Ajout de données dans RumimiR

Table des matières

[1. Recherche de publications 2](#_Toc34231429)

[2. Recherche des nouveaux microARNs 2](#_Toc34231430)

[3. Recherche et ajout des informations nécessaires 2](#_Toc34231431)

[Nom de la publication 3](#_Toc34231432)

[URL 3](#_Toc34231433)

[Tissu 3](#_Toc34231434)

[Race 3](#_Toc34231435)

[Condition d'étude 3](#_Toc34231436)

[Stade de lactation 3](#_Toc34231437)

[Age 3](#_Toc34231438)

[Espèce 3](#_Toc34231439)

[Nombre de microARNs 3](#_Toc34231440)

[Cut-off 3](#_Toc34231441)

[Méthode utilisée 3](#_Toc34231442)

[Nombre et races des animaux utilisés 3](#_Toc34231443)

[Logiciels utilisés 3](#_Toc34231444)

[Condition d'étude plus détaillée 3](#_Toc34231445)

[4. Recherche d'informations sur les microARNs 3](#_Toc34231446)

[5p/3p 3](#_Toc34231447)

[+/- 3](#_Toc34231448)

[Matching seed 3](#_Toc34231449)

[5. "Processing" des données (ajout d'informations des microARNs) 4](#_Toc34231450)

[Blaster les séquences sur le site NCBI (cf. diaporama "Blast\_RumimiR") 4](#_Toc34231451)

[Présence dans RumimiR : vérifier si une même séquence à la même localisation existe déjà 5](#_Toc34231452)

[Nom RumimiR 8](#_Toc34231453)

[Tissu spécificité 8](#_Toc34231454)

[Petits ARNs 8](#_Toc34231455)

[Identité hsa-, mmu- 10](#_Toc34231456)

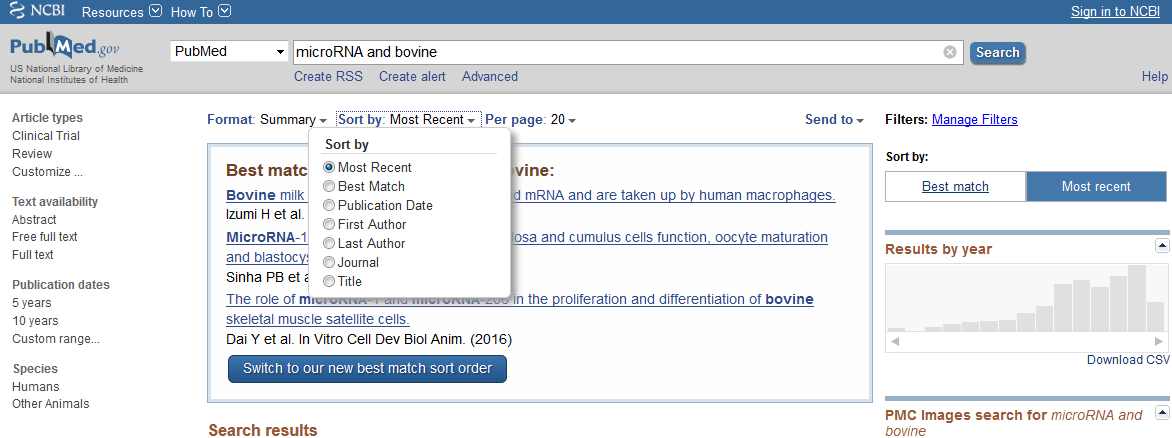
[False Positive Score avec encodage 10](#_Toc34231457)

[6. Mise à jour de génome de référence 10](#_Toc34231458)

## Recherche de publications

Recherche dans PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>) de nouvelles publications : rechercher les mentions

* microRNA and bovine
* microRNA and caprine
* microRNA and ovine
* microRNA and goat
* microRNA and sheep

Trier les résultats par "Sort by : Most Recent"

Lorsqu'une publication est trouvée, déclarant avoir trouvé de nouveaux microARNs (recherche des termes "novel" ou "new" dans le texte), passer à l'étape 2 de recherche des nouveaux microARNs.

## Recherche des nouveaux microARNs

Récupération des séquences de nouveaux microARNs : retrouver où sont les informations sur ces microARNs.

