

Subject: Re: NCBI's starts/ends vs Table_annotation_latest.xls

From: Catherine <catherine.kirsanova@crg.eu>

Date: 01/12/16 17:39

To: Maria Lluch Senar <maria.lluch@crg.eu>, Luis Serrano Pubul <luis.serrano@crg.eu>, Eva Yus Najera <eva.yus@crg.eu>

CC: Carolina Gallo Lopez <Carolina.Gallo@crg.eu>, Veronica Llorens Rico <Veronica.Llorens@crg.eu>, Samuel Miravet Verde <samuel.miravet@crg.eu>, Eduard Sabido Aguade <eduard.sabido@crg.eu>, Guadalupe Espadas-Garcia <guadalupe.espadas@crg.eu>, Bioinformatics Unit <BioinformaticsUnit@crg.eu>

Dear all,

Please discuss the format in open and constructive manner.

As there is a strict standardization and DB processing rule «Excel is forbidden» (see)

the format should be a reasonable amounts of tab-separated text files uploaded one-by-one.

Each of this text files should not exceed 10Mb to be uploaded and processed in a reasonable time.

Each column of each file should be described in detail.

File names should have meaning and described in details otherwise text field in the upload interface should be filled in.

Current processing and analysis is done for 2-spreadsheet Excel files (no more than 10Mb of size)

as a trade-off for accurate format description (done by Luis) of a **standard files of results coming from a proteomics experiment** and my good will to make all my best for the lab.

Processing of Excel files into the DB with analysis on the fly is not a common practice and generally is not accepted

due to obstruct capacity to support the standards and analysis as consistent and reasonable prompt.

So CSV only (XML/JSON for those biologists who want die hard is an option too).

However, once the standard is created and described the way it was done by Luis and with the participation/by a professional IT specialist it is still possible to make sure that service is scalable and reliable.

Thanks, —
Catherine

El 12/01/16 a las 17:00, Maria Lluch Senar escribió:

| Dear all, |

Since all the MS searches were not run in the same format and using the same parameters of Mascot, we agreed with Guadalupe that we will try to repeat the searches in the files. Guadalupe has to discuss with Eduard if it is possible to do that and when. The output tables will have a format different to the one that Luis documented in the word. I will explain in a meeting once we have agreed on that with the facility. Then, the database should be updated to read this format.

Cheers,

Maria

From: Katerina Kirsanova
Sent: jueves, 01 de diciembre de 2016 16:50
To: Maria Lluch Senar <maria.lluch@crg.eu>; Luis Serrano Pubul <luis.serrano@crg.eu>; Eva Yus Najera <eva.yus@crg.eu>
Cc: Carolina Gallo Lopez <Carolina.Gallo@crg.eu>; Veronica Llorens Rico <Veronica.Llorens@crg.eu>; Samuel Miravet Verde <samuel.miravet@crg.eu>; Eduard Sabido Aguade <eduard.sabido@crg.eu>; Guadalupe Espadas-Garcia <guadalupe.espadas@crg.eu>; Bioinformatics Unit <BioinformaticsUnit@crg.eu>
Subject: Re: NCBI's starts/ends vs Table_annotation_latest.xls

File is attached, sorry for the inconvenience.

El 12/01/16 a las 16:46, Catherine escribió:

is implemented according to Luis' format document (please see it attached)
and the example of this implementation is available by this link:
<http://mycodb.crg.es:5000/data/MS.8cf249511536cc4b0243f781e2121776-0>
This service is fully ready for tests, comments and using.

I am happy to meet with all the parts and discuss in open and constructive format the further development of the DB and the analysis.

Thanks, —
Catherine

of unique peptide identified in all the experiments that we have track
Cheers
Maria

-----Original Message-----

From: Luis Serrano Pubul
Sent: jueves, 01 de diciembre de 2016 12:42
To: Eva Yus Najera <eva.yus@crg.eu>
Cc: Katerina Kirsanova <catherine.kirsanova@crg.eu>; Carolina Gallo Lopez
Subject: RE: NCBI's starts/ends vs Table_annotation_latest.xls

Thanks i will take a look later

Luis Serrano
Director
EMBO member
CRG-Centre for Genomic Regulation
Tel. +34 933160101
Luis.serrano@crg.eu

-----Original Message-----

From: Eva Yus Najera
Sent: miércoles, 23 de noviembre de 2016 10:22 p.m.
To: Luis Serrano Pubul <luis.serrano@crg.eu>
Cc: Katerina Kirsanova <catherine.kirsanova@crg.eu>; Carolina Gallo Lopez
Subject: Re: NCBI's starts/ends vs Table_annotation_latest.xls

hi

please find attached the revision of the genes that were different from

I think it is self-explanatory, but anyway there is this traffic-light

I also attach the Table_annotation.xls with the new TSCs (as we discussed)

I add the TSS_mapping... table also as i had to correct some stuff concerning

Please check and tell me if this is all fine

I must say i tried to be very conservative and inclusive, not to lose
It would be nice if eventually these protein variants are included in the

best
e