Analysis of omics data

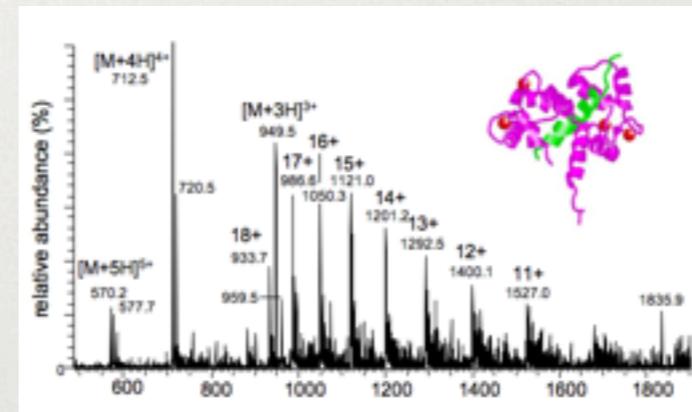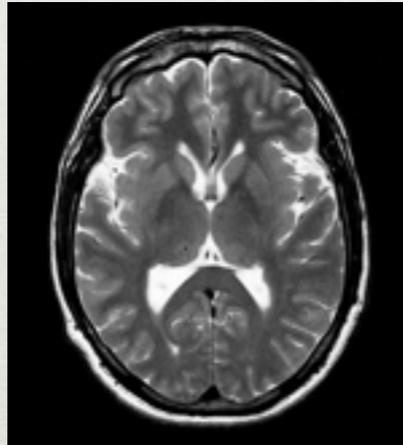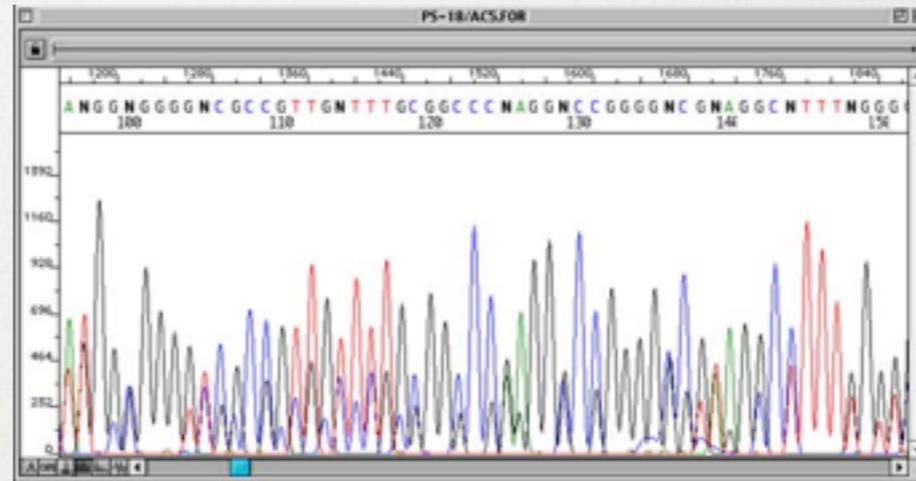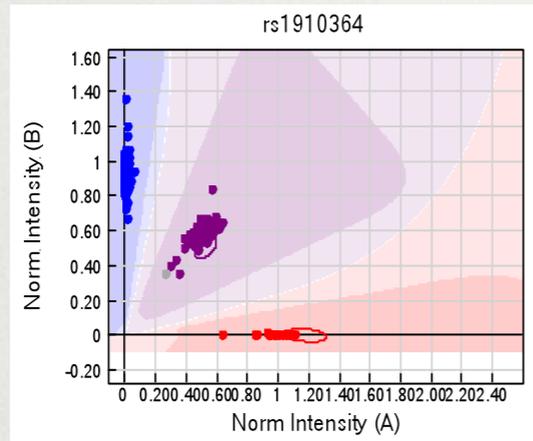## Yurii Aulchenko
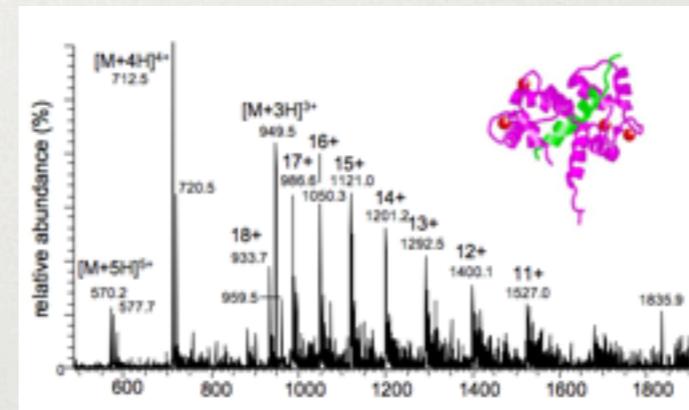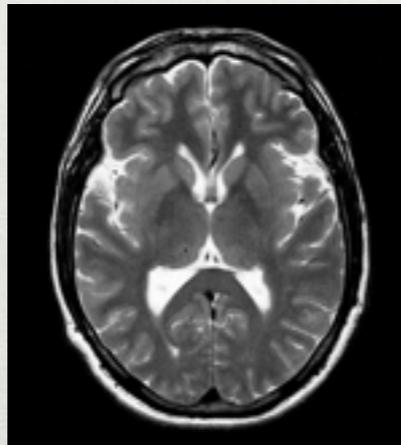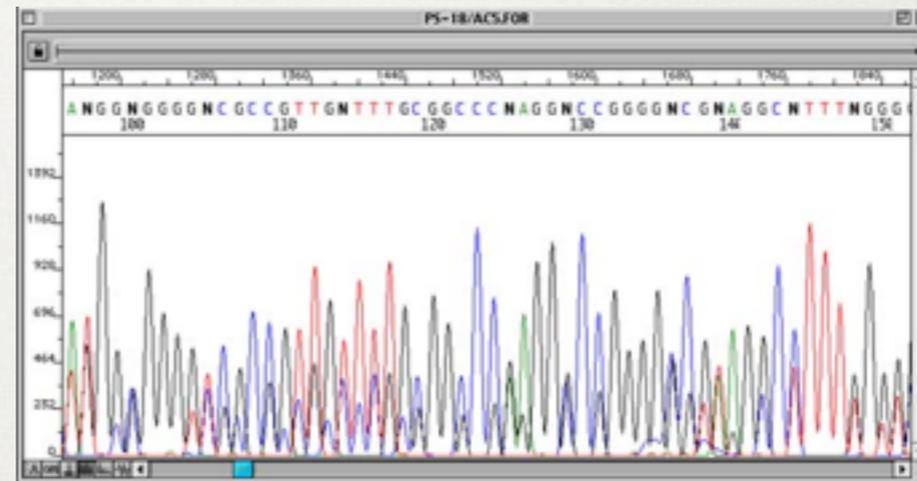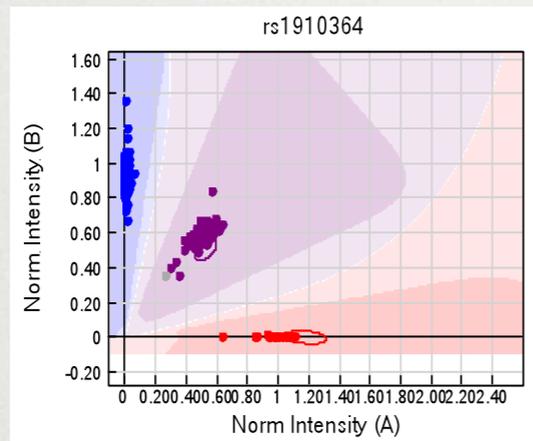Independent researcher & consultant
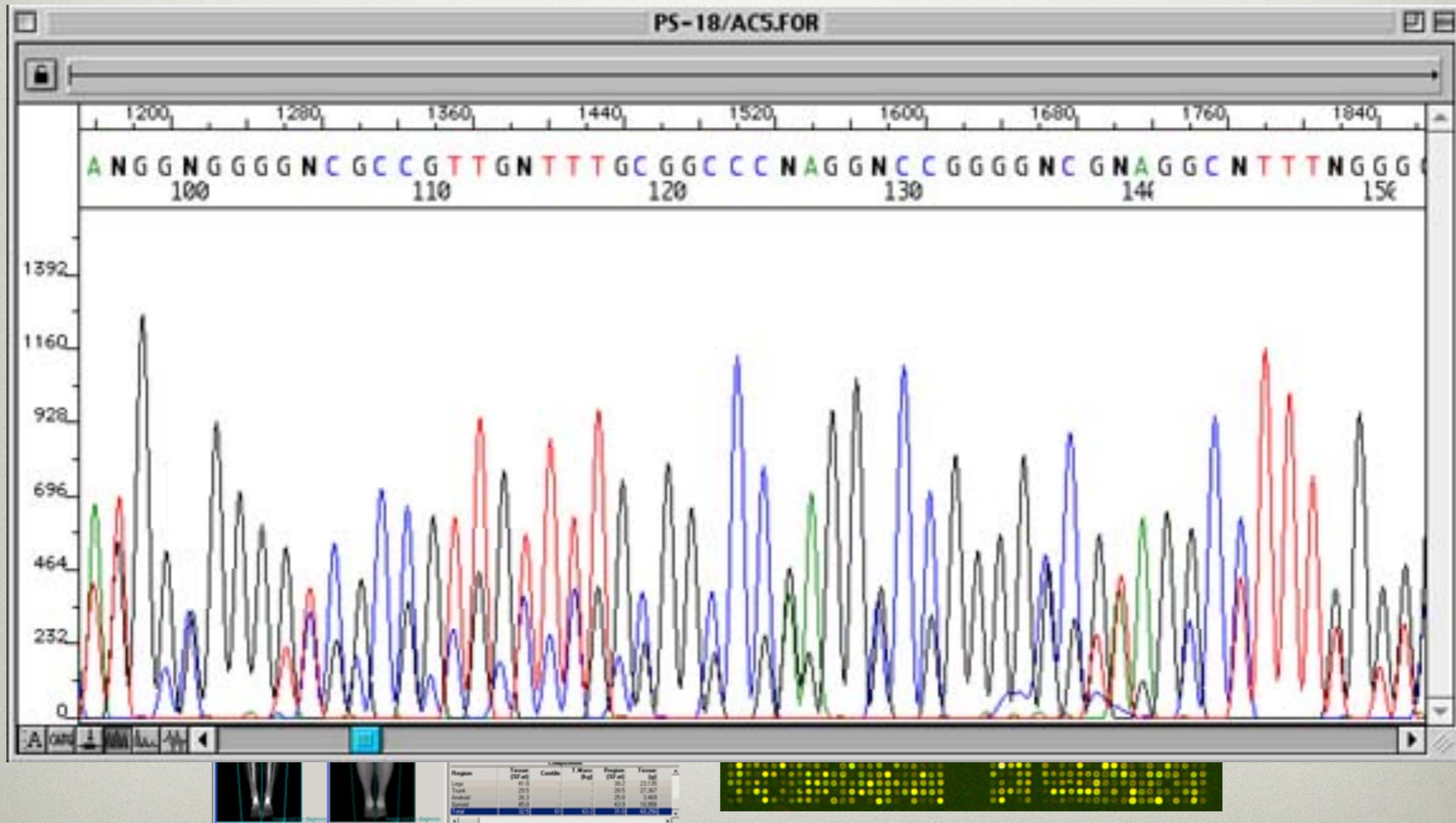yurii [dot] aulchenko [at] gmail [dot] com

