

## Commentary

# Finding the bEST Routes to Cancer

**Shao-Chun Wang**

**Mien-Chie Hung\***

Department of Molecular and Cellular Oncology; The University of Texas MD Anderson Cancer Center; Houston, Texas USA

\*Correspondence to: Mien-Chie Hung; The University of Texas MD Anderson Cancer Center; Department of Molecular and Cellular Oncology, Unit 108; 1515 Holcombe Blvd.; Houston, Texas 77030 USA; Tel.: 713.792.3668; Fax: 713.794.0209; Email: mhung@mdanderson.org

Received 11/15/04; Accepted 11/15/04

Previously published online as a *Cancer Biology & Therapy* E-publication: <http://www.landesbioscience.com/journals/cbt/abstract.php?id=1403>

### KEY WORDS

EST, Affymetrix, microarray, gene expression, pancreatic cancer

### Commentary to

#### *Identification of Novel Highly Expressed Genes in Pancreatic Ductal Adenocarcinomas through a Bioinformatics Analysis of Expressed Sequence Tags*

Dengfeng Cao, Steven R. Hustinx, Guoping Sui, P. Bala, Norihiro Sato, Sean Martin, Anirban Maitra, Kathleen M. Murphy, John L. Cameron, Charles J. Yeo, Scott E. Kern, Michael Goggins, Akhilesh Pandey and Ralph H. Hruban

The cDNA microarray gene expression analysis is a powerful technology due to its ability to examine the expression of thousands of genes simultaneously. On a single slide or chip, tens of thousands of spots are printed and their expression under different conditions or in tissues with different disease stages can be assessed at the same time, given that an appropriate analysis platform of computational algorithm is also available. Since its debut in 1990's, microarray technology has evolved tremendously and generated a revolutionary impact on the methodology of biomedical research. In the cancer research field, the pioneering study performed by Golub, Lander and the colleagues in 1999 demonstrates that the two forms of leukemia, acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL), can be distinguished based on their gene expression patterns. One year later, another study demonstrated that estrogen receptor-negative and estrogen receptor-positive breast cancers can be correctly classified using microarray-based expression profiling.<sup>2</sup> This work was followed by the first report linking gene expression profiling analysis with survival of untreated breast cancer patients.<sup>3</sup> Since then, the application of microarray in cancer diagnosis, classification, therapy, and drug discovery has been rapidly expanding. Although beyond the scope of this commentary, these areas have been extensively reviewed by others.<sup>4-6</sup>

The source and quality of the genetic elements, known as "probes", spotted on a typical nucleotide DNA microarray have a great influence on the outcome of the microarray analysis. The sequences of these nucleotide probes are derived either from the cDNA sequences of known genes or, to a good proportion of it, from random cDNA clones through large-scale sequencing. The random sequences are known as "expressed sequence tags" (ESTs), a term dubbed by Venter and his colleagues in their landmark study published in 1991.<sup>7</sup> These short single-pass sequences derived from one or both ends of random cDNA clones have been proven to be a valuable resource for the discovery of new genes.<sup>8,9</sup> Take the Affymetrix U133 GeneChip/® set as an example, it has been estimated that about 15% of the probe sets are derived from ESTs.<sup>10</sup> These expressed sequences represent the exons, or more precisely the genomic sequences likely to be the exons, of genes scattered through the immense human genome. However, due to the relative inaccuracy of the EST sequence and the yet incompleting human genome sequence project, many of the ESTs had been difficult to be identified and assigned to the corresponding source genes. Nowadays, it is still a very time-consuming task to sieve through the roughly annotated ESTs in the database, validate these ESTs, and draw useful information from them. Therefore, researchers tend to ignore the EST clones standing out from their microarray analysis results. Whether the many EST clones presented on a microarray chip contain important information regarding the biology of the disease of interest has seldom been addressed.

In this post-genome era when the human genome project is completed and continues to be refined, the EST-gene assignment problem for microarrays is ready to be tackled using the bioinformatic resources available. In the article reported by Cao et al. appears in this issue, it is demonstrated that the EST clones revealed by a microarray gene expression analysis can be further processed through bioinformatic analyses to identify potentially important genes involved in cancer progression. The authors analyzed the gene expression profiles in 64 pancreatic cancer tissues and 60 non-neoplastic tissue samples using the Affymetrix U133 array. Sixty EST clones are found overexpressed in pancreatic adenocarcinomas when compared with normal tissues. Instead of ignoring these ESTs, the authors further explored the identities of these clones by searching the human genome database. Of the 60 ESTs, 43 of them can be unambiguously assigned to a corresponding gene, as the EST sequence is located within the sequence of a known gene. For the rest clones, the authors queried the ever-expanding EST database (dbEST) maintained by National Center for Biotechnology Information (NCBI) and looked for other EST clones in the database which overlapped with the query EST sequence and would link the query EST to a nearby known gene in the genome. Using this strategy, 15 of the remaining ESTs were

