

Introduction to bioinformatics (INFO-F-208)

Article scientifique

Charlotte Nachtegaele

Septembre 2020

Lors des trois premiers projets, vous avez implémenté un sous-ensemble de méthodes communément utilisées dans l'analyse de séquences biologiques. Le dernier projet consiste à utiliser ces outils implémentés sur des données réelles et à rédiger un article avec vos résultats, observations et conclusions en suivant la démarche scientifique.

1 Problème abordé

Le gène CFTR (Cystic fibrosis transmembrane conductance regulator) code pour un canal transmembranaire des cellules épithéliales (notamment présentes dans les pancréas, intestins et voies respiratoires, ainsi que toutes les glandes exocrines). Le canal est responsable de la régulation des ions et en conséquence du maintien de l'osmose des cellules.

De nombreuses mutations à ce gène sont liées à la mucoviscidose. Ces mutations peuvent avoir des conséquences différentes comme empêcher le bon repliement 3D de la protéine, impacter l'activité du canal ou même la localisation de la protéine dans la cellule. Cela provoque une rétention d'eau dans les cellules qui a divers effets sur le corps incluant des problèmes respiratoires dû à des membranes respiratoires trop épaisses, des problèmes d'infertilité, ainsi que le blocage du duct pancréatique entraînant une insuffisance pancréatique.

Vous allez recevoir des séquences de la protéine CFTR avec diverses mutations, ainsi que la séquence originale. Ces séquences devront :

1. Être alignées globalement entre elles et avec l'originale. Quelles sont les séquences mutées les plus similaires à l'originale ?
2. Être alignées localement avec la séquence originale. Arrivez-vous à trouver les domaines-clés de la séquence originale dans les séquences mutées ?
3. Être alignées avec une PSSM construite avec les séquences du domaine AAA (partie du domaine ABC transporter dans la protéine CFTR).

Pour obtenir le fichier d'alignement multiple que vous aviez reçu dans votre projet 2, j'ai déjà fait le travail de récupérer les positions des domaines AAA à travers le site InterPro et de récupérer les séquences grâce aux positions et noms des protéines sur Uniprot sous forme de fichier FASTA. Le code pour obtenir les données et les données sont disponibles dans le dossier `get_AAA`. Les séquences obtenues dans le fichier `AAA_domain_uniprot.fasta` doivent être alignées avec un outil d'alignement multiple de votre choix (CLUSTAL, MUSCLE, etc.), disponibles sur le web. Obtenez-vous les mêmes domaines qu'indiquer sur le site d'Uniprot ?

4. Avoir leur structure secondaire prédite. Quelles sont les différences entre la séquence originale et les mutées ? Quel pourrait être l'impact de ces différences selon le cas ?

Vous êtes libres d'utiliser d'autres ressources ou faire des tests/expériences supplémentaires pour compléter vos observations ou développer vos hypothèses. Les 4 points ci-dessus sont cependant obligatoires.

En résumé, vous devez essayer avec des outils simples de bioinformatique et les séquences de protéines saine et mutées d'expliquer comment les mutations peuvent être à l'origine de la maladie.

2 Format de l'article

L'article doit être rédigé grâce au template \LaTeX à disposition sur l'UV. Vous êtes libres d'utiliser des packages supplémentaires, mais vous ne pouvez pas changer les dimensions des pages ou l'aspect général du template. L'article **doit** faire entre 4 et 6 pages, (éventuelles) références et figures non comprises.

La structure de votre article devra suivre le traditionnel IMRED : **I**ntroduction, **M**atériel et **M**éthodes, **R**ésultats **E**t **D**iscussion.

3 Conseils de rédaction

Le template contient (déjà) de nombreux conseils pour rédiger un article scientifique. Ci-dessous, vous trouverez quelques conseils plus spécifiques à notre cours.

Imaginez-vous écrire cet article pour un étudiant de BA3 qui n'a pas suivi le cours de bioinformatique. Que devez-vous écrire pour que cet étudiant puisse comprendre, suivre et reproduire votre démarche ?

3.1 Introduction

Cette section établit les connaissances nécessaires pour comprendre le contexte de l'article (par exemple la biologie de base) et explique le but de votre rapport. Quel est le sujet ? Quel est le problème que vous voulez résoudre ? Introduisez également brièvement comment vous comptez faire pour le résoudre.

3.2 Matériel et méthodes

Matériel

Typiquement, cette section contient les informations sur les jeux de données, les sites internet utilisés, etc. Vous devez décrire comment et où vous avez obtenu les données. Pour les sites internet, particulièrement ceux hébergeant les bases de données, il est souvent pertinent de spécifier la date d'accession, ou la version de la base de données utilisée.

Méthodes

Les méthodes doivent contenir suffisamment d'explications pour que le lecteur soit capable de reproduire vos résultats. Pour faire cela, des pseudocodes ou formules mathématiques sont souvent nécessaires. Vous pouvez également utiliser des outils en ligne, dont vous devez dès lors citer l'article et expliquer brièvement l'intérêt et le principe.

Les méthodes doivent être arrangées dans un ordre logique et/ou chronologique, c'est-à-dire dans l'ordre dans lequel vous allez présenter vos résultats, selon le fil conducteur de votre démarche.

3.3 Résultats

Dans cette section, vous présentez les résultats obtenus avec la méthodologie présentée à la section précédente. Ces résultats sont souvent sous la forme de tableaux ou graphiques. Portez un détail tout particulier à vos légendes : elles doivent contenir une description suffisante de votre figure, de telle façon à ce que l'on sache au minimum comment la figure a été obtenue. Vous devez décrire ce que vous avez observé, comparer par rapport à la littérature ou à ce que vous vous attendiez obtenir (hypothèse, etc.).

3.4 Discussion

Vous commentez vos résultats. Avez-vous obtenu ce que vous vous attendiez ou pas ? Pourquoi ? Quelle est votre explication ou hypothèse découlant de vos résultats ? Que suggèreriez-vous de faire pour confirmer vos soupçons

ou théories? Vous devez offrir des pistes de réflexion sur ce que vous avez observé dans les résultats.

3.5 Conclusion

Résumez ici le travail fait, depuis la question qui était posée, les méthodes utilisées pour y répondre, vos observations et conclusion sur la question.