Molecular characterisation of soft tissue tumors: 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

Home

Molecular Portraits of Soft Tissue Tumors

Explore

GenExplore the total dataset of 46 arrays x 5520 genes

Materials and Methods

Description of Materials and Methods

Figures and Tables

Figures, Supplemental Web Figures, and Tables

Supplemental Information

Supplemental information on data analysis

Download

Download complete dataset

Histology

H&E sections from the tumors

Authors

Initial Clustering

The initial set of tumors was analyzed on 22K arrays. Subsequently, larger 42K arrays replaced the 22K arrays and additional tumors were analyzed on these larger arrays. Dendrograms were obtained when both sets were analyzed separately (**Supplemental Figure 1**). The 22K group had 26 tumors, while the 42K group contained 20 tumors. In both array types, the tumors segregated into discrete groups according to pathologic diagnosis for the synovial sarcomas and the GI stromal tumors. The remaining tumors did not cluster clearly according to pathologic diagnosis, partly due to the low number of cases available for analysis.

Combining 22 and 42K Arrays

The initial gene selection procedure with the combined set of 22K and 42K arrays yielded 7425 well-measured genes that were present on both types of arrays, representing 20% of the maximum number of genes available for analysis. Selection for signal/background ratio and manually flagged spots had removed 17% of the genes, selection for 80% good data for each gene removed a further 32%, and selection for a fluorescence ratio of at least 3-fold greater than the geometric mean ratio for the specimens examined in at least 2 arrays removed another 31%. When results from 22 and 42K arrays were combined a new dendrogram was derived (**Supplemental Figure 2**). With more tumors available for analysis, additional discrete groups of tumors were noted. For example, the two schwannomas, although run on different array types, formed a tight group distinct from the remaining specimens. In addition, a group of three leiomyosarcomas (including STT516, which was run on both 22K and 42K arrays) now formed a tight cluster. All synovial sarcoma and GIST samples continued to cluster in distinct groups. However, an apparent 22K versus 42K array bias was observed that contributed to the cluster pattern. For example, in the synovial sarcoma cluster, the five specimens that were run only on 42K arrays clustered on a branch distinct from the other specimens. While four of the five tumors that had been analyzed on both arrays clustered pair-wise together one did not. GIST-STT094-A (22K array) seemed more similar to another GIST (STT219, also run on a 22K array), than to GIST-STT094-B (42K array). Finally, the correlation was quite low in three pairs, with only leiomyosarcoma STT516 showing a high degree of correlation between 22K and 42K arrays. The mean correlation coefficient, obtained with centered data, was 0.61 for the 5 pairs.

Singular Value Decomposition (SVD)

In an attempt to identify and correct the 22K versus 42K array bias, we performed SVD[1]. This analysis identified a number of eigengenes and corresponding eigenarrays in the dataset (**Supplemental Figure 3**). Several of the most significant eigengenes correlated with specific tumor groups such as the synovial sarcomas, gastrointestinal stromal tumors, and the subset of leiomyosarcomas expressing a cluster of muscle markers, including calponin. A single eigengene correlated almost perfectly with the tumors based on whether a 22K or 42K array was used for analysis (**Supplemental Figure 4**). Panel (a) shows the clustergram of all selected genes used for this report, with the arrays in the order obtained in the final dataset. Panel (b) describes the level of expression of this eigengene in each of the 46 arrays, with a near complete correlation between the expression level and the type of array used, showing a positive value found in almost all 42K arrays and a negative value in almost all 22K arrays. A different representation of these data is shown in panel (c), where the arrays have been put in the order dictated by the value for this eigengene. This shows that the vast majority of 42K arrays have a value above zero for this eigengene. Finally, panel (d) shows the contribution of each of the genes to the eigenarray that represents the 22K versus 42K array bias. Only those genes whose value is zero in this analysis are not affected; thus, it appears that the vast majority of genes are influenced by this array bias. The eigengene and eigenarray correlating with the slide bias were subtracted from the data set, and this adjusted data set was then reselected using the same criteria that generated the first data set. Because the corrected expression of those genes heavily influenced by array bias in some cases no longer varied >3 fold from the geometric mean ratio in at least two experiments, this reselection step led to the removal of 1905 genes. Thus, the initial set of 7425 genes was reduced to 5520. It should be noted however that almost all genes received some contribution to their expression levels from array bias. Subtraction of the array-type bias thus not only removed a specific set of genes but also improved the biological significance of the expression levels determined for all genes. The adjusted data set was reclustered to yield the final tumor dendrogram (**Supplemental Figure 5a**) and clustergram (**Figure 2**). Several observations can be made. First, all five tumors that had been analyzed on both array types now were located on shared

terminal branches. Second, the correlation between the pair members had improved from 0.61 before SVD to 0.73 after SVD. It should be noted that the data used for this comparison was centered, which emphasizes differences rather than similarities in gene expression. Third, a much less conspicuous clustering based on array type was noted in that the synovial sarcomas ran on 42K arrays no longer were located on a branch separate from the others. Finally by removing the array bias, the subset of calponin-expressing leiomyosarcomas that grouped tightly together had increased from 4 to 6 specimens. After singular value decomposition and subtraction of the slide bias eigengene the major gene clusters appear more condensed and readily interpretable than seen on the uncorrected clustergram (**Supplemental Figure 5b**).

Comparison of classification of genes by hierarchichal clustering, SVD and SAM

We used three complementary methods for the analysis of the data: hierarchical clustering, SVD and SAM. Clustering and SVD gave similar classifications of the tumor samples. Clustering, SVD and SAM gave similar classifications of the genes (where the supervised SAM analysis made use of the sample classification in generating gene classifications). For comparison of the classification of genes by clustering, SVD and SAM, we combined (see **Web Table 4** and **Supplemental Fig. 6**) the GIST gene cluster (**Fig. 3c**) with the SVD scores (**Web Table 2b**) and SAM score (**Web Table 3**) for the genes in this cluster.

Clustering places the gene kit almost in the center of this tight cluster of 125 genes with a correlation coefficient of 0.75. SVD ranks kit as the 41th gene, based on the high negative projection of its expression pattern onto the direction defined by eigengene B, the eigengene that distinguishes between GIST and SynSarc samples. SVD also gives kit a high anticorrelation value of 0.59 with eigengene B. Together with the high value of anticorrelation with eigengene A, the eigengene that distinguishes GISTs and SynSarcs from the rest of the tumor samples, of 0.65, kit has about 0.9 (i.e., Sqrt[0.59^2+0.65^2]) of its expression in the "GIST subspace" that is defined by these two eigengenes.

Note that, out of the 125 genes in the GIST cluster, 64 genes (or 51%) overlapped with the list of 225 genes that combined the top 125 genes ranked by SVD for highest negative projection onto eigengene B and the top 125 genes ranked by SVD for high anticorrelation with eigengene B. Also, 85 genes (or 68%) overlapped with the top 125 genes ranked by the SAM score.

ANOVA

We have also performed another type of data analysis (ANOVA) to remove the artifact induced by the use of 2 types of gene arrays. This analysis showed highly similar results to that obtained through SVD.

Identification of misplaced genes on arrays

During the analysis of this project a limited number of misplaced genes were identified. To date, only 35 genes were found in the pre-SVD dataset (a total of 7,425 genes), 27 of which remained after SVD correction (a total of 5,520 genes). These genes did not influence the data analysis and are not subject of discussion in this report. These genes theoretically could have contributed to the bias introduced by the use of 22K and 42K arrays. The misplaced genes are noted in **Web Table 5** and will be updated if additional errors are identified. The misplaced genes have been removed from the Gene Explorer dataset on this website.

Supplemental Information References

1. Alter, O. et al. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 97, 10101-10106 (2000).

cDNA clones and microarray production
The sets of 22 654 and 42 611 human cDNA genes/clones used in this study were obtained from Research Genetics (Huntsville AB, USA) (http://www.resgen.com/). The cDNA microarrays used in this study were made as previously described[1,2]. Detailed protocols are available at The Old Microarray Homepage and The Brown Lab's MGuide.

Common Reference Sample
Each of the 46 experimental samples tested here was analyzed by a comparative hybridization, using a common "reference" mRNA pool as a standard; this reference sample was composed of equal mixtures of mRNA isolated from 11 established human cell lines (MCF7, Hs578T, OVCAR3, HepG2, NTERA2, MOLT4, RPMI-8226, NB4+ATRA, UACC-62, SW872, and Colo205: see Common Reference Cell Line List for more details). The 11 cell lines were all grown to 70-90% confluence in RPMI medium containing 10% Fetal Calf Serum and Penicillin/Streptomycin. The cells were harvested either by scraping or centrifugation, and quickly resuspended in RNA lysis buffer and mRNA prepared as described in Perou et al.[3]. In each case, multiple individual mRNA preparations were collected for each cell line, which were then pooled together and analyzed via Northern analysis before final mixing to ensure the quality of the input mRNAs. The 11 mRNA samples were then mixed together in equal amounts, aliquoted in 10mM Tris (7.4), and stored at -80 C until use (2 micrograms of common reference sample was used per microarray hybridization and was always labeled using Cy3).

