



Manuel d'utilisation du pipeline RNAditing

13 Mars - 12 Septembre 2017

Benjamin DELAUNE
Master 2 Bioinformatique 2016-2017

Encadrants : Dr. Stéphane Minvielle, Dr. Florence Magrangeas
Laboratoire d'accueil : Centre de Recherche en Cancérologie et Immunologie Nantes
Angers, INSERM UMR1232/CNRS ERL6001 - 8 Quai Moncousu 44007 Nantes

Table des matières

1	Présentation du pipeline	2
2	Installation du Pipeline	3
2.1	Installation de git	3
2.2	Téléchargement du pipeline	4
3	Installation des pré-requis	4
3.1	Installation de Conda	4
3.2	L'environnement RNAEditor	5
3.2.1	Création de l'environnement RNAEditor	5
3.2.2	Installation des packages Python	5
3.2.3	Installation des logiciels pré-requis	5
3.2.4	Télécharger les fichiers nécessaires à RNAEditor	7
3.3	L'environnement REDIttools	7
3.3.1	Installation des packages Python	7
3.3.2	Installation des logiciels pré-requis	8
3.3.3	Récupération des fichiers utilisés	8
4	Utilisation du pipeline RNAditing	10
4.1	Utilisation totale du pipeline	10
4.2	Utilisation séparée des logiciels	12

1 Présentation du pipeline

RNAEditing est un pipeline implémenté en Python permettant le contrôle qualité de données RNA-Seq ainsi que la prédiction de sites d'édérations de l'ARN et la recherche de sites de RNAEditing déjà répertoriés, dans des données de RNA-Seq paired-end ou single-end.

Il utilise pour cela 2 logiciels spécialisés et disponibles en ligne.

- **RNAEditor** : outil de prédiction de sites d'édérations de l'ARN implémenté en Python par John et al.

<https://github.com/djhn75/RNAEditor>

David John, Tyler Weirick, Stefanie Dimmeler, and Shizuka Uchida. Rnaeditor : easy detection of rna editing events and the introduction of editing islands. Briefings in bioinformatics, page bbw087, 2016.

- **REDIttools** : Composé de 3 scripts implémentés en Python par Picardi et al. Dans cette version du pipeline, j'ai incorporé uniquement le script REDIttoolsKnown.py.

Ce script est utilisé pour la détection de sites d'édérations déjà référencés dans la littérature dans des données de RNA-Seq en utilisant une base de données spécialisée du RNA Editing.

Picardi, E., & Pesole, G. (2013). REDIttools : high-throughput RNA editing detection made easy. Bioinformatics, btt287.

Picardi, E., D'Erchia, A. M., Montalvo, A., & Pesole, G. (2015). Using REDIttools to detect RNA editing events in NGS datasets. Current protocols in bioinformatics, 12-12.

Les deux logiciels traitent les échantillons de façon individuelle. Cependant, il est parfois préférable de raisonner en terme de population. De ce fait, le pipeline compare les résultats des échantillons de même population afin de recenser les sites d'édérations représentatifs de celle-ci.

Des étapes d'annotation, de créations de graphiques et de récupération de gènes édités sont également réalisées afin de visualiser les résultats obtenus et les caractériser (figure 1).

