SYSTAT Application Notes

Statistical Methods in Biotechnology Using SYSTAT

Introduction

Biotechnology is the application of techniques and processes that utilize biological systems for the efficient and useful production of materials to serve human needs in agriculture, medicine, industry and daily life. Although biotechnology had its beginnings in man's earliest cultivation of crop plants, the production of wines and cheeses, and the domestication of animals, modern development in the field have been greatly stimulated by recent advances in biochemistry and molecular biology.

The history of human achievement has always been episodic. For a while, one particular field of endeavor seems to hold sway as the preserve of genius and development, before the focus shifts and development forges ahead in dizzy exponential rush in an entirely new direction. So it was with art in the renaissance, music in the 18th century, engineering in the 19th and physics in the 20th. Now it is the age of biology.

Darwin died in 1882.Ninety-nine years after his death, the first patent for a genetically modified organism was granted to Ananda Chakrabarty of General Electric, USA relating to a strain of Pseudomonas aeruginosa engineered to express the genes for certain enzymes in order to metabolize crude oil.

Twenty years later still, in the year that saw the first working draft of the human genome sequence published and the announcement of the full genetic blueprint of the fruit fly, Drosophila melanogaster, that archetype of eukaryotic genetics research, biotechnology has become a major growth industry with increasing numbers of companies listed on the world's stock exchanges.

Biocomputing and Biostatistics

As with all fields of technology, biotechnology relies heavily on computers and on the gathering and processing of data. Biotechnologists use computers to obtain, analyze and present data. In fact, with the exponential knowledge growth, the organization and retrieval of data is now becoming a critical issue in Biotechnology and this has given rise to a new branch of Biotechnology, called Biocomputing.

An example of the product of this field lies in the development of the Human Genome Project in which Biotechnologists are attempting to sequence all human DNA. As soon as a new human gene is characterized, the information is placed on the Internet. Scientists throughout the world can therefore share and use the information as part of a huge global cyber community. Biostatistics is another important element of Biotechnology.

Biotechnology experiments and their results are often very complex. Because large sums of money and often people's lives are at stake, results have to be meaningful and experiments have to be designed such that the results can be interpreted in as useful manner.

Biostatistical applications in biotechnology have increased tremendously in recent years. With the help of statistical software a biotechnologists / analysts may:

❑ Identify statistically significant differences in a single variable, between two or more groups.

❑ Design efficient single factor experiments, which are fit for purpose and economical with scarce experimental resource.

❑ Fit simple calibration curves to biological data. Produce informative data summaries for experiments with measurements on multiple variables.

❑ Compare proportions between two or more groups.

❑ Identify sources of variability in an experiment and design efficient experiments with appropriate choice of the number of replicates and level of replication.

❑ Analyze data from experiments with multiple factors.

❑ Compare the response profiles on multiple variables between groups.

Experimental Design (Smyth et al. (2003))

Before carrying out a microarray experiment one must decide how many microarray slides will be used and which mRNA samples will be hybridized to each slide. Certain decisions must be made in the preparation of the mRNA samples, for example whether the RNA from different animals will be pooled or kept separate and whether fluorescent labeling is to be done separately for each array or in one step for a batch of RNA.

Careful attention to these issues will ensure that the best use is made of available resources, obvious biases will be avoided, and that the primary questions of interest to the experimenter will be answerable. The literature on experimental design is still small. It is not possible to give universal recommendations appropriate for all situations but the general principles of statistical experiment design apply to microarray experiments.

In the simplest case where the aim is to compare two mRNA samples, A and B say, it is virtually always more efficient to compare A and B directly by hybridizing them on the same arrays rather than comparing them indirectly though a reference sample. In an experiment where the intention is to compare several mutant types with the wild type, the obvious design treats the wild type RNA effectively as a reference sample.

When more than two RNA samples are to be compared, and all comparisons are of interest, it may be appropriate to use a saturated design. In time-course experiments a loop design has been suggested.

For more complicated designs, with many samples to be compared, direct designs become more cumbersome and it may be more appropriate to use a common reference sample. Factors to be considered in designing the experiment include the relative cost and availability of reference versus treatment RNA as well as the cost of the arrays themselves. In direct comparison experiments it is generally advisable to use dye-swap pairs to minimize the effects of any gene specific dye-bias.

The choice of experiment design depends not only on the number of different samples to be compared but also on the aim of the experiment and on the comparisons, which are primary interest.

SYSTAT offers three methods for generating experimental designs: Classic DOE, the DOE Wizard, and the DESIGN command.

Classic DOE provides a standard dialog interface for generating the most popular complete (full) and incomplete (fractional) factorial designs. Complete factorial designs can have two or three levels of each factor, with two-level designs limited to two to seven factors, and three-level designs limited to two to five factors. Incomplete designs include: Latin square designs with 3 to 12 levels per factor; selected two-level designs with 3 to 1 I factors and from 4 to 128 runs; 13 of the most popular Taguchi designs; all of the Plackett and Burman two-level designs with 4 to 100 runs; the 6 three-, five-, and seven-level designs described by Plackett and Burman; and the set of 10 three-level designs described by Box and Behnken in both their blocked and unblocked versions. In addition, the Lattice, Centroid, Axial, and Screening mixture designs can be generated.

