Commentary

Gene Expression Studies on Soft Tissue Tumors

Matt van de Rijn* and Brian P. Rubin[†]

From the Department of Pathology, Stanford University Medical Center, Palo Alto, California; and the Department of Anatomic Pathology,[†] University of Washington Medical Center, Seattle, Washington*

The sequencing of the human genome along with technological achievements that allow large numbers of defined DNA sequences to be attached to various substrates in neat rows and columns, has led to the development of gene microarray analysis.¹ This recently developed technique facilitates the genome-wide screening of expression levels of tens of thousands of genes in a single experiment. The advent of this technology is especially fortuitous, as the human genome project has identified at least 30,000 genes in the human genome.² For a significant number of these genes the function is currently unknown. Many genes that play a role in pathogenesis, that could be used as targets for drug therapy, or that could aid in the diagnosis of disease, are yet to be discovered. Clearly, screening by conventional methods (such as Northern blotting) of large numbers of genes for clinically relevant markers or therapeutic targets would be an enormous task. Gene microarray analysis is being applied increasingly to a variety of human tumors, including soft tissue tumors. In the current issue of *The American Journal of Pathology*, Allander and colleagues³ report the results of gene microarray analysis of synovial sarcoma. This excellent study highlights the advantages of genome-wide screening of gene expression in synovial sarcoma, a poorly understood neoplasm.

Synovial sarcoma has a predilection for the extremities of adolescents and young adults. Despite its name there is no support for a relationship to normal synovium, and the normal cellular counterpart for this tumor is unknown. Histologically, it is a distinctive neoplasm with either a biphasic (mixed epithelioid and spindle cell morphology) or monophasic spindle-cell histology. The vast majority of synovial sarcomas have a balanced translocation t(X;18)(p11.2;q11.2) that is specific for this tumor.⁴ This translocation fuses two genes, *SYT* and *SSX* on chromosome 18 and chromosome X, respectively, to form a novel hybrid protein encoded by the 3' end of the *SYT* gene and the 5' end of the *SSX* gene.^{5,6} The t(X;18) can be detected by reverse transcriptase-polymerase chain reaction in both paraffin-embedded and frozen tissue.⁷ The translocation product is thought to have transcrip-

tional activity, but its targets are unknown. The diagnosis of biphasic synovial sarcoma generally does not present problems. However, the distinction of monophasic synovial sarcoma from other soft tissue tumors such as leiomyosarcoma, malignant peripheral nerve sheath tumor, and other spindle cell neoplasms can be difficult. Most synovial sarcomas are immunoreactive for keratins or epithelial membrane antigen (EMA) by immunohistochemistry but a significant proportion of these tumors are negative or only focally positive for these markers and are challenging to diagnose. To date there has been little progress in identifying prognostic markers in synovial sarcoma. The primary treatment for synovial sarcoma is surgical, and no specific drug therapies exist. It is clear that further study of this neoplasm is warranted and that an extensive search for diagnostic and prognostic markers and for therapeutic targets would be valuable. Gene microarray analysis allows for a characterization of gene expression levels for tens of thousands of genes in a single experiment. Thus, screening of large parts of the genome, or the entire genome, depending on the size and composition of the arrays, can be performed.

Using gene microarrays containing 6548 genes, Allander and colleagues³ examined the expression profiles of 14 synovial sarcomas, 4 malignant fibrous histiocytomas, and a fibrosarcoma. The gene arrays could differentiate the group of synovial sarcomas from the other tumors, based on the expression patterns of 153 genes that were selected for their ability to discriminate between these tumor groups. From this group of genes, 50 genes were identified that were expressed in all synovial sarcomas and these 50 genes were used to hierarchically cluster the 19 tumors examined. Hierarchical clustering is a way of representing complex data with tumor samples arranged in columns and gene expression levels arranged in rows. The clustering process puts together in one dimension the tumors that react most similar to each other across the genes tested. Likewise, in the second dimension genes are arranged in such a way that genes that have similar expression patterns in the tumors examined are put above one another. A dendrogram then depicts the degree of relatedness between tumor sam-

Accepted for publication August 26, 2002.

