



## JMP® Genomics 4.1

Discover the biological patterns in genomics data.  
From the two most trusted names in analytic software: SAS and JMP.

### What is JMP® Genomics?

JMP Genomics is a statistical discovery software tool from the two most trusted names in analytic software: SAS and JMP. Research organizations use JMP Genomics to uncover meaningful patterns in high-throughput genetics, expression, copy number and proteomics data. Dynamically interactive graphics make it easy to explore data relationships using a comprehensive set of traditional and advanced statistical algorithms.

### Why is it important?

Research organizations want to maximize their return on investment of the time, money and resources required to generate high-quality genomic data sets. Specialized statistical analyses can help identify the “nuggets of gold” hidden in long lists of candidate genes or biomarkers. Whether it’s used to identify potential drug targets, to explore the biology of a model organism or to develop a predictive disease model, JMP Genomics helps researchers gain a competitive advantage by quickly identifying key genes or proteins.

### Who should use it?

JMP Genomics is designed for biologists, biostatisticians, statistical geneticists and students engaged in analyzing the vast stores of data that are common in genomic research. It delivers a comprehensive set of analysis methods in a single desktop software package. Adopting new software across a large organization can be challenging. That’s why JMP Genomics offers analytics for many data types in the same package, making it easy for you to move into new areas of genomics as the scope of your research expands.

JMP® Genomics software from SAS brings high-powered, sophisticated genomic data exploration and analysis to the desktop. Its unique pedigree integrates the full power of the JMP statistical discovery platform with industry-leading SAS® Analytics tailored for heavy-duty processing of genomics data sets.

JMP Genomics 4.1 introduces a 64-bit edition supported on Windows XP and Vista Business and Enterprise editions.

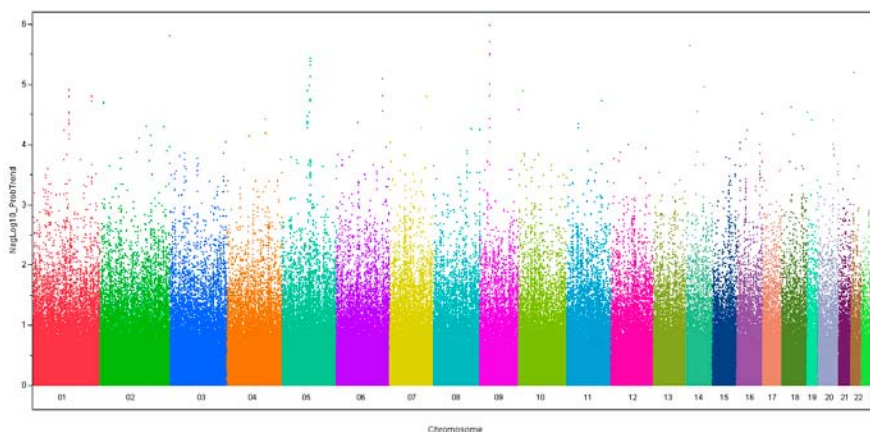
This all-in-one solution helps biologists, biostatisticians and statistical geneticists understand data generated from genetics, expression, exon, copy-number and proteomics studies. JMP Genomics dynamically links advanced statistics with graphics to provide a complete and comprehensive picture of research results, whether

“I think people are starved for software with this level of statistical power and flexibility.”

