



# JMP Genomics

Version 4.0

# 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 4.0**

Copyright ©2009, 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 3.2 to JMP Genomics 4.0. 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

### SAS and JMP

- The SAS and JMP components underlying JMP Genomics 4 have been updated to SAS 9.2 and JMP 8.0.1.
- As a result of these changes, the installation process for JMP Genomics 4 differs from that of prior versions. JMP Genomics 4.0 is an integrated SAS installer, and is available from SAS by Electronic Software Delivery.

### Configure Genomics Settings *New!*

- This new process has been added to the **File** menu.
- This process allows you to specify the default settings for a variety of parameters. These parameters include:
  - Optional specification of default folders for input, output, annotation, and settings.
  - The archiving of all of the SAS code generated by JMP Genomics. The resulting code will be placed into your output directory.
  - Specification of the proxy server and proxy port, as needed.

### Setting Personal Settings (My Default Settings)

- A new **Set As Default** button in most JMP Genomics dialogs allows you to save a set of dialog settings as your own personal default for that dialog.
- For each user, default settings can be found in the directory C:\Documents and Settings\*userid*\Local Settings\Application Data\JMPG\MyDefaultSettings. Saved JMP Genomics sample settings may be accessed by clicking on **Example Settings** in the Load Settings dialog.

### Compliance with FDCC Security Guidelines

- JMP Genomics now complies with federal government FDCC security settings, which prevent limited-access users from writing files into all directories under C:/Program Files.

### Settings Folders

- A personal settings folder has been added to the JMP Genomics installation, in addition to the original Settings folder that ships with JMP Genomics and typically installs in the C:\Program Files\SAS\JMP\8\Genomics directory. The personal Settings folder is placed in the C:\Documents and Settings\*userid*\Local

---

Settings\Application Data\JMPG directory during installation. By default, this second folder is where any user specified settings are saved. To access settings, click **Load...** in any dialog to open the Load Settings window. Click **Example Settings** to choose the original Settings folder containing the sample settings. Click **My Settings** to choose your personal Settings folder.

- *Note:* A separate personal settings folder is created for each user.

#### Dialog changes

- AP-specific icons are now displayed in the upper left hand corner of windows. The corresponding icons are also displayed in the JMP Window List (accessible from **View > Window List** in the JMP menu), making it easier to locate open AP windows by their type and also to distinguish which output types are associated with which output tables.

#### Running Processes Window *New!*

- This window appears when you click **Run** on any AP.
- This window indicates the process that is running and the time it started. A **Stop** button, which allows you to halt the process, has been added.

### Experimental Design and Data Sets

#### Experimental Design

##### Experimental Design File Column Parser *New!*

- This new utility assists in parsing columns of delimited data which appear in a column or columns of an experimental design file.
- Use this utility to select columns of delimited data for parsing. Delimited entries are then separated into individual columns in the output experimental design data set.

#### Import

##### Affymetrix Expression CEL Input Engine

- A new Probe Filter tab has been added. This tab allows you to select a SAS data set containing a list of probes to be dropped from the final data set. For exon and whole transcript arrays, only probe identifiers are required. For 3' expression arrays, the data set must contain two columns; one containing a probe set identifier and the other containing the corresponding probe. You may choose to filter out specified probes prior to background correction, normalization, or summary.
- Performance for GCRMA background correction for large data sets has been improved.

---

### Affymetrix Exon and Whole Transcript Expression CEL Input Engine

#### *New!*

- This new process is specifically tailored for import of exon and whole transcript CEL files.
- This process also allows users to filter probes during import, and choose to retain orphan probes not belonging to probesets, and single probesets not associated with specific transcript cluster identifiers.
- Please note that although this process is similar to the Affymetrix Expression CEL Input Engine, default settings are optimized for exon and whole transcript arrays. Also, certain options (e.g., GCRMA background correction) are not available for exon or whole transcript arrays in JMP Genomics at this time.

### Affymetrix Annotation CSV Files *New!*

- This new process imports comma-separated annotation files from Affymetrix NetAffx into SAS data sets which can be specified in other JMP Genomics processes.

