

DATA ANNOTATION AND PROCESSING TOOL

A USER-FRIENDLY WIZARD THAT WILL MANAGE, ANNOTATE, AND STATISTICALLY PROCESS YOUR METACORE™ DATA IN MINUTES



USE THIS METACORE TOOL FOR

- Managing gene expression data and corresponding metadata
- Annotation and organization of expression data files
- Raw microarray data processing and statistical workup
- Regulatory/signalling pathway analysis of microarray data
- Sample grouping and comparisons
- Functional analysis of clinical data

WHO CAN BENEFIT

- Systems biology researchers dealing with microarray data analyses in commercial and academic settings
- Drug discovery and diagnostics groups that deal with drug-related expression signatures
- Disease biomarker groups to elucidate key disease mechanisms and propose better biomarkers and treatment targets
- Bioinformatics departments that need a standardized way of data processing, annotation, and shared research solutions

DATA ANNOTATION AND PROCESSING TOOL OVERVIEW

This module is a part of MetaCore, our manually-curated database and software suite for functional analysis of Omics data. It is designed for anyone to quickly obtain the most significant and valuable components of the data, organize and statistically process the files, and use filters, workflows or enrichments that are based on the standard vocabulary and ontology from MetaCore.

The wizard allows fast and easy processing and includes intuitive tools for data visualization, analysis, and enrichments, filters, and easy to use workflows. Moreover, user metadata and analysis settings get captured and stored along with each sample's properties upon the completion of the wizard. This provides a unique and standardized solution for anyone dealing with functional or pathway analysis of microarray data.

The screenshot displays the 'Data Ranger 1.0' software interface. The main window is titled 'Data Processing: Define data layout'. It features a table with columns for 'ID1', 'Affymetrix tag', 'Intensity', and 'GeneGO Pathway Maps'. The 'Intensity' column contains values for various GSM IDs (e.g., GSM326901, GSM326902). A 'Signal distribution' plot is overlaid on the table, showing a histogram of gene expression levels with a threshold line at 1.4. Below the table, there are sections for 'Factors' (Sensitivity, Case/Control, Disease, Drug Dose, Tissue) and 'Values' (sensitive, resistant). A list of enriched pathways is shown on the right, including 'Cytokinesis, remodeling, TGF- β 1 and cytokinetic remodeling' and 'Cell adhesion, Chemotaxis and adhesion'. A yellow callout box in the top right corner of the screenshot reads: 'Get to what's really important in just 10 minutes!'.



HARDWARE REQUIREMENTS**CLIENT (FOR WEB PORTAL AND IN HOUSE INSTALLATIONS):**

- Mac (PPC or Intel) or Windows P4
- 2GHz CPU and 1GB RAM recommended
- Web browser: Internet Explorer version 6.0, Firefox 3.0, Safari 3.0.4 or higher
- Adobe Flash Player 8 or higher
- Java Runtime Environment (JRE) 1.5.0

SERVER (FOR IN-HOUSE INSTALLATIONS):

- 2 or more P4/XEON CPU's with 4GB of RAM recommended
- 3.2 GHz CPU and higher recommended
- SCSI HDD with minimum of 250GB of storage is recommended
- RAID recommended
- RedHat Enterprise Linux 4 or 5.1; SuSE 10; CentOS 4/5
- X development package installed
- Oracle 10.2 DBMS and client tools
- MetaCore supports x86-64bit architecture

FEATURE HIGHLIGHTS

From what the data "is" to what the data "means," this tool can quickly take you through any data forest and deliver valuable knowledge:

- Start with raw or processed data (Excel, .txt, .csv, .CEL format) or a .soft file from GEO
- Breeze through annotation and file organization steps with our controlled vocabulary terms or add your own
- Normalize and statistically process the data with RMA, FDR, filtering, DEG, and GSEA using GeneGo functional ontologies
- Get your data with all the analysis settings and metadata factors saved in the Data Manager and stored within experiment properties
- OR launch a MetaCore analysis workflow immediately at the last step with a comprehensive report summary

PARSER

Auto-recognition and semi-automated processing of Excel, .txt, .csv, or .CEL format or .soft files from GEO

STATISTICAL FILTERS

- Variable threshold filtering
- Filter by selecting top 300 or top "n" genes from each data sample
- FDR threshold filtering

FUNCTIONAL FILTERS

- GO processes
- Disease
- Tissue
- Sub cellular localization
- Interaction types

- Object types
- Metabolites
- Body fluids
- Species/orthologs (human, mouse, rat, dog, chicken, worm, chimpanzee, bovine, zebrafish, fly, mosquito, mold, bread mold, baker's yeast, fission yeast, arabidopsis, rice, candida)

ONTOLOGIES THAT CAN BE USED FOR GSEA

- Canonical Pathway Maps
- GeneGo:
 - Process Networks
 - Metabolic Networks
 - Disease Networks
 - Drug Target Networks
 - Toxicity Networks
 - Diseases
- GO:
 - Processes
 - Localization
 - Molecular Function

AUTOMATED WORKFLOWS WITH REPORTS

The last step of the wizard provides an option to launch a universal analysis workflow, which would provide functional analysis of the data with enrichments and ranking, as well as custom network building that captures the top interconnected modules for the dataset.

CONTENT

The Data Annotation & Processing Tool uses ontologies from MetaBase, which is a manually-curated proprietary database built by GeneGo PhD and M.D. employees reading full text articles that cover 2700 journals and dating back to the 60's.

THOMSON REUTERS REGIONAL OFFICES**North America**

Philadelphia +1 800 336 4474
+1 215 386 0100

Latin America

+55 11 8370 9845

Europe, Middle East and Africa

Barcelona +34 93 459 2220
London +44 20 7433 4000

Asia Pacific

Singapore +65 6775 5088
Tokyo +81 3 5218 6500

Contact us to find out more about MetaCore and the Data Annotation & Processing Tool or visit thomsonreuters.com/diseaseinsight

ABOUT METACORE

MetaCore™ is a Thomson Reuters's flagship data analysis platform, which integrates a comprehensive manually-curated database of human biology and a sophisticated systems biology toolbox. MetaCore is designed for functional analysis of microarray, re-sequencing, SNP, metabolic, proteomics and other OMICs data and is applied for validation, and/or developing of new hypotheses for target selection and biomarker research for drug discovery, translational research, and personalized medicine.

For a complete office list visit:

science.thomsonreuters.com/contact