Specimens and RNA isolation
Frozen tissue samples were archived from soft tissue tumors resected at the Vancouver Hospital & Health Sciences Centre, the Stanford University Medical Center, and the Hospital of the University of Pennsylvania in the period 1993-2000. A total of 41 specimens was used for this study, including 8 gastrointestinal stromal tumors (GIST), 8 monophasic synovial sarcomas (SS), 4 liposarcomas (1 dedifferentiated (STT563), 1 myxoid (STT419), 2 pleomorphic), 11 leiomyosarcomas (including one primary & metastatic pair), 8 malignant fibrous histiocytomas (MFH), and 2 benign peripheral nerve shealth tumors (Schwannoma). The clinical features of these tumors are shown on Supplemental Data Table 1. A frozen section was cut from each specimen prior to RNA isolation to confirm that the archived material was representative of the case. Frozen tissue specimens were anonymized and assigned an experimental code. Tissue was homogenized in Trizol reagent (GibcoBRL) and total RNA was prepared as described[3]; mRNA was then isolated using the FastTrack 2.0 method following the manufacturer's protocol (see Chuck Perou's Tumor mRNA Isolation Protocol for the detailed protocol).

mRNA Labeling and hybridization to spotted cDNA microarrays
Preparation of Cy3 (green fluorescent) labeled cDNA from reference mRNA and Cy5 (red fluorescent) labeled cDNA from each tumor specimen mRNA, hybridization to 22 000 and 42 000 (22K and 42K) spotted cDNA microarrays, and subsequent analysis was performed as described [4]. Halfway through this experiment, a new 42K gene array type replaced the old 22K gene array type and allowed expansion of the total number of genes used from 22 654 to 42 611. For this reason subsequent cases were analyzed on the larger arrays. The reference mRNA was isolated from a pool of 11 cell lines, identical to that described previously [4]. Both arrays were prepared as described [4] with detailed protocols available at The Brown Lab and http://genome-www.stanford.edu/molecularportraits/. Five specimens for whom adequate amounts of mRNA were available were analyzed on both 22K (A specimens) and 42K (B specimens) gene arrays. This allowed us to use SVD to identify and correct for the bias introduced by different array types.

Data Analysis
(Supplemental Figures 1-5b); The levels of Cy3 and Cy5 fluorescence for each gene spot on the hybridized arrays were obtained with a Genepix 4000 scanner (Axon instruments), and analyzed with Genepix 3.0 software (Axon instruments). The primary data tables and the image files are stored in the Stanford Microarray Database. Fluorescent ratios were entered in the database for analysis. Uninterpretable spots were manually flagged and excluded. A selection was made from the remaining spots to include only those with at least 80% well-measured data points among the 46 arrays, with a fluorescence ratio at least 3 fold greater than the geometric mean ratio in the specimens examined 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 green or red channels. In this manner, 7425 array elements were identified. Hierarchical clustering was then performed as described[5]. The expression pattern of the tumor set was measured using two different types of slide arrays, one with 22K genes and the other with 42K genes, which contained almost the entire gene set represented on the 22K slide plus approximately 20 000 additional cDNAs, for a total of 42 611 spots (Supplemental Figure 1). To enlarge the total data set, and thereby increase the number of tumors in any single group, the two array sets were combined. For this new combined data set, we included only those genes present on both the 22K and 42K arrays. The combined dataset yielded a similar tumor clustering of the major diagnostic groups as was observed when either of the 2 datasets was analyzed separately (Supplemental Figures 1,2). However, in the combined dataset an influence of the type of array used (22K vs. 42K) on the clustering of the tumors was evident (Supplemental Figure 2). We performed singular value decomposition (SVD) in order to correct for this artifact (Supplemental Figure 3). This technique has previously been used to detect and correct artifacts in time course experiments[6] and has been applied in many other fields of research to filter out noise from signal[7-9]. SVD determines unique dominant orthogonal (or uncorrelated) gene and corresponding array expression patterns (i.e. "eigengenes" and "eigenarrays," respectively) that can be associated with some of the independent pathways and corresponding cellular states,

that make up the similarities and differences among the distinct STT groups. A single “eigengene” was identified that correlated almost perfectly with the 22K versus 42K array bias (Supplemental Figure 4). The influence of this “eigengene” and corresponding “eigenarray” was subtracted from all data. This new data set was reselected for gene expression levels as described above and hierarchical clustering was performed (Supplemental Figure 5). Subsequently, the final data set was again analyzed by SVD (Figures 2,3). A more detailed explanation of the methods, including SVD is provided in the supplemental information section on this website (Supplemental Information). In addition to hierarchical clustering and SVD analysis, we used a supervised analytical method, SAM (Significance Analysis of Microarrays), to search for differentially expressed genes among different sarcoma diagnoses[10].

Material & Methods References

1. Ross, D. T. et al. Systematic variation in gene expression patterns in human cancer cell lines [see comments]. Nat Genet 24, 227-235 (2000).
2. Alizadeh, A. A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [see comments]. Nature 403, 503-511 (2000).
3. Perou, C. M., et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 96, 9212-9217 (1999).
4. Perou, C.M., et al. Molecular portraits of human breast tumours. Nature 406, 747-752 (2000).
5. Eisen, M.B., et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95, 14863-14868, 1998.
6. Alter, O. et al. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 97, 10101-10106 (2000).
7. Swinnen, A. et al. Detection and multichannel SVD-based filtering of trigeminal somatosensory evoked potentials. Med Biol Eng Comput 38, 297-305 (2000).
8. Zabel, M. et al. Analysis of 12-lead T-wave morphology for risk stratification after myocardial infarction. Circulation 102, 1252-1257 (2000).
9. Calamante, F. et al. Delay and dispersion effects in dynamic susceptibility contrast MRI: simulations using singular value decomposition. Magnetic Resonance in Medicine 44, 466-473 (2000).
10. Tusher, V. G. et al. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 98, 5116-5121 (2001).

Figures and Tables

Figure 1.
Representative histology of specimens used for this study, including: gastrointestinal stromal tumor, synovial sarcoma, liposarcoma, leiomyosarcoma, malignant fibrous histiocytoma, and schwannoma. Histologic sections of all specimens used can be viewed on the accompanying webpage.

Figure 2.
A: Complete clustergram of the 46 soft tissue tumor specimens. A row in the cluster represents the relative level of expression for a gene, centered at the geometric mean of its expression level among the 46 samples, and displayed using red (relative high expression) and green (relative low expression) coloration. Tumor specimens are arranged in columns. The dendrogram of the tumor clustering is displayed above and describes the degree of relatedness between tumor samples, with short branches denoting a high degree of similarity. The first three most significant eigengenes and eigenarrays are aligned with the clustergram on the bottom and along the right side, respectively. Eigengene A correlates with the combination of synovial sarcomas and GIST from the remaining specimens, with a negative value corresponding to a diagnosis of either GIST or synovial sarcoma. Eigenarray A shows the genes that contribute to this distinction. Comparisons with the clustergram show that these genes fall into gene clusters that are specific for synovial sarcoma and/or GIST specimens. Likewise eigengene B separates synovial sarcomas (positive value) from GIST specimens (negative value), with values for this eigengene around zero in the remaining specimens. Eigenarray B shows almost perfect correlation with the genes found in the synovial sarcoma and GIST clusters. Finally, eigengene C show a near perfect correlation with the subset of leiomyosarcomas that express a muscle gene cluster, including calponin. B: An essentially similar pattern of gene expression is obtained when the 22K and 42K dataset are centered separately and then combined.

