Mechanisms of disease

Molecular characterisation of soft tissue tumours: a gene expression study

Torsten O Nielsen, Rob B West, Sabine C Linn, Orly Alter, Margaret A Knowling, John X O'Connell, Shirley Zhu, Mike Fero, Gavin Sherlock, Jonathan R Pollack, Patrick O Brown, David Botstein, Matt van de Rijn

Summary

Background Soft-tissue tumours are derived from mesenchymal cells such as fibroblasts, muscle cells, or adipocytes, but for many such tumours the histogenesis is controversial. We aimed to start molecular characterisation of these rare neoplasms and to do a genome-wide search for new diagnostic markers.

Methods We analysed gene-expression patterns of 41 soft-tissue tumours with spotted cDNA microarrays. After removal of errors introduced by use of different microarray batches, the expression patterns of 5520 genes that were well defined were used to separate tumours into discrete groups by hierarchical clustering and singular value decomposition.

Findings Synovial sarcomas, gastrointestinal stromal tumours, neural tumours, and a subset of the leiomyosarcomas, showed strikingly distinct gene-expression patterns. Other tumour categories—malignant fibrous histiocytoma, liposarcoma, and the remaining leiomyosarcomas—shared molecular profiles that were not predicted by histological features or immunohistochemistry. Strong expression of known genes, such as *KIT* in gastrointestinal stromal tumours, was noted within gene sets that distinguished the different sarcomas. However, many uncharacterised genes also contributed to the distinction between tumour types.

Interpretation These results suggest a new method for classification of soft-tissue tumours, which could improve on the method based on histological findings. Large numbers of uncharacterised genes contributed to distinctions between the tumours, and some of these could be useful markers for diagnosis, have prognostic significance, or prove possible targets for treatment.

Lancet 2002; **359:** 1301–07
See Commentary page 1263

Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada (T O Nielsen MD, J X O'Connell FRCPC)**; British Columbia Cancer Agency, Vancouver, BC** (M A Knowling FRCPC, J X O'Connell)**; and Department of Pathology** (R B West MD, S C Linn MD, S Zhu MD, J R Pollack MD, M van de Rijn MD)**, Department of Genetics** (O Alter PhD, M Fero PhD, G Sherlock PhD, Prof D Botstein PhD)**, Howard Hughes Medical Institute** (Prof P O Brown MD)**, and Department of Biochemistry** (Prof P O Brown)**, Stanford University School of Medicine, Stanford, CA, USA**

Correspondence to: Dr Matt van de Rijn, Department of Pathology, Room L225, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, CA 94305, USA

Introduction

Soft-tissue tumours are neoplasms that show morphological and immunophenotypical characteristics of mesenchymal cells such as fibroblasts, adipocytes, muscle cells, or peripheral nerve-sheath cells. These rare tumours have a wide range of histological appearances, within which over 100 different entities have been defined.[1] Although many soft-tissue tumours can be distinguished from each other, the boundaries between several diagnostic groups are vague and can overlap. The classification is further complicated by the fact that there are few reliable immunohistochemical markers to aid in tumour subclassification or to help predict a patient's outcome. The specialty of soft-tissue tumour pathology is therefore hampered—perhaps more than in any other group of tumours—by uncertainty surrounding diagnosis.

Over 30 000 expressed genes in man have (at least partly) been sequenced, and we expect that the sequence of all expressed genes will be available soon. Level of expression of such large numbers of genes, impossible by old methods, can now be studied with cDNA expression MICROARRAYS.[2] Furthermore, HIERARCHICAL CLUSTERING analysis recognises groups of genes that are co-expressed, providing a new level of insight into their possible functions. Microarray hybridisation technologies have begun to identify new molecular subclassifications in breast carcinomas,[3,4] lymphomas,[5] leukaemias,[6] melanomas,[7] and prostate cancer.[8]

We did cDNA gene microarray analysis on a set of 41 soft-tissue tumours to identify gene clusters that define tumour families on a molecular level, to relate these families to histological diagnoses and known molecular markers, and to highlight new markers of potential diagnostic value.

Methods

Specimens and RNA isolation

Frozen tissue samples were obtained from soft-tissue tumour specimens resected at the Vancouver Hospital and Health Sciences Centre, the Stanford University Medical Centre, and the Hospital of the University of Pennsylvania between 1993 and 2000. 41 specimens were used for this study: these included eight gastrointestinal stromal tumours; eight monophasic synovial sarcomas; four liposarcomas (one dedifferentiated, one myxoid, two pleomorphic); 11 leiomyosarcomas (including one primary and metastatic pair); eight malignant fibrous histiocytomas; and two benign peripheral nerve-sheath tumours (schwannomas). The clinical features of these tumours are shown on supplemental web table 1.[9]

We cut a frozen section from each specimen before RNA isolation to confirm that the sample was representative of the case. Tissue was homogenised in Trizol reagent (GibcoBRL/Invitrogen, Carlsbad, USA) and total RNA was prepared; we isolated mRNA by the FastTrack 2.0 method (Invitrogen).

GLOSSARY

EIGENGENE
A trend in gene expression. If an eigengene correlates with a suspected source of artifact, than it can be deduced from the dataset.

EIGENARRAY
Represents a similar trend in array types to an eigengene.

