



# JMP Genomics

Version 3.2

# 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](http://www.jmp.com)

---

**Release Notes for JMP Genomics 3.2**

Copyright ©2008, 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<sup>®</sup>, SAS<sup>®</sup> 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 3.1 to the release of JMP Genomics 3.2. JMP Genomics 3.2 will be available as a web download for existing or new customers around April 2008.

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

- A new **Workflows** submenu has been added to the main **Genomics** menu. This section lists basic workflows used for input and analysis of genetic, expression, copy number, and exon data. The **Workflow Builder** has been moved to this folder from the main **Genomics** menu.
- A new **QTL Mapping** submenu has been added to the main **Genomics** menu. This section lists processes specific to analyses of quantitative trait loci.
- A **Documentation and Help** submenu has been added to the bottom of the main **Genomics** menu. This section provides links to the *JMP Genomics User Guide* and *Supplement* as well as SAS documentation and various JMP Genomics web links. In addition, a mechanism for requesting features and reporting problems has been added to this section.

### SAS Message Window changes

- A new button has been added to the SAS Message window that is produced as part of the output of many JMP Genomics processes. If an output data set is larger than 1 GB in size, a **View Subset** button appears in place of the normal **Open** button. Clicking this new button launches and runs the **Check Data Contents** AP behind the scenes, allowing you to view a subset of a SAS data set. This process is a useful and quick method for viewing extremely wide or tall data sets, perhaps to verify that they were imported correctly.

## Experimental Design and Data Sets

### Experimental Design

#### Affymetrix Experimental Design Wizard *New!*

- An interactive wizard that assists in creating a design file for a set of CEL or CHP files.
- The wizard can create a design file from paired ARR and data files or an existing text or Excel file that lists names of the data files,

---

## Import

### Affymetrix Expression CEL Input Engine

- Updates greatly improve performance for quantile and RMA normalization.
- Added background correction for GCRMA normalization.
- Output includes a window with Action buttons to launch either the Basic Expression Workflow or Basic Exon Workflow, using the imported data.

### Affymetrix Expression CHP Input Engine

- Added option to output detection calls from MASS5-normalized CHP files as a separate data set. This data set may be filtered directly, and/or merged with the main data table. The percentage of present calls for each row is now included in the main data table.
- Output includes a window with Action buttons to launch either the Basic Expression Workflow or Basic Exon Workflow, using the imported data.

### Affymetrix SNP CEL Input Engine

- Output includes a window with Action buttons to launch the Basic Copy Number Workflow, using the imported data.

### Affymetrix SNP CHP Input Engine

- An option to remove the AFFX control SNPs from the output data sets has been added.
- Output includes a window with Action buttons to launch the Basic Genetics Workflow, using the imported data.

### Affymetrix CN CHP Input Engine *New!*

- Imports and combines Affymetrix CN CHP files into a single SAS data set.
- Output includes a window with Action buttons to launch the Basic Copy Number Workflow, using the imported data.

### Illumina Expression Input Engine

- This input engine has been extensively modified. Modifications include the ability to select which columns to import, an option to apply *Log2* transformation to intensities, options to filter intensities based on detection *p*-values, and options for annotation data.
- Output includes a window with Action buttons to launch the Basic Expression Workflow, using the imported data.

### Illumina SNP Input Engine

- Output includes a window with Action buttons to launch the Basic Genetics Workflow, using the imported data.

**Illumina Copy Number Input Engine**

- Output includes a window with Action buttons to launch the Basic Copy Number Workflow, using the imported data.

**WinQTLCart Input Engine *New!***

- This new process allows you to import a Windows QTL Cartographer .MCD file into a SAS data set containing the marker genotypes and traits and a SAS data set containing annotation (marker map) information.

**Data Set Utilities****Data Stack *New!***

- Converts tall SAS data set and information from an experimental design file into a single stacked data set.
- Includes an option for subsetting the data, using the first  $n$  rows, for preliminary testing.
- Ideal for creating a data set for use with SAS PROC MIXED.

**Workflows *New!*****Basic Expression Workflow *New!***

- A basic workflow used for analysis of expression data files.
- A single dialog that allows you to select input expression data sets, perform quality control before and/or after normalization of the data, carry out basic analysis and upload significant results to online data bases.

**Basic Genetics Workflow *New!***

- A basic workflow used for analysis of genetics data files.
- This automated process uses the Marker Properties, Missing Genotype by Trait Summary, Subset and/or Reorder Genetics Data, and Case-Control Association APs to perform basic analyses of genetics data sets.