### BioConductor Expresso for Affymetrix

- The code has been updated so that the wrapper can now handle calling GCRMA with BioC 2.3 and 2.2 files as well as 2.1 files.

### Affymetrix SNP CHP Input Engine

- A new option has been added to code genotypes numerically as specified by the Allele Frequencies column in the Affymetrix SNP annotation data set.

### Illumina SNP Input Engine

- An option for importing a JMP G custom report (generated using the JMP G Custom report plug-in) has been added.
- An option for specifying the type of column delimiter in the input files has been added.

### Illumina Copy Number Input Engine

- An option for specifying the type of column delimiter in the input files has been added.

### PLINK Input Engine *New!*

- This new process imports a set of PLINK files into two SAS data sets. The first data set contains the marker genotypes and traits. The second data set contain annotation (marker and map) information.

## Data Set Utilities

### Reorder

- Multiple improvements to the performance of this AP have been made.

### Combine Columns *New!*

- This new process groups together columns of a tall data set and computes a summary statistic, such as the mean for each group, and creates an accompanying reduced experimental design data set (EDDS).
- This process is useful when you want to summarize data from similar columns of data (e.g., replicate runs of the same experiment) into a single column.

### Transform

- A new option has been added to allow the user to specify a shifting factor prior to transformation. This option can be used when *log*-transforming summarized count data derived from next generation sequencing platforms or other platforms where zero values may occur.

## Workflows

### Basic Genetics Workflow

- A filter to include specific markers based on annotation information has been added to the **Annotation** tab.
- A new **Subsetting** tab has been added that allows you to specify a specific number of SNPs to test, or to select individuals to include in an analysis based on values of any of the variables in the input data set.
- The workflow can now accept numerically-encoded genotypes.

### Basic Copy Number Workflow

- An option to create a journal with buttons to launch results separately for each chromosome, instead of displaying all results at once, has been added to the **One-Way ANOVA** tab. This option is useful for very large output data sets.
- A new summary plot has been added that displays the average significance index across all comparisons. The significance index for each comparison is 1 (significant) or 0 (non-significant) depending on whether the associated difference has a *p*-value exceeding the specified threshold. This summary plot was added upon request to assist in identifying shared regions of copy number differences. It is especially useful when single individuals are being compared against a control group.
- A new **EXON** tab allows you to select an exon or probeset-level annotation file with position information. If you specify chromosome and position information from this file, clicking on the **Plot Oneway Means by Chromosome and Position** Action Button overlays exon locations as bands on the mean plot.
- The **Anno 1** and **Anno 2** tabs in this workflow have been renamed **CN Anno 1** and **CN Anno 2**.
- A new **Multiple Testing** tab has been added to allow users to select a multiple testing method and *p*-value threshold.

- A new **Tracks** tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

#### Basic Expression Workflow

- A new **Multiple Testing** tab has been added to allow users to select a multiple testing method and  $p$ -value threshold.
- A new **Tracks** tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

#### Basic Exon Workflow

- A filter to include Annotation Rows for ANOVAs has been added to the **Anno 1** tab. Results are only computed for annotated rows.
- A new **Exon** tab allows you to select an exon or probeset-level annotation file with position information. If you specify chromosome and position information from this file, clicking on the **Plot Oneway Means by Chromosome and Position** Action Button overlays exon locations as bands on the mean plot.
- The **Anno 1** and **Anno 2** tabs in this workflow have been renamed **Tx Anno 1** and **Tx Anno 2**.
- A new **Multiple Testing** tab has been added to allow users to select a multiple testing method and  $p$ -value threshold.  
A new **Tracks** tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

#### Expression Quality Control Workflow *New!*

- A basic workflow used to examine and prepare expression data files for further analysis.
- Includes options for filtering (including group-wise filtering), quality control, and normalization of the data.

#### Expression Statistics Workflow *New!*

- A basic workflow used to compute statistics for expression data files.
- Includes options for clustering, ANOVA, and annotation.

### Genetics

#### All processes containing a **P-Value Plots** tab

- An option for separating and journaling results by annotation group (by chromosome, for example) has been added. Selecting this option results in the creation of a JMP journal containing **Launch** buttons specific for the results for each group. This option is useful for viewing results from individual groups instead of the full output data set (which may be very large).

