

Module 3 Semaine 2 - Sources de variabilité Génétique

A) DÉPANNAGE

Les variantes génétiques peuvent être divisées en différents types selon la taille, la cellule d'origine, la fréquence, la position du génome ou l'impact sur la protéine. Dans cet exercice, nous découvrons les types de variantes génétiques les plus importants. Nous nous entraînerons également à utiliser deux navigateurs du génome.

En génomique, nous nous référons à la variante génétique comme une différence dans la séquence d'ADN d'un individu par rapport au génome de référence.

1. Il s'agit de la séquence d'un patient dans une région à 10 nucléotides (positions 103, 271, 300 à 103, 271, 310 dans le chromosome 12 dans hg19): CTCTGAATGG

Utilisez un navigateur de génome (par exemple <https://genome.ucsc.edu/index.html>) pour afficher cette région (ATTENTION: assurez-vous toujours que vous utilisez le même assemblage de génome, dans ce cas hg19).

Obtenez cette séquence du génome de référence et comparez les deux séquences pour voir s'il y a une différence entre elles. S'il y a une différence (c'est-à-dire une variante génétique), déterminez si elle est:

- a) Un variant nucléotidique unique (SNV)
- b) Une courte insertion ou suppression (indel)
- c) un changement structurel

2. Il s'agit de la séquence du patient dans une autre région (positions 10 000 040 à 10 000 050 dans le chromosome sept, hg19): TTTAATATCTA. Voyez-vous ici une mutation utilisant la même procédure que dans la première question?

De quel type de mutation s'agit-il dans ce cas?

3. Les deux variantes génétiques mentionnées ci-dessus dans un seul gène? (Ce gène est-il lié à une maladie génétique?)

4. Pouvez-vous dire si la variante du chromosome 19 se trouve dans un exon ou dans un intron? (Zoom arrière dans le navigateur)

5. Pouvez-vous utiliser les informations du navigateur du génome pour déterminer si ce gène est codé dans le brin négatif ou dans le brin positif? (Entrez le nom du gène dans le navigateur et zoomez sur les régions de codage).

6. Regardez le fichier joint. Nous avons la séquence complète de CYP21A2, un gène responsable de l'hyperplasie surrénale congénitale, pour deux patients.

Pour déterminer si le patient présente une mutation dans ce gène, comparez-la à la séquence de référence. Pour ce faire, téléchargez la séquence complète du gène sur Ensembl (<https://www.ensembl.org/index.html>). Du côté du gène, cliquez sur les coordonnées, puis sur Exporter les données. Enregistrez la séquence dans un fichier texte qui contient à la fois les exons et les introns du gène (si vous exportez des données directement du côté du gène, vous obtiendrez les différentes transcriptions) (VIDÉO D'ÉCRAN POUR TÉLÉCHARGER LA SÉQUENCE).

Utilisez un outil d'alignement en ligne pour comparer la séquence de chaque patient avec la référence (par exemple, MUSCLE sur <https://www.ebi.ac.uk/Tools/msa/muscle/>).

La ou les mutations décrites sont-elles homozygotes ou hétérozygotes chez chaque patient?

Pouvez-vous déterminer le type de mutation chez les patients A et B?

a) Insertion

b) Suppression

c) la translocation

d) Inversion

Pouvez-vous prédire l'impact de ces mutations sur la protéine? Vous pouvez prendre les nucléotides qui composent votre mutation et utiliser l'outil feuille dans le navigateur du génome UCSC pour déterminer la position des séquences. En zoomant dans la région du gène, vous pouvez visualiser les séquences de nucléotides et d'acides aminés (VIDEO).

B. ACTIVITÉ D'APPRENTISSAGE

L'analyse génétique pour le diagnostic et la recherche nécessite l'intégration d'informations provenant de différentes sources. Comprendre les effets moléculaires d'une mutation particulière ou le schéma de transmission d'un syndrome sont des étapes nécessaires, mais souvent insuffisantes pour tirer des conclusions. Nous pourrions avoir besoin de découvrir quels gènes sont liés à une maladie ou à un phénotype particulier, quelles variantes génétiques ont été montrées pour déclencher un syndrome, ou pour trouver plus d'informations sur un ensemble spécifique de gènes ou de mutations pour déterminer leur possible rôle d'une maladie génétique.

Dans cet exercice, nous utiliserons des exemples pour pratiquer et apprendre comment obtenir des informations utiles à partir des principales bases de données publiquement disponibles sur la génétique humaine.

Commençons par l'hérédité mendélienne en ligne chez l'homme (OMIM), un catalogue constamment mis à jour des gènes humains et des troubles et traits génétiques, avec un accent particulier sur la relation gène-phénotype.

Ouvrez un navigateur et accédez à l'OMIM (<https://www.omim.org/>) Choisissez une maladie humaine. Par exemple, la fibrose kystique (FK).

- Lisez la FAQ OMIM (question 1.3, <https://www.omim.org/help/faq>) pour comprendre l'utilisation des différents symboles.