Address reprint requests to Matt van de Rijn, Stanford University Medical Center, Department of Pathology, L225, 300 Pasteur Dr., Palo Alto, CA 94305. E-mail: mrijn@stanford.edu.

ples (or genes). The group of synovial sarcomas clustered tightly together on a branch distinct from the other tumors in this study, indicating that the genes that were selected differ significantly in their expression patterns between the two groups. Furthermore, the authors identified a set of 21 genes that showed significantly different levels of expression between biphasic and monophasic synovial sarcomas. The fact that this gene set contained a number of keratin genes (highly expressed in the epithelial component of biphasic sarcomas) is as expected and provides additional support for the validity of their studies.

Although the fact that the arrays could classify these lesions correctly is impressive, it is not that surprising that a technique that examines thousands of genes at a time can distinguish tumors that are morphologically as distinct as synovial sarcoma and malignant fibrous histiocytoma (MFH). However, it is good to remember that it is not the purpose of gene arrays to classify lesions that can also be distinguished using conventional techniques such as light microscopy (as is sometimes proposed). The fact that arrays can do this is merely a verification that gene arrays work; they reflect what is seen under the microscope. The idea that gene array analysis will supplant histological examination is a simplification of the issue and in our opinion, wrong. Histopathologists have already done an amazing job of classifying lesions and will continue to do so with very high accuracy. In fact, it is comforting to see that every time a new technology is born (immunohistochemistry, reverse transcriptase-polymerase chain reaction, and so forth), it generally validates the pre-existing histological classification system. Rather, the real power of microarray analysis is to permit a genome-wide search for markers that may distinguish subtypes of tumors that remain inapparent by conventional techniques. These subtypes of tumors can then be tested for their sensitivity to specific drug therapies that when tested on the entire tumor group might go unnoticed. Clearly, many clinical studies will be necessary to determine the relevance of further subclassification of soft tissue tumors by gene arrays. In addition to subclassification, other aspects of these lesions can be identified. These include: prognosis and disease surveillance markers and molecular pathways that can be targeted therapeutically and aid in understanding the pathogenesis of these tumors. Allander and colleagues³ illustrate this very nicely in their study.

Targeting important oncogene-activated pathways can be useful in the treatment of different neoplasms. One notable recent example of such a targeted therapy is imatinib mesylate (also known as Gleevec, STI571), a specific inhibitor of the KIT receptor tyrosine kinase, which is constitutively activated in gastrointestinal stromal tumors.⁸ Several large-scale therapeutic trials are currently under way using imatinib to treat gastrointestinal stromal tumors and the preliminary results are encouraging.^{9,10} Although the association of the *KIT* gene with gastrointestinal stromal tumor was originally recognized by a hypothesis driven approach, two recent microarray gene expression-profiling studies were also able to tightly associate the *KIT* gene with gastrointestinal stromal tu-

mors, suggesting that this technique may also be able to identify therapeutic targets in diseases for which no equivalent genetic target is yet known.^{11,12} In the study by Allander and colleagues,³ the gene arrays identified high levels of expression of *ERBB2* in biphasic synovial sarcomas and a subset of monophasic synovial sarcomas. Obviously, the expression of *ERBB2* in this tumor subset suggests that this protein may be a target for specific therapies such as Herceptin.¹³ However, gene array analysis does not yield definitive information about protein expression levels, so the results of gene array analysis require validation by alternate methods. The authors of this study validated the expression of *ERBB2* by immunohistochemical analysis of tissue microarrays, another recently developed technique whereby numerous tissue samples are arrayed in the form of tissue cores in a paraffin block so that the resultant tissue sections contain as many as 500 samples.¹⁴ Furthermore they used fluorescence *in situ* hybridization to explore whether the expression of *ERBB2* was related to amplification of the gene (it was not). In addition, the arrays showed high levels of expression of *IGF1* and *IGFBP2*, which were validated by immunohistochemical studies. As the authors pointed out, these molecules are currently under study as potential clinical markers in the serum of patients with tumors that express these proteins. A serological marker that could be used to monitor the recurrence of synovial sarcomas would clearly be of immense clinical value.

Several other considerations regarding this study are important. As the authors note, the downstream transcriptional targets of the SYT-SSX fusion protein are currently unknown and it is quite possible that candidates for the targets of this transcription factor are relatively overexpressed or underexpressed as indicated in the hierarchical clustering.