---

## Genetic Data Set Utilities

### Check Data Contents

- Results include the number of SNPs and/or individuals filtered.
- An option for outputting the subset columns and rows to a separate SAS data set has been added.

### Subset and Reorder Genetic Data

- A field for specifying the variable containing the marker names has been added to the Annotation tab.
- Output from this process now includes a report detailing the number of markers and individuals (rows) failing to meet filtering criteria specified on the Annotation tab and, thus removed from the genotype data set.

### Recode Genotypes

- Multiple improvements to the performance of this AP have been made.

### Recode Missing Genotypes *New!*

- This new process creates a data set that replaces the current value that represents missing genotypes or alleles with the value that is required by the Genetics processes: blank for character genotypes or alleles and (.) for numeric genotypes or alleles.

### Population Genetic Distance Matrix *New!*

- This experimental new process outputs a symmetric matrix of dissimilarities between specified groups within a study. The resulting matrix can be used as input to Multidimensional Scaling.
- Genetic distance computation is based on differences in group allele frequencies, with groups specified by the input values of the Population Variable.
- Multiple Genetic Distance matrices can be output for separate sets of markers if an Annotation By Group variable is specified.

## Genetic Marker Statistics

### Marker Properties

- Multiple improvements to the performance of this AP have been made.

### Linkage Disequilibrium

- A new option to use the location variable for computing distances used to calculate linkage disequilibrium has been added to the Options tab.
- If the Estimate Rho and K\_Rho option is checked on the Options tab, output now includes a window enabling you to launch the Malecot LD Map AP.

- A new interactive plot<sup>1</sup> has been added. This plot allows you to visually explore LD patterns and drill down to interesting areas. *Note:* this plot is extremely memory intensive, and performance is best when used to explore relatively small regions of LD.

## Association Testing

### PCA for Population Stratification

- A filter for selecting the Null SNPs has been added to the **Annotation** tab. This filter replaces the column specification found in previous versions and provides increases control over the specification of Null SNPs.

### Marker-Trait Association

- An option for calculating the trend odds ratio has been added to the **Options** tab.
- An option for specifying the PROC GLIMMIX estimation method has been added to the **Options** tab.
- An option for computing the sandwich (empirical) covariance matrix estimator has been added to the **Options** tab.

### SNP-Trait Association

- An option for calculating the trend odds ratio has been added to the **Options** tab.
- An option for specifying the PROC GLIMMIX estimation method has been added to the **Options** tab.
- The process now allows for specification of random effects for binary, ordinal or nominal traits.
- An option for computing the sandwich (empirical) covariance matrix estimator has been added to the **Options** tab.
- An option for specifying which homozygous numerically-coded genotype as “2” has been added.

### Survey SNP-Trait Association *New!*

- This new process tests for association between various types of traits and SNP genotypes or alleles from a single SNP at a time, taking complex survey designs into account.
- You can choose between two main analysis methods: an ANOVA based on SNP genotypes or a regression testing for a linear trend of SNP alleles.
- Adjustments can be made for quantitative covariates.
- Rao-Scott *chi*-square and *F* statistics can also be computed for non-continuous traits. *P*-values from these tests, with adjustments applied if requested, are plotted along the marker map.

---

<sup>1</sup> *FYI: We are still investigating known bugs which occur with the zoom function, in which additional squares not found in the selected region appear in the zoomed plot.*

---

## Copy Number

### Bivariate One-way ANOVA (BOWA)

- The BOWA algorithm has been re-engineered to make two separate One-Way calls, enabling users to specify a second blocking variable.
- An option to segregate results by chromosome has been added to the Options tab. Normally, JMP places all of the results of the analysis in a single data set and opens that data set for inspection, a process that can place exhaustive demands upon your memory and processor. When you check this option, a JMP journal with links to individual data sets (one for each chromosome) is generated instead.
- A new drill-down option has been added that allows you to plot one-way mean expression values by chromosome and position, with adjustable confidence bands.
- An option to compute and display results for only those rows included in the Annotation Data Set has been added to the Anno 1 tab.
- A new Tracks tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

## Spectral Preprocessing

### 2D Bin

- An option for selecting variables to bin using a List-Style specification has been added.

