Fwd: [EXTERNAL] Questions about sample labeling Salmon and IGB files

1 message

Gloria K. Muday <muday@wfu.edu> Wed, Apr 19, 2023 at 4:41 PM
To: "Loraine, Ann" <Ann.Loraine@uncc.edu>, Robert Reid <RobertReid@uncc.edu>, Mark Johnson <mark_johnson_1@brown.edu>, Molly Davis <mdavi258@uncc.edu>, Anthony Postiglione <postae18@wfu.edu>, "Ali, Mohammad Foteh" <alimf@wfu.edu>

(Sorry, the last message sent before I was done)

Ann and Molly and Rob,

Anthony and I separated worked through the misordered samples and both came up with the same pattern. In the attached Excel sheet, you can see in the FIRST column what your original sample names were (before that short switch of are). In the SECOND column are what those samples already are.

When you rename the samples, we'd like you to keep the old name on the sheet someone to present in any confusion. What would be really helpful to us is if after you rename the lanes, you could put them in the intended order.

We are mining this data to look at the transcription of specific genes. Ann did note that we should look at normalized reads. We'd love to have such a spreadsheet that had the normalized reads. We'd also love this exported to an excel sheet with the SolyIDs of each gene and the gene names.

Now that we have clear sample identities, we'd love to see if the PCA plot shows groups based on temperature and genotypes!

Thank you for all the effort to try to resolve the problem.

It is unfortunate that it took Azenta so long to send the photo and that they totally ignored the sample names and ran the samples in a random and confused order. If you want to see what happened, you can look at the attached powerpoint file.

Thanks,
Gloria, Anthony, and Foteh

Thanks,
Gloria

On Sun, Apr 16, 2023 at 8:50 PM Ann Loraine <Ann.Loraine@uncc.edu> wrote:

> Hi,
>
> Molly and I are working with this file as the start of the data analysis pipeline work:
> https://bitbucket.org/hotpollen/flavonoid-rnaseq/src/main/ExternalDataSets/muday-144-SL5_salmon.merged.gene_counts.tsv
>
> The order of columns is determined by this configuration file, used to run the pipeline:
> https://bitbucket.org/hotpollen/flavonoid-rnaseq/src/main/ExternalDataSets/muday-144_samples.csv
>
> The above "original salmon" file shows the counts observed for each gene from the fastq files. Samples are named after the file (sample) that they came from.
>
> (Just making sure we're all on the same page with respect to input data files :-)