BIOINFORMATICS APPLICATIONS NOTE Vol. 26 no. 19 2010, pages 2468–2469
doi:10.1093/bioinformatics/btq410

Gene expression

Advance Access publication July 14, 2010

ANAIS: Analysis of NimbleGen Arrays Interface

Adeline Simon[1,*] and Eric Biot[2]

[1]UR1290 BIOGER-CPP, INRA, Grignon and [2]Institut Jean-Pierre Bourgin, UMR1318, INRA-AgroParisTech, Versailles, France

Associate Editor: Trey Ideker

ABSTRACT

Summary: ANAIS is a user-friendly web-based tool for the processing of NimbleGen expression data. The interface reads single-channel microarray files generated by NimbleGen platforms and produces easily interpretable graphical and numerical results. It provides biologists six turnkey analysis modules—normalization, probe to gene, quality controls, differential expression, detection, queries and clustering—to explore quickly, freely and without the need for computer programming, NimbleGen transcriptome data.

Availability: http://anais.versailles.inra.fr

Contact: simon@versailles.inra.fr

Received and revised on June 3, 2010; accepted on July 6, 2010

1 INTRODUCTION

Microarray technologies have become widely used in genome-wide gene expression studies. The NimbleGen platform (http://www.nimblegen.com/), that offers customized high-density arrays with multiple long-oligos per gene, is particularly gaining popularity among biologists. Nevertheless, few complete, ready-mixed and free analysis tools are available (Carazzolle *et al.*, 2009; Wang *et al.*, 2006). Here, we report an intuitive web interface that allows scientists to conduct themselves all the analysis steps needed to interpret a NimbleGen expression experiment. In addition to the conventional processes, such as normalization and differential expression, the tool also offers interesting modules performing quality controls, detection and queries. Moreover, the toolset was developed with a sake of rapidity compatible with a remote use via Internet, and accepts all one-colour NimbleGen expression array designs.

2 FEATURES

ANAIS is composed of eight modules. The initial module is for data loading. Subsequently, the workflow includes six steps of analyses. The final module is dedicated to downloads and project management (Fig. 1).

2.1 Data input and project definition

Users are first prompted to create, select or restore a project. The setting-up of a new project is initiated by loading a zip archive containing pair and calls standard NimbleGen files, which are quite light and therefore adequate for Internet flows. The project definition is then completed by the user's experimental design.

*To whom correspondence should be addressed.

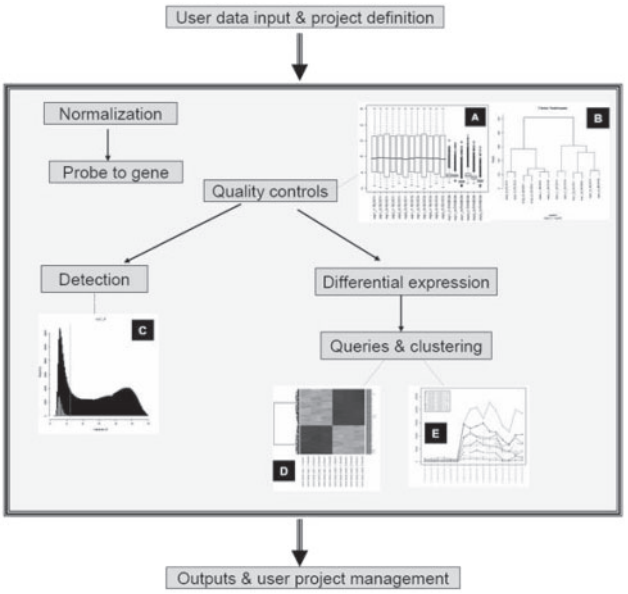

Fig. 1. ANAIS workflow and examples of graphical representations. (**A**–**B**) Box plot and clustering of arrays are useful to estimate the intensity level by probe type, and the reproducibility between replicates. (**C**) Array intensity frequency histogram, with white bar charts representing random values. Parameters for the vertical line are configured by the user, in order to define a threshold of expression. (**D**) Hierarchical clustering and heat map for queried genes. (**E**) Intensity profile curves for queried genes.