The DOE Wizard provides an alternative interface consisting of a series of questions defining the structure of the design. The wizard offers more designs than Classic DOE, including response surface and optimal designs. Optimization methods include the Federov, k-exchance and coordinate exchange algorithms with three optimally criteria available. The coordinate exchange algorithms accommodate both continuous and categorical variables. The search algorithms for fractional factorial designs allow any number of levels for any factor and search for orthogonal, incomplete blocks if requested.

The DESIGN command generates all designs found in Classic DOE using SYSTAT's command language.

Interleukin (1L)-1β is produced primarily by activated mononuclear phagocytic cells in the lung airway and functions as a potent proinflammatory cytokine. Kristin R. Coulter et al. (1999) studied the regulation of lung epithelial cell responsiveness to IL-1β, the human type 11-like airway epithelial cell line A549 and primary normal human bronchial epithelial (NHBE) cells were assayed for IL-1-specific response modifiers.

Specifically, the IL-1 type I receptor (IL-lRI), IL-1 type I1 receptor (IL-lRII), IL-1 receptor accessory protein (IL-lRAcP), and IL-1 receptor antagonist (IL-1Ra) were analyzed. Analysis of variance (ANOVA) with Tukey's honestly significant difference (HSD) testing was used to compare the IL- 1 β dose response in A549 and NHBE cells using SYSTAT.

The independent sample t-test was used to compare IL-1Ra concentrations in NHBE cells. They propose that propose that human lung epithelial cells lack the ability to down regulate IL1β activity extracellularly because of an inability to express IL-1RII.

Stratified seeds of coastal Douglas-fir (Pseudatsuga menziesii (Mirb.) Franco var, menziesii) were germinated, sown in soil, and megagametophytes were removed at various stages of early seedling development.

Yield and quality of DNA extracted from the megagametophytes were related to several morphological traits of the seedlings after 2 months of growth in a controlled environmental chamber (Krutovskii et al. (1997)).

Regression and MANOVA under the Multivariate General Linear Hypothesis (MGLH) using SYSTAT demonstrated non-linear associations between stage of megagametophyte removal and seedling size traits, DNA yield and quality, and RNA presence.

Megagametophyte removal when cotyledons had extended one-quarter of their length (about 4 nun) outside the seed coat resulted in sufficient DNA for construction of saturated PCR-(polymerase chain reaction) based genome maps and had little effect on seedling development.

Celia Carvalho et al. (2001) propose a model that predicts the intranuclear positioning of centromeric heterochromatin for each individual chromosome.

With the use of fluorescence in situ hybridization and confocal microscopy, they show that the distribution of centromeric -satellite DNA in human lymphoid cells synchronized at GO/G1 is unique for most individual chromosomes.

Regression analysis using SYSTAT revealed a tight correlation between nuclear distribution of centromeric -satellite DNA and the presence of G-dark bands in the corresponding chromosome. The data suggest that "chromosomal environment" plays a key role in the intranudear organization of centromeric heterochromatin.

The model further predicts that facultative heterochromatinization of distinct genomic regions may contribute to cell-type specific patterns of centromere localization. Research in biotechnology sector is still often seen as largely medical or pharmaceutical in nature, particularly amongst the general public.

However, while therapeutic instruments form, in many respects, the 'acceptable' face of biotechnology, elsewhere agricultural, industrial and environmental applications of biotechnology are also potentially very great.

References (in order of appearance)

Celia Carvalho et al. (2001). 'Chromosomal G-dark Bands Determine the Spatial Organization of Centromeric Heterochromatin in the Nucleus',
Molecular Biology of the Cell, Vol 12, Issue 11, 3563-3572.

Kristin R. Coulter et al. (1999).'Extracellular Regulation of Interleukin (1L)- 1 β through Lung Epithelial Cells and Defective IL-1 Type I1 Receptor Expressionl, Am. J. Respir. Cell Mol .Biol., Volume 20, Number 5, 964-975.

K.V. Krustovskii et al. (1997). 'Effects of megagametophyte removal on DNA yield and early seedling growth in coastal Douglas-fir ',Can. J. For. Res., 27,964-968.

Gordon K. Smyth et al. (2003). 'Statistical Issues in cDNA Microarray Data Analysis', In: Functional Genomics: Methods andProtocols, M. J.Brownstein and A. B. Khodursky (eds.), Methods in Molecular Biology Volume 224, Humana Press, Totowa, NJ, 2003, pages 111-136.

SYSTAT®

Corporate Headquarters
North, Central & South America
Systat Software, Inc.
2107 North First Street, Suite 360
San Jose, CA 95131-2026 USA
Phone : 800-797-7401
Fax : 800-797-7406
Email : info-usa@systat.com

UK and Ireland
Systat Software Inc
4th Floor, Block B, Vista Centre,
50, Salisbury Road,
Hounslow, - TW4 6JQ, London, UK.
Phone : +44-(0)208-538 0128
Fax : + 44-(0)208-538 0273
Email : info@systat.co.uk

Germany and Austria
Systat Software GmbH
Schimmelbuschstrasse 25
D-40699 Erkrath
Germany
Phone: +49.2104.9540
Fax: +49.2104.95410
E-mail: kontakt@systat.de

Asia-Pacific, AUS & NZ
Phone : +91 - 80 - 4128 1111
Fax : +91 - 80 - 4128 0203
Email : info-intl@systat.com