## Q-K Association Analysis

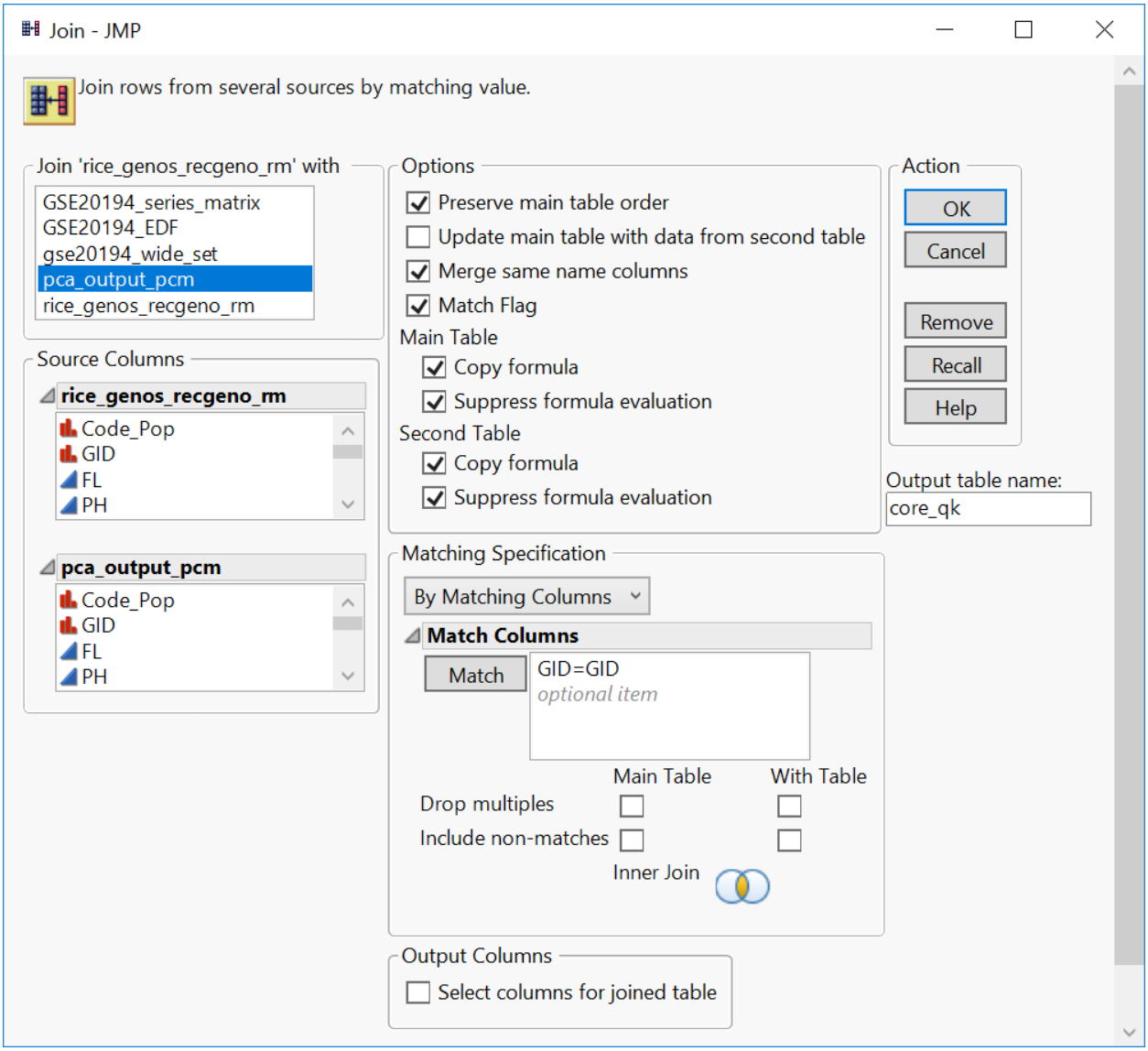
Q-K association analysis was developed to perform association mapping while controlling for population structure and/or familial relatedness. The “Q-K” in this name refers to the two kinds of information that get included in the model. The Q matrix contains information about population structure, which can come from **Multidimensional Scaling, Principal Components Analysis**, or even manual assignment of the lines or individuals into groups curated by the user. The K matrix contains more fine-grained information about relatedness, usually IBD measures calculated from the marker data using the **Relationship Matrix** procedure.

As in other kinds of association mapping, in Q-K association analysis an individual statistical model is created for each marker, using the trait as a dependent variable and the marker as an independent variable. The variables that constitute the Q and K matrices are also included in these models: Q variables as fixed effects, and K variables as random effects. Either the Q or the K variables may be omitted from the models, as desired.

To run Q-K association analysis in JMP Genomics, one needs a datafile containing the traits to be analyzed, the marker genotypes, and the Q and K variables to be used in the analysis. An annotation dataset containing map information and any other marker information can also be used in the analysis – this file is optional, but it is a good idea to include it.