To date a limited set of expression profiling studies have been published on soft tissue tumors but it is to be hoped that many others will follow. It is instructive to compare the current study by Allander and colleagues³ with a recently published gene array analysis of sarcomas including synovial sarcomas by one of the authors of this Commentary.¹² Although some of the findings correlate very well, there are also many genes that do not overlap between the two studies. Allander and colleagues³ discuss some possibilities for these discrepancies, including the fact that our study examined only monophasic synovial sarcomas. Additional explanations for these discrepancies include differences in the number of tumors and types of tumors that were compared with the expression profiles of synovial sarcoma. This is especially important because relative and not absolute gene expression levels are analyzed in these gene array studies. Moreover, the number of genes examined varied significantly. The degree of overlap between the gene sets used in the two studies is currently unknown but Allander and colleagues³ will make the data available on publication and thus it will be possible to compare the data from the two studies. One example of the differences in genes available for analysis in the two studies is the *SSX* gene, which was not included in the set of genes

available for study in the report by Allander and colleagues,³ but was included in the report by Nielsen and colleagues.¹² Another important difference is in the choice of reference RNA that was used in the two different studies. Allander and colleagues³ used RNA derived from a single osteosarcoma cell line. In contrast, a mixture of 11 human tumor cell lines was used as a source of reference RNA in the other study. This could significantly affect the results of these analyses by skewing the representation of well-measured genes. Finally, the methods of statistical analysis varied significantly. Allander and colleagues³ used a supervised technique to select 153 genes that distinguished synovial sarcoma from the other tumors. They subsequently used a subset of 50 genes from this group to perform hierarchical clustering, because they were most widely expressed in synovial sarcoma. In contrast, Nielsen and colleagues used hierarchical clustering of ~5000 genes that were selected in an unsupervised manner, ie, selected only on the adequate quality of their measurements.¹² The different approaches in statistical analysis of these large data sets each have their own pros and cons but clearly will result in different findings. The fact that some genes were identified in one but not the other study and vice versa therefore is not too surprising. This further underscores the point that gene microarrays should not be taken as a final answer but rather as a high-throughput screening measure through which genes of interest can be identified from the overwhelming number of genes present in the human genome. This is also the reason that findings from gene microarray analysis need to be validated by other techniques, as was done by Allander and colleagues.³ Nevertheless, even with major differences in gene sets, reference RNAs, and statistical analysis, it is very encouraging to see that several important genes such as *CRABP* and *EGFR* were found to be specifically expressed in synovial sarcoma in both studies. The expression of *CRABP* (cellular retinoic acid-binding protein) in synovial sarcoma suggests a role for the retinoic acid pathway in synovial sarcoma. In the study by Nielsen and colleagues, three additional genes within the retinoic acid pathway were identified in synovial sarcoma, further supporting a role for this pathway in the pathogenesis of this tumor.¹² The presence of *EGFR* (epidermal growth factor receptor) in both studies suggests that this receptor could be a target for specific inhibitors, as noted previously. Additionally, both studies suggest that neural and neurodevelopmental gene expression is a feature of synovial sarcoma.

The study of soft tissue tumors seems to be entering a new phase with high-throughput genome-wide expression profiling. These types of studies are a significant departure from more conventional hypothesis-driven research that is based on the analysis of one or a few genes/proteins of interest. In contrast, high-throughput studies such as these generate large amounts of data of which only a subset can be described in a journal article. In addition, any laboratory that generates these enormous datasets will only be able to follow up on a small portion of the findings. For these reasons it is very important that gene array studies share the complete dataset

on which an article is based via the world-wide web, allowing other investigators to mine the data for genes that hold particular interest for them. As data accumulates from multiple laboratories studying the same tumor type, software can be developed that will allow for a direct comparison and synthesis of the various datasets. The issue of data management and analysis is not restricted to gene microarrays. Tissue microarrays also generate datasets that are several orders of magnitude larger than those obtained through most conventional immunohistochemistry studies. We recently developed a method that allows for the analysis of large datasets obtained through immunohistochemistry on tissue microarrays.¹⁵ In addition, other high-throughput techniques such as array-based comparative genomic hybridization and protein arrays will further contribute to the already overwhelming amount of data that is being generated. Array-based comparative genomic hybridization allows for the identification of gene amplifications and deletions using the same gene arrays that are used for gene expression profiling.¹⁶ Array-based comparative genomic hybridization has the advantage of providing a more detailed analysis than conventional chromosome-based comparative genomic hybridization. Preliminary data from our laboratory suggests that these studies can also be performed with DNA isolated from formalin-fixed, paraffin-embedded material. Opening up archival tissue collections to this type of analysis would markedly increase the number of soft tissue tumor cases that can be examined and would allow for the retrospective study of soft tissue tumors with known clinical follow-up. Finally a variety of protein arrays are being developed.¹⁷ Some of these arrays contain large numbers of spotted immunoglobulins whereas other contains large numbers of spotted purified proteins. With the former type of arrays serum samples could be screened for the presence of synovial sarcoma-specific proteins in patients with this tumor.

