**BROAD**
INSTITUTE

**Lawrence Chung <lchung@broadinstitute.org>**

## Re: Zher Yin analysis issue on DEL analysis app

**Lawrence Chung** <lchung@broadinstitute.org>                    Tue, May 3, 2022 at 7:43 AM
To: Bruce Hua <b2hua2@gmail.com>
Cc: Paul Clemons <pclemons@broadinstitute.org>

Thank you, Bruce.
When I received from Zher Yin a config file and then a few weeks later a subsequent resequencing config file (see attached), I did not know any better when I saw the same run_ids and samp_ids in both files.

There will be a teaching moment where I inform our users to avoid using the same numbers again in config files and from now on I will do my best to catch that kind of duplication in the config files.

On Mon, May 2, 2022 at 5:43 PM Bruce Hua <b2hua2@gmail.com> wrote:
> Regarding:
> One thought is that it's not sustainable to continue with DEL analysis app's users requests to delete sample metadata and run metadata for whatever reason. Real multi-user web apps don't accept that kind of request from users.
>
> The immediate solution is to instruct users not to reuse run_ids and samp_ids if a sequencing run fails and needs to be resequenced.
>
> The long-term solution is to re-design the app to use a relational database that is implemented with auto-incremented indices as the primary keys of the run and sample metadata, which will allow users to re-use run_ids and samp_ids  when re-doing sequencing runs.
>
> I agree with this general sentiment. It was my initial intent for the app to function in this way but there are many areas for potential improvement. As one example, when resequencing samples to generate new reads (as fastq files), the user is meant to create a new entry for run_id, which auto-defaults to the next available number. I'm still not exactly sure how Zher Yin's issues came to be, but it likely arose from the app seeing metadata that it was improper in some form or another.
>
> On Mon, May 2, 2022 at 10:40 AM Paul Clemons <pclemons@broadinstitute.org> wrote:
>> I agree with both of these points -- can you convert the long-term to a GitHub issue? -- also, is there an easy way to drop this whole conversation thread into a file or something so that we can memorialize it in Slack or GitHub? Best, Paul
>>
>> On Mon, May 2, 2022 at 7:24 AM Lawrence Chung <lchung@broadinstitute.org> wrote:
>>
>>> One thought is that it's not sustainable to continue with DEL analysis app's users requests to delete sample metadata and run metadata for whatever reason. Real multi-user web apps don't accept that kind of request from users.
>>>
>>> The immediate solution is to instruct users not to reuse run_ids and samp_ids if a sequencing run fails and needs to be resequenced.
>>>
>>> The long-term solution is to re-design the app to use a relational database that is implemented with auto-incremented indices as the primary keys of the run and sample metadata, which will allow users to re-use run_ids and samp_ids  when re-doing sequencing runs.
>>>
>>>
>>> On Sun, May 1, 2022 at 2:07 PM Paul Clemons <pclemons@broadinstitute.org> wrote:
>>>> OK, I am finally caught up with emails this quarter (a month in!).
>>>> I read this recent thread with great interest, not to double-check the details (I trust y'all, especially together!), but rather to study the back-and-forth often required to get things done for real -- indeed, I contrast this case with
>>>> Larry's IDR experience (Dahlin manuscript) in that this thread unfolds over 4-5 days

instead of 7-8 weeks in the IDR case (due to asynchronous timezones, in part).

Curious to hear any thoughts either of you might have.

PAC

On Sun, May 1, 2022 at 10:35 AM Lawrence Chung <lchung@broadinstitute.org> wrote:
Great, that analysis files for er, er_lb and er_ub matched the tabulated values in the app. I will move on to checking another other analysis files just to make sure.

On Sat, Apr 30, 2022 at 3:53 PM Bruce Hua <b2hua2@gmail.com> wrote:
Check an001261_lib007_agg07_type01.csv . I might've said agg02 before, but agg007 are the trisynthons and should match with the table.

On Sat, Apr 30, 2022 at 3:44 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
Thank you. The an001261_lib007_agg02_type01.csv analysis file has one column with 290 rows of values.
Are they the first 290 non-zero enrichment values in the table? I am not finding a correspondence between the
values in the analysis file and in the table.

On Sat, Apr 30, 2022 at 2:02 PM Bruce Hua <b2hua2@gmail.com> wrote:
Nice! These enrichment values seem reasonable and are internally consistent.

Another version of the spot check would be to look at the analysis file and see if it matches what you see in the table.

On Sat, Apr 30, 2022 at 1:06 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
Bruce,
Your latest set of advice did the trick. The enrichment values look more sensible (this table's values are sorted by enrichment lower bound (er_lb). Once again, I had re-done the analysis in a duplicate version
of the DEL analysis app.
Before I re-do the analysis in the shared DEL analysis app, how do I manually calculate and spot check
the enrichment values to make sure they are correct?

Larry

On Sat, Apr 30, 2022 at 7:11 AM Bruce Hua <b2hua2@gmail.com> wrote:
If this is the issue, the minimum I think would be to go to the /meta/an_meta.csv file and delete any rows that match in the input file. That would be the first bullet I mention in the other email thread. Then re-run the analysis. The analysis files (like the one you've attached) should be written over when the new enrichments are calculated.