* Dans le texte
* Dans les figures
* En supplementary data (le plus souvent)
* Référencée sur une base avec un identifiant GEO/GSE

Traiter les informations : il faut au minimum avoir les séquences des microARNs décrits

Ajouter les séquences dans le classeur Excel

## Recherche et ajout des informations nécessaires

Informations valables pour tous les microARNs, détaillées le plus souvent dans la partie Matériel et Méthodes des articles. Quand l'information n'est pas disponible, mettre la mention "null".

### Nom de la publication

Mettre le nom de la publication suivi d'un point-virgule, du nom du premier auteur suivi de "et al. " et la date (exemple : Food Deprivation Affects the miRNome in the Lactating Goat Mammary Gland ; Mobuchon et al. 2015).

### URL

Le lien internet de la publication.

### Tissu

### Race

### Condition d'étude

Condition générale de l'étude (nutrition, développement, reproduction, organes, lactation…).

### Stade de lactation

### Age

**Convertir les âges en jours**, tout en laissant la mention inscrite dans la publication entre parenthèse.

### Espèce

### Nombre de microARNs

Nombre de microARNs décrits comme nouveaux dans la publication ("novel" ou "new" microRNAs).

### Cut-off

Critères utilisés pour la détection de microARNs (dans le matériel et méthodes, souvent partie "data processing", "sequence analysis" ou "identification of novel miRNA"). Le plus souvent : taille des microARNs, score miRDeep2, structure en tige-boucle ("stem-loop"), "Free energy of hybridization" (MFE)…

### Méthode utilisée

Séquençage ou méthodes bio-informatiques par homologies de séquences notamment.

### Nombre et races des animaux utilisés

### Logiciels utilisés

Liste des logiciels utilisés pour la détection de nouveaux microARNs (miRDeep2, Bowtie….). A classer par ordre alphabétique, séparés par un "/".

### Condition d'étude plus détaillée

Nombre d'animaux par groupe, si différence entre races : quelles races, si différence entre sain et pathologique : quelle pathologie, si alimentation : quelle alimentation, si tissus : quelles tissus, si stades de lactation : quels stades, etc…

## Recherche d'informations sur les microARNs

Informations valables pour chaque microARN individuellement, à rentrer directement dans le classeur Excel

### 5p/3p

### +/-

### Matching seed

## "Processing" des données (ajout d'informations des microARNs)

Partie de manipulations/analyses des données

### Blaster les séquences sur le site NCBI (cf. diaporama "Blast\_RumimiR")

Voir le diaporama nommé "Blast\_RumimiR" pour plus de détails sur le Blast en lui-même (et des exemples).

* [Blast bovin](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=OGP__9913__10708) pour les bovins, avec le bon génome de référence de choisi dans la liste déroulante Database : "Genome (ARS-UCD1.2 reference, Annotation Release 106)"
* [Blast caprin](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=OGP__9925__49537) pour les caprins, avec le bon génome de référence de choisi dans la liste déroulante Database : "Genome (ASM170441v1 reference Annotation Release 102)"
* [Blast ovin](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=megaBlast&BLAST_SPEC=OGP__9940__10709) pour les ovins, avec le bon génome de référence de choisi dans la liste déroulante Database : "Genome (Oar\_v4.0 reference Annotation Release 102)"

Dans le cas de localisations multiples du microARN sur le génome (ou "multi-mapping ") :

* Rechercher dans les données de la publication si ils ont noté la séquence du précurseur du microARN, et la blaster (environ 100 paires de bases : plus long donc le blast a plus de chances de trouver la localisation exacte si elle existe). En déduire donc quelle est la localisation exacte du microARN en question. (exemple : un microARN avec 2 localisations possibles : chr1 : 1002503-1002524 ou chr4 : 45825804-45825825. Le précurseur est lui localisé sur le chromosome 1 : chr1 : 1002474-1002577, la réelle localisation du microARN est donc la première).
* Si il n'y a pas la séquence du précurseur ou que ça ne donne rien : laisser les colonnes chr, start et end vides et entrer dans la colonne "multi-mapping" le nombre de localisations potentielles possibles.

