

Supplementary Information for Nature's paper:

PREDICTION OF CENTRAL NERVOUS SYSTEM EMBRYONAL TUMOUR OUTCOME BASED ON GENE EXPRESSION

Scott L. Pomeroy, Pablo Tamayo, Michelle Gaasenbeek, Lisa M. Sturla, Michael Angelo, Margaret E. McLaughlin, John Y.H. Kim, Liliana C. Goumnerova, Peter McL. Black, Ching Lau, Jeffrey C. Allen, David Zagzag, James M. Olson, Tom Curran, Cynthia Wetmore, Jaclyn A. Biegel, Tomaso Poggio, Shayan Mukherjee, Ryan Rifkin, Andrea Califano, Gustavo Stolovitzky, David N. Louis, Jill P. Mesirov, Eric S. Lander and Todd R. Golub

www.genome.wi.mit.edu/MPR/CNS

November 14, 2001

Contents:

Section I: expanded methods	2
Patient data and tumor bank	2
Microarray hybridization	2
Preprocessing and re-scaling	3
Clustering	3
Supervised learning	4
Gene marker selection	5
Permutation-based neighborhood analysis for marker gene selection and screening	6
Permutation Test for Outcome Predictor	12
Algorithms	13
Proportional chance criterion	14
Survival analysis and Kaplan-Meier plots	14
PCA and multidimensional-scaling of Brain tumor samples	14
Combined classifiers	15
Section II: datasets and clinical attributes	17
List of all samples	17
Dataset A, A1, A2 - multiple tumor samples	19
Dataset B - MD classic-desmoplastic	23
Dataset C - MD outcome	24
Section III: detailed analysis results	26
Multiple tumor PCA	26
Multiple tumor class markers	42
Multiple tumor clustering	56
Multiple tumor classes predictions (k -NN)	57
Classic vs. desmoplastic MD markers	59
Classic vs. desmoplastic MD prediction results (k -NN)	69
SOM clustering of treatment outcome samples	71
SOM-discovered C0 vs. C1 class gene markers	74
Treatment outcome markers	75
k -nearest neighbors treatment outcome prediction results	79
Permutation test for k -nearest neighbor outcome predictor	81
Weighted voting treatment outcome prediction results	82
SVM treatment outcome prediction results	84
SPLASH treatment outcome prediction results	86
TrkC treatment outcome prediction results	88
Staging treatment outcome prediction results	90
Combined treatment outcome predictors	92
Summary of medulloblastoma treatment outcome predictions	96
Improvements of multi-gene prediction algorithm (k -NN) over staging and TrkC	97
k -NN predictions in subgroup treated with vincristine, cisplatin and cytoxan	98
Comparison between signal-to-noise and t-test statistic metrics	99
References	101

Section I: expanded methods

This document provides supplementary and detailed analysis information not included in the paper. Other sources of information and the original datasets can be found in our web site www.genome.wi.mit.edu/MPR/CNS.

Patient data and tumor bank

The complete cohort for these studies consists of 68 children with medulloblastomas, 10 young adults with malignant gliomas (WHO grades III and IV), 5 children with AT/RT, 5 with renal/extrarenal rhabdoid tumors, and 8 children with supratentorial PNETs. A summary of the clinical data for the patients can be found in the [List of all samples](#) section of the document. All patients with medulloblastomas were treated with craniospinal irradiation to 2400 - 3600 centigray (cGy) with a tumor dose of 5300 - 7200 cGy. All patients with medulloblastomas were treated with chemotherapy consisting of cisplatin and vincristine, and combinations of carboplatin, etoposide, cyclophosphamide, procarbazine, lomustine (CCNU). Two patients received high dose chemotherapy at relapse, including methotrexate and thiopeta, followed by autologous bone marrow transplantation. Thirty-five of the children with medulloblastomas were part of a cohort described in previous publications (Segal et al 1994, Kim et al 1999). All tumor samples were obtained at the time of initial surgery prior to treatment. The samples were snap frozen in liquid nitrogen and stored at -80°C. The studies were done with approval of the Committee for Clinical Investigation of Boston Children's Hospital. The data were organized into three sets: Dataset A (42 samples containing: 10 medulloblastomas, 10 malignant gliomas, 5 AT/RT and 5 renal/extrarenal rhabdoid tumors, 8 supratentorial PNETs and 4 normal cerebella), Dataset B (34 samples, containing 9 desmoplastic medulloblastoma and 25 classic medulloblastoma), and Dataset C (60 samples, containing 39 medulloblastoma survivors and 21 treatment failures). There are two additional variants of Dataset A called A1 and A2 described in the second section of this document. A description of each dataset is available in the Datasets and clinical attributes section of this document.

Microarray hybridization

For a detailed protocol, see <http://www.genome.mit.edu/MPR/CNS>. Briefly, tissue samples were homogenized (Polytron, Kinematica, Lucerne) in guanidinium isothiocyanate and RNA was isolated by centrifugation over a CsCl gradient. RNA integrity was assessed either by northern blotting (Kim et al 1999) or by gel electrophoresis. The amount of starting total RNA for each reaction varied between 10 and 12 µg. First strand cDNA synthesis was generated using a T7-linked oligo-dT primer, followed by second strand synthesis. An in vitro transcription reaction was done to generate the cRNA containing biotinylated UTP and CTP, which was subsequently chemically fragmented at 95°C for 35 minutes. Ten micrograms of the fragmented, biotinylated cRNA was hybridized in MES buffer (2-[N-Morpholino]ethanesulfonic acid) containing 0.5 mg/ml acetylated bovine serum albumin (Sigma, St. Louis) to Affymetrix (Santa Clara, CA) HuGeneFL arrays at 45°C for 16 hours. HuGeneFL arrays contain 5920 known genes and 897 expressed sequence tags. Arrays were washed and stained with streptavidin-

phycoerythrin (SAPE, Molecular Probes). Signal amplification was performed using a biotinylated anti-streptavidin antibody (Vector Laboratories, Burlingame, CA) at 3 $\mu\text{g/ml}$. This was followed by a second staining with SAPE. Normal goat IgG (2 mg/ml) was used as a blocking agent. Scans were performed on Affymetrix scanners and the expression value for each gene was calculated using Affymetrix GENECHIP software. Minor differences in microarray intensity were corrected using a linear scaling method as detailed in the next section.

Preprocessing and re-scaling

The raw expression data as obtained from Affymetrix's GeneChip is re-scaled to account for different chip intensities. Each column (sample) in the dataset was multiplied by $1/\text{slope}$ of a least squares linear fit of the sample vs. the reference (the first sample in the dataset). This linear fit is done using only genes that have 'Present' calls in both the sample being re-scaled and the reference. The sample chosen as reference is a typical one (i.e. one with the number of "P" calls closer to the average over all samples in the dataset). Scans were rejected if the scaling factor exceeded a factor of 3, fewer than 1000 genes received 'Present' calls, or microarray artifacts were visible.

A ceiling of 16,000 units was chosen for all experiments because it is at this level that we observe fluorescence saturation of the scanner; values above this cannot be reliably measured. For classification problems that are very robust (e.g. distinguishing different types of brain tumors), we used a threshold of 100 units because there was a sufficiently large number of genes correlated with the distinction that the threshold could be set high, thereby minimizing noise, and maximizing potential biological interpretation of the marker genes. For the more subtle distinctions (e.g. outcome prediction), few correlates of the distinction are found, and for this reason the threshold was set at a lower level (20 units) so as to avoid missing any potentially informative marker genes.

These numbers are Affymetrix's scanner "average difference" units. After this preprocessing gene expression values were subjected to a variation filter which excluded genes showing minimal variation across the samples being analyzed. The variation filter tests for a fold-change and absolute variation over samples (comparing max/min and max-min with predefined values and excluding genes not obeying both conditions). The precise parameters of the variation filters for each dataset are provided in each analysis section of this document. Different thresholds and variation filters were used according to the purpose of the analysis (e.g. select weak marker genes for treatment outcome, strong robust marker genes for morphology, highly varying genes for PCA etc.). For example, if the maximum and minimum values of a gene across samples were max and min then the variation filter excluded those where $\text{max}/\text{min} < 5$ and $\text{max} - \text{min} < 500$. In some cases more or less stringent values were used.

Clustering.

Self Organizing Maps were performed using our GeneCluster clustering package available at www.genome.wi.mit.edu/MPR/Software. *Self-Organizing Maps* (SOMs). The Self Organizing Map is a method for performing unsupervised learning (i.e., learning models for classifying data where the true class for the data

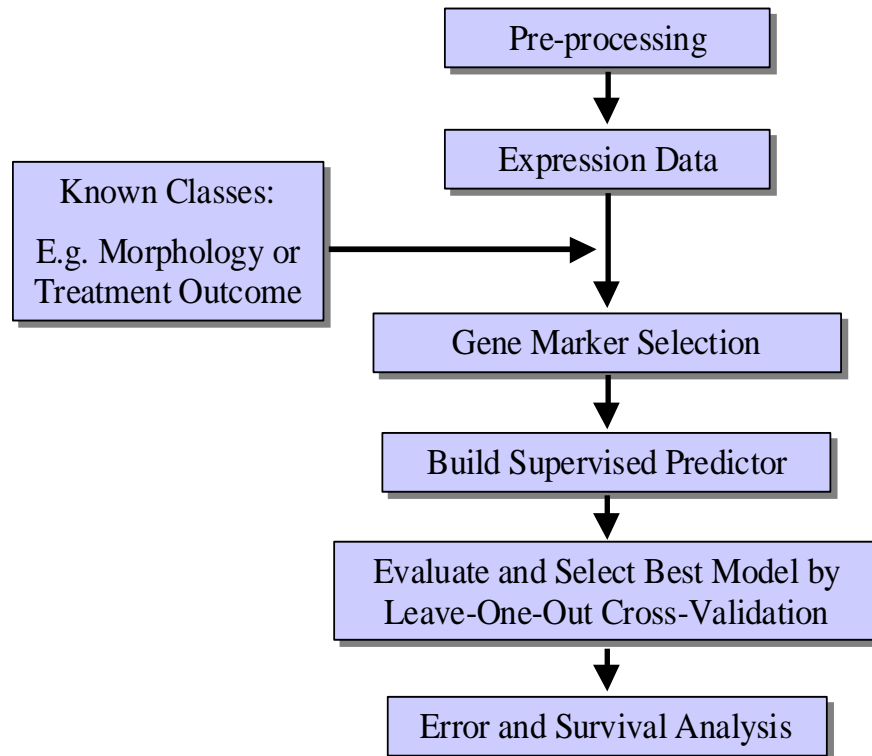
samples is assumed to be unknown prior to model training) where a grid of 2D nodes (clusters) is iteratively adjusted to reflect the global structure in the expression dataset (Tamayo et al 1999). In general, unsupervised learning presents a more difficult problem than supervised learning methods (such as weighted voting or k-NN) but is useful for discovering new classes during exploratory analysis. With the SOM, one randomly chooses the geometry of the grid (e.g., a 3 x 2 grid) and maps it into the k-dimensional feature space. Initially the features are randomly mapped to the grid but during training the mapping is iteratively adjusted to reflect the data structure. The data were first normalized by standardizing each column (sample) to mean 0 and variance 1. The SOM results for the clustering of samples can be found in the [Multiple tumor clustering](#) for multiple tumor samples and in the [SOM clustering of treatment outcome samples](#) section for the clustering of medulloblastomas.

Hierarchical Clustering is another unsupervised learning method useful for dividing data into natural groups. Data is clustered hierarchically by organizing the data into a tree structure based upon the degree of similarity between features. We used the Cluster and TreeView software (Eisen et al 1998) to perform average linkage clustering, which organizes all of the data elements into a single tree with the highest levels of the tree representing the discovered classes. The detailed clustering results can be accessed in the [Multiple tumor clustering](#) section.

Supervised learning.

This is the methodology for building a supervised classifier that we followed.

- a) define a target class based on morphology, tumor class or treatment outcome clinical information;
- b) select the “marker” genes with the highest correlation with the target class using a class separation statistic (signal-to-noise ratio). A permutation test is also applied to the top ranked genes to assess their class-correlation statistical significance.
- c) build a classifier in cross-validation (leave-one-out) by removing one sample and then used the rest as a training set
- d) several models are built using different number of marker genes and the final chosen model is the one that minimizes the total error in cross-validation
- e) evaluate prediction results, compute confusion matrices and produce Kaplan-Meier survival plots.



This methodology was used with the following algorithms: k-nearest neighbors, weighted voting, support vector machines, SPLASH, metastatic staging, TrkC gene expression and two combined predictors. The details for each algorithm are described below.

Gene marker selection

Genes correlated with a particular class distinctions (e.g. class 0 and class 1) were identified by sorting all of the genes on the array according to the signal-to-noise statistic (Golub et al 1999, Slonim et al 2000) $(\mu_{\text{class } 0} - \mu_{\text{class } 1}) / (\sigma_{\text{class } 0} + \sigma_{\text{class } 1})$ where μ and σ represent the mean (or median) and standard deviation of expression, respectively, for each class. Permutation of the column (sample) labels was performed to compare these correlations to what would be expected by chance (see the next section). These marker genes were used to build the k-nearest neighbor and weighted voting classifiers. SVM and SPLASH use different methods to select marker genes.

In Section III we described marker genes for several classifications:

- multi tumor classes ([Multiple tumor class markers](#)).
- classic vs. desmoplastic medulloblastoma morphology ([Classic vs. desmoplastic MD markers](#))
- SOM-discovered medulloblastoma classes ([SOM-discovered C0 vs. C1 class gene markers](#)) and,
- medulloblastoma treatment outcome ([Treatment outcome markers](#)).

Permutation-based neighborhood analysis for marker gene selection and screening.

Before we describe the method in detail we provide some motivation for use of the technique and put it in context with other multiple comparison and permutation test approaches.

There are two interrelated problems that we have addressed with our permutation-based neighborhood analysis first introduced in Golub et al 1999 and Slonim et al 2000. One is the problem of feature (gene) selection in terms of how many and which genes to input to a supervised learning classifier. This process is a necessary step in a supervised learning methodology as many classifier algorithms cannot deal with thousands of input variables and require some type of dimensionality reduction or prior selection. The other problem is to choose statistically significant molecular markers or differentially expressed genes that deserve more detailed biological study. For example, the ones that one may choose for further validation using a different technology or experimental technique (e.g. RT-PCR, immunochemistry, etc.).

It is important to point out that in these two problems one is basically interested in selecting the subset of genes more likely to be useful in discriminating the phenotype of interest either as single markers or in combination with others. In other words we are interested in a ranking and screening process that identifies enough of the relevant features. One can easily tolerate some amount of false positive errors in exchange for higher sensitivity. Most of the current molecular classification problems of interest, such as morphological or lineage distinction, treatment outcome prediction, drug resistance etc. fit this scenario. These problems involve the presence of a potentially weak signal, for example a few marker genes in a background of technical variation and noise, and therefore favor marker selection methods that are very sensitive and have enough statistical power to produce non-empty results. One can tolerate some number of false positive errors because the selected marker genes are usually weighted or further selected by the classifier. This is also the case in determining biological significance/relevance; in any serious follow up study the markers would have to be further validated. This need for higher sensitivity and adaptability to the dataset being analyzed is one of the main motivations behind our approach as originally applied to Leukemia subtype distinction in Golub et al 1999 and in its current form in this paper.

Marker gene selection, with the characteristics described in the previous paragraph, can be seen as an example of a Multiple Comparison Procedure (MCP) where multiple hypotheses (genes) are tested simultaneously and then accepted or rejected according to a testing procedure. For recent reviews on MCP see Hochberg and Tamhane 1997, Bender and Lange 2001, and the special issue on Multiple Comparisons of the *Journal of Statistical Planning and Inference* (vol. 82, 1999). In our case, each statistical test of a gene can be seen as testing a null hypothesis on the equivalence of the phenotype classes. In this way rejecting a subset of the hypotheses corresponds to selecting a set of statistically significant differentially expressed genes at a given significance level. A global null hypothesis would assert that no gene expression changes are significant and therefore that there is no significant difference between the biological classes as

measured by the entire set of microarrays. Notice that this situation of all null hypotheses being true is not likely to be realistic because in practice most microarrays experiments are done with biological classes with known differences that are usually reflected in multiple genes. A more realistic situation is one in which a subset of the hypotheses are false corresponds to the usual problem of selecting between a few dozens and few hundred genes. Traditional approaches to the MCP assume that all, or almost all, of the null hypotheses are true. They also control for the Family Wise Error Rate (FWER), i.e., the probability that exactly one, or at least one, type I (false positive) error occurs. In the marker selection problem this would be the case where there was only one wrong marker gene in the determined marker set. However, the models we construct are not really sensitive to a small number of false positives in the selected marker set. Thus, controlling the FWER is an overly conservative approach that does not provide enough statistical power for the purposes of marker selection and may actually yield no candidate marker genes. This situation of partial rejection is actually quite common in exploratory data analysis and in recent years alternative less conservative and more sensitive formulations of the MCP have been introduced. These methods control the False Discovery Rate (FDR) rather than the FWER (Benjamini and Hochberg 1995). The FDR is the total number of type I, or false positive errors, that are made by the MCP. Controlling for this quantity moves the MCP closer to the type of approaches used in machine learning feature selection and leads to methods with higher statistical power. Statistically the FDR is a compromise between an ultra conservative correction a la Bonferroni and making no correction at all (Benjamini et al 2001). This type of approach is clearly more appropriate for gene selection.

Regardless of the assumption on the number of true hypotheses, or the emphasis on FWER or FDR, the real problem in multiple comparisons is that the hypotheses (genes) are correlated in complex ways reflecting the structure of genetic pathways and interactions. This makes the dependence structure of the data quite difficult to analyze or capture in a test. Traditional corrections, such as Bonferroni, are too conservative and produce essentially no marker genes except for cases where the differences are overwhelming (e.g. dead tissue vs. live tissue). Less conservative approaches attempt to solve this problem by using close testing step-wise methods where the hypotheses can be tested in a specific order and decisions made in a step-wise manner. Decisions on earlier hypotheses may affect later decisions and in this way the dependent structure can be taken into account (see for example Tamhane 1996, Tamhane and Dunnett 1999, Somerville 1999, Troendle 2000). On a parallel track, MCP methods to increase the statistical power by resampling (bootstrap) have also been introduced (Westfall and Young 1993). These methods control the FWER but resample the empirical null distribution to provide less conservative corrections for the p-values. Some of these methods are included in the PROC MULTTEST procedure in SAS (Westfall and Wolfinger 1999).

The comparison of FWER vs. FDR, step-wise vs. single step, resampling vs. analytical p-value adjustments, and in general the assessment of the virtues or applicability of different MCP methods, continues today. It has generated a healthy debate in the statistics community (see for Benjamini et al 1999, Bender and Lang 2001). Another perspective on the MCP can be obtained by Bayesian methods (See Berry and Hochberg 1999 for a review).

From the perspective of machine learning and pattern recognition, the problem of optimal feature selection is intractable and one has to be content with empirical approximations that may have to be tailored to fit the application (Duda, Hart and Stork 2001). Two common approaches are based on the use of filters and wrappers (Kohavi and John 1998). Filter approaches select the best features using a score function that measures the discrimination power of the feature with respect with to the target in a way similar to the test statistic is used in MCP. Typical score functions are, for example, the mutual information, signal to noise ratios, Naïve Bayes posteriors, inner products, linear transformation (e.g. eigenvectors of the covariance matrix), or bounds on the Bayes error such as the Bhattacharyya distance. Wrapper methods involve the use of the actual classifier in the selection process and can be seen as non-linear optimization problems. For more details see Kohavi and John 1998, Cherkassky and Mulier 1998, Fukunaga 1990, Kearns and Vazirani 1997 and Duda, Hart and Stork 2001.

Our permutation based neighborhood analysis method is a direct attempt to solve the multiple hypothesis problem by comparing the actual distribution of markers (i.e. neighbors of an ideal marker separating the classes) with a reference empirical distribution obtained by permuting the phenotype class labels. It is based on a standard global permutation test (Fisher 1935, Lehman 1986, Good 1994) of the phenotype levels keeping the gene correlation information. A histogram of scores for each of the marker genes of each permutation (neighborhood) is kept and the significance of an actual gene marker is obtained by finding the appropriate percentile in the histogram of the correspondingly ranked marker (i.e. the one with the same rank, e.g. best match, second best match etc.). This empirical distribution-free method is simple, intuitive and adapts itself to the correlation structure of the data providing higher statistical power. It minimizes the total number of false positives and uses the empirical reference distribution in a similar way as FDR-based and resampling methods do. Recently general MCP methods have been proposed to combine both resampling and control of the FDR (Yekutieli and Benjamini 1999).

The application of permutation tests has also been introduced in the structural analysis of genetic linkage and detection of QTL (Quantitative Trait Linkage). In these methods (Churchill and Doerge 1994, Doerge and Churchill 1996) the traits are randomly permuted to create data sets that have random genotype-phenotype association. Those methods and ours are conceptually quite similar although in our case we consider expression “functional” rather than genotype data.

After we introduced our method in Golub *et al* 1999 other methods have been introduced in the literature. For example the SAM method of Tusher *et al* 2001 is similar to ours but includes a user-adjustable threshold to provide estimates of the FDR. Dudoit *et al* 2001 have introduced a method based on step-down adjusted p-values using Westfall and Young’s approach in the context of replicated cDNA experiments. Ideker *et al* 2000 used generalized likelihood tests to assess the statistical significance of differentially expressed genes in the context of two channel cDNA microarrays. Newton *et al* 2001 and Baldi and Long 2001 use empirical Bayes hierarchical models to assess significance of differential expression. Lee *et al* 2000 combine the data from replicates to estimate posterior probabilities and identify

differentially expressed genes. No systematic comparison of the error rates and statistical power of all these different methods has been published yet. It will be interesting to develop a better understanding of the different trade offs between sensitivity and specificity, number of false positives vs. statistical power to guide the development of future analysis methodologies.

Description of the permutation test-based neighborhood analysis method.

Permutation test based (Golub et al 1999) neighborhood analysis is used to select and screen marker genes with respect to biologically meaningful phenotypes (morphology and treatment outcome) and to assess their statistical significance. To accomplish this we compare the top signal-to-noise scores of top marker genes with the corresponding ones from data obtained by randomly permuting the class labels. Typically 500 global random permutations were used to build histograms. Based on these histograms we determined the 50% (median), 5% and 1% significance levels and compared them with the values obtained for the real dataset. As described above this procedure is motivated by considering the following question: what is the likelihood that a given set of markers genes, for example selected by signal to noise, of a phenotype of interest represent chance correlations and not biologically significant matches? If one looks down the list of markers, how many should one consider as input to a classifier or for further study? In this list of selected markers what is the best way to minimize the number of false positives but retain enough sensitivity to select a non-empty set?

In detail the permutation test procedure for a given comparison of interest (e.g. markers high in class 0 and low in class 1) is as follows:

- Generate signal-to-noise $(\mu_{\text{class } 0} - \mu_{\text{class } 1})/(\sigma_{\text{class } 0} + \sigma_{\text{class } 1})$ scores for all genes that pass a variation filter using the actual class labels (phenotype) and sort them accordingly. The best match ($k=1$) is the gene “closer” or more correlated to the phenotype using the signal to noise as a correlation function. In fact one can imagine the reciprocal of the signal to noise as a “distance” between the “phenotype” and each gene as shown in the figure (see next page). One can also use a t -statistic $(\mu_{\text{class } 0} - \mu_{\text{class } 1})/\sqrt{(\sigma_{\text{class } 0}^2 + \sigma_{\text{class } 1}^2)}$ and obtain very similar results.
- Generate 500 or more random permutations of the class labels (phenotype). For each case of randomized class labels generate signal-to-noise scores and sort genes accordingly.
- Build a histogram of signal to noise scores for each value of k . For example one for all the 500 top markers ($k=1$), another one for the 500 second best ($k=2$) etc. These histograms represent a reference statistic for the best match, second best, etc. and, for a given value of k , different genes contribute to it. Notice that the correlation structure of the data is preserved by this procedure. For each value of k , determine different percentiles (1%, 5%, 50% etc.) of the corresponding histogram. (See the bottom diagrams in the figure.)
- Compare the actual signal to noise scores with the different significance levels obtained for the histograms of permuted class labels for each value of k . This

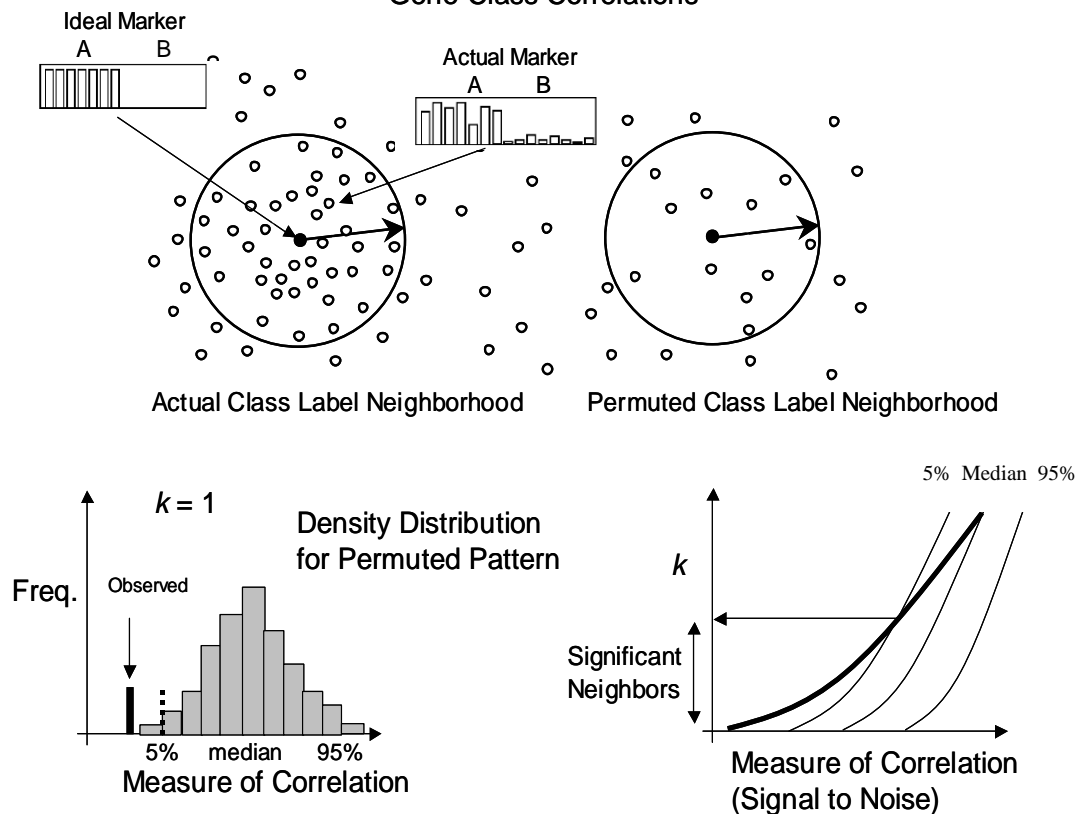
test helps to assess the statistical significance of gene markers in terms of the distribution of class-gene scores using permuted labels.

In the results section the values for permutation tests of marker genes are reported in tables with this format:

Distinction	Distance	Perm 1%	Perm 5%	Median 50%	Feature	Desc
class 0	0.96694607	1.0144908	0.8333578	0.6280173	M93119_at	INSM1 Insulinoma-associated 1
class 0	0.9096911	0.8600172	0.7669801	0.5740431	M30448_s_at	Casein kinase II beta subunit
class 0	0.90010124	0.85051423	0.7251496	0.5494933	S82240_at	RhoE
class 0	0.832689	0.84354156	0.7071885	0.5292253	U44060_at	Homeodomain protein (Prox 1)
class 0	0.83225346	0.8009565	0.68034023	0.5169537	D80004_at	KIAA0182 gene
.....						
class 1	1.6520017	0.9831643	0.84544426	0.6230137	X86693_at	High endothelial venule
class 1	1.2436218	0.88150144	0.7559189	0.5795857	M93426_at	PTPRZ Protein tyrosine phosphatase
class 1	1.2317128	0.86047184	0.70928395	0.5539352	U48705_rna1_s_at	Receptor tyrosine kinase DDR gene
class 1	1.2259983	0.8433512	0.68909335	0.5358038	X86809_at	Major astrocytic phosphoprotein PEA-15
class 1	1.214929	0.8281318	0.6849929	0.5217813	U45955_at	Neuronal membrane glycoprotein M6b
class 1	1.2095517	0.79365546	0.6711517	0.510208	U53204_at	Plectin (PLEC1)
.....						

The distinction column represents the class for which the markers are high (low in the other classes). The Distance column is the signal to noise to the actual phenotype. The Perm. 1%, 5% and 50% columns represent the percentiles (significance levels) in the histograms of signal to noise scores for permuted labels for a given value of k . The Feature column is the gene accession number and the Description column is the gene name. Permutation test results are reported in the gene markers sections: [Multiple tumor class markers, Classic vs. desmoplastic MD markers](#), [SOM-discovered C0 vs. C1 class gene markers](#), and [Treatment outcome markers](#).

Neighborhood Analysis: Assessing Statistical Significance of Gene-Class Correlations



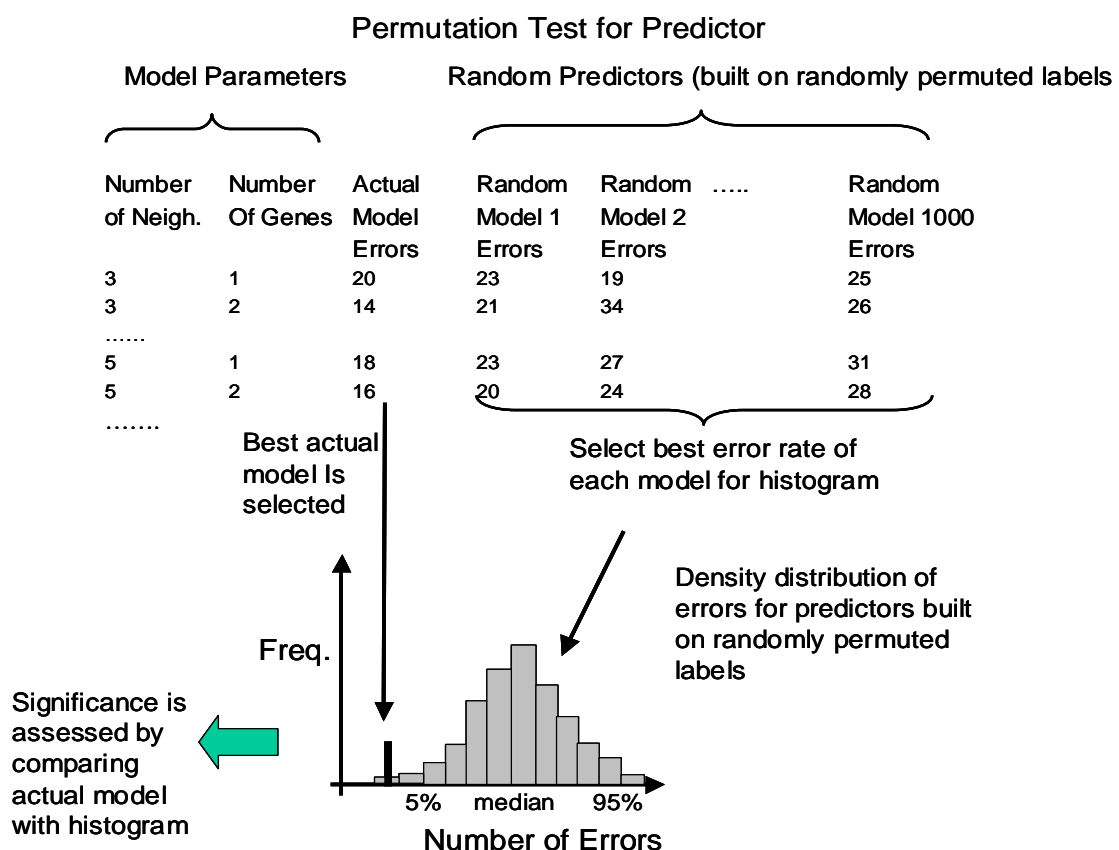
Additional Notes:

- This test helps to assess the significance of gene markers in terms of class-gene correlations but if a group of genes fails to pass the test that by itself does not necessarily imply that they cannot be used to build an effective classifier (Huberty 1994, Kearns and Vazirani 1997). For example, in contrast with the case of morphological distinctions, for treatment outcome prediction the top marker genes do not show overwhelming statistical significance (they are "weak" markers) and yet they are effective when used in combination as input to a classifier.
- The choice of the signal to noise distance is somewhat ad hoc but not unreasonable. The reason the signal to noise ratio was chosen instead of a t -statistic or other class distance measure was mainly historical and empirical: it performed slightly better in a previous study of gene expression (Leukemia) feature selection combined with a weighted voting classifier.
- In terms of feature selection our approach can be considered a filter method based on signal to noise ratios but it is important to keep in mind that when the genes selected by this method are feed to a supervised classifier there is an additional number of genes selection process based on error rates (see the *Algorithms* and *Permutation test for outcome predictor* sections of this document).
- The advantages of performing a permutation test are multiple:
 - It is a distribution-free, direct empirical method to test the significance of the matching of a given phenotype to a particular set of genes (dataset).
 - It does not assume a particular functional form for the distribution or correlation structure of genes.
 - As the permutation test is done on the entire distribution of genes (as scored by signal to noise from the phenotype) the gene-to-gene correlation structure is taken into account.
- Another more geometrical, and sometimes more intuitive, way to look at this procedure is to consider the figure above as a hypothetical projection of normalized gene expression space where each dimension represents an experiment and each data point a gene. The entire dataset of filtered genes will be represented by a collection of data points distributed in that space. Each gene is represented by a point and the closer two points are, the more correlated they are (i.e. across the set of experiments being considered). Now imagine projecting a point that corresponds to an ideal marker gene that perfectly represents the phenotype of interest. This is for example a marker gene that is high in one of the classes and low in the other. This gene will be a perfect classifier to distinguish the two classes. We are interested in finding marker genes that are, if not equal, at least similar to this ideal marker. This can be accomplished by computing a distance or correlation measure between the class labels (phenotype) and the genes. In this sense we are looking at the "neighborhood" of a phenotype in gene expression space trying to find "close" neighbors. A permutation test

in this context is equivalent to moving the ideal gene marker point at random (as the labels are permuted) and obtaining a distribution of neighbors each time it lands to a new reference point (random phenotype) in expression space. By building a histogram of distance distributions to these random reference locations one can assess how “typical” the actual neighborhood of the actual phenotype is compared to random phenotypes. For example, if only once in a thousand random tries we found a set of top 10 markers as correlated as in the actual neighborhood, then we would consider those markers to be significant. In this interpretation, the permutation test resembles a spatial correlation Mantel test in which one measures the significance of finding excess “density” of neighbors (genes) around a point (ideal marker) that represents the phenotype of interest when compared with the density at random phenotype classes.

Permutation Test for Outcome Predictor

There is an additional permutation test (Fisher 1935, Lehman 1986, Good 1994) that was developed to assess the statistical significance of the k -nearest neighbor predictor algorithm. In this test the phenotype (treatment outcome) labels are randomly permuted 1000 times and for each instance a set of models are built using the same set of parameters (e.g. k = number of neighbors, ng = number of features/genes) as the ones used in finding the actual model. Once this is done one selects the best error rate, from the results corresponding to the selected set of parameters, for each of these 1000 random predictors and makes a histogram. The significance of the predictor is assessed by the area of the histogram corresponding to random predictors with better error rates (see figure below). The results of this procedure for the k -nearest neighbor predictor of treatment outcome are reported in the [Permutation test for \$k\$ -nearest neighbor outcome predictor](#) section.



Algorithms

k-Nearest Neighbors (*k*-NN)

We developed a weighted implementation of the *k*-NN algorithm (Dasarathy 1991) that predicts the class of a new sample by calculating the Euclidean distance (*d*) of this sample to the *k* "nearest neighbor" standardized samples in "expression" space in the training set, and by selecting the predicted class to be that of the majority of the *k* samples (the method is defined in terms of Euclidean distances over standardized vectors so it is equivalent to using inner products: $\mathbf{a} \cdot \mathbf{b} / (|\mathbf{a}||\mathbf{b}|)$). We performed the marker gene selection process by which we feed the *k*-NN algorithm only the features with higher correlation with the target class. This feature selection is done by sorting the features according to the signal-to-noise statistic (Golub 1999, Slonim 2000) $(\mu_{class\ 0} - \mu_{class\ 1}) / (\sigma_{class\ 0} + \sigma_{class\ 1})$. In our version of the algorithm the weight of each of the *k* neighbors was weighted according to $1/d$. For our medulloblastoma outcome experiments, the *k*-NN models were evaluated by 60-fold leave-one-out cross-validation whereby a training set of 59 samples was used to predict the class of a randomly withheld sample. This was repeated for all samples and the cumulative error rate was recorded. Models with variable numbers of genes (1-200, selected according to their correlation with the survivor vs. treatment failure distinction in the training set) were tested in this manner. The detailed results of applying this algorithm to the different datasets can be found in the

[Multiple tumor classes predictions](#) (*k*-NN), [Classic vs. desmoplastic MD](#) and [k-nearest neighbors treatment outcome prediction results](#) sections.

Weighted Voting.

The weighted voting algorithm (Golub 1999, Slonim 2000) makes a weighted linear combination of relevant "marker" or "informative" genes obtained in the training set to provide a classification scheme for new samples. The selection of features (marker genes) is accomplished by computing the signal-to-noise statistic S_x (described above). The class predictor is uniquely defined by the initial set of samples and marker genes. In addition to computing S_x , the algorithm also finds the decision boundaries (half way) between the class means: $b_x = (\mu_{class\ 0} + \mu_{class\ 1})/2$ for each gene. To predict the class of a test sample *y*, each gene *x* in the feature set casts a vote: $V_x = S_x (g_{x'} - b_x)$ and the final vote for class 0 or 1 is *sign* ($\sum_x V_x$). The strength or *confidence* in the prediction of the winning class is $(V_{win} - V_{lose}) / (V_{win} + V_{lose})$ (i.e., the relative margin of victory for the vote). The detailed prediction results are the [Weighted voting treatment outcome prediction results](#).

Support Vector Machines.

The Support Vector Machine (SVM) for classification minimizes the generalization error rather than the training error. The basic idea behind SVMs is to construct an optimal separating hyperplane by mapping the gene expression data to a high-dimensional space (Mukherjee et al 1999, Brown et al 2000). Linear separation in this higher dimensional space corresponds to a nonlinear decision boundary in the original space. A new feature selection algorithm was developed to scale the input features to minimize the ratio of the radius around the support vectors and the

margin. The detailed results are in the [SVM treatment outcome prediction results](#) section.

SPLASH.

The Splash algorithm (Califano et al 1999) discovers efficiently and deterministically all statistically significant gene expression patterns in a target class of interest. Statistical significance is evaluated based on the probability of a “pattern,” (i.e. a subset of genes and experiments within a narrow interval of expression values) to occur by chance in the control target class. A greedy set covering algorithms is used to select an optimal subset of statistically significant patterns. These patterns are accumulated and form the basis for a likelihood ratio classification scheme to predict new samples. The detailed results are in the [SPLASH treatment outcome prediction results](#) section.

Predictors using metastatic staging and TrkC.

These classifiers were constructed by finding the decision boundary half way between the classes: $(\mu_{class\ 0} + \mu_{class\ 1})/2$ (using the staging values 0 vs. 1,2,3,4 or the continuous *TrkC* gene expression) and then predicting the unknown sample according to its gene expression value location with respect to that boundary. The detailed results can be found in the [TrkC treatment outcome prediction results](#) and [Staging treatment outcome prediction results](#) sections.

Proportional chance criterion.

In order to compute p-values for non-survival predictions, for example the p-val= 4×10^{-7} for the Classic vs. Desmoplastic classifier reported in the paper (33 out of 34 samples correctly classified) we used a “proportional chance criterion” to evaluate the probability that a random predictor will produce a confusion matrix with the same row and column counts as the gene expression predictor. For example, for a binary class (A vs. B) problem, if α is the prior probability of a sample being in class A and p is the true proportion of samples in class A then $C_p = p\alpha + (1-p)(1-\alpha)$ is the proportion of the overall sample that is expected to receive correct classification by chance alone. Then if C_{model} is the proportion of correct classifications achieved by the gene expression predictor one can estimate its significance by using a Z statistic of the form: $(C_{model} - C_p)/\text{Sqrt}(C_p(1-C_p)/n)$, where n is the total sample count. For more details see chapter VII of Huberty 1994.

Survival analysis and Kaplan-Meier plots

The Kaplan-Meier survival analysis plots are computed using the S-Plus (<http://www.insightful.com/products/splus/>) statistical software package: S-Plus 2000, Guide to Statistics Volume 2, chapter 9. The p-values for the prediction of outcome groups are computed using a log-rank test (Mantel-Haenszel method, chapter 9 in the same reference). The Kaplan Meier plots and associated rank test p-values are included at the end of each of the outcome prediction sections starting in the [k-nearest neighbors treatment outcome prediction results](#) section.

PCA and multidimensional-scaling of Brain tumor samples

Datasets of large dimensionality (i.e. large number of variables e.g. genes) are in general difficult to visualize due to the intrinsic difficulty of reducing and projecting the dataset to a small number of dimensions where standard visualization techniques are applicable. The main problem of performing a projection of that sort is that of preserving the “relevant” or “interesting” structure in the data. In our case this structure corresponds to the intrinsic similarities or the natural clustering of brain samples in the space of gene expression.

A commonly used technique for data reduction, projection and visualization is Principal Component Analysis (PCA). In this approach one finds standardized linear combinations of variables, the “principal components,” which are orthogonal and explain all of the variance in the original dataset. For more details see for example ref. 3. A typical method to obtain a simple projection (multi-dimensional scaling) of the dataset is to plot the top 2 or 3 principal components, which may account for a significant fraction of the variance, in a 2 or 3D scatter plot.

To study the natural clustering of the Brain tumor samples we performed PCA analysis and projected the top three components in 3D and 2D scatter plots (some shown in the paper as part of Figure 1). We considered two subsets of genes: *highly varying*, those with highest variation across samples that passed a variation filter (1,065 genes) and, *marker genes*, the top 10 marker genes of each tumor class by using the signal-to-noise statistics as described in the statistical analysis and prediction section. For the highest variation genes the values were thresholded to 100 from below and 16,000 from above and the variation filter selected genes with at least a 12-fold and 1,200 absolute units of variation between the minimum and maximum values across samples. This produced a subset of 1,065 highly varying genes. For the marker genes the values were thresholded to 20 from below and 16,000 from above and a variation filter selected genes with at least a 5-fold and 500 absolute units of variation between the minimum and maximum values across samples. The genes that passed this filter were ranked according to signal to noise (using medians) and the top 10 markers for each class were selected. This produced a total of 50 genes.

Once the appropriate subset of highly varying or maker genes was selected we computed the 3 principal components using the S-Plus statistical software package using default settings (S-Plus statistical software package: S-Plus 2000, Guide to Statistics Volume 2, chapter 1, <http://www.splus.mathsoft.com/products/splus/splusintro.html>). These three components were then plotted in 3D scatter plots. Figure 1 in the paper shows these plots for highly varying and marker genes where each type of brain tumor is shown in a different color. The plots show the “natural” clustering of brain tumor samples in these two subspaces of gene expression. The components and plots can also be seen in the [Multiple tumor PCA](#) section. Besides the 2D and 3D plots of the top 3 components we also include bar graphs showing the relative importance of the top components and the loadings of the top 6 genes for each component.

Combined classifiers.

The fact that sometimes the prediction algorithms make mistakes in different samples and that the class structure of the confusion matrices is different for each algorithm motivated us to combine some of them to see if the predictions can be improved in this way. We choose a simple scheme combining three algorithms according to majority. For example if the outputs of the three algorithms for a given

sample are *Survivor*, *failure*, and *Survivor*, then the output of the combined predictor will be *Survivor*. The results for two types of model combinations: using a simple majority rule: Staging, *k*-NN and TrkC and SVM, *k*-NN and TrkC can be seen in the [Combined treatment outcome predictors](#) section.

Section II: datasets and clinical attributes

The following sections of this document describe the samples, clinical attributes and datasets in detail.