Omics data

Genome (SNP chip)
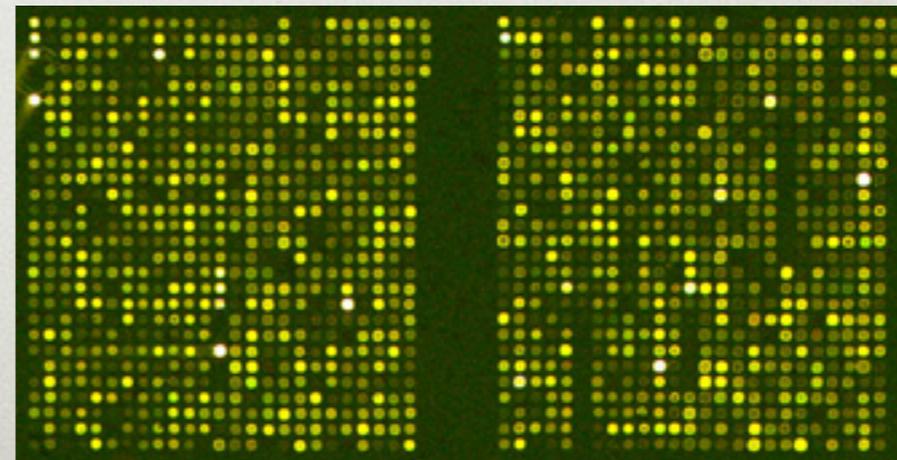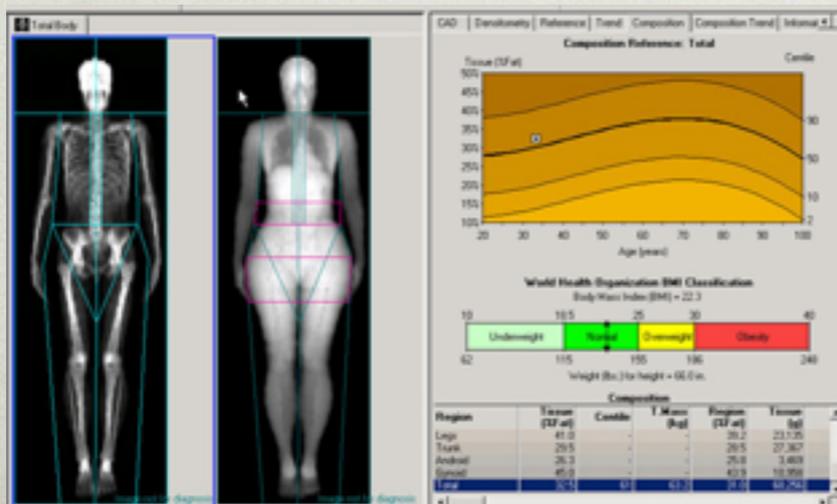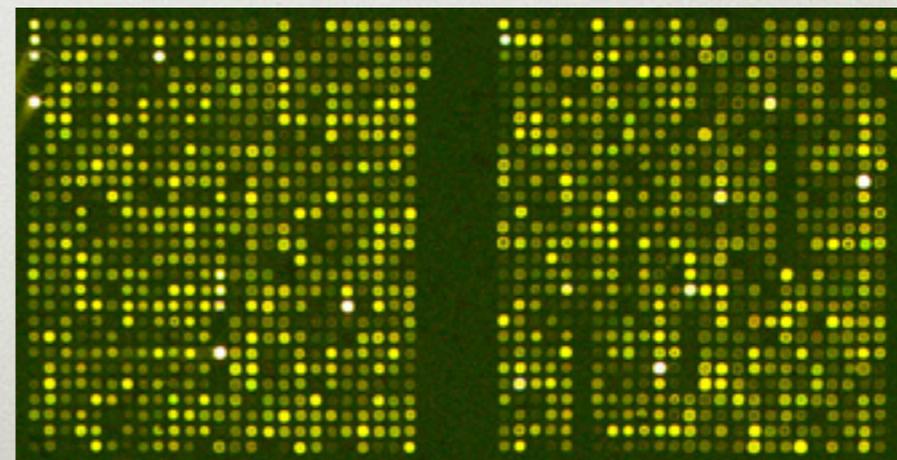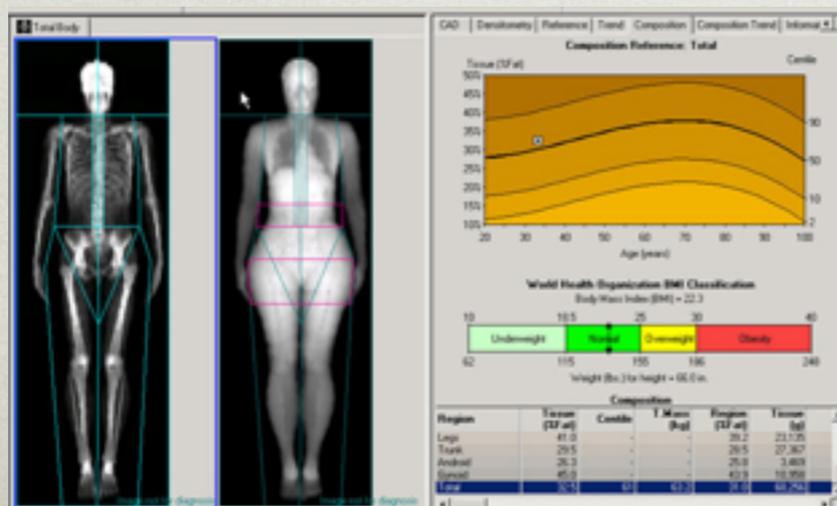
Genome (sequencing)