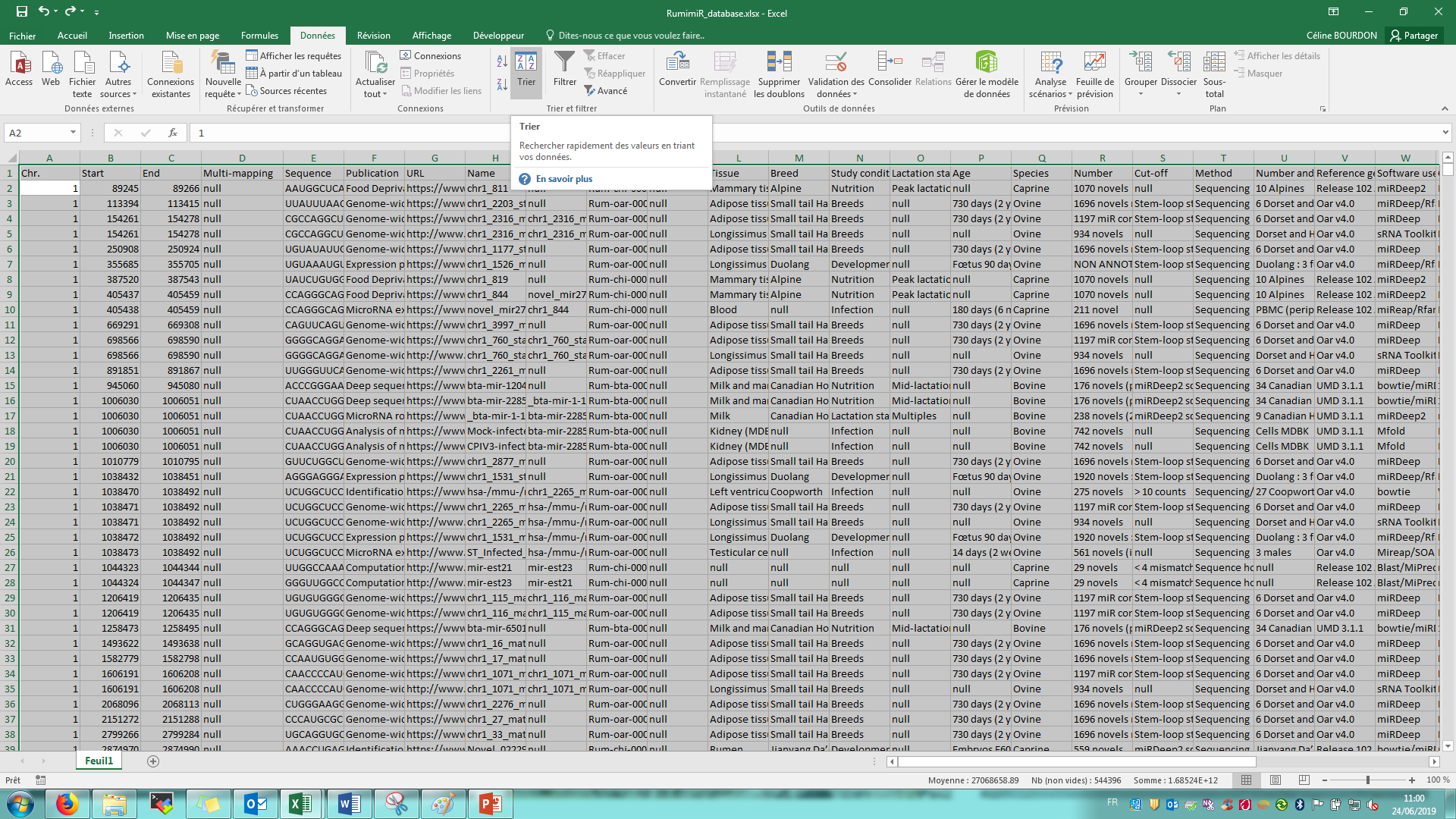
***Quand ajout des séquences dans la base de données : bien penser à remplacer les T par des U dans les séquences.*** (sélectionner la colonne des séquences ; ctrl + H, Rechercher : T, Remplacer par : U).

