



JMP® Genomics 5

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?

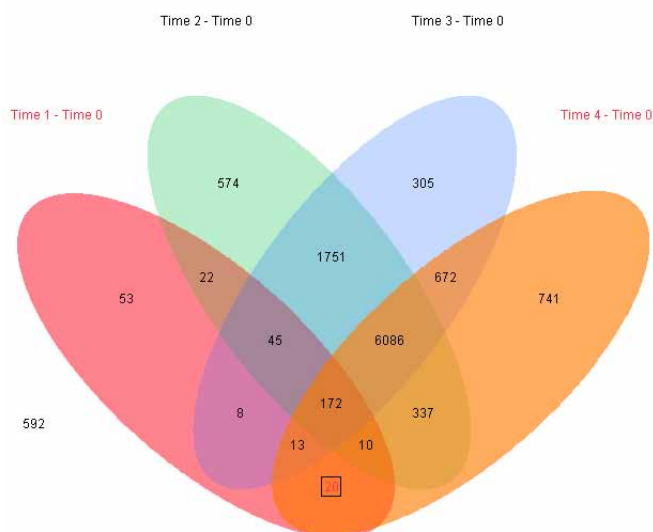
JMP Genomics is designed for biologists, biostatisticians, statistical geneticists and students engaged in analyzing the vast stores of data that are common in genomic research. It delivers a comprehensive set of analysis methods in a single desktop software package. Adopting new software across a large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

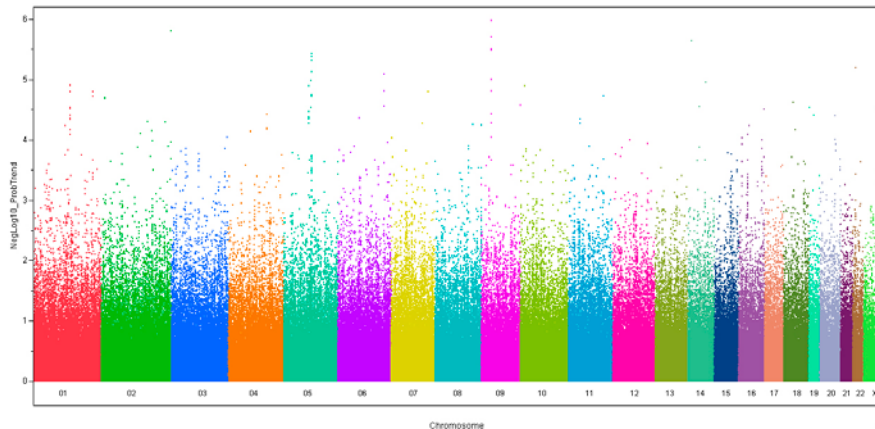
A menu-driven system simplifies analysis workflows and interactive data visualization capabilities, letting researchers see and explore their data from every angle, then easily share findings with colleagues. Even students new to genomic analysis quickly begin to discover important trends and outliers 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

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



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

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, our latest version, includes enhancements across almost all product areas. Without question, the most noticeable change is the simple elegance of the new user interface, which takes full advantage of the new windows management environment of JMP 9. 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.

This new release also introduces the JMP Genomics Starter, a customizable home window. Starter profiles can be created, exported and shared, making it simple for new and existing users to quickly access tools in JMP Genomics that are appropriate to their analysis needs. It is now easier than ever to get started with new data sets. The new JMP Genomics Wizard guides users through the import of sample information and data sets from popular genomics data platforms and text formats.

“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

SAS developers worked with members of the global JMP Genomics user community to develop an extensive list of new features for JMP Genomics 5. Additions include import routines for new data formats (e.g., SAM, CYCHP), capabilities for analysis and summarization of rare variants across loci or pathways, screening tools to detect allele-specific expression patterns, new enrichment analysis options and additional genome track types. As always, the most important source of new ideas for future versions of JMP Genomics is user feedback, so we encourage you to send your comments and suggestions to genomics@jmp.com.