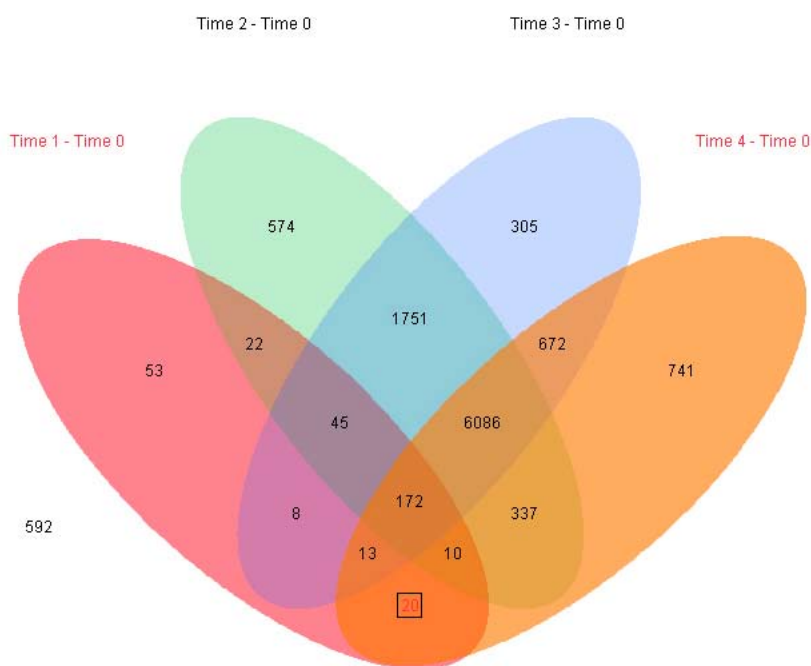
**Erik Sulman, MD, PhD**  
Cancer Researcher, Houston

generated from traditional microarray studies or summarized from next-generation technologies.

With JMP Genomics, biologists, biostatisticians and statistical geneticists enjoy a menu-driven system that simplifies the workflow throughout the research process and interactive data visualization capabilities that let them see and explore their data from every angle, then easily share findings with colleagues. Even students new to



View p-values from statistical tests individually by chromosome, or create custom, multichromosome views.



**Visualize shared patterns with multiway Venn diagrams. Overlay statistical findings and annotation categories to drill down on the most important gene sets.**

genomic analysis quickly begin to discover important trends and outliers in their data, thanks to simplified dialogs and customized workflows that eliminate the need for extensive programming skills or advanced statistical training.

Just point, click and you're on your way. It's powerful analysis made easy.

JMP Genomics brings the power of SAS to the study of genomics, whether you're screening a genome for significant genetic markers, looking for meaningful patterns from expression studies or examining spectral data in a proteomics lab. As your genomic studies expand to new areas, you can

explore new data in a familiar environment — without wasting time and money learning multiple software packages and manipulating data sets to move between them.

Beyond the comprehensive genomics menu, JMP Genomics includes full access to the extensive analysis and graphical features of the JMP platform, a powerful statistical tool for the rest of your laboratory needs. You can design experiments that are large yet efficient, and construct a variety of dynamically interactive graphics driven by a host of generic statistical methods. With JMP Genomics, you are never restricted to a small subset of specialized analysis tools.

JMP Genomics 4.1, our latest release, includes more new capabilities than any prior release. Based on new functionality from the cutting-edge SAS 9.2 and JMP 8 platforms, JMP Genomics 4.1 incorporates more tools for data exploration, enhanced visualization capabilities and a streamlined installation process. It provides new processes aimed at identifying cross-correlations between different types of measurements on the same samples (e.g., Copy Number and Expression). Additional graphical features, like Graph Builder, appeal very much to scientists with limited backgrounds in statistical analysis.

SAS developers worked with members of the global JMP Genomics user community to incorporate the most sought-after new features in JMP Genomics 4.1. These new and enhanced capabilities appeal to genomics scientists with a wide variety of research interests.

“We take the graphical features for granted. But to be able to visualize that separation [in high-dimensional data] is so wonderful. Important differences just pop right out.”

**Faye Schilkey**  
National Center for Genome Resources

“You don’t have to be a whiz-bang programmer to get your answers. It’s very visual, but also very statistically accurate.”