### 2D Detrend

- An option for selecting variables to detrend using a List-Style specification has been added.

### 2D Peak Find

- An option for selecting variables using a List-Style specification has been added.

### 2D Plot

- Options for selecting variables using List-Style specifications have been added.

## Quality Control & Normalization

### Quality Control

#### Distribution Analysis

- An option for specifying the variables for which to display distributions using a List-Style specification has been added.
- An option for filtering outliers has been added to the Action buttons generated by this AP. This option allows you to select rows in an output table corresponding to those samples which are potential outliers and then create a new subset experimental design data set

---

(EDDS) which excludes those rows. This new design data set may then be loaded into the **Distribution Analysis AP** window and the process re-run to generate output with those rows excluded.

- An option to launch select APs using the same data sets has been included in a new Action buttons window. This window, which is generated by this AP, allows you to launch and auto-load the AP dialogs for **Correlation** and **Principal Components** or **Filter Intensities** to perform further QC checks or proceed to **Data Standardize** for normalization.

### **Correlation and Principal Components**

- An option for filtering outliers has been added to the Action buttons generated by this AP. This option allows you to select rows in an output table corresponding to those samples which are potential outliers and then create a new subset experimental design data set (EDDS) which excludes those rows. This new design data set may then be loaded into the **Correlation and Principal Components AP** window and the process re-run to generate output with those rows excluded.
- An option to launch the **Correlation and Grouped Scatterplots AP** using the same data sets has been included in a new Action buttons window.

### **Correlation and Grouped Scatterplots**

- A new filtering option has been added to the **Anno 1** tab. This option allows you to apply a filter to your input data set and include only observations with specific annotation values.
- An option to include only annotated rows has been added to the **Anno 1** tab. Check this option to compute and display results only for those rows included in the Annotation data set.
- A new **UCSC Genome Browser** action button has been added to the **Correlation Scatterplots** window. Clicking this button pre-loads the **UCSC Genome Browser** process to create links to locations in the browser.

### **Filter Intensities**

- A new **Groups** tab has been added. This tab allows you to choose grouping variables from the Experimental Design SAS Data Set (EDDS) specified on the **General** tab. If a grouping variable is specified, the statistics on the **Delete Rows** tab will be calculated for groups of columns defined by that variable, rather than for all columns in the data set. You may also select a percentage cutoff to specify what percentage of groups must meet that condition for the row to be deleted. As an example, when specifying the variable `cancer_status` as a grouping variable, with the condition `NMISS > 5` on the **Delete Rows** tab, and the **Group Percentage for Deletion** set

---

to 100%, a row will only be deleted if both cancer and normal groups have a number of missing values greater than 5.

- An option has been added to the **Delete Rows** tab to allow you to choose whether to delete rows if any of the expressions on the tab are satisfied, or if all the expressions are satisfied.
- *Note:* Both of these options are also available through the Expression QC Workflow.

## Normalization

### Data Standardize

- Percentile can now be selected as a standardization method. When Percentile is selected, a user should then specify the desired percentile in Numerical Parameter for Advanced Standardization Methods field.

### Batch Normalization *New!*

- This process normalizes data using a batch profile based on averaging across within- batch-level control arrays and correcting batch profile for all arrays.

### Batch Scoring *New!*

- This process normalizes batch effect for input data based on a specified batch profile data set (generated using the Batch Normalization AP).

### Loess Normalization and Ratio Analysis

- A parameter allowing you to specify the total number of iterations to be carried out. The first iteration performs an initial LOESS fit. Subsequent iterations perform iterative reweighting. Such iterations are appropriate when there are outliers in the data or when the error distribution is a symmetric long-tailed distribution. The default number of iterations is 1.

