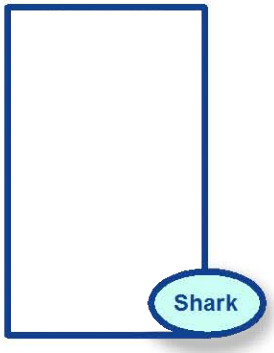
GeneGroupAnalysis 1.0.0-BetaCS

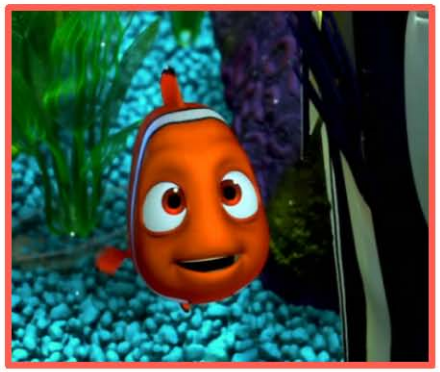
Alejandro Quiroz-Zárate

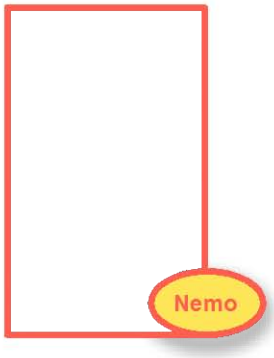
John Quackenbush lab

http://compbio.dfci.harvard.edu/compbio

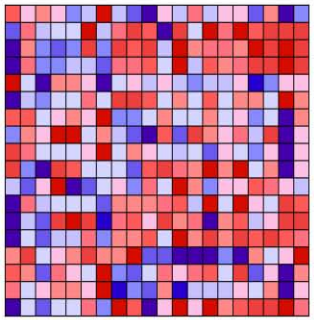
The “usual” analysis







Molecular profile

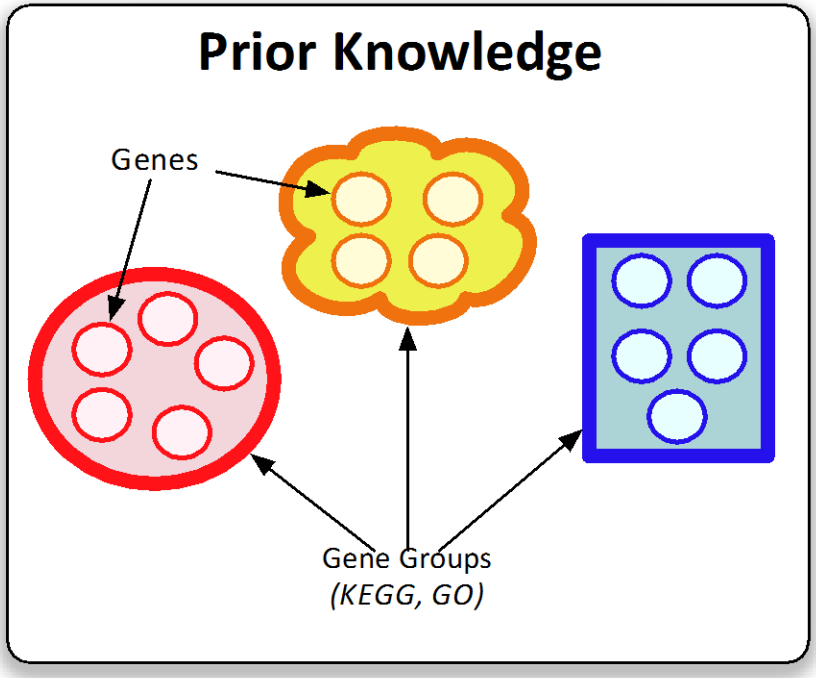


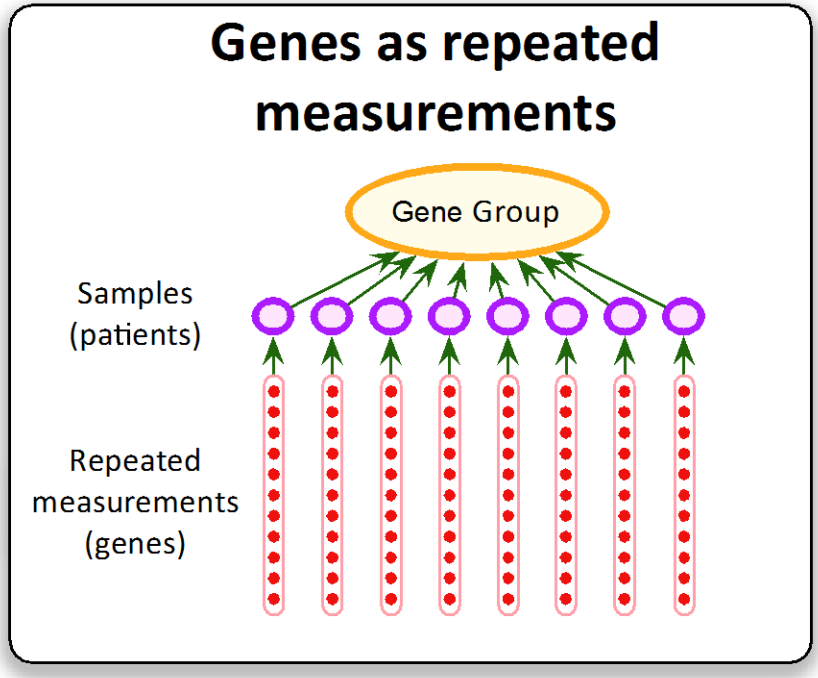
Run

GSEA
DAVID

The Key Idea

- This methodology was developed with Dr. John Quackenbush





The model

$$\left(z_{ij}^{t_1}, \ldots, z_{ij}^{t_H}\right) \sim N_H\left(\mu_i, \Sigma_i\right)$$

$$\Sigma_i \sim IW_{\nu_0}\left(\Lambda_0^{-1}\right) \text{ where,}$$

$$\mu_i = \beta_i + \alpha_i * 1(trt = B) + \gamma_i X$$

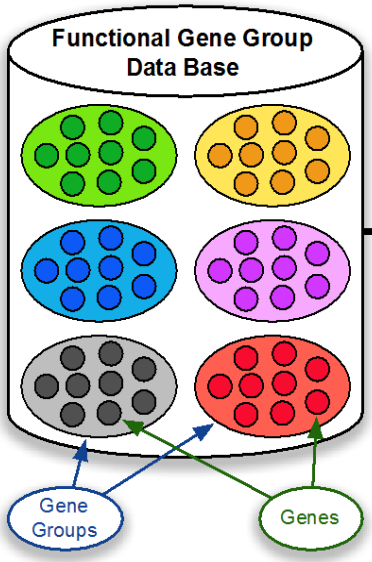
$$\beta_i = \left(\beta_i^1, \ldots, \beta_i^H\right)$$

$$\alpha_i = \left(\alpha_i^1, \ldots, \alpha_i^H\right)$$

$$\alpha_i \sim \pi_i^{\alpha} N_H\left(0, Diag\left(\sigma_i^2\right)\right) + (1 - \pi_i^{\alpha})\delta_0\left(\alpha_i\right)$$

$$\pi_i^{\alpha} \sim \rho^{\alpha} Be(am, a(1-m)) + (1 - \rho^{\alpha})\delta_0\left(\pi_i^{\alpha}\right)$$

$$\rho^{\alpha} \sim Beta(sr, s(1-r))$$



High dimensionality on the number of gene groups

Sparness priors are used to control for multiple comparison

Linear Model on the Mean of the Gene Groups

Applications

Two or more group comparison(cross-sectional design)

Time series data design

The proposed methodology

GeneGroupAnalysis 1.2.0

GeneGroupAnalysis

Gene Functional Class Analysis

Bioconductor version: Release (2.10)

R package providing functions to peform gene-set significance analysis over simple cross-sectional or time series data designs.