HIERARCHICAL CLUSTERING
Clustering in data mining is a statistical discovery process that groups a set of data in such a way that the intracluster similarity is kept to a maximum and the intercluster similarity is kept to a minimum. In the clustering process, two clusters are merged only if the interconnectivity and closeness (proximity) between two clusters are high relative to the internal interconnectivity of the clusters and closeness of items within the clusters.

MICROARRAY
A collection of spots on a solid surface (often a glass slide) arranged in neat rows and columns, so that the origin of each spot is known. Depending on the type, the spots can contain DNA sequences (as in gene microarrays), tissue fragments (tissue microarrays), protein (protein microarrays), or others. Microarrays can contain many such spots and are therefore useful in high throughput experiments.

SINGULAR VALUE DECOMPOSITION
A mathematical procedure by which trends in large datasets can be noted.

mRNA labelling and hybridisation to spotted cDNA microarrays

We prepared Cy3-labelled (green fluorescent) cDNA from reference mRNA and Cy5-labelled (red fluorescent) cDNA from mRNA from every tumour specimen. These cDNA samples were hybridised to 22 000 (22K) spotted cDNA microarrays, and subsequent analysis was done as described.[4] Halfway through this experiment, a new 42 000 (42K) gene array replaced the old 22K type. For this reason, subsequent cases were analysed on the 42K array. Reference mRNA was isolated from a pool of 11 cell lines.[4] Both arrays were prepared as described.[4,10,11] Five specimens for which adequate amounts of mRNA were available were analysed on both 22K and 42K gene arrays. We used SINGULAR VALUE DECOMPOSITION and ANOVA to identify and correct for bias introduced by different array types.

Data analysis

Intensity of Cy3 and Cy5 fluorescence for every gene spot on the hybridised arrays was measured with a Genepix 4000 scanner (Axon instruments, Foster City, USA), and was analysed with Genepix version 3.0 software (Axon instruments). The primary data tables and the image files are stored in the Stanford microarray database.[12] We entered fluorescence ratios into this database for analysis. Spots that could not be interpreted were excluded. We selected spots that had at least 80% well defined datapoints from the 46 arrays, and that had an absolute value of the fluorescence ratio at least three times greater than the geometric mean ratio of specimens looked at, in at least two arrays. A further selection criterion was that each spot should have a ratio of signal over background greater than 1·4 in either Cy3 or Cy5 channels. Hierarchical clustering was then done as described.[13] Hierarchical clustering analysis partitions genes into discrete groups, creating visually recognisable expression patterns and separating soft-tissue neoplasms into distinct groups.

We measured the expression pattern of the tumour set with two different types of slide arrays: one with 22 654 spotted cDNAs (22K array), and the other with almost all the gene set on the 22K slide plus about 20 000 additional cDNAs, for a total of 42 611 spots (42K array). To enlarge the total dataset, and thereby increase the number of tumours in any single group, the two array sets were combined. For this new combined dataset, we only included genes in both arrays. This dataset yielded closely similar tumour clustering of the major diagnostic groups, as was seen when the two datasets were analysed separately (web figures 1 and 2).[9] However, in the combined dataset, an effect of the type of array used (22K *vs* 42K) on clustering of the tumours was evident (web figure 2).[9] We did singular value decomposition to correct for this artifact (web figure 3).[9] This technique has previously been used to detect and correct for artifacts in time-course experiments[14] and has been applied in many other areas of research to filter out noise from signal.[15–17]

Singular value decomposition establishes unique dominant orthogonal (or uncorrelated) gene and corresponding array expression patterns—ie, so-called EIGENGENES and EIGENARRAYS, respectively—and then identifies genes and arrays with large projections onto or high correlations with these patterns. We identified one eigengene that correlated almost exactly with the 22K