2.2 Normalization and probe to gene

Normalization is essential to analyse experiments involving multiple arrays, whereas probe to gene methods summarize gene expression from probe sets. ANAIS is able to manage NimbleGen pre-processed files, and offers also its own processes, including the widely used quantile normalization (Bolstad *et al.*, 2003).

2.3 Quality controls

ANAIS especially focuses on quality controls and offers many graphical representations to check data before further analyses. The graphs display signal intensities, either for all arrays: box plots (Fig. 1A), histograms, scatter plots of intra-array duplicates; or for two chosen arrays: scatter plots, ratio histograms, MA plots. Moreover, the reproducibility between replicates can be well estimated by ANAIS hierarchical clustering of arrays (Fig. 1B).

2.4 Detection

Hybridization signals provide information as absolute measurements of gene expression, which may, for instance, give evidence for *in silico* predicted genes. To identify the set of genes whose intensity is above a signal-to-noise threshold, ANAIS calculates widely used cut-offs from random probes (Bilban *et al.*, 2002). The novelty of ANAIS interface is to allow users to determine visually the threshold of expression on histograms where random probes and adjustable cut-off levels are represented (Fig. 1C). ANAIS also calculates a detection *P*-value, which represents the confidence that a gene is expressed above random probes. The statistical test is adapted from Illumina detection call algorithm (Archer and Reese, 2010; http://www.illumina.com).

2.5 Differential analysis

The main goal of a transcriptome experiment is to identify the set of genes differentially expressed between conditions. Based on the experimental design provided by the user, ANAIS calculates the fold changes, and performs an analysis of variance (ANOVA) for each gene (Churchill, 2004). To deal with multiple testings, ANOVA *P*-values are further adjusted by family-wise error rate (FWER) or false discovery rate (FDR) methods (Strimmer, 2008). The obtained *P*-values are used as indicators of the strength of the evidence for differential expression.

2.6 Queries and clustering

To retrieve analysis results, ANAIS offers queries based on gene names or on differential expression issue. The selected gene values can then be downloaded, textually or graphically displayed by clustering and heat map (Fig. 1D), profile curves (Fig. 1E) or volcano plots. ANAIS also provides ready-to-use text files to access more clustering functionalities with specialized softwares like Genesis (Sturn *et al.*, 2002).

2.7 Outputs and project management

All graphical and numerical results generated during the analyses steps of the current project can be downloaded as zip files. Project can be stored on the server for subsequent analyses, cleared or downloaded for further restoration.

3 IMPLEMENTATION

ANAIS is a web-based tool written in Python and R. Dynamic interfaces are generated with python cheetah template engine (http://www.cheetahtemplate.org/) and CGIwithR package (http://cran.r-project.org/web/packages/CGIwithR/). Python/R interfacing is provided by RPy package (http://rpy.sourceforge.net/). Modules use methods from the Bioconductor project (http://www.bioconductor.org/; Gentleman *et al.*, 2004).

ACKNOWLEDGEMENTS

We thank all the biologists who tested ANAIS.

Conflict of Interest: none declared.

REFERENCES

Archer,K.J. and Reese,S.E. (2010) Detection call algorithms for high-throughput gene expression microarray data. *Brief. Bioinform.*, **11**, 244–252.

Bilban,M. (2002) Defining signal thresholds in DNA microarrays: exemplary application for invasive cancer. *BMC Genomics*, **3**, 19.

Bolstad,B.M. *et al.* (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Carazzolle,M.F. *et al.* (2009) D-MaPs - DNA-microarray projects: web-based software for multi-platform microarray analysis. *Genet. Mol. Biol.*, **32**, 634–639.

Churchill,G.A. (2004) Using ANOVA to analyze microarray data. *Biotechniques,* **37**, 173–175, 177.

Gentleman,R.C. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.,* **5**, R80.

Strimmer,K. (2008) A unified approach to false discovery rate estimation. *BMC Bioinformatics*, **9**, 303.

Sturn,A. *et al.* (2002) Genesis: cluster analysis of microarray data. *Bioinformatics*, **18**, 207–208.

Wang,X. *et al.* (2006) NMPP: a user-customized NimbleGen microarray data processing pipeline. *Bioinformatics,* **22**, 2955–2957.