

Subject: RE: MS data

From: Maria Lluch Senar <maria.lluch@crg.eu>

Date: 29/11/16 15:09

To: Katerina Kirsanova <catherine.kirsanova@crg.eu>, Luis Serrano Pubul <luis.serrano@crg.eu>, Samuel Miravet Verde <samuel.miravet@crg.eu>

Dear Katerina,

-The format of the current tables is agreed with Luis.

About your comments:

1. the existence of each sheet should pay for itself ??? **Please, explain it better**
2. each column of each sheet in the XLS has to be described **(at the moment we should maintain the name in the column and associate the number of the project)**, there should not be messy repeats **(there should not be, let me know what is for you messy repeats)**
3. there should be a clear description how different sheets are corresponding to each other **(this information is not at the moment necessary for the task you have to perform if you understood what you should do. If you did not, please come and ask)**
4. the amount of sheets should be optimized and brought to the reasonable number **(the number is the one of the projects that we have. Do you have a limit?)**
5. there should not be sheets repeating each other in format **(show me an example please)**
6. there should not be empty sheets **(there is one empty sheet because we are waiting for the results of MS)**

As I said if you need information about what is each column or data in the files just come and ask. E-mail is not the proper way to explain all the details of the files.

-If you think that changing the pipeline to consider the new format is not possible please let us know and we will find the way to implement it.

Thanks,
Maria

From: Katerina Kirsanova

Sent: martes, 29 de noviembre de 2016 14:52

To: Maria Lluch Senar <maria.lluch@crg.eu>; Luis Serrano Pubul <luis.serrano@crg.eu>; Samuel Miravet Verde <samuel.miravet@crg.eu>

Subject: Re: MS data

El 11/29/16 a las 14:22, Maria Lluch Senar escribió:

Dear Katerina

The format of the tables is not the same that Luis described because it is more difficult to standardize from the raw tables of MS.

We already have several ones like that and they can be processed as it was done before.

However, the important information of proteomics data is there: peptide sequences identifiers and amounts (areas). I have simplified very much all the files to make it easier.

I am sorry, but SummaryInfoPaper_MS & SummaryInfoPaper_MSForValues cannot be processed: they are not brought to the format/standardization level as it was done before:

1. the existence of each sheet should pay for itself
2. each column of each sheet in the XLS has to be described, there should not be messy repeats
3. there should be a clear description how different sheets are corresponding to each other
4. the amount of sheets should be optimized and brought to the reasonable number
5. there should not be sheets repeating each other in format
6. there should not be empty sheets

If you think it's necessary to change the format I would suggest you to take a look at Luis' doc: it is written correspondingly to what the standardization is and can be taken as a lorem ipsum if you'd like to create your own format description.

So please, consider this format for the webpage and come to my office and ask me if you do not understand MS data.

I understand. I have asked a professional to consult me. And I've got all comprehensive information I needed. Thanks.

Thanks

From: Katerina Kirsanova
Sent: martes, 29 de noviembre de 2016 14:14
To: Maria Lluch Senar <maria.lluch@crq.eu>; Luis Serrano Pubul <luis.serrano@crq.eu>; Samuel Miravet Verde <samuel.miravet@crq.eu>
Subject: Re: MS data

Dear Maria,

I see that some XLS files in the folder are now brought to the described in Luis' doc format, so they can be processed.

El 11/29/16 a las 13:25, Maria Lluch Senar escribió:

Dear Katerina,

I have been working in collecting the information of MS that should be implemented in the webpage. I have formatted the tables corresponding to the different project, extracting the information that is important to include in the pipeline.

In this link you will find (a part of the raw tables of MS) two excel files that contain all the data that is critical to represent the information that Luis was asking in the meeting we had:

<https://www.dropbox.com/sh/1ggtcnt0hg6hpzk/AADXxcOzoZtCTIK4NLoozh9Xa?dl=0>

1) SummaryInfoPaper_MS:

In each sheet there is the information of identified peptides with their corresponding area for each sample (header of each column) of each project (number in the sheet). The last sheet contains the descriptions for each project that each member of the lab will update but at the moment is not relevant for what you should represent in the webpage. Also links to the MS raw files and MS original tables are added in this sheet.

This cannot be taken into account because the information here is not consistent with what is described in the format mentioned above.

2) SummaryInfoPaper_MSForValues: this file contains similar information than before. But if the peptide is detected it has a 1, if it is not detected a 0.

This table can be used to extract the information of:

- Number of samples in one project (P_X) where the peptide is found
- Number of peptides identified for one identifier (id) in each sample (each column of a sheet).
- Number of projects where the same peptide has been found and which ones