



## JMP® Genomics 5.1

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 statistical discovery software 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?

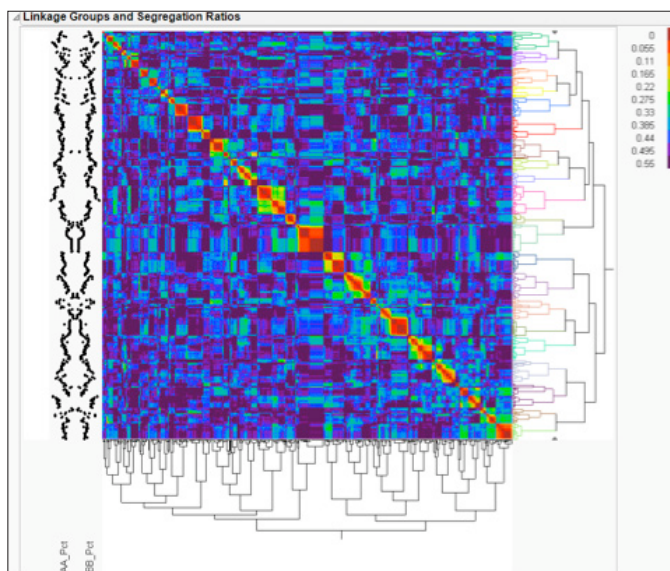
JMP Genomics is designed for biologists, biostatisticians, statistical geneticists and students engaged in analyzing the vast stores of data that are common in genomic research. It delivers a comprehensive set of analysis methods in a single desktop software package. Adopting new software across a large 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 provides an all-in-one package for high-powered, sophisticated genomic data exploration and analysis. 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.

This integrated solution helps biologists, biostatisticians and statistical geneticists understand data generated from large genetics, expression, exon and copy number studies. JMP Genomics dynamically links advanced statistics with graphics to provide a complete and comprehensive picture of research results, whether generated from traditional microarray platforms or summarized from next-generation technologies.

JMP Genomics sets itself apart with flexible statistical tools for quality control, pattern discovery and analysis of continuous intensities, counts or genotypes. As your experience grows, so does your ability to take full advantage of its powerful capabilities.

A menu-driven system simplifies analysis workflows and interactive data visualization capabilities, letting researchers see and explore their data from every angle, then easily share findings with colleagues. Even students new to 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. That’s why a growing number of professors are teaching with JMP Genomics.



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

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

JMP Genomics now uses genotype data from bi-parental crosses to calculate recombination rates, with interactive cluster heat plots to help select linkage groups.

“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

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 associations, looking for meaningful patterns from expression studies or assessing copy number differences. As your inquiries 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 its rich library of prebuilt graphics, JMP Genomics includes full access to the extensive analysis and graphical features offered by the JMP 9 platform. 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. Features like the drag-and-drop Graph Builder and interactive Data Filter provide the flexibility for all users to create customized views of their data.

JMP Genomics 5.1, our latest version, is built on the JMP 9 and SAS 9.3 platforms. With the JMP Genomics Starter, a customizable home window, new and existing users can quickly access the tools that fit their analytic needs. The JMP Genomics Wizard guides users through the import of sample information and data sets from popular genomics data platforms and text formats. Graphics and follow-up analysis options are organized into tabbed reports, with underlying tables hidden to simplify the presentation of complex analysis results. Easily recall

the hidden tables to view details, or close all tables and graphics with a single click.

SAS developers worked with members of our global JMP Genomics user community to develop the new feature set in JMP Genomics 5.1. Additions include import routines for standard data formats (e.g., BAM, VCF), several new methods for analysis and summarization of rare variants across loci or pathways, normalization and analysis routines tailored for count data, and a suite of new linkage map construction and visualization tools.