With JMP Genomics 5, you now have more freedom than ever to integrate specialized analysis tools into our platform. Additions made to the underlying JMP 9 infrastructure allow you to create add-ins that call JMP, SAS, R, or other programs seamlessly. You can create point-and-click user interfaces to call external analytics, return output tables and static graphics

“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

to JMP Genomics – or create interactive JMP reports that summarize tabular results.

When coupled with JMP’s new add-in infrastructure, 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.

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. New sample filtering

fields in many processes make it simple to restrict an analysis to a subgroup of samples that share similar characteristics.

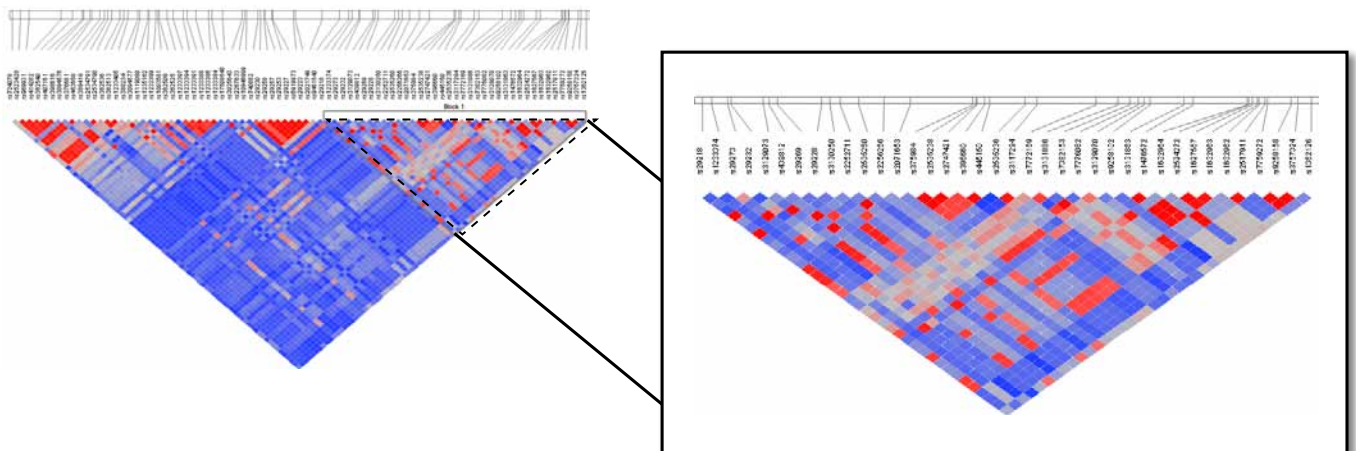
The quantile and loess normalization processes have been enhanced to use kernel density information to improve model fit. SAS® Analytics power popular normalization methods such as RMA and GCRMA, so you can apply these methods to much larger data sets than is possible with freeware or most other commercial tools.

Analysts with sophisticated modeling requirements will find even greater flexibility in JMP Genomics 5. A new process couples concurrent data from both DNA and RNA studies to explore genes showing indications of allele-specific expression. Interesting genes may then be examined further using a variety of mixed-model analysis options.

Genetics

JMP Genomics 5 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 32-bit PC workstation and possibilities to analyze even larger data sets on a 64-bit workstation or server. Case-control association is simplified with the Basic Genetics Workflow, which includes integrated SNP and individual filtering capabilities.

Sophisticated modeling methods, including generalized linear mixed models, are available for testing the association of SNPs with a variety of different trait types, with selected processes now offering experimental permutation options. You may also examine associations between SNPs and multiple continuous traits simultaneously with MANOVA, or overlay results from several single-trait tests in a Venn diagram.



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

JMP Genomics 5 introduces several new methods for discovering and examining SNP-SNP interactions, and still others that examine rare and common SNP variants collectively over a gene, pathway or positional group. New options are also available for the creation, compression and easy integration of relationship matrices into association tests that simultaneously correct for population structure and relatedness. Intermediate workflows for Q-K mixed model analysis and analysis of rare variants present a streamlined interface for these processes.