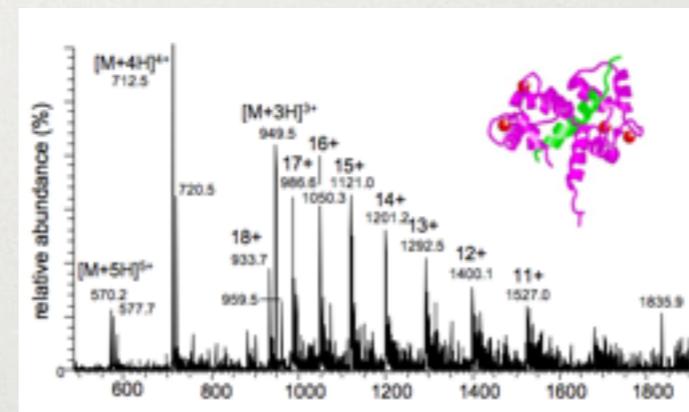# Genome (sequencing)

Transcriptome
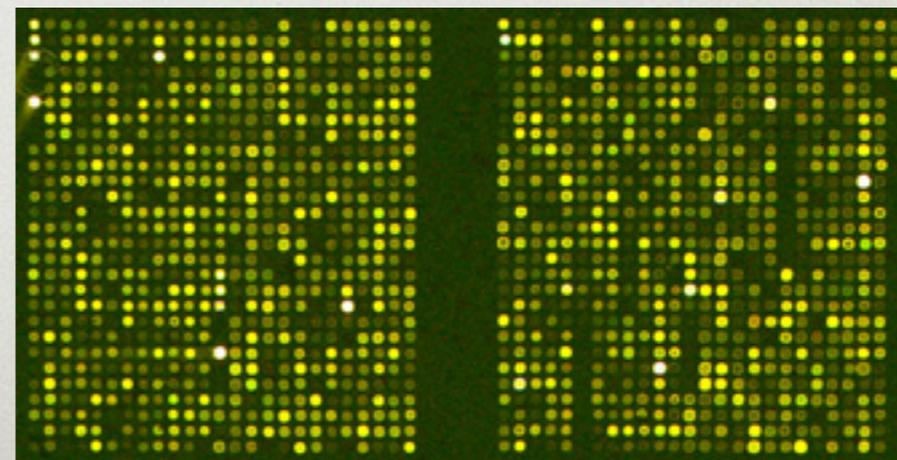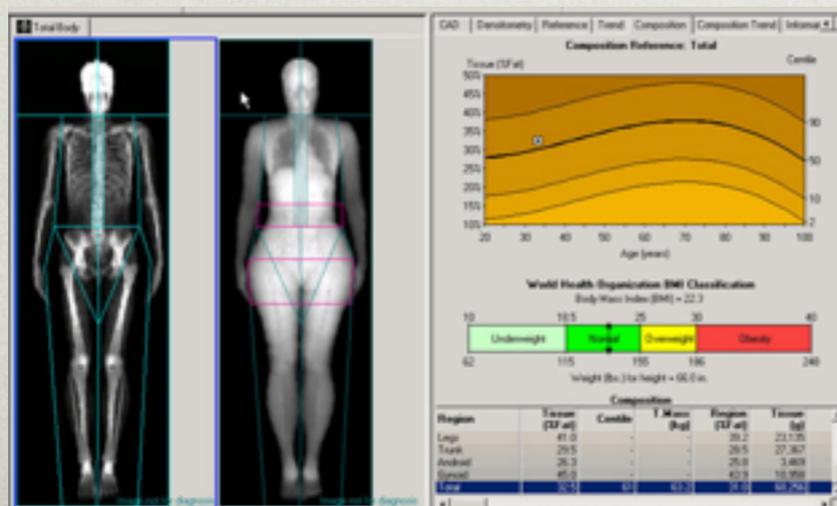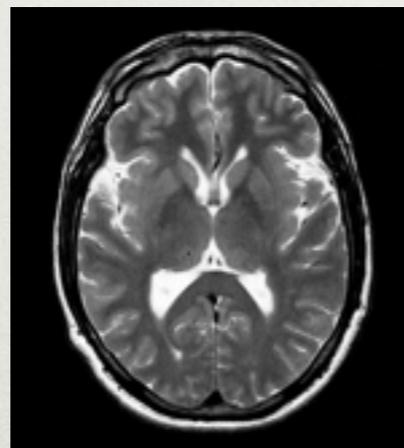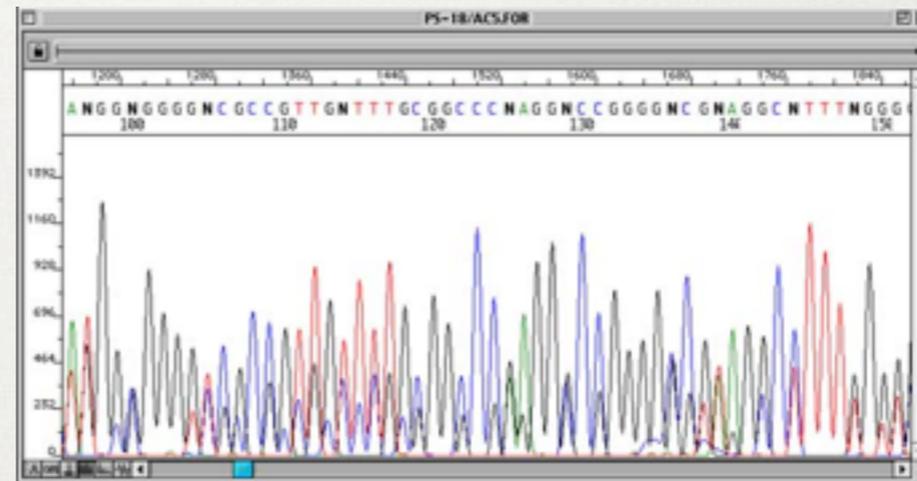
Transcriptome
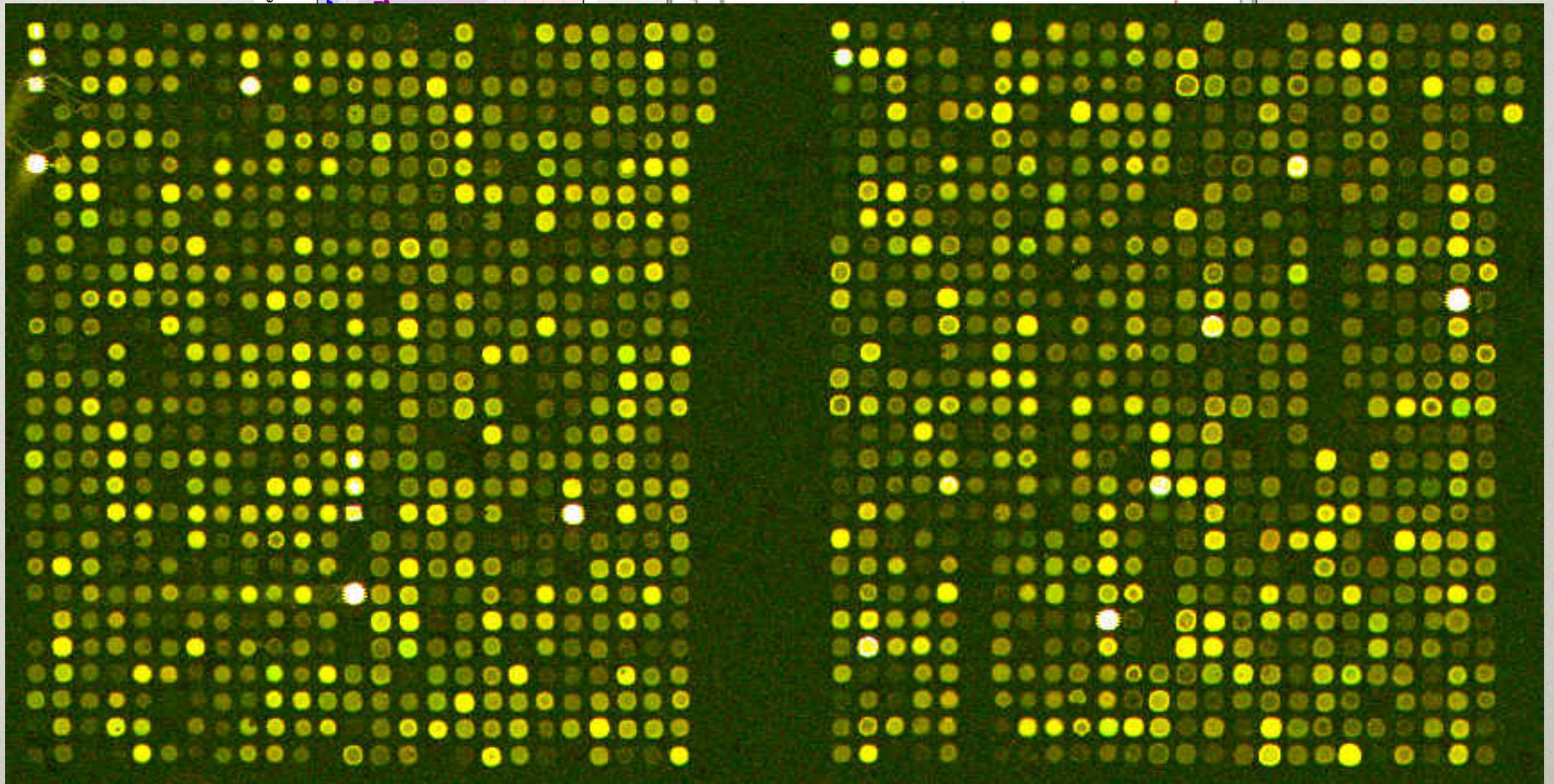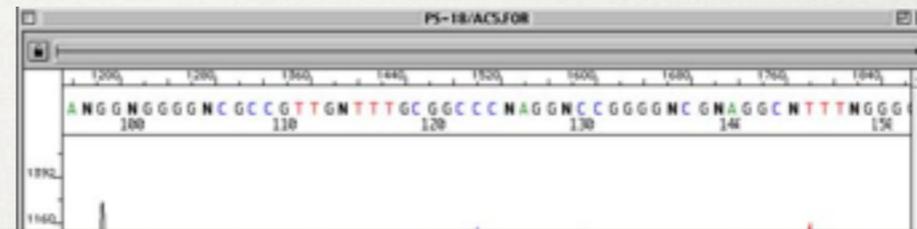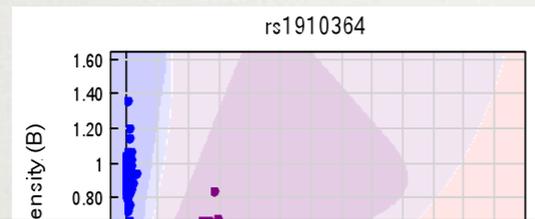