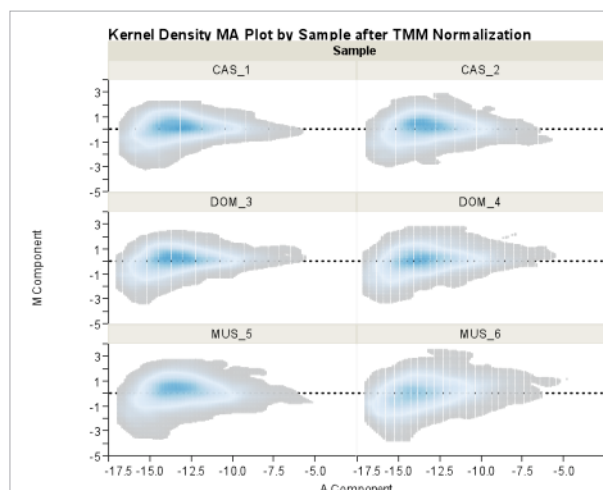
With JMP Genomics 5.1, you have the freedom to integrate specialized analysis tools into our platform by creating add-ins that call JMP, SAS, R, or other programs seamlessly. You can build point-and-click user interfaces to call external analytics, return output tables and static graphics to JMP Genomics – or create interactive JMP reports that summarize tabular results. These application development tools make it easier than ever to share your custom

analytics with a broader audience. To download add-ins created by other users and share your own add-ins with the JMP user community, please visit [www.jmp.com/addins](http://www.jmp.com/addins).

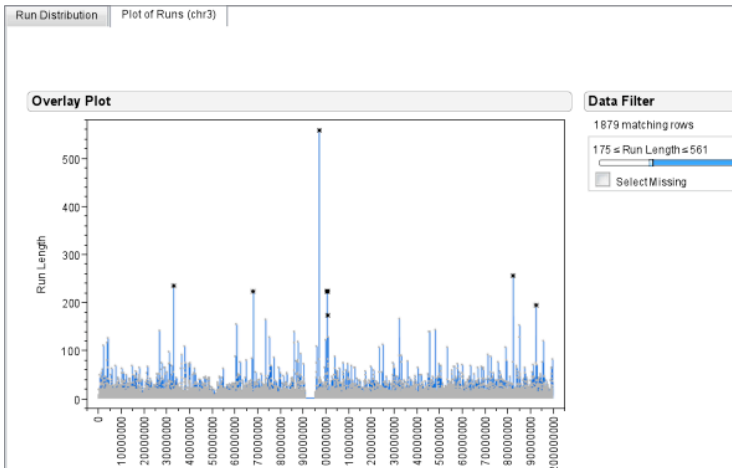
## Expression

Easy-to-use, prebuilt basic and intermediate workflows in JMP Genomics simplify quality control and statistical analysis of transcript and exon expression data sets, with options that include groupwise intensity filtering and point-and-click selection of normalization, analysis and pattern discovery methods. Sample filtering fields in many processes make it simple to restrict an analysis to a subgroup of samples that share similar characteristics.

SAS® Analytics power the normalization and analysis routines in JMP Genomics, so you can work with much larger data sets than is possible with freeware or most other commercial tools. For JMP Genomics 5.1, we have added new normalization



Scale count data across samples using TMM normalization, compare TMM factors between samples, and view kernel density plots of normalized data.



Identify genomic regions that contain marker genotypes shared identical by state between related or unrelated individuals.

methods and general linear modeling capabilities for count data summarized from next-gen studies. These routines are also featured in pre-built workflows for RNA-seq and miRNA-seq analysis.

## Genetics

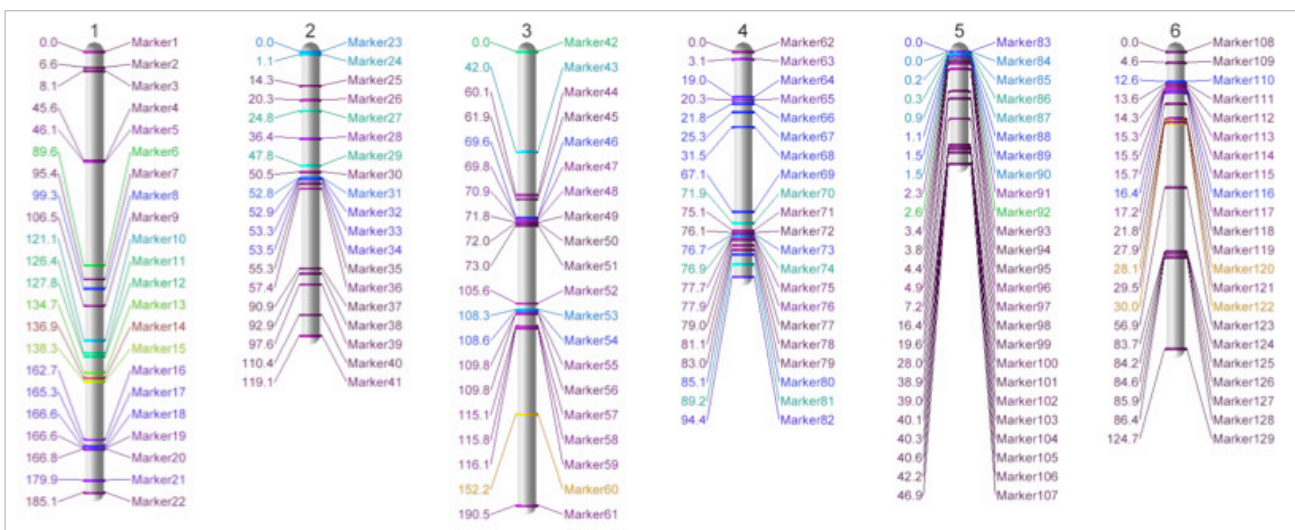
JMP Genomics 5.1 provides an exceptional toolkit for statistical geneticists. Available tools range from simplified case-control association workflows to linkage disequilibrium and sophisticated linear modeling