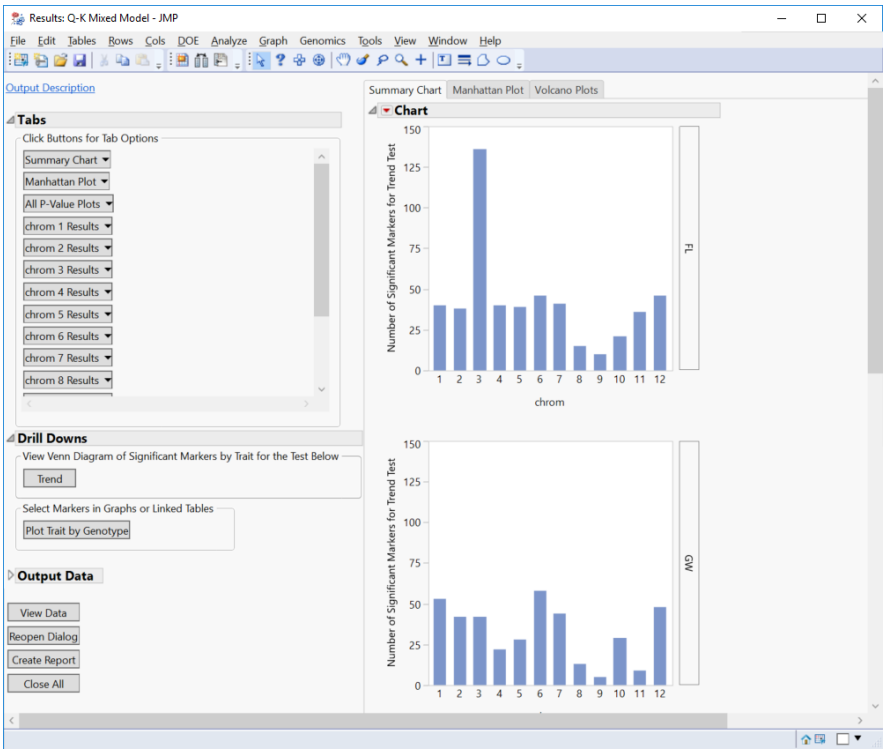
We have created the input data file for the Q-K analysis by joining the results files from the **Relationship Matrix** analysis and the **Principal Components Analysis**, outlined in the *K Matrix and Relatedness Measures Step-Guide* and the *Principal Components Analysis Step-Guide* respectively.

1. Use the **File > Open** command to open the PCA results merged file, called ***pca_output_pcm.sas7bdat***
2. Also **Open** the file with the “square root” of the IBD matrix from the **Relationship Matrix** procedure, called ***rice_genos_recgeno_rm.sas7bdat***.
3. With the file ***rice_genos_recgeno_rm.sas7bdat*** in the foreground, select **Tables > Join**.
4. In the top left box, select the file ***pca_output_pcm.sas7bdat***
5. Check the box labeled **Merge same name columns**
6. Select the variable **GID** in the two **Source Columns** boxes, then click the **Match** button.
7. Assign the **Output table name** as **core_qk**.
8. Click **OK** to create the new joined file.



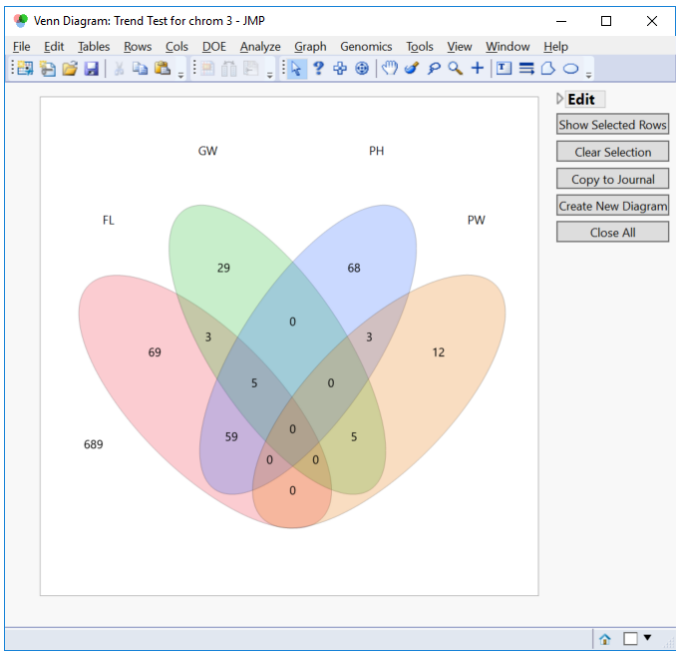

9. When the new data file appears, select **File > Save As…**
10. Change the **Save as type** option to **SAS Data Set**, and click the **Save** button.
11. Choose **Window > Close All** to close all open data files and reports.
12. From the **Genomics Starter** menu, choose **Genetics > Other Association Testing > Q-K Mixed Model.**
13. Choose ***core_qk.sas7bdat*** as the **Input SAS Data Set**.
14. Select **FL, PH, PW, GW** as **Trait Variables**.
15. Enter “recgeno:” in the box under **List-Style Specification of SNP Variables**.
16. Choose an **Output Folder**.

