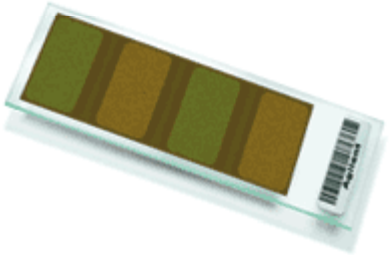
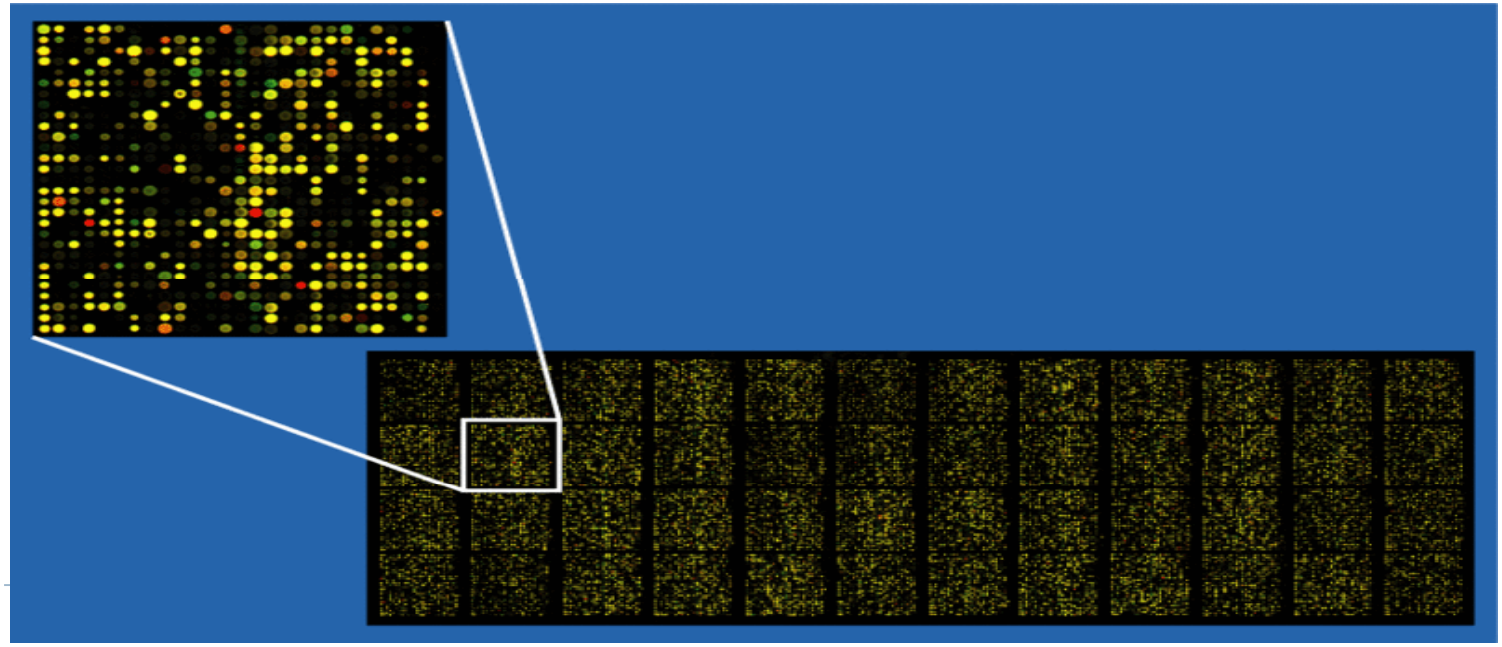
Possible Projects for CMPUT 466/551

