

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

Intro:

Les techniques de cytométrie de flux en fluorescence et de masse augmentent de plus en plus leur capacité multi-dimensionnelle. Ceci pose un réel défi à l'analyse de données traditionnelle qui était adaptée jusqu'alors à des jeux de données de 4 à 8 paramètres. La communauté scientifique a réagi rapidement à ce défi en proposant de nouveaux algorithmes, des logiciels et des procédures d'analyses. La cadence d'apparition de ces méthodologies d'analyse multiparamétrique pose un défi significatif pour les biologistes dans ce domaine complexe. En l'absence d'un pipeline méthodologique standardisé et établi visant à réduire la quantité d'information mesurée aux paramètres les plus pertinents, les nouveaux algorithmes doivent être pris en main, testés et validés avant une utilisation prudente dans les études biologiques. Cet effort complet et nécessaire prend du temps et les bio-informaticiens en France actuellement dédiés à l'analyse de données de cytométrie sont plutôt dispersés et isolés. L'évaluation critique de l'état-de-l'art des méthodologies d'analyse devrait être réalisée dans le cadre d'un travail collaboratif entre biologistes et bio-informaticiens. L'actuel travail d'évaluation est limité et répliqué au sein de plusieurs laboratoires, ce qui ralentit le transfert des méthodes vers les biologistes. Dans le but de réduire le temps de validation de chaque méthode et de transférer l'information aux biologistes des centres de recherches, l'AFC a décidé de créer un Club de Cytométrie Computationnelle (3C) avec pour objectif de regrouper ces efforts isolés sur l'analyse de données, de structurer le type de tests qui seront réalisés à l'échelle nationale et de sélectionner des jeux de données de référence pertinents qui couvriront différents types de questions biologiques.

Buts:

- regrouper les bio-informaticiens et biologistes intéressés dans l'analyse computationnelle de cytométrie autour d'un réseau
- créer une communauté d'utilisateurs avisés via des sessions régulières de formation
- créer un matériel de référence pour quiconque souhaite se lancer l'analyse de données complexes issues de cytométrie de flux
- évaluer collaborativement les nouvelles méthodologies d'une façon rationnelle. Ce travail sera effectué de manière distribuée dans le cadre de groupes de travaux virtuels
- partager avec la communauté large de l'AFC les résultats de ces tests dans le cadre d'une session parallèle au congrès annuel de l'AFC de façon à diffuser les avantages et limites de chaque approche et d'accroître les connaissances de la communauté française de cytométrie

Qui:

- Bio-informaticiens et biologistes du domaine de la recherche autant que de la clinique
- Ouvert à tous (membres de l'AFC ou non), quel que soit leur institution ou leur affiliation

Nous proposons de structurer le travail du 3C AFC dans le cadre des work packages suivants:

WP1: Architecture du réseau

Une communication du 3C AFC sera effectuée au début du second trimestre 2017 afin d'informer la communauté de cytométrie et de dénombrier les sites qui souhaitent participer activement:

1. envoyer un email général décrivant les objectifs et attentes du 3C
2. communiquer sur le site web de l'AFC
3. utiliser les réseaux respectifs des coordonnateurs du 3C

Parallèlement à cette annonce, un sondage sera établi en vue de cerner les attentes des scientifiques français au sujet de la cytométrie computationnelle et répertorier chaque besoin en analyse supervisée ou non-supervisée. Les laboratoires de la recherche et de la clinique pourront lister à cette occasion leur besoin en analyse de données et le type de problématique qui les intéresse.

Les sites qui s'engageront dans ce travail collaboratif devront prendre une part active dans la discussion concernant le WP2.

WP2: Identification de jeux de données de référence

Les jeux de données de référence devront représenter une question scientifique donnée. WP2 est de la plus grande importance parce qu'il sera utilisé pour l'analyse comparative des algorithmes.

1. Jeux de données réelles (vraies expériences réalisées) disponibles à toute la communauté
 - a. hématologie avec des profils normaux et anormaux/inattendus dans le cadre de leucémie par exemple
 - b. évènements rares
2. Jeux de données synthétiques (expériences simulées)
 - a. un jeu de données synthétique devra simuler des vrais échantillons de sang avec une représentation réaliste des populations et de leur proportion.

