

# TP L2 bioinfo

*Thomas Karaouzene*

*27 novembre 2017*

## Informations :

L'objectif de ce TP d'appréhender l'analyse bioinformatique des données issues de séquençage exomique. Les comptes rendus sont à rendre par **binomes**. Sauf exceptions les monomes et trinomes ne seront pas acceptés !!!

Les comptes rendus seront à envoyer à l'adresse tkaraouzene@gmail.com avant le **mardi 04 décembre minuit**. L'horloge de la messagerie faisant foi, les CR reçus après minuit se verront retirer 5 points.

Ce TP est très rapide si vous maîtrisez les outils présentés ( $\simeq 20$  minutes), l'objectif ici est donc de vous laisser travailler en **AUTONOMIE**.

Prenez le réflexe d'explorer par vous-même les différents outils.

1. Cours 1 :
2. Cours 2 :
  - i. Savoir télécharger des séquences sur Ensembl
  - ii. Intégrer la notion de cadre de lecture
  - iii. *Design* d'amorce pour la PCR
3. Cours 3 :
  - i. Introduction à l'analyse bioinfo des données de séquençage haut-débit
  - ii. Utilisation du logiciel *Variant Effect Predictor* pour l'annotation des variants

## Cours 2 :

Toujours sur la base de données Ensembl, téléchargez, au format FASTA la séquence ADNc du transcrit identifié au cours précédent.

Utilisez le logiciel en ligne ExPASy afin de déterminer la séquence en acides-aminés.

Le logiciel fournit 6 réponses : 5'3' Frame 1-3 et 3'5' Frame 1-3.

1. **Question 1** : Selon vous, laquelle de ces séquences correspond à celle de la protéine SPINK2? Pourquoi?

Vous cherchez désormais à séquencer l'exon 2 de ce transcrit.

2. **Question 2** : Quel est l'identifiant de cet exon sur la base de données Ensembl (ENSExxxxx) ?
3. **Question 3** : Quel est la taille de cet exon ?

Rendez-vous sur le logiciel en ligne Primer3web.

Collez dans la case prévue à cet effet la séquence de l'exon deux que vous entourerez de crochets pour indiquer qu'elle est votre cible (ex : [ATTC...GGCCGTA]). Collez ensuite, de part et d'autre de cette séquence la fin de l'intron 1 ( $\simeq 3,5$  lignes de séquence) et le début de l'intron 2 ( $\simeq 3$  lignes de séquence).

Cliquez ensuite sur le bouton **Pick Primers**

4. **Question 4** : À votre avis, quel est l'intérêt d'avoir collé les séquences introniques flanquants l'exon 2 ?

5. **Question 5** : Collez les séquences des Primers que vous avez obtenues. Quelle est leur température de fusion ?
6. **Question 6** : Quelle est la taille, en nucléotides, de la région séquencée ?

## Cours 3 :

### Liens utiles

*Variant Effect Predictor* : annotation des variants.  
 Pubmed : bibliographie

### Analyse de deux frères azoospermes

Afin d'identifier la cause génétique entraînant un phénotype d'azoospermie chez deux frères, vous décidez d'effectuer un séquençage exomique de ceux-ci. En plus de ces deux frères, vous séquencez un troisième individu sain.

En utilisant le logiciel *Variant Effect Predictor* pour annoter les exomes de ces trois individus et en appliquant les filtres classiques, identifiez le variant ayant le plus de risque d'être le responsable du phénotype des deux frères.

Vous détaillerez dans votre rapport l'ensemble des critères de filtre utilisés en justifiant leur utilisation (une simple phrase suffit).

Vous pouvez appliquer les filtres (en utilisant l'outil fourni par VEP) dans l'ordre qui vous convient, néanmoins, certaines étapes seront plus rapide si vous avez appliqué vos filtres dans un ordre logique.

NB: Dans les paramètres de VEP :

- . **Identifiers and frequency data** : cochez gnomAD (exomes) allele frequencies
- . l'outil de filtre fourni par VEP bug pour filtrer les fréquences. vous devrez effectuer ce filtre manuellement.

Utilisez bien le lien vers **VEP** fourni dans ce document pour ne pas vous retrouver sur une autre version du logiciel

### Données

L'ensemble des données ci-dessous correspondent à des variants **Homozygotes**.

Vous pouvez coller directement les variants dans la zone prévue à cet effet (ne pas coller la première ligne)

#### Frère 1 :

#CHROM	POS	ID	REF	ALT
1	92647250	.	C	T
2	97818236	.	A	C
3	239675	.	C	T
3	239798	.	G	C
4	57676326	.	G	GA
5	112824068	.	C	CCGC
5	156479452	.	GTTG	G
8	41369893	.	G	A
8	41387610	.	C	T

8	41387642	.	T	G
8	41387809	.	T	C
11	196070	.	G	C
11	199813	.	A	G
15	50540273	.	C	A
15	50546971	.	C	G
16	4924426	.	C	T
16	29912807	.	G	GTGG
16	33629700	.	G	A
20	61919110	.	C	T
22	25334178	.	G	GG
22	29921848	.	A	G
22	46656674	.	TTC	T

## Frère 2 :

#CHROM	POS	ID	REF	ALT
1	92647250	.	C	T
3	239675	.	C	T
4	57676326	.	G	GA
5	112824068	.	C	CCGC
5	156479452	.	GTTG	G
8	41369893	.	G	A
8	41387610	.	C	T
8	41387809	.	T	C
9	79118329	.	G	A
9	79118400	.	C	T
9	79118475	.	G	A
9	79118628	.	C	T
9	79118633	.	CC	C
11	196070	.	G	C
11	199813	.	A	G
15	50540273	.	C	A
15	50546971	.	C	G
16	4924426	.	C	T
16	29912807	.	G	GTGG
16	33629700	.	G	A
20	61919110	.	C	T
22	25334178	.	G	GG
22	29921848	.	A	G
22	46656674	.	TTC	T

## Individu sain :

#CHROM	POS	ID	REF	ALT
1	92647250	.	C	T
3	69113153	.	T	C
3	69113155	.	G	A
3	69113185	.	T	A
8	41369893	.	G	A
9	120176925	.	C	CGGC
9	120176929	.	G	A
9	120176967	.	T	C

15	50546971	.	C	G	
20	45355480	.	G	A	
20	45357878	.	A	G	
20	45358005	.	G	T	
22	41743863	.	T	G	
22	41743947	.	CCTC		C
22	41744022	.	C	G	
22	41744334	.	A	G	
22	41745025	.	C	T	