Everyone,

These are some points you should think about during the discussion session in tomorrow's lecture regarding the Golub paper.

- What were the goals of this project?

- What is meant by class discovery?

Class discovery refers to defining previously unrecognized tumor subtypes

- What is meant by class prediction?  
Class prediction refers to the assignment of particular tumor samples to already-defined classes, which could reflect current states or future outcomes.

- There are two datasets (initial and independent)  
   - The initial dataset only contains bone marrow samples  
   - The independent dataset contains both bone marrow and peripheral blood samples  
   - The initial dataset samples come from two labs  
   - The independent dataset samples come from four labs  
       - What are some complications that could arise due to this?  
Changes in sample handling protocols, introduction of artifacts

- The paper specifies that there are 6817 genes while the dataset contains 7129 rows, where each row describes a probe  
    - Why are there extra probes?

For redundancy, making sure multiple probes could give the same result.

- What formula was used to rank the genes best suited for differentiating the classes?