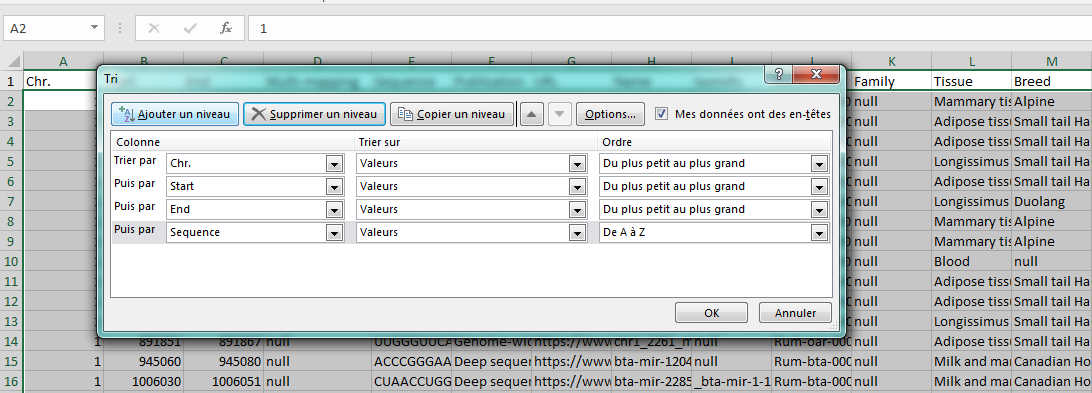
Génome de référence : ajouter le nom de la version du génome utilisée pour blaster les séquences.

Vérifier que le nombre de nouveaux microARNs ajoutés dans RumimiR pour une publication correspond bien au nombre de microARNs annoncé dans l'article.

### Présence dans RumimiR : vérifier si une même séquence à la même localisation existe déjà

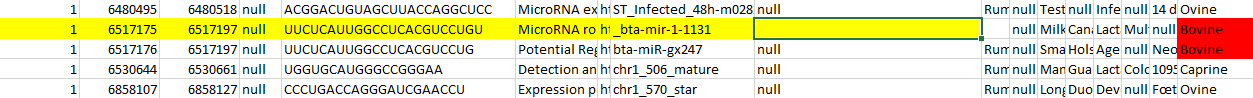
Données nouvelles complétées : surligner les lignes correspondantes (pour les retrouver plus facilement), puis trier la base entière par chr, start, end, sequence



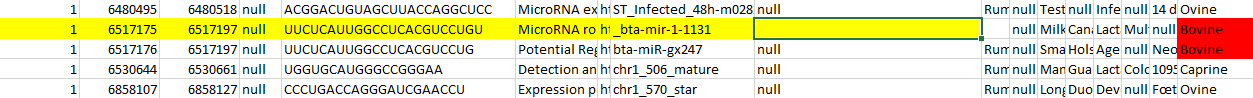


Colonne "IsomiR" encore vide pour les nouvelles données rentrées donc faciliter de les retrouver : aller sur la colonne "IsomiR", faire ctrl+flèche du bas : ça amène directement à une ligne surlignée (un microARN nouvellement ajouté). Regarder si le microARN est à une position équivalente à un autre microARN (tri par position donc regarder les lignes justes précédentes et suivantes), et si oui comparer les séquences des microARNs 🡪 si **les deux** éléments sont semblables, remplir la case "IsomiR", sinon y rentrer la mention "null".

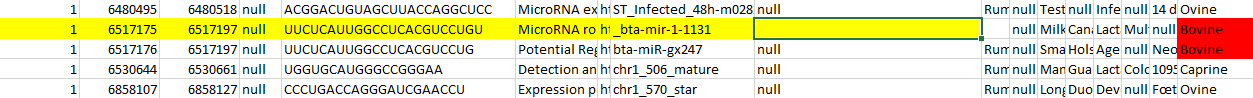
Remplissage de la case "IsomiR" : y inscrire le nom du/des microARN(s) semblables, et ajouter le nom de ce nouveau microARN dans la case "IsomiR" de ces semblables.



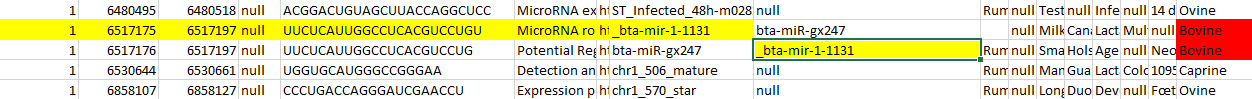
Nouveau microARN ajouté (ligne en jaune)



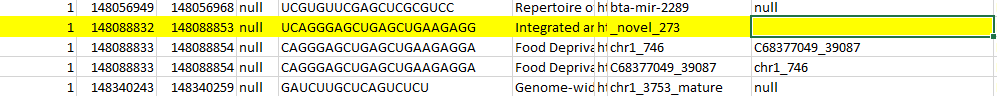
Positions génomiques semblables (encadré vert)