Metabolome
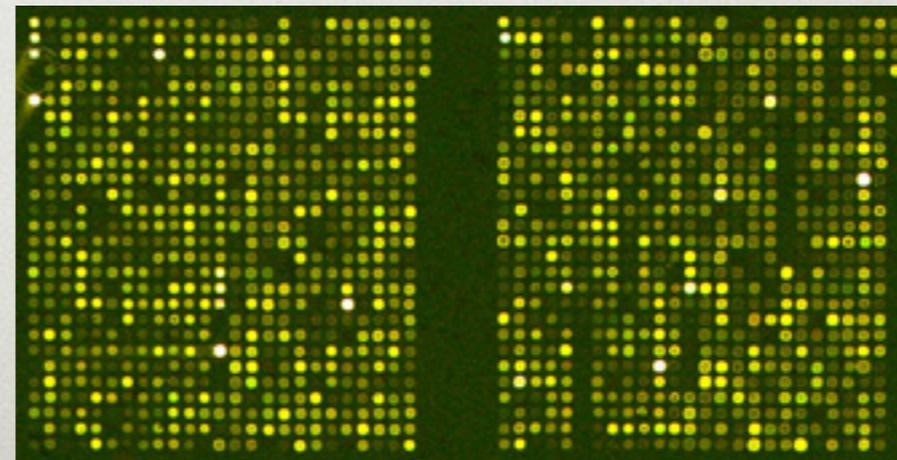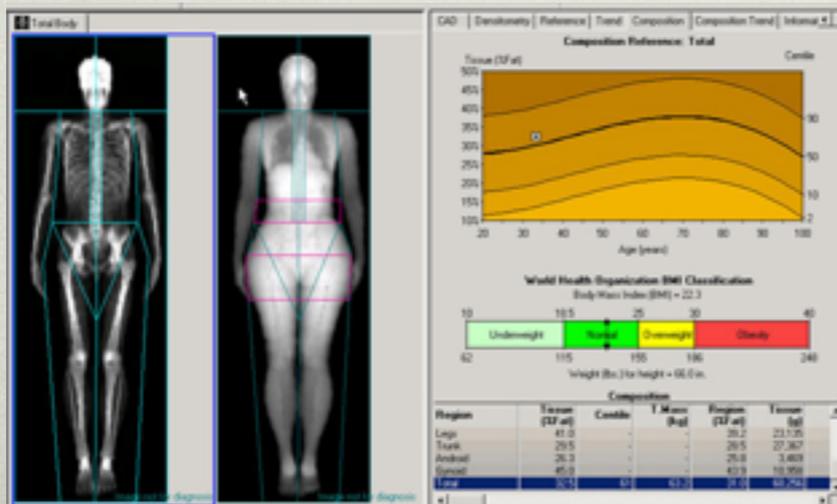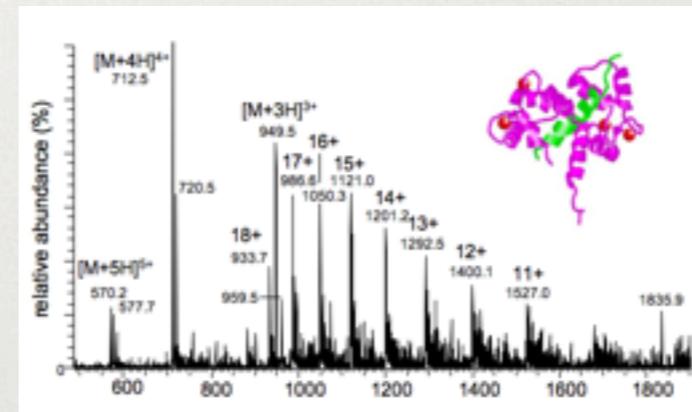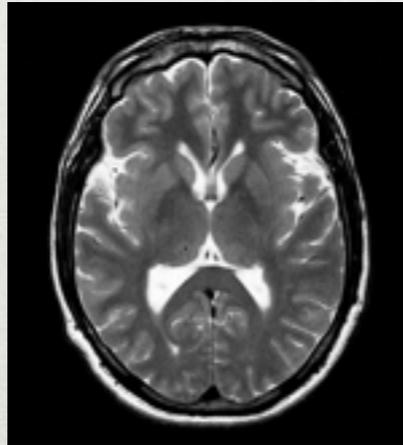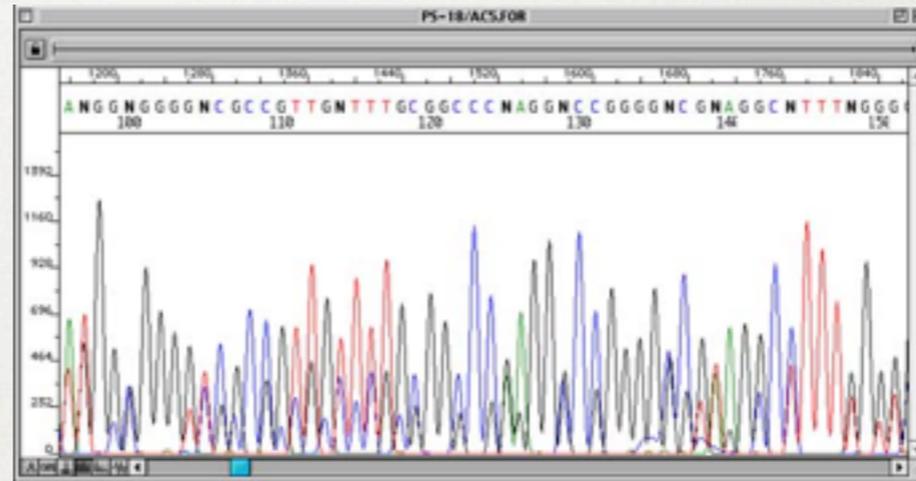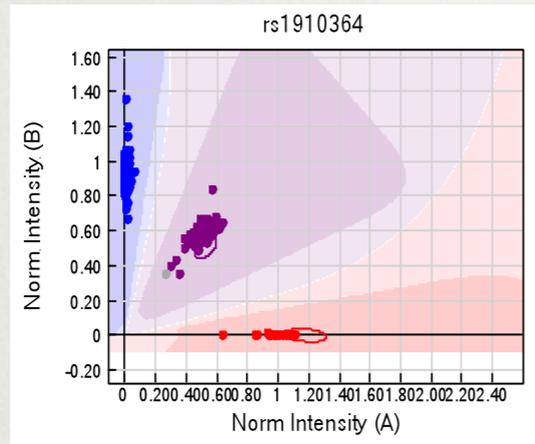
Metabolome
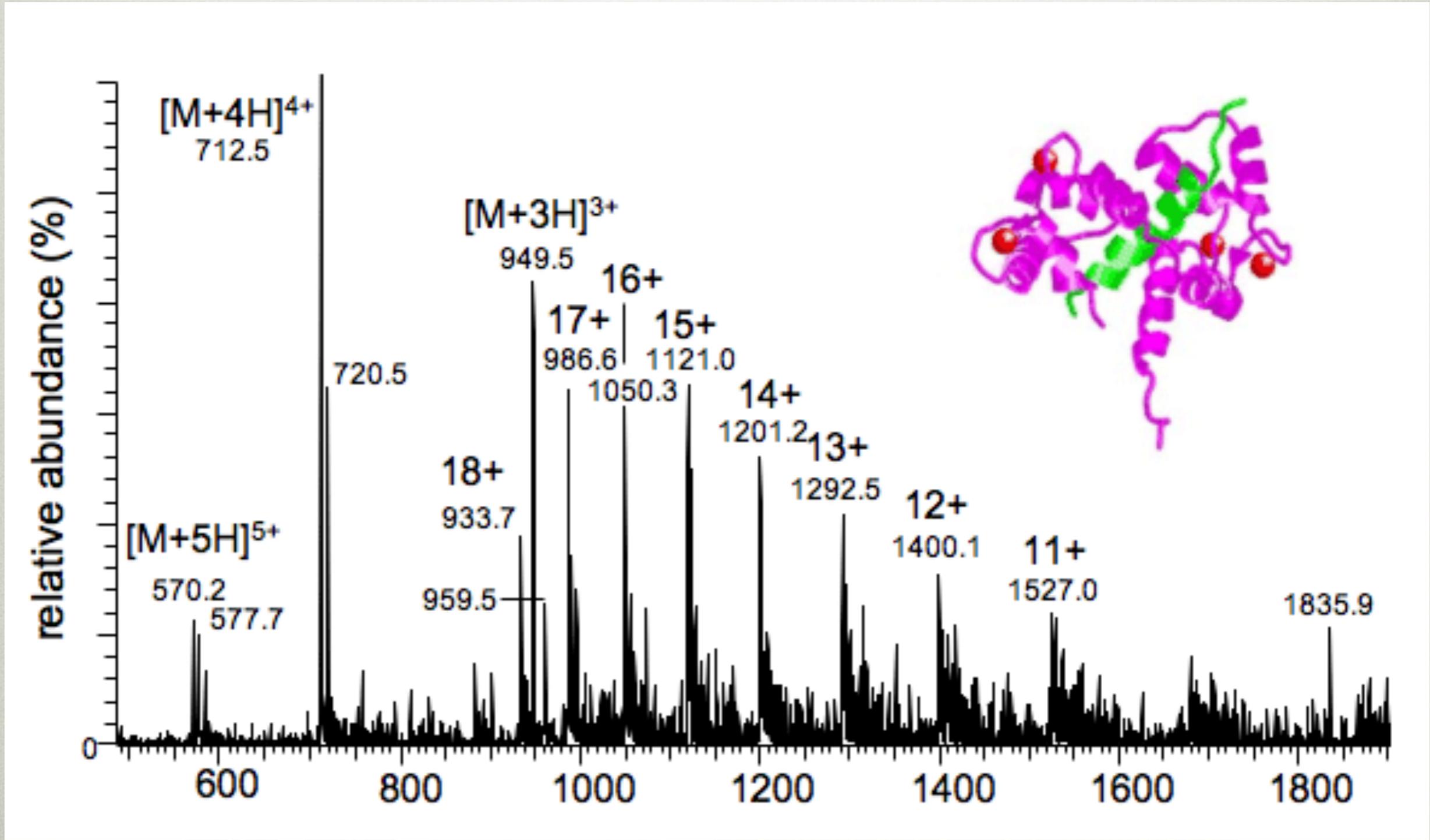
Phenome

