

3C (Computational Cytometry Club / Club de cytométrie computationnelle)

Intro:

Flow and mass cytometry are more and more expanding their multidimensional capacity posing a real challenge in the data analysis process that was till now based on technologies developed when the hardware was able to measure 4-8 parameters at the time. The scientific community promptly responded to the challenge with new algorithms, software and analysis procedures. The pace of production of these multiparametric analysis methodologies pose a significant challenge for biologists in this area of complexity. In absence of a standard and established methodological pipeline aiming to reduce the amount of information measured to the most valuable parameters, new algorithms should be considered, tested and validated before their careful use in biological studies. This comprehensive and necessary effort is time consuming and bio-informaticians currently focused on cytometry data analysis are rather scattered and isolated in France. Critical evaluation of state-to-the-art analysis methodologies should be performed in a collaborative work between biologists and bioinformaticians. Current evaluation work is scarce and replicated across several laboratories slowing down the transfer of these methods to biologists. In order to shorten validation time of each methods and transfer information to biologists in multiple research centers, the AFC has decided to create a Computational Cytometry Club with the aim to regroup these isolated efforts on data analysis, structure the type of tests that will be performed at the national level and select challenging reference datasets that are meaningful for different kind of biological questions.

Aims:

- to regroup bioinformaticians and biologists interested in computational cytometry analysis around a virtual table
- to create a community of advised users via regular teaching initiatives
- to create reference material for everyone who wants to start analysis of complex flow cytometry data
- to benchmark on a collaborative basis new analysis methodologies in a rational way. This work will be done in a distributed manner in the frame of virtual workshops
- finally, to share with the broad AFC community the results of these tests in the frame of a parallel session at the annual meeting of AFC in order to diffuse the advantage/limits of each approach and raise the overall knowledge of the french cytometry community

Who

- Bioinformaticians and biologists from the research as well as clinical field
- Open to AFC and non AFC members, irrespective of their institute or affiliation

We propose to structure the work of 3C in the frame of the following work packages:

WP1: Building up the network

An advertisement of 3C should be made in early Q2 2017 to reach out the cytometrist community and collect sites that are willing to actively participate:

1. send a general email about current aims and expectation of 3C
2. advertising on AFC website
3. use of respective network of 3C leaders

Following/concomitant of this announcement, a Survey Monkey should be set in order to sense what are the current expectation of french scientists in the field of Computational Cytometry and categorize each need in supervised and unsupervised analysis approaches. Both research and clinical laboratories should give a feedback / feeling about their need for future transfer to the their work (especially for clinic)

Sites that will engage in this work should take active part in the discussion regarding WP2.

WP2: Identification of reference datasets

Reference datasets should be representative of a given scientific question. WP1 is of outmost importance as it will be used for all subsequent algorithm benchmarking

1. Real datasets with open access to the whole community
 - a. hematology with normal and unexpected profile
 - b. rare cell events
2. Synthetic datasets / simulated datasets
 - a. the synthetic dataset should simulate a real blood samples with populations representing "real" proportions.

WP3: Cloud environment of datasets for efficient collaborative work

A set of networking tools should be set to promote data sharing, interaction and discussion.

1. Build a cloud environment to share data / code / protocols / ideas
 - a. Cloud repository will be opened to all agreed participants
 - b. Should be link to a forum/ mailing list to attract more people/share answers to common problems
 - c. Protocols that have clearly defined and approved could be disseminated using MOOC.
2. The infrastructure will be based on services provided by Renater (Réseau National de télécommunications pour la Technologie l'Enseignement et la Recherche). There are two options:
 - a. Sourcesup (<https://sourcesup.renater.fr/>) is a robust, neutral and national forge able to host code and protocols. Sourcesup is dedicated to code development but it offers many interesting companion services. Documents

can be used for reports and protocols. Wiki can guided the users through the documents. Mailing list can allow sharing ideas and organising meetings.

- b. Universaliste (<https://groupes.renater.fr/>) is a mailing list management tool. It can be linked with a Wiki, the same as Sourcesup probably. A Wiki is a quick way to share and present information and ideas. Documents can also be hosted.
3. Hosting of FCS datasets via flow repository account (access rights must be checked). Alternatively, ImmPort could also be evaluated.

WP4: Benchmarking work

Reference datasets should be representative of a given scientific question. WP2 is of utmost importance as it will be used for all subsequent algorithm benchmarking.

1. Evaluation / testing of new tools as they are available
Participant profile: bioinformaticians and biologists are welcome. Whereas computational skills are required to deal with R code or other code made available online, some methods require very few computational skills and propose a graphical user interface. Scientists at the bench are the final end users of the tools and the ones that could really validate analyses on non-synthetic datasets.
2. Testing / establish pipeline of analysis to become references for the French community
3. Develop a R-Shiny interface for biologists of these R-tools if not available and share the source code through an identified web portal.
4. Evaluation of bio-statistic analysis.

Organization of the collaborative work

Several milestones should be reached out quickly to beneficiate from the momentum at the national level brought by the Inserm Workshop and the joint annual meeting of AFC/SFI in Reims.

An email should be sent out in Q2 to recruit/identify all sites interested in 3C AFC initiative. Each sites will be free to use their collaborative system (Dropbox, Onedrive, Cytobank...) to maximize their interaction according to the disponibilities of the participants (e.g.: the coordination team uses Google Drive to create documents and share information).

Following the evaluation of the needs of each sites, the production of a work agenda with deliverables will be shared according to the rules established by the WP4. FCS files should be made public following each WP milestone through FlowRepository as cloud storage.