On Fri, Apr 29, 2022 at 11:55 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
Thank you for the clarification. When I looked for the /result/an/an001261_lib007_agg02_type01.csv file (see attached) I realized
that it has a timestamp of April 22, 2022, 2:09 PM. That was when Zher Yin ran the first analysis that had no enrichment values.
Should I have deleted it before re-running the analysis yesterday?

On Fri, Apr 29, 2022 at 10:49 PM Bruce Hua <b2hua2@gmail.com> wrote:
To clarify, could you check the analysis result file? In this case it would be /result/an/an001261_lib007_agg02_type01.csv

And see if those enrichment values match what is seen when you look up the table in the app?

On Fri, Apr 29, 2022 at 9:43 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
I was able to confirm that the targ_ct_lst and ctrl_ct_lst count files matched perfectly with the table's "targ_ct" and ctrl_ct columns.

So, would it be correct to assume that the strange enrichment values were influenced by something else in the metadata?
Let me know where else to look.

Many thanks,
Larry

On Fri, Apr 29, 2022 at 12:59 PM Bruce Hua <b2hua2@gmail.com> wrote:
Okay that one should be an_id = 1261. Let me know what you find.

On Fri, Apr 29, 2022 at 12:54 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
Thank you, Bruce.
I will look into those areas based on what Zher Yin just gave me (i.e, **res_id** = res0033, **target** = 2um BRD9 100uM Alb, **control** = beads only, **library** = 7: triazine library) and will get back to you with either a finding or more questions.

Many thanks,
Larry

On Fri, Apr 29, 2022 at 11:25 AM Bruce Hua <b2hua2@gmail.com> wrote:
> Oh. Do you need to know more than " …it was res0033 that Zher Yin was reporting the problem of large values of 800 for 0 and 0 counts."? Let me know.
Yea for things like this, would you also be able to provide related information like:

- If this is from the app, what parameters have been chosen for the visualization? Such as res_id, target, control, library.
- If this is from a file, just the name of the file should suffice.

> Anyway, I am still learning my way around the files. Which ones are relevant to the app's table and that I should spot check?
What caught my eye is that in addition to the values being abnormally large for counts of 0 and 0, they were also inconsistent between the rows. How can the same counts give rise to different enrichment calculations?

| cycle1 | cycle2 | cycle3 | ctrl_ct | targ_ct | er | er_lb | er_ub |
|---|---|---|---|---|---|---|---|
| All | All | All | All | All | All | All | All |
| 6 | 176 | 110 | 0 | 0 | 823.3 | 88.11 | 2129.8 |
| 33 | 10 | 90 | 0 | 0 | 823.3 | 88.11 | 2129.8 |
| 3 | 209 | 120 | 0 | 0 | 670.12 | 65.87 | 1704.34 |
| 6 | 27 | 246 | 0 | 0 | 670.12 | 65.87 | 1704.34 |
| 15 | 173 | 43 | 0 | 1 | 670.12 | 65.87 | 1704.34 |
| 24 | 258 | 132 | 0 | 0 | 670.12 | 65.87 | 1704.34 |
| 26 | 265 | 139 | 0 | 1 | 670.12 | 65.87 | 1704.34 |
| 28 | 128 | 106 | 0 | 0 | 670.12 | 65.87 | 1704.34 |
| 31 | 261 | 125 | 0 | 3 | 670.12 | 65.87 | 1704.34 |
| 31 | 268 | 246 | 0 | 0 | 670.12 | 65.87 | 1704.34 |

For this, it would be helpful to know what was chosen as the parameters for res_id, target, control, library. You could go to the res003X.csv file and check what an_id is associated with this combination of parameters. Then go and check the actual analysis file to see if the enrichments match what is pulled up in the app table. Just as a first step of debugging which part of the system might be the issue.

On Fri, Apr 29, 2022 at 10:00 AM Lawrence Chung <lchung@broadinstitute.org> wrote:
Oh. Do you need to know more than " …it was res0033 that Zher Yin was reporting the problem of large values of 800 for 0 and 0 counts."? Let me know. Anyway, I am still learning my way around the files. Which ones are relevant to the app's table and that I should spot check?

Many thanks,
Larry

On Fri, Apr 29, 2022 at 9:33 AM Bruce Hua <b2hua2@gmail.com> wrote:
The input looks fine.

It would be easier to troubleshoot if I could see not just the values that are off but which analysis is being looked at. For example the example that Zher Yin shared with the values in the 800s - I don't know what analysis I'm looking at.

One spot check would be to see if the values shown in the table (in the app) match up with files themselves.

On Fri, Apr 29, 2022 at 9:26 AM Lawrence Chung <lchung@broadinstitute.org> wrote:
I attached the analysis input file. What if Zher Yin never clicked on the "Run analysis" button after uploading the analysis input file? Would that explain the first problem with no enrichment values being seen?