**Basic Copy Number Workflow *New!***

- A basic workflow used for analysis of copy number data files
- A single dialog that allows you to select input copy number data sets, perform quality control before and/or after normalization of the data, carry out one-way ANOVA analysis and upload significant results to online data bases.

**Basic Exon Workflow *New!***

- A basic workflow used for analysis of exon data files
- A single dialog that allows you to select input exon expression data sets, perform quality control before and/or after normalization of the data, carry out basic analysis and upload significant results to online data bases.

---

## Genetics

### Genetic Data Set Utilities

#### Subset/Reorder Genetics Data

- An option allowing users to specify the minimum proportion of non-missing genotypes has been added to the **General** tab.
- An option allowing you to filter rows from the annotation data set only, prior to filtering marker variables from both data sets, has been added to the **Annotation** tab.
- Major performance enhancements have been made to reduce computation time and use less memory.

#### Recode Genotypes

- Major performance enhancements have been made to reduce computation time and use less memory.

#### Kinship Matrix *New!*

- This AP creates a matrix containing either kinship (coancestry) coefficients or covariance coefficients (coefficients of relationship) between pairs of related individuals. These coefficients can then be used as random effects in order to analyze family data in an association setting, using processes, such as **Marker-Trait Association** or **SNP-Trait Association**, for example, that can accommodate random effects.

### Genetic Marker Statistics

#### Marker Properties

- Default settings modified to include two additional plots: distributions of minor allele frequency (MAF plots) and missing genotype proportion plots.
- A separate set of plots is generated for each annotation group.
- Major performance enhancements have been made to reduce computation time and use less memory.

#### Missing Genotype by Trait Summary *New!*

- This process tests whether missing genotypes at a particular marker are related to a trait.

### Association Testing

#### Case-Control Association

- A new checkbox, allowing you to specify when all markers are biallelic, has been added to the **Options** tab. When this option is specified and additional criteria are met, a faster algorithm is run, substantially reducing computation time.
- Major performance enhancements have been made to reduce computation time and use less memory.

### PCA for Population Stratification

- A time-saving PCA Data Set option, which allows users to rerun analyses without having to repeat the principal component calculations, has been added to the Options tab. To rerun the PCA without having to repeat the principal component calculations, a user specifies the `_pca` data set output from a previous run. Note: this option is only applicable for runs in which the number of principal components tested is not greater than the initial run.

### Marker-Trait Association

- A field for specifying options for the PROC MIXED RANDOM statement, when random effects are included in the model, has been added to the Model Variables tab.

### SNP-Trait Association

- A field for specifying options for the PROC MIXED RANDOM statement, when random effects are included in the model, has been added to the Model Variables tab.

### TDT

- A new checkbox allowing the user to display the cell plot of Mendelian errors has been added to the Options tab. If there are any families with Mendelian errors at any of the markers, a cell plot with colors representing whether or not there is a Mendelian error at each marker for each family is displayed when this option is checked. When this option is not checked, a SAS Output window containing this information is automatically displayed in JMP.
- Major performance enhancements have been made to reduce computation time and use less memory

## QTL Mapping *New!*

### Single Marker Analysis *New!*

- This new process performs a simple regression for each marker with trait values and computes the probability of QTL evidence for each marker, providing you with a way to quickly scan the whole genome for evidence of QTL signals.
- Please note that this AP should be considered *experimental*.

### Build Genotype Probability Data Set *New!*

- This new process builds a genotype probability SAS data set. The GP (Genotype Probability) data set includes probabilities of a certain QTL genotype in each testing location (1 cM apart) for all individuals.
- The GP data set has integrated the experimental design information into its values and can be used in the IM and CIM Analysis process.
- Please note that this AP should be considered *experimental*.

### IM and CIM Analysis (Single QTL Model) *New!*

- This new process generates either interval or composite interval maps of quantitative trait loci.
- Interval Mapping (IM) is an extension of single marker analysis and it can do analysis to any genetic location flanked by two markers. IM uses complete marker linkage maps for genomic scanning (LR or LOD profile) of a QTL.
- Composite Interval Mapping (CIM) is an extension of IM analysis that uses a combination of IM with multiple regressions. CIM can improve the precision and efficiency of mapping multiple QTLs.
- Please note that this AP should be considered *experimental*.

