essai-format

Elise GUERET

16 avril 2018

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

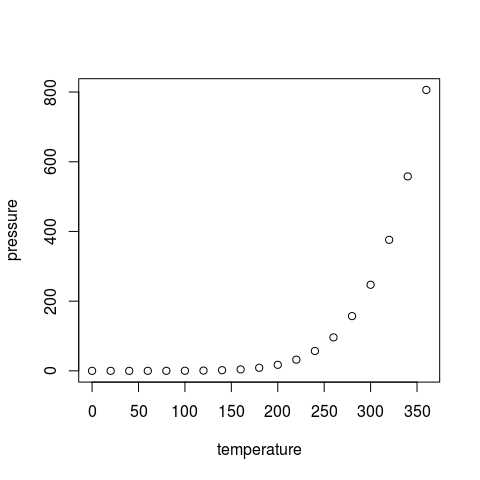
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

summary(cars)

## speed dist   
## Min. : 4.0 Min. : 2.00   
## 1st Qu.:12.0 1st Qu.: 26.00   
## Median :15.0 Median : 36.00   
## Mean :15.4 Mean : 42.98   
## 3rd Qu.:19.0 3rd Qu.: 56.00   
## Max. :25.0 Max. :120.00

## Including Plots

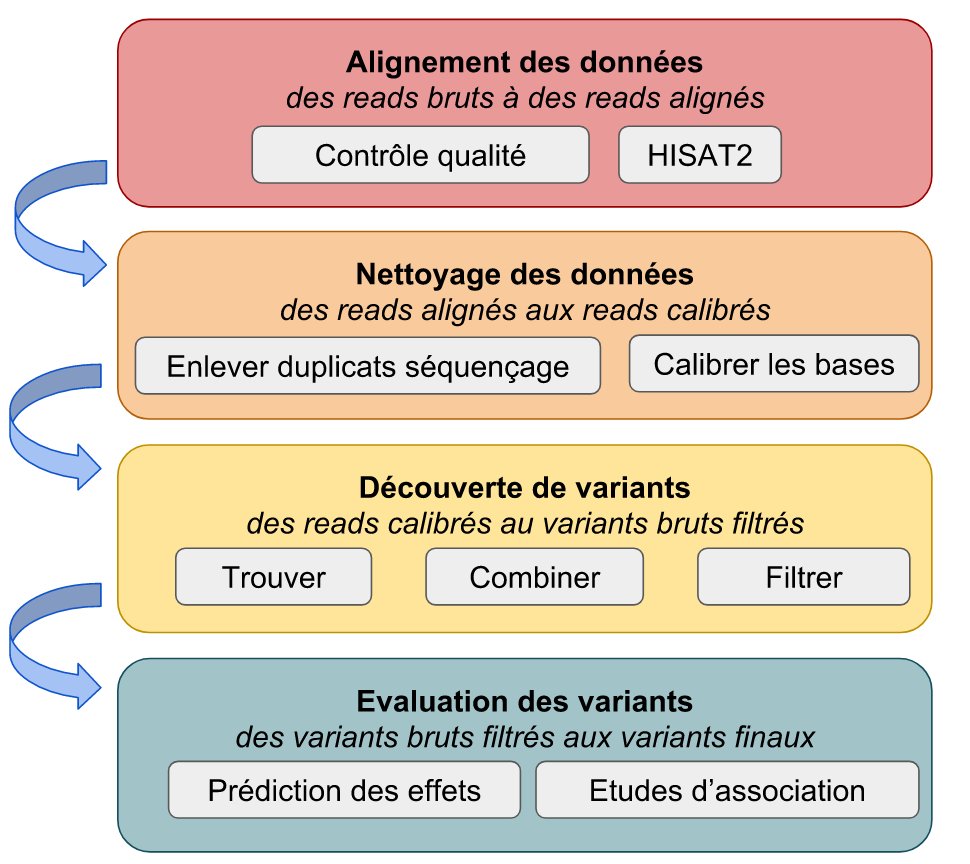
You can also embed plots, for example:



Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

|  |  |  |
| --- | --- | --- |
| Sigle du champs utilisé | Valeur si SNPs | Valeur si INDELs |
| QD | <2.0 | <2.0 |
| FS | >60.0 | >200.0 |
| SOR | >3.0 | >10.0 |
| ReadPosRankSum | <-8.0 | <20.0 |
| MQ | <40.0 |  |
| MQRankSum | <-12.5 |  |
| InbreedingCoeff |  | <-0.8 |

Ce pipeline est décomposé en 4 phases distinctes qui utilisent différentes resources et qui permettent de faire de la parallèlisation pour optimiser les temps de calculs. Ces 4 phases sont les suivantes :  
\* Alignement des données (réalisé par l’équipe de la plateforme MBB);  
\* Nettoyage des données;  
\* Découverte des variants;  
\* Evaluation des variants.

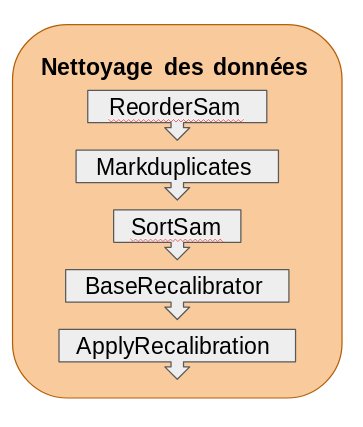


Vue générale du pipeline de génotypage

Ce pipeline est une collection de script wdl, bash et json à exécuter dans l’ordre sur une machine Linux avec une gestion des jobs de type gridengine comme un cluster de calcul. Il est également exécutable sur une machine Linux sans gestion des jobs tels qu’une station de travail.

Pour rendre les données analysables pour la découverte de variants quelques étapes des phases précédentes nécessaires sont détaillées ci-après.

#### Nettoyage des données

Le nettoyage des données est une phase nécessaire pour pouvoir réaliser la suite des analyses. En effet, elle permet d’obtenir des reads calibrés nécessaire à la phase 3. Cette phase comprend des étapes réalisées par Picard ou par GATK. 

Mon rapport commence ici.

# Titre 1

etblahblah

etblahblahblah

## Titre 2

etblahblah

etblahblahblah

### Titre 3

etblahblah

etblahblahblah