**Susan Dorsey, PhD, RN**  
**University of Maryland, Baltimore**

## Expression

Intermediate Workflows build on Basic Workflows for exon and expression analysis, providing additional options requested by power users. These easy-to-use, prebuilt workflows simplify quality control and statistical analysis of transcript and exon expression data sets, with options that include group-wise intensity filtering and point-and-click selection of normalization, analysis and pattern discovery methods. JMP Genomics includes plenty of flexibility for customized analysis by statisticians with sophisticated modeling requirements. The software uses file-based SAS Analytics to power popular normalization methods such as RMA and GCRMA, so users can apply these methods to much larger data sets than is possible with freeware or most other commercial tools.

## Genetics

Huge data sets common to genomic research pose no problem for JMP Genomics. It provides an exceptional toolkit for statistical geneticists, with analysis options for data sets as large as 1.5 million SNPs for 15,000 individuals on a PC workstation. Enhanced workflows offer streamlined SNP and individual filtering capabilities. Sophisticated linear modeling options for multiple trait types expand analysis options and are even available for analysis of genotypes from complex survey designs.

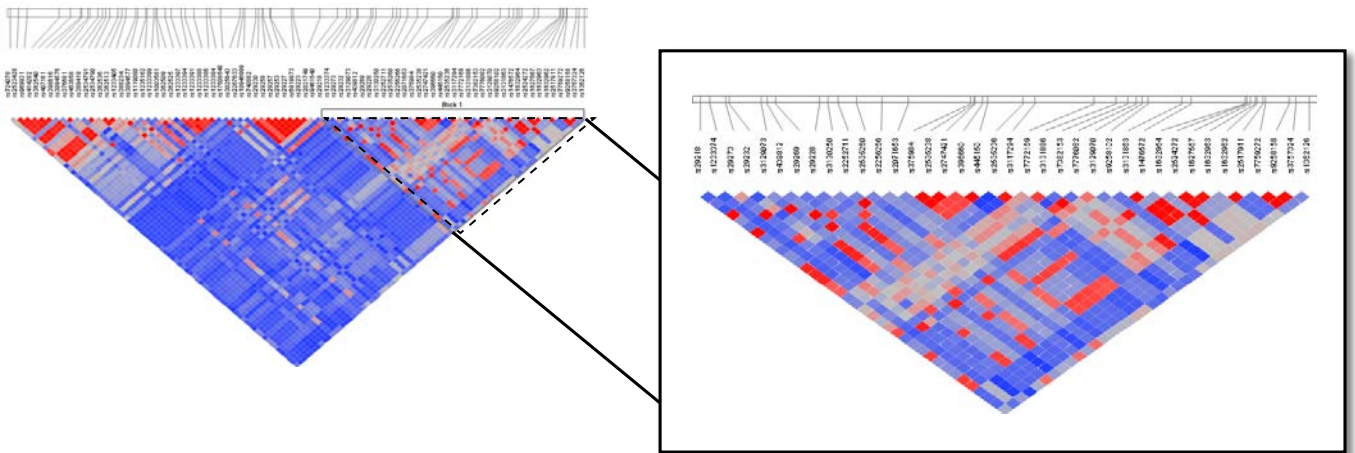
## Copy Number

Easily explore copy number differences between groups or individuals with JMP Genomics. Graphics make it easy to pinpoint common regions of variation when individual samples are compared to a reference group, and ANOVA-based methods offer statistical verification to results. Users can assess

the quality of copy number data using distribution analysis and principal components analysis to identify potential outliers and filter raw data based on statistical criteria. In addition, JMP Genomics offers copy number partitioning with circular binary segmentation.

## Predictive Modeling

The breadth and depth of predictive modeling methods, as well as options for predictor filtering and cross-validation, distinguish JMP Genomics from other genomics solutions. The software identifies key predictors within data sets incorporating multiple data types – SNP, expression, copy number – so users can build models using only the most significant biomarkers, and perform cross-validation with a choice of holdout methods and size options. Users of both JMP Genomics and SAS Grid Manager can even take advantage of experimental grid enablement of predictive modeling functions, saving time by parallelizing the analysis.