methods that support a variety of trait types. You may examine associations between SNPs and multiple continuous traits simultaneously, overlay results from several single-trait tests in a Venn diagram, correct for population structure, and discover SNP-SNP interactions. Create, compress and easily integrate relationship matrices into association tests to simultaneously correct for population structure and relatedness with Q-K mixed model analysis.

Point-and-click interfaces to an extensive and growing collection of methods streamline association analysis of rare and common variants. JMP Genomics 5.1 adds new tools to resolve strand differences across studies and perform GWAS meta-analysis. You can also identify of regions shared identical by state (IBS) between related or unrelated individuals.

## Linkage Maps and Breeding Analysis

JMP Genomics 5.1 features a new suite of interactive processes for the construction, optimization and visualization of marker linkage maps used in efforts to improve various agronomic crops or in animal breeding programs. New linkage groups are estimated using recombination fractions calculated from experimental cross data, with cluster heat plots used to guide the selection of the number of linkage groups. Markers may be ordered within these linkage groups or using pre-defined consensus groups. JMP



Visualize linkage maps created in JMP Genomics or imported from other software.

Genomics 5.1 features a default multi-dimensional scaling (MDS) marker ordering algorithm; however, JMP Genomics users who also license SAS/OR may select an advanced marker order optimization method.

Visualize and filter marker maps created in JMP Genomics or imported from other programs using simple interactive graphics or high quality multi-chromosome views. You can even use interactive tools to compare marker orders between existing maps and new maps built within JMP Genomics. In addition, you can explore genotype-environment interaction analysis in multi-environment trials, summarize phenotype information, and perform QTL analysis using newly constructed marker maps.

## Copy Number

Easily explore copy number differences between groups or within individuals with JMP Genomics. You can assess quality of copy number data using distribution analysis and principal components analysis, identify potential outliers, and filter raw data based on statistical criteria. Additionally, you can adjust copy number or LOH data using paired or grouped reference samples.

JMP Genomics offers copy number partitioning with an implementation of a fast circular binary segmentation (CBS) algorithm. ANOVA-based methods also can be used to filter data before analysis or to find statistically significant differences across groups or when comparing individual samples to a reference group. Interactive graphics with results plotted by position help pinpoint shared regions of variation.