Author: Alejandro Quiroz-Zarate and John Quackenbush

Maintainer: Alejandro Quiroz-Zarate <aquiroz at hsph.harvard.edu>

To install this package, start R and enter:

```
source("http://bioconductor.org/biocLite.R")
biocLite("GeneGroupAnalysis")
```

To cite this package in a publication, start R and enter:

```
citation("GeneGroupAnalysis")
```

Workflows »

Common Bioconductor workflows include:

- Oligonucleotide Arrays
- High-throughput Sequencing
- Annotation
- Variants
- Flow Cytometry and other assays

Mailing Lists »

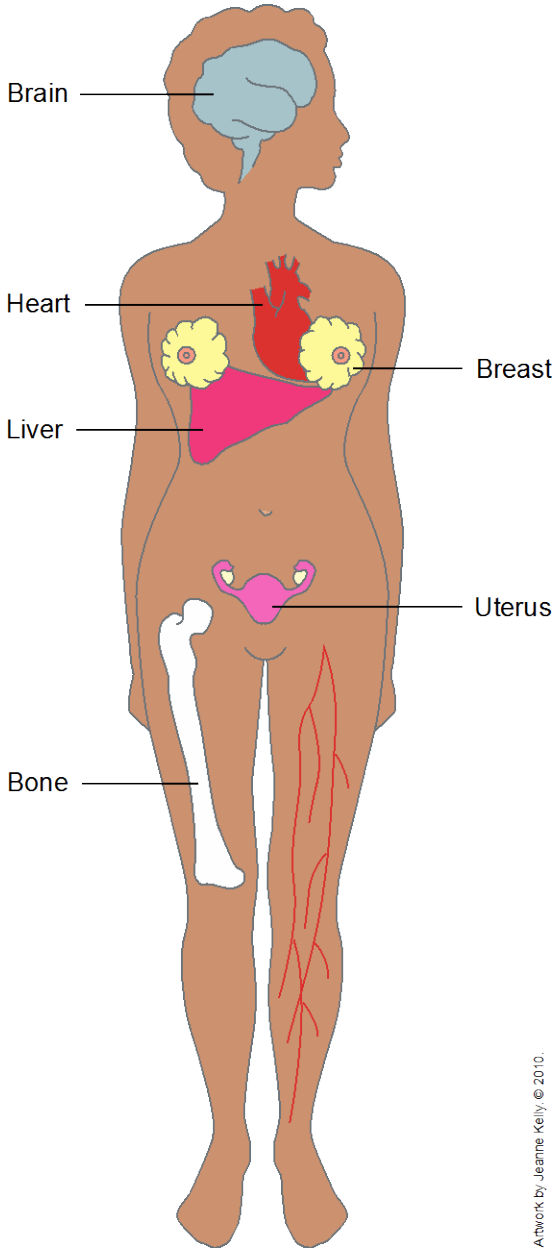
Post questions about Bioconductor packages to our mailing lists. Read the posting guide before posting!

