JMP Genomics, Version 10.0 - Release Notes

This document describes changes and enhancements from JMP Genomics, Version 9.1 to JMP Genomics, Version 10.0. In addition to general maintenance updates, several new enhancements to analytical processes were made. Processes are described in the order in which they first appear in the JMP Genomics menu[1].

General Features

JMP Genomics is an integrated solution built with JMP and SAS capabilities to deliver dedicated analysis workflows for genomic and biological data experiments.

SAS Platform Updates

JMP Genomics 10.0 is built on the latest shipping SAS release, currently this is SAS 9.4M6. For more information about the enhancements to SAS analytical software that are included in this release, please see the What's New in SAS 9.4 web page.

JMP Platform Updates

JMP Genomics 10.0 is built on the latest JMP release v15.1. JMP Genomics 10.0 now includes JMP PRO capabilities in the solution. For more information about the enhancements to JMP software that are included in this release, please see the New in JMP web page.

Software Documentation Updates

The User Guide has been updated to reflect all new and updated software features.

General Enhancements

Several Expression analytical process dialogs have been updated with revised option titles and descriptions to improve understanding and usability. Comprehensive graphical output improvements have been made to better utilize JMP Graph Builder and Data Filter capabilities.

Studies

Add Study

This utility, which identifies and adds study date sets and setting, now generates a metadata file associated with the study. Output data sets from import engines, QC processes and normalization should be included.

[1] **Note**: If you have a suggestion, comment, or encounter a bug in JMP Genomics 10.0, please click **Send a Comment or a Feature Request** under **Genomics > Documentation and Help** or email details to Genomics@jmp.com. For bugs, it is especially helpful if you can attach a settings file for the JMP Genomics process in which you encountered the problem, along with a subset of your data that can be used to reproduce the error. If you cannot share a subset of your own data but can reproduce the problem with one of our sample data sets, please send us a settings file for this so that we can replicate the error. We make every effort to address the issue promptly.

Import, QC, and normalization utilities now add newly created output datasets into the JMP Genomics study metadata for selection in downstream analytic processes.

Import

Import VCF Files

The latest versions of VCF files are now supported. This import engine now parses out information in all Format columns as numeric values, facilitating their use in downstream analyses. It also imports multiple VCF files into a single SAS data set.

Nanostring Input Engine

Enhancements include:

- Improvements to Positive Control normalization
- A new Transformation Method option enables log_2 transformation of the output data.

Reference Gene Normalization

Enhancements include:

- Improvements to Housekeeping Gene normalization
- A new Transformation Method option enables log_2 transformation of the output data.
- Addition of Stable Gene normalization to Reference Gene Normalization.

Import Feature Barcode Matrices *New!*

This import engine imports 10x Genomics Single-Cell RNA Sequencing data to SAS data sets.

Workflows

Basic Single Cell RNA-Seq Workflow *New!*

This process enables you to perform standard exploration on a Single-Cell RNA-Seq data set to analyze the expression patterns of suites of genes within single cells and facilitates identification of gene groups that may be associated within aberrant or otherwise unusual cell types. It first selects variable genes using either the Dispersion or VST method. Then it generates an interactive report including Data Overview, Variable Gene Plot, Clustering, Feature Importance Screening, and Violin Plots on individual gene expression levels. It also provides ways to navigate customized marker genes, launch ANOVA for Differential Gene Expression, and perform t-SNE or UMAP visualizations, provided the necessary R packages are installed. It then clusters cells into families based on expression levels of various marker genes and analysis of principal components affecting gene expression.

General Workflow Enhancements

Expression, RNA-seq, Alternative Splicing/Exon, Tiling, and Copy Number workflow dialog options have been enhanced and restructured to make workflows more intuitive to translational and bio-

scientists with both terminology and ease of use when setting up common differential expression QC and modeling experiments.

Genetics

One new process has been added. In addition, several enhancements to existing analytical procedures for facilitating genetic analyses have been added to JMP Genomics 10.0. All Manhattan plots have been improved to use JMP Graph Builder.