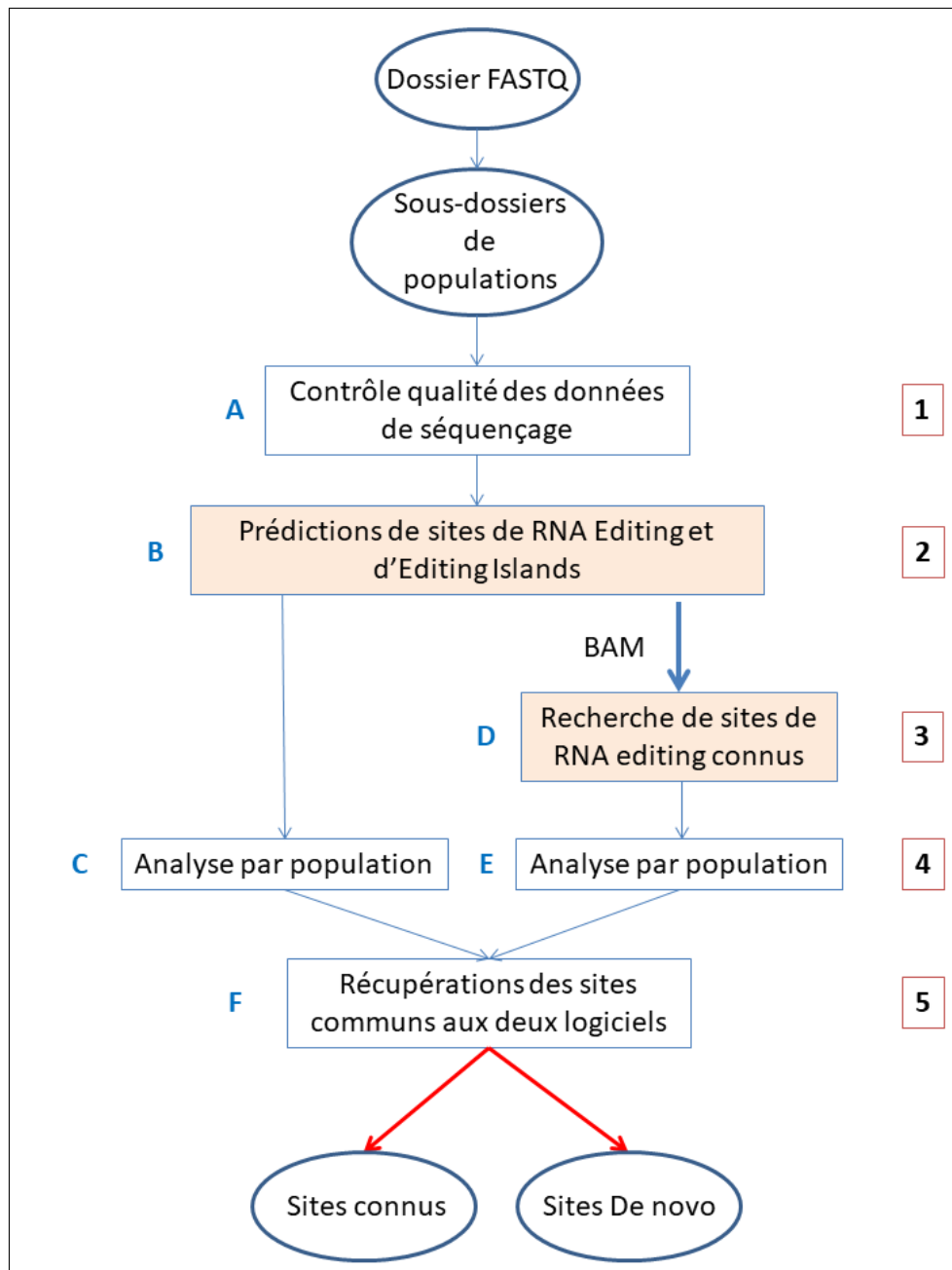


FIGURE 1 – Schéma du déroulement du pipeline

Les numéros représentent les grandes étapes et les lettres le déroulement du pipeline

2 Installation du Pipeline

Le pipeline est disponible sur gitHub à l'adresse suivante :

<https://github.com/WYSNI/RNAditing.git>

2.1 Installation de git

Pour installer git, réalisez la commande suivante :

```
$ sudo apt install git-core gitk
```

Dans le répertoire dans lequel vous voulez mettre le pipeline :

```
$ git init
```

2.2 Téléchargement du pipeline

Pour télécharger le pipeline :

```
$ git clone https://github.com/WYSNI/RNAditing.git
```

Dans le .bashrc :

— Exporter les paths

```
$ gedit /PATH/TO/.bashrc
$ export RNAditing_PATH_RNAEditor=/PATH/TO/RNAditing/RNAEditor/RNAEditor/
$ export RNAditing_PATH_REDIttools=/PATH/TO/RNAditing/REDIttools/
$ export RNAditing_PATH_RNAditing=/PATH/TO/RNAditing/RNAditing/
```

— Créer des alias

```
$ alias RNAditing="python_/PATH/TO/RNAditing/RNAditing/RNAditing.py"
$ alias TAB_modification="python_/PATH/TO/RNAditing/RNAditing/TAB_modification.py"
```

Réouvrir un terminal

3 Installation des pré-requis

3.1 Installation de Conda

Du fait que les 2 logiciels de RNA Editing utilisent des versions du package Python pysam différentes, il est nécessaire de réaliser deux environnements virtuels conda afin de pouvoir utiliser le pipeline dans son intégralité.

Aller sur le site suivant afin de télécharger le script d'installation.

<http://conda.io/miniconda.html>

Dans un terminal, réaliser les opérations suivantes :

```
$ cd /PATH/DU/DOSSIER/CONTENANT/LE/FICHER/TELECHARGE
$ chmod 755 Miniconda2-latest-Linux-x86_64.sh
$ ./Miniconda2-latest-Linux-x86_64.sh
```

Une fois conda installé, il reste à créer les 2 environnements utilisés et installer les logiciels et packages prérequis pour l'utilisation de RNAEditor et REDIttools.

3.2 L'environnement RNAEditor

3.2.1 Création de l'environnement RNAEditor

```
$ conda create -n RNAEditor numpy
$ source activate RNAEditor
```

Le dossier d'installation des futurs logiciels de l'environnement est :
/PATH/TO/miniconda/envs/RNAEditor/bin/

3.2.2 Installation des packages Python

```
$ conda install matplotlib
$ conda install pyqt=4.11.4
$ conda install -c bcbio pysam=0.8.4pre0
$ conda install -c conda-forge matplotlib-venn
$ conda install -c bioconda pybedtools
$ conda install -c bioconda mygene
```

Pour vérifier que les packages sont bien installés. Bien vérifier que la version de pyqt est la 4.11.4.

```
$ conda list
```