List of all samples

Number	Sample name	Type	Subtype	Chang Stage	Sex	Age at diagnosis [years/months]	Followup [Months]	Current status [Alive/Dead]	Chemotherapy
1	Brain_MD_1	Medulloblastoma	Classic	T4M1	M	8m	11	D	V,C,Cx,VP
2	Brain_MD_2	Medulloblastoma	Classic	T2M0	M	8yr10m	5	D	V,C,Cx,VP
3	Brain_MD_3	Medulloblastoma	Classic	T3M0	M	6yr	7	D	V,C,Cx
4	Brain_MD_4	Medulloblastoma	Classic	T3M3	M	5yr 3m	7	D	V,C,Cx,VP
5	Brain_MD_5	Medulloblastoma	Classic	M3	M	38yr 2m	7	D	V,C
6	Brain_MD_6	Medulloblastoma	Classic	T4M0	F	7m	9	D	V,C,Cx
7	Brain_MD_7	Medulloblastoma	Classic	T1M0	M	6yr 5m	14	D	V,C,Cx
8	Brain_MD_8	Medulloblastoma	Classic	T3bM1	M	6yr 1m	16	D	V,C,Cx
9	Brain_MD_9	Medulloblastoma	Classic	M0	M	8yr	18	D	V,C,Cx,VP
10	Brain_MD_10	Medulloblastoma	Classic	M0	M	3yr 10m	18	D	V,C,Cx
11	Brain_MD_11	Medulloblastoma	Classic	T2M1	M	8yr 2m	19	D	V,C,Cx,VP,Ca,T,M
12	Brain_MD_12	Medulloblastoma	Classic	M0	F	3yr 9m	25	D	V,C,Cx
13	Brain_MD_13	Medulloblastoma	Classic	T3M3	M	14yr 5m	26	D	V,C,Cx
14	Brain_MD_14	Medulloblastoma	Desmoplastic	M0	M	6yr 3m	33	D	V,C,CC
15	Brain_MD_15	Medulloblastoma	Desmoplastic	T2M0	F	11yr 7m	38	D	V,C,Cx,VP
16	Brain_MD_16	Medulloblastoma	Desmoplastic	T3M3	F	11yr 5m	39	D	V,C,VP
17	Brain_MD_17	Medulloblastoma	Classic	T3bM3	F	3yr 3m	39	D	V,C,Cx
18	Brain_MD_18	Medulloblastoma	Classic	T2M3	M	4yr 4m	42	D	V,C,Cx
19	Brain_MD_19	Medulloblastoma	Classic	M2	F	26yr 1m	65	D	V,C,Cx,VP
20	Brain_MD_20	Medulloblastoma	Classic	T3bM0	M	20yr 6m	92	D	V,C
21	Brain_MD_21	Medulloblastoma	Classic	T2M0	F	23yr 3m	102	D	V,C
22	Brain_MD_22	Medulloblastoma	Desmoplastic	M0	F	5yr 7m	24	A	V,C,CC
23	Brain_MD_23	Medulloblastoma	Desmoplastic	T4M0	M	1yr 4m	25	A	V,C,Cx
24	Brain_MD_24	Medulloblastoma	Classic	T3M0	M	10yr 10m	27	A	V,C,Cx
25	Brain_MD_25	Medulloblastoma	Classic	M0	F	5yr 4m	28	A	V,C,Cx,VP
26	Brain_MD_26	Medulloblastoma	Classic	T2M3	M	1yr	33	A	V,C,Cx,VP
27	Brain_MD_27	Medulloblastoma	Classic	M0	M	5yr 10m	34	A	V,C,Cx
28	Brain_MD_28	Medulloblastoma	Desmoplastic	T4M0	M	6yr 1m	35	A	V,C,Cx
29	Brain_MD_29	Medulloblastoma	Classic	T3M0	F	7yr 5m	35	A	V,C,Cx
30	Brain_MD_30	Medulloblastoma	Desmoplastic	T3M0	F	11yr 9m	36	A	V,C,Cx
31	Brain_MD_31	Medulloblastoma	Classic	M0	M	7yr 4m	39	A	V,C,Cx
32	Brain_MD_32	Medulloblastoma	Desmoplastic	T2M0	M	10yr 11m	39	A	V,C,Cx
33	Brain_MD_33	Medulloblastoma	Classic	T3bM0	M	12yr 9m	41	A	V,C,Cx
34	Brain_MD_34	Medulloblastoma	Classic	T3M1	M	8yr 2m	42	A	V,C,Cx
35	Brain_MD_35	Medulloblastoma	Desmoplastic	T3M0	F	2yr 3m	45	A	V,C,Cx
36	Brain_MD_36	Medulloblastoma	Classic	T3M0	M	5yr 6m	46	A	V,C,Cx
37	Brain_MD_37	Medulloblastoma	Classic	T3M0	F	12yr 7m	51	A	V,C,Cx
38	Brain_MD_38	Medulloblastoma	Desmoplastic	T3M1	F	7m	52	A	V,C,Cx
39	Brain_MD_39	Medulloblastoma	Classic	T3M0	M	10yr 9m	53	A	V,C,Cx
40	Brain_MD_40	Medulloblastoma	Desmoplastic	T4M3	M	3yr 4m	57	A	V,C,Cx
41	Brain_MD_41	Medulloblastoma	Classic	T4M0	F	4yr 8m	60	A	V,C,Cx,VP
42	Brain_MD_42	Medulloblastoma	Classic	T3M3	M	6yr	62	A	V,C,Cx,VP
43	Brain_MD_43	Medulloblastoma	Classic	T3M0	M	9yr 3m	64	A	V,C,Cx

44	Brain_MD_44	Medulloblastoma	Classic	T3M0	M	5yr 3m	66	A	V,C,Cx
45	Brain_MD_45	Medulloblastoma	Classic	T4M0	M	3yr 6m	68	A	V,C,Cx,P
46	Brain_MD_46	Medulloblastoma	Classic	T3M0	M	2yr 4m	68	A	V,C,Cx
47	Brain_MD_47	Medulloblastoma	Classic	T4M0	F	10yr 6m	70	A	V,C,Cx
48	Brain_MD_48	Medulloblastoma	Classic	T3bM0	M	5yr 5m	72	A	V,C,Cx,VP,Ca
49	Brain_MD_49	Medulloblastoma	Classic	T2M0	F	12yr 11m	74	A	V,C,Cx
50	Brain_MD_50	Medulloblastoma	Classic	T3bM0	M	9yr 11m	79	A	V,C,Cx
51	Brain_MD_51	Medulloblastoma	Classic	T3bM0	M	13yr 8m	79	A	V,C,Cx
52	Brain_MD_52	Medulloblastoma	Classic	T2M0	M	1yr 8m	80	A	V,C,Cx
53	Brain_MD_53	Medulloblastoma	Desmoplastic	T2M0	F	5yr 2m	84	A	V,C,Cx
54	Brain_MD_54	Medulloblastoma	Classic	T4M4	F	1yr 5m	85	A	V,C,Cx,VP,Ca,T,M
55	Brain_MD_55	Medulloblastoma	Classic	T3bM2	M	10yr 4m	87	A	V,C,Cx,VP
56	Brain_MD_56	Medulloblastoma	Desmoplastic	T2M0	F	28yr	87	A	V,C
57	Brain_MD_57	Medulloblastoma	Classic	T2M3	M	2yr 7m	97	A	V,C,Cx
58	Brain_MD_58	Medulloblastoma	Classic	T1M0	M	3yr 7m	108	A	V,C,Cx,VP
59	Brain_MD_59	Medulloblastoma	Classic	T3bM0	M	9yr 9m	130	A	V,C
60	Brain_MD_60	Medulloblastoma	Desmoplastic	T3M0	F	2yr	24	A	V,C,Cx
61	Brain_MD_61	Medulloblastoma							
62	Brain_MD_62	Medulloblastoma							V,C,Cx
63	Brain_MD_63	Medulloblastoma							
64	Brain_MD_64	Medulloblastoma							V,C,Cx
65	Brain_MD_65	Medulloblastoma							V,C,Cx
66	Brain_MD_66	Medulloblastoma							V,C
67	Brain_MD_67	Medulloblastoma							V,C,Cx,VP
68	Brain_MGlio_1	Malignant Glioma							
69	Brain_MGlio_2	Malignant Glioma							V= vincristine
70	Brain_MGlio_3	Malignant Glioma							C= cisplatin
71	Brain_MGlio_4	Malignant Glioma							Cx= cytoxan
72	Brain_MGlio_5	Malignant Glioma							VP= etoposide
73	Brain_MGlio_6	Malignant Glioma							CC= CCNU
74	Brain_MGlio_7	Malignant Glioma							Ca= carboplatin
75	Brain_MGlio_8	Malignant Glioma							P= procarbazine
76	Brain_MGlio_9	Malignant Glioma							M= methotrexate
77	Brain_MGlio_10	Malignant Glioma							T= thiotepe
78	Brain_Rhab_1	AT/RT (Brain)							
79	Brain_Rhab_2	AT/RT (Renal)							
80	Brain_Rhab_3	AT/RT (Renal)							
81	Brain_Rhab_4	AT/RT (Brain)							
82	Brain_Rhab_5	AT/RT (Extra Renal)							
83	Brain_Rhab_6	AT/RT (Extra Renal)							
84	Brain_Rhab_7	AT/RT (Renal)							
85	Brain_Rhab_8	AT/RT (Brain)							
86	Brain_Rhab_9	AT/RT (Brain)							
87	Brain_Rhab_10	AT/RT (Brain)							
88	Brain_Ncer_1	Normal cerebellum							
89	Brain_Ncer_2	Normal cerebellum							
90	Brain_Ncer_3	Normal cerebellum							
91	Brain_Ncer_4	Normal cerebellum							
92	Brain_PNET_1	PNET							
93	Brain_PNET_2	PNET							
94	Brain_PNET_3	PNET							
95	Brain_PNET_4	PNET							
96	Brain_PNET_5	PNET							
97	Brain_PNET_6	PNET							
98	Brain_PNET_7	PNET (pineoblastoma)							
99	Brain_PNET_8	PNET (pineoblastoma)							

Dataset A, A1, A2 - multiple tumor samples

Dataset A: 10 medulloblastomas, 10 malignant gliomas, 10 AT/RT (5 CNS, 5 renal-extrarenal), 4 normal cerebellums and 8 supratentorial PNETs.

Two of the supratentorial PNETs are pineoblastomas, which historically have been inconsistently included in the PNET category. The analysis was repeated excluding these 2 pineoblastomas.

Dataset A1: 10 medulloblastomas, 10 malignant gliomas, 10 AT/RT (5 CNS, 5 renal-extrarenal), 4 normal cerebellums and 6 supratentorial PNETs.

To test whether inclusion of a larger number of medulloblastomas might lessen the distinctions noted in Dataset A, 50 more medulloblastoma samples were added and the PCA analysis repeated.

Dataset A2: 60 medulloblastomas, 10 malignant gliomas, 10 AT/RT (5 CNS, 5 renal-extrarenal), 4 normal cerebellums and 6 supratentorial PNETs.

Dataset A

Sample number	Sample name	Type
1	Brain_MD_12	Medulloblastoma
2	Brain_MD_61	Medulloblastoma
3	Brain_MD_15	Medulloblastoma
4	Brain_MD_57	Medulloblastoma
5	Brain_MD_33	Medulloblastoma
6	Brain_MD_64	Medulloblastoma
7	Brain_MD_17	Medulloblastoma
8	Brain_MD_62	Medulloblastoma
9	Brain_MD_63	Medulloblastoma
10	Brain_MD_32	Medulloblastoma
11	Brain_MGlio_1	Malignant Glioma
12	Brain_MGlio_2	Malignant Glioma
13	Brain_MGlio_3	Malignant Glioma
14	Brain_MGlio_4	Malignant Glioma
15	Brain_MGlio_5	Malignant Glioma
16	Brain_MGlio_6	Malignant Glioma
17	Brain_MGlio_7	Malignant Glioma
18	Brain_MGlio_8	Malignant Glioma
19	Brain_MGlio_9	Malignant Glioma
20	Brain_MGlio_10	Malignant Glioma
21	Brain_Rhab_1	AT/RT (Brain)
22	Brain_Rhab_2	AT/RT (Renal)
23	Brain_Rhab_3	AT/RT (Renal)
24	Brain_Rhab_4	AT/RT (Brain)
25	Brain_Rhab_5	AT/RT (Extra Renal)
26	Brain_Rhab_6	AT/RT (Extra Renal)
27	Brain_Rhab_7	AT/RT (Renal)

28	Brain_Rhab_8	AT/RT (Brain)
29	Brain_Rhab_9	AT/RT (Brain)
30	Brain_Rhab_10	AT/RT (Brain)
31	Brain_Ncer_1	Normal cerebellum
32	Brain_Ncer_2	Normal cerebellum
33	Brain_Ncer_3	Normal cerebellum
34	Brain_Ncer_4	Normal cerebellum
35	Brain_PNET_1	PNET
36	Brain_PNET_2	PNET
37	Brain_PNET_3	PNET
38	Brain_PNET_4	PNET
39	Brain_PNET_5	PNET
40	Brain_PNET_6	PNET
41	Brain_PNET_7	PNET (pineoblastoma)
42	Brain_PNET_8	PNET (pineoblastoma)

Dataset A1

Sample number	Sample name	Type
1	Brain_MD_12	Medulloblastoma
2	Brain_MD_61	Medulloblastoma
3	Brain_MD_15	Medulloblastoma
4	Brain_MD_57	Medulloblastoma
5	Brain_MD_33	Medulloblastoma
6	Brain_MD_64	Medulloblastoma
7	Brain_MD_17	Medulloblastoma
8	Brain_MD_62	Medulloblastoma
9	Brain_MD_63	Medulloblastoma
10	Brain_MD_32	Medulloblastoma
11	Brain_MGlio_1	Malignant Glioma
12	Brain_MGlio_2	Malignant Glioma
13	Brain_MGlio_3	Malignant Glioma
14	Brain_MGlio_4	Malignant Glioma
15	Brain_MGlio_5	Malignant Glioma
16	Brain_MGlio_6	Malignant Glioma
17	Brain_MGlio_7	Malignant Glioma
18	Brain_MGlio_8	Malignant Glioma
19	Brain_MGlio_9	Malignant Glioma
20	Brain_MGlio_10	Malignant Glioma
21	Brain_Rhab_1	AT/RT (Brain)
22	Brain_Rhab_2	AT/RT (Renal)
23	Brain_Rhab_3	AT/RT (Renal)
24	Brain_Rhab_4	AT/RT (Brain)
25	Brain_Rhab_5	AT/RT (Extra Renal)
26	Brain_Rhab_6	AT/RT (Extra Renal)
27	Brain_Rhab_7	AT/RT (Renal)
28	Brain_Rhab_8	AT/RT (Brain)
29	Brain_Rhab_9	AT/RT (Brain)
30	Brain_Rhab_10	AT/RT (Brain)
31	Brain_Ncer_1	Normal cerebellum
32	Brain_Ncer_2	Normal cerebellum
33	Brain_Ncer_3	Normal cerebellum
34	Brain_Ncer_4	Normal cerebellum
35	Brain_PNET_1	PNET
36	Brain_PNET_2	PNET

37	Brain_PNET_3	PNET
38	Brain_PNET_4	PNET
39	Brain_PNET_5	PNET
40	Brain_PNET_6	PNET

Dataset A2

Sample number	Sample name	Type
1	Brain_MD_1	Medulloblastoma
2	Brain_MD_2	Medulloblastoma
3	Brain_MD_3	Medulloblastoma
4	Brain_MD_4	Medulloblastoma
5	Brain_MD_5	Medulloblastoma
6	Brain_MD_6	Medulloblastoma
7	Brain_MD_7	Medulloblastoma
8	Brain_MD_8	Medulloblastoma
9	Brain_MD_9	Medulloblastoma
10	Brain_MD_10	Medulloblastoma
11	Brain_MD_11	Medulloblastoma
12	Brain_MD_12	Medulloblastoma
13	Brain_MD_13	Medulloblastoma
14	Brain_MD_14	Medulloblastoma
15	Brain_MD_15	Medulloblastoma
16	Brain_MD_16	Medulloblastoma
17	Brain_MD_17	Medulloblastoma
18	Brain_MD_18	Medulloblastoma
19	Brain_MD_19	Medulloblastoma
20	Brain_MD_20	Medulloblastoma
21	Brain_MD_21	Medulloblastoma
22	Brain_MD_22	Medulloblastoma
23	Brain_MD_23	Medulloblastoma
24	Brain_MD_24	Medulloblastoma
25	Brain_MD_25	Medulloblastoma
26	Brain_MD_26	Medulloblastoma
27	Brain_MD_27	Medulloblastoma
28	Brain_MD_28	Medulloblastoma
29	Brain_MD_29	Medulloblastoma
30	Brain_MD_30	Medulloblastoma
31	Brain_MD_31	Medulloblastoma
32	Brain_MD_32	Medulloblastoma
33	Brain_MD_33	Medulloblastoma
34	Brain_MD_34	Medulloblastoma
35	Brain_MD_35	Medulloblastoma
36	Brain_MD_36	Medulloblastoma
37	Brain_MD_37	Medulloblastoma
38	Brain_MD_38	Medulloblastoma
39	Brain_MD_39	Medulloblastoma
40	Brain_MD_40	Medulloblastoma
41	Brain_MD_41	Medulloblastoma

42	Brain_MD_42	Medulloblastoma
43	Brain_MD_43	Medulloblastoma
44	Brain_MD_44	Medulloblastoma
45	Brain_MD_45	Medulloblastoma
46	Brain_MD_46	Medulloblastoma
47	Brain_MD_47	Medulloblastoma
48	Brain_MD_48	Medulloblastoma
49	Brain_MD_49	Medulloblastoma
50	Brain_MD_50	Medulloblastoma
51	Brain_MD_51	Medulloblastoma
52	Brain_MD_52	Medulloblastoma
53	Brain_MD_53	Medulloblastoma
54	Brain_MD_54	Medulloblastoma
55	Brain_MD_55	Medulloblastoma
56	Brain_MD_56	Medulloblastoma
57	Brain_MD_57	Medulloblastoma
58	Brain_MD_58	Medulloblastoma
59	Brain_MD_59	Medulloblastoma
60	Brain_MD_60	Medulloblastoma
61	Brain_MGlio_1	Malignant Glioma
62	Brain_MGlio_2	Malignant Glioma
63	Brain_MGlio_3	Malignant Glioma
64	Brain_MGlio_4	Malignant Glioma
65	Brain_MGlio_5	Malignant Glioma
66	Brain_MGlio_6	Malignant Glioma
67	Brain_MGlio_7	Malignant Glioma
68	Brain_MGlio_8	Malignant Glioma
69	Brain_MGlio_9	Malignant Glioma
70	Brain_MGlio_10	Malignant Glioma
71	Brain_Rhab_1	AT/RT (Brain)
72	Brain_Rhab_2	AT/RT (Renal)
73	Brain_Rhab_3	AT/RT (Renal)
74	Brain_Rhab_4	AT/RT (Brain)
75	Brain_Rhab_5	AT/RT (Extra Renal)
76	Brain_Rhab_6	AT/RT (Extra Renal)
77	Brain_Rhab_7	AT/RT (Renal)
78	Brain_Rhab_8	AT/RT (Brain)
79	Brain_Rhab_9	AT/RT (Brain)
80	Brain_Rhab_10	AT/RT (Brain)
81	Brain_Ncer_1	Normal cerebellum
82	Brain_Ncer_2	Normal cerebellum
83	Brain_Ncer_3	Normal cerebellum
84	Brain_Ncer_4	Normal cerebellum
85	Brain_PNET_1	PNET
86	Brain_PNET_2	PNET
87	Brain_PNET_3	PNET
88	Brain_PNET_4	PNET
89	Brain_PNET_5	PNET
90	Brain_PNET_6	PNET

Dataset B - MD classic-desmoplastic

Dataset B: 25 classic and 9 desmoplastic medulloblastomas.

Number	Sample name	Type	Subtype
1	Brain_MD_7	Medulloblastoma	Classic
2	Brain_MD_59	Medulloblastoma	Classic
3	Brain_MD_20	Medulloblastoma	Classic
4	Brain_MD_21	Medulloblastoma	Classic
5	Brain_MD_50	Medulloblastoma	Classic
6	Brain_MD_49	Medulloblastoma	Classic
7	Brain_MD_45	Medulloblastoma	Classic
8	Brain_MD_43	Medulloblastoma	Classic
9	Brain_MD_8	Medulloblastoma	Classic
10	Brain_MD_42	Medulloblastoma	Classic
11	Brain_MD_1	Medulloblastoma	Classic
12	Brain_MD_4	Medulloblastoma	Classic
13	Brain_MD_55	Medulloblastoma	Classic
14	Brain_MD_41	Medulloblastoma	Classic
15	Brain_MD_37	Medulloblastoma	Classic
16	Brain_MD_3	Medulloblastoma	Classic
17	Brain_MD_34	Medulloblastoma	Classic
18	Brain_MD_29	Medulloblastoma	Classic
19	Brain_MD_13	Medulloblastoma	Classic
20	Brain_MD_24	Medulloblastoma	Classic
21	Brain_MD_65	Medulloblastoma	Classic
22	Brain_MD_5	Medulloblastoma	Classic
23	Brain_MD_66	Medulloblastoma	Classic
24	Brain_MD_67	Medulloblastoma	Classic
25	Brain_MD_58	Medulloblastoma	Classic
26	Brain_MD_53	Medulloblastoma	Desmoplastic
27	Brain_MD_56	Medulloblastoma	Desmoplastic
28	Brain_MD_16	Medulloblastoma	Desmoplastic
29	Brain_MD_40	Medulloblastoma	Desmoplastic
30	Brain_MD_35	Medulloblastoma	Desmoplastic
31	Brain_MD_30	Medulloblastoma	Desmoplastic
32	Brain_MD_23	Medulloblastoma	Desmoplastic
33	Brain_MD_28	Medulloblastoma	Desmoplastic
34	Brain_MD_60	Medulloblastoma	Desmoplastic

Dataset C - MD outcome

Dataset C: 39 medulloblastomas survivors and 21 treatment failures (non-survivors)

Number	Sample name	Type	Subtype	Chang Stage	Sex	Age at diagnosis [years/months]	Followup [Months]	Current status [Alive/Dead]	Chemotherapy
1	Brain_MD_1	Medulloblastoma	Classic	T4M1	M	8m	11	D	V,C,Cx,VP
2	Brain_MD_2	Medulloblastoma	Classic	T2M0	M	8yr10m	5	D	V,C,Cx,VP
3	Brain_MD_3	Medulloblastoma	Classic	T3M0	M	6yr	7	D	V,C,Cx
4	Brain_MD_4	Medulloblastoma	Classic	T3M3	M	5yr 3m	7	D	V,C,Cx,VP
5	Brain_MD_5	Medulloblastoma	Classic	M3	M	38yr 2m	7	D	V,C
6	Brain_MD_6	Medulloblastoma	Classic	T4M0	F	7m	9	D	V,C,Cx
7	Brain_MD_7	Medulloblastoma	Classic	T1M0	M	6yr 5m	14	D	V,C,Cx
8	Brain_MD_8	Medulloblastoma	Classic	T3bM1	M	6yr 1m	16	D	V,C,Cx
9	Brain_MD_9	Medulloblastoma	Classic	M0	M	8yr	18	D	V,C,Cx,VP
10	Brain_MD_10	Medulloblastoma	Classic	M0	M	3yr 10m	18	D	V,C,Cx
11	Brain_MD_11	Medulloblastoma	Classic	T2M1	M	8yr 2m	19	D	V,C,Cx,VP,Ca,T,M
12	Brain_MD_12	Medulloblastoma	Classic	M0	F	3yr 9m	25	D	V,C,Cx
13	Brain_MD_13	Medulloblastoma	Classic	T3M3	M	14yr 5m	26	D	V,C,Cx
14	Brain_MD_14	Medulloblastoma	Desmoplastic	M0	M	6yr 3m	33	D	V,C,CC
15	Brain_MD_15	Medulloblastoma	Desmoplastic	T2M0	F	11yr 7m	38	D	V,C,Cx,VP
16	Brain_MD_16	Medulloblastoma	Desmoplastic	T3M3	F	11yr 5m	39	D	V,C,VP
17	Brain_MD_17	Medulloblastoma	Classic	T3bM3	F	3yr 3m	39	D	V,C,Cx
18	Brain_MD_18	Medulloblastoma	Classic	T2M3	M	4yr 4m	42	D	V,C,Cx
19	Brain_MD_19	Medulloblastoma	Classic	M2	F	26yr 1m	65	D	V,C,Cx,VP
20	Brain_MD_20	Medulloblastoma	Classic	T3bM0	M	20yr 6m	92	D	V,C
21	Brain_MD_21	Medulloblastoma	Classic	T2M0	F	23yr 3m	102	D	V,C
22	Brain_MD_22	Medulloblastoma	Desmoplastic	M0	F	5yr 7m	24	A	V,C,CC
23	Brain_MD_23	Medulloblastoma	Desmoplastic	T4M0	M	1yr 4m	25	A	V,C,Cx
24	Brain_MD_24	Medulloblastoma	Classic	T3M0	M	10yr 10m	27	A	V,C,Cx
25	Brain_MD_25	Medulloblastoma	Classic	M0	F	5yr 4m	28	A	V,C,Cx,VP
26	Brain_MD_26	Medulloblastoma	Classic	T2M3	M	1yr	33	A	V,C,Cx,VP
27	Brain_MD_27	Medulloblastoma	Classic	M0	M	5yr 10m	34	A	V,C,Cx
28	Brain_MD_28	Medulloblastoma	Desmoplastic	T4M0	M	6yr 1m	35	A	V,C,Cx
29	Brain_MD_29	Medulloblastoma	Classic	T3M0	F	7yr 5m	35	A	V,C,Cx
30	Brain_MD_30	Medulloblastoma	Desmoplastic	T3M0	F	11yr 9m	36	A	V,C,Cx
31	Brain_MD_31	Medulloblastoma	Classic	M0	M	7yr 4m	39	A	V,C,Cx
32	Brain_MD_32	Medulloblastoma	Desmoplastic	T2M0	M	10yr 11m	39	A	V,C,Cx
33	Brain_MD_33	Medulloblastoma	Classic	T3bM0	M	12yr 9m	41	A	V,C,Cx
34	Brain_MD_34	Medulloblastoma	Classic	T3M1	M	8yr 2m	42	A	V,C,Cx
35	Brain_MD_35	Medulloblastoma	Desmoplastic	T3M0	F	2yr 3m	45	A	V,C,Cx
36	Brain_MD_36	Medulloblastoma	Classic	T3M0	M	5yr 6m	46	A	V,C,Cx
37	Brain_MD_37	Medulloblastoma	Classic	T3M0	F	12yr 7m	51	A	V,C,Cx
38	Brain_MD_38	Medulloblastoma	Desmoplastic	T3M1	F	7m	52	A	V,C,Cx
39	Brain_MD_39	Medulloblastoma	Classic	T3M0	M	10yr 9m	53	A	V,C,Cx
40	Brain_MD_40	Medulloblastoma	Desmoplastic	T4M3	M	3yr 4m	57	A	V,C,Cx
41	Brain_MD_41	Medulloblastoma	Classic	T4M0	F	4yr 8m	60	A	V,C,Cx,VP
42	Brain_MD_42	Medulloblastoma	Classic	T3M3	M	6yr	62	A	V,C,Cx,VP
43	Brain_MD_43	Medulloblastoma	Classic	T3M0	M	9yr 3m	64	A	V,C,Cx
44	Brain_MD_44	Medulloblastoma	Classic	T3M0	M	5yr 3m	66	A	V,C,Cx
45	Brain_MD_45	Medulloblastoma	Classic	T4M0	M	3yr 6m	68	A	V,C,Cx,P
46	Brain_MD_46	Medulloblastoma	Classic	T3M0	M	2yr 4m	68	A	V,C,Cx
47	Brain_MD_47	Medulloblastoma	Classic	T4M0	F	10yr 6m	70	A	V,C,Cx
48	Brain_MD_48	Medulloblastoma	Classic	T3bM0	M	5yr 5m	72	A	V,C,Cx,VP,Ca

49	Brain_MD_49	Medulloblastoma	Classic	T2M0	F	12yr 11m	74	A	V,C,Cx
50	Brain_MD_50	Medulloblastoma	Classic	T3bM0	M	9yr 11m	79	A	V,C,Cx
51	Brain_MD_51	Medulloblastoma	Classic	T3bM0	M	13yr 8m	79	A	V,C,Cx
52	Brain_MD_52	Medulloblastoma	Classic	T2M0	M	1yr 8m	80	A	V,C,Cx
53	Brain_MD_53	Medulloblastoma	Desmoplastic	T2M0	F	5yr 2m	84	A	V,C,Cx
54	Brain_MD_54	Medulloblastoma	Classic	T4M4	F	1yr 5m	85	A	V,C,Cx,VP,Ca,T,M
55	Brain_MD_55	Medulloblastoma	Classic	T3bM2	M	10yr 4m	87	A	V,C,Cx,VP
56	Brain_MD_56	Medulloblastoma	Desmoplastic	T2M0	F	28yr	87	A	V,C
57	Brain_MD_57	Medulloblastoma	Classic	T2M3	M	2yr 7m	97	A	V,C,Cx
58	Brain_MD_58	Medulloblastoma	Classic	T1M0	M	3yr 7m	108	A	V,C,Cx,VP
59	Brain_MD_59	Medulloblastoma	Classic	T3bM0	M	9yr 9m	130	A	V,C
60	Brain_MD_60	Medulloblastoma	Desmoplastic	T3M0	F	2yr	24	A	V,C,Cx

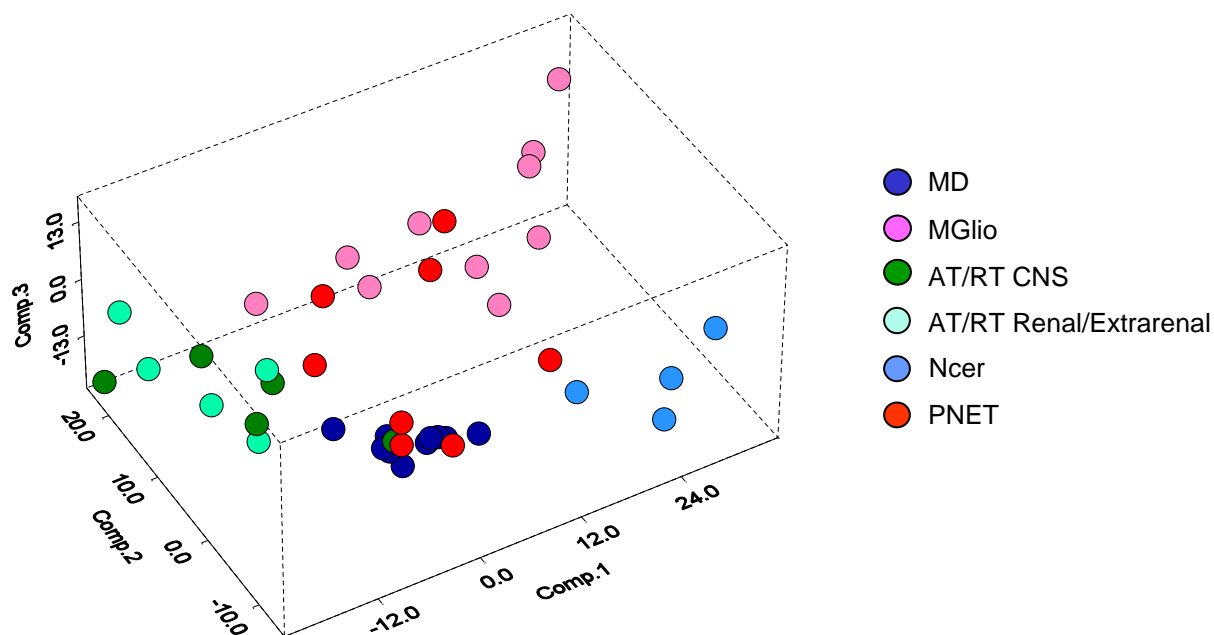
Section III: detailed analysis results

This section presents the results of applying the methods of section I to the datasets of section II. A brief comment precedes each table of results.

Multiple tumor PCA

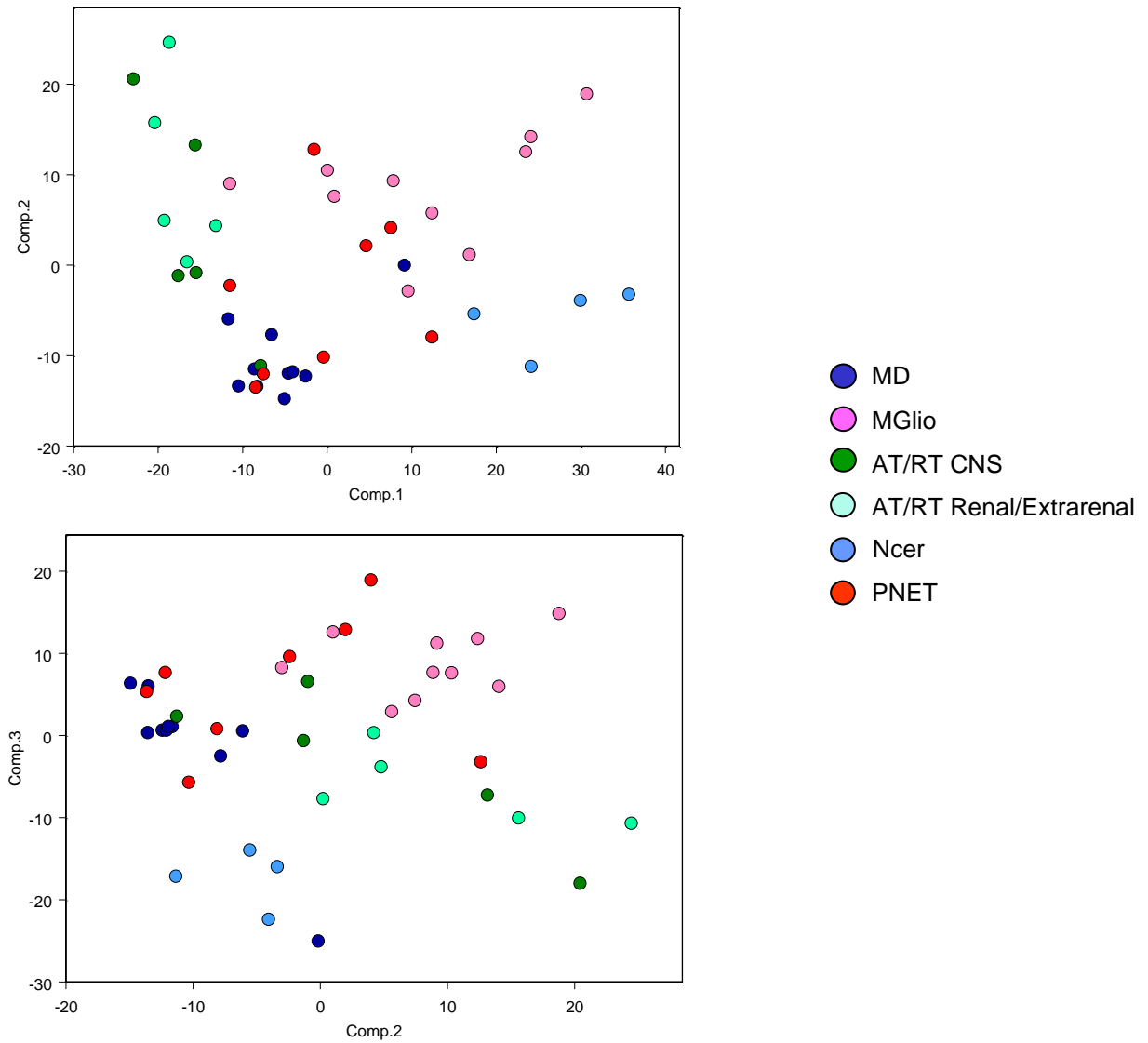
This section contains the PCA projections of highly varying and marker genes for datasets A, A1 and A2. The genes were filtered as described in the [PCA and multidimensional-scaling of Brain tumor](#) samples section.

Dataset A (42 samples) – highly varying genes

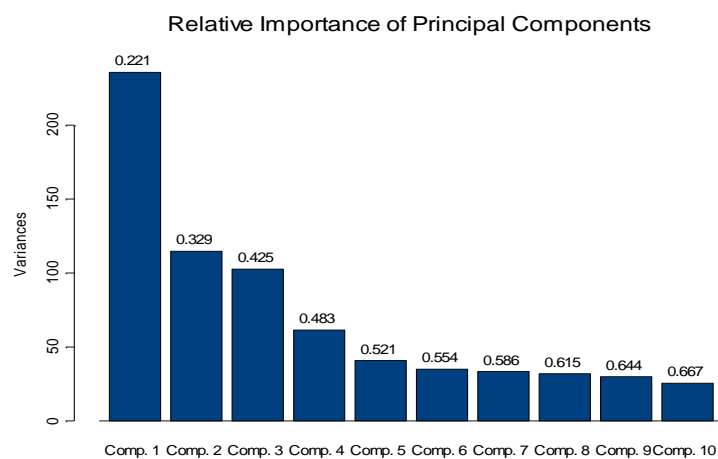


Highly varying genes were selected by using a stringent variation filter (see parameters in table below). The plot above shows the projection of the first 3 components. Notice the relative clustering of tumor samples according to tissue type. The MD samples cluster tightly while the PNET and M. Glio. appear to scatter much more. The AR/RT renal/extra-renal and CNS varieties cluster closer to each other much more than to other types.

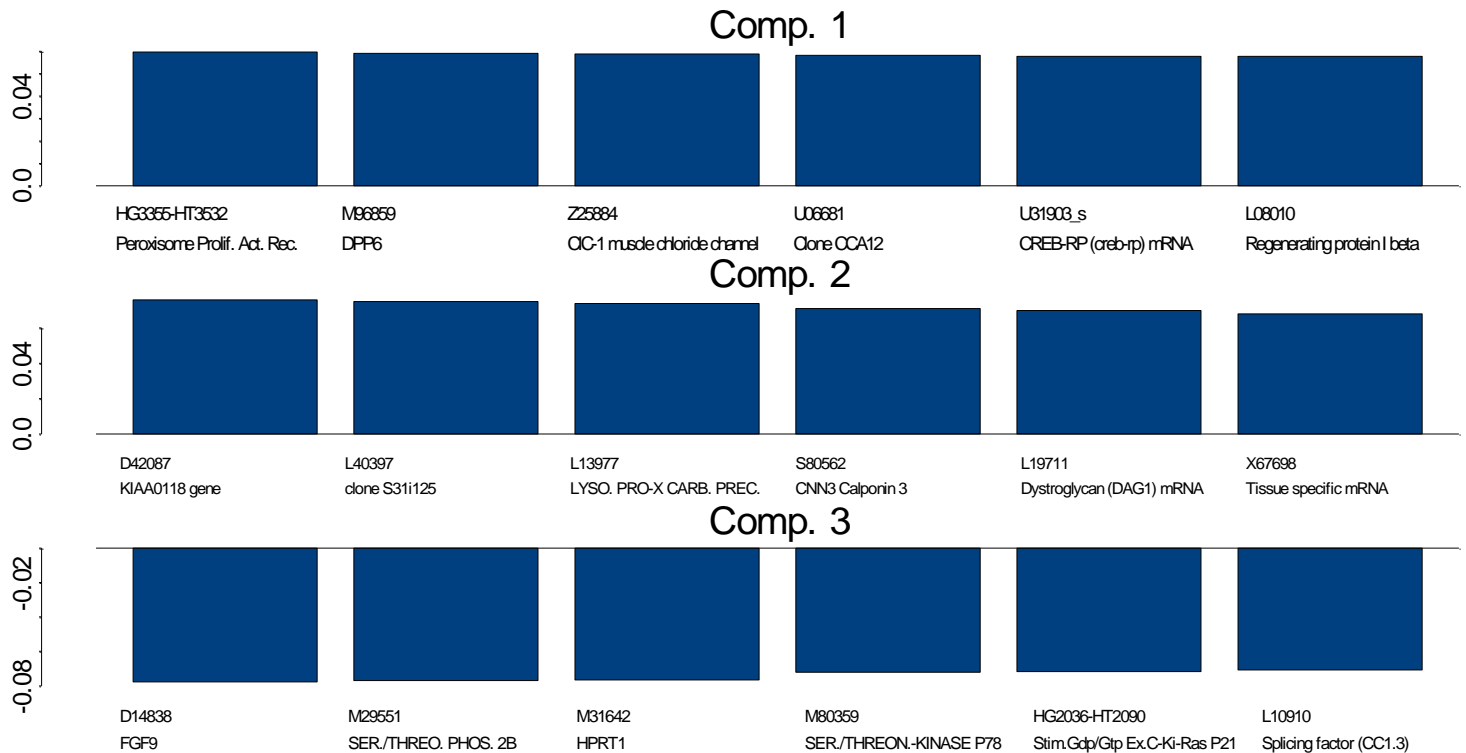
The next two plots show 2D projections of the first vs. second and second vs. third components.



The next bar graph shows the relative importance of the first components. The first three components account for 42.5% of the variance of the highly varying genes.



The bar graph below shows the contribution of the top 6 genes for each of the three principal components. Notice the almost equal weight given to multiple genes.



PCA of Multiple Tumor Samples

Dataset A

Part I: Genes with High Variation:

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 12 (12-fold), max-min= 1200 absolute units

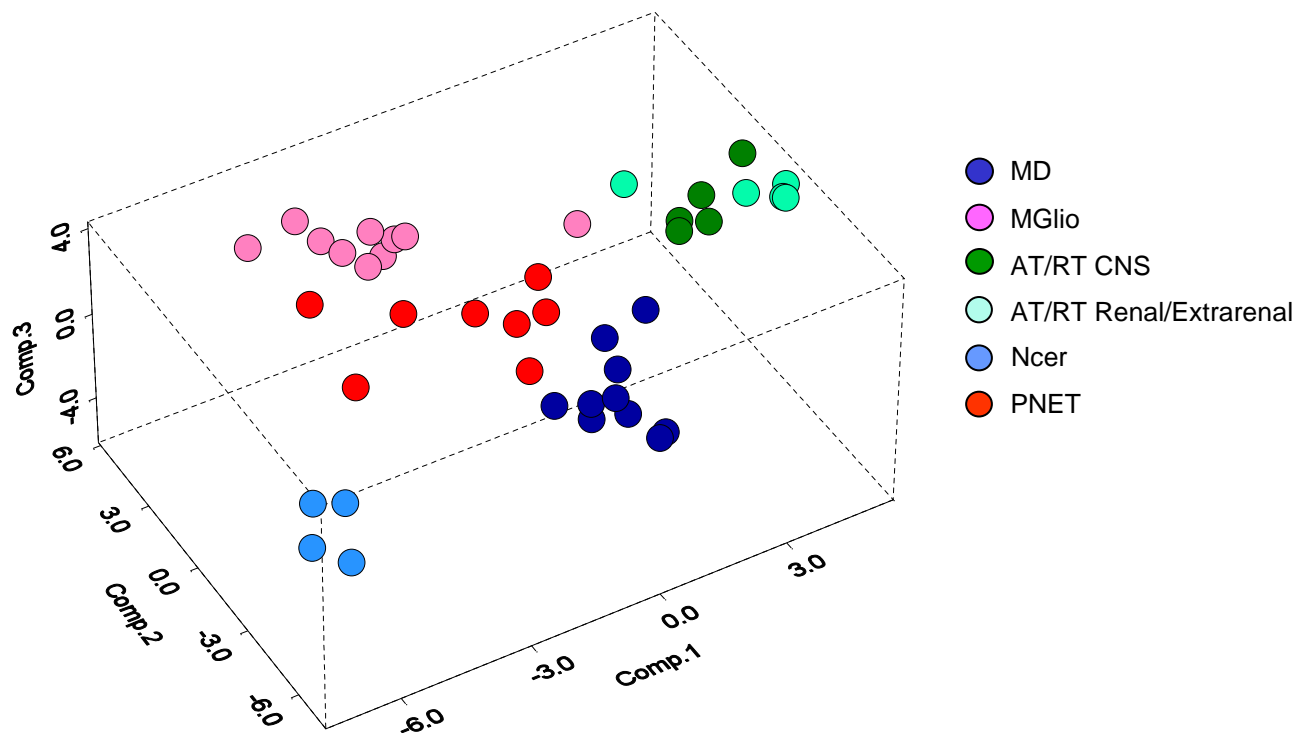
Number of features (genes) = 1065

Sample	class	PCA Components		
		C1	C2	C3

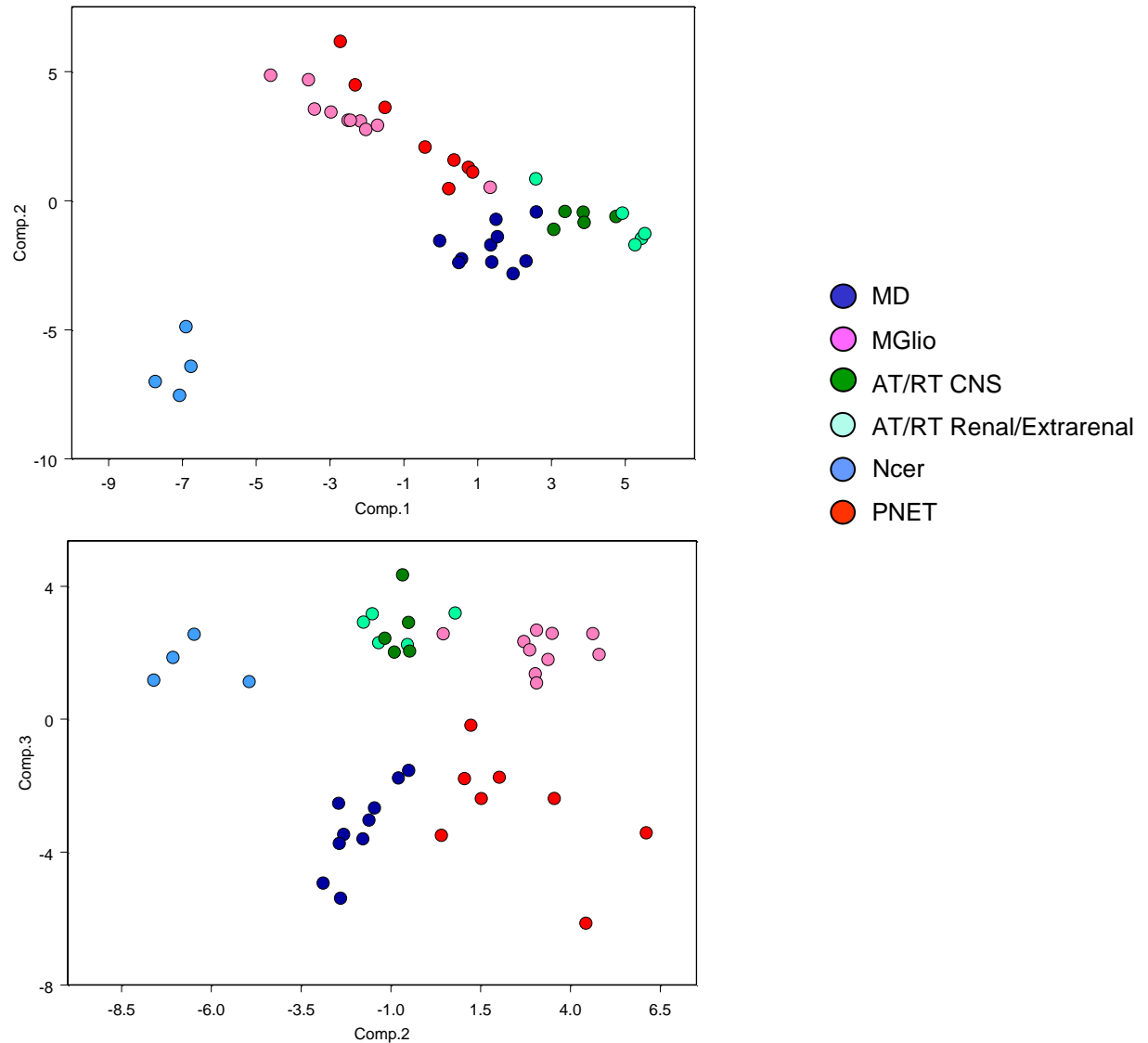
Brain_MD_12	0	-11.5892	-6.03135	0.440673
Brain_MD_61	0	-2.4498	-12.3523	0.541948
Brain_MD_15	0	-4.50062	-12.0282	0.534799
Brain_MD_57	0	-8.18619	-13.4928	0.255259
Brain_MD_33	0	-4.96798	-14.8525	6.25611
Brain_MD_64	0	-6.47181	-7.76031	-2.60217
Brain_MD_17	0	-8.49966	-11.5705	0.990409
Brain_MD_62	0	9.224503	-0.09585	-25.1327
Brain_MD_63	0	-10.4064	-13.4434	5.931149
Brain_MD_32	0	-3.99145	-11.8767	0.990155
Brain_MGlio_1	1	9.679457	-2.95242	8.168946
Brain_MGlio_2	1	30.79565	18.83608	14.74503
Brain_MGlio_3	1	23.58435	12.43997	11.68188
Brain_MGlio_4	1	12.51082	5.673459	2.789562
Brain_MGlio_5	1	7.913009	9.232989	11.14407
Brain_MGlio_6	1	0.940745	7.517246	4.158792
Brain_MGlio_7	1	0.124103	10.39103	7.501235
Brain_MGlio_8	1	-11.4252	8.928279	7.558252
Brain_MGlio_9	1	24.17954	14.10684	5.865164
Brain_MGlio_10	1	16.90079	1.072724	12.49571
Brain_Rhab_1	2	-22.8312	20.49559	-18.1079
Brain_Rhab_2	3	-16.4952	0.282439	-7.79654
Brain_Rhab_3	3	-20.2866	15.65098	-10.1592
Brain_Rhab_4	2	-15.4183	-0.91765	6.466784
Brain_Rhab_5	3	-19.1736	4.851007	-3.92123
Brain_Rhab_6	3	-18.5824	24.52529	-10.7888
Brain_Rhab_7	3	-13.0583	4.269002	0.225192
Brain_Rhab_8	2	-17.5284	-1.24629	-0.73811
Brain_Rhab_9	2	-7.77349	-11.2075	2.246888
Brain_Rhab_10	2	-15.5038	13.20266	-7.3608
Brain_Ncer_1	4	17.45483	-5.47981	-14.059
Brain_Ncer_2	4	24.21633	-11.2845	-17.2432
Brain_Ncer_3	4	30.04492	-4.00302	-22.484
Brain_Ncer_4	4	35.79165	-3.31034	-16.0721
Brain_PNET_1	5	-8.35903	-13.5797	5.241555
Brain_PNET_2	5	-11.4161	-2.351	9.501429
Brain_PNET_3	5	7.634704	4.054882	18.81268
Brain_PNET_4	5	-1.46068	12.68868	-3.29885
Brain_PNET_5	5	-7.45764	-12.1163	7.544325
Brain_PNET_6	5	-0.34124	-10.2635	-5.81399
Brain_PNET_7	5	12.47783	-8.04875	0.715326
Brain_PNET_8	5	4.700972	2.045522	12.7752

Dataset A (42 samples) – class marker genes

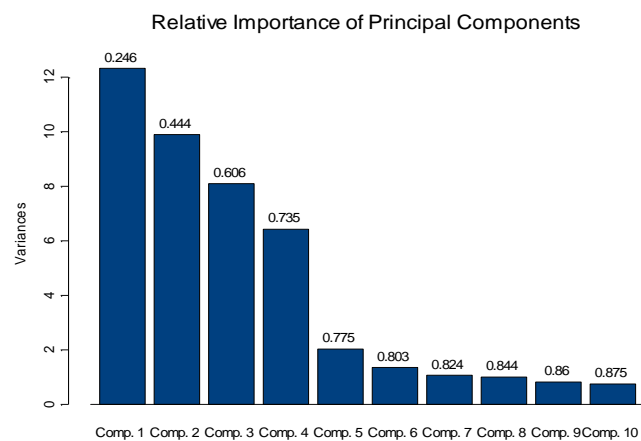
The top 10 marker genes per class were selected as described in the [Gene marker selection](#) section. The top 3 components for that set of 50 genes are shown below. Notice how the clustering and separation by tissue type is now more pronounced than in the case of highly varying genes. This is not surprising because the marker genes are some of the best class separator variables and therefore will produce one of the “cleanest” projections). Notice how the MD and PNET classes tend to occupy different areas of expression space. This was not as evident in the highly varying genes projection. The fact that the samples separate so well also implies that one should be able to build a classifier that separates the classes with low probability of error (see the results in the [Multiple tumor classes predictions \(k-NN\)](#) section).