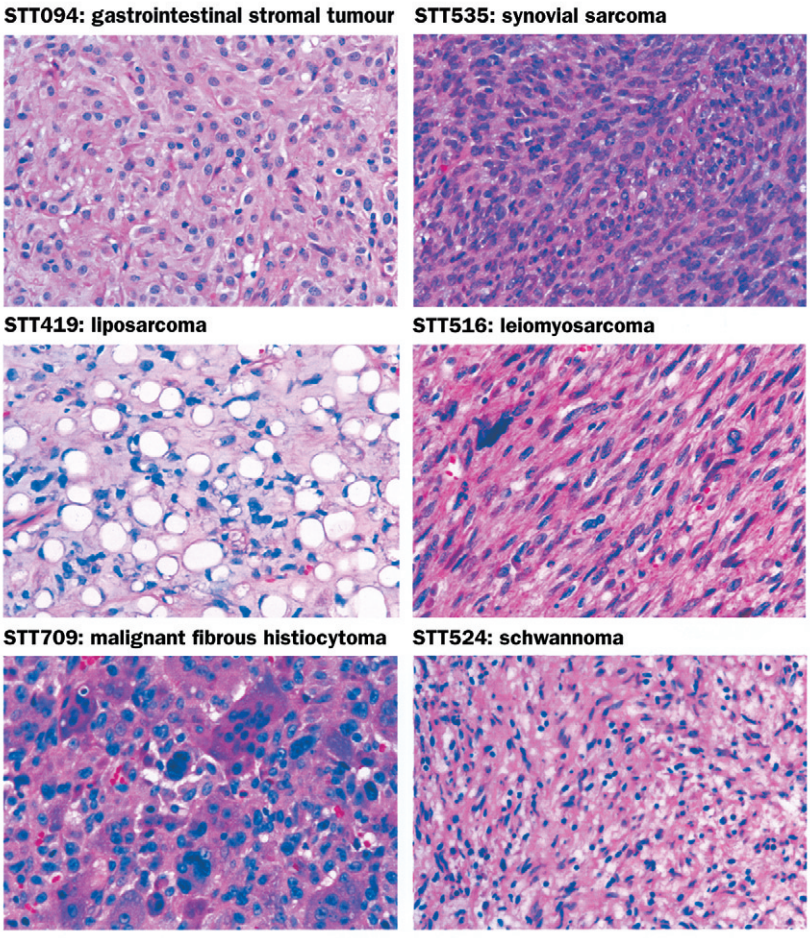

Figure 1: **Representative histological findings of specimens used for this study (×20)**

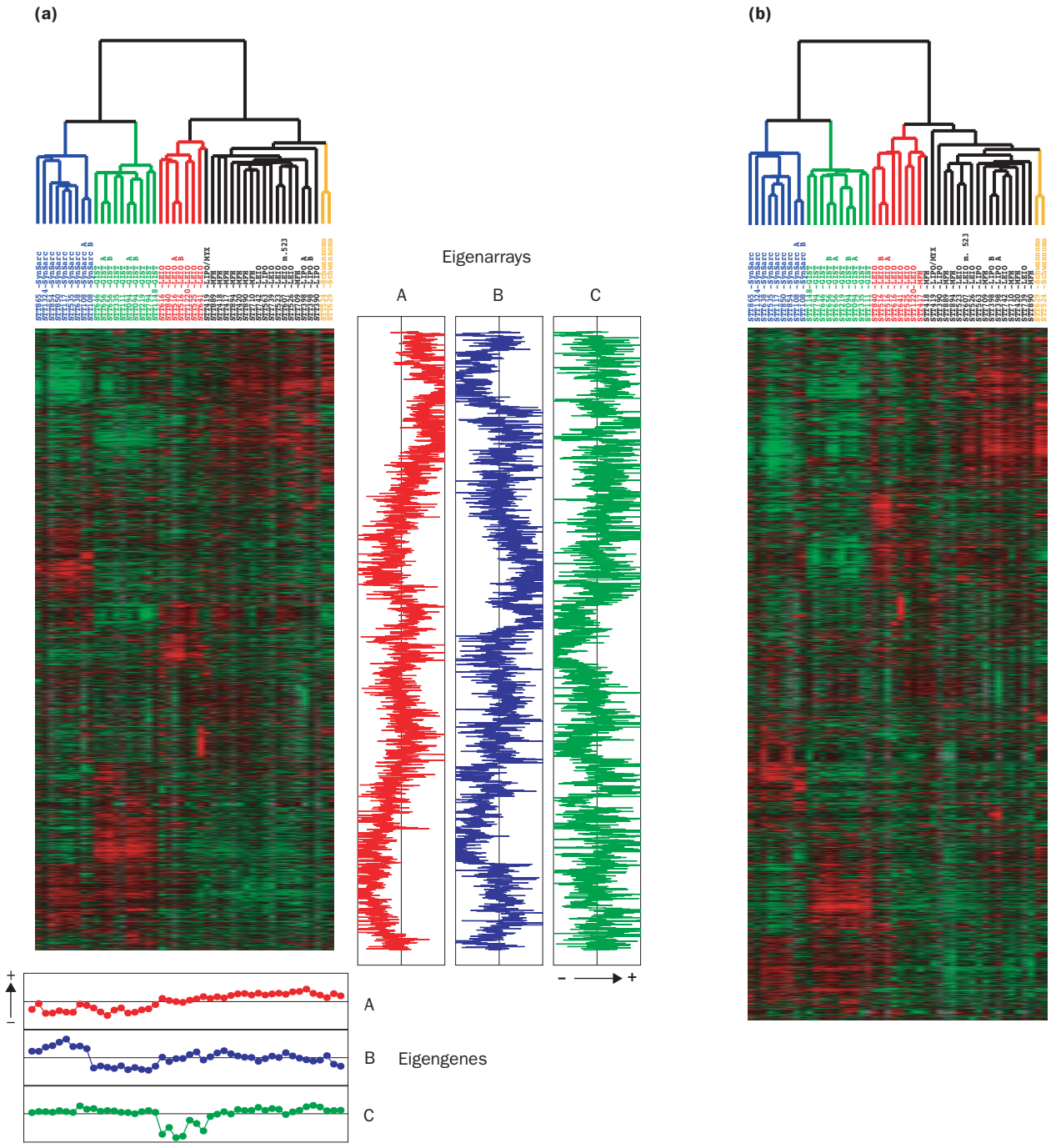

Figure 2: **Cluster analysis of 46 soft-tissue tumour specimens by singular value decomposition (a) and batch-specific centreing (b)**
SynSarc=synovial sarcoma; GIST=gastrointestinal stromal tumour; LEIO=leiomyosarcoma; MFH=malignant fibrous histiocytoma; LIPO=liposarcoma. (a) In the clustergram, a row represents the level of expression for a gene, centred at the geometric mean of its expression level in the 46 samples. Tumour specimens are arranged in columns. Red=high expression; green=low expression. The dendrogram (upper) shows degree of association between tumour samples, with short branches denoting a high degree of similarity. Five tumours, analysed on both 22K and 42K arrays (eg, STT108 SynSarc A and B, respectively) show tight pairwise clustering after correction of array bias. The first three most important eigengenes (lower) and eigenarrays (right) are aligned with the clustergram (middle), showing the relation between the two types of analyses. (b) A similar pattern of gene expression is seen when the 22K and 42K dataset are centred separately and then combined.