- bioconductor
- bioc-devel

Go to 

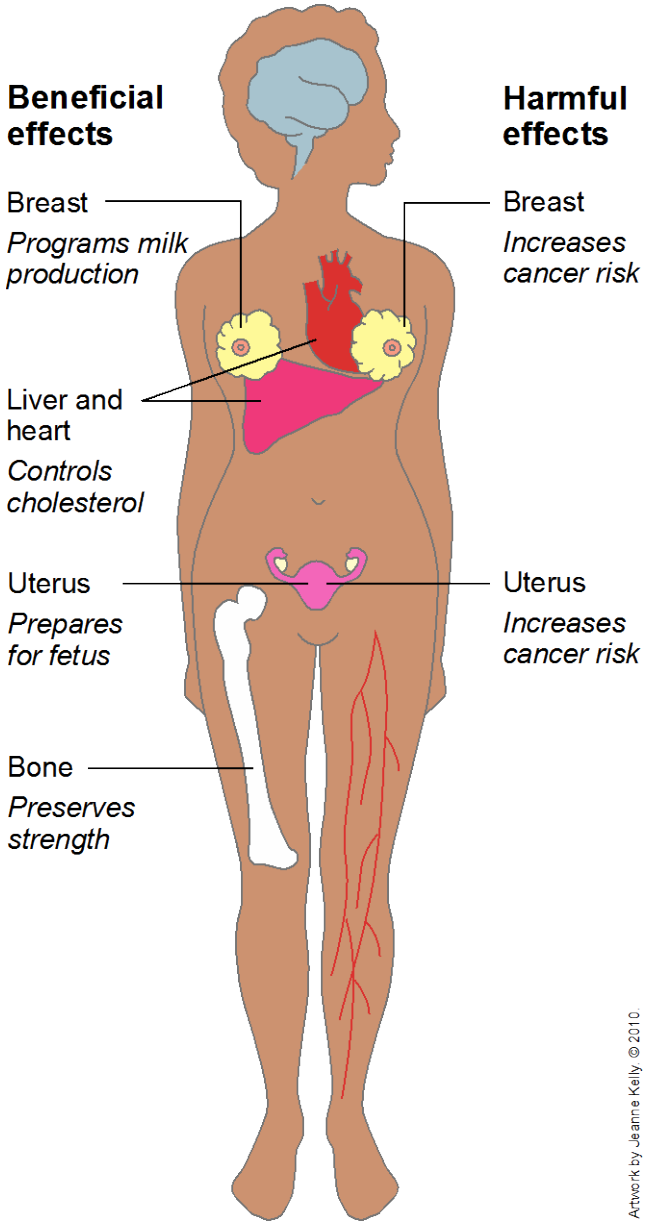
- Example (real data set)
  - Gene expression from tumor samples
    - 209 ER+ and 135 E-
    - GEO reference accession numbers:GSE2034,GSE5327
  - Affymetrix U133A
  - Conceive the data set with a Cross-Sectional Design
  - Question of interest:
    - **What are the biological mechanisms that drive the differences in Estrogen Receptor status?**