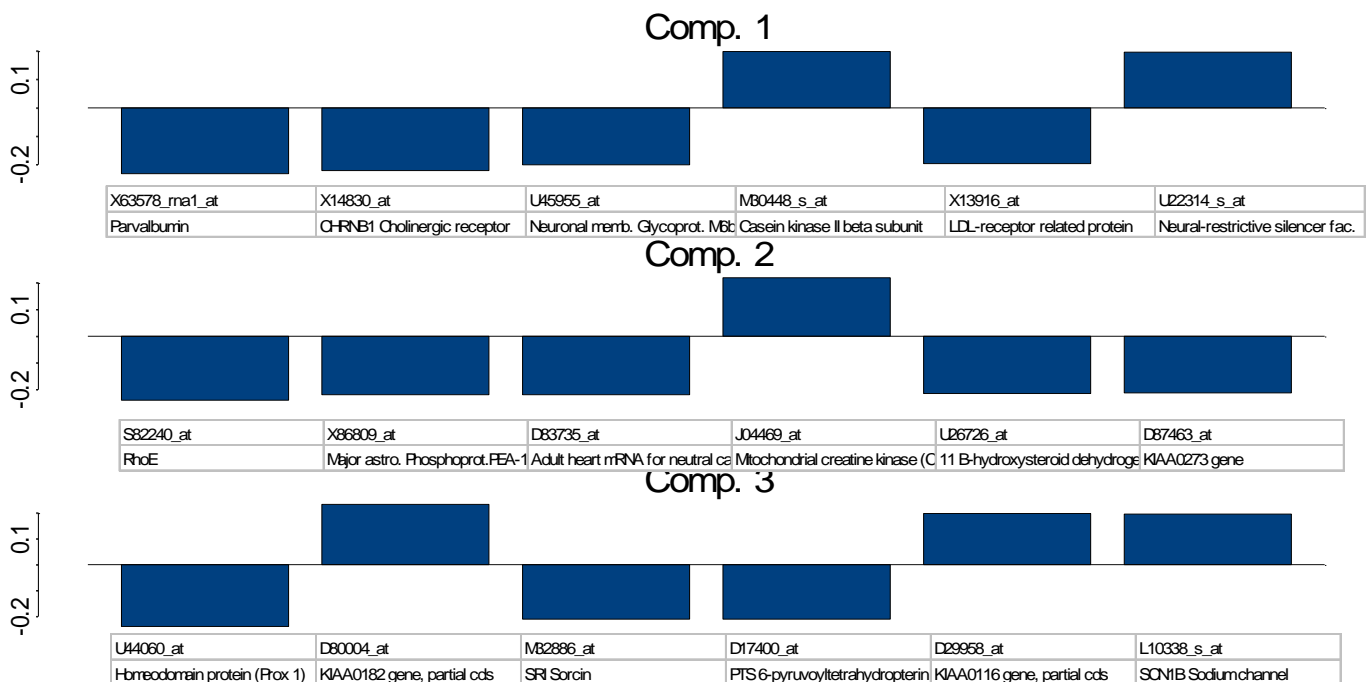
The next two plots show 2D projections of the first vs. second and second vs. third components.



The next bar graph shows the relative importance of the first components. The first three components account for 60.6% of the variance of the marker genes.



The bar graph below shows the contribution of the top 6 genes for each of the three principal components. The different combination of signs in each component is presumably a consequence of the fact that the marker genes behave as a group of correlated genes but also almost orthogonal across classes.



Part I: Top 10 Marker genes for each class (total 50 genes)

Values thresholded to 20 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

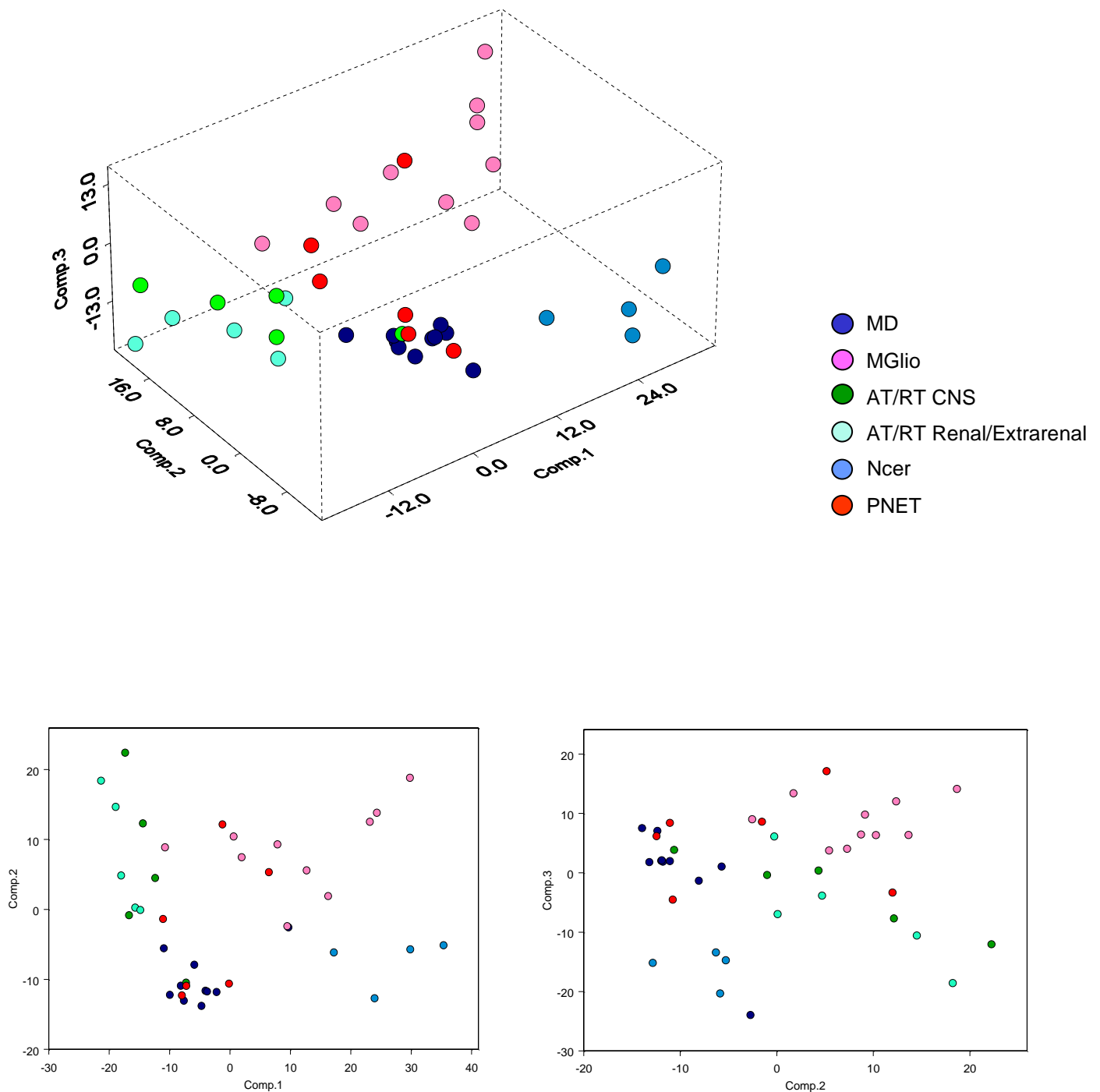
Number of features (genes) = 50

Sample	class	PCA Components		
		C1	C2	C3
Brain_MD_12	0	2.61456	-0.47658	-1.56685
Brain_MD_61	0	0.589267	-2.28945	-3.49189
Brain_MD_15	0	0.514632	-2.42775	-2.55942
Brain_MD_57	0	1.407238	-2.40925	-3.76104
Brain_MD_33	0	1.383655	-1.74997	-3.62516
Brain_MD_64	0	2.335573	-2.37497	-5.41547
Brain_MD_17	0	1.560329	-1.43053	-2.69519
Brain_MD_62	0	3.96E-05	-1.58512	-3.06371
Brain_MD_63	0	1.527174	-0.76173	-1.79533
Brain_MD_32	0	1.986229	-2.85936	-4.95683
Brain_MGlio_1	1	-2.15129	3.04742	1.340331
Brain_MGlio_2	1	-3.55696	4.648686	2.546523

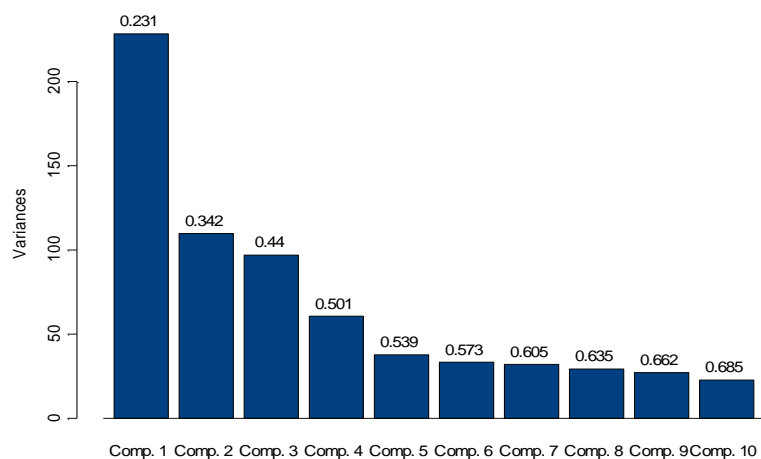
Brain_MGlio_3	1	-3.39718	3.514346	2.550463
Brain_MGlio_4	1	-2.48527	3.081104	1.062452
Brain_MGlio_5	1	-2.00525	2.728528	2.310731
Brain_MGlio_6	1	-2.41915	3.08049	2.653129
Brain_MGlio_7	1	-1.6897	2.889009	2.059978
Brain_MGlio_8	1	1.365032	0.481712	2.542437
Brain_MGlio_9	1	-4.58491	4.822834	1.92024
Brain_MGlio_10	1	-2.95011	3.399331	1.769402
Brain_Rhab_1	2	5.464601	-1.49409	3.144224
Brain_Rhab_2	3	5.292179	-1.73967	2.896035
Brain_Rhab_3	3	5.55369	-1.31506	2.274904
Brain_Rhab_4	2	2.600889	0.814271	3.167324
Brain_Rhab_5	3	4.950201	-0.51508	2.221833
Brain_Rhab_6	3	4.77626	-0.64909	4.316053
Brain_Rhab_7	3	3.392867	-0.45345	2.021606
Brain_Rhab_8	2	3.910798	-0.87741	1.990972
Brain_Rhab_9	2	3.089461	-1.14414	2.409917
Brain_Rhab_10	2	3.888992	-0.48271	2.882512
Brain_Ncer_1	4	-6.87929	-4.91617	1.099918
Brain_Ncer_2	4	-7.05358	-7.5747	1.151357
Brain_Ncer_3	4	-7.71348	-7.0406	1.830991
Brain_Ncer_4	4	-6.74515	-6.45202	2.529409
Brain_PNET_1	5	0.38724	1.541391	-2.41529
Brain_PNET_2	5	0.774589	1.253634	-0.21386
Brain_PNET_3	5	-2.69406	6.137626	-3.44386
Brain_PNET_4	5	-0.40094	2.043733	-1.7762
Brain_PNET_5	5	0.893357	1.073953	-1.81399
Brain_PNET_6	5	0.242446	0.431521	-3.51943
Brain_PNET_7	5	-2.29284	4.453621	-6.16835
Brain_PNET_8	5	-1.48216	3.575675	-2.41087

Dataset A1 (40 samples) – highly varying genes

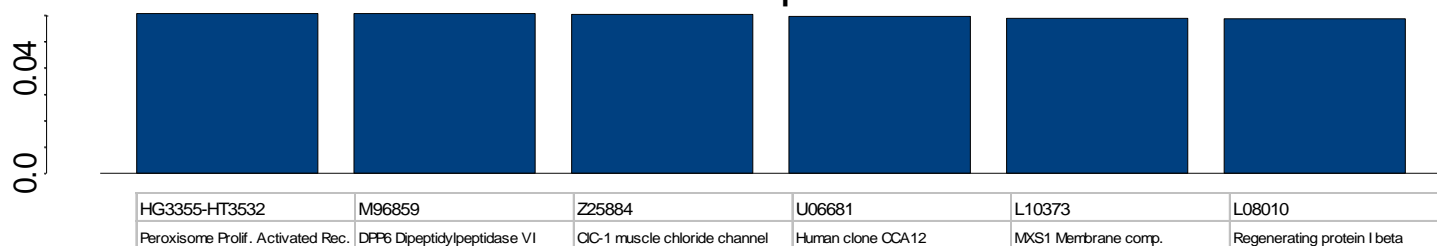
Two of the supratentorial PNETs are pineoblastomas, which historically have been inconsistently included in the PNET category. To study the difference it will make to exclude them we repeated the same PCA analysis of highly varying and marker genes but with the 2 pineoblastomas excluded (6 rather than 8 PNETs: dataset A1). The results are very similar as before.



Relative Importance of Principal Components

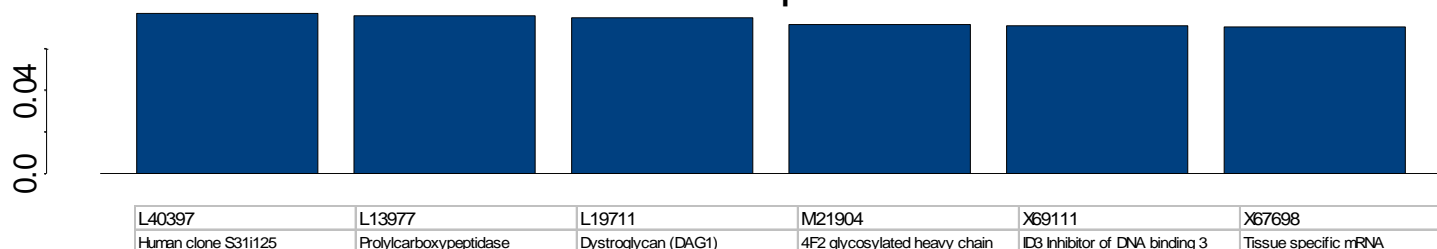


Comp. 1



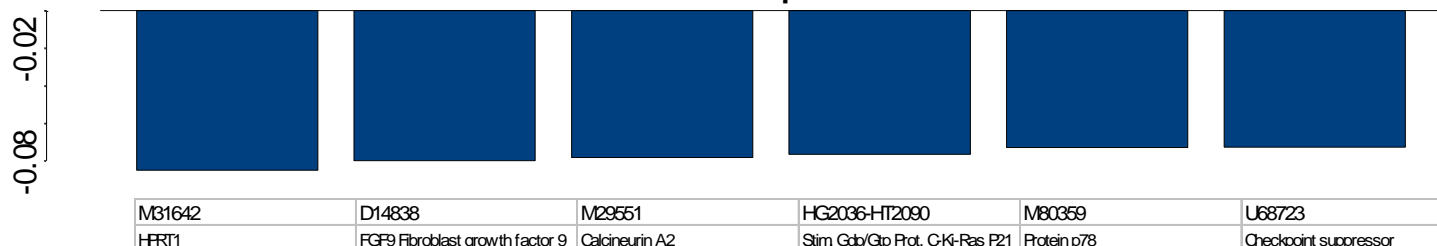
HG3355-HT3532	M96859	Z25884	U06681	L10373	L08010
Peroxisome Prolif. Activated Rec.	DPP6 Dipeptidylpeptidase VI	CIC-1 muscle chloride channel	Human clone CCA12	MXS1 Membrane comp.	Regenerating protein I beta

Comp. 2



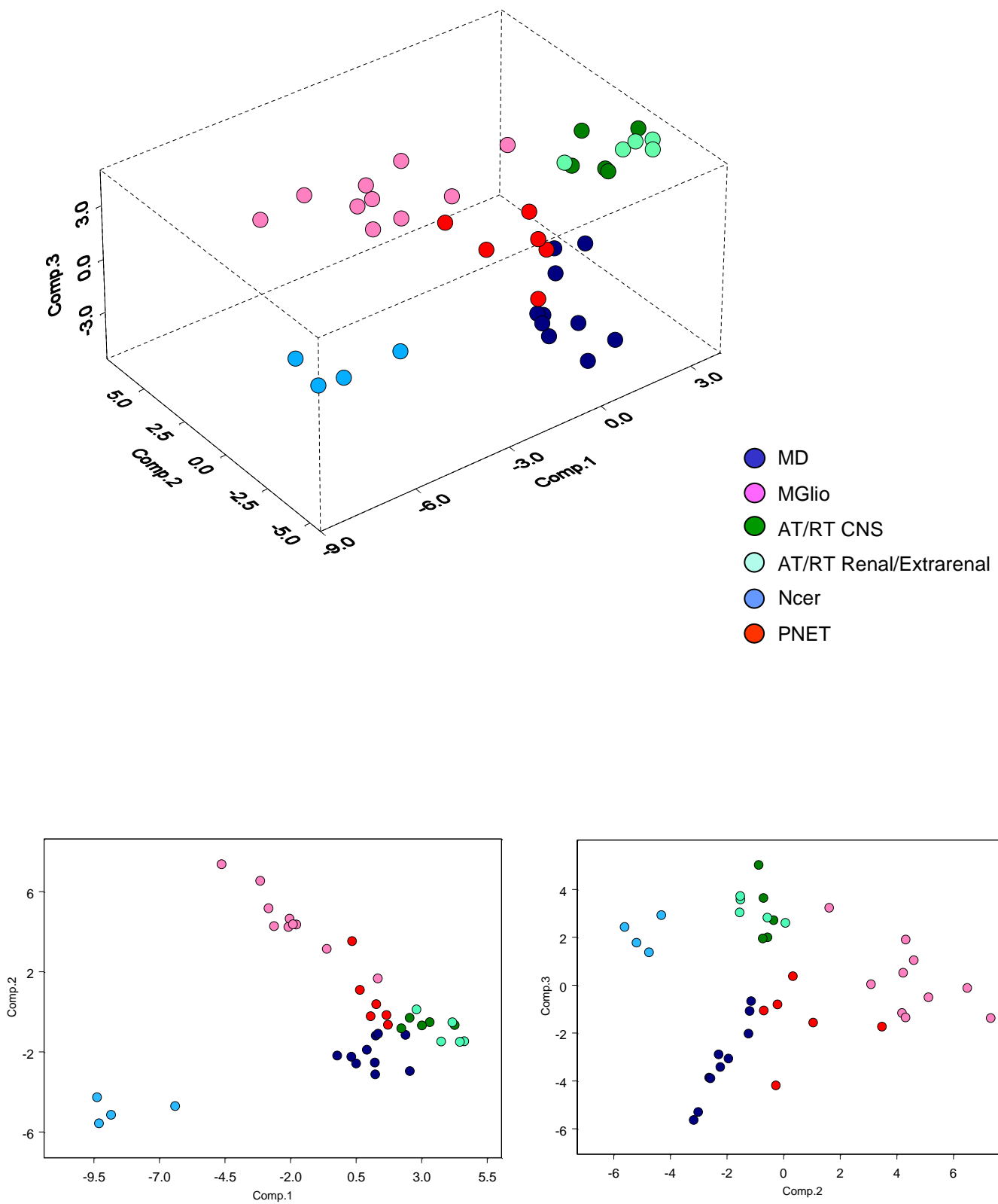
L40397	L13977	L19711	M21904	X69111	X67698
Human clone S311125	Prolylcarboxypeptidase	Dystroglycan (DAG1)	4F2 glycosylated heavy chain	ID3 Inhibitor of DNA binding 3	Tissue specific mRNA

Comp. 3

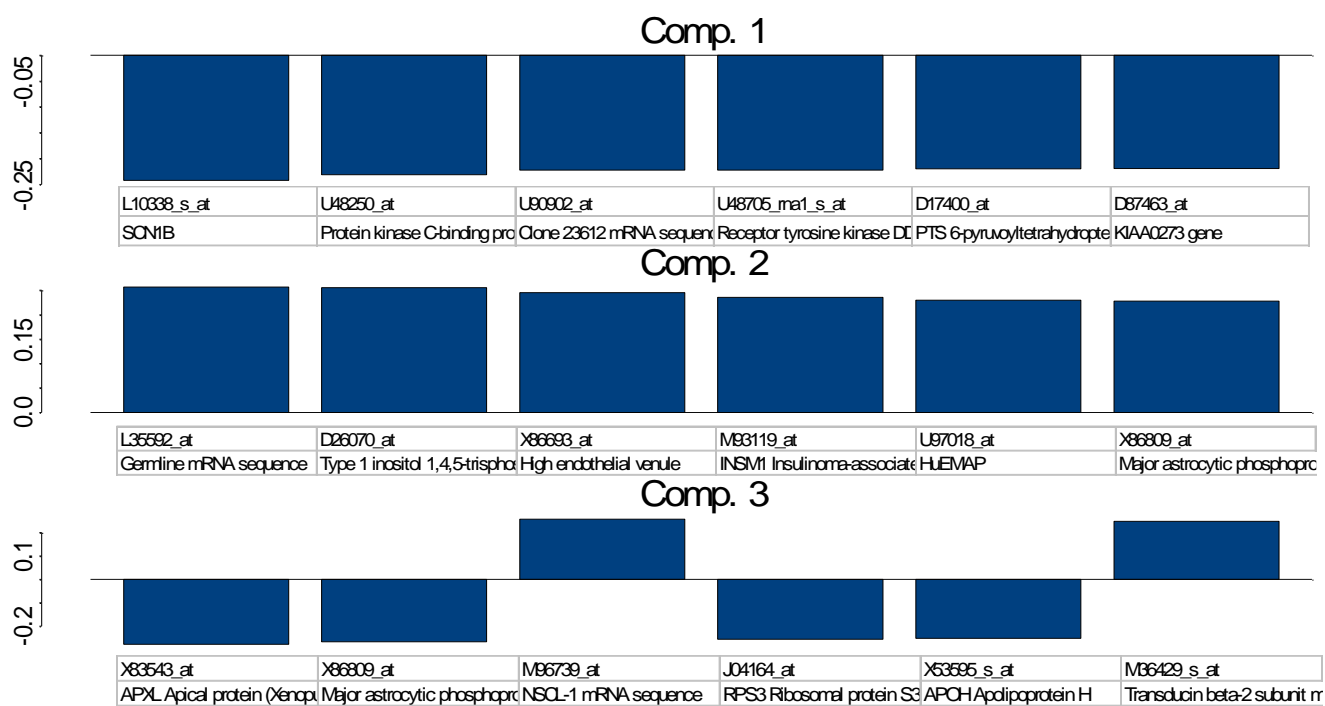
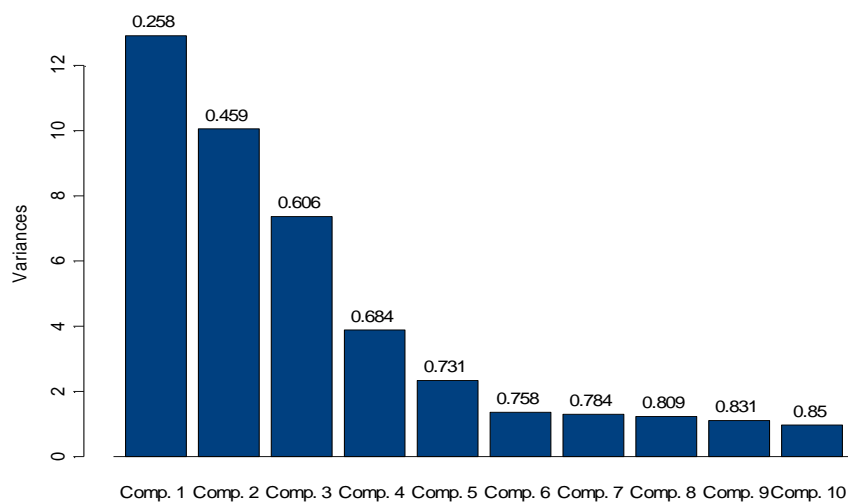


M31642	D14838	M29551	HG2036-HT2090	M80359	U68723
HPRT1	FGF9 Fibroblast growth factor 9	Calcineurin A2	Stim Gdp/Gtp Prot. C-K-Ras P21	Protein p78	Checkpoint suppressor

Dataset A1 (40 samples) – class marker genes

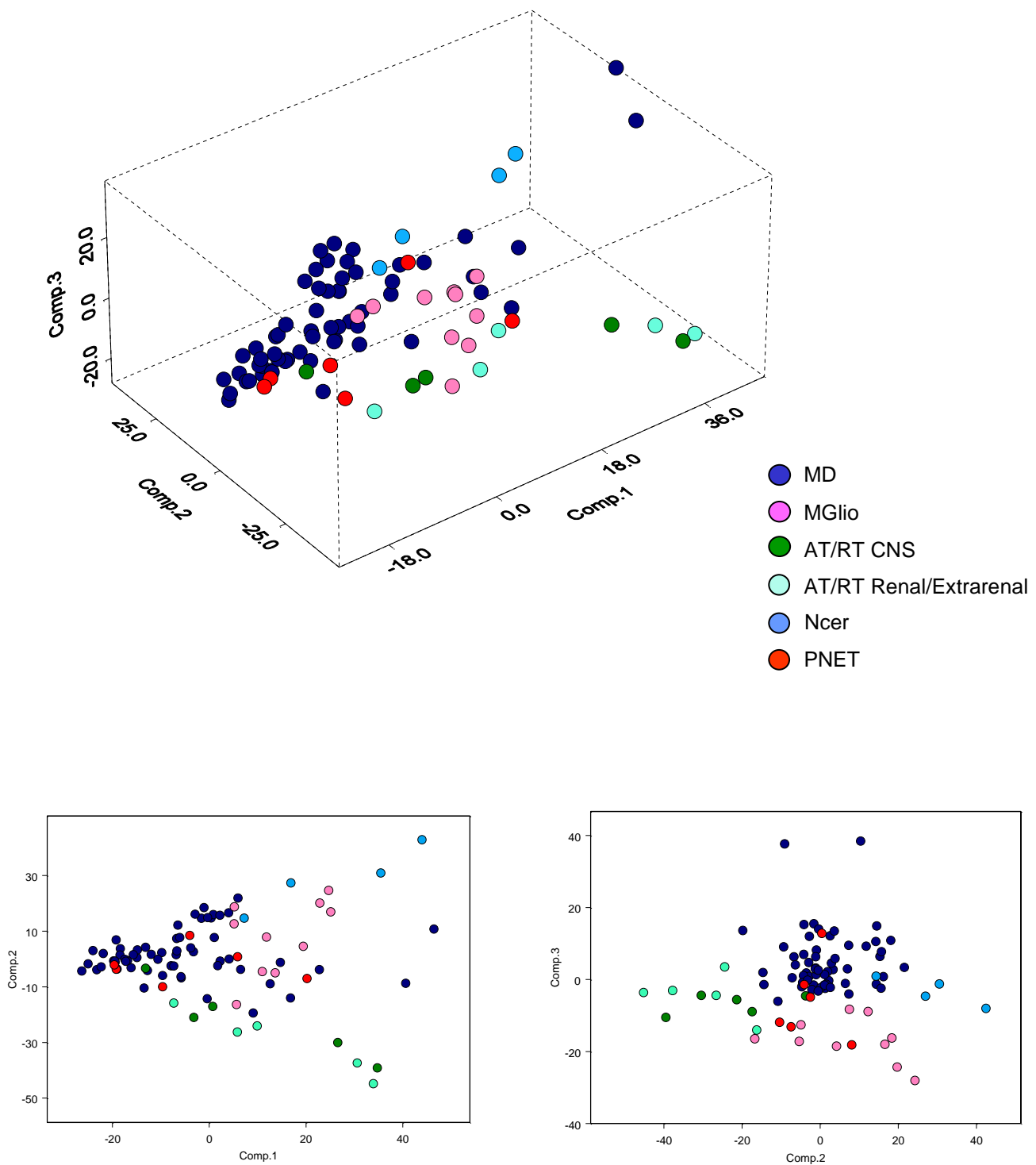


Relative Importance of Principal Components

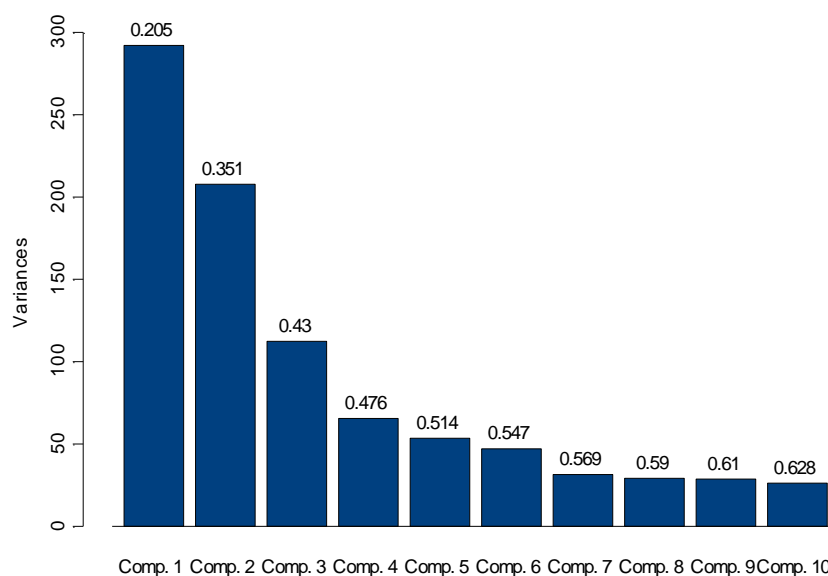


Dataset A2 (90 samples) – highly varying genes

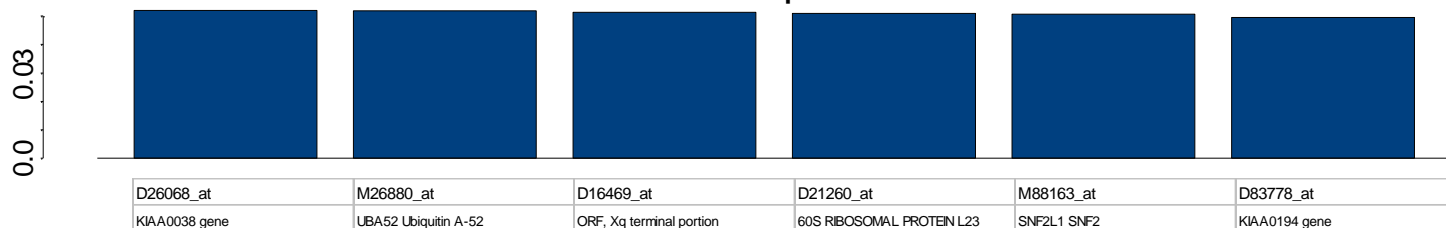
To test whether inclusion of a larger number of medulloblastomas might lessen the distinctions noted in Dataset A, 50 more medulloblastoma samples were added and the PCA analysis for highly varying and marker genes repeated.



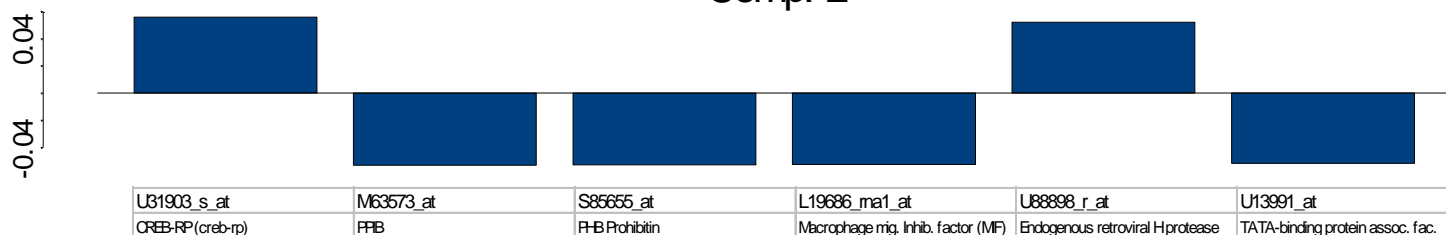
Relative Importance of Principal Components



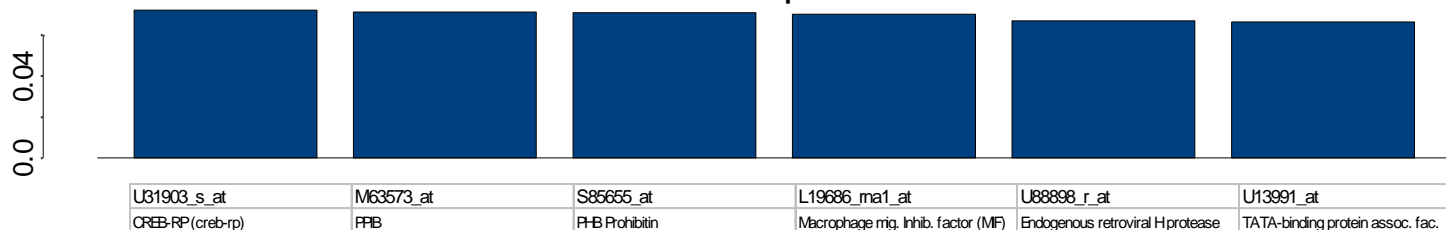
Comp. 1



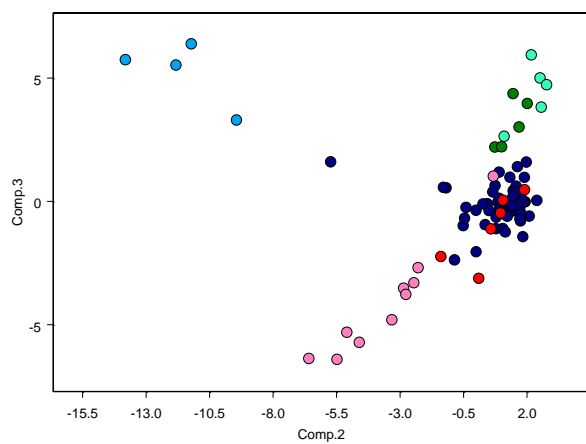
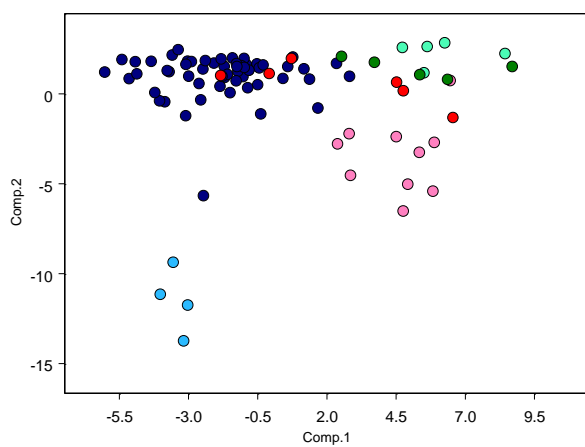
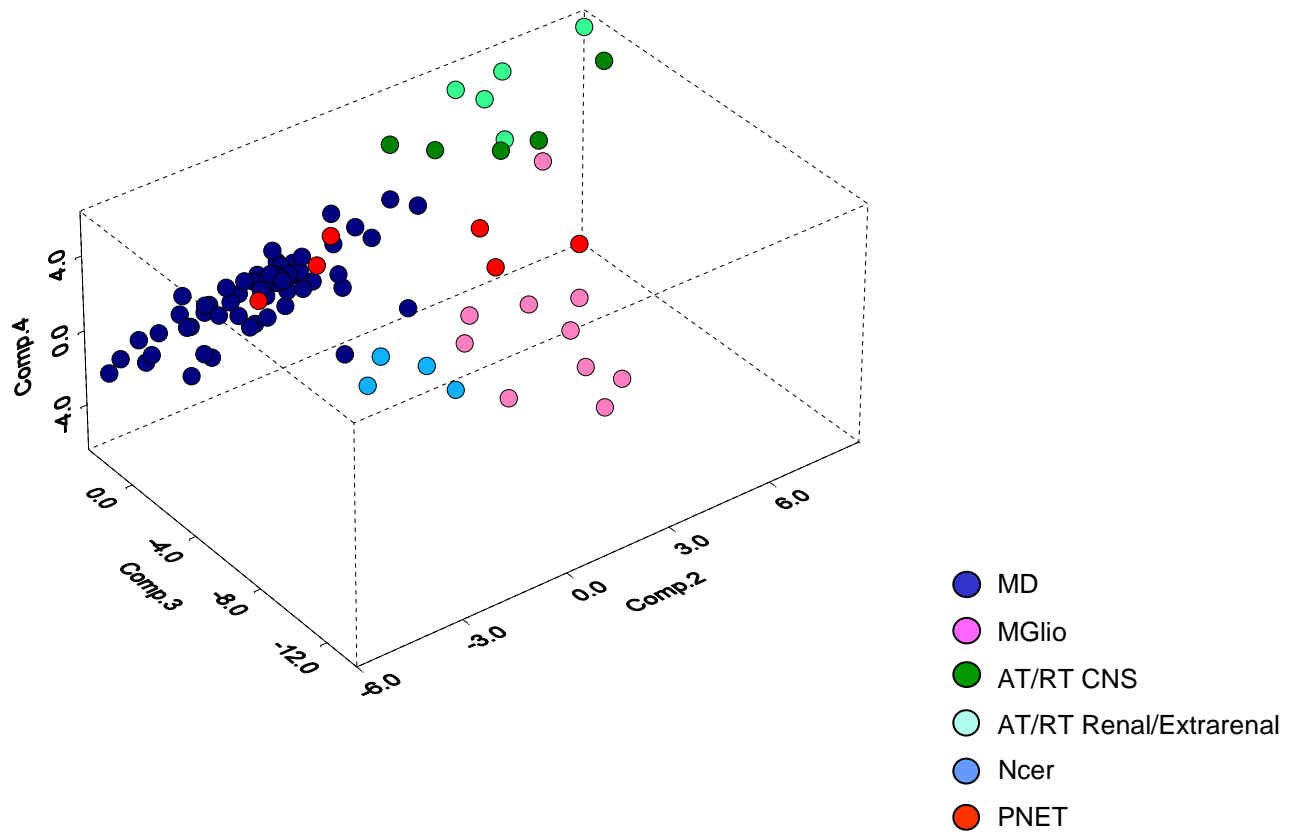
Comp. 2



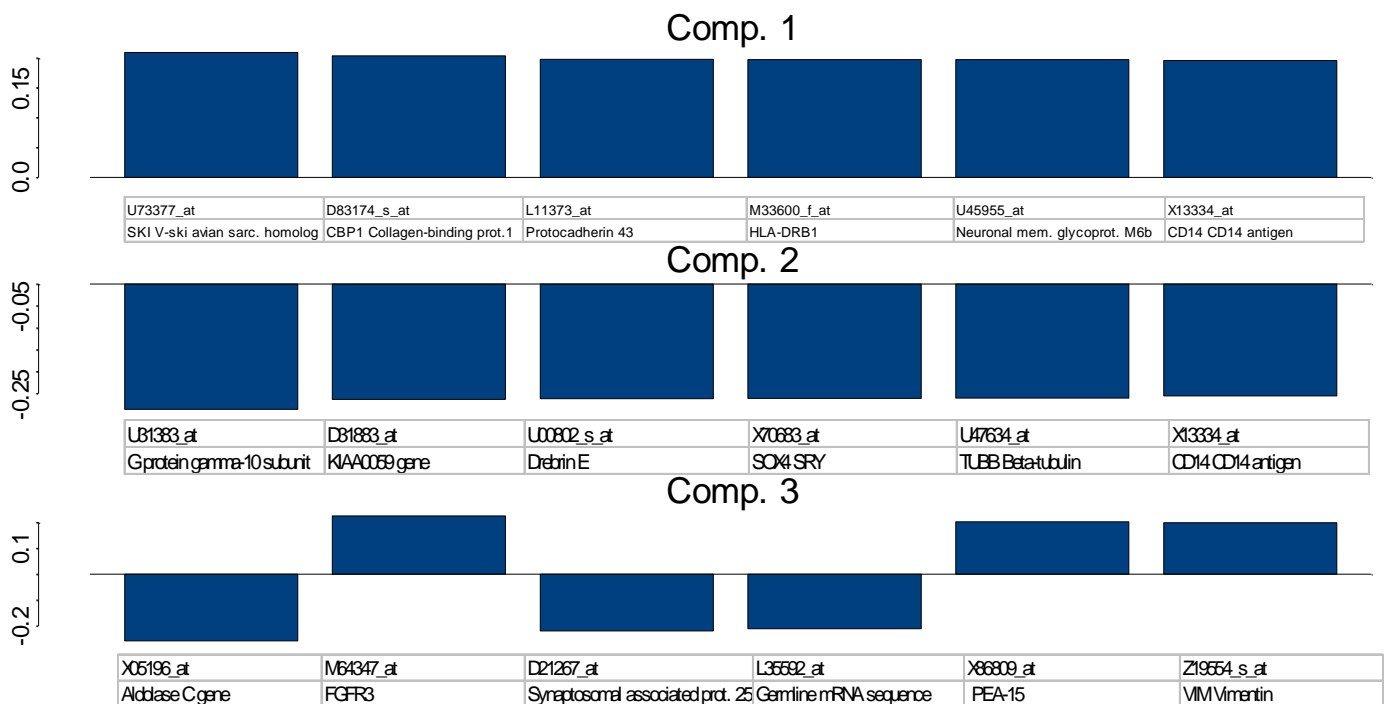
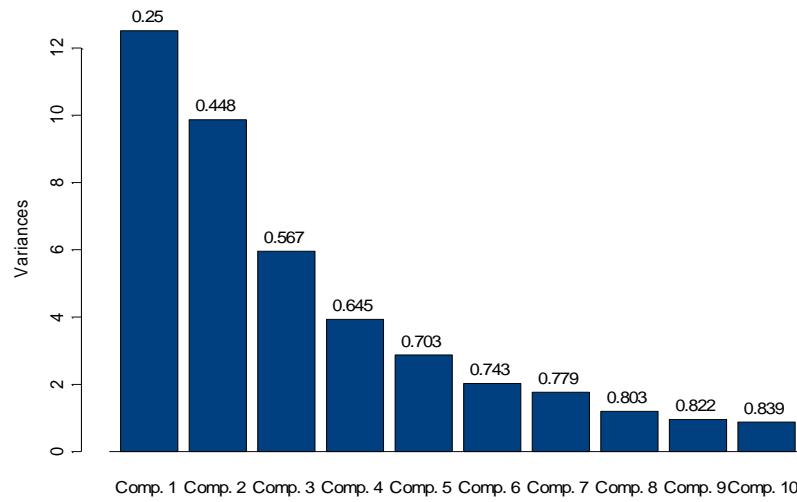
Comp. 3



Dataset A2 (90 samples) – class marker genes

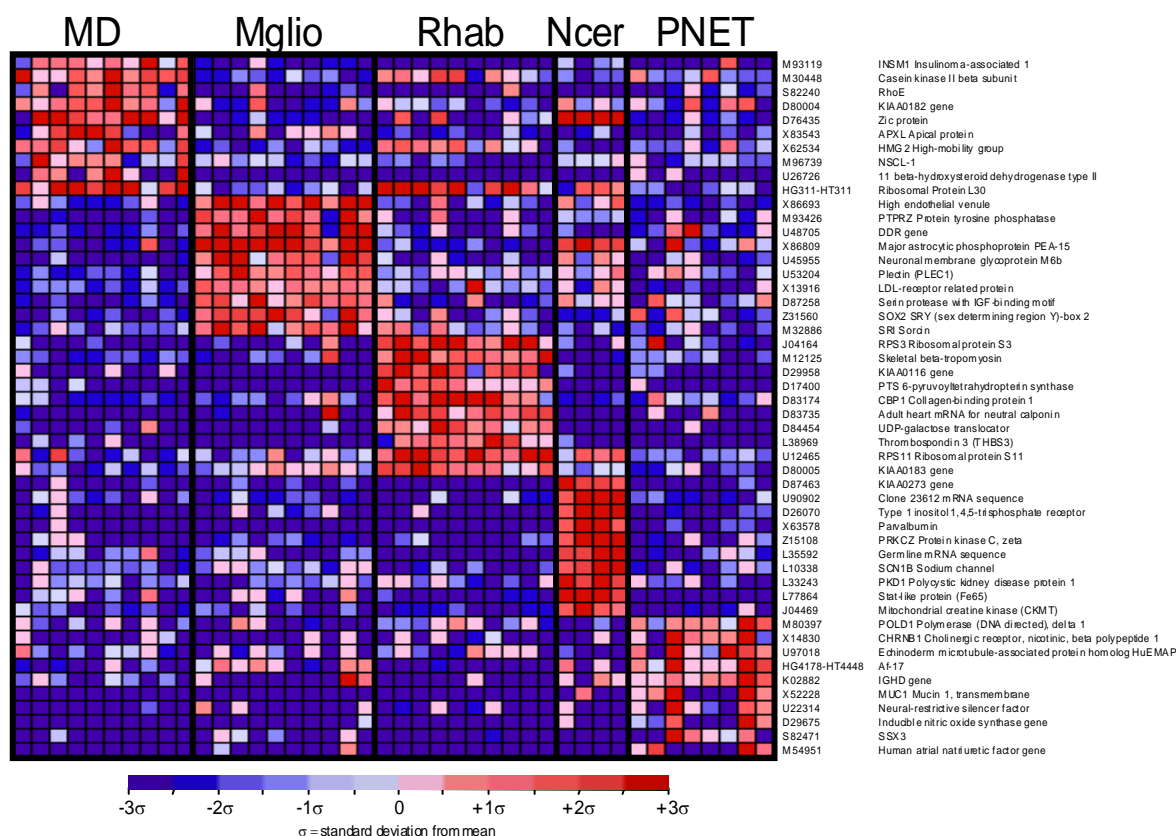


Relative Importance of Principal Components



Multiple tumor class markers

This picture shows the top 10 markers per class as sorted by their signal to noise ratios as described in [Gene marker selection](#) section. The table below shows the top 100 markers for each tumor class including the permutation test values (see [Permutation-based neighborhood analysis for marker gene](#)).