Copy Number

Easily explore copy number differences between groups or within individuals with JMP Genomics. Users 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. A new utility simplifies the adjustment of data for control samples by experimental samples, which allows for targeted loss of heterozygosity (LOH) detection by using information from paired references.

JMP Genomics offers copy number partitioning with an implementation of a fast circular binary segmentation (CBS) algorithm. Display of partition results has been significantly enhanced for this release, with new options to display segment means, shade segments relative to a reference value, and filter results to display segments that meet a pre-specified cutoff.

ANOVA-based methods 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.

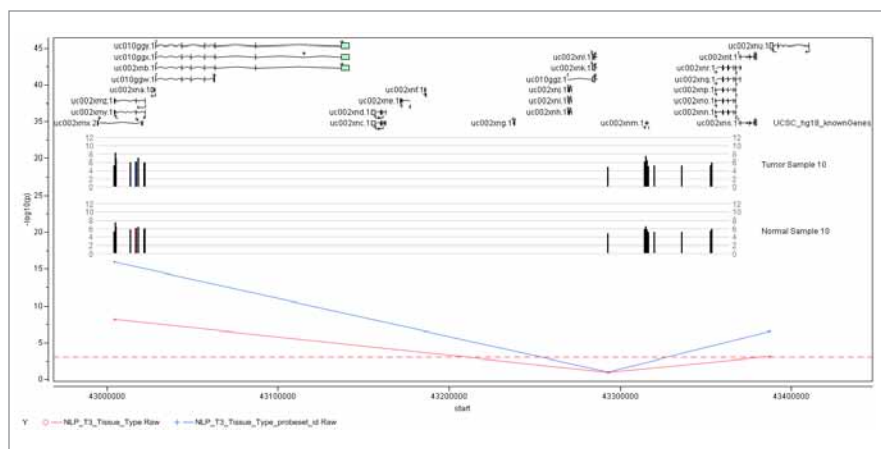
Predictive Modeling

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 users can build models using only the most significant biomarkers and perform cross-validation with a choice of hold-out methods and size options.

Next-Generation Sequencing

JMP Genomics provides sophisticated downstream statistical analysis capabilities to users of state-of-the-art sequence analysis pipelines from Illumina, GenoLogics, National Center for Genome Resources, ABI, Ion Torrent and Complete Genomics. Genotypes or summarized counts from other next-gen platforms can be imported directly from various text formats. A new SAM import process simplifies creation of analysis-ready data sets from this standard format.

The software can perform optional binning of read counts by position and summarize counts or intensities using an existing UCSC-formatted gene model. New tools allow summarization