versus 42K array bias (web figure 4).[9] The effect of this eigengene and corresponding eigenarray was subtracted from all data. This new dataset was reselected for gene-expression levels as described above, and hierarchical clustering was done (web figure 5).[9] Subsequently, the final dataset was analysed by singular value decomposition. A second technique (ANOVA) was also used to correct array bias. In this technique, the data obtained on 22K and 42K arrays was centred separately by subtracting from each gene measurement the mean fluorescence ratio for that gene in the tumours examined in the 22K dataset and doing the same procedure on the genes in the 42K dataset. Subsequently both datasets were combined and reselected for gene-expression levels as described above. The data obtained by these two methods were highly similar. The results described below are based on the findings in the dataset selected through singular value decomposition.

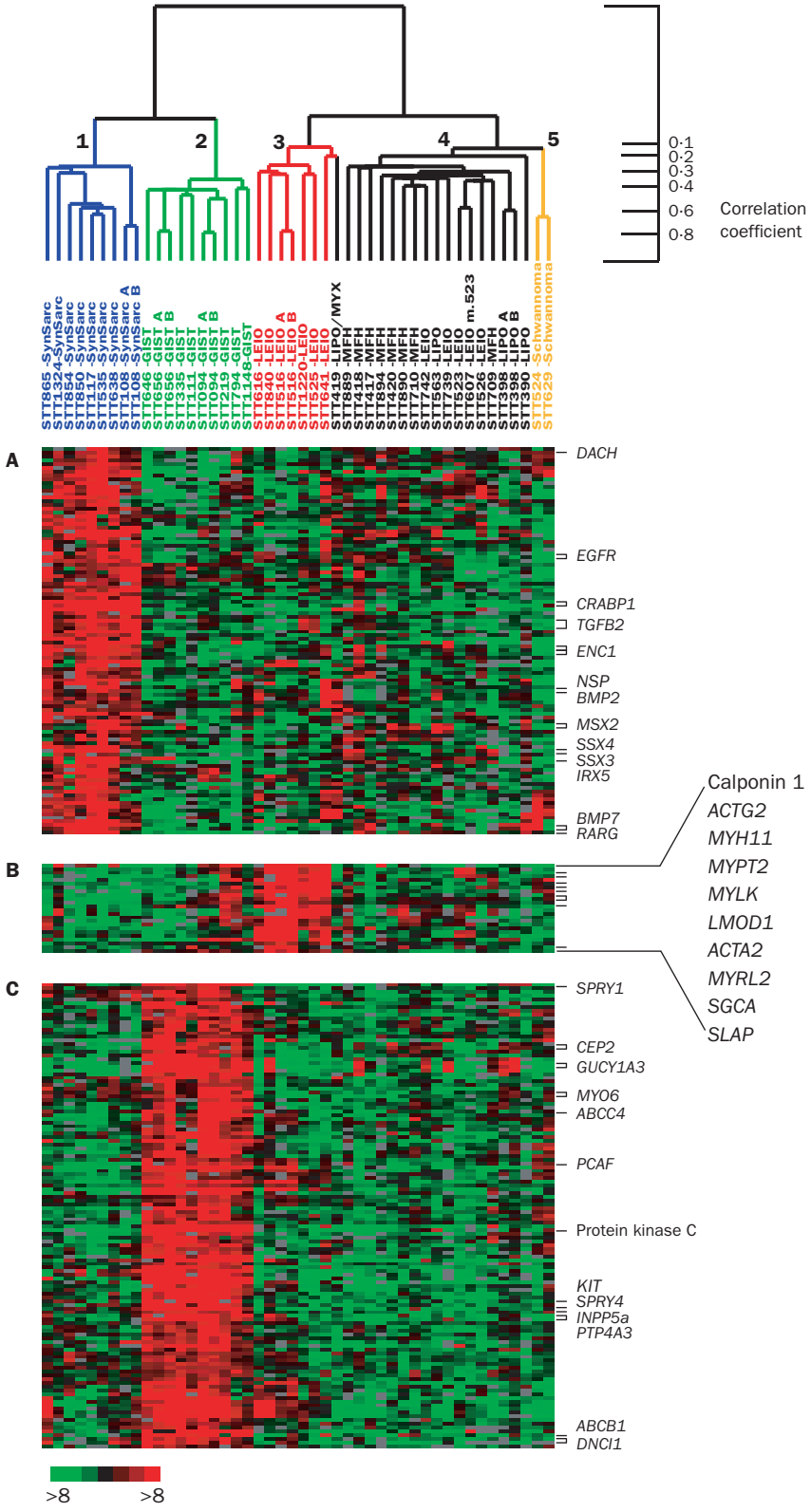

Figure 3: **Representative portions of the tumour-specific gene clusters**
Spectrum of green to red spots represents the centred expression for each gene. Sidebar shows difference from mean; selected gene names are shown on the right and defined in the panel. A=synovial sarcoma gene cluster; B=muscle gene cluster; C=gastrointestinal stromal tumour gene cluster.

