Integration of SAS® analytics and JMP® for life science applications: JMP® Genomics

Dr. Valérie Nedbal

SAS Institute

Place des Halles, Quai Kleber

Strasburg

Valerie.nedbal@eur.sas.com

Zusammenfassung

SAS has developed solutions for the life sciences companies to optimize the flow of valuable scientific data within research organization, helping to bring therapies to market faster and more profitably. JMP Genomics delivers statistically based genetics, microarray and proteomics analytical process within a collaborative, scalable and extensible framework. JMP Genomics gathers SAS'core on statistical analytical procedures and visualization technologies.

Schüsselworte: JMP Genomics, Genetics, Microarray, Proteomics, statistical discovery, analytics

Powerful JMP Genomics Software

In the rapidly advancing field of genomics, the ability to conduct sophisticated analyses on huge volumes of data and visually summarise significant findings adds a vital new dimension to scientific research.
JMP® Genomics software from SAS brings high-powered, sophisticated genomic data exploration and analysis to the desktop. Designed to help research scientists and biostatisticians understand data generated from genetics, expression or exon microarray and proteomics studies, JMP Genomics dynamically links advanced statistics with graphics to provide a complete and comprehensive picture of the research results (Figure 1).
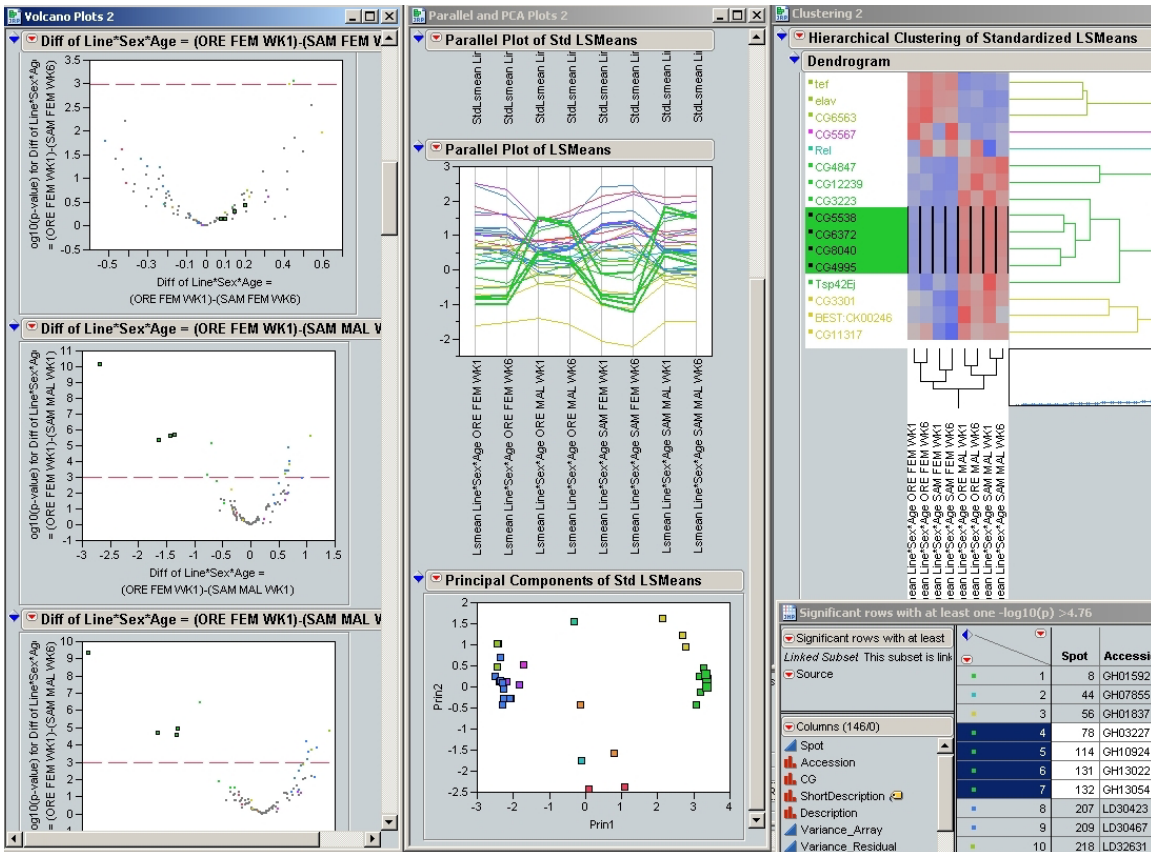
Figure 1 Interactive graphical displaying of a ANOVA output result.

With JMP Genomics, users enjoy the full power of the JMP statistical discovery platform enhanced with industry-leading SAS® analytics tailored for heavy-duty processing of genomics data sets. The menu-driven system simplifies workflow throughout the process, and JMP data visualisation capabilities translate results into interactive graphical displays at the end.