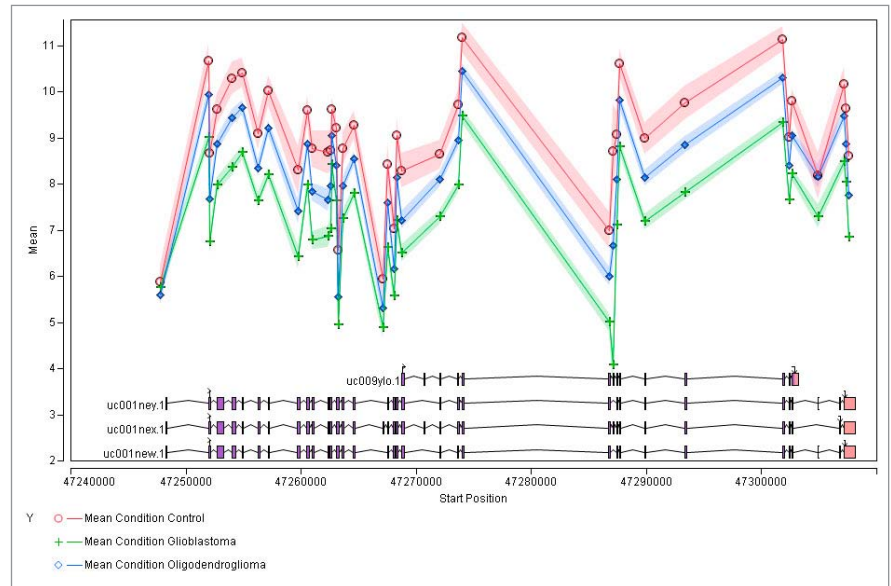
Examine patterns of linkage disequilibrium by position to identify genomic regions of greatest interest, then drill down by highlighting blocks.

## Next-Generation Sequencing

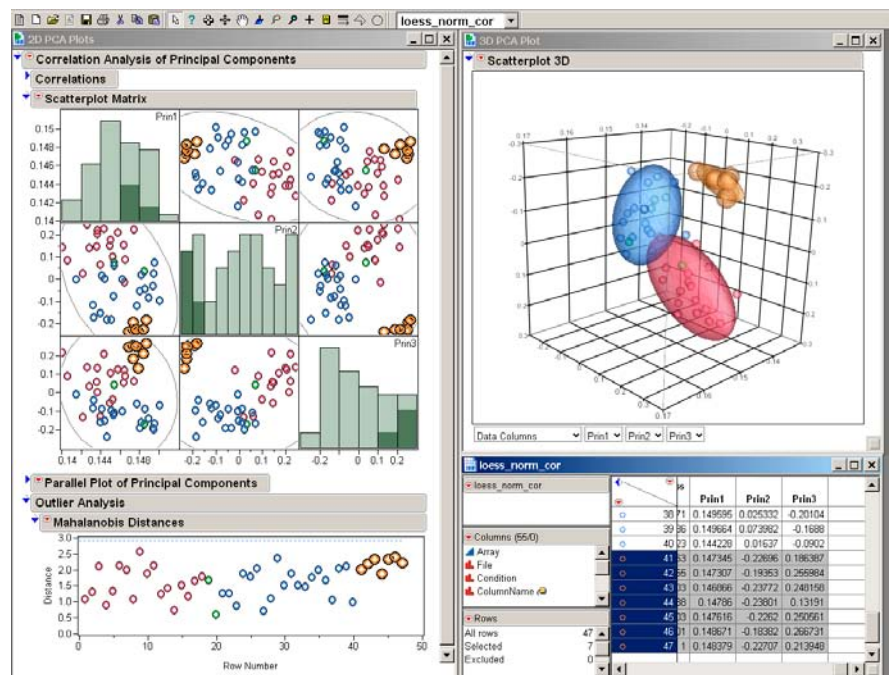
JMP Genomics frees you to follow your data wherever it leads instead of forcing your analysis down predefined paths. Whether you're working with data from traditional microarrays or transitioning to statistical analysis of data summarized from next-generation technologies, JMP Genomics sets itself apart with flexible statistical tools for quality control, normalization, pattern discovery and analysis. By partnering with leaders in next-generation sequencing, JMP Genomics provides sophisticated downstream statistical analysis capabilities to users of state-of-the-art sequence analysis pipelines from Illumina, GenoLogics and the National Center for Genome Resources.