## Microarray Analysis

### Row-by-Row Modeling

#### One-way ANOVA

- An option to segregate results by chromosome has been added to the Options tab. Normally; JMP places all of the results of the analysis in a single data set and opens that data set for inspection, a process that can place exhaustive demands upon your memory and processor. When you check this option, a JMP journal with links to individual data sets (one for each chromosome) is generated instead.
- A new drill-down option has been added that allows you to plot one-way mean expression values by chromosome and position, with adjustable confidence bands.

- 
- An option to compute and display results for only those rows included in the Annotation Data Set has been added to the Anno 1 tab.
  - A new Tracks tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

### ANOVA

- Code has been modified to allow a variable named Model to be specified as a design variable.
- An option to segregate results by chromosome has been added to the Options tab. normally; JMP places all of the results of the analysis in a single data set and opens that data set for inspection, a process that can place exhaustive demands upon your memory and processor. When you check this option, a JMP journal with links to individual data sets (one for each chromosome) is generated instead.
- A new drill-down option has been added that allows you to plot mean expression values by chromosome and position, with adjustable confidence bands.
- An option to compute and display results for only those rows included in the Annotation Data Set has been added to the Anno 1 tab.
- A new Tracks tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

### Mixed Model Analysis

- An option to segregate results by chromosome has been added to the Options tab. Normally; JMP places all of the results of the analysis in a single data set and opens that data set for inspection, a process that can place exhaustive demands upon your memory and processor. When you check this option, a JMP journal with links to individual data sets (one for each chromosome) is generated instead.
- An option to compute and display results for only those rows included in the Annotation Data Set has been added to the Anno 1 tab.
- A new Tracks tab has been added to allow users to add a graphical display of a set of genes or markers to JMP statistical results.

### Survival Analysis

- An option to compute and display results for only those rows included in the Annotation Data Set has been added to the Anno 1 tab.

### P-Value Quantile Plotter

- Additional options for how the output  $p$ -values are to be transformed and for specifying  $\alpha$  levels for the confidence intervals have been added to the dialog.

### P-Value Browser *New!*

- This experimental process allows users to adjust, transform and plot  $p$ -values by chromosome and position.
- Optionally, a setting for the Track Gene Text process may be chosen, allowing you to select points in the chromosome-position  $p$ -value plot and subsequently overlay gene tracks on the selected  $p$ -values.

## Pattern Discovery

### Cross Correlation *New!*

- The new Cross Correlation process computes all pairwise correlations between two sets of numeric variables, tests their significance, and optionally depicts them using a heat map. The default type of correlation is Pearson product-moment correlation, but you may optionally specify Hoeffding D, Kendall tau-b, or Spearman rank-order correlation.
- This process may be used to compute cross-correlations between paired genomic data types (e.g., copy number and numerically coded SNP, expression and copy number, expression and expression) and to identify patterns indicating highly correlated results. Numeric variables may be specified in a single data set or two separate data sets.

## Annotation and Power

### Annotation Analysis

#### Venn Diagram

- An option has been added to allow you to create 1-, 2-, and 3-way Venn diagrams with proportional circular areas, and/or proportional surrounding areas.
- An option has been added to allow the location of labels to be modified.
- Starting values for three-way optimization can now be specified.

#### GEO Submission Tool

- This updated 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.

#### IPA Upload

- Users may now specify a reference set when uploading to IPA from JMP Genomics. New options for selecting the array manufacturer and array type are found on the Options tab. Specifying a reference set allows IPA to use the gene set for that array to calculate  $p$ -values for genes you upload for analysis.

### Column Enrichment

- A new option has been added to allow users to specify names for output data sets containing categories and category indicators.

### Track Gene Text *New!*

- This new process creates a settings (.sas) file from a text file that defines a display track for a set of genes to be added to JMP Genomics statistical results like plots of  $-\log_{10}(p\text{-values})$  along chromosomes. You can download the text file from sources like the Table Browser of the UCSC Genome Browser and use it as input to this process. The resulting settings file can then be selected as a track file for embellishing graphics with depictions of genes.

### KEGG Pathway Analysis

- This updated process combines the previously separate KEGG Pathway and KEGG Color APs into one process.

### Configure Proxy Server

- This process has been removed from the **Annotation** submenu. The functions provided by this have been incorporated into a new Configure Genomics Settings process located under JMP's **File** menu.