# An Example of Clustering Expression Data

6.874 / 7.90

Computational Functional Genomics

Spring 2005

Results – hand generated data
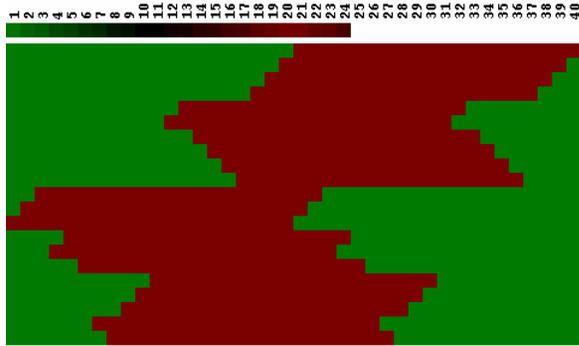

Hierarchical clustering
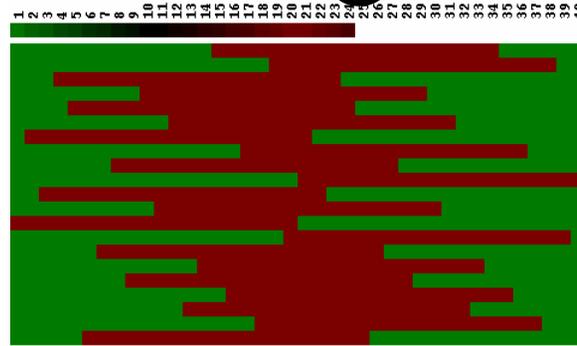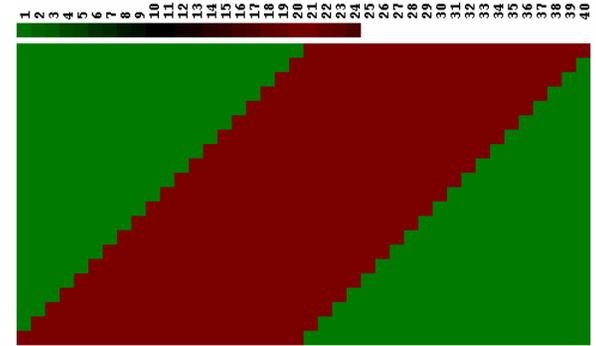
Input
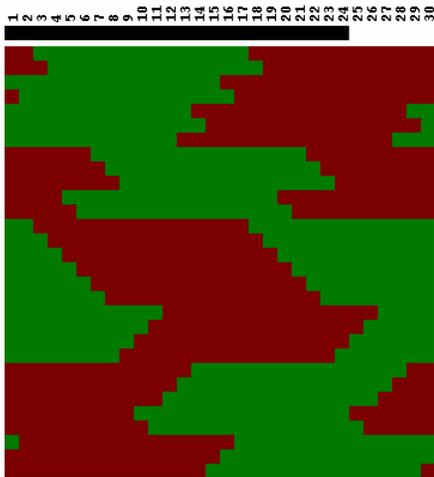
Optimal ordering

Hierarchical clustering
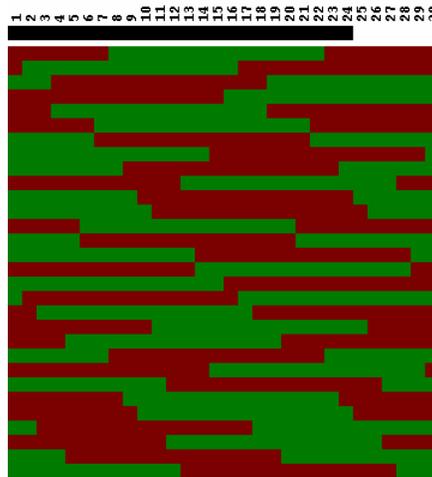
Input
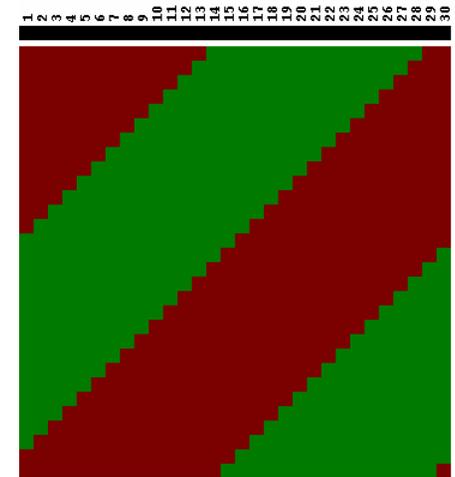
Optimal ordering

Problem definition

Denote by Φ the space of the possible linear orderings consistent with the tree.

