

One of the many challenges in diagnosing and treating cancers is that cancers which appear clinically similar can be genetically heterogeneous. The disparate gene defects can have different implications for prognosis and treatment of the cancer. You will be dealing with two different forms of acute leukemia, namely acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL). The two leukemias appear very similar morphologically. However, because the chemotherapy regimens differ for AML and ALL patients, the ability to distinguish between them is critical for successful treatment.

You will be analyzing microarray data from experiments based on 38 patients with either AML or ALL. The microarray experiments were performed by extracting RNA samples from bone marrow cells of the patients and hybridizing the RNA to a microarray chip. You can retrieve the microarray data from:

<http://cs.wellesley.edu/~cs303/assignments/ALL-AML.txt>

In order to analyze the data via clustering, you can download and install two programs, Cluster and TreeView, which are freely available via the Internet.

<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv>

<http://genetics.stanford.edu/~alok/TreeView/>

The first program, Cluster, actually clusters the data, while the program, TreeView, is used for viewing the clustering results.

- 1) If you cluster with a kMeans algorithm, and then repeat the kMeans clustering with the same parameters, do the results change? If you cluster with a hierarchical algorithm, and then repeat the hierarchical clustering with the same parameters, do the results change? Why?
- 2) Do your clustering results indicate that microarray experiments can be used to distinguish between different forms of acute leukemia? How confident would you be in diagnoses made on the basis of this small microarray data set?
- 3) Are there particular genes that whose expression patterns are good indications of the different leukemia classes, i.e., particular genes that are highly expressed in AML patients and less expressed in ALL patients or vice versa?

For further details on this research and the microarray experiments see:

Golub, Slonim, Tamayo, Huard, Gaasenbeek, Mesirov, Coller, Loh, Downing, Caligiuri, Bloomfield, and Lander, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring", *Science* **286**, page 531, 1999.