To 100 Marker Genes for each Tumor Class

Selected by signal-to-noise (median) ratio

Values thresholded to 20 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

Dataset A

Class 0 = High in Medulloblastoma, low in others

Class 1 = High in Malignant Glioma, low in others

Class 2 = High in AT/RT (renal-extrarenal), low in others

Class 3 = High in Normal Cerebellum, Low in others

Class 4 = High in PNET, low in others

Permutation test					Marker genes	
Distinction	Distance	Perm 1%	Perm 5%	Median 50%	Feature	Desc
class 0	0.96694607	1.0144908	0.8333578	0.6280173	M93119_at	INSM1 Insulinoma-associated 1 (symbol provisional)
class 0	0.9096911	0.8600172	0.7669801	0.5740431	M30448_s_at	Casein kinase II beta subunit mRNA
class 0	0.90010124	0.85051423	0.7251496	0.5494933	S82240_at	RhoE
class 0	0.832689	0.84354156	0.7071885	0.5292253	U44060_at	Homeodomain protein (Prox 1) mRNA
class 0	0.83225346	0.8009565	0.68034023	0.5169537	D80004_at	KIAA0182 gene, partial cds
class 0	0.7492524	0.7835767	0.6664746	0.5046996	D76435_at	Zic protein
class 0	0.7383032	0.77384007	0.6535448	0.4954919	X83543_at	APXL Apical protein (Xenopus laevis-like)
class 0	0.73376894	0.7426002	0.6453689	0.48881397	X62534_s_at	HMG2 High-mobility group (nonhistone chromosomal) protein 2
class 0	0.73127395	0.73893577	0.637871	0.48173288	M96739_at	NSCL-1 mRNA sequence
class 0	0.71544206	0.7368223	0.63101006	0.4792574	U26726_at	11 beta-hydroxysteroid dehydrogenase type II mRNA
class 0	0.70604366	0.7280188	0.6192388	0.47128072	HG311-HT311_at	Ribosomal Protein L30
class 0	0.66780347	0.7201379	0.6169249	0.4645534	X53331_at	MGP Matrix protein gla
class 0	0.6607844	0.7197015	0.6104924	0.46066648	M14483_rna1_s_at	PTMA gene extracted from Human prothymosin alpha mRNA
class 0	0.6518439	0.707669	0.6029651	0.4572201	Z69915_at	mRNA (clone ICRFp507L1876)
class 0	0.64655346	0.7015992	0.59863406	0.45266396	L00022_s_at	IG EPSILON CHAIN C REGION
class 0	0.64252174	0.6809993	0.5968048	0.4494678	U31382_at	G protein gamma-4 subunit mRNA
class 0	0.63783944	0.6795276	0.59584016	0.4467352	Z23064_at	HNRPG Heterogeneous nuclear ribonucleoprotein G
class 0	0.6361946	0.67848146	0.5931882	0.44210848	D82345_at	NB thymosin beta
class 0	0.60254765	0.6678708	0.5887515	0.43895075	U05012_s_at	NTRK3 Neurotrophic tyrosine kinase, receptor, type 3 (TrkC)
class 0	0.5855725	0.6635907	0.58740836	0.43586975	X87852_at	SEX gene
class 0	0.5815485	0.6612682	0.58239955	0.43329534	HG1612-HT1612_at	Macmarcks
class 0	0.5815067	0.65672636	0.58039063	0.43086103	U32315_at	Syntaxin 3 mRNA
class 0	0.5648414	0.65436137	0.5758365	0.42883328	X05855_s_at	EEF1G Translation elongation factor 1 gamma
class 0	0.56469566	0.653728	0.5696867	0.4266686	X13546_rna1_at	Put. HMG-17 protein gene extracted from Human HMG-17 gene for non-histone chromosomal protein HMG-17
class 0	0.5588449	0.6535496	0.56948066	0.4244893	M19720_rna2_at	L-myc gene (L-myc protein) extracted from Human L-myc protein gene
class 0	0.5560505	0.65252143	0.56585836	0.4232427	L33930_s_at	CD24 signal transducer mRNA and 3' region
class 0	0.550244	0.6516489	0.56495583	0.4211157	L06797_s_at	PROBABLE G PROTEIN-COUPLED RECEPTOR LCR1 HOMOLOG
class 0	0.53802216	0.6500193	0.5646127	0.41931587	M23613_at	NPM1 Nucleophosmin (nucleolar phosphoprotein B23, numatrin)
class 0	0.53711265	0.64934963	0.558147	0.41818365	X02404_at	CALCB Calcitonin-related polypeptide, beta
class 0	0.53183836	0.6488311	0.5571701	0.41529867	L10838_at	PRE-MRNA SPLICING FACTOR SRP20
class 0	0.5283321	0.6434776	0.5570184	0.41269392	S82024_at	SCG10
class 0	0.5268076	0.64197075	0.55182576	0.4111174	M11433_at	RBP1 Cellular retinol-binding protein
class 0	0.51967216	0.64013463	0.5506556	0.4098725	HG3088-HT3263_at	Splicing Factor Sc35, Alt Splice Form 3
class 0	0.5175459	0.63753295	0.5503224	0.40737128	U28686_at	Putative RNA binding protein RNPL mRNA
class 0	0.50714874	0.6362442	0.5472199	0.40469232	L40386_s_at	DP2 (Humdp2) mRNA
class 0	0.5039181	0.6355672	0.5450284	0.4028947	Z11502_at	ANNEXIN XIII
class 0	0.50178665	0.6329165	0.54309994	0.40131387	X55733_at	EUKARYOTIC INITIATION FACTOR 4B
class 0	0.50157803	0.6304573	0.54162425	0.4005091	HG4318-HT4588_s_at	Lim-Domain Transcription Factor Lim-1
class 0	0.5009597	0.6291464	0.54044646	0.39848644	U30521_at	FRAP FK506 binding protein 12-rapamycin associated protein
class 0	0.5009336	0.623241	0.53552544	0.39617687	X74330_at	PRIM1 DNA primase polypeptide 1 (49kD)

class 0	0.5008229	0.62009233	0.535334	0.3954302	X74262_at	RETINOBLASTOMA BINDING PROTEIN P48
class 0	0.49198747	0.61935806	0.5346635	0.39358965	U70862_at	Nuclear factor I B3 mRNA
class 0	0.48500997	0.618592	0.5344887	0.3920796	U79255_at	X11 protein mRNA, partial cds
class 0	0.48494342	0.6179998	0.5340308	0.390908	HG613-HT613_at	Ribosomal Protein S12
class 0	0.4806928	0.61442095	0.5297601	0.38958547	X07438_s_at	DNA for cellular retinol binding protein (CRBP) exons 3 and 4
class 0	0.47977164	0.61286503	0.52914625	0.38840055	X56465_at	Znf6 mRNA for zinc finger transcription factor
class 0	0.47824362	0.61250454	0.52643156	0.38764188	U47414_at	Cyclin G2 mRNA
class 0	0.4758209	0.6121982	0.5250884	0.38578293	L37043_at	CSNK1E Casein kinase 1, epsilon
class 0	0.47532186	0.6104803	0.5240915	0.38462755	U02031_at	Sterol regulatory element binding protein-2 mRNA
class 0	0.4750675	0.60672534	0.52364206	0.38363606	U21090_at	DNA polymerase delta small subunit mRNA
class 0	0.4699813	0.5989572	0.5215285	0.3821684	U26312_s_at	Heterochromatin protein HP1Hs-gamma mRNA
class 0	0.46889964	0.5977149	0.51875275	0.38159114	M96740_at	HELIX-LOOP-HELIX PROTEIN 2
class 0	0.46830228	0.59540033	0.51809573	0.38079777	D55716_at	DNA REPLICATION LICENSING FACTOR CDC47 HOMOLOG
class 0	0.467083	0.5951792	0.5178332	0.37955925	X52966_at	RPL35A Ribosomal protein L35a
class 0	0.4641961	0.5948711	0.51752603	0.37854284	Y09836_at	3'UTR of unknown protein
class 0	0.45807302	0.59466755	0.5157532	0.37741286	U43885_at	Grb2-associated binder-1 mRNA
class 0	0.45774597	0.5933097	0.5148932	0.37650257	X69398_at	CD47 CD47 antigen (Rh-related antigen, integrin-associated signal transducer)
class 0	0.45630834	0.58962584	0.5139473	0.37543085	X76029_at	NEUROMEDIN U-25 PRECURSOR
class 0	0.45350492	0.5884134	0.5138038	0.3743853	M13241_at	N-MYC PROTO-ONCOGENE PROTEIN
class 0	0.4528938	0.5882923	0.5130492	0.37293357	D28423_at	Pre-mRNA splicing factor SRp20, 5'UTR (sequence from the 5'cap to the start codon)
class 0	0.45122162	0.58495116	0.5122993	0.3718915	U73304_rna1_at	CB1 cannabinoid receptor (CNR1) gene
class 0	0.4486972	0.5828123	0.5106254	0.37171924	U17195_at	A-kinase anchor protein (AKAP100) mRNA
class 0	0.44740826	0.58226925	0.5075315	0.3709006	M93415_at	ACVR2 Activin A receptor, type II
class 0	0.44720528	0.58122206	0.50601745	0.36981982	M93650_at	Paired box gene (PAX6) homologue
class 0	0.44666263	0.5794936	0.50538707	0.36805078	X85545_at	Protein kinase, PKX1
class 0	0.44536316	0.5779683	0.50191903	0.36731505	S76475_at	NTRK3 Neurotrophic tyrosine kinase, receptor, type 3 (TrkC)
class 0	0.44418713	0.5775526	0.5014425	0.366797	U00802_s_at	Drebrin E
class 0	0.44110203	0.57679415	0.49992916	0.36611393	M60299_at	Alpha-1 collagen type II gene, exons 1, 2 and 3
class 0	0.43829215	0.5762043	0.49976385	0.36523366	U16954_at	(AF1q) mRNA
class 0	0.43407452	0.5753858	0.49807757	0.36481872	X99657_at	Protein containing SH3 domain, SH3GL2
class 0	0.43262523	0.5744567	0.49757218	0.3631796	X76132_at	DCC Deleted in colorectal carcinoma
class 0	0.43084678	0.5735875	0.4974876	0.36188462	U85193_at	Nuclear factor I-B2 (NFIB2) mRNA
class 0	0.43024966	0.5730583	0.4963776	0.36167425	M82919_at	GABRB3 Gamma-aminobutyric acid (GABA) A receptor, beta 3
class 0	0.42832303	0.5729875	0.4960867	0.36009976	M27691_at	CAMP-RESPONSE ELEMENT BINDING PROTEIN
class 0	0.42654628	0.57097447	0.4958619	0.3594544	L22005_at	UBIQUITIN-CONJUGATING ENZYME E2-CDC34 COMPLEMENTING
class 0	0.42234343	0.56863326	0.49558508	0.35884386	L76159_at	FRG1 mRNA
class 0	0.41938537	0.56803626	0.4927454	0.35822877	U39226_at	Myosin VIIA (USH1B) mRNA
class 0	0.41719937	0.5676926	0.49170917	0.3565564	U38810_at	Mab-21 cell fate-determining protein homolog (CAGR1) mRNA
class 0	0.41676784	0.5640895	0.49088645	0.35503575	U25034_s_at	Neuronatin alpha mRNA
class 0	0.41542533	0.56285304	0.48999164	0.35428312	U24576_at	Breast tumor autoantigen mRNA, complete sequence
class 0	0.4132539	0.5625111	0.4890851	0.35347974	U23803_at	Heterogeneous ribonucleoprotein A0 mRNA
class 0	0.4109598	0.56244606	0.48815268	0.35334146	M83822_at	Beige-like protein (BGL) mRNA, partial cds

class 0	0.4100799	0.562175	0.48733872	0.3529259	U09087_s_at	Thymopoietin beta mRNA
class 0	0.4079055	0.5613772	0.48599333	0.35199758	M91670_at	Ubiquitin carrier protein (E2-EPF) mRNA
class 0	0.4040995	0.5606459	0.48535928	0.350769	U25789_at	Ribosomal protein L21 mRNA
class 0	0.40371194	0.5600623	0.48488784	0.35030034	M62843_at	PARANEOPLASTIC ENCEPHALOMYELITIS ANTIGEN HUD
class 0	0.40356442	0.55962336	0.483534	0.34974435	U09953_at	RPL9 Ribosomal protein L9
class 0	0.40279832	0.5594289	0.48338398	0.34923753	U31814_at	Transcriptional regulator homolog RPD3 mRNA
class 0	0.4027233	0.55762196	0.48283297	0.34894606	X64229_at	DEK PROTEIN
class 0	0.40206712	0.55735785	0.4824586	0.34788424	U54999_at	LGN protein mRNA
class 0	0.40179467	0.5565962	0.48214018	0.34594992	X70683_at	SOX4 SRY (sex determining region Y)-box 4
class 0	0.3995434	0.55561066	0.48126557	0.34553674	U07919_at	ALDH6 Aldehyde dehydrogenase 6
class 0	0.39766783	0.5554421	0.48065114	0.34511587	M64358_at	Rhom-3 gene, exon
class 0	0.39684495	0.5548959	0.48008677	0.3444539	U19878_at	Transmembrane protein mRNA
class 0	0.39449787	0.55458045	0.47979704	0.3438485	AFFX-HUMRGE/M10098_M_at	AFFX-HUMRGE/M10098_M_at (endogenous control)
class 0	0.39352742	0.5533359	0.47933868	0.342717	J03827_at	DbpB-like protein mRNA
class 0	0.39003983	0.5530556	0.4786078	0.3423478	U61145_at	Enhancer of zeste homolog 2 (EZH2) mRNA
class 0	0.38979462	0.5514548	0.47775155	0.34202746	HG662-HT662_at	Epstein-Barr Virus Small Rna-Associated Protein
class 0	0.38538435	0.55026263	0.4775768	0.34102225	M73047_at	TPP2 Tripeptidyl peptidase II
class 0	0.38463777	0.5502042	0.4763407	0.34075052	D85131_s_at	Myc-associated zinc-finger protein of human islet
class 1	1.6520017	0.9831643	0.84544426	0.6230137	X86693_at	High endothelial venule
class 1	1.2436218	0.88150144	0.7559189	0.5795857	M93426_at	PTPRZ Protein tyrosine phosphatase, receptor-type, zeta polypeptide
class 1	1.2317128	0.86047184	0.70928395	0.5539352	U48705_rna1_s_at	Receptor tyrosine kinase DDR gene
class 1	1.2259983	0.8433512	0.68909335	0.5358038	X86809_at	Major astrocytic phosphoprotein PEA-15
class 1	1.214929	0.8281318	0.6849929	0.5217813	U45955_at	Neuronal membrane glycoprotein M6b mRNA, partial cds
class 1	1.2095517	0.79365546	0.6711517	0.510208	U53204_at	Plectin (PLEC1) mRNA
class 1	1.2026114	0.7930142	0.6636111	0.50219953	X13916_at	LDL-receptor related protein
class 1	1.1869695	0.77752584	0.65392506	0.49156818	D87258_at	Cancellous bone osteoblast mRNA for serin protease with IGF-binding motif
class 1	1.1676904	0.7709572	0.6380772	0.48596418	Z31560_s_at	SOX2 SRY (sex determining region Y)-box 2
class 1	1.1604098	0.76437885	0.63309973	0.47967565	M32886_at	SRI Sorcin
class 1	1.1558465	0.761579	0.62335235	0.47505242	D16181_at	PMP2 Peripheral myelin protein 2
class 1	1.1461633	0.76131815	0.61996907	0.4696402	U48250_at	Protein kinase C-binding protein
class 1	1.1236995	0.748406	0.61082774	0.46480379	D63878_at	RACK17 mRNA, partial cds
class 1	1.0904534	0.74650514	0.6089377	0.46025795	K03189_f_at	PROBABLE PROTEIN DISULFIDE ISOMERASE ER-60 PRECURSOR
class 1	1.0883032	0.74052924	0.6014309	0.4566901	U52155_at	Chorionic gonadotropin (hcg) beta subunit mRNA
class 1	1.0646937	0.73030424	0.598755	0.45272368	L11373_at	Inward rectifier potassium channel Kir1.2 (Kir1.2) mRNA, partial cds
class 1	1.0544381	0.72683483	0.5921049	0.4487451	M21551_rna1_at	Protocadherin 43 mRNA for abbreviated PC43
class 1	1.0439421	0.7250801	0.5877407	0.4458085	Z50022_at	Neuromedin B mRNA
class 1	1.0364326	0.7139357	0.58473366	0.44330922	HG620-HT620_at	Surface glycoprotein
class 1	1.0299566	0.7118054	0.5835643	0.4414981	M21904_at	Tyrosine Phosphatase, Epsilon
class 1	1.026406	0.70610374	0.58319974	0.44004935	D38522_at	MDU1 Antigen identified by monoclonal antibodies 4F2, TRA1.10, TROP4, and T43
class 1	1.0112586	0.7057017	0.5741648	0.4358882	Z50781_at	KIAA0080 gene, partial cds
class 1	1.0110288	0.703144	0.5740614	0.43265316	X54673_at	Leucine zipper protein
class 1						SLC6A1 Solute carrier family 6 (neurotransmitter transporter, GABA), member 1

class 1	0.9994149	0.6964176	0.5672535	0.43000817	M63623_at	MOG Myelin oligodendrocyte glycoprotein
class 1	0.99073607	0.68848604	0.56642866	0.42853382	M97796_s_at	ID2 Inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
class 1	0.98774695	0.68189645	0.56605524	0.4269275	L22214_at	ADORA1 Adenosine receptor A1
class 1	0.9855738	0.6807615	0.564807	0.42519456	M23254_at	CAPN2 Calpain, large polypeptide L2
class 1	0.98304486	0.6677363	0.5630516	0.42373672	S80905_f_at	PRB2 locus salivary proline-rich protein mRNA, clone cP7
class 1	0.9786299	0.6667506	0.56217647	0.4208039	M32304_s_at	TIMP2 Tissue inhibitor of metalloproteinase 2
class 1	0.97468305	0.66582626	0.5576709	0.41835085	U79272_at	Clone 23720 mRNA sequence
class 1	0.9741546	0.6632827	0.5567529	0.4152393	D25217_at	KIAA0027 gene, partial cds
class 1	0.97182816	0.6625792	0.5549602	0.41295442	U59877_s_at	Low-Mr GTP-binding protein (RAB31) mRNA
class 1	0.9682741	0.6590203	0.5520322	0.4113537	U07807_at	Metallothionein IV (MTIV) gene
class 1	0.9622625	0.6580188	0.54826355	0.4092988	D14689_at	NUCLEAR PORE COMPLEX PROTEIN NUP214
class 1	0.9585737	0.6577229	0.5479787	0.4066317	X98085_at	TNR Tenascin R (restrictin, janusin)
class 1	0.9569103	0.65765256	0.547915	0.40562418	D49817_at	Fructose 6-phosphate,2-kinase/fructose 2,6-bisphosphatase
class 1	0.95592564	0.6532676	0.543061	0.40421706	M16424_at	BETA-HEXOSAMINIDASE ALPHA CHAIN PRECURSOR
class 1	0.94691986	0.6528938	0.54168797	0.40255094	M62302_at	Growth/differentiation factor 1 (GDF-1) mRNA
class 1	0.9426242	0.65049857	0.5364322	0.4010224	L32961_at	4-AMINOBTYRATE AMINOTRANSFERASE, MITOCHONDRIAL PRECURSOR
class 1	0.9419619	0.65049285	0.53504795	0.4002328	S56151_s_at	HMFG
class 1	0.93842375	0.6503094	0.53349763	0.3990202	U90547_at	Ro/SSA ribonucleoprotein homolog (RoRet) mRNA
class 1	0.9366897	0.6481403	0.5331684	0.39706725	U76388_at	Steroidogenic factor 1 mRNA
class 1	0.93209916	0.64719605	0.53102833	0.39576787	L24559_at	POLA DNA polymerase alpha subunit
class 1	0.92833	0.64508885	0.5292084	0.39434457	D79999_at	KIAA0177 gene, partial cds
class 1	0.92812437	0.64227355	0.5264592	0.3926984	S73591_at	Brain-expressed HHCPA78 homolog [human, HL-60 acute promyelocytic leukemia cells, mRNA, 2704 nt]
class 1	0.9218112	0.64078844	0.52606547	0.39150074	X54637_at	TYK2 Protein-tyrosine kinase tyk2 (non-receptor)
class 1	0.92119294	0.63995963	0.5255806	0.39071545	U12707_s_at	WAS Wiskott-Aldrich syndrome (eczema-thrombocytopenia)
class 1	0.9184531	0.6391437	0.5252529	0.38848042	X75958_at	TrkB [alternatively spliced] [human, brain, mRNA, 1870 nt]
class 1	0.9155559	0.6370653	0.5248767	0.38763252	X04828_at	GNAI2 Guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2
class 1	0.90545046	0.63566005	0.52347106	0.38607994	S82297_at	BETA-2-MICROGLOBULIN PRECURSOR
class 1	0.8919348	0.63485956	0.52277017	0.384728	S45630_at	CRYAB Crystallin alpha-B
class 1	0.8909992	0.63345426	0.51961327	0.38367814	D13631_s_at	KIAA0006 gene
class 1	0.8867211	0.6329042	0.5160313	0.38311574	U80226_s_at	Gamma-aminobutyric acid transaminase mRNA, partial cds
class 1	0.8820352	0.63253194	0.5158146	0.3825838	L40407_at	Thyroid receptor interactor (TRIP9) gene
class 1	0.87914836	0.6318433	0.514657	0.38108054	U38980_at	PMS8 mRNA (yeast mismatch repair gene PMS1 homologue), partial cds (C-terminal region)
class 1	0.87837005	0.6309211	0.5133504	0.37988254	Y00796_at	ITGAL Integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1; alpha polypeptide)
class 1	0.87823164	0.62923473	0.5121836	0.37843332	J03040_at	SPARC SPARC/osteonectin
class 1	0.8760671	0.6286636	0.5113878	0.37749302	X55740_at	NT5 5' nucleotidase (CD73)
class 1	0.8760668	0.62734014	0.51018846	0.37679052	X92475_at	ITBA1 protein
class 1	0.87513036	0.6266346	0.5089304	0.3758733	U69263_at	Matrilin-2 precursor mRNA, partial cds
class 1	0.8745863	0.6246063	0.5083398	0.3742895	U55258_at	HBRAVO/Nr-CAM precursor (hBRAVO/Nr-CAM) gene

class 1	0.86930543	0.62451684	0.5073476	0.37309632	S50017_s_at	CNP 2',3'-cyclic nucleotide 3' phosphodiesterase
class 1	0.8662232	0.6241277	0.50710213	0.37164986	X76717_at	MT1L Metallothionein 1L
class 1	0.84933835	0.6234191	0.5067336	0.37121156	U28368_at	ID4 Inhibitor of DNA binding 4, dominant negative helix-loop-helix protein
class 1	0.8469514	0.62305886	0.5062817	0.37057114	X74794_at	CDC21 HOMOLOG
class 1	0.84659636	0.6230239	0.50575304	0.36982706	X75861_at	TEGT Testis enhanced gene transcript
class 1	0.8460863	0.6192708	0.50347847	0.36910838	M83233_at	TCF12 Transcription factor 12 (HTF4, helix-loop-helix transcription factors 4)
class 1	0.8459389	0.61905175	0.5014394	0.36789283	U73328_at	DLX7 Distal-less homeobox 7
class 1	0.8424674	0.6189927	0.5009286	0.36755785	M95936_s_at	AKT2 V-akt murine thymoma viral oncogene homolog 2
class 1	0.8386806	0.6188733	0.49939558	0.36682224	X85786_at	BINDING REGULATORY FACTOR
class 1	0.8372769	0.61818075	0.49734905	0.36528903	U14394_at	METALLOPROTEINASE INHIBITOR 3 PRECURSOR
class 1	0.8356367	0.6158615	0.49629536	0.36451015	D63486_at	KIAA0152 gene
class 1	0.83225477	0.6150287	0.4945158	0.36395043	Z68280_cds2_s_at	Erythrocyte adducin alpha subunit gene extracted from Human DNA sequence from cosmid L25A3, Huntington's Disease Region, chromosome 4p16.3 contains Human tetracycline transporter-like protein and erythrocyte adducin alpha subunit, multiple ESTs and a putative CpG island
class 1	0.83206296	0.6120459	0.4930411	0.36299255	U89335_cds2_at	NOTCH4 gene (notch4) extracted from Human HLA class III region containing notch4 (NOTCH4) gene, complete sequence
class 1	0.8239408	0.61198014	0.49286792	0.36167002	U00928_at	RNA-BINDING PROTEIN FUS/TLS
class 1	0.8234844	0.61135525	0.49238437	0.36087698	X00274_at	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DR ALPHA CHAIN PRECURSOR
class 1	0.82109606	0.6106854	0.49044895	0.36035484	M80244_at	INTEGRAL MEMBRANE PROTEIN E16
class 1	0.81676286	0.61059463	0.48990437	0.3597854	D49410_at	IL3RA Interleukin 3 receptor, alpha (low affinity)
class 1	0.8108157	0.6101961	0.48990065	0.35875112	L14813_at	CELL Carboxyl ester lipase like protein
class 1	0.80911857	0.61003023	0.48817563	0.35794193	M77016_at	TMOD Tropomodulin
class 1	0.8083867	0.60995644	0.48780048	0.35718402	Y08265_s_at	DAN26 protein, partial
class 1	0.80540043	0.6097752	0.48703083	0.35631403	M69023_at	Globin gene
class 1	0.8024884	0.6077933	0.4849392	0.35507387	M11749_at	THY-1 MEMBRANE GLYCOPROTEIN PRECURSOR
class 1	0.8011927	0.60732687	0.48423845	0.35416168	X59892_at	TRYPTOPHANYL-TRNA SYNTHETASE
class 1	0.7992353	0.6069737	0.48395756	0.3537728	HG987-HT987_at	Mac25
class 1	0.79868704	0.6057914	0.48377454	0.35269728	X83863_at	PTGER3 Prostaglandin E receptor 3 (subtype EP3) {alternative products}
class 1	0.7985791	0.6057221	0.48315245	0.35175043	X55666_at	Usf mRNA for late upstream transcription factor
class 1	0.7961344	0.604588	0.48308864	0.3515281	U30930_at	CGT UDP-galactose ceramide galactosyl transferase
class 1	0.79439163	0.6007642	0.4818679	0.35095778	U19261_at	Epstein-Barr virus-induced protein mRNA
class 1	0.79308397	0.59980625	0.48032397	0.35053596	L42601_f_at	KERATIN, TYPE II CYTOSKELETAL 6D
class 1	0.79107213	0.5997507	0.48028195	0.35026234	U46023_at	Xq28 mRNA
class 1	0.79009724	0.5991178	0.4783936	0.34935814	M10612_at	APOC2 Apolipoprotein C-II
class 1	0.78904396	0.5987258	0.4782325	0.34850967	U79528_s_at	Sigma receptor mRNA
class 1	0.788058	0.5983545	0.4774304	0.34845862	Z49825_s_at	HEPATOCYTE NUCLEAR FACTOR 4
class 1	0.7867659	0.59789264	0.4762901	0.34749532	D84145_at	WS-3 mRNA
class 1	0.78471535	0.59755576	0.47509775	0.34661785	Z11899_s_at	POU5F1 Octamer binding protein 3
class 1	0.7831359	0.59647435	0.47479355	0.34539047	D63135_at	ETS-like 30 kDa protein
class 1	0.78272295	0.5960181	0.47357324	0.3446057	U32680_at	CLN3 Ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-vogt disease)

class 1	0.78065175	0.594793	0.4729738	0.34366468	X55079_rna1_at	GAA gene extracted from Human lysosomal alpha-glucosidase gene exon 1
class 1	0.7805558	0.5940042	0.47264746	0.34325802	U94747_at	WD repeat protein HAN11 mRNA
class 2	1.5964093	0.9486641	0.8495439	0.62064433	J04164_at	RPS3 Ribosomal protein S3
class 2	1.5496515	0.87290615	0.777105	0.5718508	M12125_at	Skeletal beta-tropomyosin
class 2	1.5152686	0.827159	0.73881376	0.5467952	D17400_at	PTS 6-pyruvoyltetrahydropterin synthase
class 2	1.4285764	0.8085277	0.71054107	0.5315226	D29958_at	KIAA0116 gene, partial cds
class 2	1.406929	0.7890778	0.69446135	0.5199787	D84454_at	UDP-galactose translocator
class 2	1.3972126	0.771632	0.6803573	0.5063799	D83174_s_at	CBP1 Collagen-binding protein 1
class 2	1.3882682	0.7628288	0.6654627	0.4983154	D83735_at	Adult heart mRNA for neutral calponin
class 2	1.3158283	0.7600643	0.65747064	0.48707137	L38969_at	Thrombospondin 3 (THBS3) gene
class 2	1.2211796	0.75861675	0.6480409	0.47926977	U12465_at	RPS11 Ribosomal protein S11
class 2	1.2204406	0.74606985	0.64249825	0.47334	U47621_at	Nucleolar autoantigen No55 mRNA
class 2	1.2186558	0.744558	0.6345838	0.46700227	D80005_at	KIAA0183 gene, partial cds
class 2	1.2145118	0.7413605	0.6255762	0.4635181	X79683_s_at	LAMB2 Laminin, beta 2 (laminin S)
class 2	1.1926116	0.73349786	0.62487775	0.45888507	U73377_at	SKI V-ski avian sarcoma viral oncogene homolog
class 2	1.1885384	0.7101549	0.6203104	0.4543584	L21954_at	PERIPHERAL-TYPE BENZODIAZEPINE RECEPTOR
class 2	1.1789806	0.70968467	0.61982125	0.45166668	D85418_at	Phosphatidylinositol-glycan-class C (PIG-C)
class 2	1.1726408	0.7075023	0.6164542	0.4459662	U50523_at	BRCA2 region, mRNA sequence CG037
class 2	1.1627295	0.6977533	0.60977054	0.44309354	U13991_at	TATA-binding protein associated factor 30 kDa subunit (tafII30) mRNA
class 2	1.1177913	0.6966777	0.60499704	0.4403641	L41066_at	NF-AT3 mRNA
class 2	1.1161832	0.69525945	0.6020127	0.43546635	S80343_at	RARS Arginyl-tRNA synthetase
class 2	1.1063769	0.6934489	0.6001345	0.43275866	D78586_at	CAD PROTEIN
class 2	1.100164	0.68982345	0.59710175	0.43011752	X54304_at	Myosin regulatory light chain mRNA
class 2	1.0985785	0.6834978	0.59472984	0.4265189	X94910_at	ERp31 protein
class 2	1.0931795	0.68134505	0.59273076	0.42424324	U31383_at	G protein gamma-10 subunit mRNA
class 2	1.0755527	0.67965597	0.5872831	0.42292687	D30755_at	VIM Vimentin
class 2	1.0685377	0.67405325	0.5853282	0.42069253	U70439_s_at	PHAPI2b protein
class 2	1.0646288	0.6734994	0.5838599	0.41816902	M19645_at	78 KD GLUCOSE REGULATED PROTEIN PRECURSOR
class 2	1.0563896	0.67214715	0.58231425	0.41704497	D45248_at	Proteasome activator hPA28 subunit beta
class 2	1.0528408	0.6703584	0.5772028	0.4129843	M14338_at	PROS1 Plasma protein S
class 2	1.0516357	0.6697295	0.57507044	0.41167092	D31888_at	KIAA0071 gene, partial cds
class 2	1.0499836	0.6691372	0.57322055	0.41046417	D79996_at	KIAA0174 gene
class 2	1.0487297	0.66673857	0.5681988	0.40736142	U34683_at	GSS Glutathione synthetase
class 2	1.0459839	0.65853596	0.56721437	0.40508857	L12535_at	RSU-1/RSP-1 mRNA
class 2	1.0183622	0.65781057	0.5650841	0.40312052	X61587_at	ARHG Ras homolog gene family, member G (rho G)
class 2	1.0072109	0.6538323	0.56382716	0.40154344	X53777_at	60S RIBOSOMAL PROTEIN L23
class 2	0.99361455	0.6530042	0.56308913	0.39934984	X06700_s_at	COL3A1 Alpha-1 type 3 collagen
class 2	0.98622197	0.6510077	0.56030434	0.3967915	U41515_at	Deleted in split hand/split foot 1 (DSS1) mRNA
class 2	0.9862092	0.6501271	0.5597317	0.3955185	L14565_at	PERIPHERIN
class 2	0.9861108	0.65010846	0.5552204	0.3940714	M63573_at	PPIB Peptidylprolyl isomerase B (cyclophilin B)
class 2	0.9772656	0.64656097	0.55292964	0.39343733	Z23090_at	HSPB1 Heat shock 27kD protein 1
class 2	0.9681267	0.64562047	0.5525739	0.39118996	L25085_at	PROTEIN TRANSPORT PROTEIN SEC61 BETA SUBUNIT
class 2	0.963596	0.644272	0.551029	0.38998803	U72514_at	C2f mRNA
class 2	0.9609764	0.6431171	0.5493531	0.3888596	X15187_at	TRA1 Homologue of mouse tumor rejection antigen gp96
class 2	0.95701075	0.6428692	0.5473165	0.3870365	M29971_at	MGMT 6-O-methylguanine-DNA methyltransferase (MGMT)

class 2	0.95551085	0.6414765	0.54465944	0.3854772	D79997_at	KIAA0175 gene
class 2	0.9548831	0.64113075	0.54276997	0.38295317	Y07604_at	Nucleoside-diphosphate kinase
class 2	0.95175433	0.63506126	0.54169023	0.38189143	D78611_at	MEST Mesoderm specific transcript (mouse) homolog
class 2	0.9498536	0.6342157	0.54131114	0.380591	U84720_at	mRNA export protein Rae1 (RAE1) mRNA
class 2	0.9433024	0.63139164	0.54059446	0.37891367	U72263_s_at	EXT2 Exostoses (multiple) 2
class 2	0.94292474	0.63092226	0.5366494	0.37772393	X85373_at	Sm protein G
class 2	0.94150615	0.6304973	0.5353541	0.37663063	X98296_at	Ubiquitin hydrolase
class 2	0.9404326	0.6302717	0.534726	0.37519532	U28811_at	Cysteine-rich fibroblast growth factor receptor (CFR-1) mRNA
class 2	0.93543696	0.62990934	0.5340427	0.37421787	U41387_at	Gu protein mRNA, partial cds
class 2	0.9310851	0.62937075	0.5334321	0.37231687	L38951_at	Importin beta subunit mRNA
class 2	0.9302329	0.62831765	0.53228664	0.37106973	M11718_at	COL5A2 Collagen, type V, alpha
class 2	0.9218427	0.6272878	0.5308931	0.37046224	X02152_at	LDHA Lactate dehydrogenase A
class 2	0.91466784	0.6263975	0.5304401	0.36923638	X13839_at	LCAT Lecithin-cholesterol acyltransferase
class 2	0.91358876	0.6234602	0.53026843	0.3675315	Z25749_rna1_at	Ribosomal protein S7
class 2	0.9135651	0.62286264	0.52784836	0.36692485	D00763_at	GAPD Glyceraldehyde-3-phosphate dehydrogenase
class 2	0.91283256	0.6222702	0.52783066	0.36592916	L25270_at	XE169 PROTEIN
class 2	0.9052988	0.6209669	0.5276215	0.3650857	M64098_at	High density lipoprotein binding protein (HBP) mRNA
class 2	0.90166676	0.619184	0.52735156	0.36361563	D42041_at	KIAA0088 gene, partial cds
class 2	0.8969863	0.6182318	0.5259541	0.362975	D14043_at	PUTATIVE MUCIN CORE PROTEIN PRECURSOR 24
class 2	0.89683557	0.616351	0.52503	0.3620421	D82348_at	5-aminoimidazole-4-carboxamide-1-beta-D-ribose nucleotide transformylase/inosinase
class 2	0.8967383	0.6154521	0.5231539	0.3603715	U09587_at	GARS Glycyl-tRNA synthetase
class 2	0.89605194	0.6140075	0.5227674	0.3598843	D78275_at	Proteasome subunit p42
class 2	0.8807301	0.61248505	0.52253556	0.35895807	U15655_at	Ets domain protein ERF mRNA
class 2	0.8791751	0.6105275	0.52134347	0.3578664	M33308_at	VCL Vinculin
class 2	0.87861866	0.6101475	0.5194662	0.3572384	J04456_at	LGALS1 Ubiquinol-cytochrome c reductase core protein II
class 2	0.8758852	0.6092619	0.51829946	0.3564762	M24069_at	DNA-BINDING PROTEIN A
class 2	0.8756332	0.60834265	0.5160915	0.3562142	X66945_at	FGFR1 Basic fibroblast growth factor (bFGF) receptor (shorter form)
class 2	0.8730137	0.6069743	0.51504296	0.35525343	M22382_at	HSPD1 Heat shock 60 kD protein 1 (chaperonin)
class 2	0.871905	0.6060964	0.513713	0.3543604	J03191_at	Profilin mRNA
class 2	0.8703043	0.603569	0.51317245	0.35314965	U47926_at	Unknown protein B mRNA
class 2	0.86565924	0.6020561	0.5128095	0.35230598	M85289_at	HSPG2 Heparan sulfate proteoglycan
class 2	0.864009	0.6016002	0.51181066	0.3514165	M14636_at	PYGL Glycogen phosphorylase L (liver form)
class 2	0.854852	0.6013029	0.5113688	0.35065523	S78187_at	M-PHASE INDUCER PHOSPHATASE 2
class 2	0.85385114	0.5993562	0.5086435	0.34998524	S71018_at	Cyclophilin C [human, kidney, mRNA, 883 nt]
class 2	0.8520942	0.5992366	0.5079732	0.34953746	M14949_at	RAS-RELATED PROTEIN R-RAS
class 2	0.8491951	0.5989905	0.5072743	0.3484389	X99920_at	S100 calcium-binding protein A13
class 2	0.8471685	0.5982197	0.5060602	0.34780023	J03824_at	UROS Uroporphyrinogen III synthase
class 2	0.8435629	0.598039	0.50526935	0.3470213	L35240_at	Enigma gene
class 2	0.8416742	0.59718853	0.5044416	0.3458508	X62691_at	40S RIBOSOMAL PROTEIN S15A
class 2	0.84077215	0.59648985	0.5040173	0.34466365	L07758_at	IEF SSP 9502 mRNA
class 2	0.8404173	0.59592414	0.50253683	0.3441604	U35139_at	NECDIN related protein mRNA
class 2	0.8322671	0.59347093	0.5013147	0.34315905	U91985_at	DNA fragmentation factor-45 mRNA
class 2	0.83103186	0.59304476	0.49816543	0.34257925	D14533_at	XPA Xeroderma pigmentosum, complementation group A
class 2	0.83078337	0.59284073	0.49728844	0.34181342	D37965_at	PDGF receptor beta-like tumor suppressor (PRLTS)
class 2	0.8286705	0.5926845	0.49725458	0.341386	U72621_at	LOT1 mRNA

class 2	0.82835925	0.59190524	0.4970425	0.3410285	HG1153-HT1153_at	Nucleoside Diphosphate Kinase Nm23-H2s
class 2	0.8277117	0.59029454	0.49623272	0.3406979	M16938_s_at	Homeo box c8 protein, mRNA
class 2	0.8266656	0.5901575	0.49541354	0.34014606	D15057_at	DEFENDER AGAINST CELL DEATH 1 (clone S171) mRNA
class 2	0.8260545	0.58884704	0.49504972	0.3398025	L40393_at	Beta2-syntrophin (SNT B2) mRNA
class 2	0.8255064	0.58706975	0.49395016	0.3389208	U40572_at	HSPE1 Heat shock 10 kD protein 1 (chaperonin 10)
class 2	0.82378906	0.58681875	0.4934995	0.33787978	U07550_at	Translation initiation factor 3 47 kDa subunit mRNA
class 2	0.8221365	0.5852637	0.49296644	0.33739457	U94855_at	CRYZ Crystallin zeta (quinone reductase)
class 2	0.8211939	0.58489823	0.49241668	0.33726045	L13278_at	CHC1 Chromosome condensation 1
class 2	0.8211335	0.5832674	0.4924099	0.33619076	D00591_at	MADER mRNA
class 2	0.8189326	0.58308315	0.49072617	0.33576697	X70991_at	EEF2 Eukaryotic translation elongation factor 2
class 2	0.81821126	0.5828148	0.49018598	0.3352356	X97074_at	PABPL1 Poly(A)-binding protein-like 1
class 2	0.8171171	0.5811897	0.48924693	0.33457905	U68105_s_at	KIAA0273 gene
class 3	4.298069	1.8113496	1.542546	0.99697	D87463_at	Clone 23612 mRNA sequence
class 3	3.7472157	1.5923314	1.3552583	0.88737005	U90902_at	Type 1 inositol 1,4,5-trisphosphate receptor
class 3	3.690101	1.5091044	1.2780975	0.8364649	D26070_at	Parvalbumin
class 3	3.6179547	1.4561309	1.2265519	0.79228497	X63578_rna1_at	PRKCZ Protein kinase C, zeta
class 3	3.5801797	1.3800497	1.1935662	0.76851	Z15108_at	Germline mRNA sequence
class 3	3.1552649	1.3531709	1.1540082	0.7488579	L35592_at	SCN1B Sodium channel, voltage-gated, type I, beta polypeptide
class 3	2.98379	1.3402475	1.1488526	0.7350318	L10338_s_at	PKD1 Polycystic kidney disease protein 1
class 3	2.9386811	1.3199779	1.1085008	0.7183979	L33243_at	Stat-like protein (Fe65) mRNA
class 3	2.8076477	1.3121926	1.0983416	0.70677555	L77864_at	Mitochondrial creatine kinase (CKMT) gene
class 3	2.7285392	1.3062973	1.0917186	0.6917782	J04469_at	Metabotropic glutamate receptor 4 mRNA
class 3	2.5989084	1.2814465	1.0786043	0.6817148	U92457_s_at	SYNAPTOSOMAL ASSOCIATED PROTEIN 25
class 3	2.5474255	1.2378745	1.0633833	0.67317224	D21267_at	Clone 23682 mRNA sequence
class 3	2.4580472	1.2305647	1.0299538	0.66360885	U79288_at	DAGK4 Diacylglycerol kinase delta
class 3	2.3459284	1.2142801	1.0283259	0.65945446	D63479_s_at	DNM1 Dynamin 1
class 3	2.342861	1.2094874	1.0197432	0.65128297	L07807_s_at	KIAA0059 gene
class 3	2.280001	1.2006655	1.0105054	0.6466257	D31883_at	GRIN1 Glutamate receptor, ionotropic, N-methyl D-aspartate 1
class 3	2.2601855	1.1856893	1.0029721	0.6399636	L13266_s_at	Two P-domain K+ channel TWIK-1 mRNA
class 3	2.2338665	1.1824055	0.99219334	0.63467205	U33632_at	TUBULIN ALPHA-4 CHAIN
class 3	2.187364	1.1767278	0.9869156	0.6280837	X06956_at	Cri-du-chat region mRNA, clone NIBB11
class 3	2.1791985	1.1758486	0.9814895	0.62358665	U52827_at	TIAM1 T-cell lymphoma invasion and metastasis 1
class 3	2.1489515	1.1723996	0.96380335	0.6196844	U16296_at	Clone 23695 mRNA sequence
class 3	2.1310797	1.1686139	0.9591985	0.61347145	U79289_at	Inducible protein mRNA
class 3	2.1222842	1.1660343	0.9513746	0.60739744	L47738_at	Platelet alpha SNAP mRNA
class 3	2.0701365	1.1609877	0.94285935	0.60494643	U39412_at	MBP Myelin basic protein
class 3	2.0604408	1.1502872	0.93769294	0.59993637	M13577_at	PRKAR1B Protein kinase, cAMP-dependent, regulatory, type I, beta
class 3	2.0528634	1.147987	0.934068	0.5970684	M65066_at	ENO2 gene for neuron specific (gamma) enolase
class 3	2.0507598	1.1448451	0.9319095	0.59396625	X51956_rna1_at	GRM4 Glutamate receptor, metabotropic 4
class 3	2.0365791	1.1420877	0.9278926	0.5918934	X80818_at	Lysophospholipase homolog (HU-K5) mRNA
class 3	2.0284941	1.1411111	0.9187019	0.5886168	U67963_at	KIAA0237 gene
class 3	2.02188	1.1368862	0.91659665	0.5841766	D87074_at	KIAA0275 gene
class 3	2.0109	1.130213	0.9079699	0.581688	D87465_at	KERATIN, TYPE I CYTOSKELETAL 17
class 3	2.004002	1.1282408	0.9066855	0.5797237	S72493_s_at	