Phenome

Phenome

Phenome

Omics data

Analysis of omics data

Analysis of omics data

- Large amounts of data

- Complex models

DATA ANALYSIS

Data analysis

Data analysis

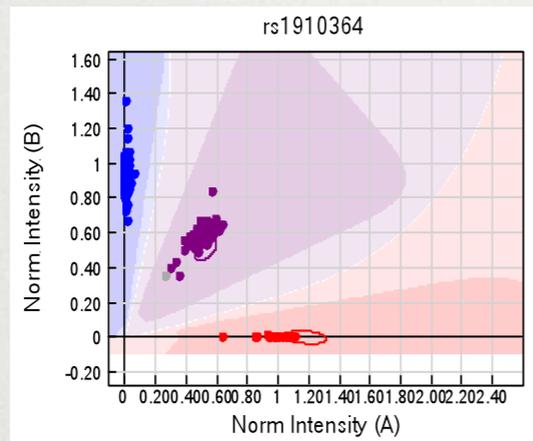**Data** + Best method =

Data analysis

Data + Best method = ⬤

Data analysis

Data + Best method =
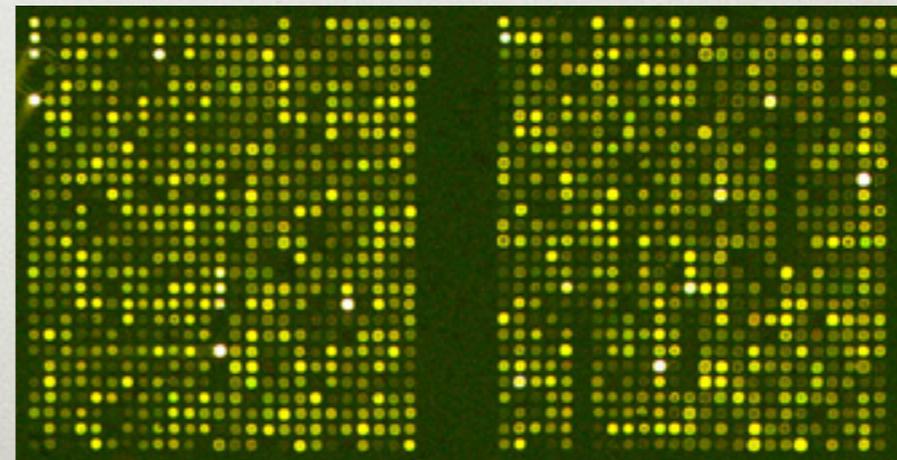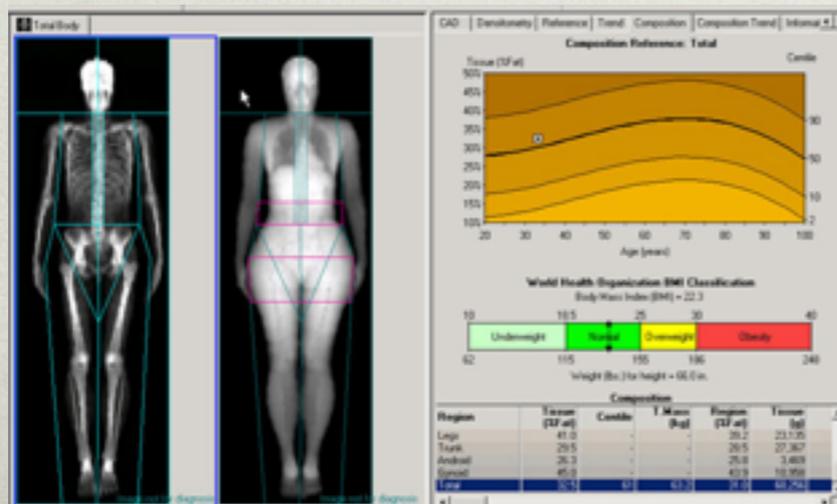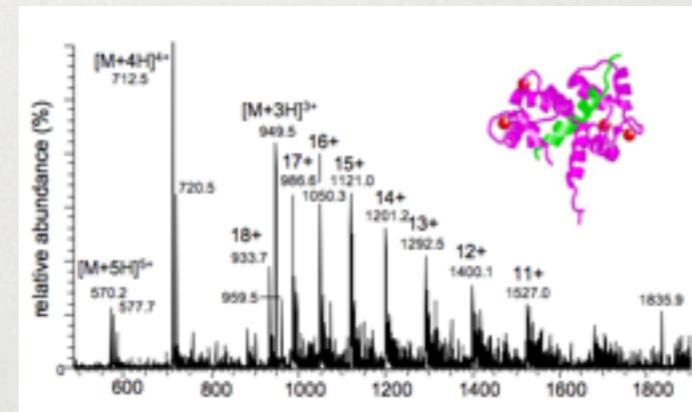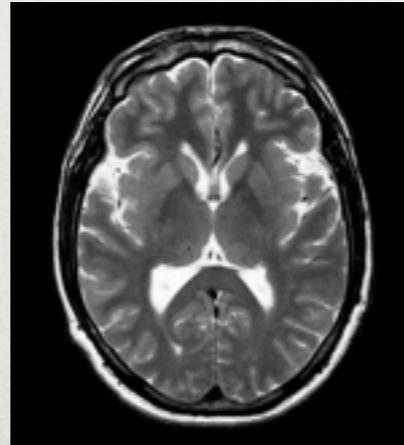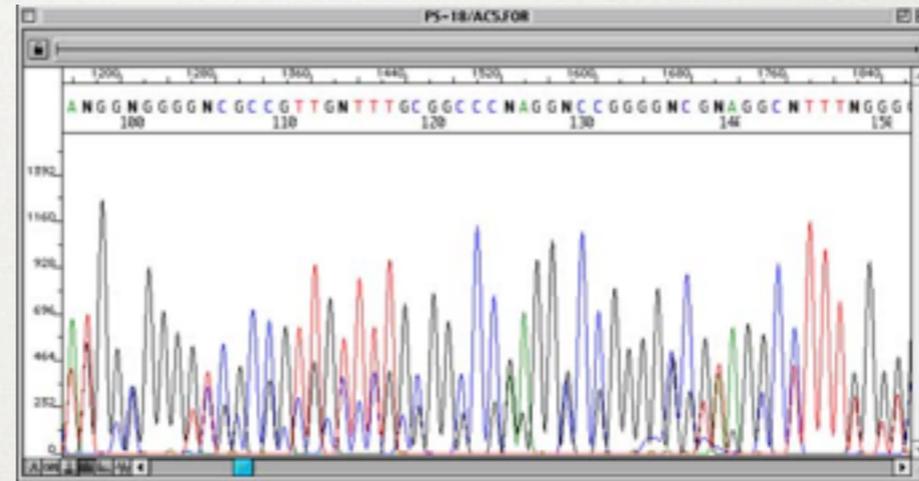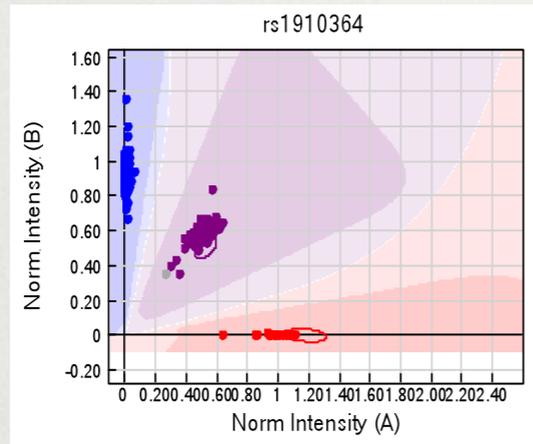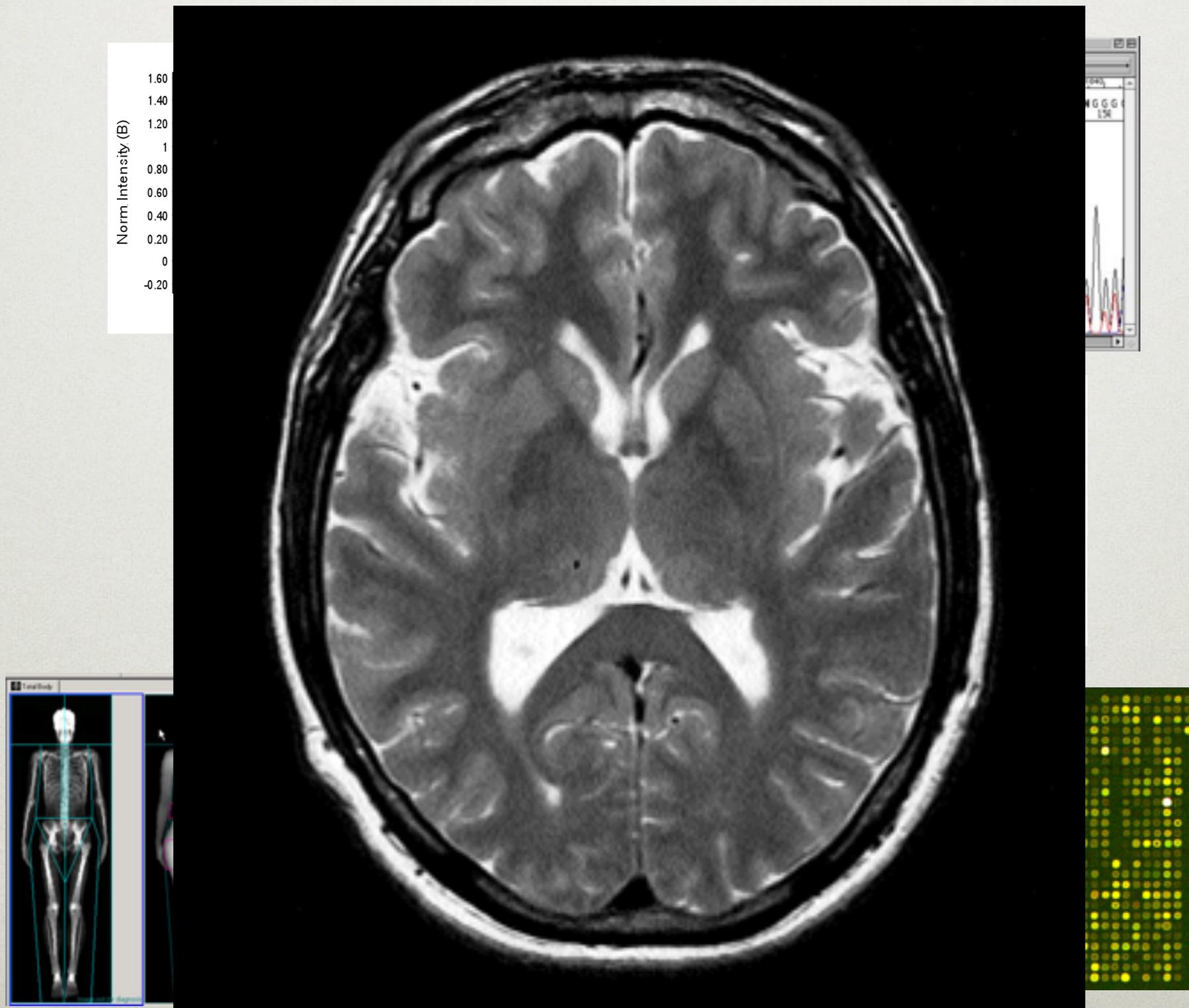