A more detailed explanation of the methods, including singular value decomposition, is provided in the supplemental information on the accompanying website.[9] In addition to hierarchical clustering and singular value decomposition analysis, we used a supervised analytical method—significance analysis of microarrays—to search for differentially expressed genes in different sarcoma diagnoses.[18] A limited number of genes were reported to be misplaced on the arrays during this project. Only 27 of these genes survived singular value decomposition correction; these genes are named in a table on the website accompanying this report (http://image.thelancet.com/01art9007webonly.pdf) and have been removed from the clusters described in this report and from the complete dataset on the website. Should more misplaced genes be discovered they will be posted as an update on the website.[9]

Role of the funding source
The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

46 specimens from 41 soft-tissue tumours were analysed for gene-expression profiles by a combination of 22K and 42K cDNA microarrays. Histological sections of representative tumours are shown in figure 1; sections of all specimens used can be viewed on the accompanying website.[9] The relation between tumour type and gene-expression profile was analysed for 5520 well defined genes that showed variation in expression across the 46 arrayed specimens (figure 2) by hierarchical cluster analysis[13] and singular value decomposition.[14]

On the basis of levels of gene expression of all 5520 genes, specimens were separated into five major groups (figures 2 and 3). The eight synovial sarcomas and the eight gastrointestinal stromal tumours formed tight clusters on two distinct dendrogram branches. A third group comprised six of 11 leiomyosarcoma specimens. The five remaining leiomyosarcomas—including one primary tumour and its pulmonary metastasis—were clustered in a large group of tumours otherwise composed of malignant fibrous histiocytomas and liposarcomas. Both the benign peripheral nerve-sheath tumours were on the same terminal branch.

Clustering is further relevant for precise pairing of the five specimens analysed with both arrays (22K and 42K) and of the primary tumour with its pulmonary metastasis. Singular value decomposition identified several highly important eigen-genes and corresponding eigenarrays that correlated with tumour type (figure 2, web table 2).[9] Eigengene A correlates with the combination of synovial sarcomas and

Definitions of gene names

Gene name	Definition
A	
DACH	Dachshund
EGFR	Epidermal growth factor receptor
CRABP1	Cellular retinoic acid binding protein-1
TGFB2	Transforming growth factor β2
ENC1	Ectodermal-neural cortex-1
NSP	Neuron-specific protein Hs 79404
BMP2	Bone morphogenetic protein 2
MSX2	msh homeo box homolog-2
SSX4	Synovial sarcoma X breakpoint-4
SSX3	Synovial sarcoma X breakpoint-3
IRX5	Iroquois homeo-box protein 5
BMP7	Bone morphogenetic protein-7
RARG	Retinoic acid receptor γ
B	
ACTG2	Actin γ2 smooth muscle enteric
MYH11	Myosin heavy polypeptide 11 smooth muscle
MYPT2	Myosin phosphatase target subunit-2
MYLK	Myosin light polypeptide kinase
LMOD1	Leiomodin-1 smooth muscle
ACTA2	Actin α2 smooth muscle aorta
MYRL2	Myosin regulatory light chain-2
SGCA	Sarcoglycan α
SLAP	Sarcolemmal-associated protein
C	
SPRY1	Sprouty homolog-1
CEP2	cdc42 effector protein-2
GUCY1A3	Guanylate cyclase 1 α3
MYO6	Myosin VI
ABCC4	ATP-binding cassette C4
PCAF	p300/CBP associated factor
KIT	c-kit/CD117
SPRY4	Sprouty homolog 4
INPP5a	Inositol polyphosphate-5-phosphatase
PTP4A3	Protein tyrosine phosphatase type 4A 3
RAB3a	RAS-associated protein RAB3A
ABCB1	ATP-binding cassette B1
DNCI1	Dynein cytoplasmic intermediate peptide 1
SYNJ1	Synaptojanin-1

gastrointestinal stromal tumours from the remaining specimens; a negative value corresponds to a diagnosis of either stromal tumour or synovial sarcoma. Eigenarray A shows the genes that contribute to this distinction. Comparison with the clustergram shows that these genes fall into gene clusters that are specific for synovial sarcoma, stromal tumour, or both. Eigengene B separates synovial sarcomas (positive) from gastrointestinal stromal tumours (negative), with values for this eigengene in the remaining specimens around zero. Eigenarray B shows almost exact correlation with the genes reported in synovial sarcoma and gastrointestinal stromal tumour clusters. Eigengene C shows near-perfect correlation with the subset of leiomyosarcomas that express a muscle gene cluster, including calponin.

Significance analysis of microarrays independently identified several diagnosis-associated genes (web table 3),[9] which overlapped with the genes obtained by clustering and singular value decomposition (web table 4 and web figure 6).[9] Synovial sarcoma specimens were differentiated by a unique pattern of expression of a cluster of 104 genes (r=0·66), which included synovial sarcoma X (*SSX*), retinoic acid pathway genes, and epidermal growth factor receptor (figure 3). Another cluster of 125 genes (r=0·75) showed high levels of expression for only the gastrointestinal stromal tumours, and thus these tumours were separated from the other soft-tissue tumours on the cluster dendrogram (figure 3).