class 3	1.959235	1.1278926	0.8963888	0.57711166 D63851_at	Unc-18 homologue
class 3	1.9334141	1.1088517	0.88973886	0.57648414 U90907_at	Clone 23907 mRNA sequence
class 3	1.9277046	1.1050158	0.8885025	0.5740537 U13616_at	ANK3 Ankyrin G
class 3	1.9067957	1.1016182	0.8826693	0.5693379 U79245_at	Clone 23586 mRNA sequence
class 3	1.8848716	1.1010174	0.8821242	0.565312 X64838_at	RSN Restin (Reed-Steinberg cell-expressed intermediate filament-associated protein)
class 3	1.8844064	1.099194	0.8803771	0.5631035 D83542_at	Cadherin-15
class 3	1.878876	1.0981528	0.8715951	0.5604119 U81607_at	GRAVIN
class 3	1.8755924	1.0932494	0.8671489	0.55765635 M64925_at	MPP1 Membrane protein, palmitoylated 1 (55kD)
class 3	1.8585938	1.090932	0.866664	0.5553793 D78577_s_at	YWHAH Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide
class 3	1.8582501	1.0852915	0.8653905	0.55245095 U47928_at	Protein A alternatively spliced form 2 (A-2) mRNA
class 3	1.8516157	1.0834761	0.86097085	0.5506602 M96859_at	DPP6 Dipeptidylpeptidase VI
class 3	1.8367375	1.0826322	0.85773826	0.5468505 U76421_at	DsRNA adenosine deaminase DRADA2b (DRADA2b) mRNA
class 3	1.8268647	1.0788103	0.85618746	0.5450503 HG2259-HT2348_s_at	Tubulin, Alpha 1, Isoform 44
class 3	1.8232589	1.0724745	0.8504435	0.54230773 X14766_at	GABRA1 Gamma-aminobutyric acid (GABA) A receptor, alpha 1
class 3	1.8144729	1.0702177	0.8499052	0.5412158 U07139_at	CAB3b mRNA for calcium channel beta3 subunit
class 3	1.8111364	1.0686209	0.8456019	0.53755283 L76627_at	Metabotropic glutamate receptor 1 alpha (mGluR1alpha) mRNA
class 3	1.7889462	1.0664026	0.8448219	0.53580296 M37400_at	GOT1 Glutamic-oxaloacetic transaminase 1, soluble (aspartate aminotransferase 1)
class 3	1.7856433	1.0653843	0.84243757	0.5344206 U27193_at	Protein-tyrosine phosphatase mRNA
class 3	1.7796848	1.063558	0.84091395	0.53190583 D63477_at	KIAA0143 gene, partial cds
class 3	1.759171	1.0625236	0.83884305	0.53067225 X92493_s_at	STM-7 protein
class 3	1.7583526	1.061523	0.8383003	0.5297999 X70940_s_at	EEF1A2 Eukaryotic translation elongation factor 1 alpha 2
class 3	1.7429492	1.0588214	0.83477515	0.52727365 D29013_at	POLB DNA polymerase beta subunit
class 3	1.7351419	1.0561141	0.8308162	0.52565986 D79998_at	KIAA0176 gene, partial cds
class 3	1.733765	1.0514225	0.83047813	0.5235324 U25029_at	GRL Glucocorticoid receptor alpha {alternative products}
class 3	1.7079235	1.0505984	0.8300135	0.52023715 J04046_s_at	CALMODULIN
class 3	1.7047126	1.0464945	0.8288378	0.5191814 M33653_at	COL4A2 Collagen, type IV, alpha 2
class 3	1.6961877	1.0404369	0.82708424	0.5182239 M58583_at	CEREBELLIN 1 PRECURSOR
class 3	1.6901898	1.034222	0.82666314	0.51679677 M32313_at	SRD5A1 Steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)
class 3	1.6850768	1.0337527	0.82482046	0.51418847 D82347_at	NEUROD1 Neurogenic differentiation 1
class 3	1.6842928	1.0330929	0.8175216	0.5133849 D83777_at	KIAA0193 gene
class 3	1.6828924	1.0288644	0.8172416	0.5115587 Z31695_at	43 kDa inositol polyphosphate 5-phosphatase
class 3	1.6768361	1.0284845	0.81461614	0.51045823 X90824_s_at	USF2a & USF2b, clone P2
class 3	1.6749456	1.0247817	0.8121015	0.50901747 D43636_at	KIAA0096 gene, partial cds
class 3	1.674618	1.0209516	0.80982834	0.5089023 L10373_at	MXS1 Membrane component, X chromosome, surface marker 1
class 3	1.6696235	1.0166084	0.8096402	0.5065178 X56411_rna1_at	ADH4 gene for class II alcohol dehydrogenase (pi subunit), exon 1
class 3	1.6669358	1.0154912	0.80860317	0.5050052 U67171_at	Selenoprotein W (selW) mRNA
class 3	1.6513395	1.0122768	0.80841583	0.5046859 Y09392_s_at	WSL-LR, WSL-S1 and WSL-S2 proteins
class 3	1.6431793	1.0100572	0.8075384	0.50179076 U85707_at	Leukemogenic homolog protein (MEIS1) mRNA
class 3	1.6431123	1.0041083	0.8070719	0.50103754 Y00067_rna1_at	Neurofilament subunit M (NF-M)
class 3	1.6419501	1.0011116	0.8067552	0.49965492 U87223_at	Contactin associated protein (Caspr) mRNA
class 3	1.6394771	0.9955323	0.80538327	0.49950275 L10333_s_at	Neuroendocrine-specific protein A (NSP) mRNA

class 3	1.6353047	0.98837924	0.7981355	0.49779063	U17838_at	Zinc finger protein RIZ mRNA
class 3	1.6318842	0.9861097	0.7980577	0.49699193	U46901_at	SNCA Synuclein, alpha (non A4 component of amyloid precursor) AUH mRNA
class 3	1.6295244	0.97382015	0.79276174	0.4967051	X79888_at	JNK3 alpha2 protein kinase (JNK3A2) mRNA
class 3	1.6287894	0.97215784	0.79182774	0.49586895	U07620_at	P21-activated protein kinase (Pak1) gene
class 3	1.6147838	0.9710149	0.7907015	0.49469632	U24152_at	GIF=growth inhibitory factor [human, brain, Genomic, 2015 nt]
class 3	1.6119615	0.966481	0.78996116	0.49367088	S72043_rna1_at	KIAA0274 gene
class 3	1.6035154	0.9626293	0.78644234	0.49325758	D87464_at	Regulator of G-protein signaling similarity (RGS7) mRNA, partial cds
class 3	1.5985726	0.9546146	0.7859533	0.4915498	U32439_at	CRYM Crystallin Mu
class 3	1.5867528	0.94753075	0.7855334	0.49130288	L02950_at	FKBP4 FK506-binding protein 4 (59kD)
class 3	1.5784372	0.94672173	0.78228825	0.48999268	M88279_at	GLRX Glutaredoxin (thioltransferase)
class 3	1.5764623	0.94641775	0.77991027	0.48929474	X76648_at	Aldolase C gene
class 3	1.5748655	0.94637823	0.77925485	0.48841372	X05196_at	ZAK1-4 mRNA in human skin fibroblast
class 3	1.5744048	0.9456003	0.7786697	0.48784614	D83407_at	Facioscapulohumeral muscular dystrophy (FSHD) gene region, D4Z4 tandem repeat unit
class 3	1.5715562	0.94528913	0.7786567	0.48684633	D38024_at	Prostaglandin D2 synthase gene
class 3	1.570073	0.94405437	0.776268	0.4850091	M98539_at	KERATIN, TYPE II CYTOSKELETAL 2 ORAL
class 3	1.5688661	0.9425145	0.775566	0.48369113	M99063_at	Clone CCA12 mRNA containing CCA trinucleotide repeat
class 3	1.5683391	0.939959	0.77452576	0.4827068	U06681_at	CYB5 Cytochrome b-5
class 3	1.5661842	0.93824065	0.7729135	0.4820274	M22976_at	AGTR1 Angiotensin receptor 1
class 3	1.5626711	0.9377882	0.7699611	0.4809086	S77410_at	SERINE/THREONINE PROTEIN PHOSPHATASE 2B CATALYTIC SUBUNIT, BETA ISOFORM
class 3	1.5524818	0.93469095	0.7696227	0.4793128	M29551_at	Diacylglycerol kinase zeta mRNA
class 3	1.5440252	0.93378896	0.76802355	0.47835198	U51477_at	DIHYDROPRYRIDINE-SENSITIVE L-TYPE, CALCIUM CHANNEL BETA-1-B1 SUBUNIT
class 3	1.5358244	0.93254817	0.7675914	0.4776862	M92303_at	OPCML Opioid-binding cell adhesion molecule
class 3	1.5323644	0.9281277	0.7668785	0.4765654	U79251_at	KIAA0216 gene
class 3	1.5293515	0.9264165	0.766238	0.47640666	D86970_at	K+ channel beta 1a subunit mRNA, alternatively spliced
class 3	1.5226866	0.9237222	0.7658018	0.47434324	L39833_at	Histidyl-tRNA synthetase homolog (HO3) mRNA
class 3	1.5180973	0.9220966	0.7589294	0.47297105	U18937_at	Phosphatidylinositol (4,5)bisphosphate 5-phosphatase homolog mRNA, partial cds
class 3	1.516081	0.9166196	0.75824547	0.47195733	U45975_at	POLD1 Polymerase (DNA directed), delta 1, catalytic subunit (125kD)
class 4	0.8734975	1.134452	0.9411052	0.67896336	M80397_s_at	CHRN1 Cholinergic receptor, nicotinic, beta polypeptide 1 (muscle)
class 4	0.8119233	0.963902	0.8420816	0.6297001	X14830_at	IGHD gene (immunoglobulin delta-chain) extracted from Human germline
class 4	0.81055194	0.9483548	0.8026216	0.6025905	K02882_cds1_s_at	IgD chain gene, C-region, C-delta-1 domain
class 4	0.79492754	0.9124361	0.7859309	0.5878106	HG4178-HT4448_at	Af-17
class 4	0.7530036	0.8959806	0.7669699	0.56836367	U97018_at	Echinoderm microtubule-associated protein homolog HuEMAP mRNA
class 4	0.7159336	0.88823646	0.7476684	0.55669653	X52228_at	MUC1 Mucin 1, transmembrane
class 4	0.7055104	0.87415504	0.73530275	0.54636556	L18920_f_at	MELANOMA-ASSOCIATED ANTIGEN 2
class 4	0.6556214	0.87129426	0.72525615	0.53645986	D29675_at	Inducible nitric oxide synthase gene, promoter and exon 1
class 4	0.6340733	0.8697119	0.71343136	0.5269011	S82471_s_at	SSX3=Kruppel-associated box containing SSX gene [human, testis, mRNA Partial, 675 nt]
class 4	0.6339644	0.84462005	0.707733	0.52023065	U22314_s_at	Neural-restrictive silencer factor, splice variant mRNA, partial cds
class 4	0.6246593	0.83731145	0.7006627	0.51330495	X74987_s_at	2-5A binding protein

class 4	0.61529684	0.82780534	0.6935719	0.50730973	M17466_at	F12 Coagulation factor XII (Hageman factor)
class 4	0.60239977	0.82662934	0.68776125	0.50441134	M29610_s_at	GYPE Glycophorin E
class 4	0.5956711	0.82629377	0.68068653	0.4996151	M54951_at	ATRIAL NATRIURETIC FACTOR PRECURSOR
class 4	0.5904324	0.825621	0.6763503	0.4960944	K02766_at	C9 Complement component C9
class 4	0.58323526	0.8113131	0.6684805	0.49252754	M36429_s_at	Transducin beta-2 subunit mRNA
class 4	0.58193743	0.8040995	0.6674828	0.4896936	M95623_cds1_at	PBGD gene (hydroxymethylbilane synthase) extracted from Homo sapiens hydroxymethylbilane synthase gene
class 4	0.579044	0.8023006	0.6626152	0.48392293	U57592_at	Jumonji putative protein (jumonji) mRNA
class 4	0.5783975	0.7968204	0.6584121	0.4814893	U40223_at	Uridine nucleotide receptor (UNR) gene
class 4	0.5780785	0.79416245	0.65169525	0.47840923	K02777_s_at	T-cell receptor active alpha-chain mRNA from Jurkat cell line
class 4	0.5670055	0.7924286	0.64660954	0.47577995	U79302_at	Clone 23855 mRNA, partial cds
class 4	0.5656941	0.7891743	0.6452447	0.47280827	U40462_at	Ikaro/LyF-1 homolog (h1k-1) mRNA
class 4	0.56489813	0.78810704	0.63897216	0.46992204	X59842_rna1_s_at	PBX2 mRNA
class 4	0.56000954	0.77741265	0.63788915	0.46515974	K02405_f_at	HLA CLASS II HISTOCOMPATIBILITY ANTIGEN, DQ(1) BETA CHAIN PRECURSOR
class 4	0.55480826	0.7769049	0.6332759	0.4604109	D25539_at	KIAA0040 gene
class 4	0.5504693	0.76869243	0.63192195	0.45713758	X57110_s_at	CBL Cas-Br-M (murine) ecotropic retroviral transforming sequence
class 4	0.5433172	0.767859	0.6289676	0.45595473	M77140_at	GALN Galanin
class 4	0.5420265	0.7672204	0.6266064	0.45416942	M91669_s_at	Bullous pemphigoid autoantigen BP180 gene, 3' end
class 4	0.5415517	0.7625079	0.6263098	0.45194966	X93921_at	Protein-tyrosine-phosphatase (tissue type: testis)
class 4	0.5387575	0.75851834	0.62277097	0.44980803	U61276_s_at	Transmembrane protein Jagged 1 (HJ1) mRNA
class 4	0.53783655	0.75743777	0.6212576	0.44735184	X53595_s_at	APOH Apolipoprotein H
class 4	0.53758186	0.754692	0.61968046	0.44606307	X16135_at	HETEROGENEOUS NUCLEAR RIBONUCLEOPROTEIN L
class 4	0.53244615	0.75203496	0.61582357	0.4445306	X98178_s_at	MACH-beta-4 protein
class 4	0.53053224	0.74949163	0.6133881	0.44177684	M81829_at	Somatostatin receptor isoform 1 gene
class 4	0.5266541	0.73750824	0.61149806	0.4408685	M21389_at	KRT5 Keratin 5 (epidermolysis bullosa simplex, Dowling-Meara/Kobner/Weber-Cockayne types)
class 4	0.5265538	0.7358704	0.60937744	0.43847492	HG2147-HT2217_r_at	Mucin 3, Intestinal (Gb:M55405)
class 4	0.52553517	0.73483276	0.6050042	0.4369384	M64572_at	PTPN3 Protein tyrosine phosphatase, non-receptor type 3
class 4	0.52480435	0.7332647	0.6033811	0.43495926	D88155_s_at	Steroidogenic factor 1 mRNA
class 4	0.52309185	0.7292017	0.6027077	0.43228412	U88964_at	HEM45 mRNA
class 4	0.5213507	0.7290054	0.60209686	0.4313736	Z84721_cds1_at	Zeta-globin 1 gene extracted from Human DNA sequence from cosmid GG1 from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3 Contains alpha and zeta globin genes and ESTs
class 4	0.5202241	0.72878873	0.5992238	0.42955112	X91103_at	Hr44 protein
class 4	0.51837975	0.72774935	0.5977254	0.4274839	U49082_at	Transporter protein (g17) mRNA
class 4	0.5167581	0.7269519	0.5975571	0.425647	U05255_at	GLYCOPHORIN B PRECURSOR
class 4	0.511345	0.7261254	0.59631515	0.42395902	D42039_at	KIAA0081 gene, partial cds
class 4	0.51000273	0.72500247	0.59339905	0.42339996	U45448_s_at	P2x1 receptor mRNA
class 4	0.5078437	0.72347194	0.5933299	0.42207658	HG3914-HT4184_s_at	Cell Division Cycle Protein 2-Related Protein Kinase (Pisslr)
class 4	0.5066535	0.7224339	0.5908667	0.4204116	D88422_at	CYSTATIN A
class 4	0.5016486	0.72010094	0.5878204	0.41856045	HG2255-HT2344_f_at	Phosphoribosyl Pyrophosphate Synthetase, Subunit Iii
class 4	0.49719426	0.7200707	0.5878162	0.4177429	U20760_at	CASR Calcium-sensing receptor (hypocalciuric hypercalcemia 1, severe neonatal hyperparathyroidism)
class 4	0.49097803	0.7180478	0.5869554	0.41703936	U51096_at	Homeobox protein Cdx2 mRNA

class 4	0.48994532	0.7152811	0.5867968	0.41594544 Y00318_at	IF I factor (complement)
class 4	0.48947743	0.714268	0.5866425	0.4146865 M76180_at	DDC Dopa decarboxylase (aromatic L-amino acid decarboxylase)
class 4	0.4875335	0.71361583	0.5847286	0.4133012 M29696_at	IL7R Interleukin 7 receptor
class 4	0.48321855	0.71123636	0.58444405	0.41136318 L13197_at	PAPPA Pregnancy-associated plasma protein A
class 4	0.48314202	0.709341	0.583089	0.40934148 M83664_at	HLA-DPB1 Major histocompatibility complex, class II, DP beta 1
class 4	0.48295903	0.70909935	0.5790559	0.40833262 X93996_rna1_at	AFX protein
class 4	0.48157865	0.7063981	0.5782582	0.4079157 HG537-HT537_at	Collagen, Type VIII, Alpha 2
class 4	0.48118794	0.7061022	0.57807314	0.40636182 M92432_at	GUC2D Guanylate cyclase 2D, membrane (retina-specific)
class 4	0.4810371	0.70342964	0.57601345	0.4028365 HG2149-HT2219_at	Mucin (Gb:M57417)
class 4	0.48017803	0.7025353	0.5742331	0.40200925 V00532_rna1_f_at	IFNA gene (interferon alpha-i) extracted from Human gene for leukocyte (alpha) interferon C
class 4	0.47445068	0.70189375	0.57402205	0.40101275 X03363_s_at	ERBB2 V-erb-b2 avian erythroblastic leukemia viral oncogene homolog 2 (neuro/glioblastoma derived oncogene homolog)
class 4	0.47439837	0.70042795	0.5727844	0.40050083 M16276_at	HLA-DQB1 Major histocompatibility complex, class II, DQ beta 1
class 4	0.4743872	0.69863063	0.57177055	0.3994889 M32598_at	RPS11 Ribosomal protein S11
class 4	0.47178313	0.6959429	0.57023054	0.39857152 U50383_at	Retinoic acid-responsive protein (NN8-4AG) mRNA
class 4	0.4714457	0.69539475	0.56962925	0.39810398 U62966_at	Na+/nucleoside cotransporter (hCNT1c) mRNA
class 4	0.47041702	0.6917778	0.56895816	0.3967692 X69950_s_at	WT1 Wilms tumor 1
class 4	0.46828622	0.69176567	0.5689231	0.39595485 M64269_s_at	Mast cell chymase gene
class 4	0.46722797	0.69092834	0.56770706	0.39445558 L07868_at	ERBB4 V-erb-a avian erythroblastic leukemia viral oncogene homolog-like 4
class 4	0.4666462	0.69058037	0.56759816	0.3939125 U66661_at	GABA-A receptor epsilon subunit mRNA
class 4	0.4662564	0.68824685	0.5647112	0.39127818 HG4677-HT5102_s_at	Oncogene Ret/Ptc2, Fusion Activated
class 4	0.46033314	0.6872949	0.5642959	0.39101368 J04599_at	BGN Biglycan
class 4	0.45942208	0.6856201	0.56342113	0.39028785 HG4236-HT4506_f_at	Zinc Finger Protein Znf138
class 4	0.4591495	0.685429	0.5624474	0.38855514 HG3264-HT3441_at	Af-6 (Gb:U02478)
class 4	0.45771	0.683483	0.56229144	0.3877895 D86965_at	KIAA0210 gene
class 4	0.457108	0.6824125	0.56183857	0.387533 M33987_at	CA1 Carbonic anhydrase I
class 4	0.45677373	0.6816408	0.56127465	0.3865487 L43576_at	(clone EST02946) mRNA
class 4	0.45494822	0.6806515	0.56029445	0.38604084 A28102_at	GABAA receptor alpha-3 subunit
class 4	0.45359454	0.6804817	0.5579805	0.38481084 M11726_at	PPY Pancreatic polypeptide
class 4	0.45104054	0.6794522	0.55785537	0.38300934 X59770_at	INTERLEUKIN-1 RECEPTOR, TYPE II PRECURSOR
class 4	0.45097542	0.6782299	0.5546474	0.382007 L11372_at	Protocadherin 43 mRNA, 3' end of cds for alternative splicing PC43-12
class 4	0.45055133	0.6776299	0.5546461	0.38108552 X07496_at	APOA1 Apolipoprotein A-I
class 4	0.45006305	0.67592806	0.55318505	0.38004473 U51241_at	CMKBR3 Chemokine (C-C) receptor 3
class 4	0.4486073	0.6751799	0.5531416	0.37975523 L29306_s_at	Tryptophan hydroxylase (Tph) mRNA
class 4	0.4483117	0.6738043	0.5521731	0.37970376 L07738_at	DIHYDROPRYRIDINE-SENSITIVE L-TYPE, SKELETAL MUSCLE CALCIUM CHANNEL GAMMA SUBUNIT
class 4	0.44712317	0.67320645	0.5497038	0.3794417 X99140_at	Hair keratin, hHb5
class 4	0.44642755	0.6727433	0.5485155	0.3787492 HG3502-HT3696_at	Homeotic Protein Hox5.4
class 4	0.44479564	0.6722694	0.5476115	0.37755352 U56244_at	HIG-1 mRNA
class 4	0.44423437	0.67023844	0.5466597	0.3766933 X82200_at	Staf50 mRNA
class 4	0.44366303	0.668862	0.5444793	0.37616846 U10323_at	Nuclear factor NF45 mRNA
class 4	0.44355562	0.6666349	0.5439289	0.37465096 X00437_s_at	TCRB T-cell receptor, beta cluster
class 4	0.44308597	0.66651595	0.5437791	0.3744109 L38707_at	Diacylglycerol kinase (DAGK) mRNA
class 4	0.44278786	0.6659957	0.54239744	0.3734854 Z29481_at	3-HYDROXYANTHRANILATE 3,4-DIOXYGENASE
class 4	0.4408863	0.66582125	0.5416646	0.37182057 S78873_s_at	Guanine nucleotide exchange factor mss4 mRNA

class 4	0.43914822	0.66549075	0.5409329	0.37104592	X75342_at	SHB SHB adaptor protein (a Src homology 2 protein)
class 4	0.43772706	0.66497093	0.53880125	0.3705083	U62433_at	CHRNA4 Cholinergic receptor, nicotinic, alpha polypeptide 4
class 4	0.43764022	0.66314214	0.5384878	0.36978632	AB002559_at	Hunc18b2
class 4	0.4371058	0.66192573	0.5379855	0.3691347	X91809_at	GAIP protein
class 4	0.43608817	0.66024655	0.537967	0.3691051	Y09980_rna4_at	HOXD3 gene
class 4	0.43561712	0.6573707	0.537837	0.36905307	HG2028-HT2082_at	Laminin, A Polypeptide
class 4	0.43479303	0.6573559	0.5359841	0.3681883	U81262_at	EPLG5 Eph-related receptor tyrosine kinase ligand 5

Multiple tumor clustering

The results of clustering the multi-tumor dataset A are shown below. Two clustering methods were used as described in the [Clustering](#) section. Notice how except for the PNETs the samples cluster mostly along tissue types. The AR/RT sample cluster together despite coming from different locations (renal, extra renal and CNS).

Hierarchical Clustering of Multiple Tumor Samples

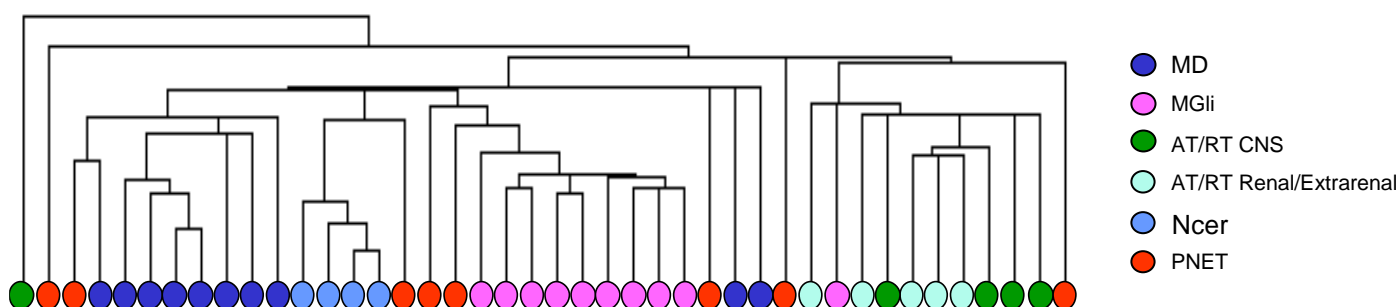
Michael Eisen's clustering algorithm

Dataset A

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 12 (12-fold), max-min= 1200 absolute units

Number of features (genes) = 1065



SOM Clustering of Multiple Tumor Samples

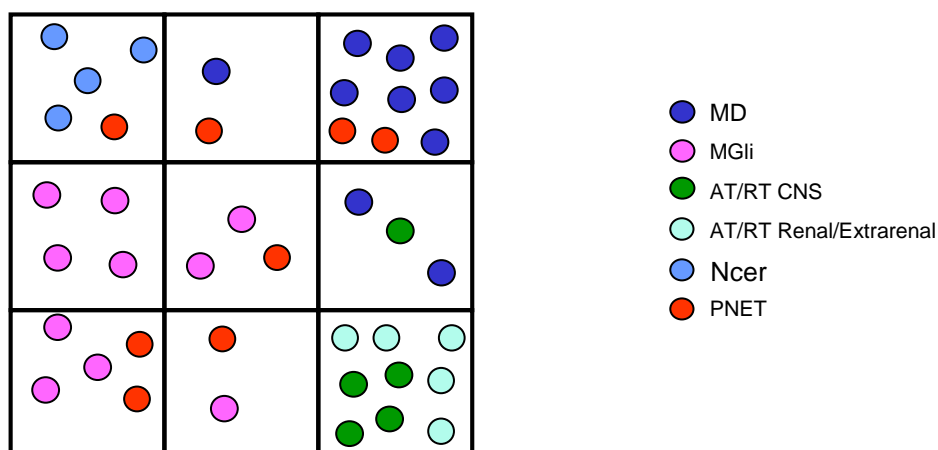
GeneCluster algorithm

Dataset A

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 12 (12-fold), max-min= 1200 absolute units

Number of features (genes) = 1065



Multiple tumor classes predictions (*k*-NN)

This section contains the detailed sample predictions and error rates of predicting the different tumor types with a *k*-nearest neighbor algorithm in leave-one-out cross-validation.

The model predicts 35 out of 42 samples correctly and it is clearly highly significant ($P\text{-val} < 10E-10$, see the calculation below and the [Proportional chance criterion.](#))

Multiple tumor classes prediction *k*-nearest neighbors algorithm

Dataset A

Values thresholded to 20 from below and 16000 from above

Variation filter: $\max/\min > 5$ (5-fold), $\max\text{-}\min = 500$ absolute units

Number of features (genes) = 10. $K=3$, 1/distance weighting

Confusion Matrix

Actual	Predicted					
	MD	MGliob	Rhab	Ncer	PNET	
MD		8	0	1	0	1
MGliob		0	10	0	0	0
Rhab		0	0	9	0	1
Ncer		0	0	0	4	0
PNET		3	0	1	0	4
		11	10	11	4	6
						42

Proportional chance criterion (see chapter VII of Huberty's Applied Discriminant Analysis)

$$C_{pro} = \frac{(10/42) \cdot (10/42) + (10/42) \cdot (10/42) + (10/42) \cdot (10/42) + (4/42) \cdot (4/42) + (8/42) \cdot (8/42)}{0.21542}$$

$$C_{pro} = 0.21542$$

$$P_{cc} = \frac{35}{42} = 0.833333$$

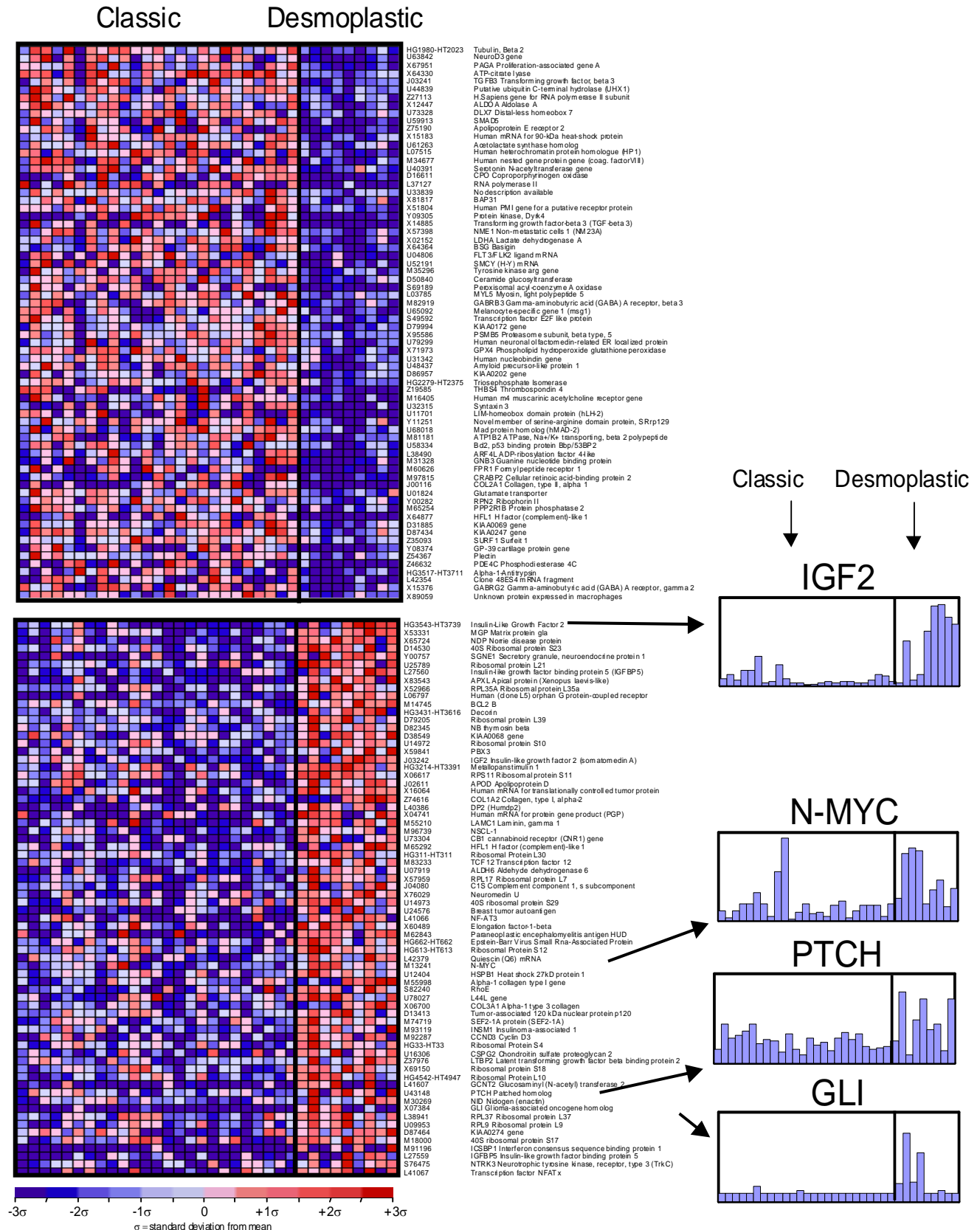
$$\frac{(P_{cc} - C_{pro})/\sqrt{C_{pro}(1-C_{pro})/n}}{Z} = 9.740725 \quad p\text{-val} < 10E-10$$

Datapoint	Predicted Class	Confidence	True Class	Error?
Brain_MD_12	2	2.80E-04	0	*
Brain_MD_61	4	0.002653	0	*
Brain_MD_15	0	0.006245	0	

Brain_MD_57	0	0.64436	0
Brain_MD_33	0	0.429685	0
Brain_MD_64	0	0.258694	0
Brain_MD_17	0	0.252698	0
Brain_MD_62	0	0.008015	0
Brain_MD_63	0	0.008603	0
Brain_MD_32	0	0.591491	0
Brain_MGlio_1	1	0.050383	1
Brain_MGlio_2	1	0.443027	1
Brain_MGlio_3	1	0.654818	1
Brain_MGlio_4	1	0.184817	1
Brain_MGlio_5	1	0.51099	1
Brain_MGlio_6	1	0.229628	1
Brain_MGlio_7	1	0.600702	1
Brain_MGlio_8	1	0.01971	1
Brain_MGlio_9	1	0.519792	1
Brain_MGlio_10	1	0.213082	1
Brain_Rhab_1	2	0.609684	2
Brain_Rhab_2	2	0.401608	2
Brain_Rhab_3	2	0.107676	2
Brain_Rhab_4	2	0.257616	2
Brain_Rhab_5	2	0.488417	2
Brain_Rhab_6	2	0.50804	2
Brain_Rhab_7	2	0.333809	2
Brain_Rhab_8	4	0.055136	2 *
Brain_Rhab_9	2	0.391122	2
Brain_Rhab_10	2	0.042137	2
Brain_Ncer_1	3	0.136345	3
Brain_Ncer_2	3	0.02372	3
Brain_Ncer_3	3	0.204818	3
Brain_Ncer_4	3	0.021819	3
Brain_PNET_1	0	2.82E-06	4 *
Brain_PNET_2	2	0.538158	4 *
Brain_PNET_3	4	0.001226	4
Brain_PNET_4	0	4.01E-04	4 *
Brain_PNET_5	4	0.090692	4
Brain_PNET_6	0	0.251837	4 *
Brain_PNET_7	4	0.003098	4
Brain_PNET_8	4	0.110559	4

Classic vs. desmoplastic MD markers

This picture shows some of the top markers of the classic vs. desmoplastic distinction sorted by signal to noise ratios as described in [Gene marker selection](#) section. The table below shows the top 200 markers including the permutation test values (see [Permutation-based neighborhood analysis for marker gene](#)). Some of the genes regulated by Shh are shown at right.



Top 200/200 Marker Genes for Classic vs Desmoplastic Medulloblastoma Distinction

Selected by signal-to-noise (mean) ratio

Values thresholded to 20 from below and 16000 from above

Variation filter: max/min > 3 (3-fold), max-min= 100 absolute units

Dataset B

Class 0 = High in Classic, low in Desmoplastic

Class 1 = High in Desmoplastic, low in Classic

Permutation test					Marker genes	
Distinction	Distance	Perm 1%	Perm 5%	Median 50%	Feature	Desc
Distinction	Distance	Perm 1%	Perm 5%	Perm (user)	Feature	Desc
class 0	0.9927214	1.100922	0.97285044	0.80008066	HG1980-HT2023	Tubulin, Beta 2
class 0	0.85944515	0.992591	0.91319513	0.7389958	U63842	Neurogenic basic-helix-loop-helix protein (neuroD3) gene
class 0	0.85575306	0.9546853	0.85402167	0.70865357	X67951	PAGA Proliferation-associated gene A
class 0	0.84818983	0.9064097	0.82473946	0.68806636	X64330	ATP-citrate lyase
class 0	0.81497514	0.8919678	0.8098986	0.66982466	J03241	TGFB3 Transforming growth factor, beta 3
class 0	0.80149823	0.85871273	0.78207725	0.6551903	U44839	Putative ubiquitin C-terminal hydrolase (UHX1) mRNA
class 0	0.7977331	0.85563594	0.77035093	0.64376813	Z27113	DNA-DIRECTED RNA POLYMERASE II 14.4 KD POLYPEPTIDE
class 0	0.77652156	0.8363159	0.75971323	0.6344267	X12447	ALDOA Aldolase A
class 0	0.7726323	0.8361712	0.7545372	0.6262046	U73328	DLX7 Distal-less homeobox 7
class 0	0.76898795	0.8337848	0.7440253	0.6194606	U59913	SMAD5 (Smad5) mRNA
class 0	0.7608277	0.8256053	0.7385865	0.6139341	Z75190	Apolipoprotein E receptor 2
class 0	0.7566224	0.8161263	0.7309519	0.60611314	X15183	60S RIBOSOMAL PROTEIN L13
class 0	0.75550735	0.8065621	0.72232366	0.60053813	U61263	Acetolactate synthase homolog mRNA
class 0	0.75087696	0.8001862	0.71438277	0.59579366	L07515	HETEROCHROMATIN PROTEIN 1 HOMOLOG
class 0	0.7489568	0.7779621	0.70930177	0.5898725	M34677	FACTOR VIII INTRON 22 PROTEIN
class 0	0.74733686	0.7772034	0.70585626	0.58594066	U40391	Serotonin N-acetyltransferase gene
class 0	0.7389886	0.77309155	0.6958213	0.5822117	D16611	CPO Coproporphyrinogen oxidase
class 0	0.72759986	0.77139044	0.692992	0.57899344	L37127	(clone mf.18) RNA polymerase II mRNA
class 0	0.72445315	0.77077353	0.6906138	0.574035	U33839	No description available for U33839
class 0	0.7202563	0.76765144	0.6882062	0.5697634	X81817	6C6-Ag mRNA
class 0	0.7089169	0.7658038	0.6853358	0.56685334	X51804	PUTATIVE RECEPTOR PROTEIN
class 0	0.7087572	0.75293714	0.68298733	0.5636675	Y09305	Protein kinase, Dyrk4, partial
class 0	0.70606977	0.7528129	0.6785786	0.5611313	X14885	Transforming growth factor-beta 3 (TGF-beta 3) exon 1
class 0	0.70567566	0.75150484	0.67536324	0.55880284	X57398	NME1 Non-metastatic cells 1, protein (NM23A) expressed in
class 0	0.7001394	0.74915314	0.67155474	0.5559985	X02152	LDHA Lactate dehydrogenase A
class 0	0.6914374	0.7464208	0.66720957	0.5522619	X64364	BSG Basigin
class 0	0.69114035	0.7378829	0.663854	0.54998326	U04806	FLT3/FLK2 ligand mRNA
class 0	0.68837094	0.7366576	0.6583117	0.5470333	U52191	SMCY (H-Y) mRNA
class 0	0.6863416	0.7331688	0.6526042	0.5439266	M35296	Tyrosine kinase arg gene mRNA
class 0	0.68461454	0.7283306	0.652604	0.54087377	D50840	Ceramide glucosyltransferase
class 0	0.68362164	0.72296304	0.64872414	0.53737545	S69189	Peroxisomal acyl-coenzyme A oxidase
class 0	0.6747787	0.7211591	0.64778924	0.53378457	L03785	MYL5 Myosin, light polypeptide 5, regulatory
class 0	0.67238975	0.71856	0.64576083	0.53147835	M82919	GABRB3 Gamma-aminobutyric acid (GABA) A receptor, beta 3
class 0	0.67033887	0.71758395	0.6423796	0.5288971	U65092	Melanocyte-specific gene 1 (msg1) mRNA

class 0	0.6700922	0.7157754	0.6378237	0.5273885 S49592	Transcription factor E2F like protein [human, mRNA, 2492 nt]
class 0	0.6688998	0.71575814	0.63683856	0.5247664 D79994	KIAA0172 gene, partial cds
class 0	0.6687492	0.71367145	0.63496286	0.52286 X95586	PSMB5 Proteasome (prosome, macropain) subunit, beta type, 5
class 0	0.6678679	0.710983	0.63297284	0.52204216 U79299	Neuronal olfactomedin-related ER localized protein mRNA, partial cds
class 0	0.6640316	0.70800865	0.6318244	0.5205372 X71973	GPX4 Phospholipid hydroperoxide glutathione peroxidase
class 0	0.66357356	0.7065394	0.6283764	0.5172761 U31342	NUCLEOBINDIN PRECURSOR
class 0	0.66107285	0.70190495	0.6281764	0.515169 U48437	Amyloid precursor-like protein 1 mRNA
class 0	0.6580634	0.6977007	0.6270069	0.5130206 D86957	KIAA0202 gene, partial cds
class 0	0.65573317	0.69658804	0.6243827	0.5114963 HG2279-HT2375	Triosephosphate Isomerase
class 0	0.65531886	0.696354	0.6234574	0.50935465 Z19585	THBS4 Thrombospondin 4
class 0	0.6551901	0.69216716	0.62214273	0.50791293 M16405	MUSCARINIC ACETYLCHOLINE RECEPTOR M4
class 0	0.65306026	0.6912271	0.61901563	0.5059022 U32315	Syntaxin 3 mRNA
class 0	0.65265644	0.69054705	0.61688703	0.50412107 U11701	LIM-homeobox domain protein (hLH-2) mRNA
class 0	0.6510852	0.6855266	0.61317533	0.50281054 Y11251	Novel member of serine-arginine domain protein, SRp129
class 0	0.6451833	0.68519264	0.611776	0.5011752 U68018	Mad protein homolog (hMAD-2) mRNA
class 0	0.6440936	0.6840746	0.6099985	0.50011754 M81181	ATP1B2 ATPase, Na ⁺ /K ⁺ transporting, beta 2 polypeptide
class 0	0.6436302	0.68304396	0.6083414	0.49924028 U58334	Bcl2, p53 binding protein Bbp/53BP2 (BBP/53BP2) mRNA
class 0	0.643048	0.6821534	0.6073584	0.4981423 L38490	ARF4L ADP-ribosylation factor 4-like
class 0	0.6426204	0.6800001	0.6052271	0.49671668 M31328	GNB3 Guanine nucleotide binding protein (G protein), beta polypeptide 3
class 0	0.64044315	0.67821306	0.6023743	0.49373135 M60626	FPR1 Formyl peptide receptor 1
class 0	0.64043903	0.67730945	0.6007907	0.49323085 M97815	CRABP2 Cellular retinoic acid-binding protein 2
class 0	0.63726294	0.675323	0.6001975	0.492158 J00116	COL2A1 Collagen, type II, alpha 1
class 0	0.63437486	0.6752639	0.5996651	0.49088508 U01824	Glutamate transporter
class 0	0.6297416	0.67464525	0.59895456	0.4887451 Y00282	RPN2 Ribophorin II
class 0	0.62950456	0.6733761	0.597548	0.48699924 M65254	PPP2R1B Protein phosphatase 2
class 0	0.6293974	0.6712747	0.59523857	0.4854714 X64877	HFL1 H factor (complement)-like 1
class 0	0.62750775	0.6657862	0.5950493	0.48436904 D31885	KIAA0069 gene, partial cds
class 0	0.6274157	0.66216063	0.5922815	0.48317143 D87434	KIAA0247 gene
class 0	0.6269861	0.6615893	0.5897296	0.48150828 Z35093	SURF1 Surfeit 1
class 0	0.6243887	0.6592147	0.588642	0.48062748 Y08374	GP-39 cartilage protein gene
class 0	0.6233924	0.6570888	0.5881375	0.4795098 Z54367	Plectin
class 0	0.6214661	0.6567585	0.58764184	0.4782393 Z46632	PDE4C Phosphodiesterase 4C
class 0	0.6162652	0.6550915	0.58412486	0.47714055 HG3517-HT3711	Alpha-1-Antitrypsin, 5' End
class 0	0.6158576	0.65411603	0.58396286	0.4756748 L42354	(clone 48ES4) mRNA fragment
class 0	0.61513585	0.65381086	0.58363634	0.47443134 X15376	GABRG2 Gamma-aminobutyric acid (GABA) A receptor, gamma 2
class 0	0.6136919	0.65283734	0.5821644	0.47330046 X89059	Unknown protein expressed in macrophages
class 0	0.6111451	0.65236974	0.5786961	0.4725207 U40998	Retinal protein (HRG4) mRNA
class 0	0.60864455	0.65140575	0.57764935	0.47161415 S72904	APK1 antigen
class 0	0.60758764	0.6513788	0.5764717	0.4703356 M28439	KERATIN, TYPE I CYTOSKELETAL 17
class 0	0.6052617	0.6509427	0.57584333	0.4691536 HG3319-HT3496	Split Gene 1 Enhancer, Tup1-Like
class 0	0.6025101	0.6475502	0.5743594	0.46873093 L20971	PDE4B Phosphodiesterase 4B
class 0	0.60209996	0.64595795	0.5726682	0.46801984 U53174	PPP1CA Protein phosphatase 1, catalytic subunit, alpha isoform
class 0	0.60148424	0.6452151	0.5725452	0.46681368 X12791	SRP19 Signal recognition particle 19 kD protein
class 0	0.59986407	0.64400727	0.572293	0.46536174 M86808	Pyruvate dehydrogenase complex (PDHA2) gene
class 0	0.59955007	0.64397943	0.5716659	0.4651167 U49250	Putative cerebral cortex transcriptional regulator T-Brain-1 (Tbr-1) mRNA
class 0	0.59948987	0.64273274	0.5710794	0.46377957 U70735	34 kDa Mov34 homolog mRNA
class 0	0.59938246	0.6419605	0.5697817	0.4631461 U36448	Ca ²⁺ -dependent activator protein for secretion mRNA
class 0	0.5992474	0.64148456	0.5664956	0.46178034 U67963	Lysophospholipase homolog (HU-K5) mRNA