Figure 3.
Representative portions of the tumor specific gene clusters. The spectrum of green to red spots represents the relative centered expression for each gene (sidebar shows fold difference from mean); selected gene names are shown on the right. The branches of the array dendrogram are numbered from 1 to 5 as indicated in the text. Correlation coefficient bar shown to the right side of the dendrogram indicates the degree of relatedness between branches of the dendrogram. **Panel a**: synovial sarcoma gene cluster (DACH: dachshund, EGFR: epidermal growth factor receptor, CRABP1: cellular retinoic acid binding protein-1, TGFB2: transforming growth factor ß2, 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, FOXC1: forkhead box C1, BMP7: bone morphogenetic protein-7, RARG: retinoic acid receptor ?). **Panel b**: muscle gene cluster (ACTG2: actin ?2 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 a2 smooth muscle aorta, MYRL2: myosin regulatory light chain-2, SGCA: sarcoglycan a: SLAP: sarcolemmal-associated protein). **Panel c**: gastrointestinal stromal tumor gene cluster (SPRY1: sprouty homolog-1, CEP2: cdc42 effector protein-2, GUCY1A3: guanylate cyclase 1 a3, MYO6: myosin VI, ABCC4: ATP-binding cassette C4, PCAF: p300/CBP associated factor, prot kinase C: protein kinase C ?, kit: c-kit/CD117, SPRY4: sprouty homolog 4, INPP5a: inositol polyphosphate-5-phosphatase, PTP4A3: protein tyrosine phosphatase type 4A 3, ABCB1: ATP-binding cassette B1, DNCI1: dynein cytoplasmic intermediate peptide 1.

Supplemental Figure 1.
Hierarchical clustering dendrograms of the initial tumor sets hybridized to 22K and 42K arrays: A) 22K slide tumor set and B) 42K slide tumor set.

Supplemental Figure 2.
Hierarchical clustering dendrogram of the combined 22K and 42K arrays, before singular value decomposition. Tumor samples ran on both array types are identified by A for 22K arrays and B for 42K arrays. For each experiment, the type of array is also noted.

Supplemental Figure 3.
3A. SVD analysis of the combined dataset of both 22K and 42K arrays. Raster display of the expression data, with overexpression (red), no change in expression (black), and underexpression (green) around the geometric mean of relative expression, showing linear transformation of the data from the 7425-genes x 46-arrays space to the reduced diagonalized 46-eigenarrays x 46-eigengenes space using the 7425-genes x 46-eigenarrays and 46-eigengenes x 46-arrays basis sets. **3B.** Eigenarrays of the combined dataset of 22K and 42K arrays. (a) Complete clustergram of the 46 specimens. (b) Eigenarrays expression in all 7425 genes. At least the top 4 significant eigenarrays, corresponding to the top 4 significant eigengenes, display some order, when the genes are ordered in the clustergram order. **3C.** Eigengenes of the combined dataset of 22K and 42K arrays. (a) Raster display of the expression of 46 eigengenes in 46 arrays, with overexpression (red), no change in expression (black), and underexpression(green) around the geometric mean of the relative expression. (b) Bar chart of the probability of eigenexpression of each eigengene, showing about 16% of the overall relative expression in the most significant eigengene, that can be associated with the array-type bias, and about 14%, 10% and 6% of overall relative expression in the next 3 most significant eigengenes, that can be associated with the separation of synovial sarcomas and GIST from the remaining specimens, the separation of synovial sarcomas from the GISTs, and the separation of the subset of leiomyosarcomas that expresses a muscle gene cluster from the rest of the specimens, respectively.

Supplemental Figure 4.
Combined clustergram (panel a), eigengene (panel b, c), and eigenarray (panel d) specific for 22K/42K array bias.

Supplemental Figure 5a.
Hierarchical clustering dendrogram of the combined 22K and 42K array, after subtraction of the eigengene (and corresponding eigenarray) that is associated with the 22K/42K array bias, and after repeating the gene selection procedure (see methods).

Supplemental Figure 5b.
Comparison of the clustergrams from the cluster analysis of the initial combined data set and the subsequent data set that has undergone subtraction of the array-type bias followed by reselection of genes: initial combined data set before SVD (7425 genes), and data set after slide bias subtraction and reselection (5520 genes). The sidebars indicate the areas that encompass the gene sets unique for the GI stromal tumors (green), the synovial sarcomas (blue), and the calponin-positive subset of leiomyosarcomas (red).

Supplemental Figure 6.
Magnified section of hierarchical cluster of 125 genes, including kit (CD117), that correlated with GIST. The first 2 columns show the negative projection rank order and the correlation rank order for each of the 125 genes of eigengene 2. The third column shows the rank order for each gene after SAM analysis that identified those genes resposible for separating GIST from all other tumors. GIST are highlighted in green on the array dendrogram. The rank order of the genes in Web Figure 6 can be correlated with those reported in Web Tables 2 and 3. Please note that in those tables only named genes are included and duplicate genes were removed, hence there is no perfect correlation between rank order number in Web Figure 6 and those in Web Tables 2, 3. (See also **Web Table 4**).

Web Table 1.
Clinical features of tumors.

Web Table 2.

SVD sorting of genes by projection and correlation with eigengenes.
SVD defines the expression pattern of each gene to be a superposition i.e., a weighted sum of the expression patterns of all eigengenes. The projection of a gene onto an eigengene is the amplitude, i.e., the weight of this eigengene pattern in the expression of the given gene. The projection, therefore, measures the variation in expression of the gene along the direction defined by the eigengene. The correlation of a gene with an eigengene is the ratio between the corresponding projection and the overall amplitude of the expression pattern of the gene. The correlation, therefore, measures the similarity (or distance) between the expression pattern of the gene and that of the eigengene, that is independent of the overall amplitude of the expression pattern of the gene.

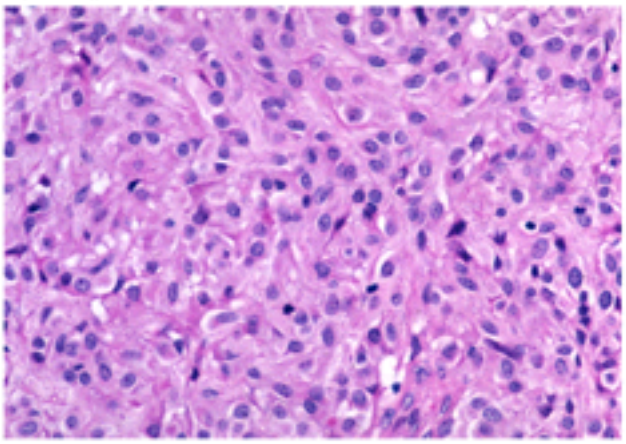
Web Table 2a. Eigengene A.
Web Table 2b. Eigengene B.
Web Table 2c. Eigengene C.

Web Table 3.
Significance of Microarray Analysis genes.

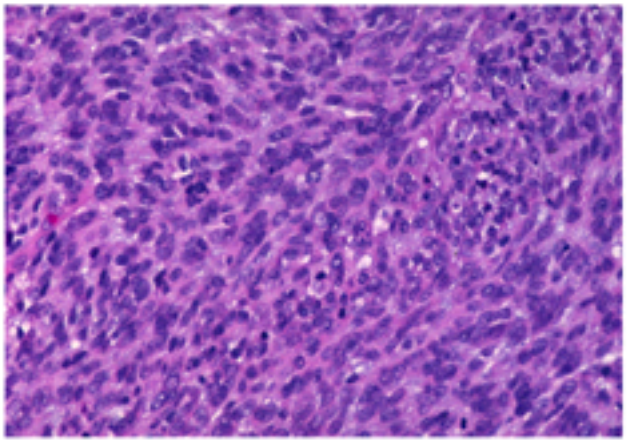
Web Table 4.
Comparison of classification of genes by hierarchical clustering, SVD and SAM.

Web Table 5.
List of misplaced genes in raw dataset and final dataset, obtained after SVD.