17. On the **Q and K** tab, type “pca:” (without the quotation marks) in the **List-Style Specification of Q Matrix Variables** box, and type “IBD:” (without the quotation marks) in the **List-Style Specification of K Matrix Square Root Variables** box.
18. On the **Model Variables** tab, select **Continuous** for **Type of Trait.**
19. On the **Annotation** tab, Choose **rice_anno_recgeno.sas7bdat** as the **Annotation SAS Data Set**.
20. Fill out the **Annotation** tab with RS as the **Annotation Label Variable**, chrom as the **Annotation Group Variable** and pos as the **Annotation Location Variable.**
21. On the **Options** tab, change the **Format of SNP Variables** to **Numeric Genotypes**.
22. Click **Run** to start the analysis. Note that the process will take some time.
23. When the results dashboard opens, scroll through the charts on the **Summary Chart** page.
    - The summary charts show the number of significant markers on each chromosome, separated by the four different traits
    - The blue bars represent results from the Trend test. The trend test looks for a linear relationship in the trend scores when moving from homozygous minor to heterozygous to homozygous major.

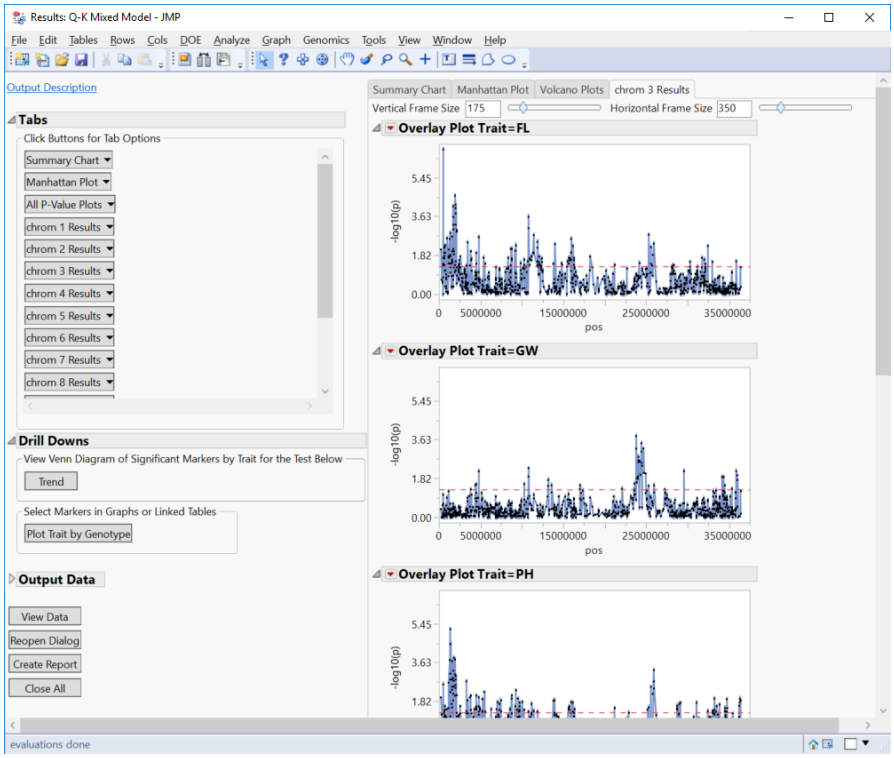



24. To see the significant markers that overlapped across the four traits for the genotype test, click the **Trend** button in the **Drill Downs** section, and select **All**.

- There are 5 markers that are significant in common between GW, FL, and PH. Let's investigate.




25. Click on the blue sector in the Venn diagram with the value "5". This sector is the intersection of the GW, FL, and PH traits.
26. Select **Tables > Subset,** and click **OK** in the new dialog.
    - A new window opens with a data file consisting of 5 rows. These are the markers from the Venn diagram.
27. Back on the **Q-K Mixed Model** results dashboard, click the button labeled **Chromosome Ch_3 Results** and select **View Tab.** This tab shows the significant parkers for each trait by position on chromosome 3.




28. Choose **Tables > Subset** to see information for the association test for markers on a selected trait.

## Summary

This guide outlined the the **Q-K Mixed Model** process, which is a general tool that has a great deal of flexibility, but requires you to construct the input dataset yourself. QK analysis can be run in two different ways in JMP Genomics. . A simpler option is the **Genetics Q-K Analysis Workflow**, which performs all the steps for the **Q-K** analysis and merges the data automatically. It is, however, less flexible than the **Q-K Mixed Model** process: The workflow only allows a Q matrix computed from PCA, and a K matrix from IBD calculations from the **Relationship Matrix** process. For detailed information on this process see the *Q-K Analysis Workflow Step-Guide.* For further steps in the **Q-K Association Analysis** pathway, view the *K Matrix Compression Step-Guide.*