class 0	0.59904045	0.6412486	0.564561	0.4611744 U61981	MSH3 MutS (E. coli) homolog 3
class 0	0.59882957	0.6405094	0.5638733	0.45987302 U94585	Requiem homolog (hsReq) mRNA
class 0	0.5963293	0.6375841	0.5611557	0.45866847 Y09216	Protein kinase, Dyrk2
class 0	0.59433	0.63753694	0.56094676	0.45810974 X72790	Endogenous retrovirus mRNA for ORF
class 0	0.5938023	0.63601726	0.5596878	0.45745686 D16593	HPCA Hippocalcin
class 0	0.5936155	0.6337328	0.5571045	0.45668325 L25876	Protein tyrosine phosphatase (CIP2)mRNA
class 0	0.59203666	0.63306457	0.55586565	0.45624977 X77197	CLCN4 Chloride channel 4
class 0	0.5918071	0.6329611	0.55581313	0.45501217 M88279	FKBP4 FK506-binding protein 4 (59kD)
class 0	0.5912278	0.6317565	0.5549752	0.45395684 X66087	MYBL1 V-myb avian myeloblastosis viral oncogene homolog-like 1
class 0	0.5896591	0.63113517	0.55333066	0.45320654 U63455	SUR Sulfonylurea receptor (hyperinsulinemia)
class 0	0.58764905	0.63048935	0.5528846	0.4522073 S81003	L-UBC
class 0	0.58715725	0.63013744	0.55262655	0.4510834 M12959_s	TCRA T cell receptor alpha-chain
class 0	0.5870895	0.6294	0.5509657	0.4502766 U82310	Unknown protein mRNA, partial cds
class 0	0.5868637	0.627622	0.55062884	0.44969308 M28879	GRANZYME B PRECURSOR
class 0	0.58671856	0.6271133	0.5484541	0.4487445 U29607	EIF-2-associated p67 homolog mRNA
class 0	0.585308	0.62496793	0.5479123	0.44770333 X69962_s	FMR1 Fragile X mental retardation 1
class 0	0.58297056	0.6241309	0.5472998	0.44696727 U09607	JAK3 Janus kinase 3 (a protein tyrosine kinase, leukocyte)
class 0	0.5828922	0.6234299	0.54680544	0.44528863 U51587	Golgi complex autoantigen golgin-97 mRNA
class 0	0.5822135	0.62112784	0.5465898	0.44471288 U04270	Putative potassium channel subunit (h-erg) mRNA
class 0	0.58068085	0.6209773	0.5451401	0.44420785 D45371	ApM1 mRNA for GS3109 (novel adipose specific collagen-like factor)
class 0	0.5791186	0.6202937	0.5446089	0.44336843 U79528_s	Sigma receptor mRNA
class 0	0.5788491	0.62027115	0.54382896	0.44273722 U52112_ma5	RbP gene (renin-binding protein) extracted from Human Xq28 genomic DNA in the region of the L1CAM locus containing the genes for neural cell adhesion molecule L1 (L1CAM), arginine-vasopressin receptor (AVPR2), C1 p115 (C1), ARD1 N-acetyltransferase related protein (TE2), renin-binding protein (RbP), host cell factor 1 (HCF1), and interleukin-1 receptor-associated kinase (IRAK) genes, and Xq28lu2 gene
class 0	0.5769936	0.6195113	0.5428214	0.44138107 AB002314	KIAA0316 gene
class 0	0.57692134	0.6191036	0.54082024	0.44056216 U36341_ma1	SLC6A8 gene (creatine transporter) extracted from Human Xq28 cosmid, creatine transporter (SLC6A8) gene, and CDM gene, partial cds
class 0	0.57615805	0.61866623	0.5405354	0.43898973 X97544	TIM17 preprotein translocase
class 0	0.57606333	0.61844265	0.53912455	0.4380664 U31628	IL15RA Interleukin 15 receptor alpha chain
class 0	0.57538915	0.61622304	0.5377508	0.43786755 AB003102	Proteasome subunit p44.5
class 0	0.5750055	0.6160982	0.53766185	0.43737975 L00058	MYC V-myc avian myelocytomatosis viral oncogene homolog
class 0	0.5747552	0.61550885	0.5372219	0.436866 L35249_s	ATP6B2 ATPase, H+ transporting, lysosomal (vacuolar proton pump), beta polypeptide, 56/58kD, isoform 2
class 0	0.57380915	0.6145756	0.5359526	0.43563268 M81933	CDC25A Cell division cycle 25A
class 0	0.5736341	0.61447406	0.53550535	0.43463433 D13666_s	Osteoblast specific factor 2 (OSF-2os)
class 0	0.57303506	0.6129059	0.5350379	0.43410015 M37457_s	Na+,K+ -ATPase catalytic subunit alpha-III isoform gene
class 0	0.5719543	0.6121835	0.5343064	0.4337287 J04027	Adenosine triphosphatase mRNA
class 0	0.5713075	0.6111451	0.53332496	0.43279478 L19058	Glutamate receptor (GLUR5) mRNA
class 0	0.57126564	0.6107866	0.53288037	0.43174514 U33821	Tax1-binding protein TXBP151 mRNA
class 0	0.57023346	0.61025316	0.53199846	0.43114722 Y10807_s	Suppressor for yeast mutant
class 0	0.570124	0.6099161	0.53194267	0.43112728 L10955_cds1_s	Carbonic anhydrase IV gene extracted from Human carbonic anhydrase IV gene, promoter region and
class 0	0.569515	0.607602	0.53076357	0.4303349 U32324	Interleukin 11 receptor isoform (incomplete)
class 0	0.56919754	0.6072801	0.52994424	0.4296401 M28882_s	CELL SURFACE GLYCOPROTEIN MUC18 PRECURSOR
class 0	0.56461215	0.6062743	0.5283741	0.42871588 U96629_ma2	2A8.3 gene (hereditary multiple exostoses gene isolog) extracted from Human chromosome 8 BAC clone CIT987SK-2A8 complete sequence
class 0	0.56319255	0.6061561	0.52635425	0.42734823 L40395	(clone S20iii15) mRNA, 3' end of cds

class 0	0.5631556	0.6061188	0.5262581	0.4266111 D82343	AMY
class 0	0.5628468	0.6055819	0.5256596	0.4263719 M74491	ARF3 ADP-ribosylation factor 3
class 0	0.5626699	0.60536665	0.52549785	0.42572156 U20758_rna1	Osteopontin gene
class 0	0.5626479	0.60445035	0.5245646	0.42506737 D26528	RNA helicase
class 0	0.56226707	0.60398006	0.52378994	0.42434895 X85750	Transcript associated with monocyte to macrophage differentiation
class 0	0.56212807	0.6039531	0.5228068	0.4232765 X92762	Tafazzins protein
class 0	0.56127095	0.6039013	0.522677	0.42294946 U03397_s	Receptor protein 4-1BB mRNA
class 0	0.56082225	0.60215974	0.5224149	0.42230636 D79984_s	KIAA0162 gene
class 0	0.5607825	0.6020119	0.52201355	0.42194042 X15341	CYTOCHROME C OXIDASE POLYPEPTIDE VIA-LIVER PRECURSOR
class 0	0.5607145	0.6019508	0.52181506	0.42124954 U52155	Inward rectifier potassium channel Kir1.2 (Kir1.2) mRNA, partial cds
class 0	0.5601084	0.60187197	0.52122813	0.4205988 U70323	SCA2 Spinocerebellar ataxia 2 (olivopontocerebellar ataxia 2, autosomal dominant)
class 0	0.5591348	0.60052645	0.5201653	0.42014632 U89336_cds3	RAGE gene (receptor for advanced glycosylation end products) extracted from Human HLA class III region containing NOTCH4 gene, partial sequence, homeobox PBX2 (HPBX) gene, receptor for advanced glycosylation end products (RAGE) gene, and 6 unidentified cds, complete sequence
class 0	0.55810887	0.59824055	0.51929057	0.4197526 U79734	Huntingtin interacting protein (HIP1) mRNA
class 0	0.5563211	0.59629256	0.5186152	0.41924495 J04823_rna1	Cytochrome c oxidase subunit VIII (COX8) mRNA
class 0	0.554648	0.59499043	0.5172463	0.41886258 X96506_s	NC2 alpha subunit
class 0	0.5546377	0.59397686	0.5169297	0.4185497 X95073	Translin associated protein X
class 0	0.55415726	0.592642	0.5165416	0.41753975 M37457	Na+,K+ -ATPase catalytic subunit alpha-III isoform gene
class 0	0.5532569	0.5926305	0.5156605	0.41711715 HG2797-HT2905	Clathrin, Light Polypeptide B, Alt. Splice 1
class 0	0.5528954	0.59211236	0.515271	0.41627023 Y09980_rna4	HOXD3 gene
class 0	0.5528874	0.5918755	0.51416284	0.415643 L08485	GABRA5 Gamma-aminobutyric acid (GABA) A receptor, alpha 5)
class 0	0.55221575	0.5918504	0.5131231	0.41515335 HG2797-HT2906_s	Clathrin, Light Polypeptide B, Alt. Splice 2
class 0	0.5518999	0.591169	0.512098	0.4149368 M99435	TRANSDUCIN-LIKE ENHANCER PROTEIN 1
class 0	0.5515173	0.5907996	0.5111602	0.4138937 X92720	Phosphoenolpyruvate carboxykinase
class 0	0.55130094	0.58944005	0.5108808	0.41327104 M96995_s	GRB2 Growth factor receptor-bound protein 2
class 0	0.55055606	0.58905005	0.50878483	0.413232 U38904	Zinc finger protein C2H2-25 mRNA
class 0	0.5502738	0.5887711	0.50823826	0.41261023 Y11174	RP3 gene
class 0	0.5501896	0.5877102	0.5071817	0.41216868 U59321	DEAD-box protein p72 (P72) mRNA
class 0	0.5500676	0.5870229	0.50688064	0.4114269 X60592	CD40 CD40 antigen
class 0	0.54953444	0.5865171	0.50673527	0.41095555 U78524	Gu binding protein mRNA, partial cds
class 0	0.5494719	0.5863608	0.50618243	0.41064388 X17648	GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR RECEPTOR ALPHA CHAIN PRECURSOR
class 0	0.54937106	0.5850061	0.50587565	0.40965393 U15932	Protein tyrosine phosphatase mRNA
class 0	0.5493632	0.58423793	0.50495315	0.40887225 X06389	SYP Synaptophysin
class 0	0.5488799	0.58391607	0.5047827	0.40855095 U20499	Estrogen sulfotransferase mRNA
class 0	0.5484502	0.58388555	0.50416595	0.40830544 U21049	DD96 mRNA
class 0	0.5463388	0.5823978	0.5032712	0.407351 AB002380	KIAA0382 gene, partial cds
class 0	0.546322	0.5818097	0.5023653	0.40690655 Y07565_s	RIN Ric (Drosophila)-like (expressed in neurons)
class 0	0.5454462	0.5815752	0.5009951	0.40647876 X72632_s	Rev-ErbAalpha protein (hRev gene)
class 0	0.54503894	0.58104753	0.49995357	0.4060531 M16801	MLR Mineralocorticoid receptor (aldosterone receptor)
class 0	0.5444173	0.58083797	0.49962866	0.40523773 M60922	Surface antigen mRNA
class 0	0.5433324	0.5802529	0.49946222	0.40488547 D88213	Retina-specific amine oxidase
class 0	0.54245716	0.57975286	0.49915946	0.40433422 L07548	ACY1 Aminoacylase 1
class 0	0.54151386	0.5794049	0.49902466	0.40362412 M25269	ELK1 ELK1, member of ETS oncogene family
class 0	0.5409097	0.57867557	0.4990148	0.40250564 J02645	EIF2A Eukaryotic translation initiation factor 2A
class 0	0.5405903	0.57867485	0.49736193	0.40202776 M86933_s	AMELY Amelogenin (chromosome Y encoded)
class 0	0.54021716	0.5782	0.4971733	0.40150848 S43646	KERATIN, TYPE II CYTOSKELETAL 2

					EPIDERMAL
class 0	0.54005593	0.5776254	0.49683756	0.4008628 AF009368	Luman mRNA
class 0	0.5399942	0.5769982	0.49621394	0.40025672 X01630	ASS Argininosuccinate synthetase
class 0	0.53959274	0.5767422	0.495073	0.39979956 U78735	ABC3 ATP-binding cassette 3
class 0	0.5386063	0.57520956	0.49495786	0.39922723 U01038	PLK mRNA
class 0	0.5366796	0.5748313	0.4945322	0.39906326 M55040	ACHE Acetylcholinesterase (YT blood group)
class 0	0.5362463	0.574607	0.49252602	0.39807323 Z48541	Protein tyrosine phosphatase
class 0	0.53534234	0.57422996	0.4920491	0.3976656 X54637	TYK2 Protein-tyrosine kinase tyk2 (non-receptor)
class 0	0.53511363	0.574129	0.49168175	0.39719057 K02777_s	T-cell receptor active alpha-chain mRNA from Jurkat cell line
class 0	0.5336789	0.5738576	0.49123472	0.39659634 U64197	CC chemokine LARC precursor
class 0	0.5328535	0.57379687	0.49086955	0.3959535 X52426_s	KRT13 Keratin 13
class 0	0.53211504	0.5699487	0.490195	0.39569798 X99459	Sigma 3B protein
class 0	0.53210926	0.56966376	0.48971868	0.39414003 L11573	Surfactant protein B mRNA
class 0	0.532053	0.56789446	0.4893898	0.3937184 HG4668-HT5083_s	Transcription Factor Mef2, Alt. Splice 2
class 0	0.53199273	0.56780976	0.48924723	0.39327776 U82979	Immunoglobulin-like transcript-3 mRNA
class 0	0.53159195	0.56724167	0.48922685	0.39241347 M27281	VEGF Vascular endothelial growth factor
class 0	0.5308711	0.56693166	0.48836374	0.3918192 M76559	Neuronal DHP-sensitive, voltage-dependent, calcium channel alpha-2b subunit mRNA
class 0	0.53077525	0.56625986	0.4882309	0.39136678 X59892	TRYPTOPHANYL-TRNA SYNTHETASE
class 0	0.52997464	0.56559825	0.4878811	0.39132828 D16688_s	LTG9/MLLT3 mRNA, C-terminal
class 0	0.528599	0.5648617	0.48774076	0.39066023 U10439	ADAR Double-stranded RNA adenosine deaminase
class 0	0.5282795	0.56396955	0.48766053	0.39063936 X86779	FAST kinase
class 0	0.52772367	0.5631641	0.4875238	0.390095 U08998	TAR RNA binding protein (TRBP) mRNA
class 0	0.5259058	0.56073135	0.48746073	0.38921615 U00944	Clone A9A2BRB6 (CAC)n/(GTG)n repeat-containing mRNA
class 0	0.52554536	0.56066585	0.48733485	0.38901442 U44799_s	U1-snRNP binding protein homolog mRNA
class 0	0.52523357	0.5604787	0.48671508	0.38824853 X76228	ATP6E ATPase, H+ transporting, lysosomal (vacuolar proton pump) 31kD
class 0	0.5247069	0.5601869	0.48609465	0.3877893 AFFX-HUMRGE/M10098_5	AFFX-HUMRGE/M10098_5 (endogenous control)
class 0	0.524563	0.5596961	0.48504826	0.38749382 L43579	L43579 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone 110298, mRNA sequence
class 0	0.52441233	0.5587105	0.48397085	0.38674313 HG2797-HT2905_s	Clathrin, Light Polypeptide B, Alt. Splice 1
class 0	0.5241388	0.55815023	0.48341292	0.3864582 X83416_s	PrP gene, exon 2
class 0	0.5238535	0.5580398	0.48227537	0.38593346 X06268	COL2A1 Collagen, type II, alpha 1 (primary osteoarthritis, spondyloepiphyseal dysplasia, congenital)
class 0	0.52373844	0.55795956	0.48225394	0.38527587 U66619	SWI/SNF complex 60 KDa subunit (BAF60c) mRNA
class 0	0.52261454	0.5571336	0.48142985	0.38485435 X56494	PKM2 Pyruvate kinase, muscle
class 1	0.9920165	1.1499524	1.0030181	0.75236905 HG3543-HT3739	Insulin-Like Growth Factor 2
class 1	0.97661066	0.9616979	0.86222804	0.691033 X53331	MGP Matrix protein gla
class 1	0.9585667	0.8834012	0.81371576	0.6553457 X65724	NDP Norrie disease (pseudoglioma) protein
class 1	0.88205796	0.84205025	0.78712994	0.6343712 D14530	40S RIBOSOMAL PROTEIN S23
class 1	0.880672	0.8215255	0.7616048	0.620643 Y00757	SGNE1 Secretory granule, neuroendocrine protein 1 (7B2 protein)
class 1	0.8662454	0.794612	0.7433427	0.6053828 U25789	Ribosomal protein L21 mRNA
class 1	0.81006235	0.7840071	0.73380023	0.59553546 L27560	Insulin-like growth factor binding protein 5 (IGFBP5) mRNA
class 1	0.8061057	0.7732427	0.71599203	0.5836112 X83543	APXL Apical protein (Xenopus laevis-like)
class 1	0.7963144	0.7665661	0.7039129	0.5745494 X52966	RPL35A Ribosomal protein L35a
class 1	0.7821859	0.76450706	0.69823205	0.56959337 L06797	PROBABLE G PROTEIN-COUPLED RECEPTOR LCR1 HOMOLOG
class 1	0.7820521	0.75147885	0.6901111	0.5607618 M14745	BCL2 B cell lymphoma protein 2
class 1	0.7767357	0.7358279	0.6859018	0.5550981 HG3431-HT3616	Decorin, Alt. Splice 1
class 1	0.77340704	0.7356313	0.6778774	0.548212 D79205	Ribosomal protein L39
class 1	0.7371966	0.72674626	0.6712165	0.5430719 D82345	NB thymosin beta
class 1	0.7175167	0.72600675	0.66350347	0.53756976 D38549	KIAA0068 gene, partial cds

class 1	0.71659106	0.7239192	0.65854394	0.53445965 U14972	Ribosomal protein S10 mRNA
class 1	0.71585566	0.71054196	0.651889	0.5302352 X59841	PRE-B-CELL LEUKEMIA TRANSCRIPTION FACTOR-3
class 1	0.7044602	0.70438224	0.65061134	0.5265213 J03242	IGF2 Insulin-like growth factor 2 (somatomedin A)
class 1	0.7026445	0.70189273	0.6452029	0.5213764 HG3214-HT3391	Metallopanstimulin 1
class 1	0.69295913	0.7003029	0.6411657	0.51902276 X06617	RPS11 Ribosomal protein S11
class 1	0.69290376	0.6944331	0.6401815	0.515944 J02611	APOD Apolipoprotein D
class 1	0.6838895	0.6939349	0.6379803	0.51201856 X16064	TRANSLATIONALLY CONTROLLED TUMOR PROTEIN
class 1	0.6776533	0.69227374	0.632398	0.5075445 Z74616	COL1A2 Collagen, type I, alpha-2
class 1	0.67510426	0.6836942	0.6293568	0.5051634 L40386	DP2 (Humdp2) mRNA
class 1	0.67472875	0.6825732	0.6266333	0.5020662 X04741	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L1
class 1	0.6703414	0.6777735	0.6220077	0.49927366 M55210	LAMC1 Laminin, gamma 1 (formerly LAMB2)
class 1	0.66915506	0.6772996	0.6169228	0.49835536 M96739	NSCL-1 mRNA sequence
class 1	0.66442496	0.67441475	0.6109734	0.4961448 U73304	CB1 cannabinoid receptor (CNR1) gene
class 1	0.6633565	0.67025864	0.6074427	0.49294007 M65292	HFL1 H factor (complement)-like 1
class 1	0.66297793	0.66304475	0.6040194	0.49020073 HG311-HT311	Ribosomal Protein L30
class 1	0.6621168	0.6628506	0.60389394	0.4881786 M83233	TCF12 Transcription factor 12
class 1	0.65106505	0.6614697	0.6003128	0.4862768 U07919	ALDH6 Aldehyde dehydrogenase 6
class 1	0.6483891	0.660038	0.5960164	0.48381075 X57959	RPL17 Ribosomal protein L7
class 1	0.64783746	0.6586179	0.5918964	0.482616 J04080	C1S Complement component 1, s subcomponent
class 1	0.6444846	0.6569774	0.5899946	0.4810916 X76029	NEUROMEDIN U-25 PRECURSOR
class 1	0.6431463	0.6527849	0.5874605	0.47882786 U14973	40S RIBOSOMAL PROTEIN S29
class 1	0.64175445	0.64769405	0.5862101	0.47660354 U24576	Breast tumor autoantigen mRNA, complete sequence
class 1	0.63975364	0.6461164	0.58385	0.47526005 L41066	NF-AT3 mRNA
class 1	0.63081163	0.6454896	0.58339643	0.47360942 X60489	Elongation factor-1-beta
class 1	0.62705797	0.6454799	0.5795754	0.4714921 M62843	PARANEOPLASTIC ENCEPHALOMYELITIS ANTIGEN HUD
class 1	0.6269411	0.6337208	0.5765729	0.4702747 HG662-HT662	Epstein-Barr Virus Small Rna-Associated Protein
class 1	0.62111175	0.63143706	0.5743106	0.46925655 HG613-HT613	Ribosomal Protein S12
class 1	0.6195382	0.630508	0.57252336	0.46820134 L42379	Quiescin (Q6) mRNA, partial cds
class 1	0.6190998	0.62700677	0.5705358	0.46622977 M13241	N-MYC PROTO-ONCOGENE PROTEIN
class 1	0.6153043	0.62574947	0.5684561	0.46493605 U12404	HSPB1 Heat shock 27kD protein 1
class 1	0.6119269	0.62252253	0.56778365	0.4618914 M55998	Alpha-1 collagen type I gene, 3' end
class 1	0.6047074	0.6202116	0.5667907	0.46059835 S82240	RhoE
class 1	0.5998984	0.61648834	0.5644458	0.45931065 U78027	L44L gene
class 1	0.5997888	0.61597705	0.56125534	0.457502 X06700	COL3A1 Alpha-1 type 3 collagen
class 1	0.5986623	0.6130399	0.55806524	0.45596722 D13413	Tumor-associated 120 kDa nuclear protein p120
class 1	0.59865105	0.6114418	0.556301	0.4549787 M74719	SEF2-1A protein (SEF2-1A) mRNA, 5' end
class 1	0.5893714	0.60740095	0.5559502	0.45420048 M93119	INSM1 Insulinoma-associated 1 (symbol provisional)
class 1	0.5881829	0.6069417	0.55419403	0.45310345 M92287	CCND3 Cyclin D3
class 1	0.5843839	0.60632783	0.5535847	0.4518154 HG33-HT33	Ribosomal Protein S4, X-Linked
class 1	0.58205545	0.60504496	0.55251557	0.4503006 U16306	CSPG2 Chondroitin sulfate proteoglycan 2 (versican)
class 1	0.57964575	0.60460865	0.55130297	0.44919688 Z37976	LTBP2 Latent transforming growth factor beta binding protein 2
class 1	0.57818264	0.6033843	0.548655	0.4471849 X69150	Ribosomal protein S18
class 1	0.5697472	0.6027349	0.54859287	0.44544208 HG4542-HT4947	Ribosomal Protein L10
class 1	0.56644344	0.60101384	0.54806477	0.4440113 L41607	GCNT2 Glucosaminyl (N-acetyl) transferase 2, I-branching enzyme
class 1	0.56627625	0.5981668	0.54536986	0.44266245 U43148	PTCH Patched (Drosophila) homolog
class 1	0.56476176	0.5980047	0.54293185	0.4405897 M30269	NID Nidogen (enactin)
class 1	0.5646692	0.5970898	0.54111236	0.4390807 X07384	GLI Glioma-associated oncogene homolog (zinc finger protein)
class 1	0.5645039	0.5934657	0.5405502	0.4387671 L38941	RPL37 Ribosomal protein L37
class 1	0.5623646	0.5929238	0.5385063	0.436879 U09953	RPL9 Ribosomal protein L9

class 1	0.5605586	0.59292054	0.5378748	0.43572348 D87464	KIAA0274 gene
class 1	0.5601569	0.5907391	0.53619635	0.43403146 M18000	40S RIBOSOMAL PROTEIN S17
class 1	0.55792636	0.5882164	0.53568417	0.43313977 M91196	ICSBP1 Interferon consensus sequence binding protein 1
class 1	0.5550686	0.5858174	0.5354681	0.432014 L27559	IGFBP5 Insulin-like growth factor binding protein 5
class 1	0.5531683	0.5847639	0.5341864	0.4305753 S76475	NTRK3 Neurotrophic tyrosine kinase, receptor, type 3 (TrkC)
class 1	0.5519678	0.5820916	0.5326356	0.4299089 L41067	Transcription factor NFATx mRNA
class 1	0.5519046	0.5801515	0.53112125	0.42853457 X15940	RPL31 Ribosomal protein L31
class 1	0.54831284	0.57951117	0.5298944	0.42775545 M13934	RPS14 gene (ribosomal protein S14)
class 1	0.54646355	0.57748485	0.52962327	0.42654306 J04164	RPS3 Ribosomal protein S3
class 1	0.5452459	0.5765971	0.5289784	0.4260657 D14678	Kinesin-related protein, partial cds
class 1	0.54087687	0.57616264	0.52835536	0.425178 M31520	Ribosomal protein S24
class 1	0.53762615	0.57376033	0.52642053	0.42397496 Z25749	Ribosomal protein S7
class 1	0.53477347	0.5735111	0.5255214	0.42304212 D87735	CAG-isl 7 (trinucleotide repeat-containing sequence)
class 1	0.5327469	0.57256067	0.5238885	0.42232138 X74295	ITGA7 Integrin, alpha 7B
class 1	0.5313974	0.56816363	0.5215355	0.42137793 M77232	Ribosomal protein S6 gene and flanking regions
class 1	0.52145535	0.5667701	0.5209896	0.42066354 U29195	NPTX2 Neuronal pentraxin II
class 1	0.51958996	0.565056	0.5194922	0.4199424 X67734	AXONIN-1 PRECURSOR
class 1	0.5176871	0.56338716	0.51761514	0.41912028 HG4319-HT4589	Ribosomal Protein L5
class 1	0.51742566	0.5625303	0.51681256	0.4180165 M14764	NGFR Nerve growth factor receptor
class 1	0.515282	0.5620647	0.5142898	0.41645116 X69391	RPL6 Ribosomal protein L6
class 1	0.5145462	0.5597948	0.5122884	0.41600755 L37043	CSNK1E Casein kinase 1, epsilon
class 1	0.51452804	0.5585209	0.5122436	0.4148983 HG3364-HT3541	Ribosomal Protein L37
class 1	0.5133517	0.55822176	0.50993776	0.41432607 D82348	5-aminoimidazole-4-carboxamide-1-beta-D-ribonucleotide de transformylase/inosinase
class 1	0.51271385	0.55787575	0.50949436	0.41301724 M64716	RPS25 Ribosomal protein S25
class 1	0.5108828	0.55782914	0.50928015	0.4125411 M81757	40S RIBOSOMAL PROTEIN S19
class 1	0.5108767	0.557722	0.50865674	0.41186035 X79234	Ribosomal protein L11
class 1	0.5083245	0.5567942	0.5078859	0.41117817 D13627	KIAA0002 gene
class 1	0.5066584	0.555303	0.5063182	0.41037014 M17254	TRANSFORMING PROTEIN ERG
class 1	0.50393724	0.55510676	0.50627005	0.40977025 D63476	KIAA0142 gene
class 1	0.5006533	0.55366576	0.5056975	0.40813982 Z46629	SOX9 SRY (sex-determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)
class 1	0.4997523	0.5518851	0.5054221	0.40693 X53777	60S RIBOSOMAL PROTEIN L23
class 1	0.49702138	0.55170476	0.50420064	0.4063421 M98045	Folypolyglutamate synthetase mRNA
class 1	0.49669904	0.55143607	0.5025265	0.40599987 D23660	RPL4 Ribosomal protein L4
class 1	0.4962502	0.55131245	0.502122	0.40507796 X17254	GATA1 Transcription factor Eryf1
class 1	0.495279	0.550734	0.5010209	0.4050369 D79989	KIAA0167 gene
class 1	0.49403584	0.5506776	0.49984783	0.4037516 U25165	Fragile X mental retardation protein 1 homolog FXR1 mRNA
class 1	0.49375808	0.55031294	0.49885884	0.40304828 U02031	Sterol regulatory element binding protein-2 mRNA
class 1	0.49297127	0.5496973	0.49817213	0.4027928 HG3510-HT3704	V-Erba Related Ear-3 Protein
class 1	0.4915902	0.54927313	0.49810657	0.40183237 Z74615	COL1A1 Collagen, type I, alpha 1
class 1	0.4908611	0.5478711	0.49570867	0.4006703 U58682	RPS28 Ribosomal protein S28
class 1	0.4892058	0.54716134	0.49524176	0.39926985 HG384-HT384	Ribosomal Protein L26
class 1	0.48803678	0.5469765	0.49498755	0.39867598 X67247_rna1	RpS8 gene for ribosomal protein S8
class 1	0.48778662	0.5467125	0.49310726	0.39855596 L37868_s	POU-domain transcription factor (N-Oct-3)
class 1	0.4868394	0.5464856	0.49219152	0.3974036 Y09836	3'UTR of unknown protein
class 1	0.48679835	0.5445919	0.49152628	0.39659202 U03105	B4-2 protein mRNA
class 1	0.48197228	0.54264337	0.49125198	0.39608747 X80909	Alpha NAC mRNA
class 1	0.48074037	0.5422034	0.490522	0.39525136 D86982	KIAA0229 gene, partial cds
class 1	0.47754747	0.54169583	0.49043605	0.39439687 M64099	GAMMA-GLUTAMYLTRANSPEPTIDASE 5 PRECURSOR
class 1	0.47734782	0.54160315	0.48973796	0.39351916 D15050	Transcription factor AREB6
class 1	0.47724903	0.54061115	0.487093	0.39287397 U38846	Stimulator of TAR RNA binding (SRB) mRNA

class 1	0.47694737	0.5393285	0.4867437	0.39244702 U26726	11 beta-hydroxysteroid dehydrogenase type II mRNA
class 1	0.47501153	0.53837866	0.48582202	0.39175433 U43901_rna1_s	37 kD laminin receptor precursor/p40 ribosome associated protein gene
class 1	0.47327766	0.537135	0.48561665	0.3909066 M30448_s	Casein kinase II beta subunit mRNA
class 1	0.47271132	0.53606623	0.48516303	0.38992578 HG821-HT821	Ribosomal Protein S13
class 1	0.47148794	0.53591806	0.48515683	0.38910928 X56932	LCAT Lecithin-cholesterol acyltransferase
class 1	0.47017133	0.5355234	0.48478094	0.38841477 D87433	KIAA0246 gene, partial cds
class 1	0.46981534	0.53283656	0.48465174	0.38810182 L77886	Protein tyrosine phosphatase mRNA
class 1	0.46942267	0.5318093	0.48331326	0.38732716 X04325	GJB1 Gap junction protein, beta 1, 32kD (connexin 32, Charcot-Marie-Tooth neuropathy, X-linked)
class 1	0.46733344	0.53141624	0.4826737	0.3858285 X62691	40S RIBOSOMAL PROTEIN S15A
class 1	0.46700636	0.52987796	0.4821765	0.38530836 M23613	NPM1 Nucleophosmin (nucleolar phosphoprotein B23, numatrin)
class 1	0.46547055	0.52970636	0.4814417	0.3844768 HG2994-HT4850_s	Elastin, Alt. Splice 2
class 1	0.46264017	0.52904475	0.480897	0.3832074 HG2873-HT3017	Ribosomal Protein L30 Homolog
class 1	0.461879	0.528705	0.48009557	0.38286644 U80628	Thymidine kinase 2 (TK2) mRNA
class 1	0.46098718	0.5282426	0.48005548	0.3827526 X55715	RPS3 Ribosomal protein S3
class 1	0.45965302	0.5280224	0.47943622	0.3822391 X07173	INTER-ALPHA-TRYPSIN INHIBITOR COMPLEX COMPONENT II PRECURSOR
class 1	0.45465934	0.52741534	0.47873193	0.381494 Z11793	Selenoprotein P
class 1	0.4498619	0.5268396	0.47792363	0.38107353 M14058	C1R Complement component C1r
class 1	0.44982657	0.52639043	0.47577056	0.38061887 X56997_rna1	UbA52 gene coding for ubiquitin-52 amino acid fusion protein
class 1	0.44927323	0.5257839	0.47518098	0.3798872 M31520_rna1_s	Unknown protein gene extracted from Human ribosomal protein S24 mRNA
class 1	0.44878542	0.5257578	0.4742035	0.37922156 HG4716-HT5158	Guanosine 5'-Monophosphate Synthase
class 1	0.44854823	0.5250102	0.47346252	0.3788581 U29943_s	ELAV-like neuronal protein-2 Hel-N2 mRNA
class 1	0.44723308	0.52477926	0.47323424	0.37794998 X03342	RPL32 Ribosomal protein L32
class 1	0.446246	0.5233783	0.47306406	0.37778178 D86961	KIAA0206 gene, partial cds
class 1	0.4457633	0.5233542	0.47277132	0.3767854 M87789_s	(hybridoma H210) anti-hepatitis A IgG variable region, constant region, complementarity-determining regions mRNA
class 1	0.44380325	0.52199423	0.47275352	0.37630454 X12671_rna1	Hnmp a1 protein gene extracted from Human gene for heterogeneous nuclear ribonucleoprotein (hnRNP) core protein A1
class 1	0.44341132	0.52142936	0.47238418	0.3753195 M62402	IGFBP6 Insulin-like growth factor binding protein 6
class 1	0.44313014	0.52122724	0.47210145	0.37470126 D42123	ESP1/CRP2
class 1	0.4424557	0.5204168	0.4712368	0.37411138 M13450	ESD Esterase D/formylglutathione hydrolase
class 1	0.44135985	0.52034736	0.47011894	0.37278944 L41349	PLCB4 Phospholipase C, beta 4
class 1	0.4395942	0.52032745	0.46904048	0.37265533 U27655	RGP3 mRNA
class 1	0.43931964	0.52012825	0.46877444	0.37242573 Y08915	Alpha 4 protein
class 1	0.43852326	0.5195769	0.46851683	0.37194404 J03507	C7 Complement component 7
class 1	0.43827492	0.5183492	0.467564	0.37121144 D13370	DNA-(APURINIC OR APYRIMIDINIC SITE) LYASE
class 1	0.43730646	0.5180041	0.46579874	0.37099004 D87460	KIAA0270 gene, partial cds
class 1	0.43694746	0.5175007	0.4651442	0.3703059 U14970	RPS5 Ribosomal protein S5
class 1	0.43632242	0.51683414	0.46464762	0.3700658 X99325	Alpha-tubulin mRNA
class 1	0.4341169	0.51539576	0.46385196	0.36979818 S79522	UBA52 Ubiquitin A-52 residue ribosomal protein fusion product 1
class 1	0.43373224	0.51529187	0.46378323	0.36874855 U83411	Carboxypeptidase Z precursor, mRNA
class 1	0.43370542	0.5148967	0.46348524	0.3684767 U13616	ANK3 Ankyrin G
class 1	0.43326634	0.514756	0.4634733	0.36746866 L04483_s	RPS21 Ribosomal protein S21
class 1	0.4322948	0.513936	0.46344924	0.3671752 J00314	mRNA fragment encoding beta-tubulin. (from clone D-beta-1)
class 1	0.42978635	0.51332444	0.46247655	0.36685425 X04347_s	Liver mRNA fragment DNA binding protein UPI homologue (C-terminus)
class 1	0.42779183	0.5132742	0.4624057	0.36615053 U08096	Peripheral myelin protein-22 (PMP22) gene, non-coding exon 1B
class 1	0.4273552	0.5130134	0.4621326	0.36527485 U31814	Transcriptional regulator homolog RPD3 mRNA
class 1	0.42513227	0.51109487	0.46213248	0.36480188 D13988	Rab GDI mRNA
class 1	0.42448258	0.51027435	0.4615859	0.3641077 HG1515-	Transcription Factor Btf3b

HT1515_f					
class 1	0.4230558	0.51021904	0.4607458	0.3631419 M84711	RPS3A Ribosomal protein S3A
class 1	0.42282543	0.5098846	0.45967656	0.36288628 L13698	GAS1 Growth arrest-specific 1
class 1	0.4221029	0.5096207	0.45967004	0.36251333 L07919	Homeodomain protein DLX-2 mRNA, 3' end
class 1	0.4207693	0.50906163	0.45930198	0.36156428 L06505	RPL12 Ribosomal protein L12
class 1	0.4203779	0.5077076	0.45851415	0.36111102 L07648	MX11 mRNA
class 1	0.42026496	0.5075645	0.45740777	0.360508 X07438_s	DNA for cellular retinol binding protein (CRBP) exons 3 and 4
class 1	0.41933295	0.5074675	0.45598942	0.36015892 HG2383-HT4824_s	Cystathionine Beta Synthase, Alt. Splice 3
class 1	0.41929802	0.5062536	0.45589927	0.3594407 M86667	HnRNP C2 protein mRNA
class 1	0.41805196	0.50479084	0.45509034	0.3591927 S77094	Muscle acetylcholine receptor alpha-subunit
class 1	0.41662169	0.50435406	0.45392206	0.35893115 U18422	DP2 (Humdp2) mRNA
class 1	0.41592053	0.5040681	0.45345718	0.35854766 X58529	IGHM Immunoglobulin mu
class 1	0.41580322	0.503959	0.452932	0.35770673 M60854	RPS16 Ribosomal protein S16
class 1	0.4155998	0.5024803	0.45250317	0.3574924 D87292	Rhodanese
class 1	0.41409877	0.5021771	0.4519568	0.3567802 M22382	HSPD1 Heat shock 60 kD protein 1 (chaperonin)
class 1	0.4123219	0.50217485	0.45182624	0.35641807 X64707	60S RIBOSOMAL PROTEIN L13
class 1	0.4120609	0.5017571	0.4515104	0.3557218 D90209	ATF4 CAMP-dependent transcription factor ATF-4 (CREB2)
class 1	0.4119028	0.5017274	0.45102918	0.35539606 D61391	Phosphoribosypyrophosphate synthetase-associated protein 39
class 1	0.4078506	0.5014528	0.45040438	0.3544836 L29008	SORD Sorbitol dehydrogenase
class 1	0.40782854	0.50134575	0.44985783	0.3539076 S69265_s	Neuron-specific RNA recognition motifs (RRMs)-containing protein [human, hippocampus, mRNA, 1992 nt]
class 1	0.407235	0.50109625	0.44926503	0.35373962 X86570	Acidic hair keratin 1
class 1	0.40639907	0.5003693	0.44891816	0.3526457 U31556	E2F5 E2F transcription factor 5, p130-binding
class 1	0.40480205	0.49891683	0.4488706	0.35251245 U09937_rna1_s	Urokinase-type plasminogen activator receptor gene extracted from Human urokinase-type plasminogen receptor
class 1	0.40304074	0.498528	0.44867498	0.35219634 Z26653	LAMA2 Laminin, alpha 2 (merosin, congenital muscular dystrophy)
class 1	0.40300912	0.4983526	0.44849887	0.35170025 HG3039-HT3200	Adp-Ribosylation-Like Factor
class 1	0.40173146	0.49770778	0.4484872	0.35161403 J04621	SDC2 Syndecan 2 (heparan sulfate proteoglycan 1, cell surface-associated, fibroglycan)
class 1	0.40006796	0.4975498	0.4481869	0.35051855 U77846_rna1_s	Elastin gene, partial cds and partial 3'UTR
class 1	0.39813024	0.49706382	0.44777563	0.35011983 X03689_s	mRNA fragment for elongation factor TU (N-terminus)
class 1	0.3974015	0.4966702	0.4467744	0.3493167 L22548	COL18A1 Collagen, type XVIII, alpha 1
class 1	0.39731237	0.49588135	0.44673792	0.34877178 U53445	Ovarian cancer downregulated myosin heavy chain homolog (Doc1) mRNA
class 1	0.39550346	0.49571368	0.4466238	0.34861916 HG3549-HT3751	Wilm'S Tumor-Related Protein
class 1	0.39484447	0.4953815	0.44644707	0.34847924 M55409_s	EEF1G Translation elongation factor 1 gamma
class 1	0.3947713	0.4953512	0.44526893	0.34809443 X03100_cds2	HLA-SB alpha gene (class II antigen) extracted from Human HLA-SB(DP) alpha gene
class 1	0.3944759	0.49516183	0.44511965	0.34766483 K03430	C1QB Complement component 1, q subcomponent, beta polypeptide
class 1	0.39365962	0.49513546	0.44488803	0.3471049 X65867	ADENYLOSUCCINATE LYASE
class 1	0.3936173	0.4951319	0.4441889	0.34675756 D50926	KIAA0136 gene, partial cds
class 1	0.39248642	0.49410003	0.4416117	0.346463 M17863_s	IGF2 Insulin-like growth factor 2 (somatomedin A)
class 1	0.39113295	0.4940771	0.44132063	0.34596905 X63527	GAPD Glyceraldehyde-3-phosphate dehydrogenase
class 1	0.39072797	0.49286294	0.44131553	0.3458115 HG2686-HT2782	Ryanodine Receptor 3
class 1	0.38937965	0.49171317	0.44126588	0.34559935 U94855	Translation initiation factor 3 47 kDa subunit mRNA
class 1	0.3876931	0.4914396	0.4409931	0.34535253 X95876	G-protein coupled receptor

Classic vs. desmoplastic MD prediction results (*k*-NN).

This section contains the detailed sample predictions and error rates of predicting classic vs. desmoplastic in leave-one-out cross-validation with a *k*-nearest neighbor algorithm.

The model predicts 33 out of 34 samples correctly and it is clearly highly significant (P-val = 0.0000008630, see the calculation below and the [Proportional chance criterion](#)).

Classic vs. desmoplastic medulloblastoma prediction *k*-nearest neighbors algorithm

Dataset B

Values thresholded to 20 from below and 16000 from above

Variation filter: max/min > 3 (3-fold), max-min= 100 absolute units

Confusion Matrix

Actual	Predicted		
	Classic	Desmoplastic	
Classic	24	1	25
Desmoplastic	0	9	9
	24	10	34

Proportional chance criterion (see chapter VII of Huberty's Applied Discriminant Analysis)

$$\begin{aligned}
 C_{pro} &= (25/34) * (25/34) + (9/34) * (9/34) \\
 C_{pro} &= 0.610727 \\
 P_{cc} &= 33/34 = 0.970588 \\
 (P_{cc} - C_{pro}) / \sqrt{C_{pro}(1-C_{pro})/n} &= Z = 4.783099 \quad P\text{-val} = 0.0000008630
 \end{aligned}$$

Num Data	Num Right	Num Wrong	Threshold	Num Abstain	Abs Error	ROC Error
	34	33	1	0	0	0.029412
	34	5				0.02
Datapoint	Predicted Class	Confidence	True Class	Error?		
Brain_MD_7	0	0.302768	0			
Brain_MD_59	0	0.548534	0			

Brain_MD_20	0	0.61025	0
Brain_MD_21	0	0.283638	0
Brain_MD_50	0	0.149308	0
Brain_MD_49	1	0.09796	0 *
Brain_MD_45	0	0.369326	0
Brain_MD_43	0	0.59691	0
Brain_MD_8	0	0.41708	0
Brain_MD_42	0	0.233207	0
Brain_MD_1	0	0.012653	0
Brain_MD_4	0	0.014019	0
Brain_MD_55	0	0.321458	0
Brain_MD_41	0	0.511905	0
Brain_MD_37	0	0.33274	0
Brain_MD_3	0	0.042915	0
Brain_MD_34	0	0.364898	0
Brain_MD_29	0	0.599312	0
Brain_MD_13	0	0.546489	0
Brain_MD_24	0	0.438043	0
Brain_MD_65	0	0.706403	0
Brain_MD_5	0	0.249954	0
Brain_MD_66	0	0.398238	0
Brain_MD_67	0	0.464782	0
Brain_MD_58	0	0.576908	0
Brain_MD_53	1	2.22E-04	1
Brain_MD_56	1	0.079909	1
Brain_MD_16	1	0.033374	1
Brain_MD_40	1	0.008221	1
Brain_MD_35	1	0.35569	1
Brain_MD_30	1	0.141191	1
Brain_MD_23	1	0.307987	1
Brain_MD_28	1	6.87E-04	1
Brain_MD_60	1	0.269587	1

SOM clustering of treatment outcome samples.

In order to study the unsupervised intrinsic structure of the medulloblastoma data we clustered the samples using a SOM algorithm. We performed multiple clusterings to make the sure the results were robust and reproducible and selected the most common clustering results as a representative. This is the 2-cluster scheme shown below that separates the medulloblastomas in two groups: C0 with 23 and C1 with 37 samples.

The only clinical attribute or observation that appears to correlate with this discovered classes is the abundance of ribosomal protein-encoding genes. See for example the list of marker genes for these classes in the [SOM-discovered C0 vs. C1 class gene markers](#) section.

There is a correlation of these C0 and C1 groups with outcome but it is barely significant and it does not provide an accurate predictor of outcome (confusion matrix Fisher test P-value=0.104 and survival rank-log test P-value=0.027, see calculations below). The error rate of such a predictor will be worse than the corresponding to multi-gene models, staging or TrkC (see [Summary of medulloblastoma treatment outcome predictions](#) section.)

At the same time it is clear these unsupervised classes provide a background on which the outcome markers behave differently (see the [Treatment outcome markers](#) section).

