**Demarrer notebook : a lancer dans terminal**

genefish:~ srlab$ cd /Volumes/web/scaphapoda/Yanouk/ipython\_nb

genefish:ipython\_nb srlab$ ipython notebook --pylab inline

**BSMAP** Dans notebook :

!/Users/Shared/Apps/bsmap-2.73/bsmap -a /Volumes/web/scaphapoda/Yanouk/Sample\_Y54\_Mix/filtered\_Y54\_Mix\_GTGGCC\_L007\_R1.fastq.gz -d /Volumes/web/cnidarian/oyster.v9.fa -d /Volumes/web/cnidarian/oyster.v9.fa -o /Volumes/web/scaphapoda/Yanouk/Mixbsmapoutput.sam -p 4

ou alors :

!/Users/Shared/Apps/bsmap-2.73/bsmap -a /Volumes/web/scaphapoda/Yanouk/Sample\_Y54\_Mix/filtered\_Y54\_Mix\_GTGGCC\_L007\_R1.fastq.gz -b /Volumes/web/scaphapoda/Yanouk/Sample\_Y54\_Mix/filtered\_Y54\_Mix\_GTGGCC\_L007\_R2.fastq.gz -d /Volumes/web/cnidarian/oyster.v9.fa -o /Volumes/web/scaphapoda/Yanouk/Testbsmapoutput.sam -p 4

*2 fichiers d’entrée(qui sont les resultats du bisulfite sequencing) :R1 et R2 lus dans les 2 sens, 1 fichier correspondant au genome, le programme bsmap, et le fichier de sortie que je nomme avec l’extension* ***.sam***

**METHRATIO:**

!python /Users/Shared/Apps/bsmap-2.73/methratio.py -d /Volumes/web/cnidarian/oyster.v9.fa -u -g -z -o /Volumes/web/scaphapoda/Yanouk/DecData/Mix54methratio.txt -s /Users/Shared/Apps/bsmap-2.73/samtools /Volumes/web/scaphapoda/Yanouk/DecData/Mix54bsmapoutput.sam

*programme methratio, 1 fichier correspondant au genome, le fichier* ***.txt*** *correspondant au fichier de sortie que l’on cree, l’utilisation d’une extension samtools permettant de lire le fichier d’entree (.sam)*

**Sqlshare**

Permet de mettre des filtres a mon fichier texte… parfois le fichier txt a charger sur SQLSHARE est trop lourd, donc il faut passer par un terminal, et si c’est trop complique, demander a Mackenzie de le faire pour moi( en lui donnant le fichier sur eagle sans doute)

**Convertir le fichier txt en fichier csv sur SQL Share**

SELECT

chr as [Column1],

pos as [Column2],

pos + 1 as [Column3],

'CpG' as [Column4],

cast(ratio as float) as [Column5]

FROM [sr320@washington.edu].[clean\_BiGill\_methratio\_v9\_A]

where

context like '\_\_CG\_'

and

CT\_Count >= 5​ *il faut que la sequence CpG soit au moins repetee 5 fois pour etre sur qu’il y a bien un enchainement de C puis G*

and

ratio > 0.5 *Cpg pour lesquels on assure qu’il y a bien methylation…. Si c’etait inferieur a 0.5 ca correspondrait a des CpG unmethylated*

*J’utilise plutot le code d’apres sur derive dataset qui me permet d’obtenir le fichier sous la forme qu je desire pour la suite*

SELECT   
chr as seqname,    
'methratio' as source,    
'CpG' as feature,   
pos as start,  
pos + 1 as [end],  
ratio as score,    
strand,    
'.' as frame,    
'.' as attribute    
  
FROM [mgavery@washington.edu].[Mix54methratio\_final]       
where   
context like '\_\_CG\_' --\_=single character wildcard  
and  
CT\_Count > 5​

**A ce moment, j’obtiens un fichier .csv avec toutes les zones putatives de methylations, mais il faut rajouter un code (precedemment :and ratio > 0.5)(ou alors sur excel)pour ne conserver que les zones avec un score sup ou egal a 50%**

**A partir du fichier csv qui est en realite un format gff (9colonnes), je l’enregistre en .txt (tab delimited text) je l’ouvre avec “text wrangler”, j’enregistre en changeant le mac en unix file.**

**Je change ensuite le fichier .txt en fichier .gff (en verifiant bien que je ne change pas juste le nom , mais aussi l’extension).**

**Le but de l’operation suivante est de faire tourner bedtools, cad a partir de mon fichier gff avec tous mes scaffolds, je peux savoir toutes les zones des scaffolds correspondants a un gene en particulier…**

!/Users/Shared/Apps/bedtools-2.17.0/bin/intersectBed -wb -a /Volumes/web/scaphapoda/Yanouk/Table\_Mix54forGO.gff -b /Volumes/web/trilobite/Crassostrea\_gigas\_v9\_tracks/Cgigas\_v9\_gene.gff >/Volumes/web/scaphapoda/Yanouk/Mix54intersectbed.txt

*le premier fichier gff est le fichier d’entrée qui est le fichier.txt que l’on a transforme en gff. L second fichier gff correspond au fichier du labo correspondant aux regions intrageniques ( on aurait pu faire la meme avec les exons, les introns, les promoteurs…). On obtient donc l’ID de nos scaffolds touches par les methylations. Le fichier .txt de la fin est le fichier de sortie.*

*On ouvre ce fichier sous excel, et ne pas oublier de virer dans la colonne ID le “ID=” ainsi que le “;” a la fin.*

**ID sur SQLSHARE**

SELECT \* FROM [noukitus@gmail.com].[Mix54intersectbedfinal.txt]  
      share\_​

*A partir du fichier de Mackenzie (disponible sur SQLshare :* [mgavery@washington.edu].[TJGR\_Gene\_SPID\_evalue\_Description.txt] ca me permet d’avoir l’ID et la GO des genes ayant des methylations!!!

**TRAVAIL AVANCE SUR SQL SHARE ou R Package**

*Il faut (encore une fois) modifie la forme du fichier excel, changer les colonnes sur sqlshare avec le code suivant*

SELECT   
  chr as chr,  
  pos as start,  
  '+' as strand,  
  cast (CT\_count as float) as CT\_count,  
  cast (ratio as float) as ratio

FROM [mgavery@washington.edu].[Mix54methratio\_final]  
  where   
context like '\_\_CG\_'  
and  
  CT\_Count >= 5   
and   
  ratio <> 'NA'

*Une fois que l’on a le fichier, il faut l’enregistrer en format .txt après avoir supprime la 1ere ligne avec le nom des colonnes!*

***R***

*Installer les differents packages*

*Puis lancer les differents codes 1 a 1*

**NOTEBOOK of ROBERTSLAB**

<https://genefish.wikispaces.com/sr320+Notebook>

**Connect to eagle http://eagle.fish.washington.edu/**

Rechercher eaglefish Washington, se connecter : srlab mdp:scallop

**e21207006**

**Votre mot de passe : MKN4D9Ab**