



JMP Genomics

Version 3.1

Release Notes

"Creativity involves breaking out of established patterns in order to look at things in a different way." Edward de Bono



JMP. A Business Unit of SAS
SAS Campus Drive
Cary, NC 27513
www.jmp.com

Release Notes for JMP Genomics 3.1

Copyright ©2007, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

JMP[®], SAS[®] and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

This document describes changes and enhancements from JMP Genomics 3.0 to the release of JMP Genomics 3.1. JMP Genomics 3.1 will be available as a web download for existing or new customers around October 31, 2007.

New and improved features in JMP Genomics Analytical Processes (APs) are described in the following sections. Changes to specific analytical processes are organized according to the JMP Genomics main menu.

General Features

Menu changes

- The name of the **Data Set Creation** submenu has been changed to **Import**.
- A new **Copy Number** submenu has been added to the main **Genomics** menu. This section lists processes specific to analyses of copy number and loss of heterozygosity (LOH) data.

Dialog changes

- The *Where clause* option found on many of the dialogs has been changed to *Filter to Include Observations where* in order to eliminate confusion with regards to the function of this option. Improved help for this filtering statement helps clarify how users should specify observations to keep in the analysis using this statement. Please click on the ? buttons next to the Filter statement in the dialog for more details.

Experimental Design and Data Sets

Import

Affymetrix Expression CHP Wizard *New!*

- A JSL script that assists import of expression data contained in Affymetrix's CHP files into JMP Genomics and automatically creates a workflow for analyzing that data
- Highlights of the workflow include automatic import of design information from ARR files or text files, import of expression CHP files, selection of important design variables, and download of NetAffx annotation information if desired
- Results are captured as scripts in a journal, and may be launched to review results.

Download NetAffx Files *New!*

- A JSL script that facilitates your identification and download of annotation, library, map, or other accessory files from NetAffx.

Affymetrix SNP CHP Input Engine

- Supports import of large sets of SNPv5 and SNPv6 CHP files

Affymetrix Expression CEL Input Engine

- Updates fix a bug which caused incorrect import of 93 control probe sets for Human exon array.
- Improved import speed for CEL data sets, particularly for large data sets.

Affymetrix SNP CEL Input Engine *New!*

- Imports CEL files from Affymetrix SNP GeneChips, including SNPv5 and SNPv6. Data can be imported at the probe level, or users may apply a number of options for transformation, background correction, normalization and summarization of probe level data. A SNP annotation data set is required to merge in chromosome and position information into the output dataset, and users of SNPv5 and SNPv6 may also specify a copy number annotation data set to merge in chromosome and position information for copy number probes. Several datasets are produced, including a `_snp` dataset where intensity information for SNP probe sets is summarized at the allele level and a `_cn` dataset which contains summarized CN probesets and intensity information for SNP probesets summarized to the SNP level.

Affymetrix CNAT *New!*

- Imports output files from Affy CNAT 4.0 into a single, tall, SAS data set.

Illumina SNP Input Engine

- Added ability to import map and sample files and increased selection of optional settings.

Illumina Copy Number Input Engine *New!*

- Imports BeadStudio final report files generated from the Report Wizard into a single, tall, SAS data set

Import Individual Text, CSV or Excel File

- Now imports very wide text files up to 1 million columns. Users may also specify types and lengths of variables as desired, and specify number of columns to scan to automatically determine variable attributes

Data Set Utilities

Statistics for Rows

- Added box for optional specification of a grouping variable to allow computation of statistics across groups of rows.

Genetics

Genetic Data Set Utilities

Check Data Contents *New!*

- Allows you to view a subset of a wide data set that would otherwise be too large to view in JMP.

Genetic Marker Statistics

Marker Properties

- A new option for creating a data set with numeric genotypes with or without an accompanying cell plot has been added.

Linkage Disequilibrium

- Improved handling of data sets with large numbers of markers.
- A checkbox for performing LD calculations on all pairs of markers has been added.

LD tagSNP Selection

- Improved handling of data sets with large numbers of markers.
- A checkbox for performing LD calculations on all pairs of markers has been added.

Association Testing

Case-Control Association

- Additional association tests have been added. These include: Dominant, Recessive, and MAX.
- The Dominant and Recessive tests are the Cochran-Armitage trend tests using dominant and recessive codings for genotypes.
- The MAX test uses the maxima of the dominant, recessive, and trend statistics in simulations to estimate p -values.

PCA for Population Stratification *New!*

- Performs the trend test while correcting for population stratification using principal components via the EIGENSTRAT method.

TDT

- The ability to use wide data sets has been added.

Haplotype Analysis

Haplotype Estimation

- The ability to accommodate wide data sets has been added.

htSNP Selection

- An option for creating an output data set containing an htSNP indicator variable has been added.

Multiple Genetic Analytical Processes

- The ability to accommodate numeric (0, 1, and 2) genotypes in the input data set has been added.