JMP Genomics provides a cost-effective solution for flexible analysis of multiple data types from a variety of platforms, so you spend more time analyzing your data and less time reformatting it. As your analysis expertise grows, so does your ability to take full advantage of its powerful capabilities.

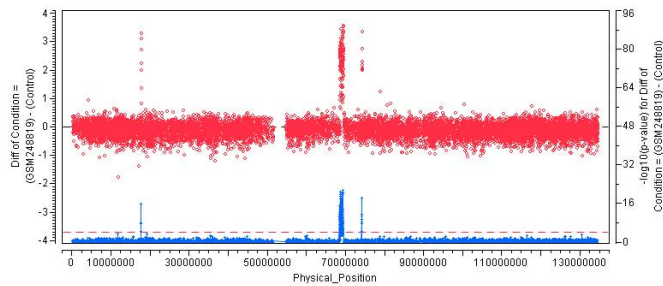
*Corporate, government and academic licenses for JMP Genomics are available by annual subscription. For more information about our software and complimentary Webcast training for prospective customers, please visit our Web site: [jmp.com/genomics](http://jmp.com/genomics).*



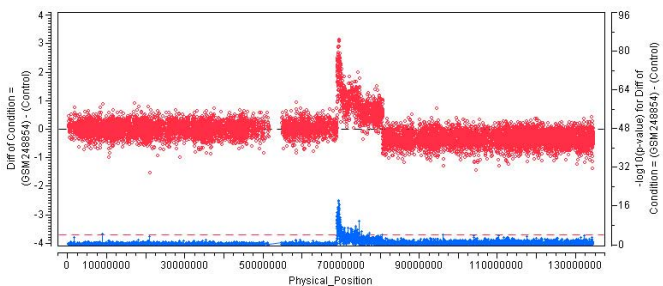
View statistical results in genomic context with overlay of gene and SNP tracks. Here, mean intensity values for probe sets from an Affymetrix exon array are displayed for control samples and two different tumor types.



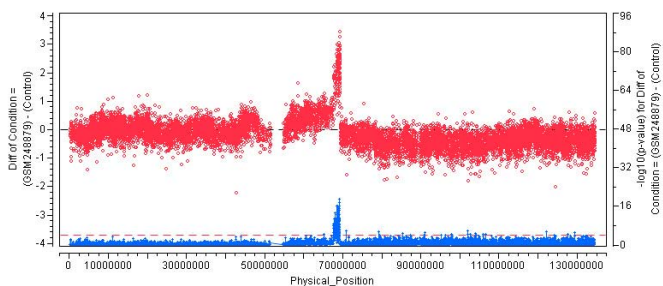
Visual representations of genomics data make it simple to see patterns in your data. JMP Genomics links graphics and data tables for easy browsing of significant results. Output of principal components analysis in JMP Genomics includes both 2-D and 3-D plots, both shown here, where enlarged markers in the graphics correspond to selected rows in the data table.