WP3: Solutions de partage de données pour un travail collaboratif efficace

Un ensemble d'outils de travail collaboratif sera mis en place de façon à promouvoir le partage de données, l'interaction et la discussion.

1. Construire un espace internet pour partager les données / codes / protocoles / idées
 - a. un dépôt sur internet sera ouvert à tous les participants agréés
 - b. un forum/ mailing list permettra de regrouper plus de personnes et de partager les réponses à des problèmes communs
 - c. les protocoles qui auront été clairement définis et approuvés pourront être diffusés à travers un MOOC.
2. L'infrastructure sera basée sur les services proposés par Renater (Réseau National de télécommunications pour la Technologie l'Enseignement et la Recherche). Deux options sont possibles:
 - a. Sourcesup (<https://sourcesup.renater.fr/>) est un dépôt de code et de documents robuste, neutre et national pour héberger le code de tests et les protocoles. Sourcesup is dédié au développement de programme mais offre plusieurs services d'accompagnement intéressants. Les documents peuvent héberger les rapports et les protocoles. Le Wiki peut guider the utilisateurs vers les documents qu'ils cherchent. Une mailing list peut permettre de partager des idées et d'organiser des télé-conférences.
 - b. Universaliste (<https://groupes.renater.fr/>) est un gestionnaire de mailing listes. Un espace Wiki peut être activé, le même que Sourcesup probablement. Un Wiki est façon rapide de partager et de présenter des informations et des idées. Des documents y être hébergés aussi.
3. L'hébergement de jeux de données FCS pourra s'effectuer sur un compte/espace FLOWRepository (<https://flowrepository.org/>, les droits d'accès doivent être vérifiés). Alternativement, ImmPort (<https://importgalaxy.org/>) pourra être évalué.

WP4: Travail d'évaluation

Les jeux de données de référence doivent être représentatifs d'une question scientifique donnée. WP2 est crucial en ce sens.

1. Évaluer / tester les nouveaux outils quand ils deviennent disponibles
Profil des participants : bio-informaticiens et biologistes sont bienvenus. Alors que des compétences en programmation sont requises pour faire fonctionner du code R ou autre disponible en ligne, certaines méthodes ne demandent que très peu de connaissances en programmation (voire aucune) et proposent une interface utilisateur graphique. Les scientifiques à la paillasse sont les utilisateurs de ces outils et donc ceux qui peuvent valider réellement les analyses des jeux de données non-synthétiques.
2. Tester / établir un pipeline d'analyse pour être la référence dans la communauté française.
3. Développer une interface R-Shiny à l'attention des biologistes pour les outils R n'en proposant pas et partager le code source au travers d'un portail web identifié.
4. Evaluer les analyses d'un point de vue bio-statistique.

Organisation du travail collaboratif

Plusieurs étapes devraient être atteintes rapidement afin de bénéficier de la conjoncture actuelle à l'échelle nationale du fait de l'atelier Inserm et du congrès annuel joint AFC/SFI à Reims.

Un email devra être envoyé au deuxième quarter Q2 afin d'identifier tous les sites intéressés par l'initiative 3C de l'AFC. Chaque site sera libre d'utiliser son propre système collaboratif (Dropbox, Onedrive, Cytobank...) afin de maximiser son interaction en fonction des disponibilités des participants (e.g.: l'équipe de coordination utilise Google Drive pour créer des documents et partager l'information).

En suivant les besoins énoncés par chaque site, un agenda de travail sera établi et partagé afin de réaliser les tests du WP4 et de rendre progressivement le résultat des tests. Après évaluation du besoin de chaque site partenaire de 3C AFC, un agenda de travail sera établi et discuté avec les membres. Des bornes temporelles de livraison de résultats d'une évaluation seront décidées afin que cet effort soit effectivement réalisé et mis en oeuvre de la manière la plus efficace possible. Les fichiers FCS requis dans chaque WP seront rendus publics à travers la solution d'archivage dans le nuage FLOWRepository.