**Assignment 1:**

Read papers 1–5 and write a reflection, on perhaps two pages, answering the following questions:

* What are these authors in agreement on?
* For each paper, what does it contribute relative to the other papers?

**Answer:**

The following papers were read and a reflection is written about them :

* Paper 1: [A Quick Guide to Organizing Computational Biology Projects](http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000424) by William Stafford Noble.
* Paper 2: [Ten Simple Rules for Reproducible Computational Research](http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003285) by Sandve *et al.*
* Paper 3: [So you want to be a computational biologist?](http://www.nature.com/nbt/journal/v31/n11/full/nbt.2740.html) by Loman & Watson.
* Paper 4: [Best Practices for Scientific Computing](http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745) by Wilson *et al.*
* Paper 5: [Ten Simple Rules for the Open Development of Scientific Software](http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002802) by Prlić & Procter.

The papers stated above are written for the support of the young computational biologists to stratify, plan, execute their research plans in a well organized way .

The authors in the papers do agree upon several things pertaining to the research activity especially for a computational biologist. The utmost important things they all agree upon are:

1. Well defined structure: whether it is analysis or production of new bioinformatics tool, the things pertaining to it should be well organized and structured in a way that helps the researcher to handle errors and make complete utilization of the resources such as computers.

The first paper “A quick Guide to organizing computational Biology projects” suggests how to organize the project information in a well defined structure and also introduces concepts of using a electronic lab note book , execution of a single computer experiment . These serve as basics for a computational biologist in order to produce valuable research. The remaining papers emphasize the importance of having such an organization of the project.

1. Replicability and Reproducibility: a well-written code when it is not reproducible by yourself or not understood by others is of no use for future research practices. So every computational biologist needs to write his scripts and codes in such a fashion that it is easily replicated by you and reproducible & understood by others who want to read and execute them.

Once the files are organized in a well-structured format as specified in paper I, the reproducibility and replicability which are the corner stones of the scientific research specified in the paper II becomes quite easy. Avoiding manual manipulation steps and tracking of every time point become the pavement stones for the best practices fro scientific computing.

1. Use of Version control like platforms: it is very important to maintain and coordinate a complex bioinformatics research. And this is only possible if you store the whole project as a repository and can have a log record of errors & results. And version control also helps collaborators to work simultaneously. The version control tools available now are: github, bitbucket, Gitlab.

The version control due to its above stated advantages helps out a computational biologist to overcome the situations of repeating the codes for different kinds of data and to track the exact results in the future. This part of the version control is introduced in the paper I, emphasized about it in paper II, III and paper IV, thus laying the correct path for becoming a successful computational biologist.

1. Constructive criticism and judgment: As a scientist it is quite important to understand the difference between being a programmer and researcher . As a researcher you have to be critical in understanding goals of the project and to choose which tool & language are appropriate for a given aim.

Understanding well organization, reproducibility, version control is necessary, but how does an aspiring novice computational biologist get the idea of these? It is achieved by usage of appropriate sources of learning such as MOOC’s, problem solving, web forums, blogs and lists. These concepts were introduced by paper III. A special emphasis is also given about the term “pipeline” which is often used term by every bioinformatician. One should actually understand when it is necessary to construct such pipeline in terms of saving time & its use to others.

1. Storage: every kind of data whether it is a raw file, plot, code , script , written documents must be stored appropriately and chronologically for easy understanding, reproducibility , replicability. Every author specified in above papers does agree that storage in well-executed manner makes an immense impression of the research that is being conducted.
2. Follow principles: Taking in a famous quote: “obey the principles without being bound by them”(Bruce lee) , there are always different ways of performing a designated research , but one should always keep in mind the principles & practices in order to do a valuable research in time. Every author of the paper laid out principles and practices for better scientific computing but do also advise on being flexible at every time point of the research and development. This emphasis is clearly described in paper IV: “Best practices for scientific computing “

Once we have learnt all the above stated 6 points, every computational biologist like me should understand that a computational biology research needs to be simple revisable, readable, replicable, reproducible, recordable. Thus the work doesn’t stop there, this research tool should be available for open development in future. This is possible by following the rules laid out in the paper V: “[Ten Simple Rules for the Open Development of Scientific Software](http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002802) “. The list of rules are quite simple .one should always look for the tool if it has been already produced and available. Simplicity & transparency in the tool are requirements achieved through using version control tools. Perfectionism should not be the main motto for writing any script.

As a researcher, we have to be always ready to collaborate with other researchers and promote the project tool, which ensures value beyond the lifetime by continuous maintenance of the toll either through sponsors or through more advanced scientific academic writing.