For your last question, it was res0033 that Zher Yin was reporting the problem of large values of 800 for 0 and 0 counts.

Sorry, I will be offline for the rest of the morning. I will get back to you if you need anything else from me. Thank you very much again for your guidance and support.

Larry

On Fri, Apr 29, 2022 at 9:12 AM Bruce Hua <b2hua2@gmail.com> wrote:
Can I see the analysis input file that was used for res0033? Even between res0032 and res0033 the metadata doesn't match up. For example, the an_id = 1256 for each of them has different samples for the targ_ct_lst column. It's difficult to think of ways to end up in this situation - one might be if the an_id was included as part of the analysis input file. In this case res0033 matches up with an_meta.csv whereas res0032 doesn't.

I can double check, but I think the script does the following:

- Take the user-provided analysis input file
- Read in the existing analysis metadata (an_meta)
- For each row of the input file, if the combination of targ_ct_lst, ctrl_ct_lst, agg_id, an_type, and ci already exists, then assign that an_id to the row
- For everything else, assign it a new an_id and perform the analysis and generate a new analysis file (the ones with the enrichment ratios)

I'm pretty sure if the analysis is not in the metadata, the script will still go ahead and run the analysis and replace any existing files. I'm not sure how res0032 could have been run, and then res0033 run and end up with this metadata.

What do the enrichment values look like now? Are there still large values of 800 for 0 and 0 counts?

On Fri, Apr 29, 2022 at 6:50 AM Lawrence Chung <lchung@broadinstitute.org> wrote:
Thank you for taking a look. The latest re-analysis was res0033. Sorry for not telling you that.

On Thu, Apr 28, 2022 at 10:24 PM Bruce Hua <b2hua2@gmail.com> wrote:
There seems to be stuff missing from an_meta.csv. IDs 560 through 574 are missing.

| 555 | run031_samp000679_lib002.csv, run031_samp0006 | run031_samp000655_lib002.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 554 |
| 556 | run031_samp000679_lib004.csv, run031_samp0006 | run031_samp000655_lib004.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 555 |
| 557 | run031_samp000679_lib005.csv, run031_samp0006 | run031_samp000655_lib005.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 556 |
| 558 | run031_samp000679_lib006.csv, run031_samp0006 | run031_samp000655_lib006.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 557 |
| 559 | run031_samp000679_lib007.csv, run031_samp0006 | run031_samp000655_lib007.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 558 |
| 560 | run031_samp000679_lib100.csv, run031_samp0006 | run031_samp000655_lib100.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 559 |
| 561 | run031_samp000657_lib002.csv, run031_samp0006 | run031_samp000655_lib002.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.99 | 575 |
| 562 | run031_samp000657_lib004.csv, run031_samp0006 | run031_samp000655_lib004.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.99 | 576 |
| 563 | run031_samp000657_lib005.csv, run031_samp0006 | run031_samp000655_lib005.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.99 | 577 |
| 564 | run031_samp000657_lib006.csv, run031_samp0006 | run031_samp000655_lib006.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.99 | 578 |
| 565 | run031_samp000657_lib007.csv, run031_samp0006 | run031_samp000655_lib007.csv, run031_samp | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.99 | 579 |

This corresponds to the first half of the analyses according to res0032.csv:

| targ_ct_lst | ctrl_ct_lst | targ_description | ctrl_description | agg_id | an_type_i | ci | an_id |
|---|---|---|---|---|---|---|---|
| run029_samp000639_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 0uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 560 |
| run029_samp000641_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 2uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 561 |
| run029_samp000643_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 5uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 562 |
| run029_samp000645_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 10uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 563 |
| run029_samp000647_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 50uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 564 |
| run029_samp000649_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 100uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 565 |
| run029_samp000653_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 2um BRD9 524uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 567 |
| run029_samp000641_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 2uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 568 |
| run029_samp000643_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 5uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 569 |
| run029_samp000645_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 10uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 570 |
| run029_samp000647_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 50uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 571 |
| run029_samp000649_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 100uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 572 |
| run029_samp000651_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 300uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 573 |
| run029_samp000653_lib007.csv | run029_samp000639_lib007.csv | 2um BRD9 524uM Alb | 2um BRD9 0uM Alb | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 574 |
| run029_samp000640_lib007.csv | run029_samp000637_lib007.csv, run029_samp000638_lib007.csv | 5um BRD9 0uM Alb | beads only | 1, 2, 3, 4, 5 | 1, 2, 3 | 0.95 | 1256 |

I don't know how this could happen. Do the actual analysis files with enrichment values exist? Again it seems like something that could be fixed by redoing the analysis.

On Thu, Apr 28, 2022 at 10:10 PM Lawrence Chung <lchung@broadinstitute.org> wrote:

> Thank you for the offer to look into the metadata. The meta folder is over 25 MB in size so the attached zip file is actually a link to the file in Google Drive.

