

CCB Bioinformatic Survey

Theo Killian

2/18/2021

Introduction

The heads of labs were surveyed with a list of questions in order to assess the state of bioinformatic needs at CCB, and to ascertain analytical expertise and infrastructure use.

What bioinformatic analyses does your lab primarily perform?

- Max - We primarily perform single cell RNAseq, ATACseq, and CRISPR/Cas screening.
- Sarah - We primarily perform bulk RNAseq, spatial transcriptomics, MS/proteomic and metabolomic analyses.
- Anna - We primarily perform RNAseq, proteomic analyses and docking.
- Georg - We will do many bulk and single cell RNAseq analyses. Mainly from mouse tissues but a few fly and human also.
- Patrizia - We primarily perform FACS, lipidomics and image analysis.

What are types of analyses or infrastructure that you may require assistance with?

- Max - We do often Crop-seq. This is a need: so far Barnard is doing all the analysis but cannot keep doing this. We have also large datasets of histone marks, metabolites and so on.
- Sarah - Juan is doing the analysis in my lab. We also will do spatial metabolomics (mass spec imaging).
- Anna - We are planning on also doing 5C and Riboseq, as well as comprehensive splicing analysis.
- Georg - We will not have enough bioinformatic capacity in the future.
- Patrizia - Yes, we are planning in near future (next months) scRNAseq analysis (tumor EC) and proteomics (lysosomes, organelles in general), lipidomics also in tissue and metabolomics (in coll with Bart and Sarah)

Where do members of your lab typically perform bioinformatic analyses?

- Max - Lab members perform analyses on their own laptop.
- Sarah - Juan on the VSC, the rest currently local but this is expected to change.
- Anna - Lab members mostly use the VSC.
- Georg - I have one student that is doing this. She is very good at it but will leave in about 2 years I expect. And with the increasing load she may not have enough time.
- Patrizia - Lab members perform analyses on their own laptop.

Where do you typically store or archive your research data?

- Max - KU Leuven L-Drive.
- Sarah - KU Leuven L-Drive.
- Anna - We are using L- and K-drives, external drives. All published data are submitted to public repositories.
- Georg - I think we use a server and the account of Stein Aerts.
- Colinda - KU Leuven L-Drive.
- Patrizia - KU Leuven L- and K-drives.

Do you experience any problems with data storage in the near future?

- Max - Not really.
- Sarah - In the future our data is too large, and we need a solution for this.
- Anna - Storage is more problematic and expensive.
- Colinda - I foresee that data storage will become a problem for us in the near future. Right now we are storing our data on the network drive (L-drive) of the KU Leuven. However, within the first months we have reached the limit of the storage we bought and had to expand to 20TB (which costs us 100euros/TB/year). We are very close to the 20TB limit again and we have noticed that the network drive is very slow. If we continue to acquire data with this speed or even more (since the lab will for sure grow in the next months/years.), the costs of our data storage will become problematic. Therefore, we are now considering to purchase a NAS system for offline data storage. But of course it would be safer to have an online solution in the end, so we are open to other solutions as well.
- Patrizia - Yes because we often experience L- and K- drives getting out of service.

If a new dedicated CCB node was created, would members of your lab use it?

- Max - I think so: usage and need are in constant increase.
- Sarah - Yes, mainly Juan.
- Anna - We are happy with the VSC service. There are some limitations, in terms of time or the number of running jobs. On the other hand, it is affordable, they are advising on the job optimization and the maintenance of the VSC is ok.
- Georg - Yes!
- Patrizia - It will be nice to have a dedicated storage location for department or lab (if possible) where members can easily access the data without actually altering the original data. It depends on the use and output a user wants from it (as the new tool demands more time to learn how to use it, unless it's very user friendly) and unless it's going to give some additional outputs.

Summary

- Mostly labs appear to perform bulk RNAseq and single cell RNAseq, as well as other types of analyses which are unique to each lab.
- Most labs who responded to the survey seemed to indicate that they suffer from a lack of bioinformatic capacity.
- Labs either rely on local laptops or VSC for computational tasks.
- Use of the L- and K- drives is common, however there appear to be problems with data storage of large data files and a more economical solution is necessary.

- Many labs who use the VSC for larger computational tasks indicate that there are limitations such as jobs timing out as well as limitations on numbers of cores, etc. If a dedicated CCB server/node was made available, labs indicate that it would be put to good use.