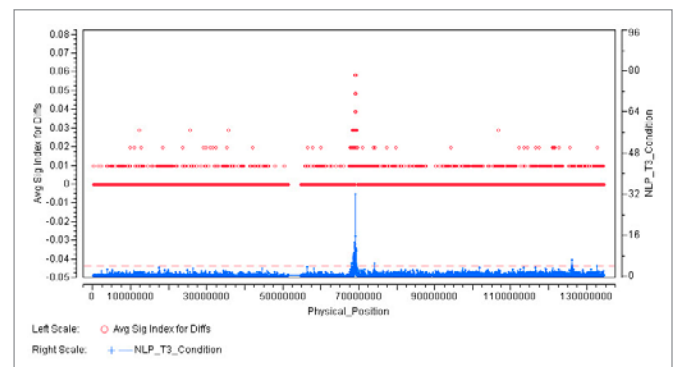
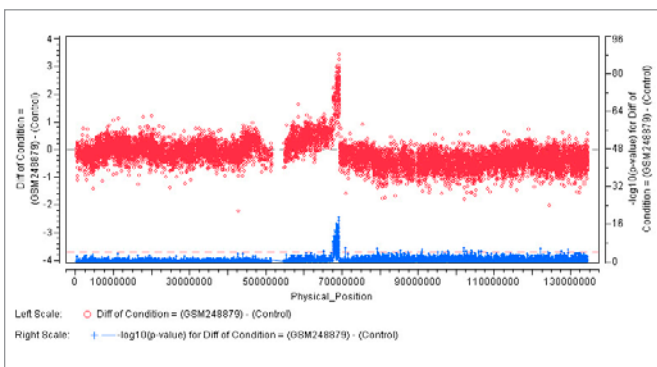
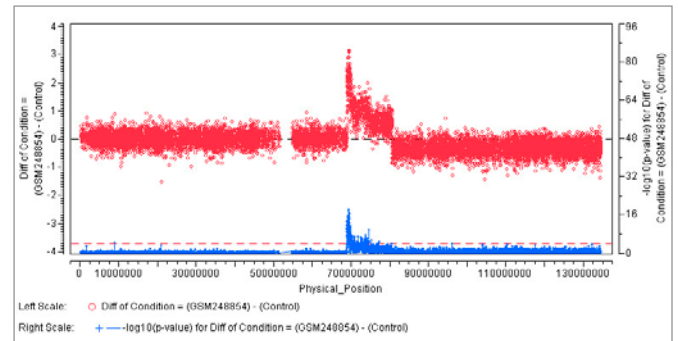
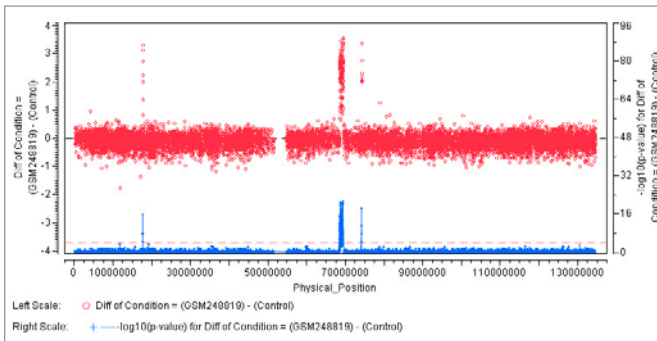
Display summaries of your statistical analyses in genome context to identify interesting regions with pre-built settings for commonly used genomes, or by creating custom genome views. Then drill down to overlay gene, histogram, SNP and heat plot tracks on statistical results. Here, p-values across a genomic region are overlaid with co-localized genes and a histogram track that summarizes raw exon-level data for two samples of particular interest.

and analysis of rare and common variants. Sequence counts or statistical analysis results may now be viewed in the JMP Genomics Browser, with new histogram and heat plot tracks overlaid. These new track types can be used to capture individual- or group-level summaries that complement

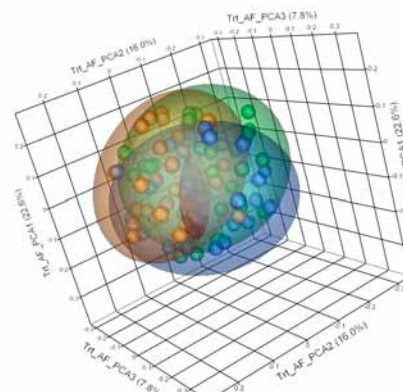
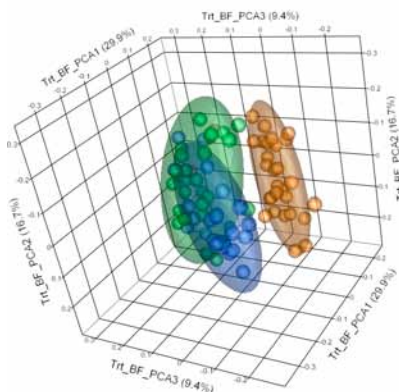
genome information conveyed by existing SNP and gene tracks.

JMP Genomics 5 is supported on most 32- and 64-bit business versions of Windows XP, Vista and Windows 7 desktop and server operating systems.

Corporate, government and academic licenses for JMP Genomics are available by annual subscription. For more information about our software and complimentary Getting Started webcasts, please visit our website: jmp.com/genomics.