Estrogen Targets Tissues

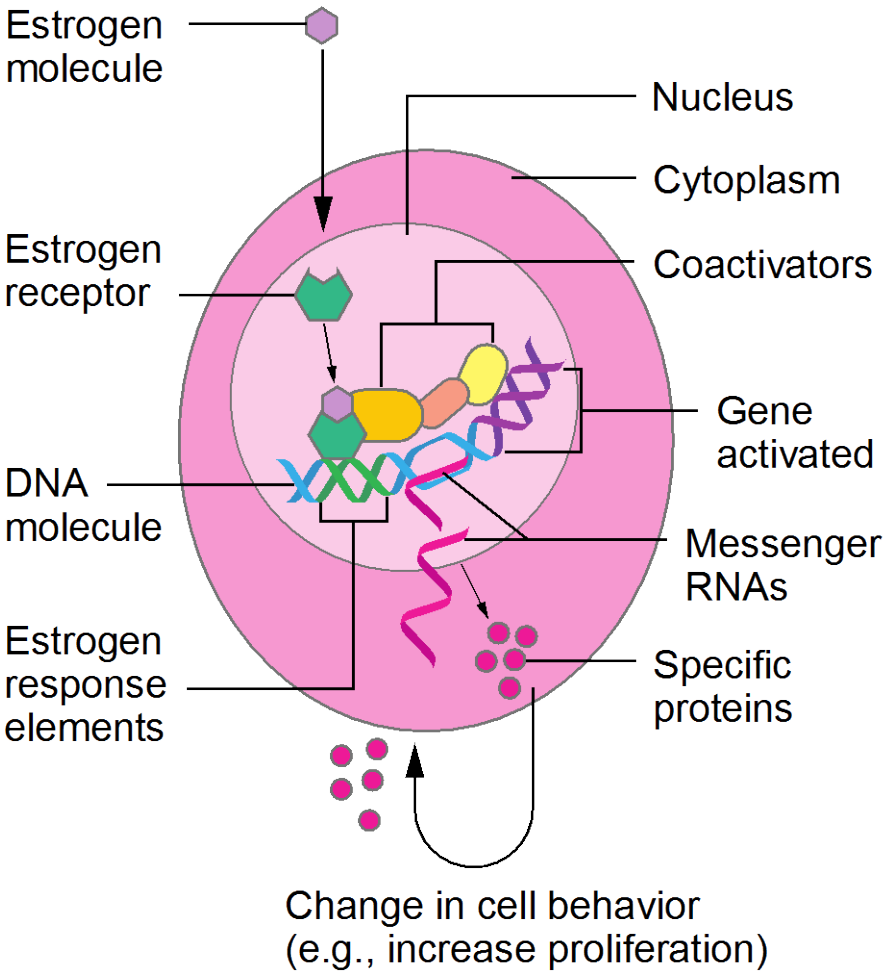


NATIONAL CANCER INSTITUTE

Estrogen and Cancer

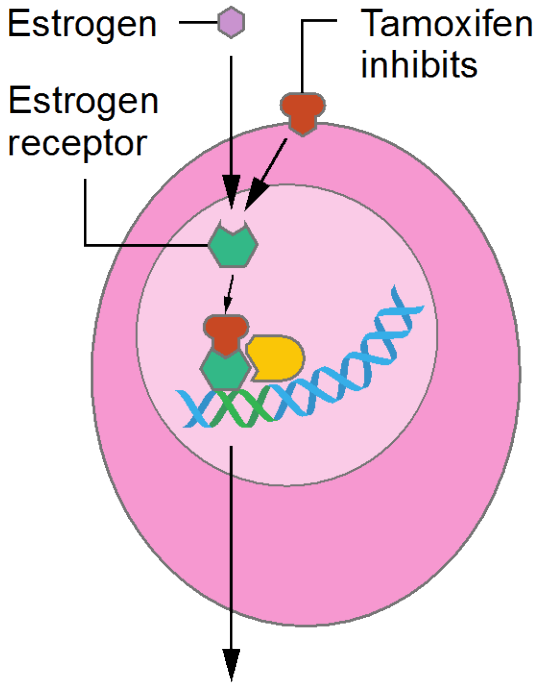


Estrogen Receptors Trigger Gene Activation



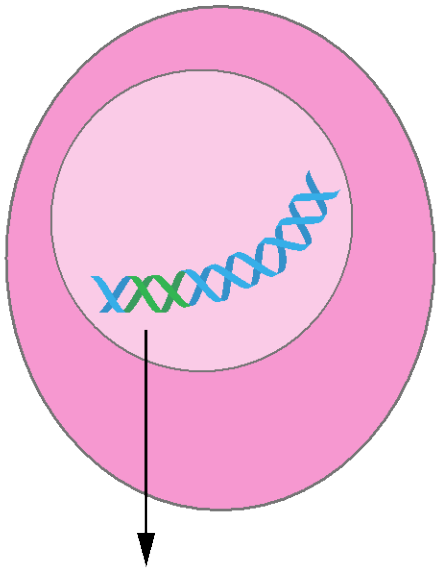
Estrogen Receptor-Negative Breast Cancer

Estrogen receptor-positive breast cancer



Cell proliferation
- Controlled by estrogen
- Inhibited by tamoxifen

Estrogen receptor-negative breast cancer



Cell proliferation
- Not controlled by estrogen
- Not inhibited by tamoxifen

Analysis