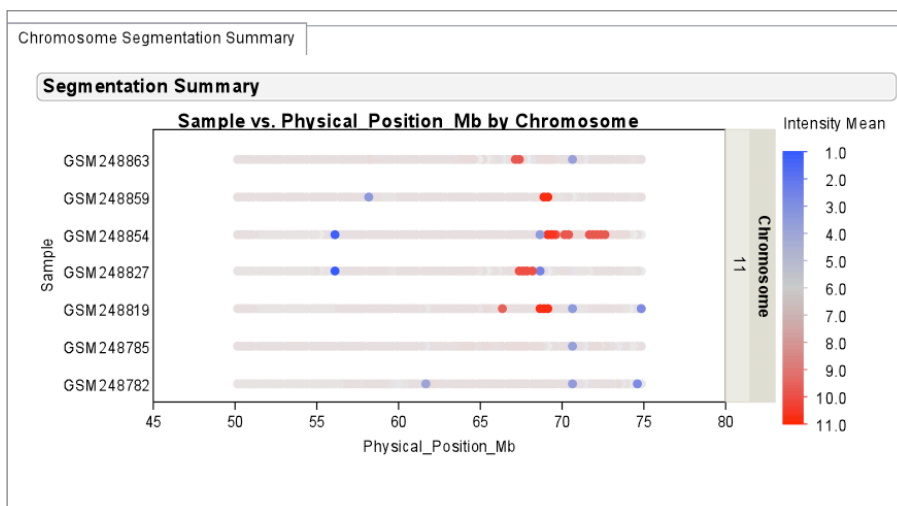
## Predictive Modeling

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 wide data sets incorporating multiple data types – SNP, expression, copy number – so you can build models using only the most significant biomarkers. Replication and iteration strategies implemented in the software seek to reduce bias, with honest cross-validation approaches that can accurately assess the relative performance of hundreds of different models at a time.

Enhancements introduced in this new release include the ability to lock in key predictors, compute survival residuals to use as input for a variety of predictive modeling methods, and specify custom costs for classification with binary or nominal outcomes.

## Next-Generation Sequencing

JMP Genomics 5.1 provides sophisticated downstream statistical analysis capabilities for aligned reads from state-of-the-art sequence analysis pipelines. Import counts from text formats or summarize counts from SAM, BAM, and Eland input files to take advantage of new normalization and general linear modeling methods. Workflows for RNA-seq and miRNA-seq streamline the steps in the analysis process.



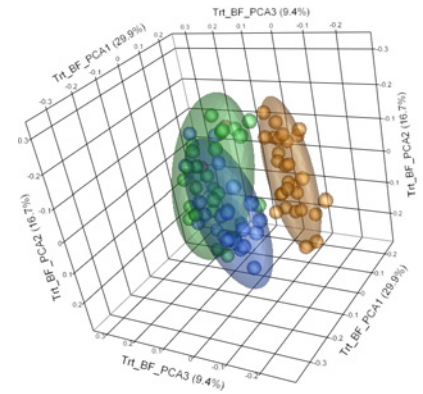
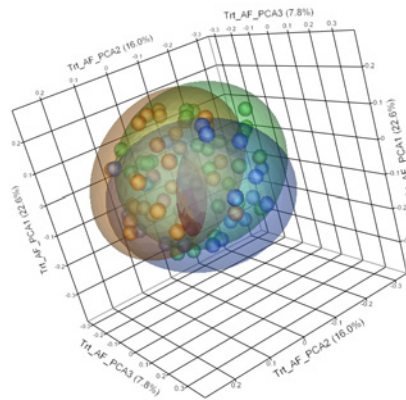
New segmentation summary plots can be filtered interactively to identify shared regions of copy number loss or gain.

You can import genotypes directly from a variety of text formats or VCF files, or elect to call variants from BAM files using a reference genome. This new release supports an expanded list of methods for association analysis of rare and common SNP variants and a new process for identifying regions identical by state (IBS) between related or unrelated individuals.

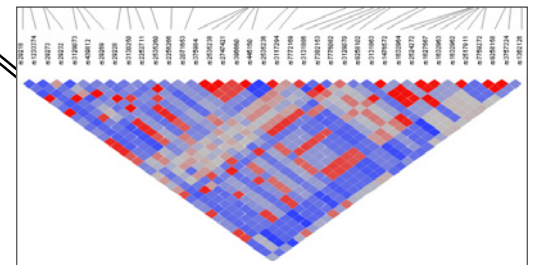
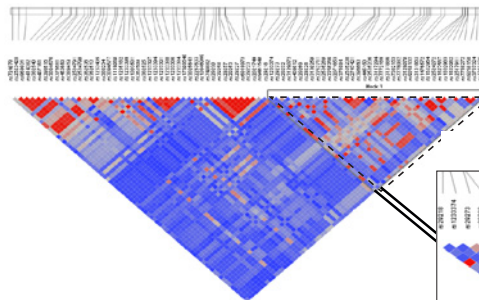
Take advantage of the rich information provided by sequencing experiments to screen for significant correlations between different data types. View counts or statistical analysis results in the JMP Genomics Browser, and overlay histogram and heat plot tracks with individual- or group-level summaries to complement SNP and gene tracks.

Corporate, government and academic licenses for JMP Genomics are available by annual subscription. For more information about our software and complimentary Getting Started webcasts or to request new features for future releases, please visit our website: [jmp.com/genomics](http://jmp.com/genomics).