A distinct cluster of 24 genes (r=0·75) was highly expressed in six of 11 leiomyosarcoma specimens. This group, part of a much larger cluster of genes, contained many genes known to be implicated in muscle structure and function, including actin, myosin, leiomodin, myosin phosphatase, and calponin (figure 3). Most of the other five leiomyosarcoma specimens did not express these markers, but were positive for desmin by immunohistochemistry (data not shown). The separation of the calponin-positive leiomyosarcoma subgroup from the calponin-negative tumours was dominant, and resulted in an important eigengene and corresponding eigenarray (figure 2, web figure 3).[9]

The two peripheral nerve-sheath tumours were clustered together, and were characterised by high expression of five nerve-sheath genes (r=0·68), including neuroligin, L1 cell-adhesion molecule, neurexin, and semaphorin 3B precursor, with S100b present at a separate location (data not shown). The gene-expression profiles for malignant fibrous histiocytomas, liposarcomas, and the remaining leiomyosarcomas grouped these tumours together on a broad branch of the dendrogram (figure 3).

Discussion

We have reported gene-expression profiles of 41 soft-tissue tumours with cDNA microarrays; the complete dataset is available in a searchable format on the website accompanying this report.[9] We have shown that singular value decomposition analysis can be used to overcome bias introduced by use of different batches of arrays. The two methods used for removal of array bias showed strikingly similar results: 5520 genes survived reselection after removal of array bias by singular value decomposition. The centreing method was slightly more permissive and allowed 5925 genes to remain after reselection. The overlap between the two methods was impressive, with 5282 genes (96%) of the 5520 singular value decomposition dataset being present in the centred dataset. All genes described in this report were present in both datasets.

We identified a cluster of genes, including *SSX*, retinoic acid pathway genes, and epidermal growth factor receptor, that showed specific expression for synovial sarcoma. *SSX* is a fusion partner with *SYT* in the t(X;18) translocation, which is reported to arise in most synovial sarcomas.[19] *SSX3* and *SSX4* cDNA spots in our dataset were highly homologous with other *SSX* variants, including *SSX1* and *SSX2*, which are most often implicated in this translocation. This association means that we cannot establish type of *SSX* gene with the cDNA hybridisation method. The *SSX* sequence used on the array contains the 3′ end of the gene, and hence cannot discern native *SSX* from *SYT-SSX* translocation product. Other genes in this cluster that showed synovial-sarcoma-specific expression could be associated with the effects of this translocation.

Two components of the retinoic acid pathway were also present in this gene cluster—cellular retinoic acid binding protein-1 and retinoic acid receptor-γ. Many genes known to be experimentally induced in vertebrates by retinoic acid also correlated with this expression pattern, including *ENC1*,[20] *IRX5*,[21] and *TGF*β2.[22,23] The synovial-sarcoma gene cluster also contained the gene for the epidermal growth factor receptor, raising the possibility that synovial sarcomas could be amenable to treatment by small molecule tyrosine kinase inhibitors such as ZD1839,[24] or the epidermal growth factor receptor antibody C225.[25]

Within this group of sarcomas, a further subdivision of cases was noted that could correlate with differences in clinical behaviour.

We reported a cluster of genes with specific expression in gastrointestinal stromal tumours. This tightly associated group of genes was centred on *CD117* (*KIT*), a gene in which most of these stromal tumours have activating mutations.[26] Within the group of soft-tissue tumours, this gene is known to be highly expressed in gastrointestinal stromal tumours.[27] Also included in this cluster was the protein kinase C gene, and immediately adjacent was *PIK3CG*—these genes code for two proteins that have been reported to transduce and inhibit *KIT* signalling, respectively, in various model systems.[28,29]

Gastrointestinal stromal tumours are known to be more resistant to chemotherapy than other sarcomas.[30] Genes coding for two members of the superfamily of ATP-binding-cassette transporters (*ABCB1* and *ABCC4*) were recorded within the cluster of genes specific for these stromal tumours; both are known to be implicated in multidrug resistance.[31] ABCB1 protein expression in gastrointestinal stromal tumours has been reported.[32] A gene set expressed in both this group of tumours and synovial sarcomas included *BCL2*, which is consistent with known immunohistochemical staining patterns of these tumours.[33]

The prevalence of activating *KIT* mutations in gastrointestinal stromal tumours,[26] and recognition of genes from the *KIT* signalling pathway in this cluster, lend further support to the idea that aberrant *KIT* activity is important in transformation of these tumours. The presence of genes known to counteract the function of *KIT* in this gene set suggests that gastrointestinal stromal tumours have retained various normal cell functions that supply inhibitory feedback to, now aberrant, *KIT* activity. These tumour samples have also reduced cell-cycle activity, as shown by underexpression of proliferation genes, with an increase in antiapoptotic activity suggested by relatively high *BCL2* expression (data not shown). Taken together, these findings suggest that gastrointestinal stromal tumours are low-grade and of low complexity, which are driven by few mutations, or perhaps one mutation, in *KIT*.