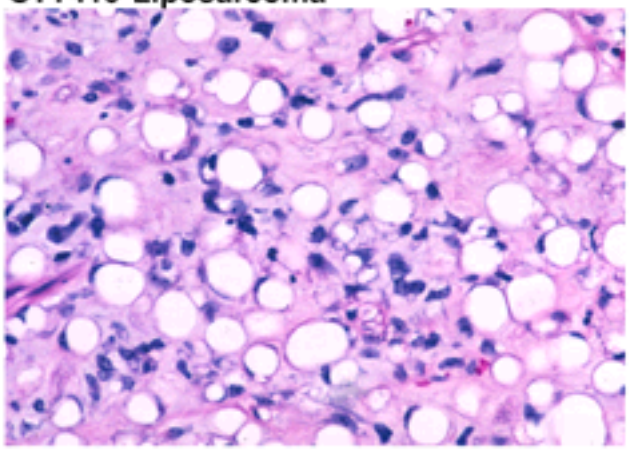
STT094- Gastrointestinal stromal tumor



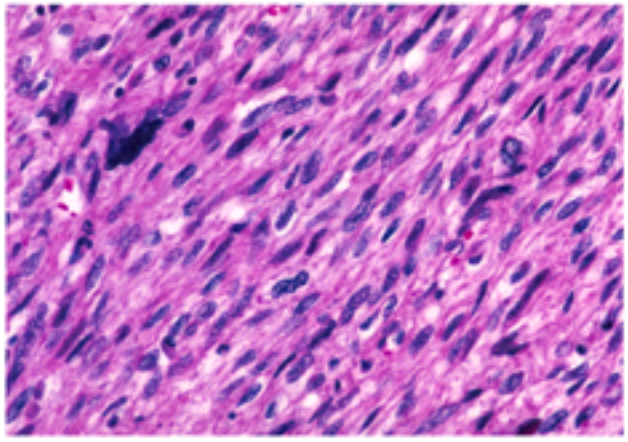
STT535-Synovial sarcoma



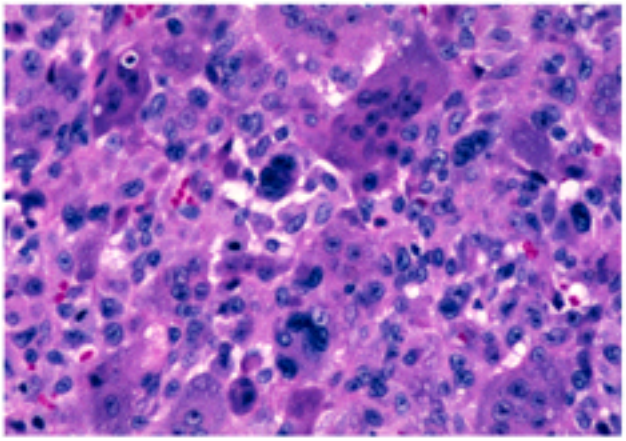
STT419-Liposarcoma



STT516-Leiomyosarcoma



STT709-Malignant fibrous histiocytoma



STT524-Schwannoma

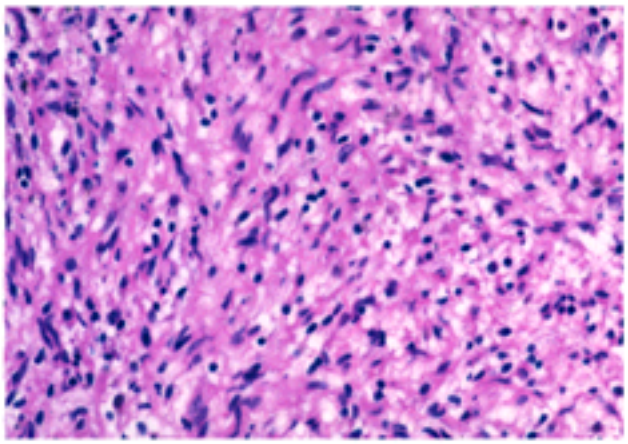
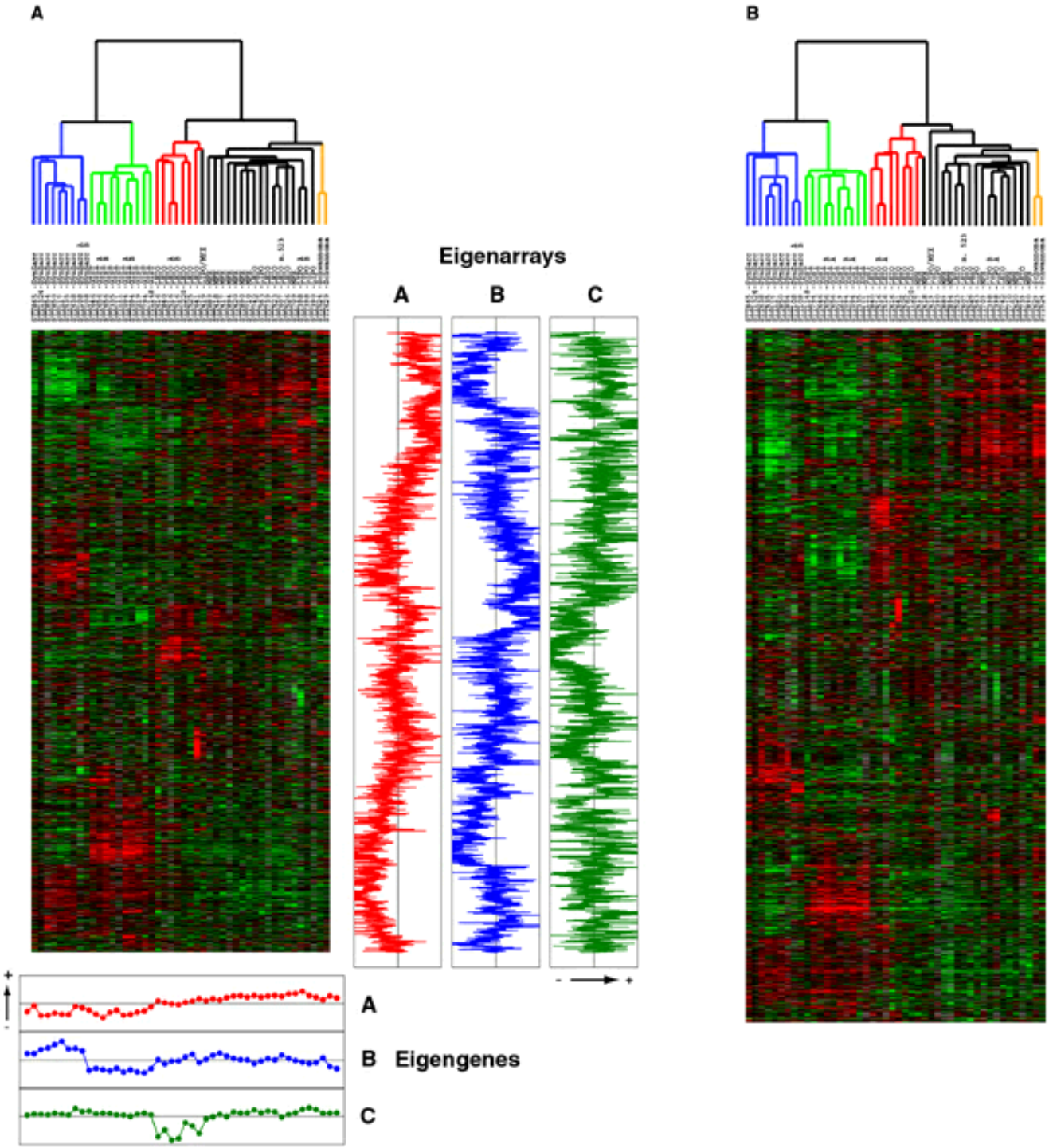


