COMPUTATIONAL MOLECULAR BIOLOGY
HOMEWORK 2
SHUBHRA GUPTA
ASU ID 993755974

Follicular non-Hodgkin's lymphoma (NHL) is an indolent B cell malignancy with an annual incidence exceeding 10,000 cases in the United States.
Rituximab, a chimeric IgG1 monoclonal antibody directed at the B cell antigen.

Column and row represent:
Each row represents a particular cDNA and each column is an individual follicular lymphoma (FL) or normal lymphoid tissue sample.
CLID: Clone id
FL1_CR_LC: follicular lymphoma (patient number)_complete response_lymphochip microarrays
FL1_NR_LC: follicular lymphoma (patient number)_no response_lymphochip microarrays
FL1_PR_LC: follicular lymphoma (patient number)_partial response_lymphochip microarrays
GWEIGHT (gene weight) and **EWEIGHT** (experiment weight)
Name: clone identifier
IMAGE and **LCP:** clone name or gene

Three different mining tasks using microarray raw data set:
1. **Preprocessing:** preprocessing to improve power of multiple testing.
   Filtering, transformations, normalization of data and etc can do preprocessing.
2. **Association Rule Mining:**
   Association Rule Mining is an advanced data mining technique that is useful in deriving meaningful rules from microarray expression data.
3. **Visualization:**
   Using visualization tools one can find the structure and patterns or knowledge discovery.

Data set corresponds to article:
"Variation in gene expression patterns in follicular lymphoma and the response to rituximab".
Sean P. Bohen, Olga G. Troyanskaya, Orly Alter, Roger Warnke, David Botstein, Patrick O. Brown, and Ronald Levy
Published February 18, 2003
Proc. Natl. Acad. Sci. USA, **vol.100, no. 4**

How data set was used:
Expression data from 24 independent follicular lymphoma lymph node samples on

lymphochip (LC) microarrays was used to generate the gene lists. The lists and p-values were generated using the Wilcoxon rank sum test (non-parametric alternative to the two sample t-test which is based solely on the order in which the observations from the two samples) in supervised analysis. Prior to supervised analysis, data were SVD filtered and imputed. Prior to SVD and supervised analysis, clones from the LC data were selected based on data quality; clones were required to have signal greater than 2.5 fold.

Prior to analysis, individual data points were median centered for each cDNA clone. Hierarchical cluster analysis applied to the gene axis and to the sample axis. Prior to cluster analysis, the data was filtered for data quality and variance, as follows. Data from 24 arrays representing malignant and normal tissue was selected for signal greater than 1.5 fold. Clones were identified by Stanford Unique Identifier (SUID), with averages data if multiple copies of the same clone are present on the array. Data were then filtered based on variance such that clones were included if expression was greater than two fold above or below the median on at least three arrays. Clones were required to have good data, by the previous criteria, on at least 80% of the arrays to be included in hierarchical cluster analysis. Genes and arrays were then clustered by Pearson Correlation. Hierarchical cluster analysis of LC data revealed a technical artifact that resulted is samples segregating by the date of the experiment. Further investigation revealed that this artifact was likely due to differences in the calibration of the two scanners used to analyze the microarrays. Singular value decomposition (SVD) was used to filter data for remove the pattern corresponding to this artifact prior to analysis. Data from arrays was filtered prior to supervised analysis, using the same criteria as above. Supervised analysis taking into account known outcome to rituximab treatment was performed using Wilcoxon rank sum test to generate a rank list of genes whose corresponding mRNA levels differ significantly in rituximab responders versus non-responders. For this analysis, patients were divided into two groups, rituximab responders (composed of CR and PR) and non-responders (composed of NR and MR).

Patients were included for treatment based on the availability of freshly frozen lymph node biopsy material containing enough mRNA to allow cDNA microarray analysis. Only patients with samples that had been obtained before any systemic therapy were included. In all cases the pathological diagnosis was follicular non-Hodgkin's lymphoma (follicular small cleaved (grade 1), follicular mixed (grade 2), or follicular large cell (grade 3)). Each patient received rituximab treatment with documentation of clinical outcome. In all cases, biopsy and pathology review were performed at Stanford University Medical Center.

Analysis of data:

Analysis of the patterns of gene expression in follicular lymphomas from 24 patients (no

significant differences in age or treatment) suggested that two groups of tumors might be distinguished. All patients, whose biopsies were obtained before any treatment, were treated with rituximab, a monoclonal antibody directed against the B cell antigen. Or more specifically, whether gene expression patterns in tumors might predict sensitivity to rituximab treatment. Data suggest that analysis of gene expression patterns in tumors may allow prediction of the sensitivity of tumors to particular anti tumor agents and elucidate the biology underlying resistance to a given therapy.

Conclusion:

Gene expression patterns in the tumors that subsequently failed to respond to rituximab appeared more similar to those of normal lymphoid tissues than to gene expression patterns of tumors from rituximab responders. According to authors these findings suggest the possibility that the response of follicular lymphoma to rituximab treatment may be predicted from the gene expression pattern of tumors.

5 query using Entrez:

Database name

Pubmed	Nucleotide	Protein	Genome	Books
Structure	SNP	Journals	Unigene	PDB

Identity

Genebank (some Id)	Cluster ID	Gene	Accession, Score	NLM ID	Sequence ID

Links

B Links, Related Article	Item, Sections	Article	Literature	Cited Reference	Heading	Title

Search item

Clone Id	Patent record	G-protein	Alzheimer ’s	Chromosomes	Journal of theoretical biology	Binding protein

Organism

Mus Musculus	Drosophila	Streptococcus mutans UA159

Author

Anderson, C.W. et	Searls, D.B.	Beres, SB et	Deafness

1. List the names of database come with books of Alzheimer's (disease) and Asthma, which contain at least three items of these diseases?
In United State about 4million people have Alzheimer's disease.

2. Search all article of authors "Anderson, C.W. and Beres, SB**"** which comes with some database and cited references. One can search what are those database and cited references?

3. Search all types of identity that comes with organism (Drosophila fruit fly)?

4. Identity NLM comes in some journals that is database journal with some articles. What are those journals and articles?

5. List Identities of this entire search item (G-protein, patent records, clone id and chromosomes) which include some common database name?