The central prominence of *KIT* in the gene cluster defining gastrointestinal stromal tumours accords with the finding that imatinib mesylate (formerly STI571), a tyrosine-kinase inhibitor with activity against *KIT*, seems highly effective in slowing of tumour growth.[34] Only one of eight stromal tumours was negative for *KIT* by immunohistochemistry. Despite this case's failure to stain for *KIT*, it was clinically diagnosed as a gastrointestinal stromal tumour by its histological findings and the location of the tumour in the gastric wall. On gene-array analysis, this case had high levels of *KIT* mRNA, and most of the genes that are associated with *KIT* in the gene cluster specific for this group of tumour were present.

This case is one example of the increase in reliability of tumour classification by genome-wide analysis compared with immunohistochemistry with limited numbers of markers. We noted that gastrointestinal stromal tumours are widely separate from leiomyosarcomas. This observation lends support to the hypothesis that these gastrointestinal stromal tumours are derived not from the smooth muscle fibres of the intestinal tract but from cells associated with pacemaker activity in the bowel wall.[35]

We showed the existence of two subgroups of leiomyosarcoma, both distinct from gastrointestinal stromal tumours. Too few cases were available to allow for meaningful comparison based on histological findings between the two groups. Clinical features, such as tumour location, did not account for the separation of the leiomyosarcomas into these separate groups.

We also reported a group of genes associated with malignant fibrous histiocytomas, liposarcomas, and the remaining leiomyosarcomas. Division of this group was not correlated with morphological diagnosis, but rather was the result of a collection of genes with fibrous and histiocytic features.[9] The histiocytic part included genes characteristically expressed by macrophages, including *CD11b* and *CD68*. Also prominent were genes of the interferon-responsive cluster, identified previously in breast carcinoma,[3] and genes associated with other inflammatory processes. The fibrous part of the gene set included many genes for collagen, collagen metabolism, and constituents of the extracellular matrix and angiogenesis. Overall, these genes overlapped with the stromal/fibroblast and endothelial gene clusters identified in molecular profiling of breast tumours.[4] The absence of a clear distinction between malignant fibrous histiocytomas, liposarcomas, and a subset of the leiomyosarcomas correlates with ongoing debate on the degree to which malignant fibrous histiocytomas can be diagnosed as a separate tumour category.[36]

In all our tumour-specific clusters, we identified several candidate tumour markers and potential drug targets. Genes that help define tumour-specific clusters have potential value for resolution of differential diagnoses, but need to be validated at the immunohistochemical level. The expression levels identified with spotted microarray technology might not always correspond to the amount of protein product, or might be represented by small quantities of protein that are difficult to detect by immunohistochemistry. However, it is reassuring to note that known immunohistochemical markers (*S100b*, *BCL2*, and *KIT*) are present in the identified clusters, suggesting that other, new, useful markers will also be present.

Genome-wide expression profiling could lead to improved soft-tissue tumour classification. Molecular profiling of various tumours of this type will shed light on the cells of origin for many of these tumours, and could also allow better understanding of poorly described normal connective-tissue counterparts of these tumours, such as fibroblasts, myofibroblasts, and pericytes. Some of the highly expressed genes we identified are implicated in the pathogenesis of these tumours. Our analysis has recorded numerous genes that cluster with these markers, and thus that could also be associated with pathogenesis. Further study of these genes could lead to new diagnostic or prognostic markers and new therapeutic targets.

Contributors

T O Nielsen, R B West, and S C Linn contributed equally to this project with tissue selection, study design, and data analysis. O Alter did singular value decomposition analysis. M A Knowling and J X O'Connell did tissue selection and data analysis. S Zhu did array hybridisation. M Fero and G Sherlock supervised array synthesis and database design and management. P O Brown, D Botstein, M van de Rijn, and J R Pollack were responsible for data analysis and study design.

Conflict of interest statement

None declared.

Acknowledgments

This work was supported by National Institutes of Health grants CA85129 and CA84967 and the Howard Hughes Medical Institute. POB is an associate investigator of the Howard Hughes Medical Institute. OA is an Alfred P Sloan and US Department of Energy postdoctoral fellow in Computational Molecular Biology, and has been awarded a National Human Genome Research Institute Individual Mentored Research Scientist Development grant in genomic research and analysis (National Institutes of Health, 1 K01 HG00038-01). SCL receives a Dutch Cancer Society postdoctoral research fellowship.

