AMDA 2.10, an automated cross-platform pipeline for Affymetrix and illumina data analysis

Kapetis D[1], Vitulli F[1], Tuana G[1], Pelizzola M[2], Ricciardi-Castagnoli P[3], Foti M[4], Zolezzi F[1]

Motivation

Expression profiling using microarrays has become a widely used method for the study of gene-expression patterns. We published in 2006, the Automated Microarray Data Analysis (AMDA, version 2.3.5) pipeline [1] that was developed under R version 2.4 and Bioconductor 1.9 release, capable to process Affymetrix 3' IVT arrays. The availability of newer technologies such as Affymetrix GeneChips® Gene/Exon 1.0 ST as well as the cheaper Illumina Bead array technology prompt us to develop a multi-platform version of AMDA: AMDA 2.10 (Figure 1). The upgraded AMDA has been implemented in R-2.10 for Bioconductor 2.5. Furthermore, to improve the understanding of the biological data, differentially expressed genes (DEGs) have been mapped into the KEGG pathways.

Methods

Packages implemented in AMDA 2.10: eXpression Profiling System (XPS) [2], affyPLM [2], genefilter [2], Lumi [3] and Aroma Affymetrix [4]. Additional wrapper functions have been further implemented to include the RankProduct, Limma Paired, KEGG pathways and Alternative Splicing (AS) algorithm: Splice Index (SI) [5]; an AS robust prediction method based on entropy (ARH) [6]; Pattern-Based Correlation (PAC) [8]; two ANOVA

Results

The following features have been implemented in AMDA 2.10: Quality Control metrics Relative Log Expression (RLE) and Normalized Un-scaled Standard Error (NUSE) are calculated and plotted. These quality controls make very easy the identification of microarrays that stand out due to poor RNA quality or failed hybridization to the microarray (not for illumina). Pre-filtering methods IQR function has been added for gene selection (not for Exon 1.0 ST arrays) to have a flexible and experimental design oriented data set filtering method. For Gene/Exon 1.0 ST

[1] Genopolis Consurtium for functional genomics, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy [2] Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, Connecticut 06520, USA [3] Singapore Immunology Network (SIgN), Biomedical Sciences Institutes, Agency for Science, Technology and Research (A*STAR), 8A Biomedical Grove, IMMUNOS, 138648, Singapore [4] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy. Corresponding address: info@genopolis.it

arrays, Presence/absence Calls of probe sets/exons are calculated by DABG using surrogate background intensities [5]. Rank Product Rank Product (RP) [2] provides a straightforward and statistically meaningful way to determine the significance of differential expression for each gene. The RP approach is powerful for both identifying biologically relevant expression changes and controlling the false discovery rate (FDR). Limma Paired Sample Function This method performs comparison between a common baseline and an experimental condition of a dataset containing paired samples. The implementation of a paired samples test makes AMDA useful also for analyzing data from clinical studies (e.g.: human lymphocytes before and after a drug treatment). AS identification algorithms Detection of AS is still a challenging effort. In order to identify reliable AS events, several different methods have been implemented. Their choice will depend on the structure of the data set and the goals of the experiment. It is also possible to use more than one of these algorithms in parallel. KEGG pathway Maps Enrichment of DEGs in KEGG gene sets and their mapping in the pathway diagram assesses the potential functional convergence of gene-signatures on basis of the KEGG pathway modules. The updated design of AMDA is a mutli-platform pipeline that allows the analysis of new Affymetrix Gene 1.0 ST and Exon 1.0 ST GeneChips® and Illumina Bead arrays. Furthermore, intends to better respond to various experimental designs that can occur in microarray experiments, and to deliver more insightful biological understanding. and up-to-date annotations.

Contact e-mail

dimos.kapetis@unimib.it

Image