

Subject: Project

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

Date: 28/09/16 14:23

To: Katerina Kirsanova <catherine.kirsanova@crg.eu>

CC: Luis Serrano Pubul <luis.serrano@crg.eu>

Dear Katerina,

I have talked with Luis and he said me that he has given you the opportunity to work for three months in the lab and to finish the project.

I do not know if you have done the tasks that I asked in my previous mail. We need to know when you are planning to have them. We expect them at the begging of next week.

We can sit and discuss if you have any doubt or question. As I said, please, just propose when you want to meet.

As reminder:

- 1) Then for the comparative. We need a different output file. We would like having an excel file with this information:
 - a. id of gene
 - b. number of compared genomes where the protein has been found
 - c. Category of conservation:
 - i. Mycoplasma: if it is conserved in other Mycoplasmas different to M. pneumoniae
 - ii. Gram positive: if it appears conserved in Bacillus subtilis and or Lactobacillus
 - iii. Gram negative: if it appears conserved in E. coli
- 2) Also, the last section that we discussed long time ago about the pipeline is the representation of the results.
 - a. There should be a section of Analysis of smORFs and it should represent in histograms or pies the results of:
 - i. Conservation of smORFs (%of the total putative smORFs that can be classified in any of the three categories)
 - ii. Transcriptionally active (% of the total putative smORFs that show significant levels of RNA)
 - iii. ncRNAs (% of ncRNAs that could encode for proteins (ORF inside); % of real ncRNAs). Of each section of the pie there should be two subpies representing for each category: the % that overlap in antisense with annotated ORFs (A) the % that overlap in sense (S) and the % that are in intergenic regions (IR).

Thanks

Maria