Nasimeh Asgarian
Russ Greiner

Microarray

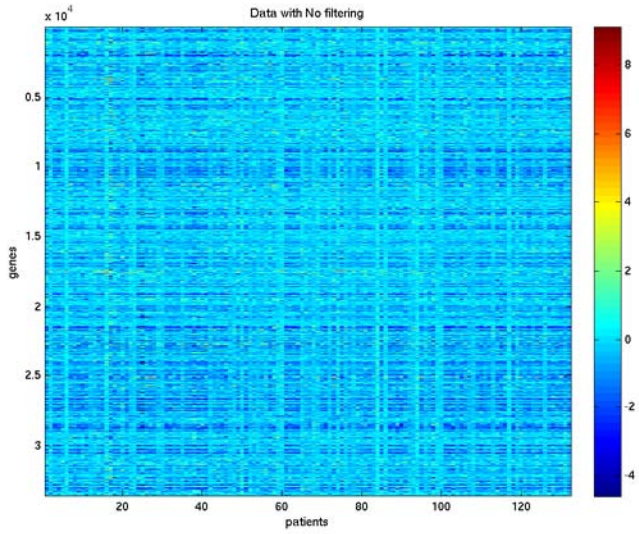


50,000 genes for
4 Patients on this slide



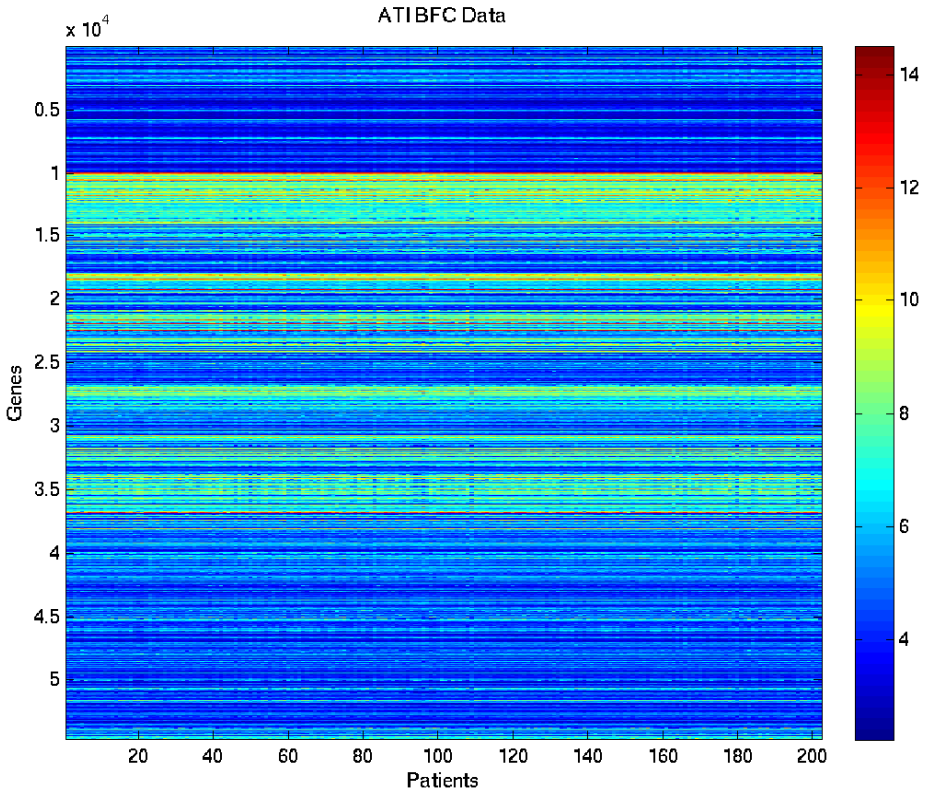
Microarray: BREAD Data

- **B**reast cancer **R**elapse: **EA**rly **D**eterminants
  - Predict Hormone Receptor Status (ER, PR, HER2)
    - Determines which treatment is best for each patients
  - Predict Relapse
    - For which patients their cancer will come back within 3 years
  - Predict Recurrence
    - For which patients their cancer will come back
- Data
  - 132 patients
  - 30 patients as validation set
  - ~33K genes (no flitering)
  - 40 Clinical features



