Public data - for projects

Data sets:
- Smoking stop
- Leukemia
- Response to chemotherapy [breast cancer]
- Testicular carcinoma in situ
- Prenatal Trisomy
- Human Tissues

Or find one yourself !?

Try GEO at NCBI or ArrayExpress
Effects of cigarette smoking and smoking stop

Human airway epithelial samples from
34 smokers
23 never-smokers
18 former smokers

Data
RNA from 75 Human airway epithelial cell samples on Affymetrix U133A
Metadata: demographics, Smoking history, lung function
http://pulm.bumc.bu.edu/siegeDB/siege_html/downloaddatatables.html

References:
GEO GSE994
PubMed 15210990
Leukemia Data

102 bone marrow samples from children with acute lymphoblastic leukemia (ALL)

- 12 BCR-ABL
- 15 E2A-PBX1
- 17 TEL-AML
- 17 MLL
- 11 Psuedodiploid
- 20 Hyperdiploid
- 10 Normal

Additional 12 test samples (a total of 114 samples)
(BCR-ABL, E2A-PBX1, MLL, and TEL_AML)

Platform: Affymetrix HGU133A

The data set has previously been used for classification and discovery of new classes

For more information:
Ross et al., Blood, 2003
Gene Expression and Chemotherapy Response

Data set contains:

- **133 Breast cancer** samples collected by MD Anderson Cancer Centre (Affymetrix hgu133a)
- Detailed Clinical Information and Demographics: Age, Ethnicity, Histology, etc.
- Chemo therapy response (TFAC, TH/FAC etc.)
- Article: Hess et al. 2006 (PMID: 16896004)

Data

http://bioinformatics.mdanderson.org/Supplements/Datasets/PredictorValidation/MDA133-CELFiles.zip
Testicular carcinoma in situ (CIS) Data

- 21 samples designed for finding the origin of testicular cancer
  - 3 embryonic stem cells (ESC)
  - 3 fetal gonocytes (Gon1-3)
  - 3 fetal oogonia (Oo1-3)
  - 3 CIS (mixed cell types) (CIST1-3)
  - 3 CIS purified by laser microdissection and staining (CIS1-3)
  - 3 normal testis (NT1-3)
  - 3 sertoli cells (Sert1-3)

- Platform: Agilent

- Dataset has previously been used for comparison of CIS cells to other cell types in order to describe the origin of CIS

For more information:
Sonne et al., Cancer Research, 2009
Prenatal Trisomy Data

Human prenatal samples from amn and cvc
- 3x no trisomy
- 3x trisomy 13 - Patau syndrome
- 3x trisomy 18 - Edward syndrome
- 3x trisomy 21 - Down syndrome

Data
21 GeneChips (HG-U133_Plus_2)

Used for investigation of chromosomal organization of transcription and more.

Altug-Teber et al. 2007
GEO: GSE6283
Human Tissue Atlas

79 human tissues (in duplicates)
Among the tissues are:
• brain samples
• heart and liver samples
• fetal samples

22215 probe sets (HGU133A subset)
Preprocessed data:
• normalized
• expression index calculated

Used for investigation of global trends and chromosomal organization of transcription, evaluation of gene prediction

Su et al., 2004
GEO: GDS596
Practical stuff

Today
• Group formation (3-4 individuals per group)
• Each group send a name + student number list to hbjorn@cbs.dtu.dk

Week 2:
• Monday-Wednesday: Group discussions on problem formulations
• Wednesday: 15:00, deadline for problem formulation
  hand-in maximum half a page
• Thursday: presentation of project (1 slide)

Exam – week 3, friday
  present poster