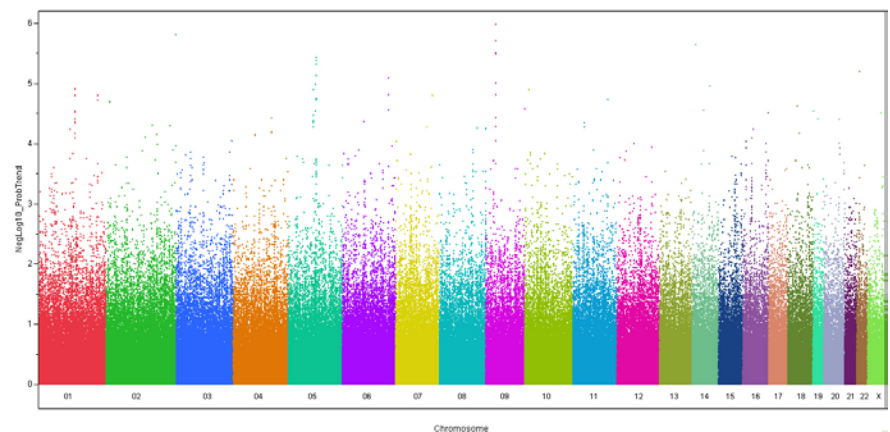
JMP Genomics 5.1 is supported on most 32- and 64-bit business versions of Windows XP, Vista and Windows 7 desktop and server operating systems.



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.



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



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

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

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

- Support 32- and 64-bit Professional, Business and Enterprise editions of Windows XP, Vista and Windows 7 desktop and server operating systems.
- Offer point-and-click menus and options so users can get started quickly.
- Drive powerful and robust data import, quality control, analyses, annotation and pattern discovery features using well-documented and innovative methods.
- Requires no previous SAS programming experience.

### The JMP software platform provides:

- Dynamic, drag-and-drop interface for visual exploration of data patterns with Graph Builder.
- Point-and-click creation of custom graphics: 2-D and 3-D scatter-plots, box, parallel, overlay, contour, trellis, and bubble plots.
- Easy copy-and-paste into Word and PowerPoint.
- An add-in infrastructure that simplifies the integration of external analytics into JMP.
- R integration capabilities that let R users leverage JMP's interactive graphics to display analytic results.
- Tools for R programmers to build and package user interfaces that let them share customized R analytics with a broader audience.
- 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 with extensive drill-down capabilities.
- Are organized into tabbed reports and linked to underlying data tables.
- Offer point-and-click selection and easy subset creation.
- Can be queried dynamically to create tailored views of your data, using the JMP Data Filter or a variety of other selection tools.
- [Can be converted to static RTF or PDF reports with new Create Report buttons on tabbed reports.](#)

### Flexible workflows offer options for all users:

- JMP Genomics Wizard guides the import of new data sets.
- Basic workflows for expression, exon, genetics, copy number, tiling and miRNA.
  - [New Basic RNA-seq workflow offers normalization and modeling options for count and continuous data.](#)
  - [Updated Basic miRNA expression/miRNA-seq workflow offers greater flexibility for miRNA analysis.](#)
- Intermediate workflows for expression quality control and analysis.
- Q-K Analysis and Genetics Rare Variants workflows.

- Expression and copy number workflows include variance components analysis to help identify important factors to include in statistical models.
- Workflow Builder offers complete control for expert users who wish to create their own custom workflows.
- [Journal Builder creates a journal that links results from analyses not originally run in a workflow.](#)