📄 del_metadata.zip

On Thu, Apr 28, 2022 at 8:01 PM Bruce Hua <b2hua2@gmail.com> wrote:

> Yea that should never be the case. Based on your last email I think the issue might be in the metadata. If you want to send me all the metadata I can take a look at it to see if I can find anything.
>
> On Thu, Apr 28, 2022, 7:07 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
>
>> Paul and Bruce,
>>
>> Zher Yin raised an interesting question about enrichment values when there are no counts.
>>
>> *I took a closer look at some data and ranked according to er_lb. it seems to give the highest values for some building blocks with 0 counts? This seems odd to me..*

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Table** | | | | | | | | | |
| Show 10 ∨ entries | | | | | | | | Search: | |
| cycle1 | cycle2 | cycle3 | | ctrl_ct | targ_ct | er | er_lb ▾ | | er_ub |
| All | All | All | All | All | All | All | All | | All |
| 6 | 176 | 110 | | 0 | 0 | 823.3 | 88.11 | | 2129.8 |
| 33 | 10 | 90 | | 0 | 0 | 823.3 | 88.11 | | 2129.8 |
| 3 | 209 | 120 | | 0 | 0 | 670.12 | 65.87 | | 1704.34 |
| 6 | 27 | 246 | | 0 | 0 | 670.12 | 65.87 | | 1704.34 |
| 15 | 173 | 43 | | 0 | 1 | 670.12 | 65.87 | | 1704.34 |
| 24 | 258 | 132 | | 0 | 0 | 670.12 | 65.87 | | 1704.34 |
| 26 | 265 | 139 | | 0 | 1 | 670.12 | 65.87 | | 1704.34 |
| 28 | 128 | 106 | | 0 | 0 | 670.12 | 65.87 | | 1704.34 |
| 31 | 261 | 125 | | 0 | 3 | 670.12 | 65.87 | | 1704.34 |
| 31 | 268 | 246 | | 0 | 0 | 670.12 | 65.87 | | 1704.34 |
| Showing 1 to 10 of 5,655,000 entries | | | | | | Previous 1 2 3 4 5 ... 565500 Next | | | |

On Wed, Apr 27, 2022 at 11:20 PM Bruce Hua <b2hua2@gmail.com> wrote:

> That's great! Hopefully it works on the online app with the updated metadata. I can't quite piece together what the consequences of using the incorrect run_id would be, but since data tables and metadata tables are joined together at various stages in the app, that could be giving rise to the issue if there's incompatible tables being joined.
>
> On Wed, Apr 27, 2022 at 11:13 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
>
>> Hi Bruce,
>>
>> Thank you, that was great advice because while preparing for re-running Zher Yin's analysis I noticed that the run_meta.csv has not been updated since March 30, 2022.
>>
>> ```
>> (base) –bash:chembio-dev-02:/chembio/
>> datasets/proj/del/meta/run 1013 $ ls –lt
>> total 800
>> -rw-rw-r-- 1 cbshiny cbshinyusers 321958 Apr 27 21:10
>> ct_file_meta.csv
>> -rw-rw-r-- 1 cbshiny cbshinyusers   2119 Mar 30 11:07
>> run_meta.csv
>> (base) –bash:chembio-dev-02:/chembio/
>> datasets/proj/del/meta/run 1014 $
>> ```
>>
>> The last 10 rows in that run_meta.csv file include run_id #29, 220303_Albumin_CIP-DEL_Opt:

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | run_id | seq_type | seq_location | seq_completion_d | fastq_folder_name | |
| 24 | 23 HiSeq | NIBR | 6/23/21 | 210623_titration_v1v2_BAF_HiSeq | | |
| 25 | 24 MiSeq | NIBR | 8/6/21 | 210806_cip_del_opt_and_ha_idh1 | | |
| 26 | 25 HiSeq | NIBR | 8/30/21 | 210806_CiP-DEL_opt_and_HA-IDH1-HiSeq | | |
| 27 | 26 MiSeq | NIBR | 10/5/21 | 211005_MiSeq_macrodomain_superpool_and_CiP-DEL | | |
| 28 | 27 MiSeq | NIBR | 10/7/21 | 211007_MiSeq_macrodomain_superpool_and_CiP-DEL_Repeat | | |
| 29 | 28 HiSeq | NIBR | 10/25/21 | 211025_HiSeq_macrodomain_superpool_and_CiP-DEL | | |
| 30 | 29 MiSeq | Xavier lab | 3/3/22 | 220303_Albumin_CIP-DEL_Opt | | |
| 31 | 30 MiSeq | Broad | 3/10/22 | 220310_DYRK1A_BRDs | | |
| 32 | 31 HiSeq | Broad | 2/24/22 | 220224_Choudhary_DARPA_DEL_Screen_1-5-HiSeq | | |
| 33 | 32 HiSeq | Broad | 3/28/22 | 220328_DYRK1A_BRDs | | |