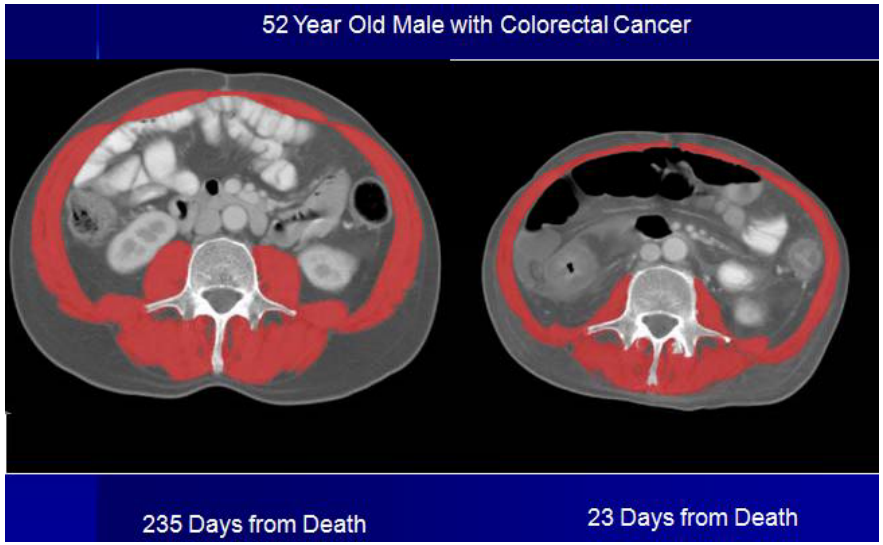
Microarray: ATI

- Kidney transplant data
  - Goal: predict who will reject the transplant
  - Predict kidney function
  - Predict lesions
- Data
  - 173 unique patients
  - 54K genes
  - 38 Gene sets (PBTs)
  - Some clinical features



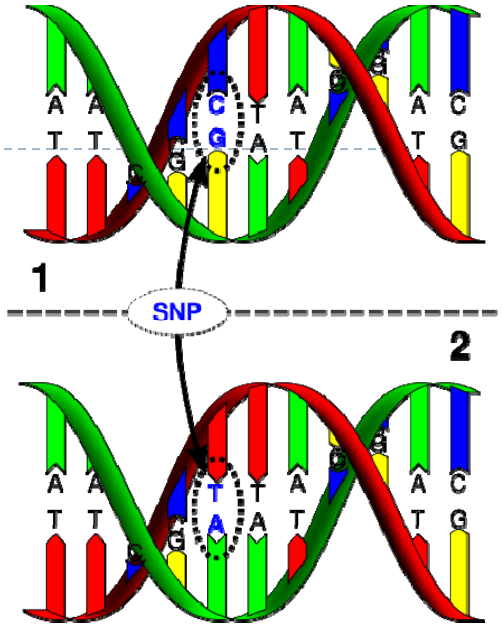
Microarray: Cachexia

- A significant and rapid loss of adipose and skele tissue
  - Difficult to detect and quantify
- Use microarray data
  - To predict which patient will become cachexic
  - Find the genes involved
- Data
  - 138 patients
  - 41K genes
  - Some clinical features
    - age,
    - gender,
    - Height,/weight,
    - type of cancer, …

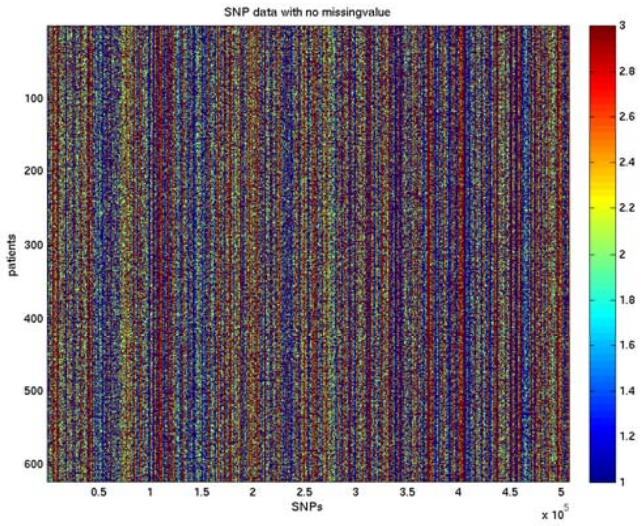


SNP

- **Single-Nucleotide Polymorphism** is a DNA sequence variation occurring when a single nucleotide -- A, T, C, or G-- in the genome(or other shared sequence) differs between members of a species.



- SNP Data for Breast Cancer
  - Predict who will get breast cancer
- Data
  - 623 patients
  - 506,836 SNPs (with no missing value)
  - 782,838 (with 99% known values)



Other

- Breast Cancer subtypes/MUC1
  - Can we predict breast cancer *recurrence* based on localization of MUC1 protein?
  - Can we predict breast Cancer *sub-types* based on other clinical parameters?
    - Luminal A
    - Luminal B (Initial, Expanded)
    - Triple negative
    - HER2 Positive
- Data
  - 1350 patients
  - ~ 80 features