

## Overview of DIANE 1.0

Software requirements:

DIANE 1.0 is a microarray analysis platform written in JSL (JMP Scripting language – a SAS product). It is spreadsheet based, is fast and efficient, and unlike EXCEL can accommodate very large file sizes with no decrease in processing speed. Since DIANE is written in the JSL language, it takes advantage of the computational power and visualization options found in JMP. Additionally, DIANE microarray analysis modules are free, while the basic [JMP](#) software on which DIANE is designed is inexpensive, technically well supported, and integrated with the SAS statistical package.

Modules:

DIANE has six modules and they operate in a similar manner; learning The first module aids in learning the other five. This enables the user to learn the program in a straightforward manner.

In each module one proceeds in a similar way. First one opens the data file. Then a box is provided to choose data columns that must be subject to calculation. A menu pops up for choices of thresholds of fold change or p-value. Having specified these choices, a click of the button operates the program. The results are appended as subsequent columns to the master file. Thus, at each stage of the analysis one file contains all the data that progressively builds the analysis.

DIANE can read data in a variety of options principally through text files or Excel sheets. The output can also be as Excel worksheets amongst a variety of choices.

### Module 1: **Data Normalization**

DIANE provides a menu of options to normalize with such as Z-normalization, log-median normalization, log transforms etc. Also it is convenient to add any user specified normalization.

### Module 2: **Principal component analysis**

This module facilitates clustering at the level of whole arrays while identifying arrays that are to be deemed outliers. PCA can be performed in 2 dimensions as well as in 3-dimensions. The user has the option of examining the clustering in a live rotating view. Output of .jpg images is provided as which can be used for publication.

### Module 3: **Generating Gene Lists**

Finds genes differentially expressed between treatment and control groups both through Z-difference formalism and ANOVA methods. As output, one receives gene lists in Excel for each comparison. Additionally, a summary sheet examines each gene for the number of distinct treatments in which it is over or under expressed. This often forms the first step in an investigator's study of one's experimental set up for it provides a comparative glance at the behavior of a gene over various treatment control pairs.

### Module 4: **Standard Gene Clustering, N-Tuple Clustering**

This module clusters genes, which are differentially expressed, via traditional schemes such as k-means clustering or hierarchical clustering. Additionally it provides options to explore new clustering methods such as binary clustering or N-tuple clustering.

Here again, one can produce cell plots for hierarchical clustering or line plot diagrams for k-means clustering which can be made parts of manuscripts without further processing.

### Module 5: **Grouping by Gene Ontology (GO) categories**

Having determined a list of differentially expressed genes between treatments, the investigator is at pains to determine their distribution into biologically relevant groups such as GO functional categories, GO molecular or cellular functions, GO Biological functions or any other user determined categories. Upon specifying the functional categories of interest, DIANE produces pie charts depicting percentage distribution of genes into categories of interest. The pie chart mode can be converted to other graphical depictions that are suitable to the investigator. Efforts are underway to link clusters of Module 4 and pie charts of this module.

### Module 6: **Web connection to gene identifiers such as GenBank, Entrez.**

Module 3 generates lists of genes in several formats in successive columns. Typically they are gene names, identifiers from recognized databases such as GenBank, Entrez, Locuslink, Unigene etc. JMP enables opening relevant web pages corresponding to the Identifiers. The user can obtain annotations and published data relevant

to the genes in module 3.

#### Module 7: **PAGE: parametric analysis of gene set enrichment (1)**

This module tests for the enrichment of groups of genes between two treatments and identifies functional groups that are significantly enhanced. This facilitates a departure from conventional microarray analysis focused on movement of individual genes and enables the user to explore functional groups of genes.

For information on obtaining **DIANE 1.0** contact:

Vinayakumar Prabhu Ph.D.

Gene Expression and Genomics Unit  
TRIAD Technology Center, Room 2-207  
333 Cassell Drive  
National Institute on Aging  
National Institutes of Health  
Baltimore, MD 21224  
TEL: 410-558-8141  
FAX: 410-558-8236  
prabhuvi@grc.nia.nih.gov  
<http://www.grc.nia.nih.gov/branches/rrb/dna/dna.htm>

#### References:

1. Kim SY, Volsky DJ.  
PAGE: parametric analysis of gene set enrichment.  
BMC Bioinformatics. 2005 Jun 8;6:144.