

Bioinformatics merges recent advances in molecular biology and genetics with advanced statistics and computer science technology. The goal is increased understanding of the complex web of interactions linking the individual components of a living cell to the integrated behaviour of the entire organism.

Functional genomics

The genome is the genetic material in a cell; the blueprint of life, and the functional units of the genome are called genes. Sequencing the genome consists of identifying the position and DNA-sequence of each gene. The genome of many organisms is now fully sequenced, and important milestones in sequencing the humane genome have been reached. We have thus moved from the pre-genome era to the post-genome era. The most important challenge in the post-genome era is to understand the role of each gene, *functional genomics*. Functional genomic research can e.g. provide a better understanding of how an individual's genetic inheritance affects the body's response to drugs, providing better diagnosis and treatment of different diseases. New techniques, for example cDNA microarrays, make it possible to study the role of thousands of genes simultaneously. Such techniques produce enormous amounts of data. Bioinformatics merges these new techniques with advanced statistical methods and

computer science technology to organize, analyse and interpret data.

Statistical methods for analysing gene expression data

NR has identified problems where our statistical competence is vital in transforming the huge amounts of functional genomics data into important pieces of knowledge.

- ✓ Designing a cDNA microarray experiment: How should the samples be distributed on the arrays to get the most accurate results?
- ✓ From images to datasets: Finding the optimal way to do gridding, segmentation, relating spot intensities to background intensities and normalize intensity data within- and between arrays.
- ✓ Find similar groups of samples and/or genes using statistical clustering methods.
- ✓ Discriminate between groups of samples or classify a new sample to one of many groups. This can be done using statistical methods for discrimination and classification.
- ✓ Find genes that are differentially expressed based on hierarchical Bayesian inference, classification modelling or statistical tests.

The need for statistical thinking

To achieve the best results close cooperation between researchers in statistics and researchers in genetics, biochemistry, medicine and/or biology is needed. We currently cooperate with the The Norwegian Radium Hospital Microarray Project, we participate in several FUGE projects, and we are working to establish further cooperation with academic and commercial actors in the field in Norway and abroad. (FUGE is a programme in the Norwegian Research Council for functional genome research.)

For more information about statistical bioinformatics at NR, please contact Senior Research Scientist Marit Holden at Marit.Holden@nr.no or visit our web pages at <http://www.nr.no/documents/samba>.

The Norwegian Computing Center (NR) is a leading research institute in statistical-mathematical modelling, pattern recognition, image analysis and central areas of information and communication technology. Our 100 researchers work with applied problems for domestic private companies, public sector, Norwegian Research Council, as well as EU and international companies. We develop research results that make a difference!

**Norwegian Computing Center, P.O. Box 114 Blindern, NO-0314 Oslo, Norway
Tel: +47 22 85 25 00, Fax: +47 22 69 76 60, Web: www.nr.no**