Lisez le texte, puis essayez de répondre aux questions suivantes:

1. Quel est le gène affecté?
2. Que fait le gène?
3. Quelle est la mutation la plus courante dans la mucoviscidose
4. Décrivez l'anomalie moléculaire associée à la mucoviscidose.
5. Pouvez-vous expliquer l'apparence de l'infertilité chez les hommes atteints de mucoviscidose?
6. Quel est le génotype spécifique associé à l'infertilité chez les hommes atteints de mucoviscidose?

Essayons maintenant d'autres maladies. Vérifiez si vous pouvez obtenir ces informations auprès de l'OMIM.

1. Quel est le schéma héréditaire du cancer de la thyroïde?

2. Énumérez les gènes qui seraient liés à tous les cancers thyroïdes non médullaires.

3. Quelles maladies ont été liées au gène CF: seulement CFTR?

4. Quels types de variantes du gène xx sont liés au syndrome yy?

Comme vous pouvez le voir, la base de données OMIM fournit des informations utiles sur les gènes et les mutations qui sont déjà associés à un syndrome particulier. Dans d'autres cas, nous pouvons avoir besoin d'informations plus spécifiques sur une mutation particulière. Imaginez que vous découvriez une variante génétique d'un gène que vous soupçonnez de déclencher un syndrome génétique chez un patient. Si vous avez de la chance, cette variante a été signalée comme causale dans OMIM, mais ce n'est souvent pas le cas.

Ces dernières années, les scientifiques ont identifié, catalogué et examiné les variations génétiques chez l'homme. Un grand nombre de bases de données publiques fiables ont été développées et constituent une ressource importante pour l'interprétation des variantes (dbSNP, ClinVar, CADD, ExAC, Decipher etc.)

Continuons avec l'exemple de la fibrose kystique. Imaginez que nous ayons séquencé le gène entier chez certaines personnes en bonne santé, juste pour savoir si certains d'entre eux sont porteurs de la maladie. Nous n'avons trouvé aucune des variantes génétiques répertoriées dans OMIM, mais il existe d'autres variantes génétiques dans ce gène

- 7: 117120163 T / C
- 7: 117120179 G / A
- 7: 117199644 A / G (rs113993960)
- 7: 117120217 A / T
- 7: 117120229 G / T
- 7: 117144299 T / A
- 7: 117144400 A / G 7: 117149101 G / T (rs77284892)
- 7: 117267694 C / G (rs121908763)
- 7: 117171096 C / G

Certaines des variantes génétiques de nos patients ont déjà été identifiées et ont un SNP-ID. Vous pouvez vérifier ces variantes directement dans la base de données dbSNP (<https://www.ncbi.nlm.nih.gov/snp>).

La base de données de polymorphisme à nucléotide unique (dbSNP) est une base de données sur les variations du National Center for Biotechnology Information (NCBI). Il s'agit d'une archive publique de toutes les variations de séquences courtes, pas seulement des substitutions de nucléotides individuelles, mais comprend également une large collection de variations génétiques simples telles que les substitutions de nucléotides à base unique, les suppressions à petite échelle ou les insertions de petites bases et les répétitions de microsatellites.

Pour les postes avec un ID SNP, entrez simplement le nom du SNP dans le moteur de recherche. Vous serez redirigé vers la page d'informations pour ce SNP particulier. Vous pouvez trouver beaucoup d'informations sur cette variante génétique ici.

Ce SNP explique un grand nombre de cas de FC.

- Quel est le type de mutation?
- Quelle est leur fréquence dans tous les groupes de population?
- Dans quelle population cette variante est-elle la plus courante?

Entre autres, vous trouverez également des informations sur la signification clinique, la position dans le génome et les publications dans lesquelles il est rapporté.

- Toutes les significations cliniques signalées sont-elles liées à certaines maladies?
- Que trouvez-vous d'autre?

Vérifiez maintenant les autres identifiants SNP que vous avez reçus dans votre étude et indiquez le type de mutation provoquée par le changement et sa fréquence dans les différentes populations.

- Trouvez-vous une tendance dans tous les SNP qui vous semble étrange?

Pour examiner le reste des variantes obtenues, nous pouvons bénéficier d'ExAC, une base de données de données de séquençage d'exome provenant d'une variété de grands projets de séquençage. L'ensemble de données fourni sur ce site Web comprend 60 706 personnes non apparentées, qui ont été séquencées dans le cadre de diverses études spécifiques à la maladie et à la génétique des populations.

Allez sur ExAC (<http://exac.broadinstitute.org/>) et vérifiez si les variantes génétiques obtenues ont déjà été décrites chez un individu.

- Laquelle de vos variantes figure dans la base de données?

Ici, vous trouverez également des informations sur la fréquence, le type de mutation, la position dans le gène, etc.

- Y a-t-il une mutation que vous n'avez trouvée nulle part?
- Pouvez-vous dire quel changement vous souhaitez apporter?