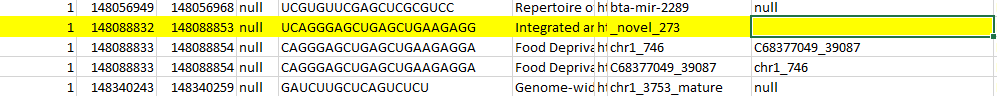
Séquences semblables (encadré vert)



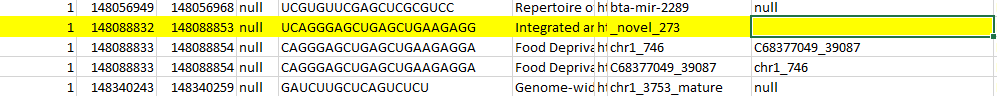
Ajout de l'information dans la colonne IsomiRs

Lorsque le microARN est semblable à plusieurs autres :

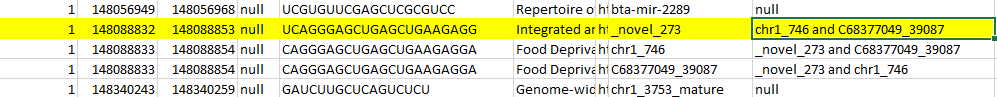
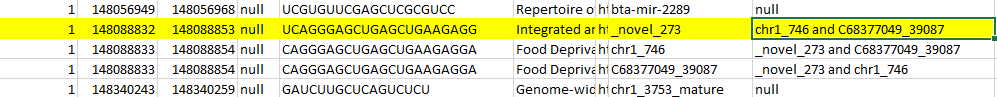
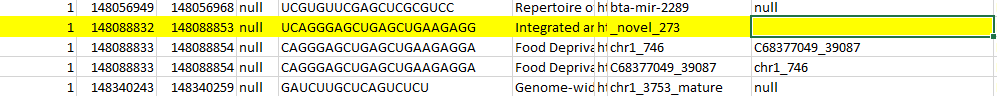
Nouveau microARN ajouté (ligne en jaune)



Positions génomiques semblables (encadré vert)



Séquences semblables (encadré vert)



Ajout de l'information dans la colonne IsomiRs

### Nom RumimiR

Nomenclature : "Rum-", suivi de trois lettres spécifiques à l'espèce ("bta-", "chi-" ou "oar-") puis d'un nombre à 5 décimales, attribué au microARN en fonction de leur position dans les données de la publication où il a été décrit (une nouvelle publication aura donc des microARNs avec des n° qui se suivent, à l'exception de ceux déjà retrouvés dans RumimiR) :

* Classer les microARNs de la nouvelle publication par chr, start, end, sequence.
  + Pour les microARNs n'ayant pas de doublon (mention "null" dans la colonne isomiR) : reprendre la numérotation à partir du dernier numéro attribué dans la base de données pour l'espèce concernée.
  + Pour les microARNs déjà présents dans RumimiR, l'identifiant sera le même que celui qui a été attribué à celui (ceux) déjà présent(s) (= ses doublons).

### Tissu spécificité

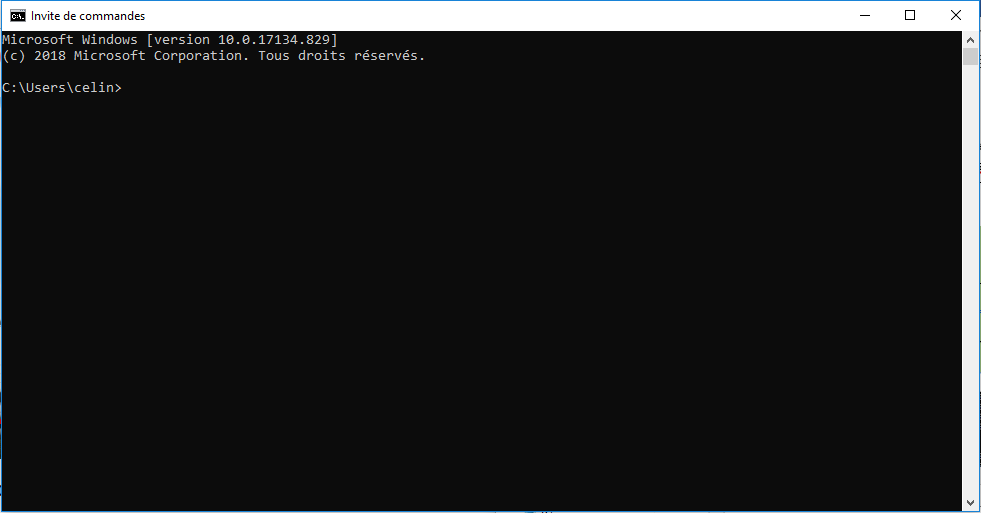
Tissu-spécificité (colonne permettant de mettre en évidence les microARNs détectés dans un seul tissu – et donc spécifique de ce tissu) :