Figure 1

FIGURE 2



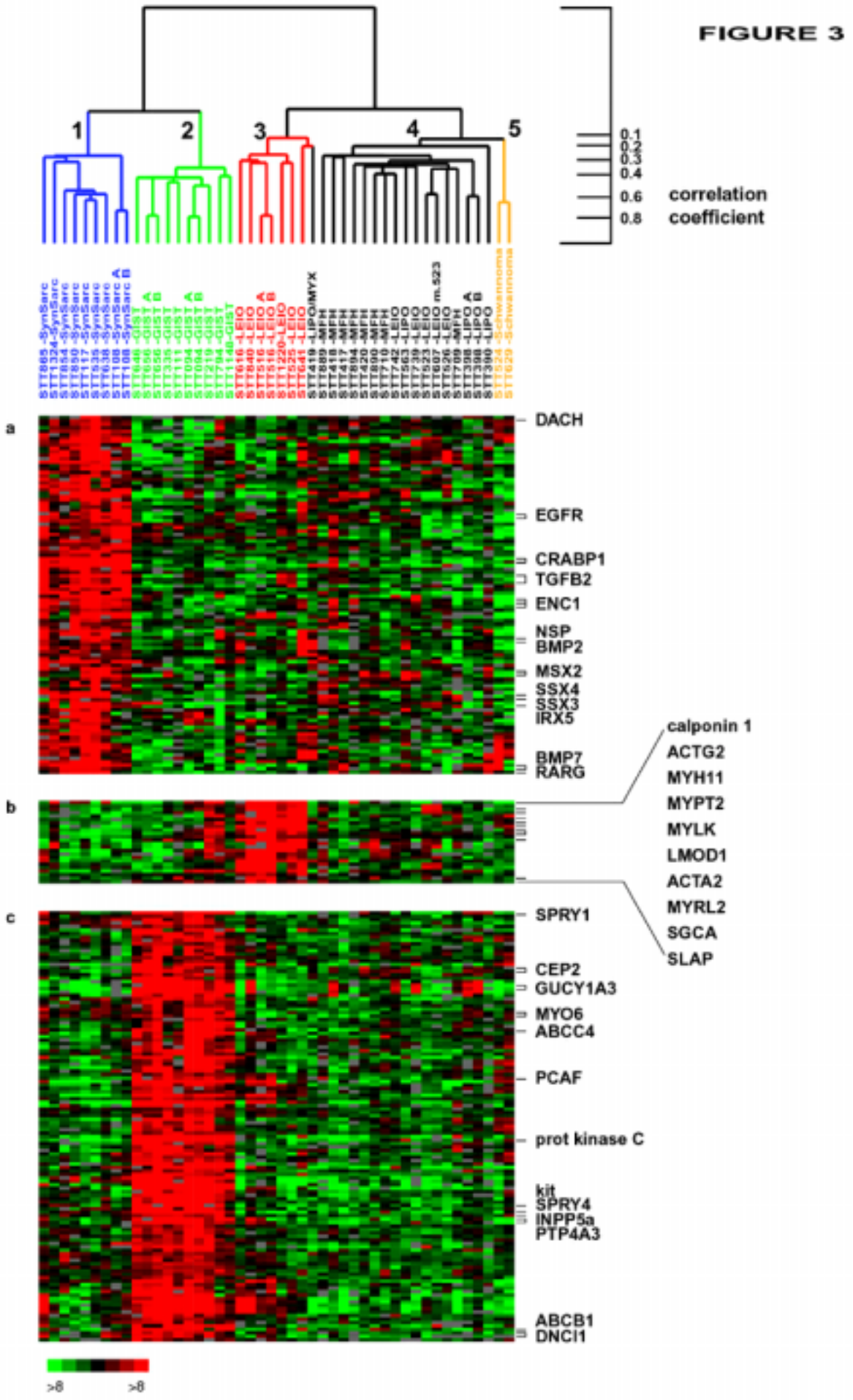
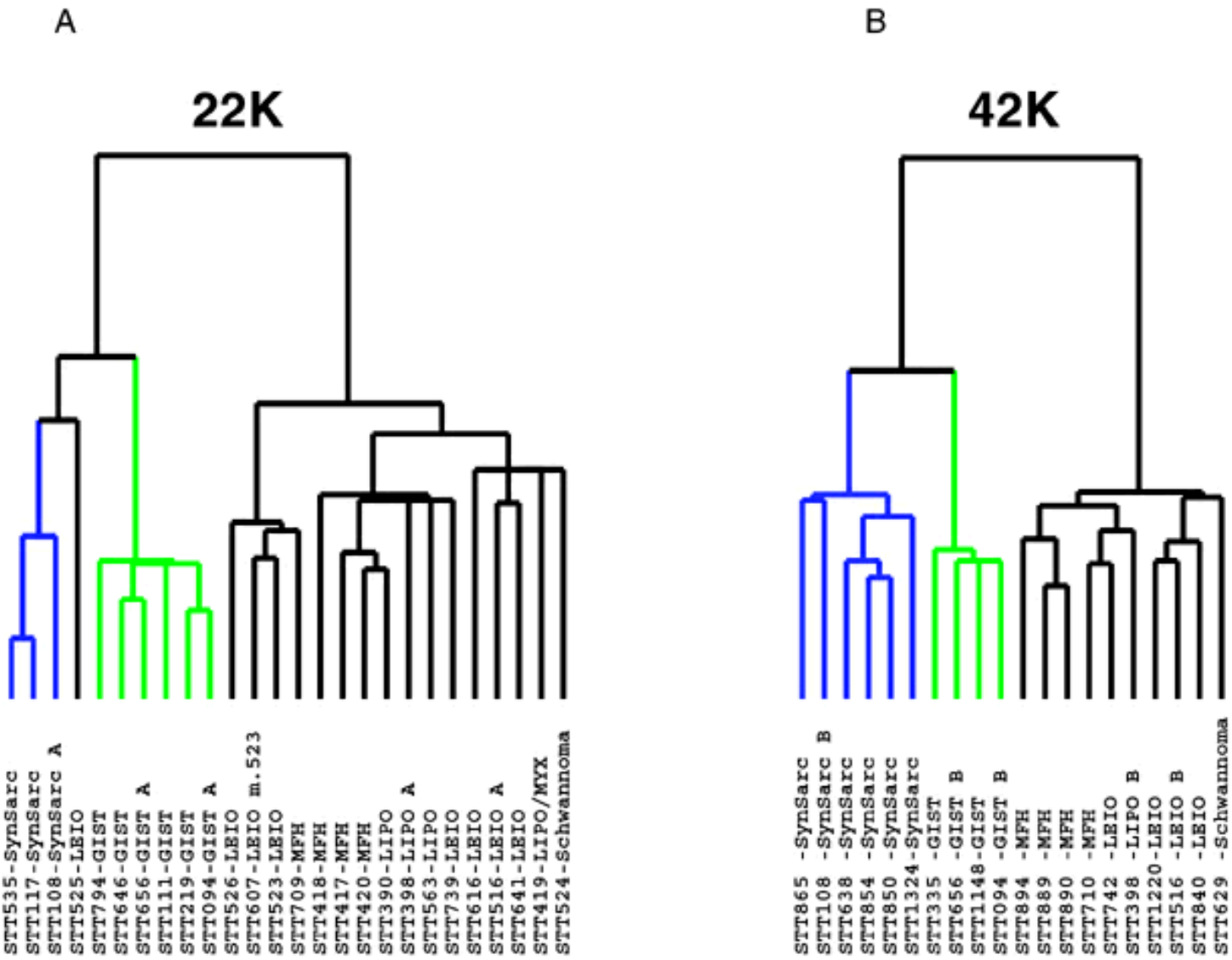
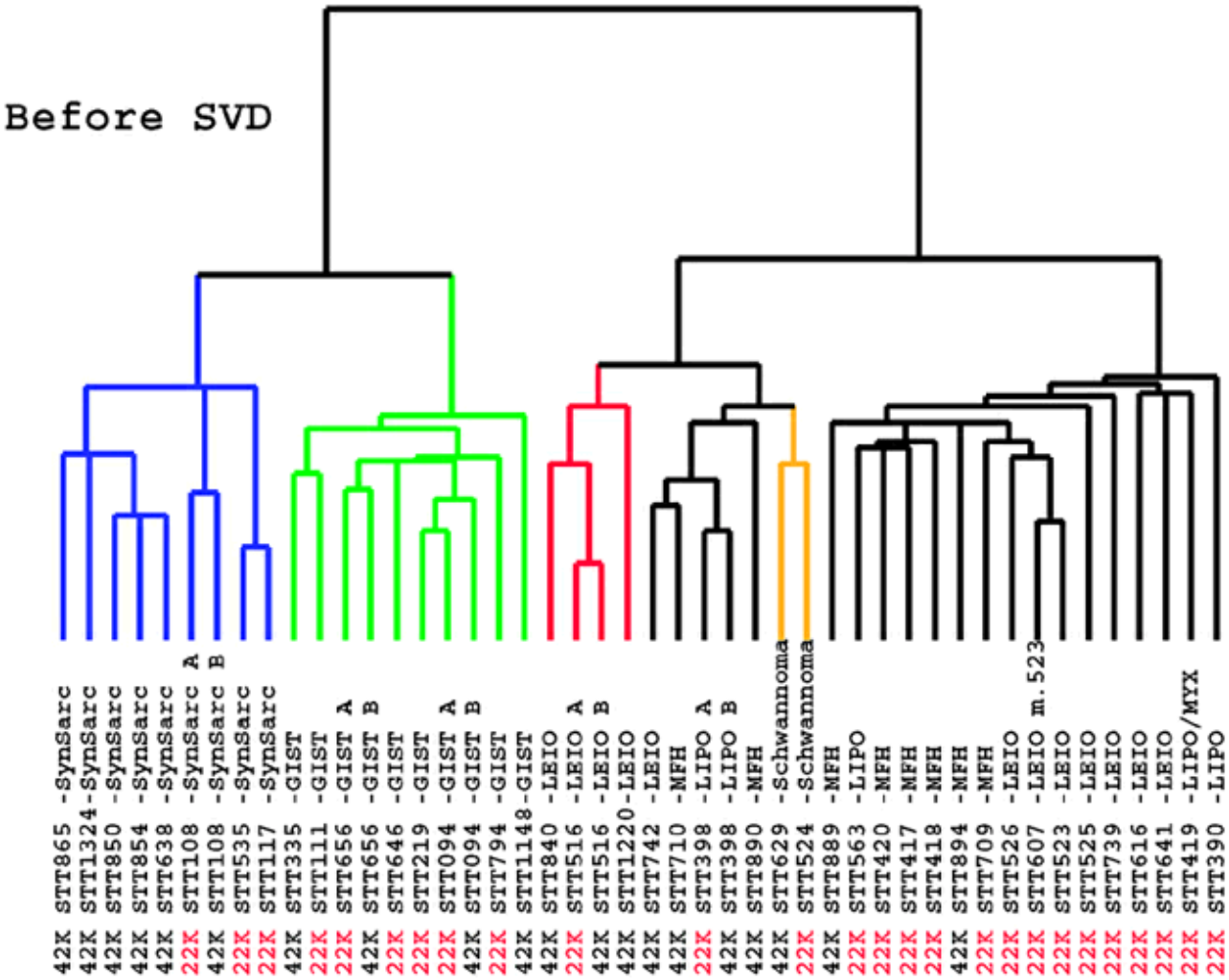