## Haplotype Analysis

### Haplotype Estimation

- The default output now includes buttons to launch either the Haplotype Trend Regression (if the option of creating a Phase Assignment Data Set is selected) and/or htSNP Selection (if the option for creating a data set containing haplotype frequency estimates is selected) processes.

### Haplotype Trend Regression

- Analyses are now automatically performed by haplotype window. There is no longer a need to specify the Window variable as a By Variable.
- The name of the Filter to Include Observations field has been changed to Filter to Include Windows.

## Copy Number

### Copy Number Partitioning *New!*

- This AP uses recursive partitioning to find potential breakpoints of copy number changes across a chromosome.

## Quality Control & Normalization

## Quality Control

### Distribution Analysis

- Default output now includes buttons to launch either the Data Standardize or Correlation and Principal Components APs.

### Correlation and Principal Components

- In prior releases this process generated an .rtf file illustrating the partitioning of the variance components in a pie chart, by default. In this version, a bar graph is generated instead, but only if a name for the .rtf file is specified on the Options tab.
- Default output now includes a button to launch the Correlation and Grouped Scatterplots AP

## Normalization

### Ratio Analysis

- The ability to segregate rows into distinct groups defined by a **By Variable** has been added. Distinct Loess curves are then generated and fit to each to the rows in each group so that print-tip Loess normalization may be performed.

## Microarray Analysis

### Pattern Discovery

#### Hierarchical Clustering

- Options for centering and/or scaling the rows before analysis have been added.

#### K-Means Clustering

- Instead of having to specify a radius before the analysis, the **Automated Radius K-Means** method for joining clusters now allows you to select a correlation value that is used automatically to determining a minimum radius.
- Options for centering and/or scaling the rows before analysis have been added.

#### Distance Matrix

- An Action button to launch the **Multidimensional Scaling AP** has been added to the default output.

#### Multidimensional Scaling

- Action buttons to plot residuals and display the clustered heat map and 2D and 3D scores have been added to the default output.

#### Partial Correlation Diagram *New!*

- This new AP helps you infer potential causal relationships between a set of variables by plotting each variable as a node and connecting the nodes with line segments.

---

## Row-by-Row Modeling

### One-Way ANOVA

- A suite of new Action buttons has been added to the output dialogs. Options include the ability to subset the data in either wide or tall format, construct one-way plots for selected rows, and to fit models to the input data for selected rows and then plot the LS Means (using the JMP Fit Model platform) for that data.
- An option enabling the use of estimate statements has been added.

### ANOVA

- A suite of new Action buttons has been added to the output dialogs. Options include the ability to subset the data in either wide or tall format, construct plots for selected rows, and to fit models to the input data for selected rows and then plot the LS Means (using the JMP Fit Model platform) for that data.
- An option for including untransformed  $p$ -values and exponentiated differences as part of the output data set has been added.
- An option to show simple differences has been added to the LS Means tab. Simple differences are a subset of All Pairwise Differences for which only one level changes, and pertain only to LSMeans involving more than one factor.
- A new **Baseline Covariate** tab has been added. The parameters on this tab allow you to specify a subset of responses to be a baseline covariate. For example, in a dose-response study, you can specify the dose 0 values to be covariates instead of modeled responses.

### Mixed Model Analysis

- An option for including untransformed  $p$ -values and exponentiated differences as part of the output data set has been added.
- An option to show simple differences has been added to the LS Means tab. Simple differences are a subset of All Pairwise Differences for which only one level changes, and pertain only to LSMeans involving more than one factor.
- A new **Baseline Covariate** tab has been added. The parameters on this tab allow you to specify a subset of responses to be a baseline covariate. For example, in a dose-response study, you can specify the dose 0 values to be covariates instead of modeled responses.

### Survival Analysis *New!*

- Survival Analysis tests association of each row of the input data set with a censored response. It fits a Cox proportional hazards model on a row-by-row basis to a normalized data set.

## Predictive Modeling

### Learning Curves *New!*

- This AP uses cross-validation to evaluate a model using different sample sizes, thereby revealing the influence of sample size on accuracy and variability of the model.
- A wide variety of distance metrics are available for making the predictions.

## Annotation and Power

### Annotation Analysis

#### GEO Submission Tool *New!*

- This new process helps the user format an experiment for submission to the Gene Expression Omnibus (GEO) database. Data is formatted in MiniML format and written to an .xml file for batch submission to GEO.

## 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 log<sub>10</sub>  $p$ -values.