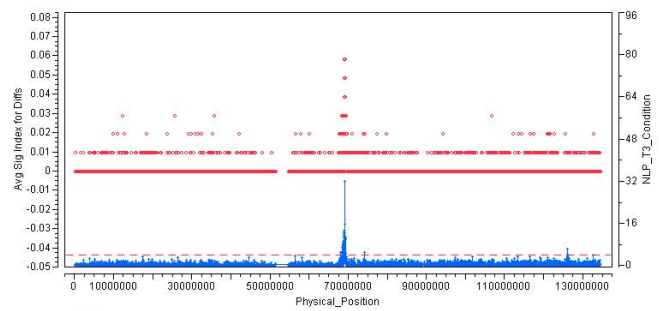
Left Scale: ○ Diff of Condition = (GSM248819) - (Control)  
 Right Scale: +  $-\log_{10}(\text{p-value})$  for Diff of Condition = (GSM248819) - (Control)



Left Scale: ○ Diff of Condition = (GSM248854) - (Control)  
 Right Scale: +  $-\log_{10}(\text{p-value})$  for Diff of Condition = (GSM248854) - (Control)



Left Scale: ○ Diff of Condition = (GSM248879) - (Control)  
 Right Scale: +  $-\log_{10}(\text{p-value})$  for Diff of Condition = (GSM248879) - (Control)

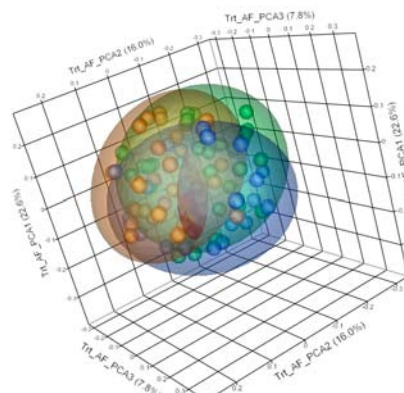
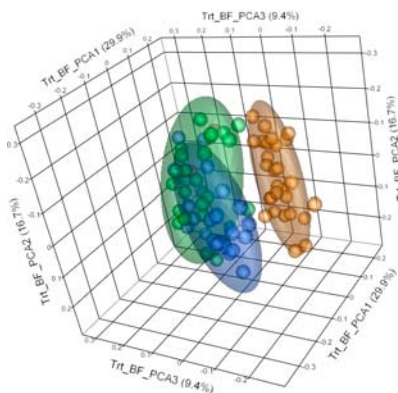


Left Scale: ○ Avg Sig Index for Diffs  
 Right Scale: + NLP\_T3\_Condition

“The ability to instantly recreate any analysis by opening the saved script means that I can determine exactly what analysis was done, even when the student forgets what options they selected. And JMP Genomics uses SAS data sets, allowing me to move easily between SAS and JMP Genomics tools.”

**Lauren McIntyre, PhD**  
**University of Florida Genetics Institute**

Examine individual variations in copy number (left) or guide your search for shared regions with summary plots (right). JMP Genomics also allows examination of group-level differences.



See batch effects in your data and remove them prior to statistical analysis. JMP Genomics offers several different options for batch normalization. At left, samples collected in different batches group closely together, outweighing treatment effects. At right, the same samples are shown after batch effect removal.

## See and explore your genomics data from every angle. Features in JMP® Genomics include:

### Customized SAS Analytics running behind a JMP user interface:

- Offer point-and-click menus and options so users can get started quickly.
- Power robust data import, quality control, analyses, annotation and pattern discovery features using well-documented methods.
- Require no previous SAS programming experience.

### The JMP software platform provides:

- Unparalleled flexibility for point-and-click creation of custom graphics: 2-D and 3-D scatterplots, parallel, overlay, contour and bubble plots.
- Dynamic, click-and-drag interface for visual exploration of data patterns with Graph Builder.
- Built-in JMP Scripting Language (JSL) and auto-generated graphics scripts that make it easy to capture and share important findings.
- Options for creating tailored dialogs for custom analysis processes.

### Interactive graphics generated automatically during analysis:

- Produce easy-to-understand summaries of large data sets.
- Link graphical features to data for point-and-click selection and easy creation of subset data tables.
- Can be queried dynamically to create tailored views of your data, using the JMP Data Filter or a variety of other selection tools.