>> However, last Friday Zher Yin submitted 220422_Albumin_CIP-DEL_Opt that was also marked as run_id #29. Wouldn't that cause an issue with the analysis?
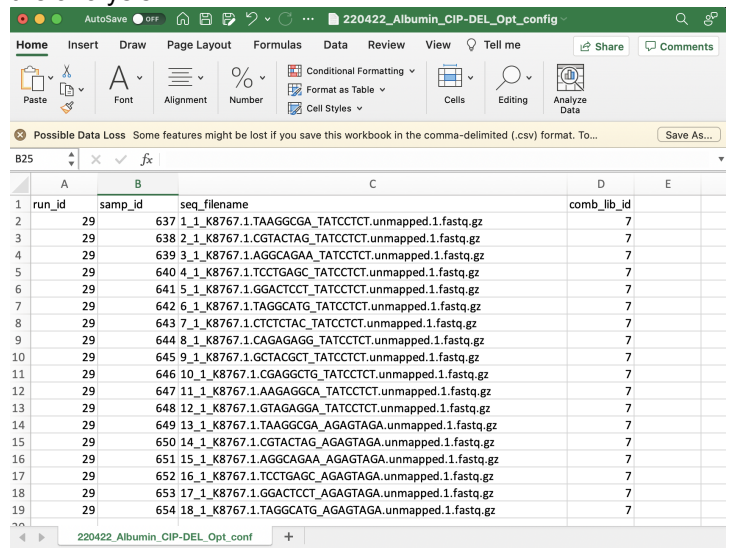
Anyway, I took your advice but re-ran the analysis on my local laptop's DEL analysis app instead
of the shared app. (The local run_meta.csv file did not have the older run_id #29).

No errors popped up during the analysis. Only 232 of the same warnings appeared in the console:

```
Warning: Problem with `mutate()` column `value`.
?~D? `value = get_er(...)`.
?~D? NaNs produced
```

Happily, after the analysis finished I retrieved a table for target 2um BRD9 100uM Alb that had enrichment values (unlike the screenshot that Zher Yin gave us earlier today).

| cycle1 | cycle2 | cycle3 | ctrl_ct | targ_ct | er | er_lb | er_ub |
|--------|--------|--------|---------|---------|------|-------|--------|
| 45 | 154 | 140 | 2 | 71 | 18.1 | 6.53 | 137.58 |
| 45 | 289 | 140 | 1 | 64 | 28.2 | 8.1 | 1048.06 |
| 45 | 279 | 140 | 1 | 50 | 22.07 | 6.28 | 820.8 |
| 45 | 289 | 240 | 0 | 41 | 66.45 | 9.46 | 182.8 |
| 40 | 279 | 140 | 3 | 38 | 6.85 | 2.71 | 32.14 |
| 45 | 287 | 140 | 1 | 38 | 16.81 | 4.72 | 526 |
| 45 | 279 | 9T | 11 | 32 | 1.71 | 0.91 | 3.64 |
| 40 | 279 | 9T | 20 | 29 | 0.87 | 0.5 | 1.57 |
| 45 | 173 | 140 | 0 | 29 | 47.18 | 6.61 | 129.31 |
| 11 | 273 | 140 | 14 | 28 | 1.19 | 0.64 | 2.36 |
| 31 | 279 | 140 | 1 | 28 | 12.43 | 3.42 | 463.67 |
| 63 | 279 | 9T | 27 | 25 | 0.56 | 0.32 | 0.96 |
| 11 | 276 | 140 | 11 | 24 | 1.29 | 0.66 | 2.79 |
| 64 | 279 | 9T | 29 | 24 | 0.5 | 0.29 | 0.86 |
| 42 | 279 | 140 | 2 | 23 | 5.93 | 1.99 | 45.76 |
| 53 | 279 | 140 | 2 | 23 | 5.93 | 1.99 | 45.76 |
| 66 | 279 | 140 | 10 | 23 | 1.36 | 0.67 | 3.05 |
| 67 | 279 | 9T | 25 | 23 | 0.55 | 0.31 | 0.98 |
| 70 | 279 | 9T | 20 | 23 | 0.69 | 0.38 | 1.27 |
| 76 | 287 | 140 | 0 | 23 | 37.54 | 5.19 | 102.57 |
| 11 | 274 | 140 | 9 | 22 | 1.44 | 0.7 | 3.38 |
| 29 | 279 | 140 | 4 | 22 | 3.08 | 1.25 | 11.42 |
| 63 | 273 | 9T | 38 | 22 | 0.35 | 0.2 | 0.59 |
| 71 | 279 | 9T | 25 | 22 | 0.53 | 0.3 | 0.94 |
| 11 | 23 | 140 | 5 | 21 | 2.4 | 1.02 | 7.61 |

Showing 1 to 25 of 5,655,000 entries          Previous  1  2  3  4  5  ...  226,200  Next