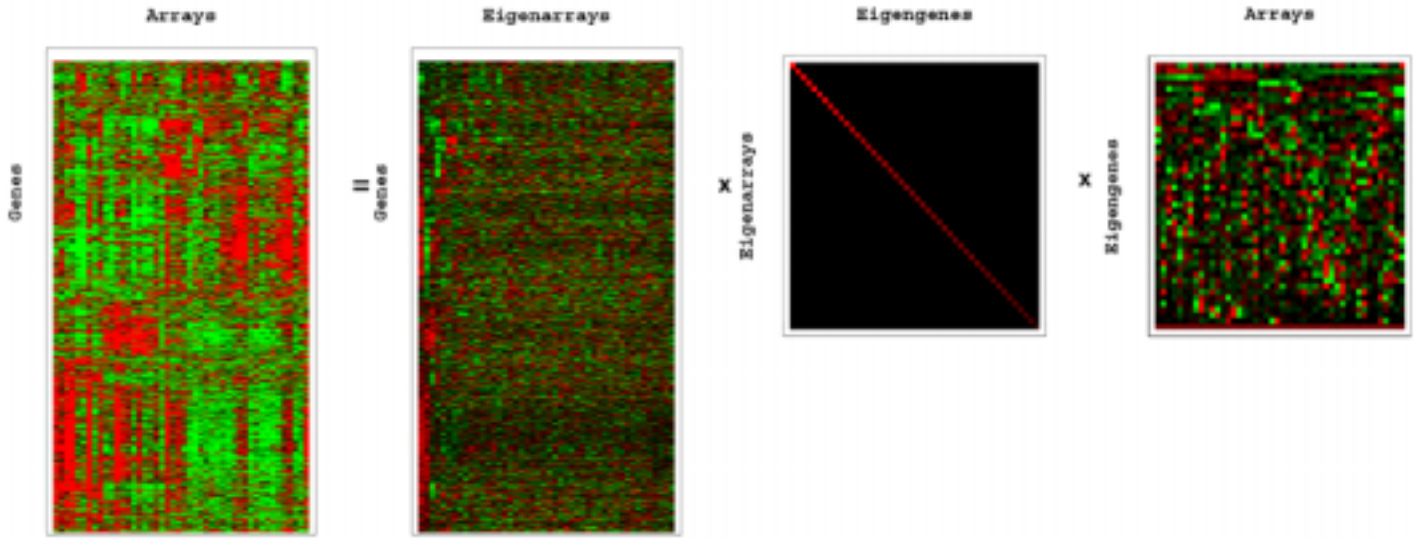
FIGURE 3
1
2
3
4
5
0.1
0.2
0.3
0.4
0.6
0.8
correlation
coefficient
a
DACH
EGFR
CRABP1
TGFB2
ENC1
NSP
BMP2
MSX2
SSX4
SSX3
IRX5
BMP7
RARG
b
calponin 1
ACTG2
MYH11
MYPT2
MYLK
LMOD1
ACTA2
MYRL2
SGCA
SLAP
c
SPRY1
CEP2
GUCY1A3
MYO6
ABCC4
PCAF
prot kinase C
kit
SPRY4
INPP5a
PTP4A3
ABCB1
DNCI1
>6
>6


A
22K
STT535-SynSarc
STT117-SynSarc
STT108-SynSarc A
STT525-LEIO
STT794-GIST
STT646-GIST
STT656-GIST A
STT111-GIST
STT219-GIST
STT094-GIST A
STT526-LEIO
STT607-LEIO m.523
STT523-LEIO
STT709-MFH
STT418-MFH
STT417-MFH
STT420-MFH
STT390-LIPO
STT398-LIPO A
STT563-LIPO
STT739-LEIO
STT616-LEIO
STT516-LEIO A
STT641-LEIO
STT419-LIPO/MYX
STT524-Schwannoma
B
42K
STT865 -SynSarc
STT108 -SynSarc B
STT638 -SynSarc
STT854 -SynSarc
STT850 -SynSarc
STT1324-SynSarc
STT335 -GIST
STT656 -GIST B
STT1148-GIST
STT094 -GIST B
STT894 -MFH
STT889 -MFH
STT890 -MFH
STT710 -MFH
STT742 -LEIO
STT398 -LIPO B
STT1220-LEIO
STT516 -LEIO B
STT840 -LEIO
STT629 -Schwannoma