* Si la publication décrit des microARNs détectés dans plusieurs tissus : inscrire "No" à la tissu-spécificité
* Si la publication décrit des microARNs détectés dans un seul tissu :
  + Si le microARN n'a aucun isomiR (pas de doublon) 🡪 inscrire "Yes" (dans un seul tissu, détecté dans une seule publication)
  + Si le microARN a un/des isomiR(s) 🡪 comparer les tissus dans lesquels les tissus ont été détectés : si plusieurs tissus différents 🡪 "No", si un seul tissu (toujours le même) 🡪 "Yes"

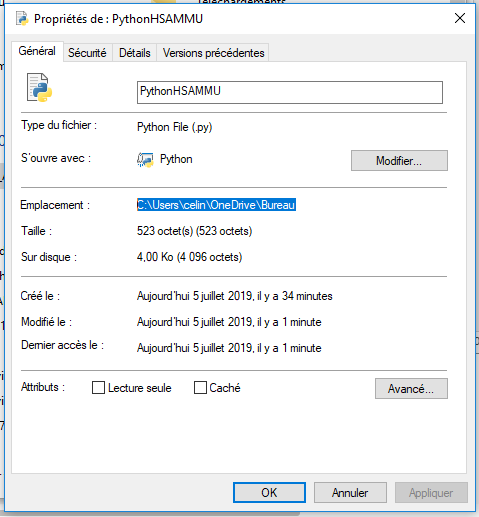
### Petits ARNs

Utiliser le script python "RumimiR\_smallRNAs.py" dans la console de commande (cmd) :

* Coller les noms et séquences des nouveaux microARNs dans le fichier texte "nveaux\_miRs"
* Installer Python (<https://www.python.org/downloads/>)
* Lancer l'invite de commande : dans windows, menu démarrer, chercher cmd.exe

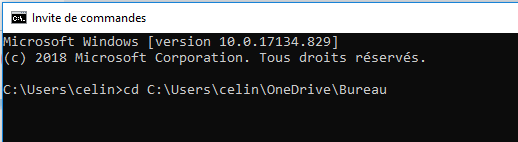


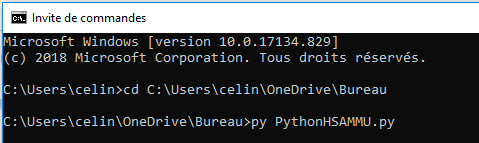
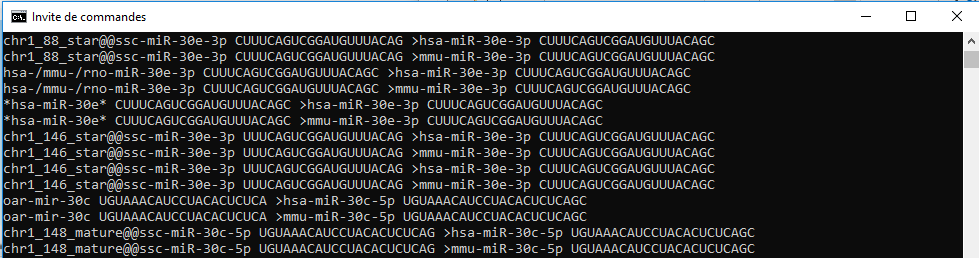
* Chercher l’emplacement du fichier RumimiR\_smallRNAs.py (clique droit sur le fichier, propriété, copier l’emplacement du fichier)



* Se positionner dans le dossier où se trouve le programme Python (cd emplacementquiaétécopié)

Exemple :



* Lancer le programme (taper « py RumimiR\_smallRNAs.py »)
* Coller les résultats dans un classeur Excel (clic droit, sélectionner, griser résultats, clic droit ; puis aller dans Excel et faire ctrl+V -coller-)
* Remplir la colonne dans RumimiR avec le nom du petit ARN associé.