Examine individual variations in copy number (left and upper right) or guide your search for shared regions with summary plots (lower 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:

- 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.
- 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:

- New integration capabilities 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.
- A new add-in infrastructure that simplifies the integration of external analytics into JMP.
- Dynamic, drag-and-drop interface for visual exploration of data patterns with Graph Builder.
- Unparalleled flexibility for point-and-click creation of custom graphics: 2-D and 3-D scatterplots, parallel, overlay, contour and bubble plots.
- 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.
- Are organized into tabbed reports and linked to underlying data tables.
- Offer 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:

- JMP Genomics Wizard guides the import of new data sets.
- Basic Workflows for expression, exon, genetics, copy number, tiling and miRNA.
- Intermediate Workflows for expression quality control and analysis.
- New Q-K Analysis and Rare Variants workflows.
- Expression and copy number workflows 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:

- Aligned sequence reads stored in SAM files.
- Illumina BeadStudio or GenomeStudio output files for expression, SNP, genetic marker, copy number and other data types.

- 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.

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, including 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.
 - Now identify and visualize linkage disequilibrium blocks.
- 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 on a 32-bit desktop work station.
 - Tackle even larger data sets on a 64-bit desktop or server.
- Summarize across common or rare variants to perform statistical testing and test grouped SNPs within a locus or pathway.
- Explore associations between genetic markers and binary or quantitative traits while adjusting for covariates.
 - New options to output residuals and R-square for each SNP.
 - Compute LS Means and differences when performing genotype tests with continuous traits or random effects.
 - Experimental permutation options now available.
- Test for association between SNPs and multiple traits, either separately or jointly, while adjusting for covariates.
- Correct association tests for relatedness and population structure simultaneously.
 - New workflow streamlines the steps in performing Q-K Mixed Model analysis.
 - Save computational time by creating a compressed K matrix and using it as input to Q-K analysis.
- Test for associations using imputed SNP data.
- Visualize and correct for population structure prior to association tests with Principal Components Analysis (PCA) or Multidimensional Scaling (MDS).

“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

Expand analysis options for marker data to incorporate:

- Haplotype estimation and discovery of haplotype-trait associations.
- Selection of tagSNPs for haplotypes or linkage disequilibrium blocks.
- Computation and clustering of genetic distance matrices for individuals or populations.
- Calculation of IBD, IBS and allele-sharing individual relationship matrices.
- New option to output pairs of individuals exceeding a user-specified IBS threshold.
- Single-marker, interval and composite-interval QTL mapping.

Assess large expression data sets with confidence to:

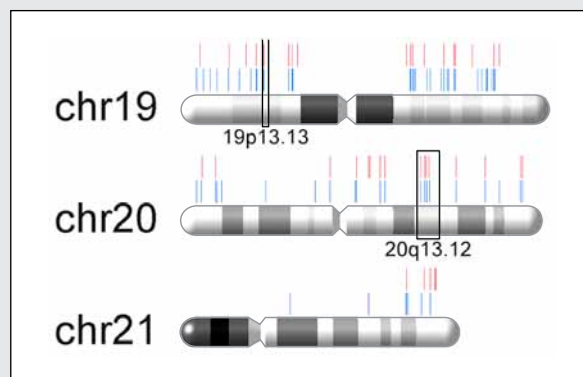
- Identify data quality issues and remove outlier arrays prior to statistical analysis.
- Visualize intensity distributions, 2-D and 3-D PCA plots, and sample clustering patterns to explore the impact of experimental and technical effects.
- Pinpoint experimental and technical factors that contribute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

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).
 - [Standardize using a shifting factor and perform \$\log_2\$ 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.](#)
- 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 and adjusting for covariates and random effects.
- Use sample characteristics to easily specify subsets for analysis.
- Output adjusted p-values and t-statistics for statistical tests of differential expression.
- [Screen paired DNA and RNA intensities for allele-specific expression.](#)
- 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.