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

- Summarized read count, RPM, RPKM, or genotype data in text format.
  - Raw read counts for direct analysis or summarization using gene model information in UCSC format or BED files.
    - [Aligned sequence reads in SAM, BAM, and Eland file formats.](#)
    - [Generate counts, RPM, and RPKM values at the exon or gene level using gene models in BED, UCSC or text format or for fixed positional bins of user-specified size.](#)
  - [Call variants in BAM files using bcftools.](#)
  - [Complete Genomics variant, dbSNP, and gene variant summary files.](#)
  - [CLC Bio SNP and indel summary files.](#)
  - [VCF v4.0 files, a standard format of the 1000 Genomes Project.](#)
  - Illumina BeadStudio or GenomeStudio output files for expression, SNP, genetic marker, copy number and other data types.
    - [Multiple expression and miRNA Final Report files and their associated sample files may now be imported and combined.](#)
    - [Multiple SNP Final Report or Full Data tables may now be imported simultaneously using the same map file.](#)
  - Exon, whole transcript, miRNA and 3' expression CEL and CHP files from GCOS and Affymetrix Command and Expression Console.
  - Tiling CEL files and BAR files from Affymetrix Tiling Array Software.
  - CEL, CHP, LOHCHP and CNCHP files from Affymetrix Genotyping Console, and CNAT files.
  - Cytogenetics CEL and CHP files.
  - GenePix, QuantArray, one-color and two-color Agilent files.
  - Genomics data contained within single text files or multiple text files.
  - Excel and comma-separated files, including data formats from multiple NimbleGen platforms.
- ### Integrate statistics into next-gen sequencing workflows to:
- Analyze sequence counts at the SNP, exon or transcript level generated by pipelines from Illumina, the National Center for Genome Resources, GenoLogics, CLC Bio, or summarized by other software.
  - Normalize and analyze RNA-seq data using point and click workflows.
  - Test for association between rare and common variant alleles and traits using a variety of different methods.
  - Perform cross-correlation analysis to relate sequence counts to other numeric genomic measures.
    - [Now incorporating multiple testing adjustment.](#)

“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

#### Assess genome-wide data sets to:

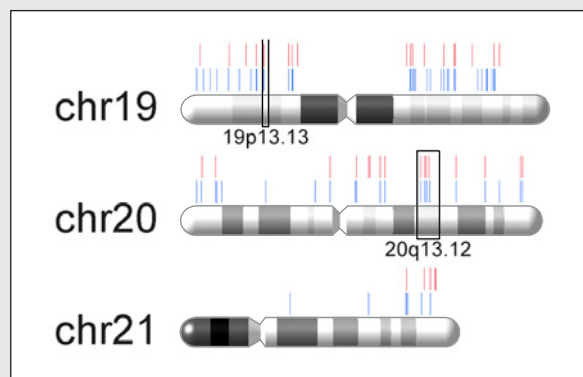
- Examine missing data patterns for individuals and genetic markers.
- Summarize allele and genotype frequencies, HWE, number of missing values, heterozygosity and diversity.
- Filter data sets by marker properties prior to statistical analysis, including filtering by HWE values for a subgroup (e.g.,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

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

- Analyze GWAS data sets as large as 1.5 million SNPs for 15,000 samples on a 32-bit desktop work station.
- Tackle even larger data sets on a 64-bit desktop or server.
- Explore associations between genetic markers and binary or quantitative traits while adjusting for covariates, with experimental permutation options.
  - [Association trend tests now include volcano plots.](#)
- Test for association between SNPs and multiple traits, either separately or jointly, while adjusting for covariates.
- Test for associations using imputed SNP data.
  - [Perform meta-analysis to combine results for the same SNP across multiple GWAS studies using p-values or effects as input.](#)
    - [Reconcile strand differences by comparing study annotations and flipping strands of major and minor alleles as needed.](#)
- Visualize and correct for population structure prior to association tests with Principal Components Analysis (PCA) or Multidimensional Scaling (MDS).

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

- Statistical testing on common or rare SNP variants collapsed within a locus or pathway in Genetics Rare Variants Workflow.
- [Five new methods and a general framework for rare variants association that allows modifications or combinations of methods, using permutation to assess significance.](#)
  - [The Rare Variants Tutorial explains approaches implemented in Genetics Rare Variants Workflow and Rare Variants Association.](#)
- Computation and clustering of genetic distance matrices for individuals or populations.
  - [Calculate Fst as a measure of distance between populations.](#)
- Calculation of IBD, IBS and allele-sharing individual relationship matrices.
  - [Output pairs exceeding a given IBS or IBD threshold.](#)
- [Identification of genomic regions that contain marker genotypes shared identical by state between related or unrelated individuals.](#)
- Compression of K matrices to save computational time during Q-K analysis.
- Correction of association tests for relatedness and population structure simultaneously with Q-K Mixed Model analysis.



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

- Haplotype estimation and discovery of haplotype-trait associations.
  - [New option to output only the most probable haplotype pair.](#)
- Selection of tagSNPs for haplotypes or linkage disequilibrium blocks.