Denote by $v_1 \ldots v_n$ the tree leaves.

Our goal is to find an ordering that maximizes the similarity of adjacent elements:

$$\max_{\phi \in \Phi} \sum_{i=1}^{n-1} S(v_i^{\phi}, v_{i+1}^{\phi})$$

where S is the similarity matrix

59 experiments, combining cdc15, cdc28 and α factor arrest
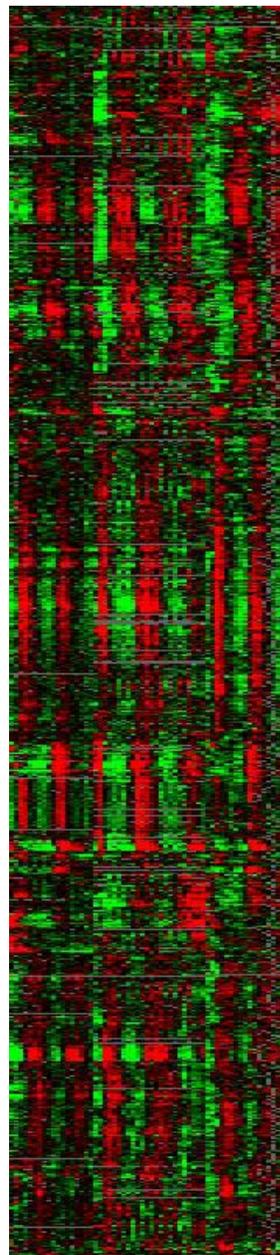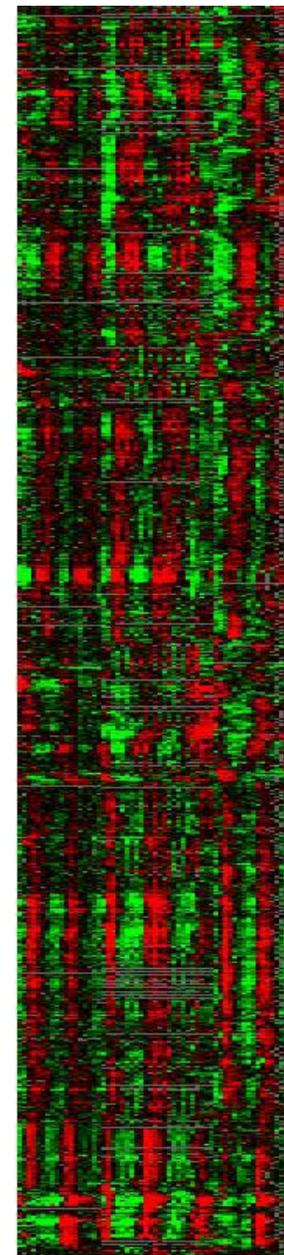
Discovering new types of cancer

Alizadeh, A., Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature 403, pp. 503-511 (February 3, 2000)

(Does not use optimal ordering)

Goal

- Discover cause for different disease courses for diffuse large B-cell lymphoma (DLBCL)
  - 40% of patients respond to therapy
  - 60% succumb to disease
- Provide diagnostic / prognostic tool
- DLBCL is most common subtype of non-Hodgkin's lymphoma

Questions

- Can we create a molecular portrait of distinct types of B-cell malignancy?

- Can we identify types of malignancy not yet recognized?

- Can we relate malignancy to normal stages in B-cell development and physiology?

Lymphochip

- 17,856 cDNA clones
  - 12,069 from germinal B-cell library
  - 2,338 from DLBCL, follicular lymphoma (FL), mantle cell lymphoma, and chronic lymphocytic leukaemia (CLL)
  - 3,186 genes important to lymphocyte and/or cancer biology
  - B- and T-lymphocyte genes that respond to mitogens or cytokines

Data sources

- Rearranged immunoglobulin genes in DLBCL are characteristic of germinal center of secondary lymphoid organs

- 96 normal and malignant lymphocyte samples

Figures removed for copyright reasons.

See Alizadeh, A., et al. "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling." Nature 403 (February 3, 2000): 503-511.

Summary

- DLBCL groups are still diverse – some members of GC B-like DLBCL group die
  - 5 in first 2 years
- May be able to find informative features for more groups
- If can find constitutive genes in cancers, target upstream regulators