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

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.
- [Predictive modeling for survival analysis with Harrell's assessment method and integration with Cross-Validation Model Comparison.](#)
- Calculation of principal components on a primary data set and scoring of components in a secondary data set.
- Comparison of results across eight different predictive modeling methods.
 - [Depict partition tree information graphically for standard models with new Tree Viewer.](#)
- 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 distribution analysis.
- Analyze SNP intensities 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.](#)
- Look for shared genomic areas that display statistically significant differences using ANOVA.
- Compare breakpoints within and between samples identified by circular binary segmentation.
 - [Filter or shade segments by mean intensity, with an option to display segment mean intensity and set a reference value for shading.](#)

Continued on next page

*Blue text denotes feature introduced in JMP Genomics 5.

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.
- Summarize counts using gene model information downloaded in UCSC format.
- 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.
- **Perform enrichment analysis using functional information from Ingenuity Pathways Analysis.***
- Retrieve KEGG IDs and pathways to visualize sets of co-regulated genes and perform enrichment analysis.

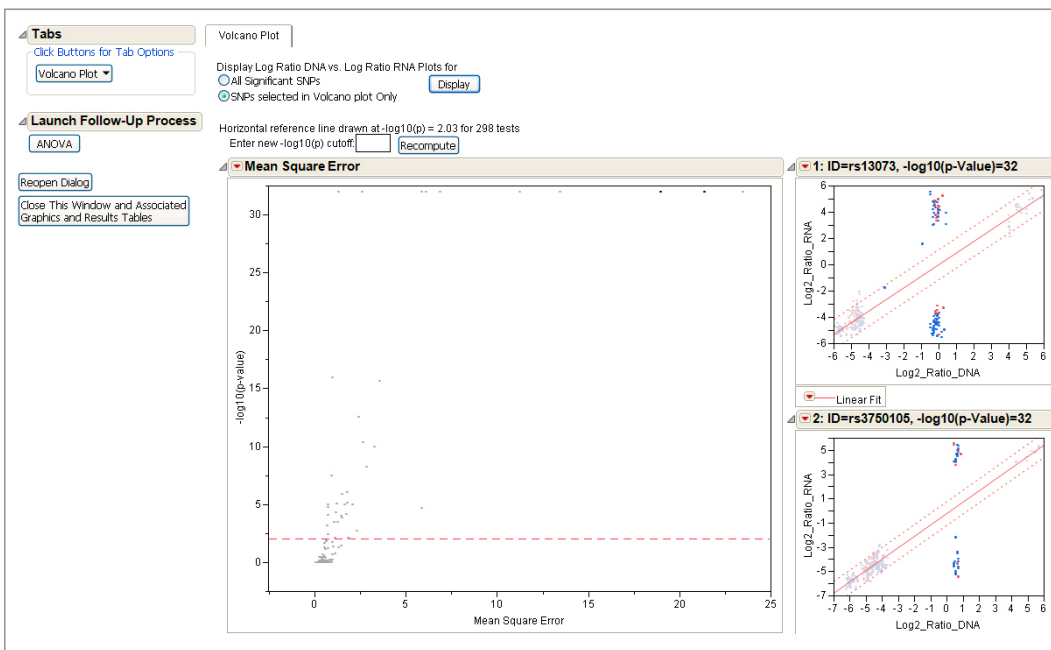
- Using significance indicators, create Venn diagrams to assess overlap of up to five categories simultaneously, with proportional area option for one-, two- and three-way diagrams.
 - **Using a common identifier, compare list membership for up to five groups and display overlaps with Venn diagrams.**

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 or SNP tracks.
 - **New bar chart track allows summarization of reads or intensities.**
 - **New color map track displays heat plots of information for individual subjects.**

*Blue text denotes feature introduced in JMP Genomics 5.

+ This feature is available to licensed users of both JMP Genomics and IPA. Existing customers can contact support@ingenuity.com. To obtain a trial license, go to www.ingenuity.com.



Use paired RNA expression and DNA hybridization data to screen for indications of allele-specific expression with the new Allele Specific Expression Filter process. Examine a summary volcano plot, then drill down to display detailed information for specific SNPs.