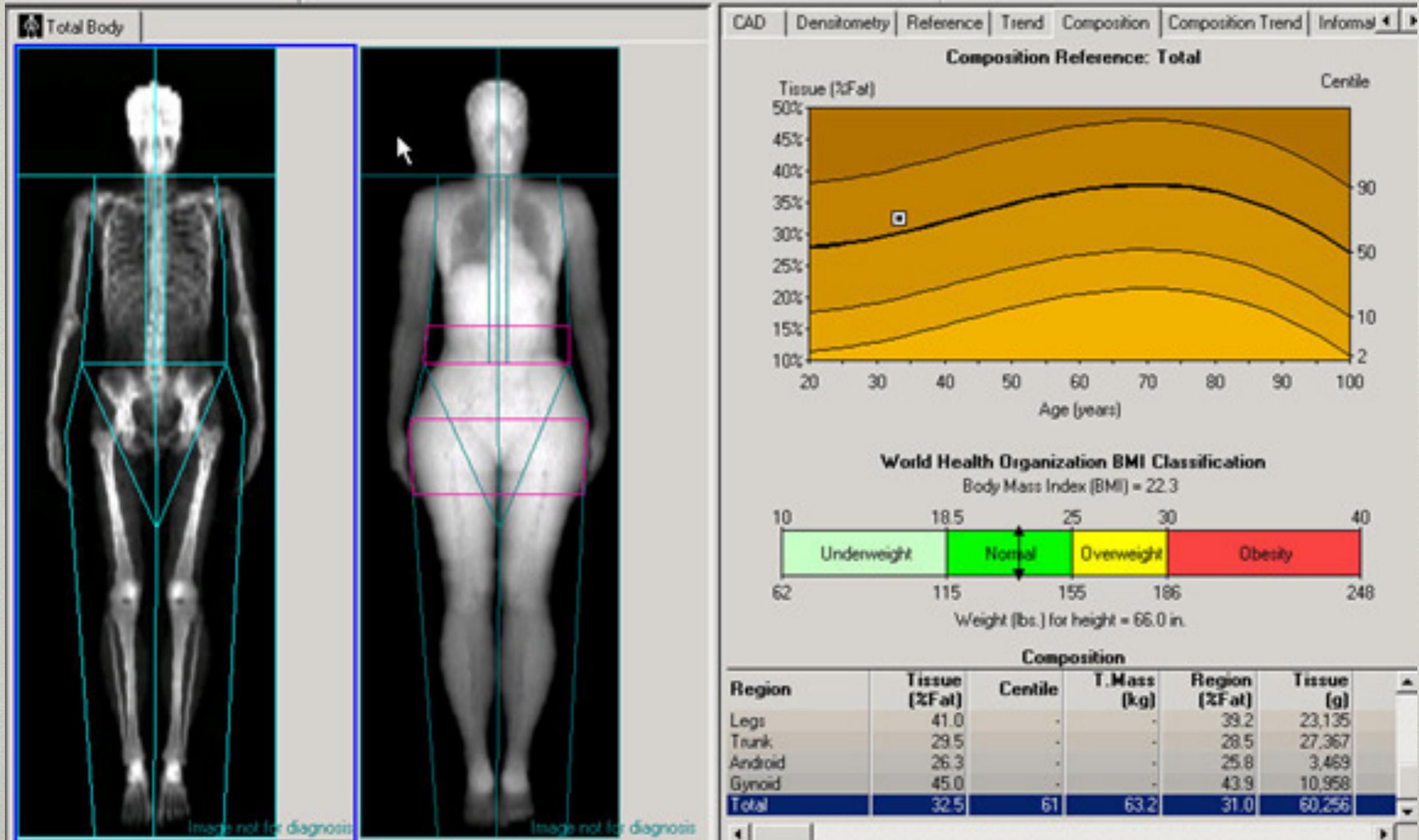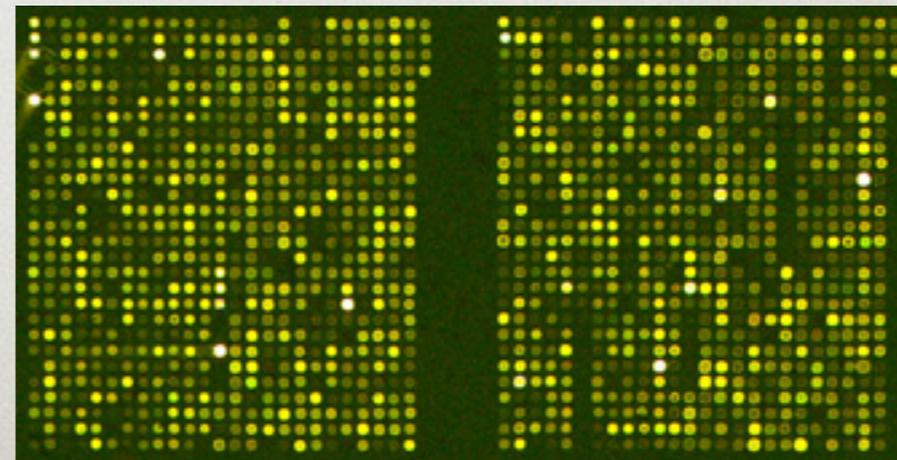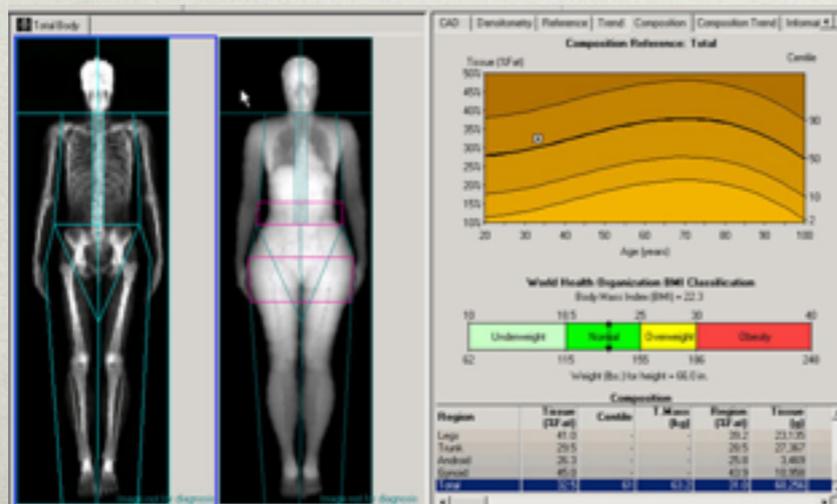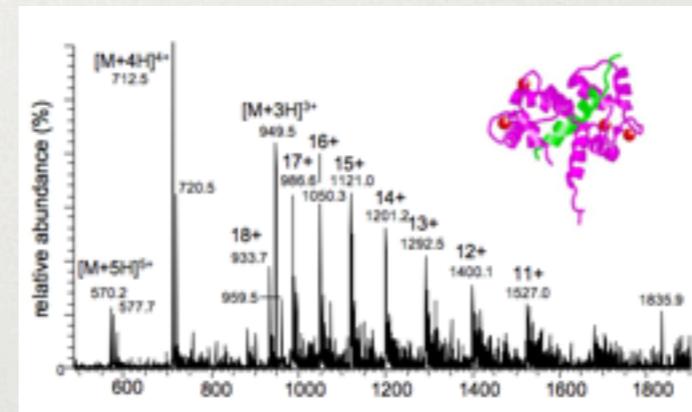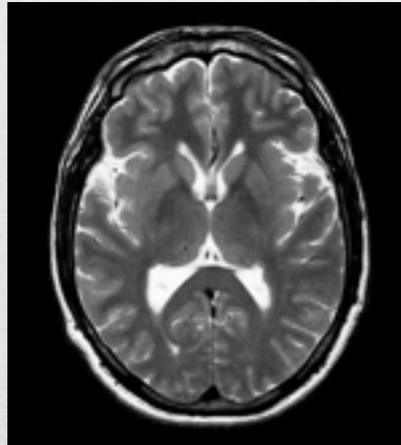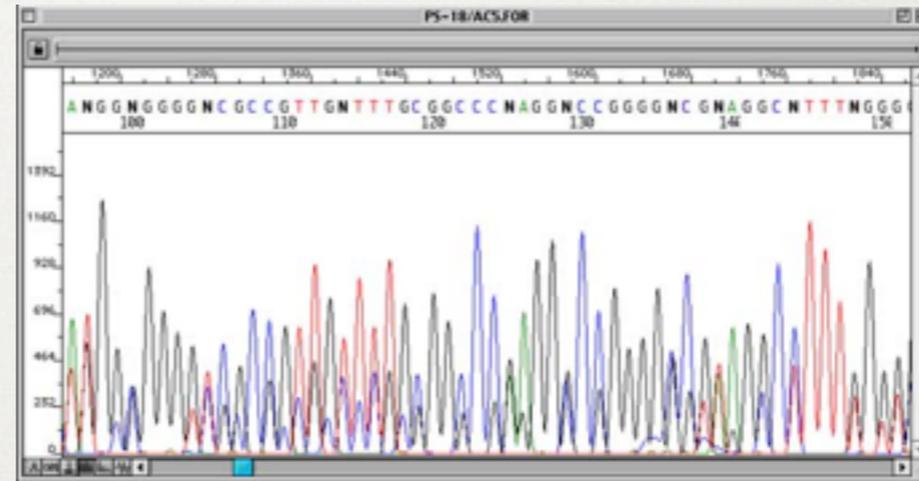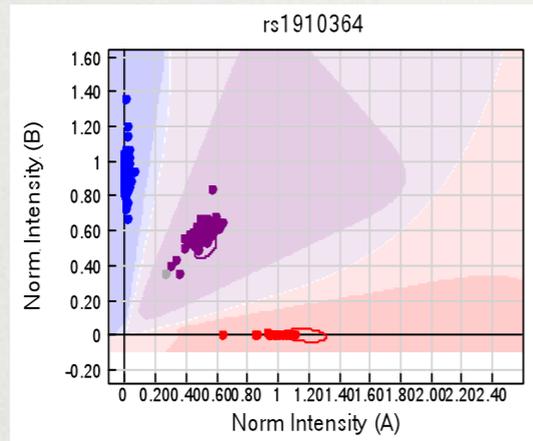
DATA ANALYSIS