#### Improve crop and livestock breeding strategies by:

- Examining distributions of categorical and continuous phenotypes among different individuals, genotypes, or lines.
- [Analyzing genotype performance in a multi-environment trial, displaying stability measures, genotype and GxE least square means from linear-bilinear models, PCA biplots, and heritabilities.](#)
- [Identifying linkage groups using experimental cross design information and clustering of pairwise recombination rates.](#)
- [Generating marker order solutions within linkage groups using multidimensional scaling \(MDS\) or optimization methods in PROC OPTMODEL from SAS/OR®.](#)
  - [Drill down to visualize pairwise recombination rates for adjacent markers with interactive triangle plots.](#)
- [Visualizing linkage maps created in JMP Genomics or other mapping software packages.](#)
- Performing single-marker, interval and composite-interval QTL mapping.

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

- Identify data quality issues and remove outliers prior to statistical analysis.
- Visualize intensity distributions, 2-D and 3-D PCA plots, and sample clustering patterns to explore the impact of experimental and technical effects.
- Pinpoint experimental and technical factors that contribute to the variance explained by each principal component.

Continued on next page

### Normalize within and across samples to remove confounding sources of variation to:

- Perform batch normalization and scoring, or utilize PLS normalization to remove known technical effects.
- [Adjust for sample-to-sample variability in count data using TMM and KDMM normalization.](#)
- Use loess (within or between arrays), quantile, RMA, GCRMA, factor analysis, and ANOVA normalizations as well as standardization to a variety of statistics (e.g., mean, median, IQR).
- Standardize using a shifting factor and perform log2 transformation after standardization.
- Specify a baseline data set to apply reference information to a new data set during between-array loess or quantile normalization.
- Use kernel density information in loess and quantile normalization.
- 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, exon, or gene level while correcting for multiple tests and adjusting for covariates and random effects.
  - [Analyze count data using general linear models supported by SAS PROC GLIMMIX with automatic use of a multiplicative over-dispersion parameter.](#)
- Screen paired DNA and RNA intensities for allele-specific expression.
- Perform row-by-row analysis of censored survival data.
  - [Specify least squares means effects, custom differences and estimates, and class and continuous covariates.](#)
- Use sample characteristics to easily specify subsets for analysis.
- Output adjusted p-values and t-statistics for statistical tests of differential expression.
- Select sets of comparisons for inclusion in output and reverse the order of differences.
- 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.
- Assessment of multiple data types from different experiments.
- Customized predictor filtering during model construction.
  - [Now lock in key class or continuous predictors.](#)
- Comparison of relative performance across eight different predictive modeling methods using cross-validation with adjustable hold-out and iteration options.
  - [New option for specifying different costs for classification with binary or nominal outcomes.](#)
- Predictive modeling for survival analysis with Harrell's C statistic and integration with Cross-Validation Model Comparison.
- [Computation of survival residuals for further analysis with other predictive modeling methods.](#)

- Calculation of principal components on a primary data set and scoring of components in a secondary data set.
- Graphical depiction of partition tree information.
- Learning Curve analysis assessing the impact of sample size.

### Assess copy number data sets to:

- Examine data quality with PCA and distribution analysis.
- Analyze intensities or counts directly or import copy number values generated by a variety of algorithms.
- Adjust intensities or counts for experimental samples using paired or grouped control samples.
- Compare breakpoints within and between samples identified by circular binary segmentation.
  - [Visualize shared patterns of copy number loss or gain in new interactive summary graphics.](#)
  - [Explore segment means across samples in new output data sets using dynamic filtering tools.](#)
- Filter or shade segmentation results by mean intensity, with optional segment mean intensity lines.
- Find genomic areas that display statistically significant differences between groups (or individuals and a control group) using ANOVA.

### 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 and color pathways.
- Perform enrichment analysis using functional information from Ingenuity Pathways Analysis. +
- Merge pathway information from mSigDB, KEGG or other sources to perform enrichment analysis or gene set scoring.
  - [New option to specify multiple annotation categories.](#)
- Visualize sets of co-regulated genes in KEGG pathways
- Create Venn diagrams to assess overlap of up to five categories of significant results simultaneously, with proportional area option for one-, two- and three-way diagrams.
- Compare list membership for up to five groups and display overlaps with Venn diagrams using common gene identifiers.

### Create genome-level views to:

- Color chromosomes using custom themes based on annotation information or summarized statistical results.
- Use a variety of continuous measures for summarization.
- 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, SNP, bar chart, and color map tracks.

\* Blue text denotes features introduced in JMP Genomics 5.1.

+ Must be licensed separately.



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 © 2010, SAS Institute Inc. All rights reserved. 103112\_S73579.0911