### Flexible workflows offer options for all users:

- Affymetrix CHP Wizard, which guides new users through analysis of CHP and ARR files through a simple, interactive interface.
- Basic Workflows for expression, exon, genetics and copy number.
  - [New Basic Workflows for Tiling and miRNA.\\*](#)
- Intermediate Workflows for expression quality control and analysis.
- [Expression and copy number workflows now incorporate variance components analysis to guide statistical model selection.](#)
- Workflow Builder, which offers complete control for expert users who wish to create their own custom workflows.

### JMP Genomics imports data from a variety of formats, including:

- Illumina BeadStudio or GenomeStudio output files for expression, SNP, genetic marker, copy number and other data types.
- Exon, whole transcript and 3' expression CEL and CHP files from GCOS and Affymetrix Command and Expression Console.
- CEL, CHP, LOHCHP and CNCHP files from Affymetrix Genotyping Console, and CNAT files.
  - [Now with option to base normalization on autosomes only to reduce impact of gender imbalance within sample groups.](#)
  - [Support added for summarization of SNP6 files with MPS files and import of cytogenetics CEL files.](#)

- miRNA CEL, cytogenetics CEL, tiling CEL files, and BAR files from [Affymetrix Tiling Array Software](#).
- GenePix, QuantArray, one-color and two-color Agilent files.
- Genomics data contained within single text files or multiple text files.
- Text, Excel and comma-separated files.

### Assess genome-wide data sets to:

- Examine missing data patterns for individuals and genetic markers.
- Summarize characteristics of genetic marker data sets: allele and genotype frequencies, HWE, number of missing values, heterozygosity and diversity.
- Filter data sets by marker properties prior to statistical analysis.
  - [New option permits filtering by HWE values for a subgroup \(e.g., controls only\).](#)
- Calculate and visualize linkage disequilibrium measures to zoom into interesting regions with interactive triangular plots.
- Generate distributions of categorical and continuous phenotypes.

### Perform candidate-gene or whole-genome SNP analysis to:

- Analyze data sets as large as 1.5 million SNPs for 15,000 samples.
- Explore associations between genetic markers and binary or quantitative traits while adjusting for covariates.
- [Correct association tests for relatedness and population structure simultaneously.](#)
- [Test for associations using imputed SNP data.](#)
- Visualize and correct for population structure prior to association tests with PCA for Population Stratification or Multidimensional Scaling.

### Expand analysis options for marker data to incorporate:

- Detection of, and corrections for, relatedness among samples.
- Haplotype estimation and discovery of haplotype-trait associations.
- Selection of tagSNPs for haplotypes or linkage disequilibrium blocks.
- Computation of genetic distance matrices for individuals or populations.
  - [Calculate IBD, IBS and allele-sharing individual relationship matrices.](#)
- Single-marker, interval and composite-interval QTL mapping.

### Assess large expression data sets with confidence to:

- Identify data quality issues and remove outlier arrays prior to statistical analysis.
- Visualize intensity distributions, 2D and 3D PCA plots, and sample clustering patterns to explore the impact of experimental and technical effects.
  - [Now with enhanced visualizations to pinpoint which experimental and technical factors contribute to the variance explained by each principal component.](#)

“When you’re going from looking at 10 genes to looking at thousands of genes, making biological sense of the results isn’t easy – it’s impossible to do if you don’t have the tools that help you easily visualize and explore the annotation of the results. JMP is great for that.”

Tom Juenger, PhD  
University of Texas, Austin

- Normalize within and across arrays to remove confounding sources of variation to:
  - Perform batch normalization and scoring, or utilize PLS normalization to remove known technical effects.
  - Use Loess (within or between arrays), Quantile, Factor Analysis, and ANOVA normalizations as well as standardization to a variety of statistics (e.g., mean, median, IQR).
  - Specify a set of reference quantiles to apply to new data during between-array Loess or Quantile Normalization.
  - Apply RMA and GCRMA for Affymetrix 3’ expression arrays.
  - Use RMA, mean and median standardization and summary for Affymetrix exon and miRNA arrays.
  - Perform MAT and quantile normalization for Affymetrix tiling arrays.