Population Admixture *New!*

This new process takes as input a table of marker genotypes for a sample of individuals that may originated from different ancestral populations. The matrix of marker genotypes is used to estimate a matrix F of dimension m x n (m=number of markers and n=number of individuals) encoding all Binomial parameters f_{ij} (admixture probabilities an individual i=1,2,…,n carrying a reference allele from population j=1,2…,d) and factorize F in the form of $F=PQ$ (Cabreros and Storey, 2019[2]). P is an m x d matrix in which each row of P can be interpreted as the frequency of a single marker's reference allele in each of the estimated ancestral populations, and Q is a d x n matrix in which each column of Q can be interpreted as the admixture populations for a single individual.

Recombination and Linkage Groups

This process now computes *LOD* score values estimated from the marker data, much the same way as the recombination data set, and these values can be used for grouping. A data set with the *LOD* scores values is generated and can be used in the **Linkage Map Order** process for ordering purpose.
A new *Minimum Number of Linkage Groups* option has been added.

Linkage Map Order

The *LOD* Score SAS Data Set created in the **Recombination and Linkage Groups** process can be loaded in the Linkage Map Order for ordering purpose. Now, markers can be ordered based on both recombination and LOD values.
To facilitate the use of *LOD* scores:

- A new *LOD Score Data Set* tab has been added to the process dialog. Options enable you to specify the LOD Score data set and relevant variables for the analysis
- An option to specify a minimum *LOD* score value threshold for breaking linkage groups has been added.
- An option for specifying a pairwise LOD score for use as a constraint in the Map Order Optimization method has been added.

[2] Cabreros, I, and Storey, J.D. 2019. A Likelihood-Free Estimator of Population Structure Bridging Admixture Models and Principal Components Analysis. *Genetics* **212**, 1009–1029.

Cross Evaluation
New options have been added. These include:

- Progeny can now be simulated from outcrosses across multiple generations.
- A selection index can now be computed. This option is useful because it allows users to build super-traits that are a linear combination of multiple traits of interest.
- Partition trees can now be used to evaluate crosses through the prediction of trait values of their simulated progenies.
- Trait values of individuals in the input data set can now be predicted. This option is valuable when a user has a set of individuals that have marker measurements without phenotyping and wants to predict the trait values of each row (individual) in the input data set.
- Progeny can now be selected for crosses into the next generation via threshold, index, or percent values.

Progeny Simulation
New options have been added. These include:

- Progeny can now be simulated from outcrosses across multiple generations.
- A selection index can now be computed. This option is useful because it allows users to build super-traits that are a linear combination of multiple traits of interest.
- Partition trees can now be used to evaluate crosses through the prediction of trait values of their simulated progenies.

Expression

The Expression sub-menu "Row-by-Row Modeling" has been renamed to "Differential Expression".

Distribution Analysis
Parallel plots have been removed. Instead, data are stacked and plotted using JMP's Graph Builder option. Individual distributions are now easily selectable for further analysis. In addition, a local data filter has been added to the dashboard enabling filtering and exploration of results using experimental design information.

Variable Gene Selection *New!*
This process enables you to select a subset of genes that exhibit high cell-to-cell variation in a Single-Cell RNA-Seq data set. It provides two methods, Dispersion and Variance-stabilizing transformation (VST), to calculate variability of the genes.

ANOVA and One-Way ANOVA
The dialog options have been restructured, renamed and improved to be more intuitive and easier to set up differential expression model tests. Output in several drill downs have been improved.

Predictive Modeling

One new process and one new Add-In have been added.

Model Summary and Ensemble ***New!***
This process helps to evaluate results from Predictive Modeling Review and perform ensemble on the selected best models.

Genomic Bayesian Regression ***New!***
This Add-In builds predictive models using the Bayesian methods present in the BGLR package (Pérez and de los Campos, 2013[3]). Genomic Bayesian regression is a form of regularized regression that allows for numerous, potentially correlated, predictors and shrinks them using a common variance component model.

Genomic BLUP
Improved analysis of markers with missing values prevents loss of samples.

XG Boost Regression ***New!***
The new Add-In process builds predictive models using the extreme gradient boosting machine learning method that allows for numerous, potentially correlated predictor variables. Computations are performed using the XGBoost Python Package, and both linear and tree regression can be fitted in this process.

General Utilities

R Package Manager ***New!***
This utility helps to install and uninstall recommended R packages. It also enables you to install newest packages that available on the R CRAN site (https://cran.r-project.org/web/packages/available_packages_by_name.html).

[3] Pérez P, de los Campos G. BGLR: a statistical package for whole genome regression and prediction. R package version 1. 0.2. 2013.