References

1 Weiss SW, Goldblum JR. Soft tissue tumors. St Louis: Mosby, 2001.
2 Phimister B, ed. The chipping forecast. *Nat Genet* 1999; **21** (suppl)**:** 3–60.
3 Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA* 1999; **96:** 9212–17.
4 Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; **406:** 747–52.
5 Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403:** 503–11.
6 Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286:** 531–37.
7 Bittner M, Meltzer P, Chen Y, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 2000; **406:** 536–40.
8 Dhanasekaran SM, Barrette TR, Ghosh D, et al. Delineation of prognostic biomarkers in prostate cancer. *Nature* 2001; **412:** 822–26.
9 Molecular portraits of soft-tissue tumours. http://genome-www.stanford.edu/sarcoma/ (accessed Jan 11, 2002).
10 The Brown lab. http://cmgm.Stanford.EDU/pbrown/ (accessed Jan 14, 2002).
11 Molecular portraits of human breast tumours. http://genome-www.stanford.edu/breast_cancer/molecularportraits/ (accessed Jan 14, 2002).
12 Stanford microarray database. http://genome-www4.stanford.edu/MicroArray/SMD/ (accessed Jan 14, 2002).
13 Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998; **95:** 14863–68.
14 Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* 2000; **97:** 10101–06.
15 Swinnen A, Van Huffel S, Van Loven K, Jacobs R. Detection and multichannel SVD-based filtering of trigeminal somatosensory evoked potentials. *Med Biol Eng Comput* 2000; **38:** 297–305.
16 Zabel M, Acar B, Klingenheben T, Franz MR, Hohnloser SH, Malik M. Analysis of 12-lead T-wave morphology for risk stratification after myocardial infarction. *Circulation* 2000; **102:** 1252–57.
17 Calamante F, Gadian DG, Connelly A. Delay and dispersion effects in dynamic susceptibility contrast MRI: simulations using singular value decomposition. *Magn Reson Med* 2000; **44:** 466–73.
18 Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 2001; **98:** 5116–21.
19 dos Santos NR, de Bruijn DR, van Kessel AG. Molecular mechanisms underlying human synovial sarcoma development. *Genes Chromosomes Cancer* 2001; **30:** 1–14.
20 Hernandez MC, Andres-Barquin PJ, Holt I, Israel MA. Cloning of human ENC-1 and evaluation of its expression and regulation in nervous system tumors. *Exp Cell Res* 1998; **242:** 470–77.
21 Gomez-Skarmeta JL, Glavic A, de la Calle-Mustienes E, Modolell J, Mayor R. Xiro, a Xenopus homolog of the Drosophila Iroquois complex genes, controls development at the neural plate. *EMBO J* 1998; **17:** 181–90.
22 Glick AB, Flanders KC, Danielpour D, Yuspa SH, Sporn MB. Retinoic acid induces transforming growth factor-beta 2 in cultured keratinocytes and mouse epidermis. *Cell Regul* 1989; **1:** 87–97.
23 Choudhury A, Singh RK, Moniaux N, El-Metwally TH, Aubert JP, Batra SK. Retinoic acid-dependent transforming growth factor-β2-mediated induction of MUC4 mucin expression in human pancreatic tumor cells follows retinoic acid receptor-alpha signaling pathway. *J Biol Chem* 2000; **275:** 33929–36.
24 Ryan PD, Chabner BA. On receptor inhibitors and chemotherapy. *Clin Cancer Res* 2000; **6:** 4607–09.
25 Shin DM, Donato NJ, Perez-Soler R, et al. Epidermal growth factor receptor-targeted therapy with C225 and cisplatin in patients with head and neck cancer. *Clin Cancer Res* 2001; **7:** 1204–13.
26 Hirota S, Isozaki K, Moriyama Y, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* 1998; **279:** 577–80.
27 Miettinen M, Sobin LH, Sarlomo-Rikala M. Immunohistochemical spectrum of GISTs at different sites and their differential diagnosis with a reference to CD117 (KIT). *Mod Pathol* 2000; **13:** 1134–42.
28 Shearman MS, Herbst R, Schlessinger J, Ullrich A. Phosphatidylinositol 3′-kinase associates with p145c-kit as part of a cell type characteristic multimeric signalling complex. *EMBO J* 1993; **12:** 3817–26.
29 Blume-Jensen P, Siegbahn A, Stabel S, Heldin CH, Ronnstrand L. Increased Kit/SCF receptor induced mitogenicity but abolished cell motility after inhibition of protein kinase C. *EMBO J* 1993; **12:** 4199–209.
30 Brennan MF, Alektiar KM, Maki RG. Sarcomas of the soft tissue and bone. In: DeVita VT Jr, Hellman S, Rosenberg SA, eds. Cancer: principles and practice of oncology. Philadelphia: Lippincott, Williams, and Wilkins, 2001; 1841–91.
31 Klein I, Sarkadi B, Varadi A. An inventory of the human ABC proteins. *Biochim Biophys Acta* 1999; **1461:** 237–62.
32 Plaat BE, Hollema H, Molenaar WM, et al. Soft tissue leiomyosarcomas and malignant gastrointestinal stromal tumors: differences in clinical outcome and expression of multidrug resistance proteins. *J Clin Oncol* 2000; **18:** 3211–20.
33 Suster S, Fisher C, Moran CA. Expression of bcl-2 oncoprotein in benign and malignant spindle cell tumors of soft tissue, skin, serosal surfaces, and gastrointestinal tract. *Am J Surg Pathol* 1998; **22:** 863–72.
34 Joensuu H, Roberts PJ, Sarlomo-Rikala M, et al. Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med* 2001; **344:** 1052–56.
35 Kindblom LG, Remotti HE, Aldenborg F, Meis-Kindblom JM. Gastrointestinal pacemaker cell tumor (GIPACT): gastrointestinal stromal tumors show phenotypic characteristics of the interstitial cells of Cajal. *Am J Pathol* 1998; **152:** 1259–69.
36 Fletcher CD, Gustafson P, Rydholm A, Willen H, Akerman M. Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. *J Clin Oncol* 2001; **19:** 3045–50.