#### Apply trusted statistical methods with flexible options to:

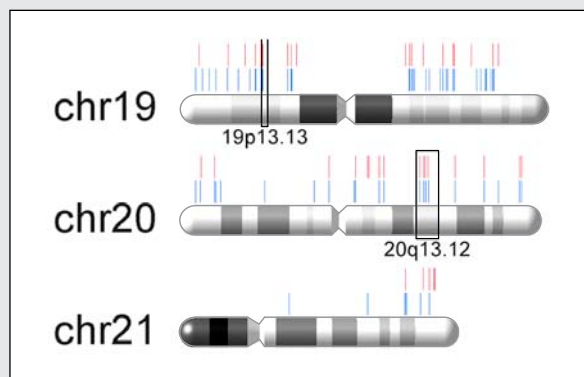
- Perform gene-by-gene modeling to discover statistically significant differences at the probe, transcript or exon level while correcting for multiple tests.
  - [New Adaptive Holm, Adaptive Hochberg, Adaptive FDR and Dependent FDR methods for multiple testing correction now available in ANOVA-related and other processes.](#)
- Select sets of comparisons for inclusion in output and reverse the order of differences with new [Difference Chooser](#).
- Reveal biological insights with [pattern discovery tools](#).
  - [Plot and color profiles of raw or normalized intensities by sample or by group with dynamic data filtering to pinpoint key patterns.](#)
  - Cluster samples or genes with hierarchical and K-means analyses.

#### Apply advanced predictive modeling analysis tools to allow:

- Identification of reliable biomarkers from large, wide data sets.
  - [New predictive modeling for survival analysis.](#)
- Assessment of multiple data types from different experiments.
- [Comparison of results across eight different predictive modeling methods.](#)
- Customized predictor filtering during model construction.
- Cross-validation with adjustable hold-out and iteration options to enable comparison of relative performance across multiple models.
- Learning Curve analysis assessing the impact of sample size.

#### Assess copy number data sets to:

- Examine data quality with PCA and data distribution analysis.
- Analyze SNP intensities directly or importing copy number values generated by a variety of algorithms.
- Look for shared genomic areas that display statistically significant differences using ANOVA.
- Compare breakpoints within and between samples identified by circular binary segmentation.



Create custom genome color themes and overlay statistical results, then zoom and drill-down to visualize gene and SNP tracks.

#### Integrate statistical analysis into next-gen sequencing workflows to:

- Import sequence counts at the SNP, exon or transcript level generated by partner software from Illumina, the National Center for Genome Resources or GenoLogics, or summarized by other software.
- Examine mRNA-seq data using trusted methods implemented in existing exon and expression workflows.
- Test for association between variant alleles and traits.
- Perform cross-correlation analysis to relate sequence counts to other genomic measures.

#### Use JMP Genomics annotation tools to:

- Merge functional information with statistical results.
- Download annotation and library files from Affymetrix NetAffx.
- Upload results to Ingenuity Pathways Analysis to seek points of interaction between SNP, gene and protein lists.
- [Retrieve KEGG IDs and pathways to visualize sets of co-regulated genes and perform enrichment analysis.](#)
- Create Venn diagrams to assess overlap of up to five categories simultaneously, with proportional area option for one-, two- and three-way diagrams.

#### Create genome-level views to:

- [Color chromosomes using custom themes based on annotation information or summarized statistical results.](#)
- [Overlay information from multiple comparisons or experiments to find regions of shared significance.](#)
- [Drill down on interesting regions to plot p-values and view gene or SNP tracks.](#)

\*Blue text denotes feature introduced in JMP Genomics 4.1.



**SAS Institute Inc. World Headquarters** +1 919 677 8000

To contact your local JMP office, please visit: [www.jmp.com](http://www.jmp.com)

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 trademarks of their respective companies. Copyright © 2009, SAS Institute Inc. All rights reserved. 103112\_553487.1209