Users don't need programming skills or advanced statistical training to take full advantage of the unique capabilities of JMP Genomics. Point-and-click menus let users select from a variety of powerful, prebuilt SAS analytic processes and easily choose customized options appropriate for your genomic data sets.

More than 100 procedures for genetics, microarray and proteomics analysis make JMP Genomics an all-in-one solution, whether users are screening a genome for significant genetic markers, looking for meaningful patterns from expression microarrays, or examining high-throughput spectral data in a proteomics lab. As the genomic studies expand to new areas, users can explore new data in a familiar environment—without wasting time and money on new software packages and the training required to master 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. Users can design experiments that are large yet efficient, and intuitively construct a variety of dynamically interactive graphics driven by a host of generic statistical methods. For example, basic copy-number analysis is

straightforward. With JMP Genomics, users are never restricted to a small subset of specialized analysis tools, and they don't need to mix and match multiple software packages to handle different data types.

Research scientists use JMP® Genomics to:

- Enhance the quality of genomics research. Get a clear picture of data quality and relationships, even in massive volumes of data. Pinpoint potential problems with research methods early on, then take appropriate steps to optimize lab procedures and enhance data quality. JMP Genomics lets users identify and remove the effects of uninteresting experimental variability.
- Generate graphical summaries automatically. Users don't need advanced technical skills to produce graphical representations of their data. JMP Genomics provides standard graphical outputs that are dynamically linked to your data, plus the flexibility to create their own summaries using JMP's extensive graphing capabilities. You can even cut and paste selected graphics into your publications or presentations.
- Provide context to research findings. Annotation features in JMP Genomics make it easy to link to functional information from a variety of sources, including publicly available databases and tools.

Biostatisticians use JMP® Genomics to:

- Plan experiments. Make the most of limited resources by increasing efficiency in the design-of-experiment process. Estimate power and sample size, and construct and customize powerful designs that include sufficient replication over all important factors.

Adapt existing SAS® code. Run users own customized SAS programs from JMP Genomics menus to accomplish specialized tasks. Easily construct GUIs for customized code with a convenient XML syntax, then share these files with others to standardize commonly used procedures.

- Remove bottlenecks. Import data quickly from a variety of sources, assess its quality, and analyze it. Then distribute results to colleagues using JMP journals and JMP Scripting Language (JSL) scripts. Or help more people in less time by creating custom tutorials to guide other users through standard workflows.

Corporate, government and academic licenses for JMP Genomics are available by annual subscription. For more information, visit JMP Genomics Web site at
www.jmp.com/genomics

Analyze large data sets. Discover biological patterns. Translate results into interactive graphical displays. All with JMP® Genomics.

Customized SAS® analytics running behind a JMP® user interface:

- Allow all users to get started quickly.
- Require no previous SAS programming knowledge.
- Present point-and-click options: selection boxes, menus and radio buttons.
- Offer 100 analytic processes (APs) for data import, quality control, preprocessing, analysis, annotation and pattern discovery, plus all standard features of the JMP software platform.

Interactive graphics and summary tables that are generated automatically during analysis:

- Link points in a graph to lines in a corresponding table for easy viewing.
- Allow selection of sets of interesting points to subset to new tables.
- Produce easy-to-understand summaries of large data sets.

JMP® Genomics imports a variety of data formats, including:

- Affymetrix and Illumina SNP and standard genotype data.
- Affymetrix, Agilent and Illumina expression data.
- GenePix, QuantArray and several other popular image processors.
- ABI Analyst, single- and multiple-text files.

- Text, Excel and comma-separated formats.

Mine genetic marker data from families and unrelated individuals to:

- Summarize information about phenotypes and genetic markers.
- Explore associations between genetic markers and binary or quantitative traits.
- Select tagSNPs for haplotypes or areas of high linkage disequilibrium.
- Reconstruct haplotypes and discover haplotype-trait associations.

Identify key genes from large microarray data sets to:

- Assess quality-control metrics to identify and remove outlier arrays.
- Normalize within and across arrays to remove effects of experimental biases.
- Perform gene-by-gene modeling to discover statistically significant differences.
- Reveal biological insight with pattern discovery and predictive modeling tools.

Find protein biomarkers, using spectral data to:

- Bin, detrend and find peaks in 2-D spectra.
- Plot 2-D and 3-D spectra for graphical browsing.
- Align 3-D spectra across multiple data sets.
- Analyze preprocessed proteomics data with ANOVA and other methods.

Annotate results using public or proprietary sources of information to:

- Merge in annotation files during analysis.
- Search public databases to construct tables of annotation hyperlinks.
- Upload results to Ingenuity Pathways Analysis to seek points of interaction between significantly changing gene and protein lists.
- Color KEGG pathways to identify co-regulated genes.