mapped in clusters very close to a known gene, which are then assumed to be the source genes of these EST clones. It is noteworthy that two of them, the integrin  $\beta$ -like 1/CD29 gene and the protein tyrosine phosphatase PTPN14 gene, are also identified in the group of the 43 EST clones which are unambiguously mapped to known genes. The potential of this bioinformatic strategy is further supported by the observation that 3 of the EST clones identified in this study, the integrin  $\beta$ -like 1/CD29 gene, the retinoid acid induced 3 (RAI3) gene, and the TGF  $\beta$  ligand inhibin  $\beta$  A (INHBA) gene, have also been identified in the previous study of the authors. Among them, CD29 is found to be overexpressed in 16 of 18 (89%) of pancreatic cancer tissue as tested by immunohistochemistry. Although the sample group is relatively small, the high overexpression rate as well as the recurrence of this gene in independent analyses make it a promising tumor marker and should warrant a more extensive study for CD29 expression in more tumor samples and different tumor types. The authors also report a dramatic increase of the expression of a scaffold protein, A-kinase anchoring protein 12 (AKAP12), in the tumor tissues and in a variety of cancer cell lines. Since it is known that AKAP12 acts as a metastasis inhibitor of prostate cancer,<sup>11</sup> the observation of the increased expression of AKAP12 in pancreatic cancer is intriguing and further investigation would be required. Interestingly, of the 60 EST clones identified, at least ten of them are proteins known to be involved in cell attachment or migration, such as molecules of cell adhesion, extracellular matrix, metalloprotease, and integrins. This may reflect the pathological stages of the tumor samples included in this paper. Of the 50 tumor tissues, 42 of them are from infiltrating pancreatic adenocarcinoma and 8 of them are metastatic pancreatic adenocarcinoma. The finding of the many genes involved in cell migration and adhesion may explain the invasive nature of these tumors, and indirectly supports the accuracy of profiling analysis using ESTs.

Although the biology linking pancreatic cancer to the identified genes needs to be further explored, the study done by Cao et al. demonstrates that the often overlooked ESTs resulted from microarray gene expression analysis may actually be a valuable gold mine to uncover important genes involved in cancer development. This study also highlights the need for a more informative annotation system for the EST probes used in microarrays. Such effort can be initiated by both the private and the public sectors. In addition, with the continuous updating of the human genome database and the UniGene cluster,<sup>12</sup> it is foreseeable that the source gene identities of more and more EST probes on the microarray will be readily determined and well-annotated, facilitating the exploration for disease-associated markers or novel therapeutic targets. In the many routes of research leading to the molecular mechanisms and therapy of cancer, ESTs could become important landmarks along the path, just like the critical contribution they have made toward gene discovery and genome mapping.

#### References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh KL, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286:531-7.
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Bostein D. Molecular portraits of human breast tumours. *Nature* 2000; 406:747-52.
3. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GL, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; 415:530-6.
4. Lakhani, SR and Ashworth A. Microarray and histopathological analysis of tumours: the future and the past? *Nature Rev Cancer* 2001; 1:151-7.
5. Liu ET. Classification of cancers by expression profiling. *Curr Opin in Gene Dev* 2003; 13:97-103.
6. Luo J, Issac WB, Trent JM, and Duggan DJ. Looking beyond morphology: cancer gene expression profiling using DNA microarrays. *Cancer Invest* 2003; 21:937-49.
7. Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF, Kerlavage AR, McCombie WR, Venter JC. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252:1651-6.
8. Boguski MS, Tolstoshev CM, Bassett DE Jr. Gene discovery in dbEST. *Science* 1994; 265:1993-4.
9. Boguski MS. The turning point in genome research. *Trend Biol Sci* 1995; 20:295-6.
10. Lacobuzio-Donahue CA, Ashfaq R, Maitra A, Adsay NV, Shen-Ong GL, Berg K, Hollingsworth MA, Cameron JL, Yeo CJ, Kern SE, Goggins M, Hruban RH. Highly expressed genes in pancreatic ductal adenocarcinomas: A comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. *Cancer Res* 2003; 63:8614-22.
11. Xia W, Unger P, Miller L, Nelson J, Gelman IH. The src-suppressed C kinase substrate, SSeCKS, is a potential metastasis inhibitor in prostate cancer. *Cancer Res.* 2001; 61:5644-51.
12. Pontius JU, Wagner L, Schuler GD. UniGene: a unified view of the transcriptome. In: *The NCBI Handbook*. Bethesda (MD): National Center for Biotechnology Information; 2003.