Copy Number *New!*

Bin *New!*

- Allows you to combine multiple probesets or SNPs into groups to reduce the total number of observations in a very large data set to a manageable number suitable for preliminary analysis.

Bivariate One-Way ANOVA *New!*

- Uses paired, allele-level intensity observations to simultaneously test for differences in copy number and allele frequency between groups
- To prevent *out of memory* errors for large data sets, a check box enabling you to turn off clustering of significant means profiles and the subsequent subset table generation has been added to the Options tab.
- Positive FDR (pFDR) has been added as an option to the methods for multiple testing.

Quality Control & Normalization

Quality Control

Distribution Analysis

- The default bandwidth setting was changed to improve performance when working with large data sets. As a result, appearance of curves generated using default settings may appear smoother than those generated with earlier versions of JMP Genomics.
- To revert to prior settings, adjust the bandwidth on the Options tab.

Correlation and Principal Components

- New option for centering the rows has been added to the Analysis tab. This option subtracts the mean from each row before doing PCA, which greatly reduces the percentage of variation explained by the first principal component. The check box for this option is selected by default. Uncheck the box if you prefer to retain the first component.
- Improved labeling of axes in the 3D plots generated by this AP include percentage of variation explained by each component.

Filter Intensities *New!*

- Allows you to pre-filter data sets by row or column statistics prior to normalization to remove poor-performing probesets.

Microarray Analysis

Pattern Discovery

Hierarchical Clustering

- When grouping variables are specified from the design file, the right side cluster dendrogram is colored. Different levels of the design variable are colored with different colors. This option is not available when a *By Variable* is selected during clustering.

Row-by-Row Modeling

One-Way ANOVA

- An option for specifying chromosome and position variables has been added to the **General** tab. This option allows you to generate plots of differences by chromosome and position.
- To prevent *out of memory* errors for large data sets, a check box enabling you to turn off clustering of significant means profiles and the subsequent subset table generation has been added to the **Options** tab.
- Positive FDR (pFDR) has been added as an option to the methods for multiple testing

ANOVA

- An option for specifying chromosome and position variables has been added to the **General** tab. This option allows you to generate plots of differences by chromosome and position.
- To prevent *out of memory* errors for large data sets, a check box enabling you to turn off both LSMeans clustering and the subsequent subset table generation has been added to the **LSMeans** tab.
- Positive FDR (pFDR) has been added as an option to the methods for multiple testing

Mixed Model Analysis

- An option for specifying chromosome and position variables has been added to the **General** tab. This option allows you to generate plots of differences by chromosome and position.
- Positive FDR (pFDR) has been added as an option to the methods for multiple testing.

P-Value Quantile Plotter *New!*

- JSL Script that allows you to plot a column of p -values from a JMP table against the expected p -values from a uniform distribution.

Predictive Modeling

Distance Scoring *New!*

- This AP generates predictions for an observation based on the distance between that observation and observations in a training set.
- A wide variety of distance metrics are available.

Cross Validation Model Comparison

- Several new methods for cross-validation have been added to the analysis tab.

Test Set Model Comparison *New!*

- This AP evaluates one or more test sets, with predictive model setting that you select, and then compares the results.

All Predictive Modeling APs

- The Predictor Reduction tab has been split into two tabs (Predictor Reduction 1 and Predictor Reduction 2).
- New filtering, standardization, and testing options have been added to the predictor reduction tabs to allow for further reduction in the number of predictors.

Annotation and Power

Annotation Analysis

Venn Diagram *New!*

- Creates up to five-way clickable Venn Diagrams using 0-1 variable columns from a SAS dataset, such as the significance indices generated by ANOVA, **Mixed Model**, **One-Way ANOVA**, or **Bivariate One-Way ANOVA** APs.
- Users may also create their own 0-1 variables from selected rows in an open table by using the **Data Filter** process found under the **Rows** menu or other selection options, and then choosing the **Create 0-1 Indicator from Selected Rows** AP under the **Annotation** submenu.

IPA Upload

- The dialog has been modified to resemble the dialog in the IPA Java tool. Uploading data through JMP Genomics 3.1 is similar to uploading data through the IPA web interface.
- All expression types supported by IPA are now supported by JMP Genomics.
- The IPA entry point is now configurable. Users may choose whichever entry point best suit their needs

Affymetrix Download NetAffx Files *New!*

- A JSL script that facilitates your identification and download of annotation, library, map, or other accessory files from NetAffx.

Column Enrichment

- A PAGE test (parametric analysis of gene enrichment) which performs analysis of p -values and differences has been added.
- The ability to specify a secondary delimiter has been added to the General tab. This secondary delimiter is needed for parsing some files (e.g., Affymetrix annotation files) that contain multiple delimiters per column.

Other Processes**Transform**

- Found under JMP's Table menu
- A 1/EXP function has been added to allow for easy conversion back to p -values from log p -values or log₁₀ p -values.