And I was able to get a 3D plot
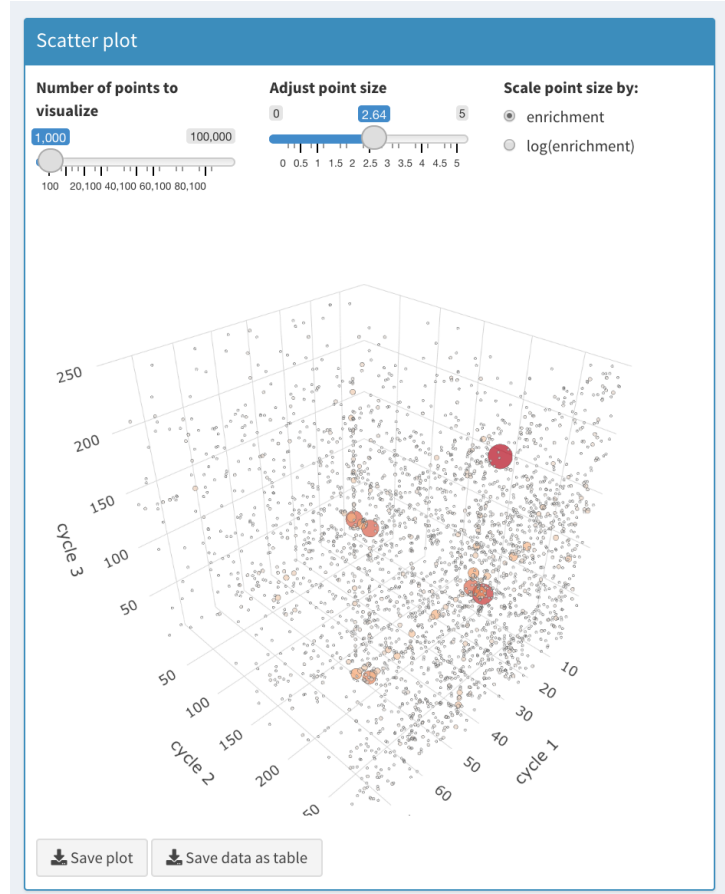


Tomorrow, I will try to replicate this success on the shared version of the DEL analysis app.

Thank you very much for your patient guidance and help.
Larry

On Wed, Apr 27, 2022 at 6:21 PM Bruce Hua <b2hua2@gmail.com> wrote:
> It might still be an issue with the app. But if it is an issue with the app, the issue occurred during the running of the analysis.
>
> A quick potential troubleshoot could be to rerun the analysis using the same analysis input file and keep an eye out for any errors that pop up during that time period.
>
> On Wed, Apr 27, 2022 at 6:18 PM Lawrence Chung <lchung@broadinstitute.org> wrote:
> > Bruce, what are your thoughts or advice about the absence of enrichment values where there are control counts and total counts? I need to explain to Zher Yin that there is not a failure of the app. Let me know what

I can search for you.
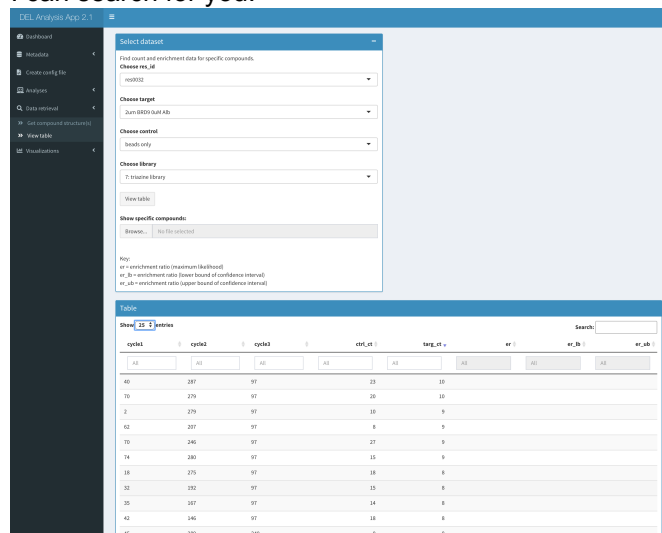


Many thanks,
Larry


On Wed, Apr 27, 2022 at 2:10 PM Lawrence Chung
<lchung@broadinstitute.org> wrote:
> Thanks Bruce, but there is no confusion. The point is
> that the 3D plot problem does occur for all the
> targets in Zher Yin analysis #res0032.
>
> So yes, the screenshot I captured and shared on
> Apr 26, 2022 at 11:49 PM was for target 2um BRD9
> 0uM Alb.   But today I shared Zher Yin's  screenshot
> of  target 2um BRD9 100uM Alb table, which he first
> sorted in descending order for "targ_ct" to claim that
> there were sufficient counts for that target. It also has
> the same 3D plot problem anyway.
>
> Many thanks,
> Larry
>
> On Wed, Apr 27, 2022 at 1:51 PM Bruce Hua
> <b2hua2@gmail.com> wrote:
>> I think this table might be different from the one
>> you shared before, and is different from the input
>> used to (attempt to) visualize the 3D plot. For the
>> 3D plot the target was chosen as 2um BRD9 0uM
>> Alb. And the table you shared in your second
>> email has many more 0 counts than the one you're
>> sharing now. Is there perhaps a confusion with the
>> selection of the targets and controls?
>>
>> On Wed, Apr 27, 2022 at 1:13 PM Lawrence
>> Chung <lchung@broadinstitute.org> wrote:
>>> Thank you, Bruce, for analyzing that data file
>>> and for the great explanation.
>>>
>>> Yet, Zher Yin had sent me this screenshot and
>>> claimed that it shows good target and control
>>> counts for enrichment calculations. But isn't it
>>> still true that because the vast majority (97%) is

NA, therefore there can be no enrichment calculations?