Webfigure 1

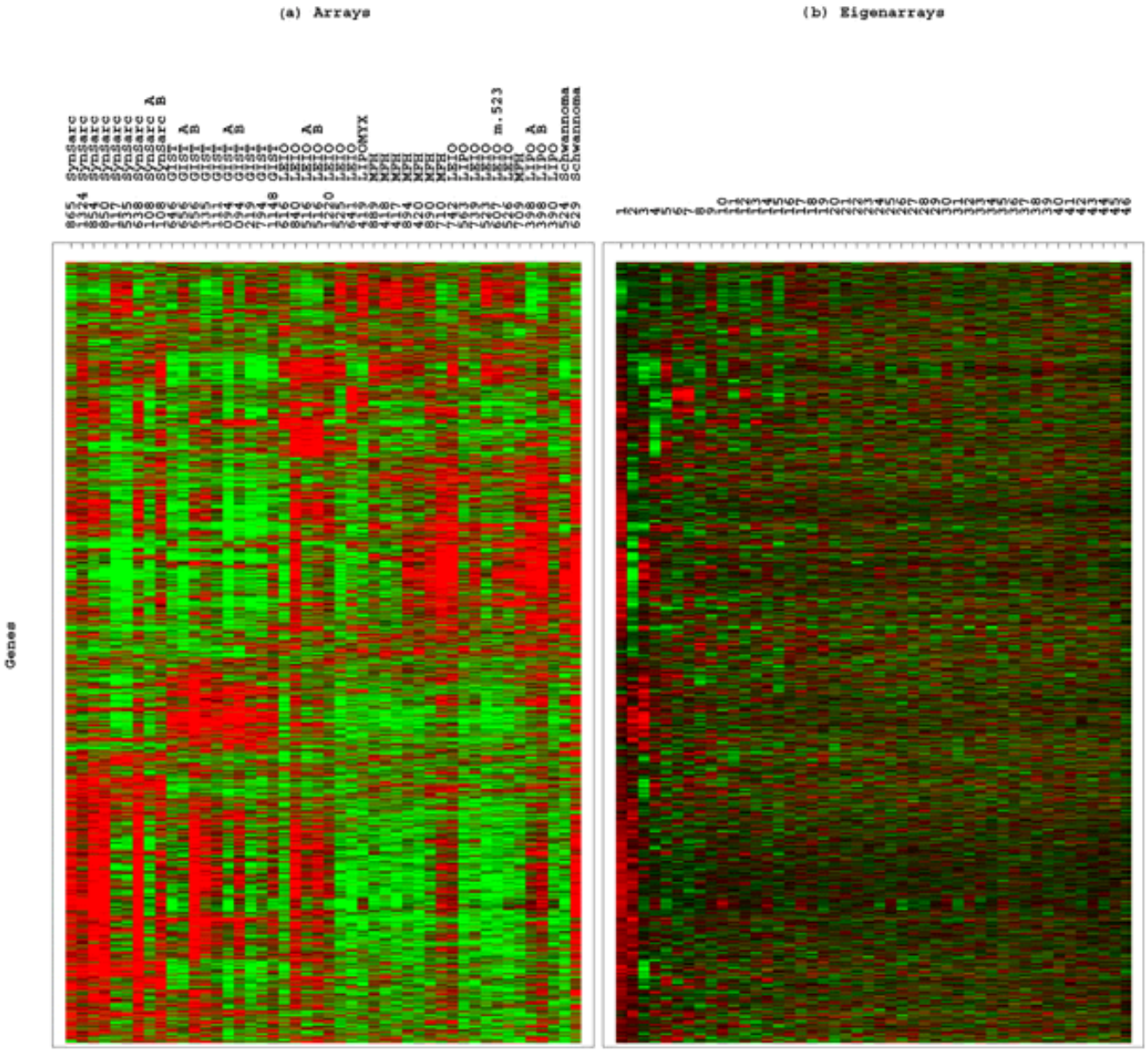

Before SVD
42K STT865 -SynSarc
42K STT1324-SynSarc
42K STT850 -SynSarc
42K STT854 -SynSarc
42K STT638 -SynSarc
22K STT108 -SynSarc A
42K STT108 -SynSarc B
22K STT535 -SynSarc
22K STT117 -SynSarc
42K STT335 -GIST
22K STT111 -GIST
22K STT656 -GIST A
42K STT656 -GIST B
22K STT646 -GIST
22K STT219 -GIST
22K STT094 -GIST A
42K STT094 -GIST B
22K STT794 -GIST
42K STT1148-GIST
42K STT840 -LEIO
22K STT516 -LEIO A
42K STT516 -LEIO B
42K STT1220-LEIO
42K STT742 -LEIO
42K STT710 -MFH
22K STT398 -LIPO A
42K STT398 -LIPO B
42K STT890 -MFH
42K STT629 -Schwannoma
22K STT524 -Schwannoma
42K STT889 -MFH
22K STT563 -LIPO
22K STT420 -MFH
22K STT417 -MFH
22K STT418 -MFH
42K STT894 -MFH
22K STT709 -MFH
22K STT526 -LEIO
22K STT607 -LEIO m.523
22K STT523 -LEIO
22K STT525 -LEIO
22K STT739 -LEIO
22K STT616 -LEIO
22K STT641 -LEIO
22K STT419 -LIPO/MYX
22K STT390 -LIPO