Il est parfois difficile de savoir si les changements génétiques observés sont réellement impliqués dans la maladie. De plus, certaines de vos variantes n'ont peut-être jamais été décrites auparavant. La priorisation des variantes pour de nouvelles investigations expérimentales est donc un défi central dans les études de séquençage actuelles pour la recherche de maladies complexes. Les indices de pathogénicité sont des algorithmes de bioinformatique qui nous indiquent si la substitution d'acides aminés est tolérée ou nocive. Divers outils in silico ont été utilisés pour cette tâche, notamment PolyPhen-2, SIFT, FatHMM, MutationTaster-2, MutationAssessor, CADD, GERP et PhyloP. Par exemple, CADD est généralement utilisé pour prédire la nocivité des variantes nucléotidiques individuelles et pour les insertions / suppressions.

CADD Score est un outil puissant pour prédire l'effet d'une mutation où plusieurs bases de données sont intégrées dans une seule métrique.

Si le changement d'acides aminés devrait être nocif, c'est-à-dire la fonction protéique est altérée, le score CADD est plus élevé (généralement > 30). D'un autre côté, des valeurs de score CADD plus faibles prédisent des changements tolérés.

Veuillez lire ce document. Nous allons ensuite mener une activité sur cette base.

scores CDAO. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3992975/>

Pouvez-vous prédire l'effet des mutations trouvées avec CADD?

(ATTENTION, assurez-vous toujours que vous utilisez le même assemblage de génome, dans ce cas hg19)

Vous pouvez laisser la référence et les allèles alternatifs vides et recevoir toutes les combinaisons possibles à ce stade. Faites juste attention à la variante qui nous intéresse.

Lancez la recherche d'une variante nucléotidique unique (SNV) (<https://cadd.gs.washington.edu/snv>). Ici, vous pouvez accéder au score d'une seule variante de nucléotide. Essayons certains que nous avons reçus plus tôt:

Quel génome de référence avons-nous utilisé?

Quelles sont les valeurs CADD pour toutes les variantes que nous avons identifiées?

Selon vous, lequel est le plus nocif et serait le meilleur candidat

pour d'autres études?

C. QUIZ

1. Lequel des gènes suivants se trouve dans la plage positive du génome humain?
 - a. LRBA
 - b. ABO
 - c. SLC24A5
 - d. HGD
 - e. aucun

2. Voici la séquence pour une personne pour les positions chr8: 55.540.510 à 55.540.520: GGATTCAAAGT. Existe-t-il une variante génétique par rapport au génome de référence (hg19)?
 - a. aucun
 - b. Oui, une seule variante de nucléotide (SNV)
 - c. Oui, une insertion d'un nucléotide
 - d. Oui, une suppression de deux nucléotides
 - e. Oui, une suppression de dix nucléotides

3. Pour la même région que dans la question précédente, il s'agit de la séquence d'un individu avec un variant nucléotidique unique (SNV): GTATTCAAAGT. Cette variante provoque:
 - a. Un changement équivalent
 - b. Un nouveau codon stop qui détruit la protéine
 - c. Un changement faux-sens de l'acide aspartique à la tyrosine.
 - d. Un virage faux-sens de l'asparagine à la cystéine
 - e. Un changement faux-sens de la glycine à la valine

4. La drépanocytose est causée par le remplacement de A pour T sur le 17e nucléotide du gène de la chaîne bêta de l'hémoglobine, changeant ainsi le codon GAG en GTG. Pouvez-vous dire quel type de mutation est à l'origine de la maladie?
 - a. Mutation silencieuse
 - b. Copier variantes numéro
 - c. mutation frameshift
 - d. mutation nonsen
 - e. mutation faux-sens

5. Quel type de mutation un codon d'arrêt précoce introduit-il dans un gène?
 - a. Mutation absurde
 - b. Mutation silencieuse
 - c. faux-sens

- d. frameshift
- e. inversion

6. Laquelle des mutations suivantes est susceptible de provoquer des changements de trame?

- a. synonyme
- b. faux-sens
- c. empiècements
- d. suppressions
- e. c et d

7. La source ultime de variation génétique est:

- a. mutation
- b. La combinaison des chromosomes parentaux pendant la reproduction sexuelle
- c. Variantes chromosomiques
- d. épissure
- e. Tout ce qui précède

8. Les nouvelles mutations chez l'homme se produisent généralement à un taux d'environ ____ par personne chez des personnes normales et en bonne santé.

- a. 2-3
- b. 1000-2000
- c. 0
- d. million
- e. On ne sait pas

9. Qu'est-ce qui est vrai au sujet des mutations aléatoires?

- a. Ils sont toujours négatifs
- b. Tu es toujours positif
- c. Vous supprimez les variations d'une population
- d. Ils apportent de nouvelles variations à une population
- e. Ils rendent impossible la mesure des variations

10. Le défaut génétique du syndrome de Huntington et du syndrome de l'X fragile consiste en:

- a. une série de séquences nucléotidiques répétées
- b. un décalage des paires de bases
- c. une destruction importante d'une section importante d'un gène
- d. un bloc métabolique
- e. une variante du gène faux-sens