer Institute, National Institutes of Health

We developed a procedure for discovering genes that are expressed only in a given organ or tissue. Products of such genes can be used for cancer diagnosis and imaging. They can be useful also for developing new anti-tumor vaccines and new anti-cancer drug molecules that kill only a particular organ or cancer cells. Our procedure takes advantage of the extensive Expressed Sequence Tag (EST) database, which in turn is partly the result of the CGAP (Cancer Genome Anatomy Project). There are two stages to this procedure. In the first stage, the EST database is analyzed to produce a list of candidate genes. This analysis is done using BLAST-based sequence comparisons and a clustering algorithm that is designed to collect all ESTs derived from the same gene into one cluster. We have developed a computer program to do this, but we also use the Unigene clusters provided by the NCBI. In the second stage, promising candidates are selected from this list and examined experimentally to verify their tissue specificity and to obtain the full sequence of the expressed gene. When the EST sequence can be located in the human genome sequence, the process becomes faster.

We found and characterized some five genes to date by this method, which seem to produce proteins that are unique to an organ or that are nearly so and which have not been described in the literature. Work on many other promising genes is in progress.

One of the first new genes that we found by this new procedure is called PAGE4. It is expressed in normal and malignant prostate, testicular, and uterine tissues. It codes for a small protein of 102 amino acids and is related to a family of tumor-associated antigens called GAGE. XAGE1 is another new gene found, which shares some sequence homology with the PAGE4 gene. XAGE1 is expressed in testis and in sarcomas of various types. This gene product is a potential target for vaccine development against osteosarcoma. Another gene product, which we named TARP, corresponds to an alternate reading frame of

the T-cell receptor g-locus (TCRg). This form of the gene product is not expressed in any lymphocytes, but produced only in prostate and breast cancer cells. ERGL is another new gene, which is related to a known gene called ERGIC-53. ERGL is expressed abundantly in prostate, but is also expressed in other vital tissues. The protein that this gene produces is likely to be membrane-bound and could be useful for imaging. Another gene, PRAC, encodes a rather small protein of 58 amino acid residues. It appears to be a nuclear protein and is expressed in prostate, rectum, and descending colon. Nothing is known about the function of any of the proteins found so far.