Webfigure 2

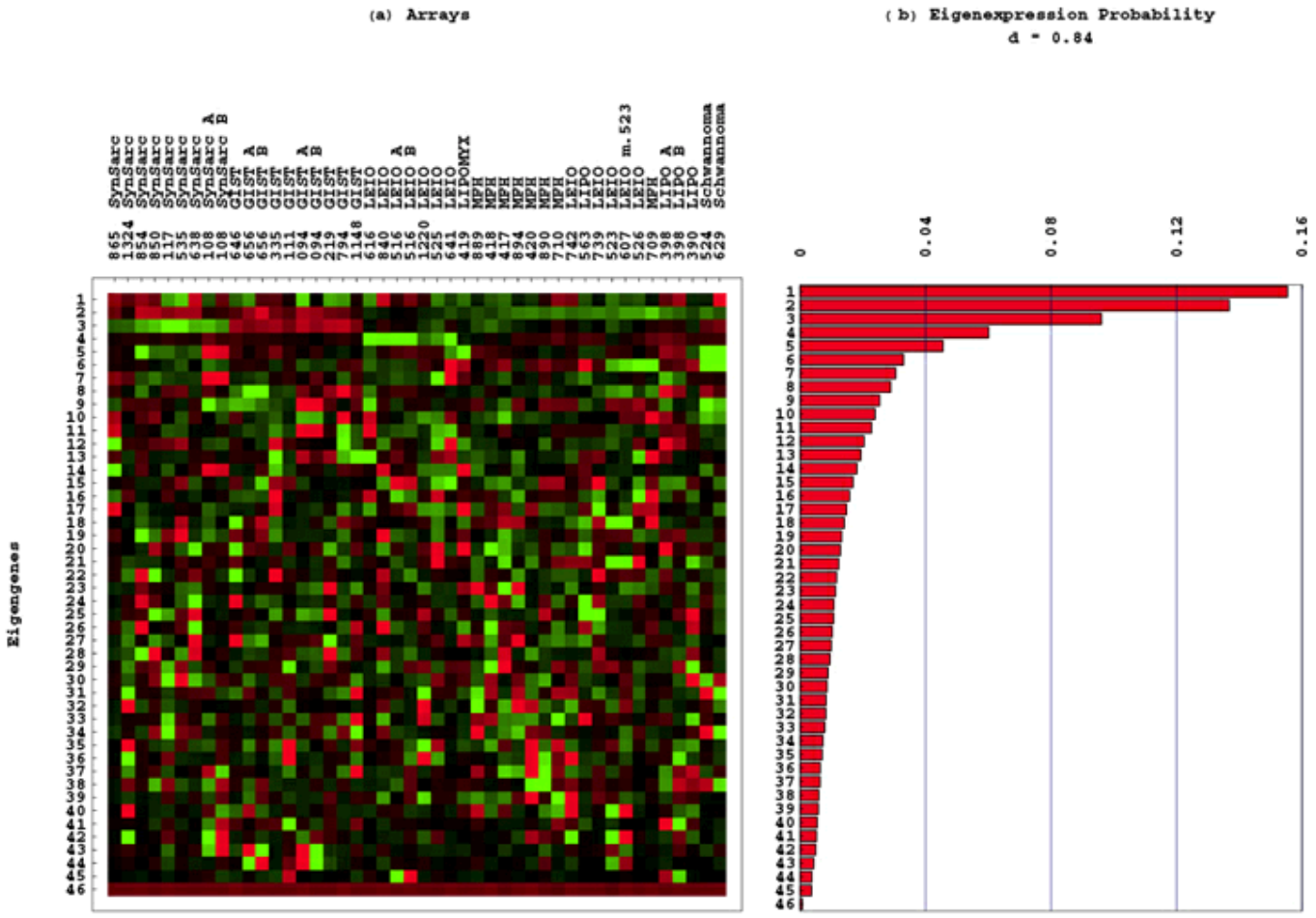


Webfigure 3a

http://genome-www.stanford.edu/sarcoma/Webfigure3b.jpg

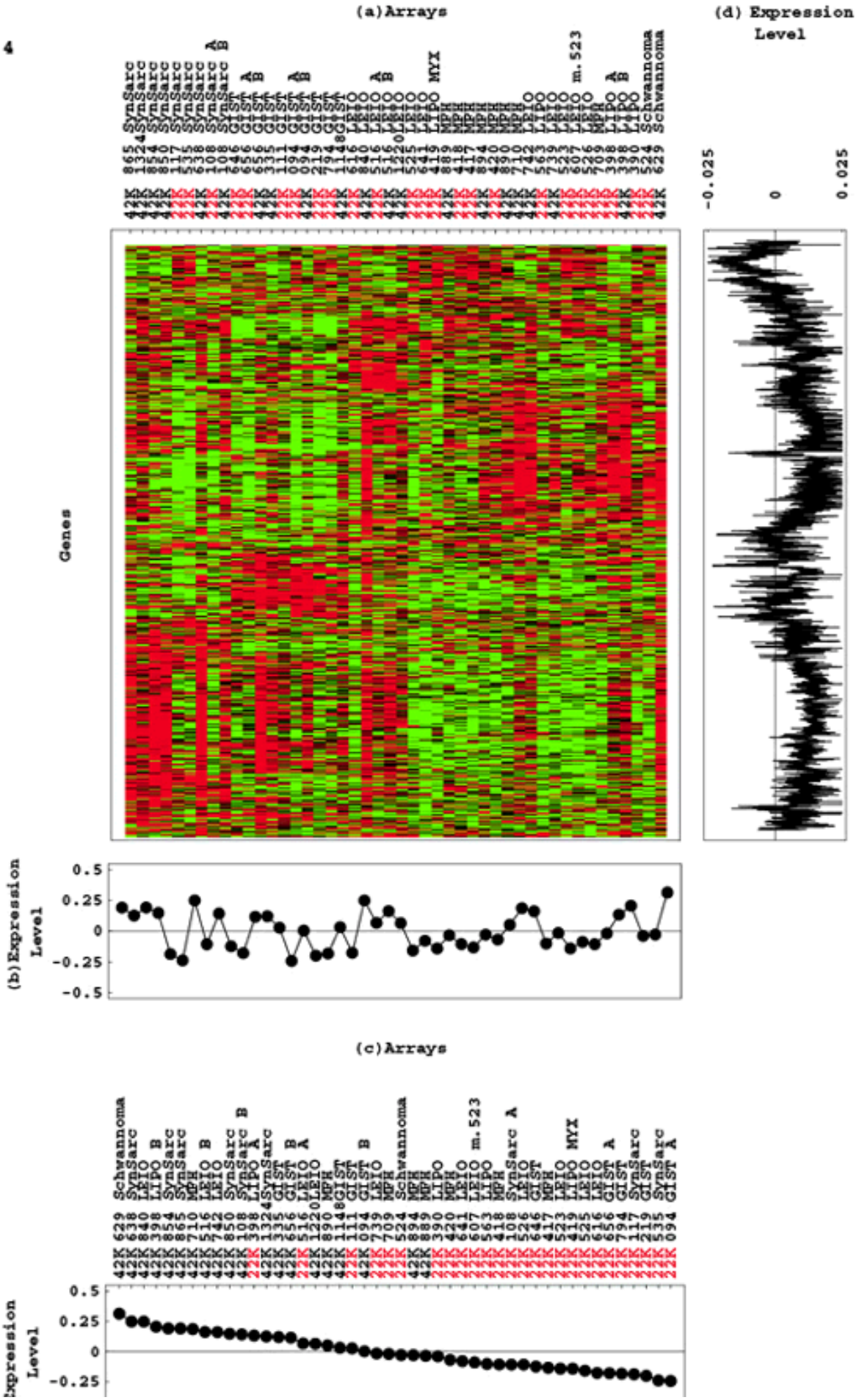


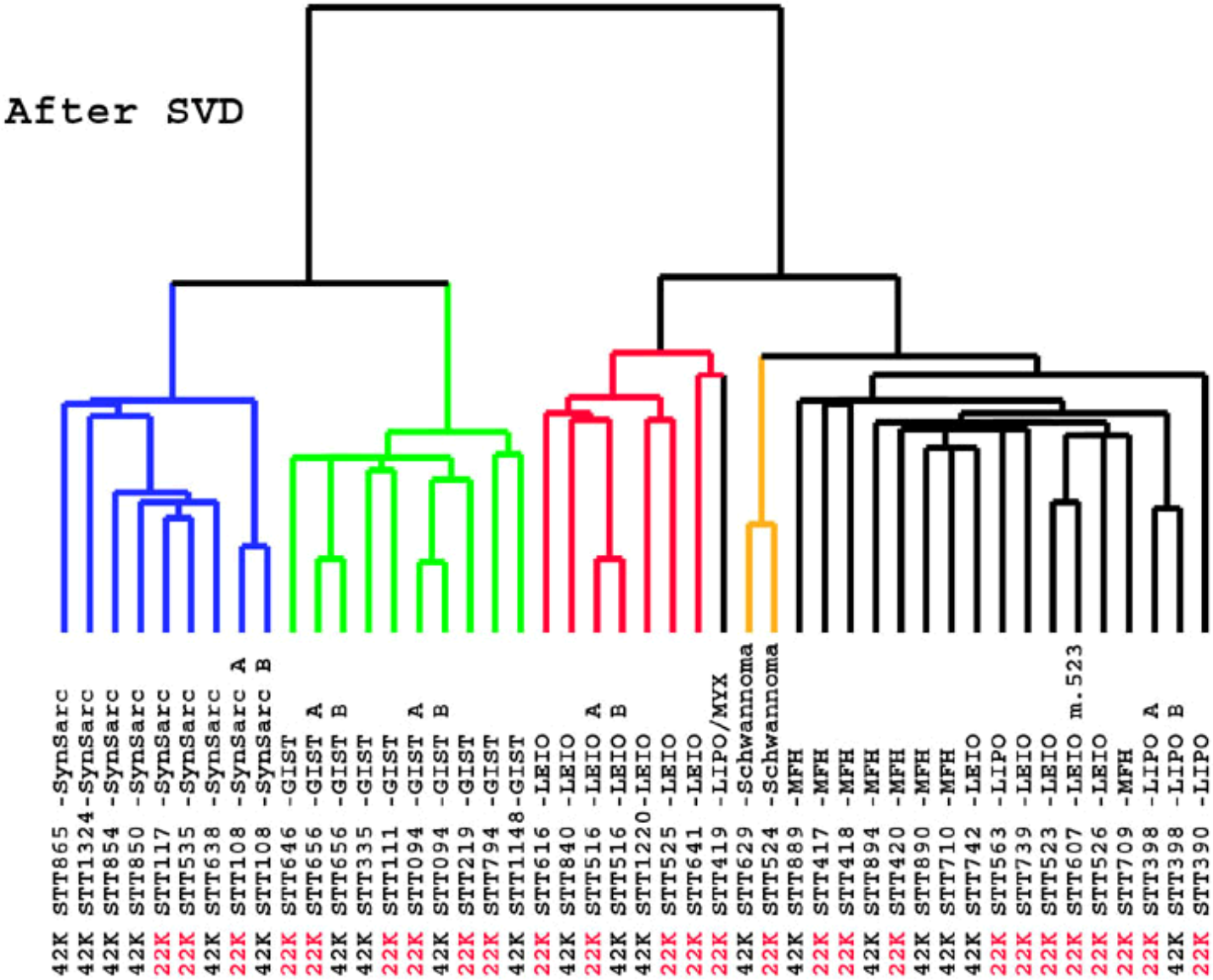
Webfigure 3b



Webfigure 3c

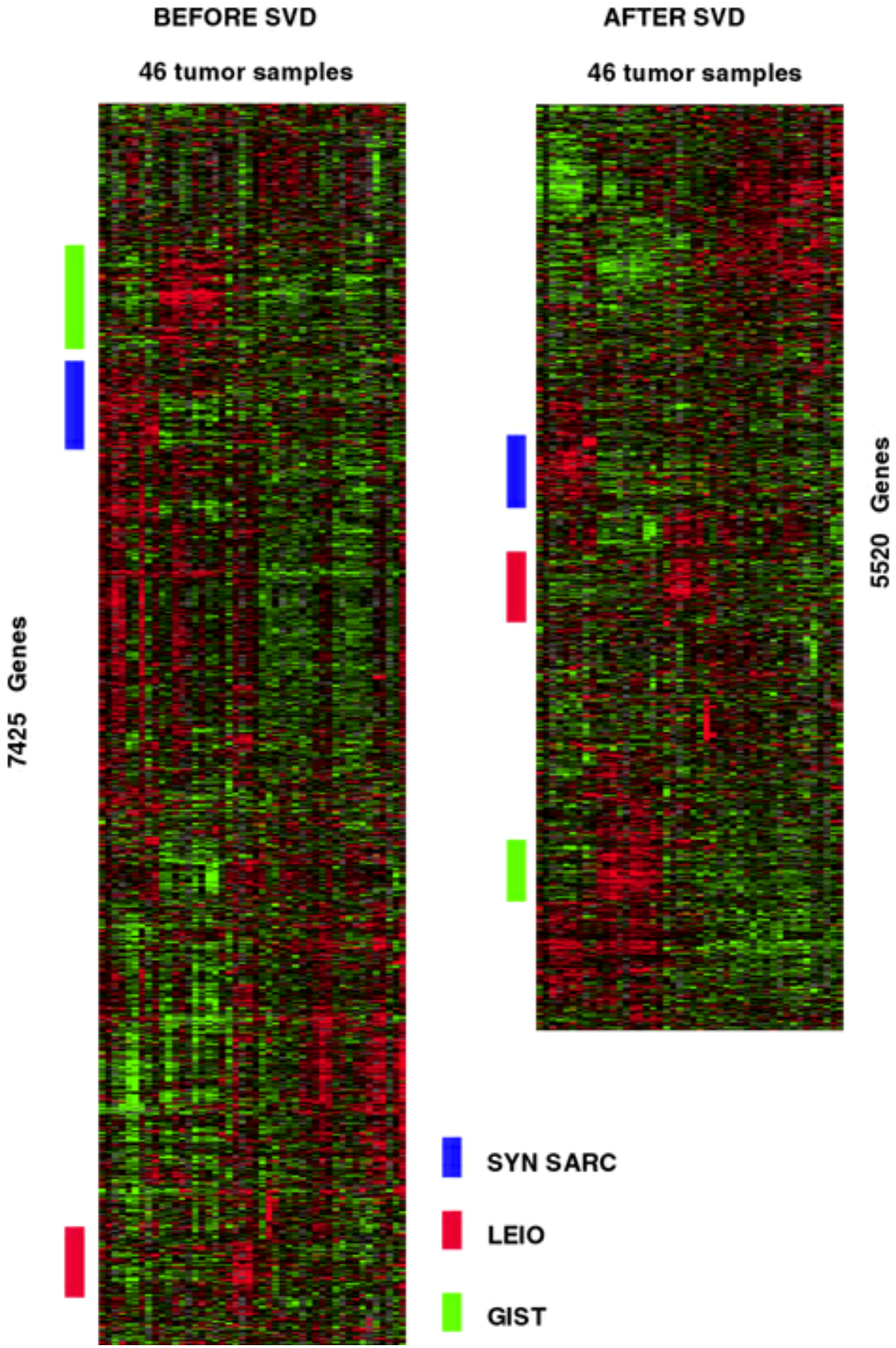
WEBFIGURE 4


(a) Arrays
(d) Expression Level
Genes
(b) Expression Level
(c) Arrays
Expression Level



Webfigure 5a

http://genome-www.stanford.edu/sarcoma/
webfigure5b.jpg



Web Figure 6