

Ben Ong^{1,3,4}, Pankaj Prakash⁵, Silvina Felitti^{1,2}, Tracie Webster^{1,3}, Pei Tian^{1,2}, Marc Ramsperger^{1,2}, Kate Shields^{1,2}, David Edwards^{1,4}, Soheil Shams⁵ and German Spangenberg^{1,2,3,4}

¹Plant Biotechnology Centre, Primary Industries Research Victoria, Department of Primary Industries, La Trobe University, Bundoora, Victoria 3086, Australia

²Molecular Plant Breeding CRC, Australia

³Victorian Microarray Technology Consortium, Australia

⁴Victorian Bioinformatics Consortium, Australia

⁵BioDiscovery Inc., USA

The aims of our research are to undertake: a) the application of GeneDirector system to manage data generated along the entire gene expression process, b) demonstrate the use of this system on managing data associated with experiments on plant and fungi and c) analysing data associated with experiments on fungi.

INTRODUCTION

With the rapid increase in the use of gene expression microarray technology for plant and animal research experiments, a significant challenge has emerged on how to best manage the large volumes of data generated by this technology. It is well known that a single gene expression array experiment can produce measurement data for tens of thousands of genes in parallel. Managing this so called quantification data is just one of the challenges faced by researchers today. An even more significant problem is managing the data associated with the entire experiment, from the spotted probes, to sample extraction protocols, to image quantification parameters while maintaining the relationships inherent among all of these diverse data types. Additionally, as more scientists utilise microarray tools as part of their research, a centralised enterprise-wide system is required to allow for secure collaboration among researchers.

This work is collaboration between Victorian Microarray Technology Consortium and BioDiscovery. For more detail biological information of the microarray experiment, please refer to concurrent poster by Felitti et al.

METHODS AND RESULTS

1) GeneDirector System

GeneDirector is a microarray data management system that has both laboratory information management system (LIMS) and data analysis capabilities (Figure 1). GeneDirector is compatible with arrays, arrayers, and scanners from various manufacturers and possesses platform-independent hardware integration. It is embedded with Oracle RDBMS.

GeneDirector possesses management capabilities for data, project, array, sample, and results (Figure 1). GeneDirector tracks the entire gene expression experiment process within a project in a step by step work flow format. The information between each step is interlinked and information within each step needs to be completed to proceed to the next step. This ensures all relevant information is captured. In addition, it has three computational modules: CloneTracker (microarray design and management), ImaGene (microarray image analysis), and GeneSight (microarray data analysis).

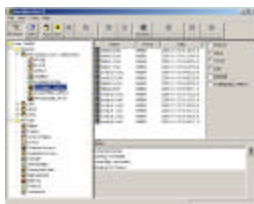


Figure 1. Overview of GeneDirector data management software. The entire gene expression experimental process is captured step by step in folders format from probe data to analysed data. In addition, relevant information related to the microarray data itself such as protocols used, biological materials used (cell lines, tissues, RNA, oligo, cDNA, etc.), labelling dyes used (Cy3, Cy5), and the information related to manufacturers, materials, compounds are stored in the database.

2) *Neotyphodium* Gene Expression Experiments Data Integration In GeneDirector

First, a project within the GeneDirector's user directory was created. Second, data relevant to the *Neotyphodium* gene expression experiments were entered into the project according to the work flow i.e. PLATE, ARRAY DESIGN, ARRAY, SAMPLE and HYBRIDISATION. Text files containing gene id in 384-well format were imported into the folder PLATE. These gene id were linked onto the microarray slides using Clone Tracker within the folder Array Designs. The number of microarray slides used for the experiments was entered (folder ARRAY). In the folder SAMPLE, information relating to the tissues used, harvest date, labelling, Cy3 and Cy5 dyes used were recorded. Then hybridisations were setup virtually by pairing the labelled targets with the microarray slides.

3) Microarray Image Analysis

The images of the *Neotyphodium* gene expression experiments were entered into the folder SCANNED_ARRAY. Each pair images (reference and experiment) were selected for quantification using ImaGene (Figure 2). During quantification, the microarray design and the gene id were automatically placed over the two overlay images. The outputs of the quantification were stored in the database and captured in the folder QUANTIFIED_ARRAY.

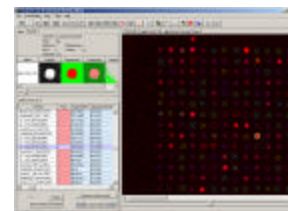


Figure 2. ImaGene performs spot quantification on superimposed images by demarcating spots with circle from a larger quantification area that defines the background. ImaGene also has a feature that perform quality of spot analysis. The output of ImaGene analysis is a text file linking the gene id with signal and background values, along with flagging values showing the quality of microarray spots as empty, poor or negative.

4) Microarray Data Analysis

The quantified arrays were then selected for data mining and statistical analysis using GeneSight. The data were prepared using a sequence of several transformations, including Lowess normalization. Confidence analysis test (bootstrap algorithm) was conducted to determine the differentially regulated genes between different experimental conditions. The data were visualized using scatter plot (Figure 3) and further mined using partitioning k-means clustering and hierarchical clustering (Figure 4).

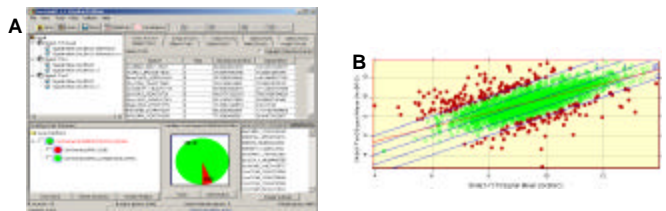


Figure 3. (A) Confidence analysis test (bootstrap algorithm) was conducted between *N. lolii* liquid and solid grown cultures. Genes that were switched-on and -off at a 99% confidence level were sub-selected (red section of the pie-chart along with their gene ids on the right). (B) A 2-dimensional scatter plot of the entire gene population displaying the differentially regulated genes at a 99% confidence level (red filled circles) between *N. lolii* liquid and solid grown cultures.

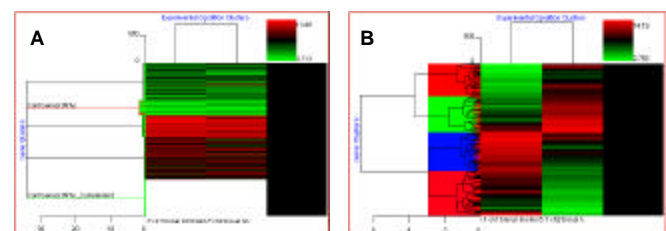


Figure 4. (A) Kmeans partitioning clustering of *N. lolii* solid grown culture (column 1) and *N. lolii* liquid grown culture (column 2). Cluster enrichment analysis indicated cluster 2 to be enriched with 99% confident genes at a probability of 0.01. (B) Hierarchical clustering of the genes found differentially regulated at a confidence level of 99% between *N. lolii* (column 1) and *N. coenophialum* (column 2) species.

EXPECTED OUTCOMES

- A single environment of GeneDirector LIMS to store microarray data.
- A single environment of GeneDirector LIMS to analyse microarray data.
- Discovery and functional analysis of novel genes.

Acknowledgements

This work was supported by the Victorian Department of Primary Industries, the Victorian Microarray Technology Consortium, Victorian Bioinformatics Consortium and Molecular Plant Breeding CRC.