On Wed, Apr 27, 2022 at 11:59 AM Bruce Hua <b2hua2@gmail.com> wrote:

> I'm not positive why the app would error out, since there are values that it could potentially plot. That being said, the calculated enrichments are quite low for this analysis, and also with the vast majority (97%) as NA, meaning there were no counts observed in target or control. Of the remaining, 1.8% of the barcodes have a value of 0 (meaning no counts were observed in the target but were observed in the control samples), only 0.008% of the barcodes (482 total) have enrichments above 1, and only 0.0001% of the barcodes (6 total) have enrichments above 2. So regardless of the reason why the app is unhappy with this as input, it's unlikely that there's enough resolution to do any sort of analysis with these data. More counts are needed.

```
# A tibble: 19 x 3
    value         n        pct
    <dbl>     <int>      <dbl>
 1  0        101570    1.80
 2  0.0131        1    0.0000177
 3  0.0168        3    0.0000531
 4  0.0234       82    0.00145
 5  0.0385     1858    0.0329
 6  0.110     51033    0.902
 7  0.176        16    0.000283
 8  0.272       225    0.00398
 9  0.380         2    0.0000354
10  0.571        22    0.000389
11  0.636      5458    0.0965
12  0.891         4    0.0000707
13  1.23        427    0.00755
14  1.56          1    0.0000177
15  1.85         48    0.000849
16  2.48          4    0.0000707
17  3.11          1    0.0000177
18  4.37          1    0.0000177
19  NA       5494244  97.2
```

On Wed, Apr 27, 2022 at 10:30 AM Lawrence
Chung <lchung@broadinstitute.org> wrote:

> Sorry, I grabbed the wrong files to show
> you. So, in
> an001270_lib007_agg07_type02.csv (see
> attached), the values do look atypical
> because most of them are "NA". Does that
> make sense therefore that no enrichments
> could be calculated?
>
> Many thanks,
> Larry
>
> On Wed, Apr 27, 2022 at 9:39 AM Bruce
> Hua <b2hua2@gmail.com> wrote:
>
>> If that's the relevant analysis ID, can you
>> check if the agg07 files exist? agg07
>> corresponds to trisynthons which is what
>> would be used for the 3D boxplot. So
>> something like
>> an001270_lib007_agg07_type02.csv
>>
>> On Wed, Apr 27, 2022 at 9:36 AM
>> Lawrence Chung
>> <lchung@broadinstitute.org> wrote:
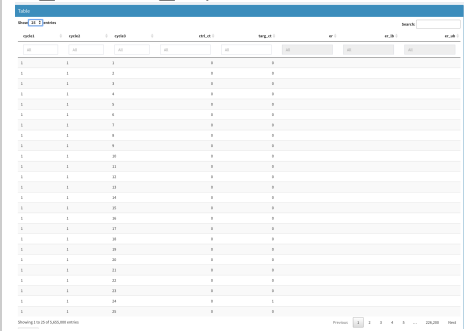>>
>>> Hi Bruce,
>>>
>>> Just to follow up, Zher Yin had asked if
>>> the enrichment calculations could
>>> possibly return nothing if values are
>>> under a certain threshold. However,
>>> these aggregate files seem to have
>>> typical values, for example. Is there
>>> anything else I should be checking?
>>>
>>> Many thanks,
>>> Larry
>>>
>>> On Tue, Apr 26, 2022 at 11:49 PM
>>> Lawrence Chung
>>> <lchung@broadinstitute.org> wrote:
>>>
>>>> Thank you for the valuable tip. I did
>>>> as you suggested and see that there

are no enrichment ratio values (er, er_lb and er_ub)



Many thanks,
Larry

On Tue, Apr 26, 2022 at 10:07 PM
Bruce Hua <b2hua2@gmail.com>
wrote:

> Hi Larry,
>
> It might be difficult to troubleshoot
> from my end, but here's an
> attempt. I think the error that's
> popping up can happen when you
> ask R to take a max of something
> that doesn't exist.
>
> Can you try to use the Data
> retrieval > View table tab to look at
> what the input would be for that 3D
> boxplot? Using the same input of
> res_id, target, control, and library.
> If the above is true then you would
> probably see no rows associated
> with that combination of inputs, or
> maybe it'll error out if the analysis
> file doesn't exist.
>
> Thanks,
> Bruce
>
>
> On Tue, Apr 26, 2022 at 9:33 PM
> Lawrence Chung
> <lchung@broadinstitute.org>
> wrote:
>
>> Hi Bruce,
>>
>> I hope all is going well with you.
>>
>> In the DEL analysis app, Zher
>> Yin ran into a problem with
>> getting a 3D box plot of counts
>> from his MiSeq sequencing run.
>> I tried troubleshooting but cannot
>> find the cause.  Zher Yin's data
>> and metadata looked fine to me.
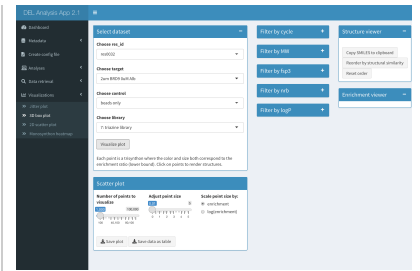
1. The following was written to the log file when the visualization failed for Zher Yin:

```
Warning: Error in seq.default:
'from' must be a finite number
  176: stop
  175: seq.default
  168: renderPlot
[/local/shiny-server/del-app/
app.R#809]
  166: func
  126: drawPlot
  112: <reactive:plotObj>
   96: drawReactive
   83: renderFunc
   82:
output$dashboard_screens_
timeline
    1: runApp
Selecting by scaled_size
Joining, by = c("cycle1",
"cycle2", "cycle3")
Warning: Unknown or
uninitialised column:
`scaled_size`.
Selecting by scaled_size
Selecting by scaled_size
Selecting by scaled_size
Joining, by = c("cycle1",
"cycle2", "cycle3")
Warning: Unknown or
uninitialised column:
`scaled_size`.
Selecting by scaled_size
Selecting by scaled_size
Warning: Unknown or
uninitialised column:
`scaled_size`.
Selecting by scaled_size
Selecting by scaled_size
Selecting by scaled_size
Warning: Unknown or
uninitialised column:
`scaled_size`.
Selecting by scaled_size
Selecting by scaled_size
Selecting by scaled_size
```

2. When I replicated the problem on a copy of the app on my local computer, the following was written to the output when the visualization failed:

```
    1: runApp
|------------------------------
----------------------|
|=============================
===================|
|------------------------------
----------------------|
|=============================
==================|
```

```
|----------------------------
--------------------|
|============================
==================|
|----------------------------
--------------------|
|============================
==================|
Joining, by = c("cycle1",
"cycle2", "cycle3")
|----------------------------
--------------------|
|============================
==================|
|----------------------------
--------------------|
|============================
==================|
|----------------------------
--------------------|
|============================
==================|
|----------------------------
--------------------|
|============================
==================|
|----------------------------
--------------------|
|============================
==================|
Warning in max(rv$rv_3d_er_sp_
interactive_tbl$value, na.rm =
T) :
  no non-missing arguments to
max; returning -Inf
Warning: The 'plotly_click'
event tied a source ID of
'plotly_3d_er_interactive' is
not registered. In order to
obtain this event data, please
add `event_register(p,
'plotly_click')` to the plot
(`p`) that you wish to obtain
event data from.
Warning: The 'plotly_click'
event tied a source ID of
'plotly_3d_er_interactive' is
not registered. In order to
obtain this event data, please
add `event_register(p,
'plotly_click')` to the plot
(`p`) that you wish to obtain
event data from.
Warning: The 'plotly_click'
event tied a source ID of
'plotly_3d_er_interactive' is
not registered. In order to
obtain this event data, please
add `event_register(p,
'plotly_click')` to the plot
(`p`) that you wish to obtain
event data from.
```

3. The error condition seems to be located around this R code in the app:

```
  observe({
    req(rv$rv_3d_er_sp_
interactive_tbl)
    input$visualize_3d_er_sp_
interactive_scale
    input$visualize_3d_er_sp_
interactive_scale_log_radio
```

```
if(input$visualize_3d_er_sp_
interactive_scale_log_radio ==
"enrichment") {

rv$rv_3d_er_sp_interactive_tbl
<-
rv$rv_3d_er_sp_interactive_tbl
%>%options(shiny.error =
browser)
        mutate(scaled_size =
value*input$visualize_3d_er_
sp_interactive_scale)
```

4. I attached Zher Yin's generated stats file if it helps show that the reads and counts look normal.

5. No other sequence count analysis has this same problem with 3D plot visualization.

Please let me know what can fix this problem or where else I need to check.

Many thanks,

--
Larry Chung
Senior Software Engineer
Chemical Biology & Therapeutics Science
Broad Institute of MIT and Harvard
415 Main Street
Cambridge, MA 02142
lchung@broadinstitute.org

--
Larry Chung
Senior Software Engineer
Chemical Biology & Therapeutics Science
Broad Institute of MIT and Harvard
415 Main Street
Cambridge, MA 02142
lchung@broadinstitute.org

--
Larry Chung
Senior Software Engineer
Chemical Biology & Therapeutics Science
Broad Institute of MIT and Harvard
415 Main Street
Cambridge, MA 02142
lchung@broadinstitute.org

--
Larry Chung
Senior Software Engineer
Chemical Biology & Therapeutics Science
Broad Institute of MIT and Harvard
415 Main Street
Cambridge, MA 02142
lchung@broadinstitute.org

**2 attachments**

**220422_Albumin_CIP-DEL_Opt_config.csv**
2K

**220303_Albumin_CIP-DEL_Opt_config.csv**
1K