

Variations - Travaux Pratiques

Le laboratoire qui vous emploie souhaite mettre en évidence des facteurs de risque de développer un cancer du sein. Pour cela, une vingtaine d'exomes d'ADN constitutionnel ont été réalisés chez des patientes caucasiennes ayant développé un tel cancer avant 30 ans, et vous vous intéressez maintenant aux variants récurrents mis en évidence par cette approche.

La substitution d'un A par un G en position 37.053.568 du chromosome 3 ressort particulièrement dans votre cohorte. Utilisez le système de requêtes avancées de la base *dbSNP* du NCBI pour retrouver ce variant.

- 1.1. Commentez le nombre de résultats obtenus, et relevez l'identifiant de ce variant.
- 1.2. Toujours sur cette page de résultats, relevez l'impact clinique de ce variant. Commentez la provenance et la fiabilité de cette information (cliquez dessus pour obtenir de l'aide).
- 1.3. De la même façon, relevez la MAF de ce variant, expliquez son principe et sa provenance.

Rendez vous maintenant sur la page détaillée de ce variant.

- 2.1. Parmi les populations proposées, choisissez-en une de taille suffisante qui reflète mieux votre cohorte. Cliquez sur son nom, et décrivez brièvement son contenu et sa taille.
- 2.2. Relevez la MAF dans cette population. Correspond-elle à celle relevée en 1.3 ? Expliquez.
- 2.3. Sur quel gène ce variant se situe-t-il ? Suivez les liens vers la base de données *Genes* et décrivez rapidement l'intérêt que cette information peut représenter.
- 2.4. Décrivez l'impact de ce variant dans les différents transcrits de ce gène. Qu'en est-il des protéines qui en résultent (regardez avec attention) ?

Retrouvez ce variant dans les données de l'*Exome Variant Server*.

- 3.1. En vous limitant à la population la plus pertinente, affichez et relevez la MAF.
- 3.2. Précisez l'impact de ce variant sur la protéine donné en 2.4, en relevant et commentant les scores de Grantham et PolyPhen.
- 3.3. A partir de l'ensemble des informations collectées jusqu'ici, concluez quand à l'intérêt que peut représenter ce variant pour votre étude.

*Le gène *BRCA1* étant bien connu dans les cancers du sein et insuffisamment couvert dans vos données, vous vous proposez de reséquencer en Sanger des SNP connus pour compléter votre étude. Pour cela, vous allez mettre en évidence des SNPs situés dans le domaine fonctionnel de ce gène (nous nous limiterons à un unique candidat ici).*

Rendez vous sur le *Genome Browser* de l'UCSC, et affichez les pistes suivantes : *UCSC genes*, *Pfam domains in UCSC genes* et *Flagged SNP (138)*.

- 4.1. Quels domaines observez vous sur ce gène ? Suivez les liens jusqu'à la base Pfam pour donner une description succincte de ces domaines.
- 4.2. Choisissez un variant susceptible d'avoir un impact sur la fonction de la protéine parmi les nombreux proposés. Justifiez votre choix.
- 4.3. Suivez les liens jusqu'à dbSNP. Dans le cadre « GeneView », utilisez l'outil « Primer BLAST » pour designer des primers de séquence qui permettront de vérifier le statut de ce SNP dans les ADN constitutionnels de vos patientes. Ajoutez la piste « Clinical channel » au graphique de sortie pour vérifier votre couple de primers et prenez une capture d'écran. Recopiez également le tableau contenant les séquences correspondantes.

Variations - Travaux Pratiques

L'analyse d'ADN tumoraux par puces CGH vient d'être réalisée dans votre laboratoire, et vous êtes chargé d'interpréter les anomalies récurrentes mises en évidence. Les puces ayant été hybridées contre un pool d'ADN de référence, vous n'êtes pas certains s'il s'agit d'anomalies acquises par la tumeur ou de polymorphismes, ce que vous vous proposez de vérifier.

Le premier événement récurrent que vous avez pu mettre en évidence est le suivant : pour une sélection de patients se trouvent dans le tableau suivant les régions que l'analyse a permis de déclarer comme délétée, de manière hétérozygote ou homozygote.

chrom	start	end	sample	copies
9	19 845 236	22 245 212	LNH048	1
9	20 523 656	22 145 893	LNH075	1
9	20 325 412	28 745 215	LNH089	0
9	14 854 236	23 545 856	LNH112	1
9	7 852 265	45 212 365	LNH136	0
9	21 125 423	22 148 523	LNH141	0

Rendez vous sur le Genome Browser de l'UCSC, et visualisez ces données en temps que « custom track ». Vous pourrez être amenés à modifier leur format pour respecter l'un des nombreux standards supportés. Limitez l'affichage à une piste de gènes, ces données et au catalogue COSMIC (dense).

- 5.1. Grâce à une recherche sur internet, décrivez brièvement la taille et le contenu de ce catalogue, et justifiez son emploi ici.
- 5.2. Qu'observez-vous ? Relevez l'identifiant d'un ou plusieurs enregistrements COSMIC qui vous semblent pertinents, leurs fréquences et le nom du gène impliqué.

Rendez vous sur la base COSMIC pour continuer cette analyse.

- 6.1. Recherchez le gène relevé à la question précédente. Grâce aux trois premiers onglets, citez quelques uns des événements les plus fréquemment observés dans ce gène.
- 6.2. Nous nous intéressons ici à des tumeurs du sein. A quelle fréquence l'événement observé en 5.2 est-il décrit dans ce type de tissus ?
- 6.3. Concluez : l'événement récurrent étudié ici vous semble-t-il somatique ou constitutionnel ?

Un second événement récurrent a été mis en évidence dans cette série :

chrom	start	end	sample	copies
11	5 421 325	5 449 045	LNH012	0
11	5 398 452	5 449 045	LNH068	1
11	5 421 325	5 456 458	LNH079	1
11	5 398 452	5 449 045	LNH094	0
11	5 398 452	5 449 045	LNH142	1

Reprenez la même démarche avec la piste « DGV Struct Var ».

- 7.1. Idem que la 5.1 au sujet de la *Database of Genomic Variants*.
- 7.2. Idem que la 5.2. Vous suivrez également les liens jusqu'au site de la DGV pour y relever les références des publications ayant mis en évidence ces altérations.
- 7.3. Idem que la 6.3.