### Identité hsa-, mmu-

Utiliser le script python "RumimiR\_hsammu.py" dans la console de commande (cmd) :

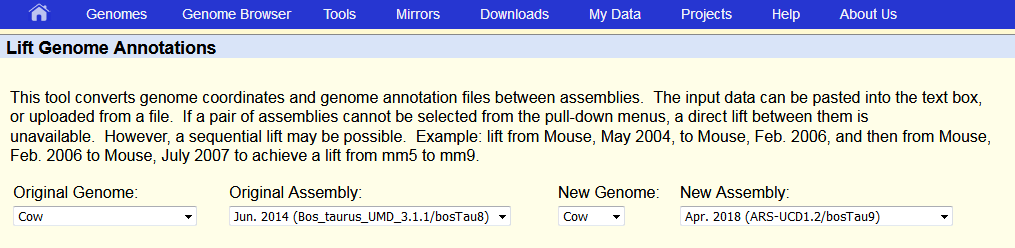
* Installer Python (<https://www.python.org/downloads/>)
* Lancer l'invite de commande : dans windows, menu démarrer, chercher cmd.exe
* Lancer l'invite de commande : dans windows, menu démarrer, chercher cmd.exe
* Chercher l’emplacement du fichier RumimiR\_smallRNAs.py (clic droit sur le fichier, propriété. Copier l’emplacement du fichier)
* Se positionner dans le dossier où se trouve le programme Python (cd emplacementquiaétécopié)
* Lancer le programme (python RumimiR\_hsammu.py)
* Coller les résultats dans un classeur Excel
* Remplir les colonnes dans RumimiR

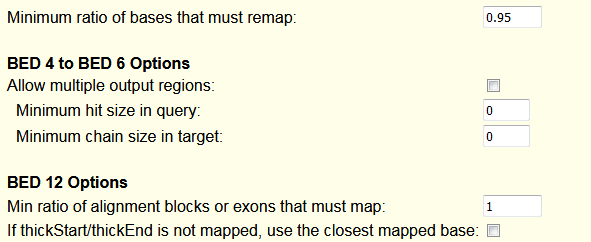
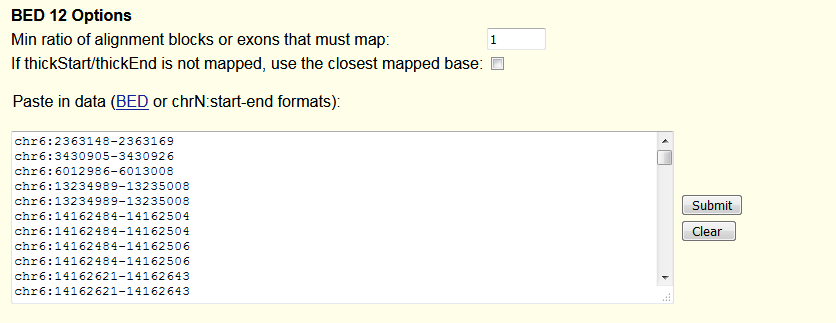
### False Positive Score avec encodage

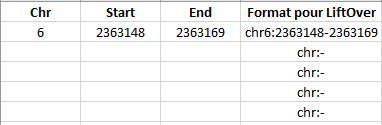
La colonne "False Positive Code" se complète automatiquement dans le classeur Excel lorsque les données sont ajoutées.

## Mise à jour de génome de référence

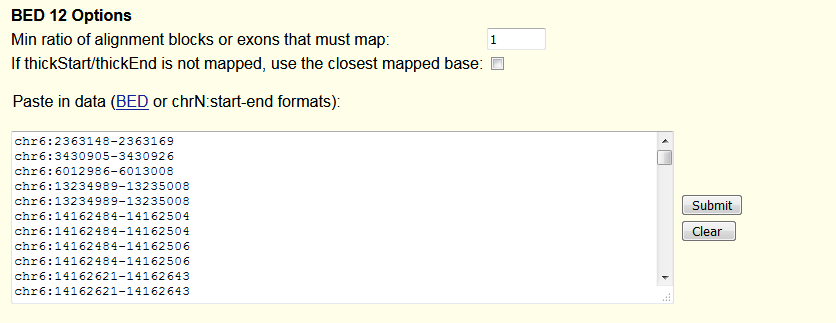
Afin de transposer les séquences dans un nouveau génome de référence, l'outil LiftOver de UCSC est utilisé : <https://genome.ucsc.edu/cgi-bin/hgLiftOver>.