Medulloblastoma Treatment outcome Clustering

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

SOM 2x1 250,000 iterations

(most common clustering of samples in 10 independent clusterings)

	C0	C1		
A	18	21	39	
D	5	16	21	
	23	37	60	
Fisher exact test	p-value = 0.1042			

Proportional chance criterion (see chapter VII of Huberty's Applied Discriminant Analysis)

$$C_{pro} = (39/60) * (39/60) + (21/60) * (21/60)$$

$$C_{pro} = 0.545$$

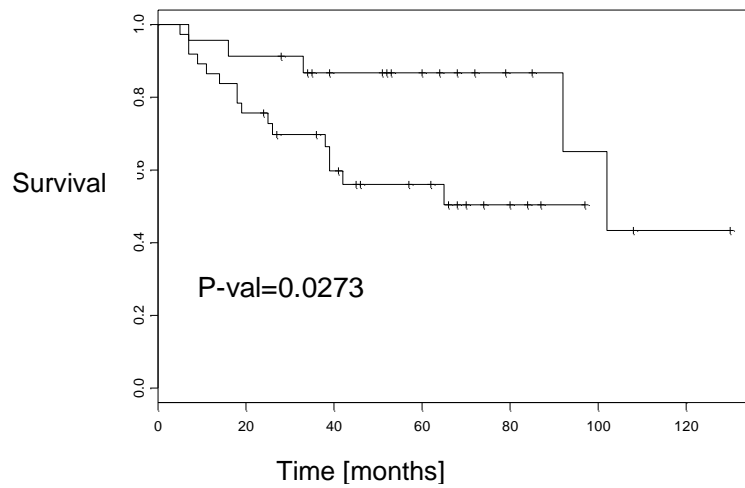
$$P_{cc} = 34/60 = 0.5666667$$

$$(P_{cc} - C_{pro}) / \sqrt{C_{pro}(1 - C_{pro})/n} = Z = 0.281976 \quad p\text{-val} = 0.388980856$$

Datapoint	C0=0/C1=1	distance	Outcome (1=D,0=A)
Brain_MD_1	1	46686.465	1
Brain_MD_2	1	62372.953	1
Brain_MD_3	1	53835.16	1
Brain_MD_4	1	38884.28	1
Brain_MD_5	0	89883.66	1
Brain_MD_6	1	40543.027	1
Brain_MD_7	1	46187.44	1
Brain_MD_8	0	51966.36	1
Brain_MD_9	1	42860.812	1
Brain_MD_10	1	51593.727	1
Brain_MD_11	1	33562.94	1
Brain_MD_12	1	58727.18	1
Brain_MD_13	1	52192.195	1
Brain_MD_14	0	99314.85	1
Brain_MD_15	1	44546.17	1
Brain_MD_16	1	40649.52	1
Brain_MD_17	1	51659.758	1
Brain_MD_18	1	53809.836	1
Brain_MD_19	1	51028.08	1
Brain_MD_20	0	49023.414	1
Brain_MD_21	0	59584.203	1
Brain_MD_22	1	77951.586	0
Brain_MD_23	1	49253.03	0
Brain_MD_24	1	43594.652	0
Brain_MD_25	0	80744.79	0
Brain_MD_26	0	71646.86	0
Brain_MD_27	0	45744.094	0
Brain_MD_28	0	59030.145	0
Brain_MD_29	0	52226.234	0
Brain_MD_30	1	48958.145	0
Brain_MD_31	0	48885.035	0
Brain_MD_32	1	60523.902	0
Brain_MD_33	1	42020.992	0
Brain_MD_34	1	44006.707	0
Brain_MD_35	1	59958.125	0
Brain_MD_36	1	40265.812	0
Brain_MD_37	0	71715.74	0
Brain_MD_38	0	73755.11	0
Brain_MD_39	0	45035.742	0
Brain_MD_40	1	51523.34	0
Brain_MD_41	0	54920.566	0
Brain_MD_42	1	44616.004	0
Brain_MD_43	0	49523.766	0
Brain_MD_44	1	47706.28	0
Brain_MD_45	0	41383.582	0
Brain_MD_46	1	74529.46	0
Brain_MD_47	1	44987.453	0
Brain_MD_48	0	50744.977	0
Brain_MD_49	1	43264.297	0
Brain_MD_50	0	50196.22	0
Brain_MD_51	0	74102.5	0
Brain_MD_52	1	73643.53	0

Brain_MD_53	1	33567.434	0
Brain_MD_54	0	54177.47	0
Brain_MD_55	1	45652.695	0
Brain_MD_56	1	47614.305	0
Brain_MD_57	1	45471.82	0
Brain_MD_58	0	63106.32	0
Brain_MD_59	0	104183.734	0
Brain_MD_60	1	62373.688	0

Survival Analysis

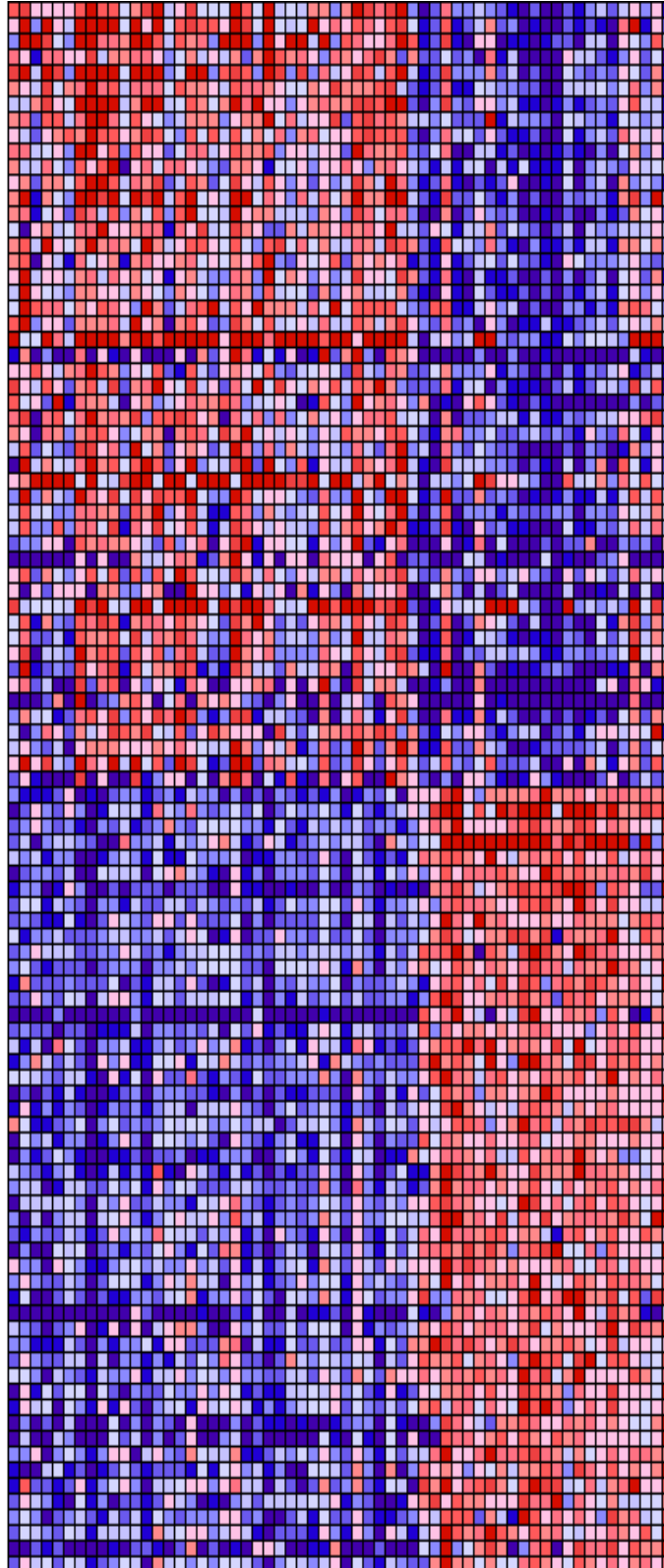


SOM-discovered C0 vs. C1 class gene markers

This picture shows some of the top markers that differentiate the C0 and C1 discovered classes sorted by their signal to noise ratios as described in the [Gene marker selection](#) section.

C1

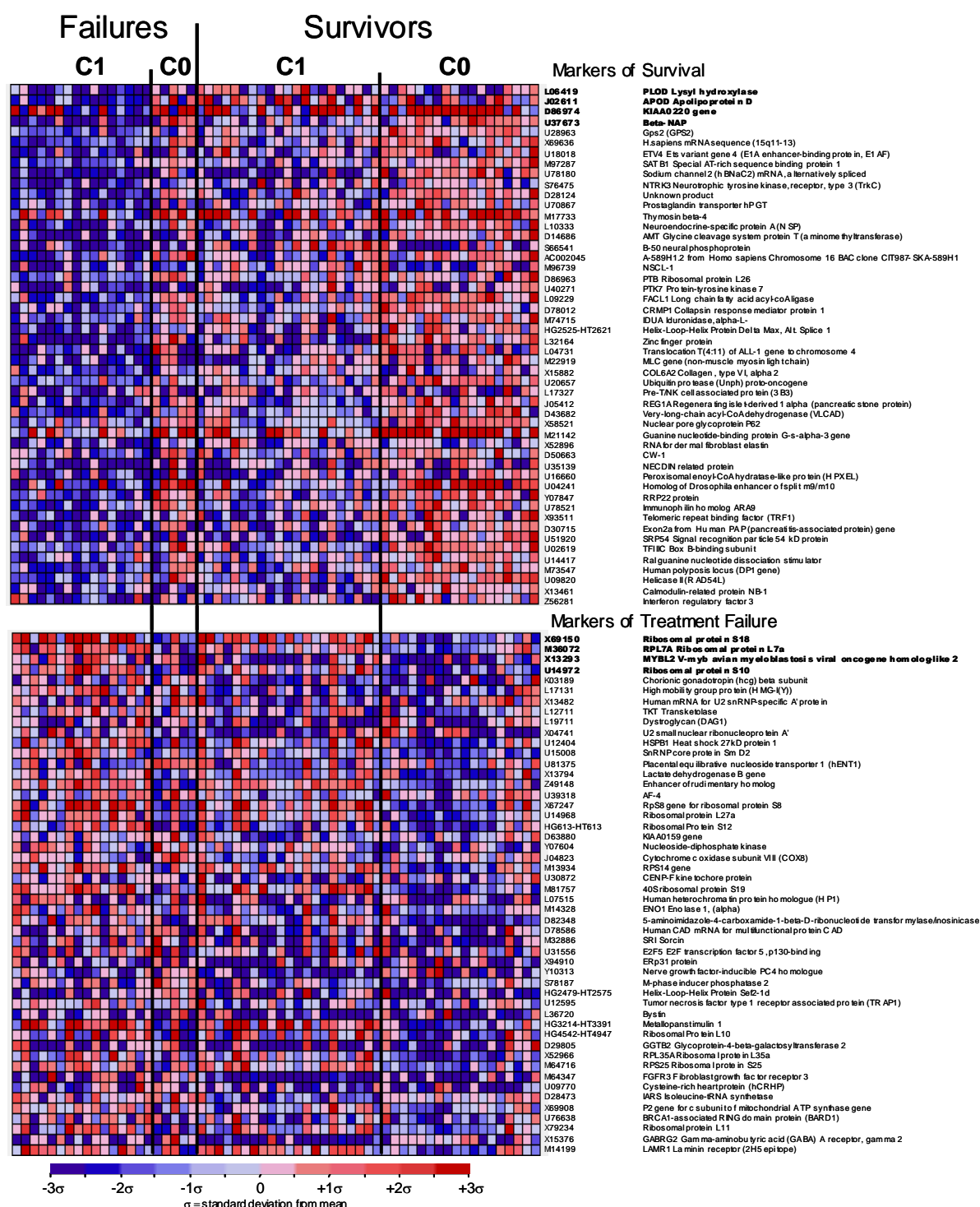
C0



M81757_at	40S RIBOSOMAL PROTEIN S19
M18000_at	40S RIBOSOMAL PROTEIN S17
HG3214-HT339	Metallopanstimulin 1
L06505_at	RPL12 Ribosomal protein L12
X69150_at	Ribosomal protein S18
Z49148_s_at	Enhancer of rudimentary homolog mRNA
U14970_at	RPS5 Ribosomal protein S5
X15940_at	RPL31 Ribosomal protein L31
L11566_at	RPL18 Ribosomal protein L18
Z25749_rna1_a	Ribosomal protein S7
L33842_rna1_a	IMPDH2 gene
U14972_at	Ribosomal protein S10 mRNA
X67247_rna1_a	RpS8 gene for ribosomal protein S8
X53777_at	60S RIBOSOMAL PROTEIN L23
D87735_at	CAG-isl 7
M13934_cds2_	RPS14 gene (ribosomal protein S14)
X64707_at	60S RIBOSOMAL PROTEIN L13
X62691_at	40S RIBOSOMAL PROTEIN S15A
M31520_rna1_s	Unknown n protein extracted from Human ribosomal protein S24
M14199_s_at	LAMR1 Laminin receptor (2H5 epitope)
U14968_at	Ribosomal protein L27a mRNA
M24194_at	Alpha-tubulin mRNA
L06797_s_at	PROBABLE G PROTEIN-COUPLED RECEPTOR LCR1 HOMOLOG
U15008_at	SnRNP core protein Sm D2 mRNA
Z28407_at	RPL8 Ribosomal protein L8
D82348_at	5-aminoimidazole-4-carboxamide-1-beta
M36072_at	RPL7A Ribosomal protein L7a
M64716_at	RPS25 Ribosomal protein S25
X52966_at	RPL35A Ribosomal protein L35a
L04483_s_at	RPS21 Ribosomal protein S21
X80822_at	60S RIBOSOMAL PROTEIN L18A
U12404_at	HSPB1 Heat shock 27kD protein 1
U58682_at	RPS28 Ribosomal protein S28
U14971_at	RPS9 Ribosomal protein S9
K03460_at	Alpha-tubulin isotype H2-alpha gene, last exon
U26726_at	11 beta-hydroxysteroid dehydrogenase type II mRNA
X55954_at	RPL17 Ribosomal protein L17
U21090_at	DNA polymerase delta small subunit mRNA
D23660_at	RPL4 Ribosomal protein L4
X55715_at	RPS3 Ribosomal protein S3
M77232_rna1_a	Ribosomal protein S6 gene and flanking regions
X79234_at	Ribosomal protein L11
U78027_rna3_a	L44L gene
L21954_at	PERIPHERAL-TYPE BENZODIAZEPINE RECEPTOR
L40386_s_at	DP2 (Humdp2) mRNA
HG613-HT613_	Ribosomal Protein S12
Z48501_s_at	Polyadenylate binding protein II
U25789_at	Ribosomal protein L21 mRNA
HG1800-HT182	Ribosomal Protein S20
U07919_at	ALDH6 Aldehyde dehydrogenase 6
J05243_at	SPTAN1 Spectrin, alpha, non-erythrocytic 1 (alpha-fodrin)
Z48950_at	HISTONE H3.3
Y08200_at	Rab geranylgeranyl transferase, alpha-subunit
L11672_at	ZNF91 Zinc finger protein 91 (HPF7, HTF10)
M74491_at	ARF3 ADP-ribosylation factor 3
L36151_at	PHOSPHATIDYLINOSITOL 4-KINASE ALPHA
X99664_at	Protein containing SH3 domain, SH3GL3
U79260_at	Clone 23745 mRNA
X04526_at	GNB1 Guanine nucleotide binding protein (G protein)
U68142_at	RaiGDS-like 2 (RGL2) mRNA, partial cds
U44839_at	Putative ubiquitin C-terminal hydrolase (UHX1) mRNA
Y09616_at	Carboxylesterase (hCE-2) mRNA
U49089_at	Neuroendocrine-dlg (NE-dlg) mRNA
U20285_at	Gps1 (GPS1) mRNA
L20971_at	PDE4B Phosphodiesterase 4B, cAMP-specific
L11373_at	Protocadherin 43 mRNA for abbreviated PC43
X98801_at	Dynactin
U40391_rna1_a	Serotonin N-acetyltransferase gene
M62762_at	ATP6C Vacuolar H+ ATPase proton channel subunit
D63851_at	Unc-18 homologue
Z71460_at	Vacuolar-type H(+)-ATPase 115 kDa subunit
X65784_s_at	Cellular adhesion regulatory molecule [human, mRNA, 429 nt]
X99688_at	mRNA from TYL gene
M60165_cds1_	HLA-DQB1 gene
J04615_at	SNRPN Small nuclear ribonucleoprotein polypeptide N
M69013_at	GNA11 Guanine nucleotide binding protein (G protein)
U66617_at	SWI/SNF complex 60 KDa subunit (BAF60a)
U93237_rna2_a	MEN1 gene (menin)
U81802_at	Phosphatidylinositol 4-kinase
U24152_at	P21-activated protein kinase (Pak1) gene
HG651-HT4201	Adducin, Alpha Subunit, Alt. Splice 2
X96484_at	DGCR6 protein
U95740_rna2_a	362G6.2 gene extracted from Human chromosome 16p13.1
D87465_at	KIAA0275 gene
X75962_at	OX40L RECEPTOR PRECURSOR
U59302_at	Steroid receptor coactivator (SRC-1) mRNA
U27831_at	Striatum-enriched phosphatase (STEP) mRNA, partial cds
U04811_at	Trophinin mRNA
X12492_at	CCAAT BOX-BINDING TRANSCRIPTION FACTOR 1
U89336_cds3_	RAGE gene
M24899_at	THRA Thyroid hormone receptor
D87743_at	KIAA0267 gene, partial cds
D87437_at	KIAA0250 gene
X90530_at	RagB protein
U15172_at	Nip1 (NIP1) mRNA

Treatment outcome markers

This picture shows some of the top 50 markers of the treatment failure vs. survival distinction. The genes are sorted by their signal to noise ratios as described in [Gene marker selection](#) section. The table below shows the top 100 markers for each tumor class including the permutation test values (see [Permutation-based neighborhood analysis for marker gene](#)). The samples are sorted according to treatment outcome status and then by membership to the unsupervised SOM-discovered classes C0 and C1. Notice the different behavior of markers according to the sample membership in those classes. For example the *low in failures/high in survivors* markers do not distinguish very well the failure samples that belong to the C0 class.



Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

Dataset C

class 0 = High in failure class, low in survivors class

class 1 = High in survivors class, low in failure class

Class	Distance	Perm 1%	Perm 5%	Median (50%)	Feature	Desc
class0	0.79	0.8575851	0.7458573	0.57256466	X69150_at	Ribosomal protein S18
class0	0.58	0.7196305	0.6636454	0.5213177	M36072_at	RPL7A Ribosomal protein L7a
class0	0.52	0.6967163	0.62734157	0.49769974	X13293_at	MYBL2 V-myb avian myeloblastosis viral oncogene homolog-like 2
class0	0.43	0.6554627	0.6037056	0.48016676	U14972_at	Ribosomal protein S10 mRNA
class0	0.39	0.6267573	0.5743042	0.4491265	K03189_f_at	Chorionic gonadotropin (hcg) beta subunit mRNA
class0	0.37	0.6124803	0.5631084	0.44041193	L17131_rna1_at	High mobility group protein (HMG-I(Y)) gene exons 1-8
class0	0.37	0.6079583	0.5574408	0.4283559	X13482_at	U2 SMALL NUCLEAR RIBONUCLEOPROTEIN A'
class0	0.36	0.5920371	0.5417171	0.42088938	L12711_s_at	TKT Transketolase (Wernicke-Korsakoff syndrome)
class0	0.36	0.5873178	0.5382286	0.41427517	L19711_at	Dystroglycan (DAG1) mRNA
class0	0.35	0.5774571	0.5246414	0.40221128	X04741_at	UBIQUITIN CARBOXYL-TERMINAL HYDROLASE ISOZYME L1
class0	0.35	0.5691193	0.52006763	0.39952728	U12404_at	HSPB1 Heat shock 27kD protein 1
class0	0.35	0.564594	0.5123353	0.39575094	U15008_at	SnRNP core protein Sm D2 mRNA
class0	0.34	0.5508108	0.5075411	0.39274555	U81375_at	Placental equilibrative nucleoside transporter 1 (hENT1) mRNA
class0	0.34	0.545049	0.5041218	0.3862151	X13794_rna1_at	Lactate dehydrogenase B gene exon 1 and 2 (EC 1.1.1.27) (and joined CDS)
class0	0.33	0.5437621	0.4992256	0.38400882	Z49148_s_at	Enhancer of rudimentary homolog mRNA
class0	0.33	0.5414132	0.49528137	0.380192	U39318_at	AF-4 mRNA
class0	0.33	0.5383373	0.49330828	0.37714508	X67247_rna1_at	RpS8 gene for ribosomal protein S8
class0	0.33	0.5350176	0.4877165	0.37104735	U14968_at	Ribosomal protein L27a mRNA
class0	0.33	0.5349308	0.48364687	0.36859724	HG613-HT613_at	Ribosomal Protein S12
class0	0.32	0.5341373	0.48146704	0.36665422	D63880_at	KIAA0159 gene
class0	0.32	0.5304447	0.47949836	0.3642997	Y07604_at	Nucleoside-diphosphate kinase
class0	0.32	0.5302321	0.47619662	0.3608576	J04823_rna1_at	Cytochrome c oxidase subunit VIII (COX8) mRNA
class0	0.31	0.5268929	0.47357216	0.35861334	M13934_cds2_at	RPS14 gene (ribosomal protein S14) extracted from Human ribosomal protein S14 gene
class0	0.3	0.5230379	0.4727932	0.35644948	U30872_at	CENP-F kinetochore protein mRNA
class0	0.3	0.517165	0.471055	0.3537757	M81757_at	40S RIBOSOMAL PROTEIN S19
class0	0.3	0.510794	0.46896455	0.35145545	L07515_at	HETEROCHROMATIN PROTEIN 1 HOMOLOG
class0	0.3	0.5095285	0.46605366	0.35070398	M14328_s_at	ENO1 Enolase 1, (alpha)
class0	0.29	0.508716	0.46462357	0.34898144	D82348_at	5-aminoimidazole-4-carboxamide-1-beta-D-ribose nucleotide transformylase/inosinase
class0	0.29	0.508341	0.46285483	0.3464003	D78586_at	CAD PROTEIN
class0	0.29	0.5031112	0.46092945	0.34413984	M32886_at	SRI Sorcin
class0	0.28	0.5001085	0.45781645	0.3421644	U31556_at	E2F5 E2F transcription factor 5, p130-binding
class0	0.27	0.4985688	0.4531834	0.34044018	X94910_at	ERp31 protein
class0	0.27	0.4967446	0.45302796	0.33841276	Y10313_at	Nerve growth factor-inducible PC4 homologue
class0	0.27	0.4962007	0.45097864	0.3369098	S78187_at	M-PHASE INDUCER PHOSPHATASE 2
class0	0.26	0.4961145	0.4478716	0.3364148	HG2479-HT2575_s_at	Helix-Loop-Helix Protein Sef2-1d

class0	0.26	0.4889722	0.44644976	0.334412	U12595_at	Tumor necrosis factor type 1 receptor associated protein (TRAP1) mRNA, partial cds
class0	0.26	0.4863545	0.44478703	0.33138537	L36720_at	Bystin mRNA
class0	0.26	0.4863338	0.444101	0.32964143	HG3214-HT3391_at	Metallopanstimulin 1
class0	0.26	0.4848765	0.44319925	0.3288533	HG4542-HT4947_at	Ribosomal Protein L10
class0	0.25	0.4844726	0.44242045	0.32686168	D29805_at	GGTB2 Glycoprotein-4-beta-galactosyltransferase 2
class0	0.25	0.4840101	0.4404198	0.32587272	X52966_at	RPL35A Ribosomal protein L35a
class0	0.25	0.4818713	0.43920797	0.3253415	M64716_at	RPS25 Ribosomal protein S25
class0	0.25	0.4806192	0.4372189	0.3233852	M64347_at	FGFR3 Fibroblast growth factor receptor 3 (achondroplasia, thanatophoric dwarfism)
class0	0.25	0.4799493	0.4344101	0.32173064	U09770_at	Cysteine-rich heart protein (hCRHP) mRNA
class0	0.24	0.478935	0.43330073	0.32071036	D28473_s_at	IARS Isoleucine-tRNA synthetase
class0	0.24	0.4770883	0.43237802	0.319521	X69908_rna1_at	P2 gene for c subunit of mitochondrial ATP synthase gene extracted from H.sapiens gene for mitochondria
class0	0.24	0.4735355	0.43201992	0.31852308	U76638_at	BRCA1-associated RING domain protein (BARD1) mRNA
class0	0.24	0.4734543	0.4286573	0.31740758	X79234_at	Ribosomal protein L11
class0	0.24	0.4714663	0.42765373	0.31620368	X15376_at	GABRG2 Gamma-aminobutyric acid (GABA) A receptor, gamma 2
class0	0.24	0.4707498	0.42616725	0.31488287	M14199_s_at	LAMR1 Laminin receptor (2H5 epitope)
class1	0.68	0.8203825	0.71479475	0.54178566	L06419_at	PLOD Lysyl hydroxylase
class1	0.68	0.6813562	0.61684877	0.4866072	J02611_at	APOD Apolipoprotein D
class1	0.62	0.6525392	0.5892669	0.46121535	D86974_at	KIAA0220 gene, partial cds
class1	0.58	0.6052752	0.55940294	0.4377526	U37673_at	Neuron-specific vesicle coat protein and cerebellar degeneration antigen (beta-NAP) mRNA
class1	0.54	0.5903411	0.5472241	0.4268065	U28963_at	Gps2 (GPS2) mRNA
class1	0.53	0.5833272	0.5351269	0.41475436	X69636_at	mRNA sequence (15q11-13)
class1	0.52	0.5681573	0.5214865	0.40589228	U18018_at	ETV4 Ets variant gene 4 (E1A enhancer-binding protein, E1AF)
class1	0.51	0.5620877	0.5118141	0.39511982	M97287_at	SATB1 Special AT-rich sequence binding protein 1 (binds to nuclear matrix/scaffold-associating DNA's)
class1	0.51	0.5489072	0.5050725	0.3881474	U78180_at	Sodium channel 2 (hBNAC2) mRNA, alternatively spliced
class1	0.5	0.5460772	0.49833363	0.38249692	S76475_at	NTRK3 Neurotrophic tyrosine kinase, receptor, type 3 (TrkC)
class1	0.49	0.5402584	0.49208724	0.37573755	D28124_at	Unknown product
class1	0.47	0.5376918	0.49007148	0.37205702	U70867_at	Prostaglandin transporter hPGT mRNA
class1	0.47	0.5376203	0.48863164	0.36825827	M17733_at	Thymosin beta-4 mRNA
class1	0.47	0.5331166	0.48450905	0.36425692	L10333_s_at	Neuroendocrine-specific protein A (NSP) mRNA
class1	0.46	0.5309144	0.47723606	0.35968268	D14686_at	AMT Glycine cleavage system protein T (aminomethyltransferase)
class1	0.46	0.5209687	0.47456706	0.35637328	S66541_s_at	B-50=neural phosphoprotein [human, Genomic, 778 nt, segment 3 of 3]
class1	0.46	0.5196613	0.47301164	0.3522324	AC002045_xpt2_s_at	A-589H1.2 from Homo sapiens Chromosome 16 BAC clone CIT987-SKA-589H1 ~complete genomic sequ
class1	0.46	0.5092424	0.47074348	0.34869587	M96739_at	NSCL-1 mRNA sequence
class1	0.45	0.502768	0.46288192	0.3469092	D86963_at	PTB Ribosomal protein L26
class1	0.44	0.5001023	0.46171203	0.3424973	U40271_s_at	PTK7 Protein-tyrosine kinase 7
class1	0.44	0.4967905	0.4602382	0.34013447	L09229_s_at	FACL1 Long chain fatty acid acyl-coA ligase
class1	0.43	0.4930968	0.457573	0.33526275	D78012_at	CRMP1 Collapsin response mediator protein 1
class1	0.43	0.4905391	0.455364	0.33248144	M74715_s_at	IDUA Iduronidase, alpha-L-
class1	0.43	0.4900592	0.45425656	0.32911295	HG2525-HT2621_at	Helix-Loop-Helix Protein Delta Max, Alt. Splice 1
class1	0.43	0.4890777	0.4522365	0.32650998	L32164_at	Zinc finger protein mRNA, 3' end
class1	0.42	0.4890256	0.44947585	0.32418716	L04731_at	Translocation T(4:11) of ALL-1 gene to chromosome 4
class1	0.42	0.4885243	0.44821537	0.32321975	M22919_rna2_at	MLC gene (non-muscle myosin light chain) extracted from Human nonmuscle/smooth muscle alkali myosin
class1	0.42	0.4878781	0.44515502	0.3214133	X15882_at	COL6A2 Collagen, type VI, alpha 2
class1	0.42	0.484461	0.4439538	0.3199113	U20657_at	Ubiquitin protease (Unph) proto-oncogene mRNA
class1	0.42	0.4837765	0.442202	0.31707913	L17327_at	Pre-T/NK cell associated protein (3B3) mRNA, 3' end
class1	0.41	0.4793581	0.4388235	0.3152247	J05412_at	REG1A Regenerating islet-derived 1 alpha (pancreatic stone protein, pancreatic thread protein)
class1	0.41	0.4784921	0.43725225	0.3134606	D43682_s_at	Very-long-chain acyl-CoA dehydrogenase (VLCAD)
class1	0.41	0.4770678	0.43665424	0.3111634	X58521_at	NUCLEAR PORE GLYCOPROTEIN P62
class1	0.41	0.4761319	0.43582407	0.3108418	M21142_cds2_s_at	Guanine nucleotide-binding protein G-s-alpha-3 gene extracted from Human guanine nucleotide-binding pr
class1	0.4	0.4710167	0.43269396	0.30824196	X52896_s_at	RNA for dermal fibroblast elastin
class1	0.4	0.4703343	0.43064603	0.30662587	D50663_at	CW-1 mRNA

class1	0.4	0.4676111	0.42923966	0.30482316	U35139_at	NECDIN related protein mRNA
class1	0.4	0.4663965	0.4282913	0.30239248	U16660_at	Peroxisomal enoyl-CoA hydratase-like protein (HPXEL) mRNA
class1	0.4	0.464419	0.42767128	0.30059886	U04241_at	Homolog of Drosophila enhancer of split m9/m10 mRNA
class1	0.4	0.4624396	0.42370704	0.29999858	Y07847_at	RRP22 protein
class1	0.4	0.4604367	0.4224223	0.2977264	U78521_at	Immunophilin homolog ARA9 mRNA
class1	0.39	0.4591597	0.4217393	0.29728124	X93511_s_at	Telomeric repeat binding factor (TRF1) mRNA
class1	0.39	0.4569826	0.4172285	0.2956864	D30715_xpt5_s_at	Exon2a from Human PAP (pancreatitis-associated protein) gene, 5'-flanking region./ntype=DNA /annot=ex
class1	0.39	0.4565771	0.4171863	0.2946754	U51920_at	SRP54 Signal recognition particle 54 kD protein
class1	0.39	0.4550281	0.41622347	0.29249102	U02619_at	TFIIIC Box B-binding subunit mRNA
class1	0.39	0.454998	0.41258836	0.2910374	U14417_at	Ral guanine nucleotide dissociation stimulator mRNA, partial cds
class1	0.39	0.4549076	0.41155508	0.28867468	M73547_at	POLYPOSIS LOCUS PROTEIN 1
class1	0.39	0.4530375	0.41020757	0.28755918	U09820_s_at	Helicase II (RAD54L) mRNA
class1	0.39	0.4513509	0.40929455	0.2865325	X13461_s_at	CALMODULIN-RELATED PROTEIN NB-1
class1	0.39	0.4505063	0.40849882	0.28548497	Z56281_at	Interferon regulatory factor 3

k-nearest neighbors treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for the *k*-nearest neighbor algorithm.

Medulloblastoma Treatment outcome prediction

k-Nearest Neighbors Algorithm

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

Number of features (genes) = 8. Median based feature selection. K=5, 1/distance weighting

4459 genes pass the filter.

Dataset C

Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	37	2	39
Failures	11	10	21
	48	12	60

Fisher exact test P-val= 0.0002

Proportional chance criterion (see chapter VII of Huberty's Applied Discriminant Analysis)

Cpro= $(39/60)*(39/60) + (21/60)*(21/60)$

Cpro= 0.545

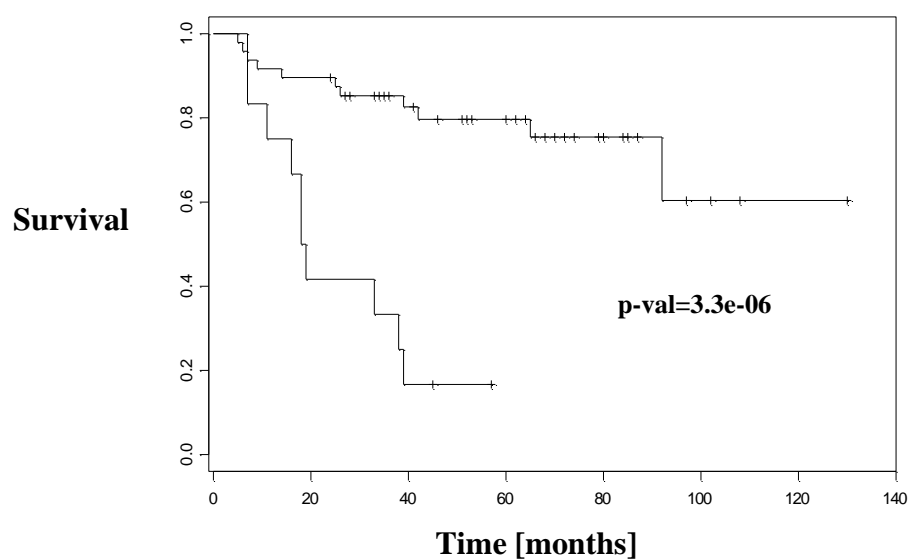
Pcc= $47/60 = 0.783333$

$(Pcc - Cpro)/\sqrt{Cpro(1-Cpro)/n} = Z = 3.101741$ p-val = 0.000962000

Num Data	Num Right	Num Wrong	Threshold	Num Abstain	Abs Error	ROC Error
60	47	13	0	0	0.21667	0.288
60	5					
Datapoint	Predicted Class	Confidence	True Class	Error?		
Brain_MD_1	0	0	0			
Brain_MD_2	1	0.014714	0 *			
Brain_MD_3	1	0.051905	0 *			
Brain_MD_4	0	0.016268	0			
Brain_MD_5	1	0.244053	0 *			
Brain_MD_6	0	0.00339	0			
Brain_MD_7	1	0.174131	0 *			
Brain_MD_8	1	0.021684	0 *			
Brain_MD_9	0	0	0			
Brain_MD_10	0	0.077769	0			
Brain_MD_11	0	0.238361	0			
Brain_MD_12	0	0.395321	0			
Brain_MD_13	1	0.117485	0 *			
Brain_MD_14	1	0.621401	0 *			
Brain_MD_15	0	0.163458	0			
Brain_MD_16	0	0.001465	0			
Brain_MD_17	0	0.368863	0			

Brain_MD_18	1	0.196423	0 *
Brain_MD_19	1	0.083515	0 *
Brain_MD_20	1	0.131556	0 *
Brain_MD_21	1	0.236	0 *
Brain_MD_22	1	0.119483	1
Brain_MD_23	1	0.449442	1
Brain_MD_24	1	0.087128	1
Brain_MD_25	1	0.002469	1
Brain_MD_26	1	0.004054	1
Brain_MD_27	1	0.229156	1
Brain_MD_28	1	0.214794	1
Brain_MD_29	1	0.132556	1
Brain_MD_30	1	0.004142	1
Brain_MD_31	1	0.071982	1
Brain_MD_32	1	0.15699	1
Brain_MD_33	1	0.070619	1
Brain_MD_34	1	0.086266	1
Brain_MD_35	1	0.095713	1
Brain_MD_36	0	0.134619	1 *
Brain_MD_37	1	0.115611	1
Brain_MD_38	1	1.27E-04	1
Brain_MD_39	1	0.085404	1
Brain_MD_40	1	0.175227	1
Brain_MD_41	0	0.001709	1 *
Brain_MD_42	1	0.434137	1
Brain_MD_43	1	0.042809	1
Brain_MD_44	1	0.038684	1
Brain_MD_45	1	0.012557	1
Brain_MD_46	1	0.190361	1
Brain_MD_47	1	0.078001	1
Brain_MD_48	1	0.028872	1
Brain_MD_49	1	0.209988	1
Brain_MD_50	1	0.440045	1
Brain_MD_51	1	0.186536	1
Brain_MD_52	1	0.32828	1
Brain_MD_53	1	0.01044	1
Brain_MD_54	1	0.096563	1
Brain_MD_55	1	0.00146	1
Brain_MD_56	1	0.310485	1
Brain_MD_57	1	0	1
Brain_MD_58	1	0.00709	1
Brain_MD_59	1	0.010633	1
Brain_MD_60	1	0.059727	1

Kaplan Meier Plot



Permutation test for k-nearest neighbor outcome predictor

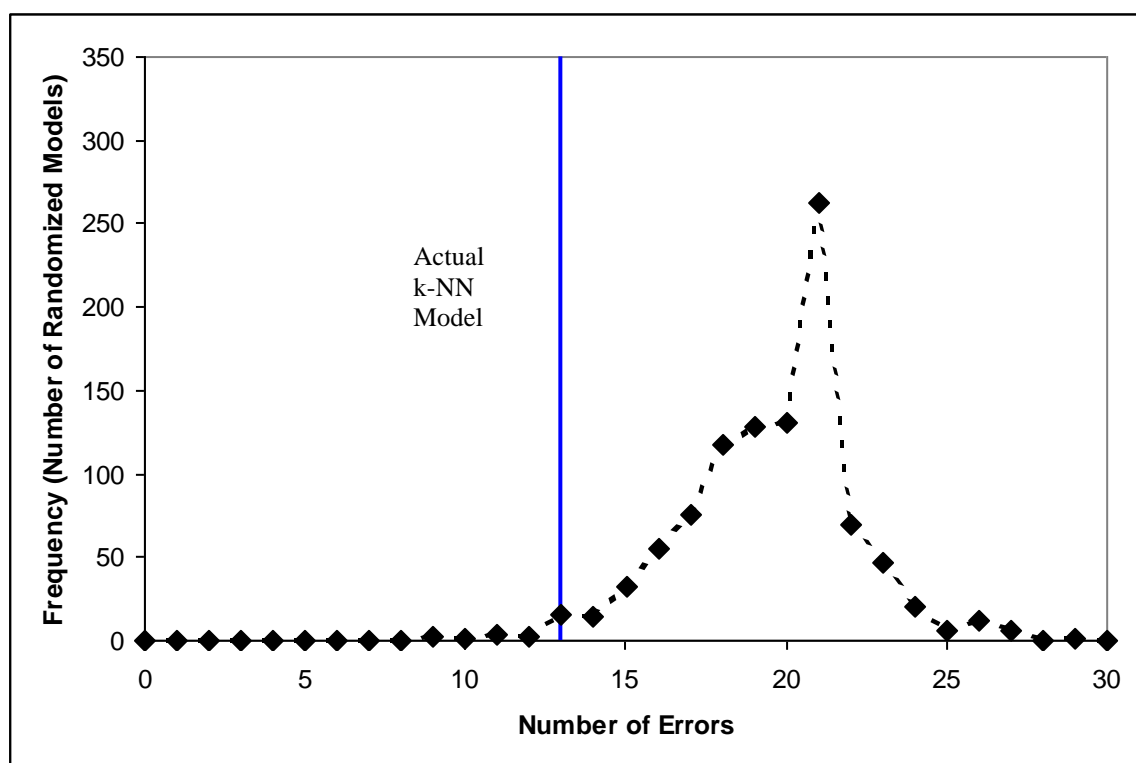
The picture below shows the permutation test for the k-nearest neighbor predictor using the method described in the [Permutation Test for Outcome Predictor](#) section.

Number of genes parameter (ng) values: 1,2,3,4,5,6,7,8,9,10,15,25,50,100

Number of neighbors (k) parameter values: 3, 5

Number of random permutations was 1000

There are 9 k-NN random models better (lower error rates) than the actual k-NN model ($k=5$, $ng=8$) that achieves 13 errors. The significance is $9/1000 = 0.009$.



Weighted voting treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for the weighted voting algorithm.

Medulloblastoma treatment outcome prediction

Weighted voting algorithm

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

4459 genes pass the filter.

Dataset C

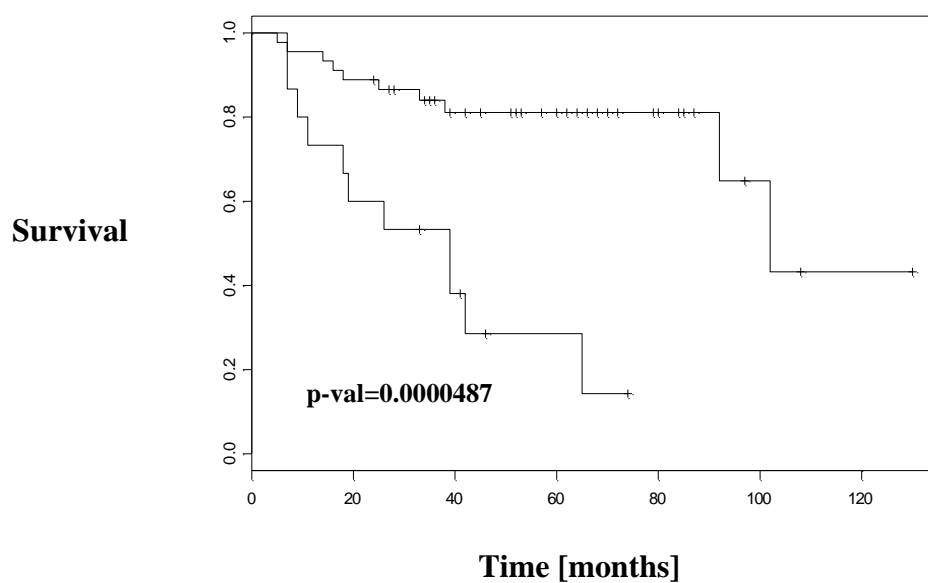
Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	35	4	39
Failures	10	11	21
	45	15	60

Datapoint	Predicted Class	Confidence	True Class	Error?	Final	Pred Thres=0.3
Brain_MD_1	0	0.3952327	0			0
Brain_MD_2	1	0.4901857	0 *			1*
Brain_MD_3	0	0.7813152	0			0
Brain_MD_4	0	0.5952852	0			0
Brain_MD_5	1	0.3449895	0 *			1*
Brain_MD_6	0	0.6714032	0			0
Brain_MD_7	1	0.0853546	0 *			1*
Brain_MD_8	1	0.0744516	0 *			1*
Brain_MD_9	1	0.339672	0 *			1*
Brain_MD_10	0	0.8249887	0			0
Brain_MD_11	0	0.4271654	0			0
Brain_MD_12	0	0.1754278	0			1*
Brain_MD_13	0	0.3934312	0			0
Brain_MD_14	1	0.7060123	0 *			1*
Brain_MD_15	1	0.2107395	0 *			1*
Brain_MD_16	0	0.3573091	0			0
Brain_MD_17	0	0.3510691	0			0
Brain_MD_18	0	0.4114339	0			0
Brain_MD_19	0	0.3019681	0			0
Brain_MD_20	1	0.5024587	0 *			1*
Brain_MD_21	1	0.4159824	0 *			1*
Brain_MD_22	1	0.0565264	1			1
Brain_MD_23	1	0.3096231	1			1
Brain_MD_24	1	0.0712419	1			1
Brain_MD_25	1	0.3827674	1			1
Brain_MD_26	0	0.3012823	1 *			0*

Brain_MD_27	1	0.3565034	1	1
Brain_MD_28	1	0.2837476	1	1
Brain_MD_29	1	0.1805679	1	1
Brain_MD_30	0	0.2770377	1 *	1
Brain_MD_31	0	0.0251648	1 *	1
Brain_MD_32	1	0.2175978	1	1
Brain_MD_33	0	0.4701843	1 *	0*
Brain_MD_34	1	0.6038642	1	1
Brain_MD_35	0	0.2912066	1 *	1
Brain_MD_36	0	0.6073407	1 *	0*
Brain_MD_37	1	0.0385693	1	1
Brain_MD_38	1	0.1748749	1	1
Brain_MD_39	1	0.1087455	1	1
Brain_MD_40	1	0.0164151	1	1
Brain_MD_41	1	0.0581653	1	1
Brain_MD_42	1	0.0824984	1	1
Brain_MD_43	1	0.0644064	1	1
Brain_MD_44	1	0.0228837	1	1
Brain_MD_45	1	0.4116783	1	1
Brain_MD_46	1	0.3631051	1	1
Brain_MD_47	1	0.4536451	1	1
Brain_MD_48	1	0.3325118	1	1
Brain_MD_49	0	0.3832514	1 *	0*
Brain_MD_50	1	0.3337134	1	1
Brain_MD_51	1	0.6800034	1	1
Brain_MD_52	1	0.5439253	1	1
Brain_MD_53	0	0.2974122	1 *	1
Brain_MD_54	1	0.166894	1	1
Brain_MD_55	0	0.0892733	1 *	1
Brain_MD_56	1	0.1744895	1	1
Brain_MD_57	1	0.0187014	1	1
Brain_MD_58	1	0.1895063	1	1
Brain_MD_59	1	0.6713425	1	1
Brain_MD_60	1	0.2848721	1	1

Kaplan Meier Plot



SVM treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for the SVM algorithm.

Medulloblastoma treatment outcome prediction SVM algorithm

150 genes

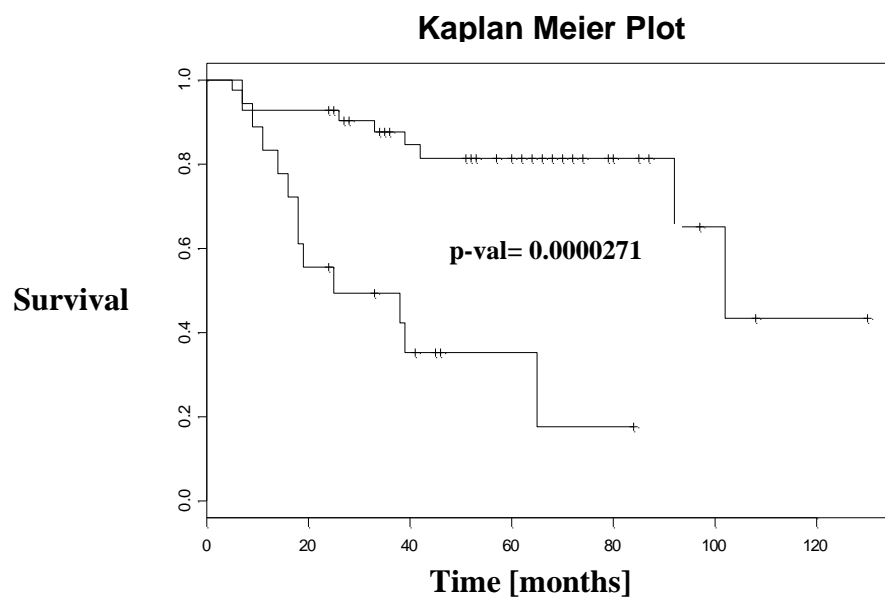
Dataset C

Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	33	6	39
Failures	9	12	21
	42	18	60

Datapoint	Predicted Class	Confidence	True Class	Error?
Brain_MD_1	0	-0.793188		0
Brain_MD_2	1	1.35161		0*
Brain_MD_3	0	-0.0212459		0
Brain_MD_4	1	0.349387		0*
Brain_MD_5	1	2.80968		0*
Brain_MD_6	0	-0.0663716		0
Brain_MD_7	0	-0.442181		0
Brain_MD_8	0	-0.155642		0
Brain_MD_9	0	-0.259622		0
Brain_MD_10	0	-0.82724		0
Brain_MD_11	0	-0.921629		0
Brain_MD_12	0	-0.267656		0
Brain_MD_13	1	1.08133		0*
Brain_MD_14	1	1.50771		0*
Brain_MD_15	0	-0.338922		0
Brain_MD_16	1	0.460525		0*
Brain_MD_17	0	-0.587502		0
Brain_MD_18	1	1.04191		0*
Brain_MD_19	0	-0.303932		0
Brain_MD_20	1	0.472321		0*
Brain_MD_21	1	2.14838		0*
Brain_MD_22	0	-0.908658		1*
Brain_MD_23	1	1.6444		1
Brain_MD_24	1	0.290781		1
Brain_MD_25	1	0.991721		1
Brain_MD_26	0	-0.850044		1*
Brain_MD_27	1	0.786961		1

Brain_MD_28	1	2.54499	1
Brain_MD_29	1	0.305534	1
Brain_MD_30	1	1.18804	1
Brain_MD_31	1	0.692038	1
Brain_MD_32	1	0.0597337	1
Brain_MD_33	0	-0.601687	1*
Brain_MD_34	1	0.660088	1
Brain_MD_35	0	-0.557487	1*
Brain_MD_36	0	-1.30296	1*
Brain_MD_37	1	1.46924	1
Brain_MD_38	1	1.02544	1
Brain_MD_39	1	0.547241	1
Brain_MD_40	1	0.391706	1
Brain_MD_41	1	0.0201054	1
Brain_MD_42	1	1.87217	1
Brain_MD_43	1	0.440148	1
Brain_MD_44	1	1.07468	1
Brain_MD_45	1	0.70975	1
Brain_MD_46	1	0.92651	1
Brain_MD_47	1	1.06011	1
Brain_MD_48	1	0.443325	1
Brain_MD_49	1	1.0668	1
Brain_MD_50	1	0.610242	1
Brain_MD_51	1	1.65304	1
Brain_MD_52	1	1.09599	1
Brain_MD_53	0	-0.166939	1*
Brain_MD_54	1	0.481764	1
Brain_MD_55	1	0.522411	1
Brain_MD_56	1	1.1148	1
Brain_MD_57	1	0.445794	1
Brain_MD_58	1	0.00549191	1
Brain_MD_59	1	0.5419	1
Brain_MD_60	1	0.820917	1



SPLASH treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for the SPLASH algorithm.