It is clear therefore that genome-wide expression studies and other high-throughput analyses of soft tissue tumors will lead to the identification of diagnostic markers, prognostic markers, and the identification of specific cellular pathways that can be targeted therapeutically. The promise therefore of gene arrays is not in the replacement of the microscope to classify different disease processes, but in the power of the technique to screen the genome of neoplasms to promote a better understanding of the biology of these complex neoplasms. This will enable advancements in diagnosis, disease surveillance and prognostication, and treatment.

References

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999, 286: 531-537
2. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860-921
3. Allander SV, Illei PB, Chen Y, Antonescu CR, Bittner M, Ladanyi M, Meltzer PS: Expression profiling of synovial sarcoma by cDNA micro-

- arrays: association of ERBB2, IGFBP2, and ELF3 with epithelial differentiation. *Am J Pathol* 2002, 161:1587–1595
4. Turc-Carel C, Dal Cin P, Limon J, Li F, Sandberg AA: Translocation X;18 in synovial sarcoma. *Cancer Genet Cytogenet* 1986, 23:93
5. Clark J, Rocques PJ, Crew AJ, Gill S, Shipley J, Chan AM, Gusterson BA, Cooper CS: Identification of novel genes, SYT and SSX, involved in the t(X;18)(p112;q112) translocation found in human synovial sarcoma. *Nat Genet* 1994, 7:502–508
6. Crew AJ, Clark J, Fisher C, Gill S, Grimer R, Chand A, Shipley J, Gusterson BA, Cooper CS: Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homology to the Kruppel-associated box in human synovial sarcoma. *EMBO J* 1995, 14:2333–2340
7. Argani P, Zakowski MF, Klimstra DS, Rosai J, Ladanyi M: Detection of the SYT-SSX chimeric RNA of synovial sarcoma in paraffin-embedded tissue and its application in problematic cases. *Mod Pathol* 1998, 11:65–71
8. Hirota S, Nishida T, Isozaki K, Taniguchi M, Nakamura J, Okazaki T, Kitamura Y: Gain-of-function mutation at the extracellular domain of KIT in gastrointestinal stromal tumours. *J Pathol* 2001, 193:505–510
9. Demetri GD, von Mehren M, Blanke CD, Van den Abbeele AD, Eisenberg B, Roberts PJ, Heinrich MC, Tuveson DA, Singer S, Janicek M, Fletcher JA, Silverman SG, Silberman SL, Capdeville R, Kiese B, Peng B, Dimitrijevic S, Druker BJ, Corless C, Fletcher CD, Joensuu H: Efficacy and safety of imatinib mesylate in advanced gastrointestinal stromal tumors. *N Engl J Med* 2002, 347:472–480
10. van Oosterom AT, Judson I, Verweij J, Stroobants S, Donato DP, Dimitrijevic S, Martens M, Webb A, Sciort R, Van Glabbeke M, Silberman S, Nielsen OS: Safety and efficacy of imatinib (STI571) in metastatic gastrointestinal stromal tumours: a phase I study. *Lancet* 2001, 358:1421–1423
11. Allander SV, Nupponen NN, Ringner M, Hostetter G, Maher GW, Goldberger N, Chen Y, Carpten J, Elkahoul AG, Meltzer PS: Gastrointestinal stromal tumors with KIT mutations exhibit a remarkably homogeneous gene expression profile. *Cancer Res* 2001, 61:8624–8628
12. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Zhu S, Fero M, Sherlock G, Pollack JR, Brown PO, Botstein D, van de Rijn M: Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 2002, 359:1301–1307
13. Baselga J, Norton L, Albanell J, Kim YM, Mendelsohn J: Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts. *Cancer Res* 1998, 58:2825–2831
14. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998, 4:844–847
15. Liu CL, Prapong W, Natkunam Y, Alizadeh A, Montgomery K, Gilks CB, van de Rijn M: Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. *Am J Pathol* 2002, 161:1557–1565
16. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999, 23:41–46
17. Kodadek T: Development of protein-detecting microarrays and related devices. *Trends Biochem Sci* 2002, 27:295–300