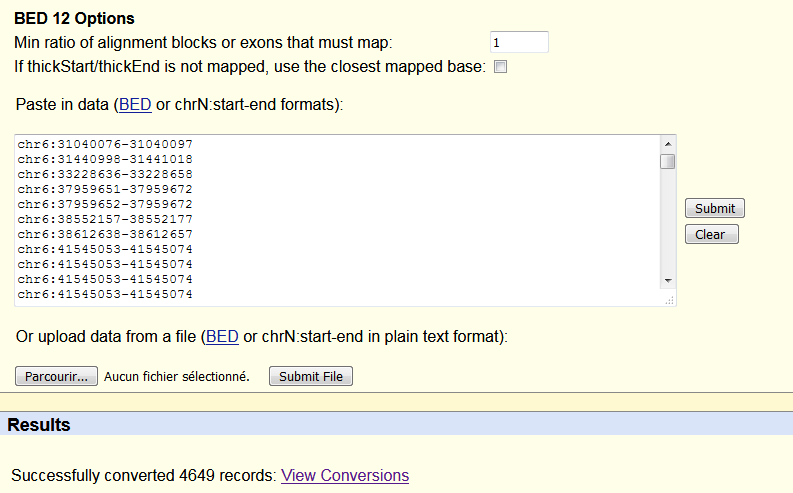
* Choisir le génome de référence utilisé auparavant (ici le génome UMD3.1.1 en bovin) et le nouveau génome de référence (ici ARS-UCD1.2)
* Garder les conditions par défaut :
* Coller les positions des microARNs dont le génome de référence doit être changé dans un format chrN:start-end :

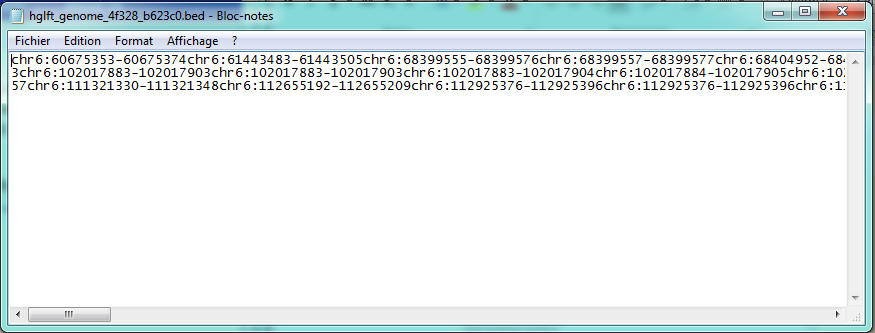
Pour obtenir le bon format, possibilité d'utiliser le classeur Excel "Pour\_LiftOver".

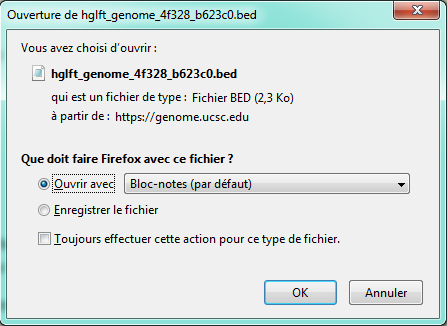
* Cliquer sur "Submit" :

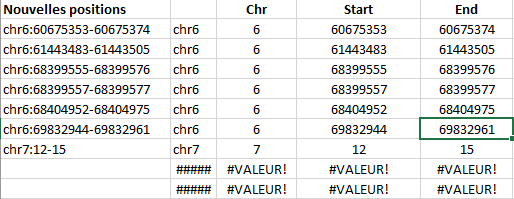


* Regarder les nouvelles positions :



* Un fichier .txt s'ouvre alors dans le bloc-note :



* Copier ces positions (ctrl a ; ctrl c pour toutes les sélectionner et copier) dans le classeur Excel "Conversion\_NvellesPositions" afin de les transposer dans le bon format pour RumimiR (avec les 3 colonnes : Chr., Start, End) :
* Trier les données du classeur RumimiR par espèce puis copier dans le classeur, en ayant au préalable coloré les cases concernant les séquences n'ayant pas pu être transposées (pour ne pas copier les résultats dans ces cases, et possibilité de trier par couleur de case dans cette colonne)

Lors d'un changement de génome de référence, penser à reblaster les séquences qui n'ont pu être transposées (en cliquant sur "Display failure file" ces séquences apparaissent) :