Medulloblastoma treatment outcome prediction

SPLASH algorithm

Dataset C

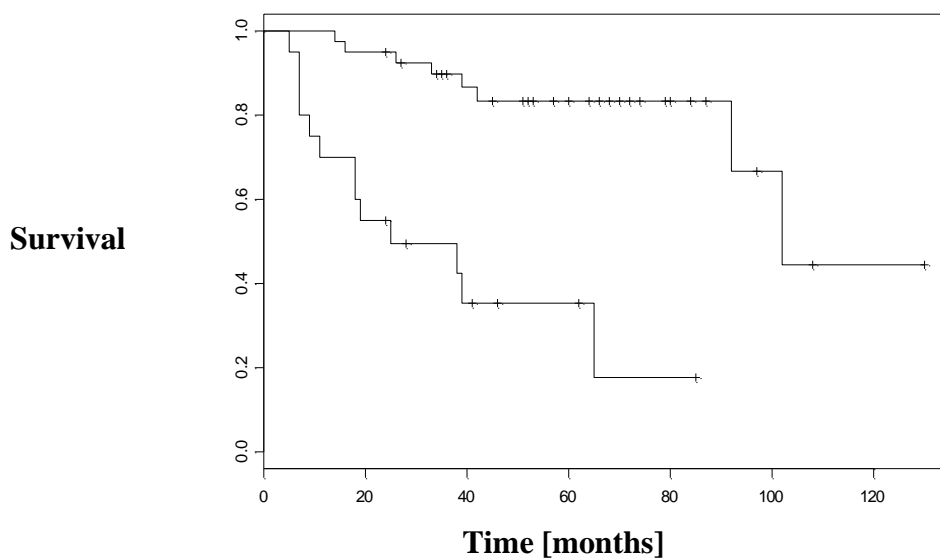
Confusion Matrix

Actual	Predicted	Failures		
		Survivors	Failures	
Survivors		32	7	39
Failures		8	13	21
		40	20	60

Datapoint	Predicted Class	Confidence	True Class	Error?
Brain_MD_1	0	0.778		0
Brain_MD_2	0	1.382		0
Brain_MD_3	0	0.451		0
Brain_MD_4	0	1.017		0
Brain_MD_5	0	1.516		0
Brain_MD_6	0	1.854		0
Brain_MD_7	1	0.114		0*
Brain_MD_8	1	-0.731		0*
Brain_MD_9	0	1.41		0
Brain_MD_10	0	1.582		0
Brain_MD_11	0	0.385		0
Brain_MD_12	0	1.293		0
Brain_MD_13	1	-0.363		0*
Brain_MD_14	1	-0.307		0*
Brain_MD_15	0	1.116		0
Brain_MD_16	1	-0.101		0*
Brain_MD_17	0	2.622		0
Brain_MD_18	1	-0.057		0*
Brain_MD_19	0	0.795		0
Brain_MD_20	1	-0.859		0*
Brain_MD_21	1	-1.609		0*
Brain_MD_22	0	0.428		1*
Brain_MD_23	0	0.599		1*
Brain_MD_24	1	-1.198		1
Brain_MD_25	0	2.303		1*
Brain_MD_26	1	-0.061		1
Brain_MD_27	1	-1.265		1
Brain_MD_28	1	-0.843		1

Brain_MD_29	1	-1.033	1
Brain_MD_30	1	-0.137	1
Brain_MD_31	1	-0.967	1
Brain_MD_32	1	-1.013	1
Brain_MD_33	0	2.019	1*
Brain_MD_34	1	-1.526	1
Brain_MD_35	1	-0.516	1
Brain_MD_36	0	3.17	1*
Brain_MD_37	1	0.238	1
Brain_MD_38	1	0.068	1
Brain_MD_39	1	0.211	1
Brain_MD_40	1	-0.714	1
Brain_MD_41	1	-1.248	1
Brain_MD_42	0	0.725	1*
Brain_MD_43	1	-0.419	1
Brain_MD_44	1	-1.157	1
Brain_MD_45	1	-0.333	1
Brain_MD_46	1	0.142	1
Brain_MD_47	1	-1.403	1
Brain_MD_48	1	-1.849	1
Brain_MD_49	1	-0.628	1
Brain_MD_50	1	0.232	1
Brain_MD_51	1	-0.622	1
Brain_MD_52	1	-0.473	1
Brain_MD_53	1	0.284	1
Brain_MD_54	0	0.44	1*
Brain_MD_55	1	-0.719	1
Brain_MD_56	1	-1.287	1
Brain_MD_57	1	-0.631	1
Brain_MD_58	1	0.263	1
Brain_MD_59	1	-0.97	1
Brain_MD_60	1	-0.828	1

Kaplan Meier Plot



TrkC treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for the single-gene TrkC predictor.

TrkC single-gene predictor

Values thresholded to 100 from below and 16000 from above

Variation filter: max/min > 5 (5-fold), max-min= 500 absolute units

Number of features (genes) = 1 = TrkC

Dataset C

Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	23	16	39
Failures	4	17	21
	27	33	60

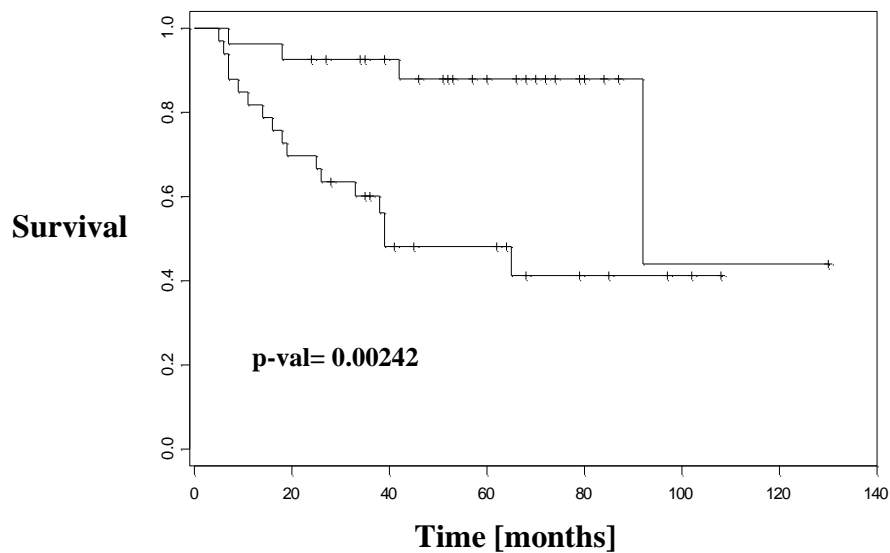
Num Data	Num Right	Num Wrong	Threshold	Num Abstain	Abs Error	ROC Error
60	40	20	0	0	0.333333	0.300366
60	5					

Datapoint	Predicted Class	Confidence	True Class	Error?
-----------	-----------------	------------	------------	--------

Brain_MD_1	0	1	0	
Brain_MD_2	0	1	0	
Brain_MD_3	0	1	0	
Brain_MD_4	0	1	0	
Brain_MD_5	1	1	0	*
Brain_MD_6	0	1	0	
Brain_MD_7	0	1	0	
Brain_MD_8	0	1	0	
Brain_MD_9	0	1	0	
Brain_MD_10	0	1	0	
Brain_MD_11	1	1	0	*
Brain_MD_12	0	1	0	
Brain_MD_13	0	1	0	
Brain_MD_14	0	1	0	
Brain_MD_15	0	1	0	
Brain_MD_16	0	1	0	
Brain_MD_17	0	1	0	
Brain_MD_18	0	1	0	
Brain_MD_19	1	1	0	*
Brain_MD_20	0	1	0	
Brain_MD_21	1	1	0	*

Brain_MD_22	0	1	1 *
Brain_MD_23	1	1	1
Brain_MD_24	0	1	1 *
Brain_MD_25	1	1	1
Brain_MD_26	0	1	1 *
Brain_MD_27	0	1	1 *
Brain_MD_28	1	1	1
Brain_MD_29	1	1	1
Brain_MD_30	0	1	1 *
Brain_MD_31	0	1	1 *
Brain_MD_32	1	1	1
Brain_MD_33	0	1	1 *
Brain_MD_34	0	1	1 *
Brain_MD_35	1	1	1
Brain_MD_36	0	1	1 *
Brain_MD_37	1	1	1
Brain_MD_38	1	1	1
Brain_MD_39	1	1	1
Brain_MD_40	1	1	1
Brain_MD_41	1	1	1
Brain_MD_42	1	1	1
Brain_MD_43	0	1	1 *
Brain_MD_44	0	1	1 *
Brain_MD_45	1	1	1
Brain_MD_46	1	1	1
Brain_MD_47	0	1	1 *
Brain_MD_48	1	1	1
Brain_MD_49	1	1	1
Brain_MD_50	1	1	1
Brain_MD_51	0	1	1 *
Brain_MD_52	1	1	1
Brain_MD_53	1	1	1
Brain_MD_54	1	1	1
Brain_MD_55	0	1	1 *
Brain_MD_56	1	1	1
Brain_MD_57	1	1	1
Brain_MD_58	0	1	1 *
Brain_MD_59	0	1	1 *
Brain_MD_60	1	1	1

Kaplan Meier Plot



Staging treatment outcome prediction results

This section contains the detailed sample predictions, error rates and survival analysis results for staging as a predictor.

Staging predictor

M0 = no metastases, Mx = the following, M1 is positive CSF cytology, M2 is local metastases, M3 is metastases throughout the central nervous system, M4 is metastases beyond the central nervous system

Dataset C

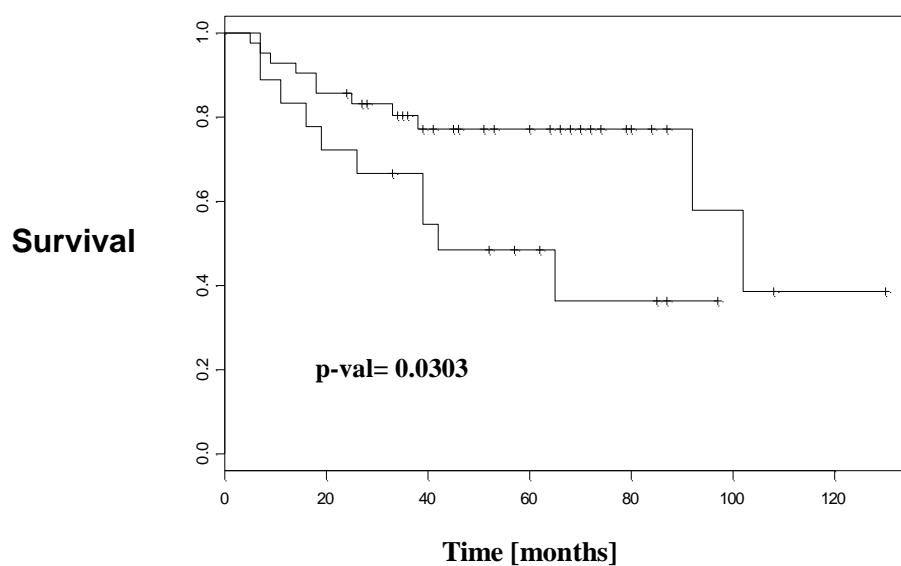
Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	31	8	39
Failures	11	10	21
	42	18	60

Datapoint	Predicted Class	Confidence	True Class	Error?	Chang stage
Brain_MD_1	0		0		T4M1
Brain_MD_2	1		0 *		T2M0
Brain_MD_3	1		0 *		T3M0
Brain_MD_4	0		0		T3M3
Brain_MD_5	0		0		M3
Brain_MD_6	1		0 *		T4M0
Brain_MD_7	1		0 *		T1M0
Brain_MD_8	0		0		T3bM1
Brain_MD_9	1		0 *		M0
Brain_MD_10	1		0 *		M0
Brain_MD_11	0		0		T2M1
Brain_MD_12	1		0 *		M0
Brain_MD_13	0		0		T3M3
Brain_MD_14	1		0 *		M0
Brain_MD_15	1		0 *		T2M0
Brain_MD_16	0		0		T3M3
Brain_MD_17	0		0		T3bM3
Brain_MD_18	0		0		T2M3
Brain_MD_19	0		0		M2
Brain_MD_20	1		0 *		T3bM0
Brain_MD_21	1		0 *		T2M0
Brain_MD_22	1		1		M0
Brain_MD_23	1		1		T4M0
Brain_MD_24	1		1		T3M0

Brain_MD_25	1	1	M0
Brain_MD_26	0	1 *	T2M3
Brain_MD_27	1	1	M0
Brain_MD_28	1	1	T4M0
Brain_MD_29	1	1	T3M0
Brain_MD_30	1	1	T3M0
Brain_MD_31	1	1	M0
Brain_MD_32	1	1	T2M0
Brain_MD_33	1	1	T3bM0
Brain_MD_34	0	1 *	T3M1
Brain_MD_35	1	1	T3M0
Brain_MD_36	1	1	T3M0
Brain_MD_37	1	1	T3M0
Brain_MD_38	0	1 *	T3M1
Brain_MD_39	1	1	T3M0
Brain_MD_40	0	1 *	T4M3
Brain_MD_41	1	1	T4M0
Brain_MD_42	0	1 *	T3M3
Brain_MD_43	1	1	T3M0
Brain_MD_44	1	1	T3M0
Brain_MD_45	1	1	T4M0
Brain_MD_46	1	1	T3M0
Brain_MD_47	1	1	T4M0
Brain_MD_48	1	1	T3bM0
Brain_MD_49	1	1	T2M0
Brain_MD_50	1	1	T3bM0
Brain_MD_51	1	1	T3bM0
Brain_MD_52	1	1	T2M0
Brain_MD_53	1	1	T2M0
Brain_MD_54	0	1 *	T4M4
Brain_MD_55	0	1 *	T3bM2
Brain_MD_56	1	1	T2M0
Brain_MD_57	0	1 *	T2M3
Brain_MD_58	1	1	T1M0
Brain_MD_59	1	1	T3bM0
Brain_MD_60	1	1	T3M0

Kaplan Meier Plot



Combined treatment outcome predictors

This section describes the results for two combinations of models using a simple majority voting rule. The two combinations are Staging + k-NN + TrkC and SVM + k-NN + TrkC. These combinations achieve better performance than any single method alone.

Combined model I: staging, k-NN and TrkC

Medulloblastoma treatment outcome prediction Combined predictor: Staging, k-NN and TrkC

Dataset C

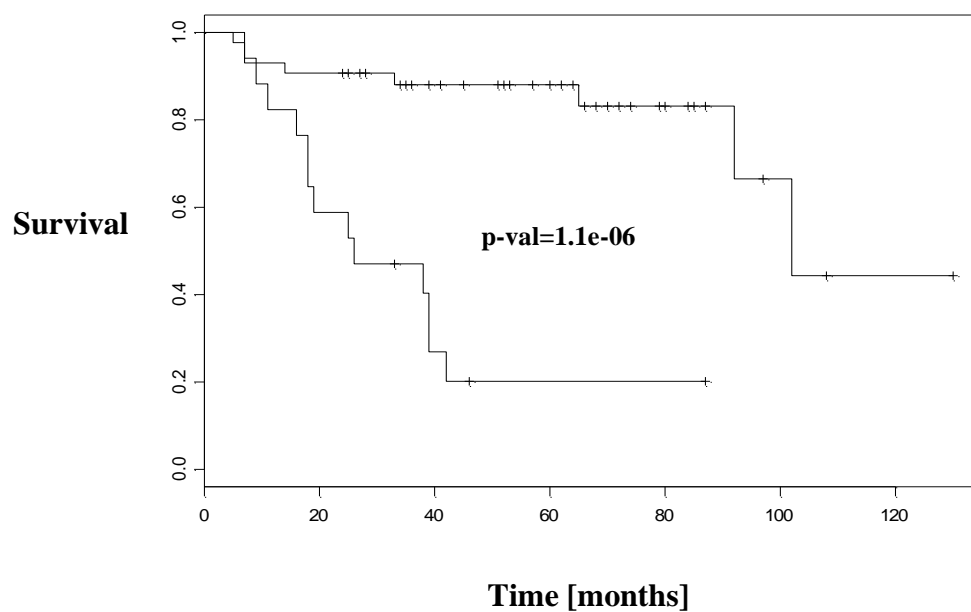
Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	35	4	39
Failures	8	13	21
	43	17	60

Datapoint	Staging	k-NN pred.	TrkC	True Class	Combined majority predictor	error?
Brain_MD_1	0	0	0	0	0	0
Brain_MD_2	1	1	0	0	0	1*
Brain_MD_3	1	1	0	0	0	1*
Brain_MD_4	0	0	0	0	0	0
Brain_MD_5	0	1	1	0	0	1*
Brain_MD_6	1	0	0	0	0	0
Brain_MD_7	1	1	0	0	0	1*
Brain_MD_8	0	1	0	0	0	0
Brain_MD_9	1	0	0	0	0	0
Brain_MD_10	1	0	0	0	0	0
Brain_MD_11	0	0	1	0	0	0
Brain_MD_12	1	0	0	0	0	0
Brain_MD_13	0	1	0	0	0	0
Brain_MD_14	1	1	0	0	0	1*
Brain_MD_15	1	0	0	0	0	0
Brain_MD_16	0	0	0	0	0	0
Brain_MD_17	0	0	0	0	0	0
Brain_MD_18	0	1	0	0	0	0
Brain_MD_19	0	1	1	0	0	1*
Brain_MD_20	1	1	0	0	0	1*
Brain_MD_21	1	1	1	0	0	1*
Brain_MD_22	1	1	0	1	1	1

Brain_MD_23	1	1	1	1	1
Brain_MD_24	1	1	0	1	1
Brain_MD_25	1	1	1	1	1
Brain_MD_26	0	1	0	1	0*
Brain_MD_27	1	1	0	1	1
Brain_MD_28	1	1	1	1	1
Brain_MD_29	1	1	1	1	1
Brain_MD_30	1	1	0	1	1
Brain_MD_31	1	1	0	1	1
Brain_MD_32	1	1	1	1	1
Brain_MD_33	1	1	0	1	1
Brain_MD_34	0	1	0	1	0*
Brain_MD_35	1	1	1	1	1
Brain_MD_36	1	0	0	1	0*
Brain_MD_37	1	1	1	1	1
Brain_MD_38	0	1	1	1	1
Brain_MD_39	1	1	1	1	1
Brain_MD_40	0	1	1	1	1
Brain_MD_41	1	0	1	1	1
Brain_MD_42	0	1	1	1	1
Brain_MD_43	1	1	0	1	1
Brain_MD_44	1	1	0	1	1
Brain_MD_45	1	1	1	1	1
Brain_MD_46	1	1	1	1	1
Brain_MD_47	1	1	0	1	1
Brain_MD_48	1	1	1	1	1
Brain_MD_49	1	1	1	1	1
Brain_MD_50	1	1	1	1	1
Brain_MD_51	1	1	0	1	1
Brain_MD_52	1	1	1	1	1
Brain_MD_53	1	1	1	1	1
Brain_MD_54	0	1	1	1	1
Brain_MD_55	0	1	0	1	0*
Brain_MD_56	1	1	1	1	1
Brain_MD_57	0	1	1	1	1
Brain_MD_58	1	1	0	1	1
Brain_MD_59	1	1	0	1	1
Brain_MD_60	1	1	1	1	1

Kaplan Meier Plot



Combined model II: SVM, k-NN and TrkC

Medulloblastoma treatment outcome prediction

Combined predictor: SVM, k-NN and TrkC

Dataset C

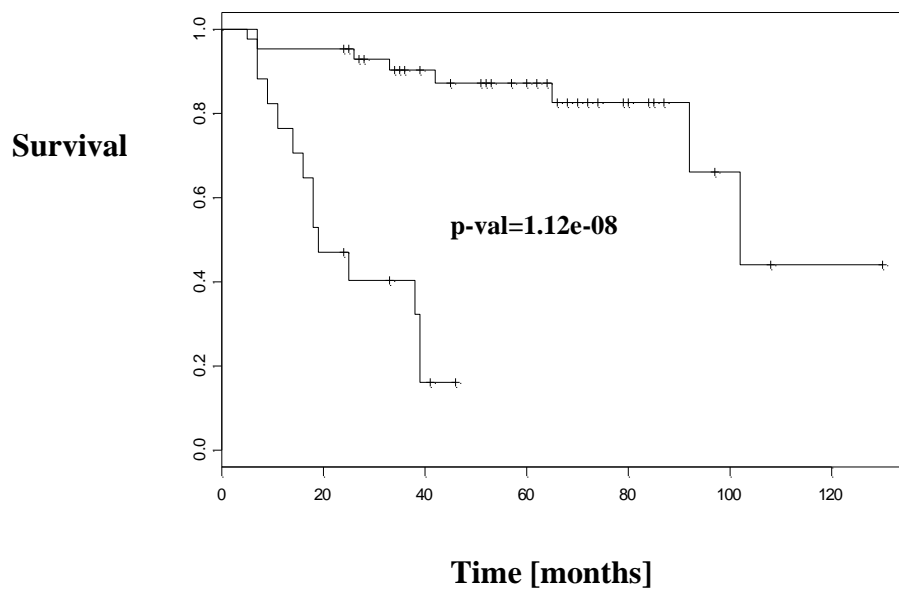
Confusion Matrix

Actual	Predicted		
	Survivors	Failures	
Survivors	35	4	39
Failures	8	13	21
	43	17	60

Datapoint	SVM	k-NN pred.	TrkC	True Class	Combined majority predictor	error?
Brain_MD_1	0	0	0	0	0	0
Brain_MD_2	1	1	0	0	0	1*
Brain_MD_3	0	1	0	0	0	0
Brain_MD_4	1	0	0	0	0	0
Brain_MD_5	1	1	1	0	0	1*
Brain_MD_6	0	0	0	0	0	0
Brain_MD_7	0	1	0	0	0	0
Brain_MD_8	0	1	0	0	0	0
Brain_MD_9	0	0	0	0	0	0
Brain_MD_10	0	0	0	0	0	0
Brain_MD_11	0	0	1	0	0	0
Brain_MD_12	0	0	0	0	0	0
Brain_MD_13	1	1	0	0	0	1*
Brain_MD_14	1	1	0	0	0	1*
Brain_MD_15	0	0	0	0	0	0
Brain_MD_16	1	0	0	0	0	0
Brain_MD_17	0	0	0	0	0	0
Brain_MD_18	1	1	0	0	0	1*
Brain_MD_19	0	1	1	0	0	1*
Brain_MD_20	1	1	0	0	0	1*
Brain_MD_21	1	1	1	0	0	1*
Brain_MD_22	0	1	0	1	1	0*
Brain_MD_23	1	1	1	1	1	1
Brain_MD_24	1	1	0	1	1	1
Brain_MD_25	1	1	1	1	1	1
Brain_MD_26	0	1	0	1	1	0*
Brain_MD_27	1	1	0	1	1	1
Brain_MD_28	1	1	1	1	1	1
Brain_MD_29	1	1	1	1	1	1
Brain_MD_30	1	1	0	1	1	1

Brain_MD_31	1	1	0	1	1
Brain_MD_32	1	1	1	1	1
Brain_MD_33	0	1	0	1	0*
Brain_MD_34	1	1	0	1	1
Brain_MD_35	0	1	1	1	1
Brain_MD_36	0	0	0	1	0*
Brain_MD_37	1	1	1	1	1
Brain_MD_38	1	1	1	1	1
Brain_MD_39	1	1	1	1	1
Brain_MD_40	1	1	1	1	1
Brain_MD_41	1	0	1	1	1
Brain_MD_42	1	1	1	1	1
Brain_MD_43	1	1	0	1	1
Brain_MD_44	1	1	0	1	1
Brain_MD_45	1	1	1	1	1
Brain_MD_46	1	1	1	1	1
Brain_MD_47	1	1	0	1	1
Brain_MD_48	1	1	1	1	1
Brain_MD_49	1	1	1	1	1
Brain_MD_50	1	1	1	1	1
Brain_MD_51	1	1	0	1	1
Brain_MD_52	1	1	1	1	1
Brain_MD_53	0	1	1	1	1
Brain_MD_54	1	1	1	1	1
Brain_MD_55	1	1	0	1	1
Brain_MD_56	1	1	1	1	1
Brain_MD_57	1	1	1	1	1
Brain_MD_58	1	1	0	1	1
Brain_MD_59	1	1	0	1	1
Brain_MD_60	1	1	1	1	1

Kaplan Meier Plot



Summary of medulloblastoma treatment outcome predictions

The following table summarizes the results for the different prediction algorithms in dataset C.

All of the multi-gene algorithms achieve similar performance and are better classifier of treatment outcome than staging of TrkC alone.

Notice the asymmetry in the number of false negatives and positives between TrkC and the other algorithms.

Summary of treatment outcome prediction performance

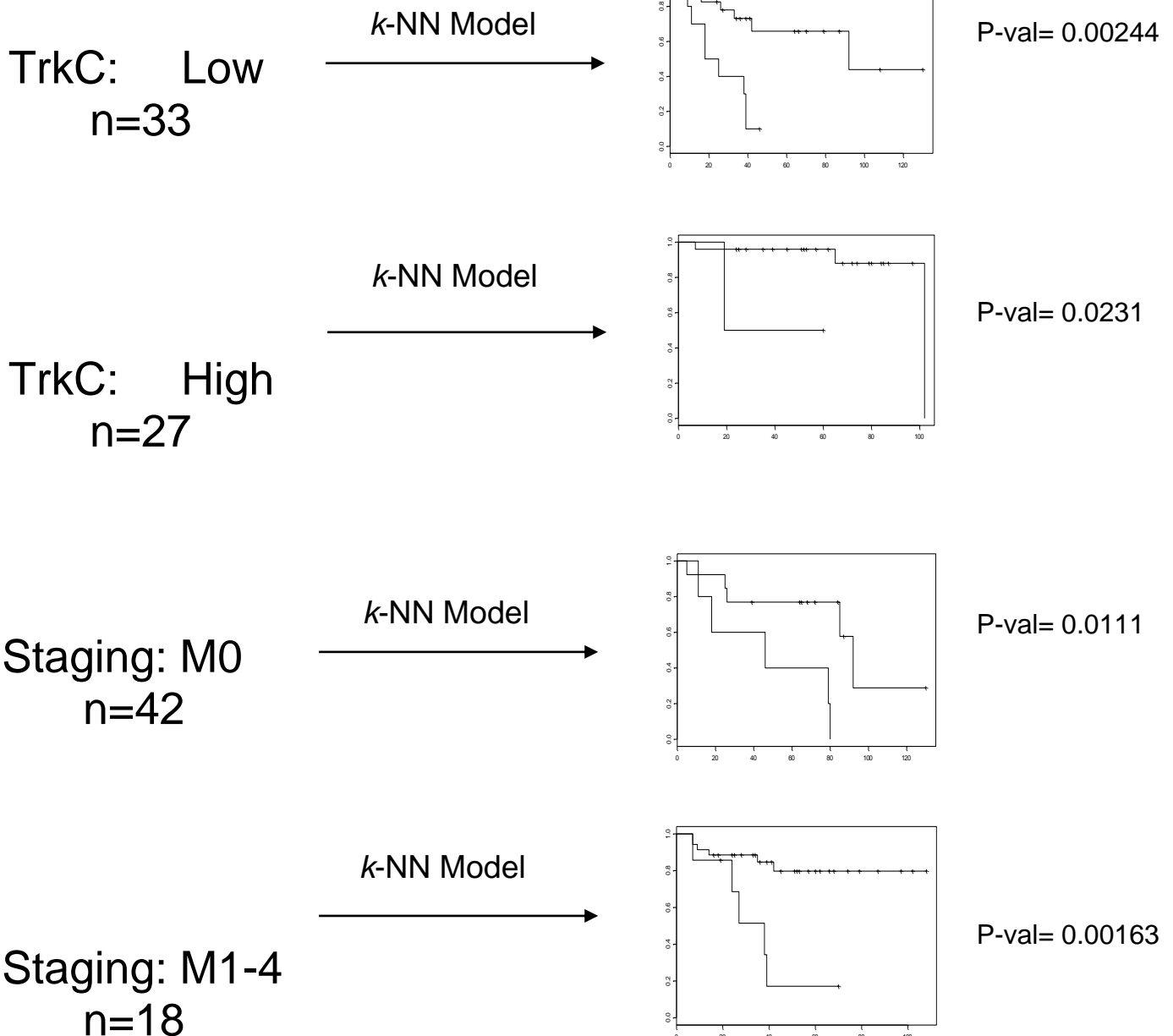
Dataset C

Algorithm	Total Correct	Total Errors	Errors in Failure Class	Errors in Survival Class	KM Rank Test P-value
Staging	41	19	11	8	0.03
TrkC	40	20	4	16	0.0024
Weighted Voting	46	14	10	4	0.00005
SVM	45	15	9	6	0.000027
<i>k</i> -nearest neighbors	47	13	11	2	3.30E-06
SPLASH	45	15	8	7	2.89E-06
Combined model I Staging, <i>k</i> -NN and TrkC	48	12	8	4	1.10E-06
Combined model II SVM, <i>k</i> -NN +and TrkC	48	12	8	4	1.12E-08

Improvements of multi-gene prediction algorithm (*k*-NN) over staging and TrkC.

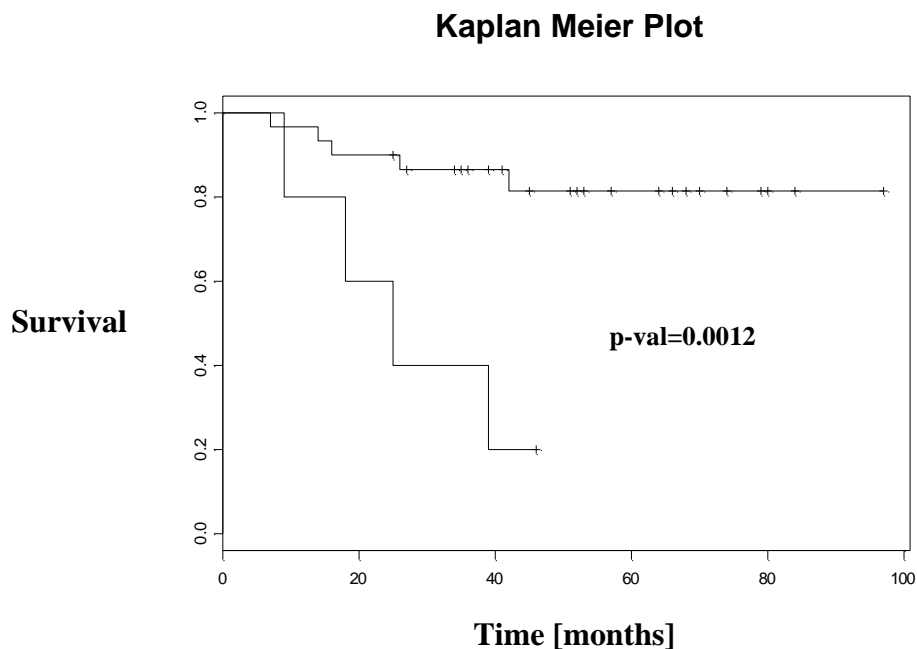
The following graph shows the results of taking low and high risks as defined by staging and TrkC expression and further classified them by using the *k*-NN algorithm results. As can be seen on the right side survival plots the multi-gene algorithm (*k*-NN) is able to resolve "survival" with additional resolution that translates in significant p-values for some of the *k*-NN subgroups. Notice in particular how the *k*-NN algorithm appears to correct the mistakes made by the staging classifier in low staging risk patients (e.g. those with small tumors but bad outcomes) and low TrkC patients (e.g. those with low TrkC expression which respond to treatment.). This confirms that there is additional information contained in the multi-gene expression profiles not necessarily contained in Staging and TrkC expression alone (the *k*-NN model did not use TrkC as one of its marker genes)

Kaplan Meier Plots



***k*-NN predictions in subgroup treated with vincristine, cisplatin and cytoxan.**

All patients received a chemotherapy regime of vincristine and cisplatin. Some patients also received cytoxan or a combination of some of the following: cytoxan, etoposide, CCNU, carboplatin, procarbazine, methotrexate and thiotepa. In order to test if our multi-gene prediction algorithm was somehow predicting outcome based on differences in these additional detail of the chemo regime we decided to analyzed the predictions in the a subset that received an identical regime of vincristine, cisplatin and cytoxan. The survival Kaplan Meier plot below shows that the model clearly resolves the failure and survival class inside this group and it is therefore not a proxy of the type of chemotherapy.



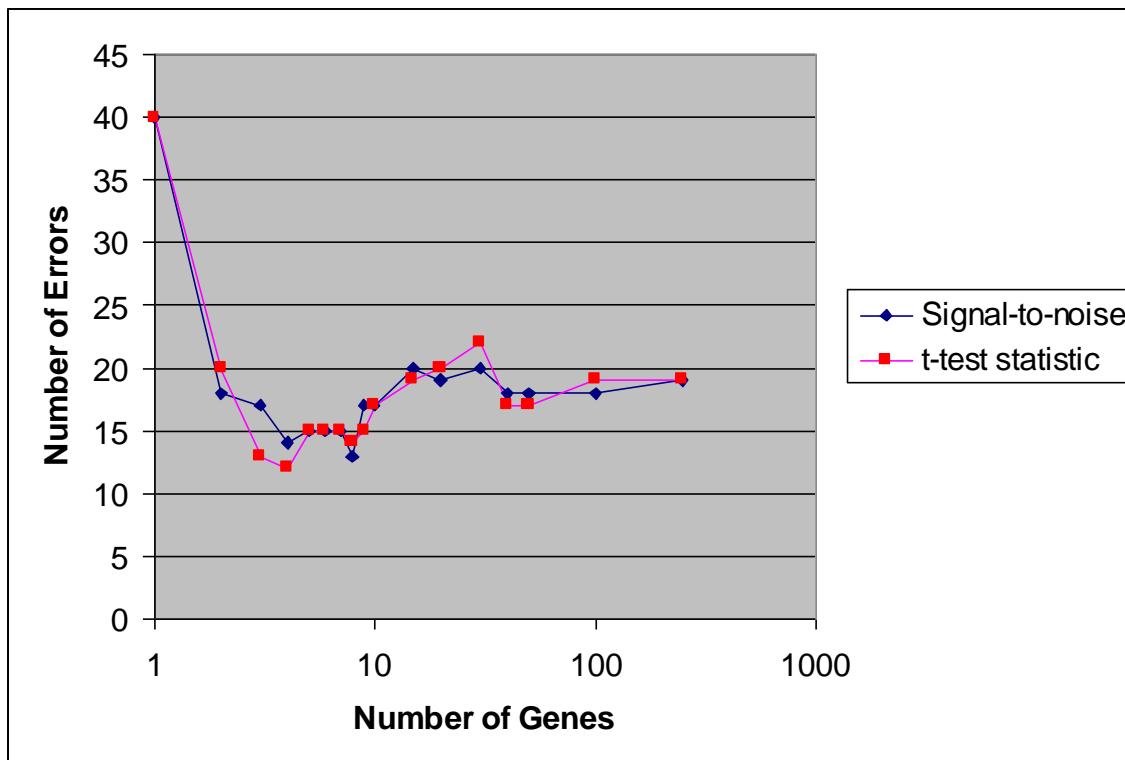
Comparison between signal-to-noise and t-test statistic metrics

In order to better understand the effect of using a more standard metric for gene selection we repeated some of the analysis of the Medulloblastoma treatment outcome dataset using a t-test statistic metric:

$$t = ((\mu_0 - \mu_1) / \sqrt{(\sigma_0^2 / N_0 + \sigma_1^2 / N_1)}),$$

where N_1 and N_2 represents the number of samples in each class.

The results obtained using this metric are very similar to the ones obtained by the signal-to-noise metric used in the paper. This is something we had observed in other cases (datasets) as well. The following plot shows the total error rate in cross-validation as a function of the number of genes for k-NN models using both metrics. These models used $k=5$ and the same filtering parameters as those used in the “k-nearest neighbors treatment outcome prediction results” section earlier in this document.



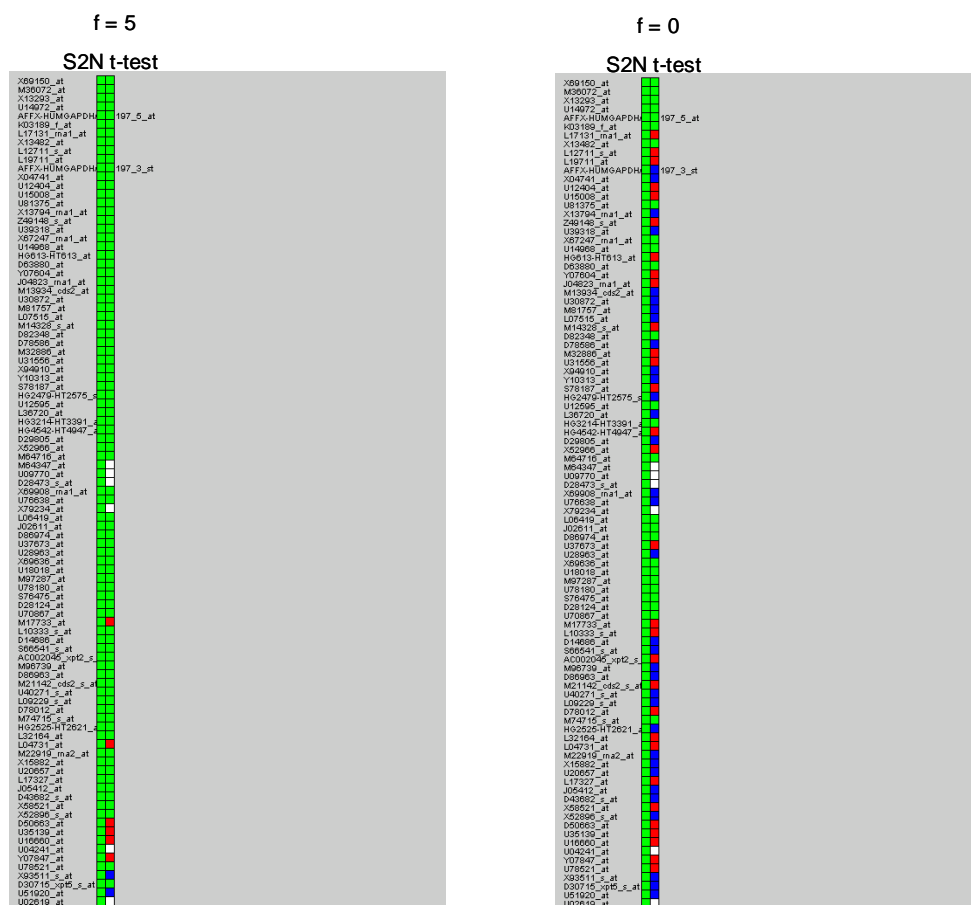
Signal-to-noise: $(\mu_0 - \mu_1) / (\sigma_0 + \sigma_1)$

T-test statistic: $(\mu_0 - \mu_1) / \sqrt{(\sigma_0^2 / N_0 + \sigma_1^2 / N_1)}$

The similarity of the k-NN model results is a consequence of the fact that the two metrics produce very similar ranking of features. The next plot compares the rankings produced by both metrics using a color scheme. The parameter f determines how many differences in rank are allowed to consider the gene to be in the “same” rank (green). We show results for $f = 5$ and 0 (exact matching). This comparison shows the close similarity in the rankings obtained using the two metrics.

Ranking comparison for Signal to noise and t-test metrics

green -> $\text{abs}(\text{difference in rank}) \leq f$
red -> rank is higher in signal to noise
blue -> rank is lower in signal to noise
white -> missing in t-test (found in signal to noise)



References

- Baldi and Long 2001.** P. Baldi and A. D. Long. A Bayesian Framework for the Analysis of Microarray Expression Data: Regularized t-Test and Statistical Inferences of Gene Changes. *Bioinformatics*, in press (2001).
- Bender and Lang 2001.** R. Bender and S. Lange. Adjusting for Multiple Testing – When and How? *Journal of Clinical Epidemiology* 54 (2001) 343-349.
- Benjamini and Hochberg 1995.** Benjamini, Y., Hochberg, Y. (1995). "Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing", *Journal of the Royal Statistical Society B*, 57 289-300.
- Benjamini et al 1999.** Y. Benjamini, L. Hothorn and P. K. Sen. Preface. *Journal of Statistical Planning and Inference* 82 (1999) 1-4.
- Benjamini et al 2001.** Yoav Benjamini, Dan Drai, Greg Elmer, Neri Kafkafi, Ilan Golani. Controlling the False Discovery Rate in Behavior Genetics Research. 2001. <http://www.math.tau.ac.il/~ybenja/ExB.html>.
- Berry and Hochberg 1999.** D .A. Berry and Y. Hochberg. Bayesian Perspectives on Multiple Comparisons. *Journal of Statistical Planning and Inference* 82 (1999) 215-227. Kohavi and John 1998.
- Brown et al 2000.** Brown M.P.S., et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. (USA)* 97, 262-267 (2000).
- Califano et al 1999.** Califano et al. Analysis of Gene Expression Microarrays for Phenotype Classification. *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, San Diego, California, August 19-23, p75-85, 1999.
<ftp.sdsc.edu/pub/sdsc/biology/ISMB00/154.pdf>
- Cherkassky and Mulier 1998.** V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory and Methods*. John Wiley and Sons. Inc. 1998.
- Churchill and Doerge 1994.** G. A. Churchill and R. W. Doerge. Empirical Threshold Values for Quantitative Trait Mapping. *Genetics* 138: 963-971 (1994).
- Dasarathy 1991.** Dasarathy V.B. (ed), *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE computer society press, Los Alamitos, Calif., December 1991. ISBN: 0818689307.
- Doerge and Churchill 1996.** R. W. Doerge and G. A. Churchill. Permutation Tests for Multiple Loci Affecting a Quantitative Character. *Genetics* 142: 285-294 (1996).
- Duda, Hart and Stork 2001.** R. O. Duda, P. E. Hart and D. G. Stork. *Pattern Classification*, 2ed. John Wiley and Sons. 2001.
- Fisher 1935.** R. Fisher. *The Design of Experiments*. 3ed. Oliver and Boyd Ltd. London.
- Fukunaga 1990.** *Introduction to Statistical Pattern Recognition*, 2ed. Academic Press 1990.
- Golub et al 1999.** Golub T.R., et al. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537 (1999).
- Good 1994.** P. Good. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypothesis*. Springer –Verlag New York 1994.
- Hochberg and Tamhane 1997.** Yosef Hochberg, Ajit C. Tamhane. *Multiple Comparison Procedures*. John Wiley and Sons. 1997.
- Huberty 1994.** C. J. Huberty, "Applied Discriminant Analysis," John Wiley and Sons Inc. (1994).
- Ideker et al 1999.** Trey Ideker; Vesteinn Thorsson; Andrew F. Siegel; Leroy E. Hood. Testing for Differentially-Expressed Genes by Maximum-Likelihood Analysis of Microarray Data. *Journal of Computational Biology* Vol. 7 Num. 6 (2000) 805-817.

Kearns and Vazirani 1997. M. J. Kearns and U. V. Vazirani, "An Introduction to Computational Learning Theory", MIT Press. 1997.

Kim et al 1999. Kim JYH, et al. Activation of neurotrophin-3 receptor TrkC induces apoptosis in medulloblastomas. *Cancer Res.* 59, 711-719 (1999).

Kohavi and John 1998. Kohavi, R. & John, G.H., (1998) The Wrapper Approach, in Feature Selection for knowledge Discovery and Data Mining, H. Liu & H. Motoda (eds.), Kluwer Academic Publishers, pp33-50.

Lee et al 2000. M. T. Lee, F. Kuo, G. A. Whitmore and J. Sklar. Importance of Replication in Microarray Gene Expression Studies: Statistical Methods and Evidence from Repetitive cDNA hybridizations. *PNAS* 2000 97: 9834-9839

Lehman 1986. E. C. Lehman. Testing Statistical Hypothesis. 2ed. John Wiley and Sons. New York. 1986.

Mukherjee et al 1999. Mukherjee S. et al. Support vector machine classification of microarray data. CBCL Paper #182/AI Memo #1676, Massachusetts Institute of Technology, Cambridge, MA, December 1999. <http://www.ai.mit.edu/projects/cbcl/publications/ps/cancer.ps>

Newton et al 2001. M.A. Newton, C.M. Kendzierski, C.S. Richmond, F.R. Blattner, K.W. Tsui. On differential variability of expression ratios: Improving statistical inference about gene expression changes from microarray data. *Journal of Computational Biology* Vol. 8 Num. 1 (2001) 37-52.

Segal et al 1994. Segal, R.A., Goumnerova, L.C., Kwon, Y.K., Stiles, C.D., Pomeroy, S.L. Expression of the neurotrophin receptor TrkC is linked to a favorable outcome in medulloblastoma. *Proc. Natl. Acad. Sci. (USA)* 91, 12867-12871 (1994).

Slonim et al 1999. Slonim, D. K. et al. Class prediction and discovery using gene expression data. *Procs. of the Fourth Annual International Conference on Computational Molecular Biology* Tokyo, Japan April 8 - 11, p263-272, 2000. http://www.genome.wi.mit.edu/MPR/publications/cancer_class_preprint_version.rtf

Somerville 1999. P. N. Somerville. Critical Values for Multiple Testing and Comparisons: One Step and Step Down Procedures. *Journal of Statistical Planning and Inference* 82 (1999) 129-138.

Tamayo et al 1999. Tamayo P, et al. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. (USA)* 96, 2907-2912 (1999).

Tamhane and Dunnett 1999. A. C. Tamhane and C. W. Dunnett. Stepwise Multiple Test Procedures with Biometric Applications. *Journal of Statistical Planning and Inference* 82 (1999) 55-68.

Troendle 2000. J.F. Troendle. Stepwise Normal Theory Multiple Test Procedures Controlling the False Discovery Rate. *Journal of Statistical Planning and Inference* 84 (2000) 139-158.

Tusher et al 2001. V. G. Tusher, R. Tibshirani and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. *PNAS* 2001 98: 5116-5121.

Westfall and Young 1993. P. H. Westfall and S. S. Young. Resampling-Based Multiple Testing. John Wiley and Sons. Inc. 1993.

Westfall and Wolfinger 1999. P. Westfall, R. Tobias, R. D., and R. Wolfinger, Multiple Comparisons and Multiple Tests using the SAS System, SAS Institute Inc, Cary, NC, 1999.

Yekutieli and Benjamini 1999. D. Yekutieli and Y. Benjamini. Resampling-Based False Discovery Rate Controlling Multiple Test Procedures for Correlated Test Statistics. *Journal of Statistical Planning and Inference* 82 (1999) 171-196.