Data $+$ $=$

Data analysis

Data $+$ Software $=$
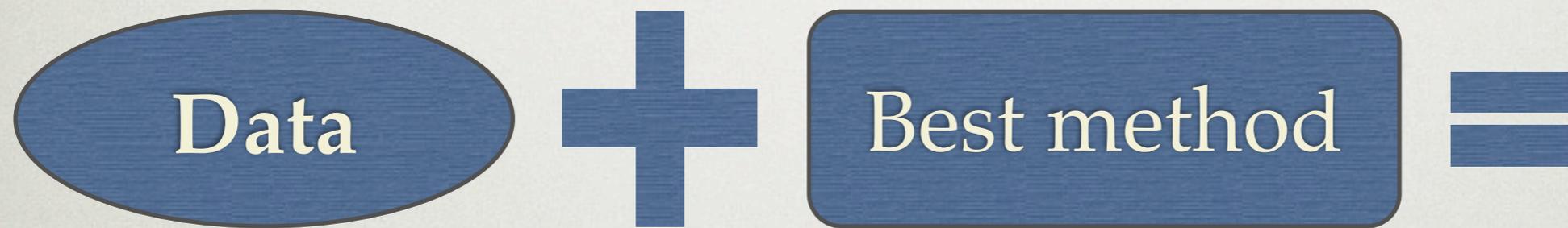
DATA ANALYSIS
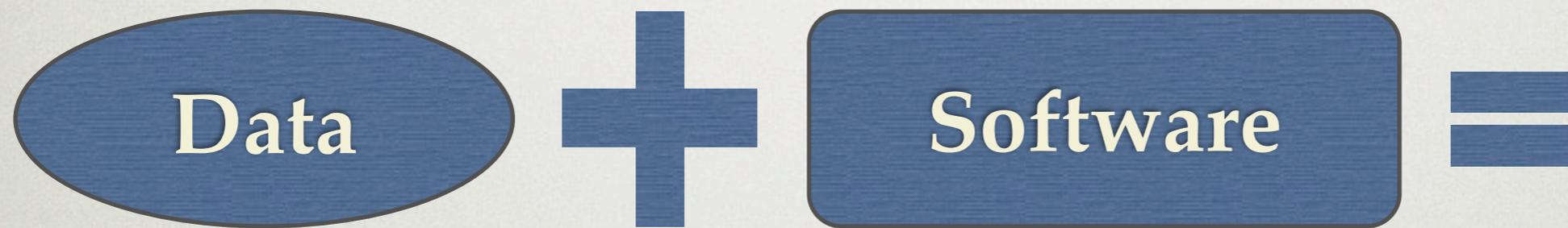
Data + **Software** = **Results**

Method to software

Method

Method to software

Approximation

Method

Method to software

Algorithm

Approximation

Method

Method to software

Implementation

Algorithm

Approximation

Method

Method to software

Software

Implementation

Algorithm

Approximation

Method

The GenABEL project

www.GenABEL.org

Example problem

- Genome-wide analysis of association of "omics" data in genetically isolated population

- 2,500,000 genetic locations

- 30,000 traits

- Hence ~ 10^{11} tests (in 3,000 individuals)

State in 2007

- Singe test takes ~5 minutes (ASReml)

- ... single GWAS takes 24 years

- ... 700k years to analyze 30k phenotypes

State in 2009

- Singe test takes <0.01 second (mmscore)

- ... single GWAS takes 5 hours

- ... 17 years to analyze 30k phenotypes

State in 2012

- Singe test takes ... (depends)

- ... single GWAS takes <1 hour

- ... few days to analyze 30k phenotypes