3.2.3 Installation des logiciels pré-requis

Installation avec Conda

— Blat

```
$ conda install -c bioconda blat=35
```

— Bwa

```
$ conda install -c bioconda bwa=0.7.15
```

— Samtools

```
$ conda install -c bioconda samtools=1.4
```

— Bedtools

```
$ conda install -c bioconda bedtools=2.26.0
```

— Trimmomatic

```
$ conda install -c bioconda trimmomatic
```

— FASTQC

```
$ conda install -c bioconda fastqc=0.11.4
```

— Vérification de l'installation des logiciels

```
$ conda list
```

Installation sans Conda

— JAVA

```
$ sudo apt install default-jre
```

— GCC

```
$ sudo apt install gcc
```

— Make

```
$ sudo apt install make
```

— Les outils Picards

Télécharger l'archive à cette adresse :

<https://sourceforge.net/projects/picard/files/picard-tools/1.119/picard-tools-1.119.zip/>

```
$ unzip picard-tools-1.119.zip
$ cd picard-tools-1.119
$ mkdir /PATH/TO/.conda/envs/RNAEditor/bin/picard-tools/
$ mv *.jar /PATH/TO/.conda/envs/RNAEditor/bin/picard-tools/
```

— GenomeAnalyseToolKit3.5

Télécharger l'archive à cette adresse :

<https://software.broadinstitute.org/gatk/download/archive>

Il est nécessaire d'avoir un compte sur le site pour pouvoir télécharger une archive.

Choisir la version 3.5.

```
$ mkdir /PATH/TO/.conda/envs/RNAEditor/bin/GATK/
$ mv GenomeAnalysisTK-3.5.tar.bz2 /PATH/TO/.conda/envs/RNAEditor/bin/GATK/
$ cd /PATH/TO/.conda/envs/RNAEditor/bin/GATK/
$ bunzip2 GenomeAnalysisTK-3.5.tar.bz2
$ tar -xvf GenomeAnalysisTK-3.5.tar
```

— Samstat 1.5.1

Télécharger l'archive à cette adresse :

<https://sourceforge.net/projects/samstat/>

```
$ gunzip samstat-XXX.tar.gz
$ tar -xvf samstat-XXX.tar
$ cd samstat-XXX
$ ./configure
$ make
$ make check
```

Si tout c'est bien passé.

```
$ cp src/samstat /PATH/TO/.conda/envs/RNAEditor/bin/
```

— Homer v4.9

Installer le logiciel :

```
$ wget http://homer.ucsd.edu/homer/configureHomer.pl
$ mkdir PATH/TO/Homer
$ cd PATH/TO/Homer
$ mv PATH/TO/configureHomer.pl
$ perl configureHomer.pl -install
$ perl configureHomer.pl -install hg19
```

Mettre le chemin d'accès des scripts dans le .bashrc

```
$ gedit PATH/TO/.bashrc
$ PATH="$PATH:PATH/TO/Homer/bin"
```

A la fin de toutes les installations et récupérations, ne pas oublier de sortir de l'environnement

```
$ source deactivate
```

3.2.4 Télécharger les fichiers nécessaires à RNAEditor

Les fichiers sont disponibles à cette adresse :

<http://rnaeditor.uni-frankfurt.de/download.php>

Je recommande de télécharger directement les archives GRCH37 ou GRCH38.

Voici comment télécharger l'archive pour hg19 :

```
$ wget http://141.2.194.197/rnaeditor/_annotations/GRCH37.tar.gz
```

3.3 L'environnement REDIttools

```
$ conda create -n REDIttools python
$ source activate REDIttools
```

Le dossier d'installation des futurs logiciels de l'environnement est :

/PATH/TO/miniconda/envs/REDIttools/bin/

3.3.1 Installation des packages Python

```
$ conda install -c bioconda pybedtools=0.6.9
$ conda install -c conda-forge numpy=1.12.1
$ conda install -c bioconda mygene
$ conda install -c anaconda pysam=0.6
```

Pour vérifier que les packages sont bien installés. Bien vérifier que la version de pysam est la 0.6.

```
$ conda list
```

3.3.2 Installation des logiciels pré-requis

Pour l'installation de samtools, matplotlib et matplotlib-venn suivre paragraphe 3.2.3

```
$ conda list
```

— Récupération du script psl_splicesites

Ce script permet la récupération des sites d'épissage dans un fichier refGene

```
$ wget http://research-pub.gene.com/gmap/src/gmap-gsnap-2017-04-13.tar.gz
$ gunzip gmap-gsnap-2017-04-13.tar.gz
$ tar -xvf gmap-gsnap-2017-04-13.tar
$ cd gmap-2017-04-13
$ ./configure
$ make
$ make check
$ make install
$ cp util/psl_splicesites /PATH/TO/miniconda/envs/REDIttools/bin/
```

3.3.3 Récupération des fichiers utilisés

— Fichier répertoriant les sites d'épissage

```
$ wget http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/refGene.txt.gz
$ gunzip -c refGene.txt.gz | psl_splicesites -s 1 > mysplicesites

$ gawk -F"|" '{split($2,a,":"); split(a[2],b,".");
if (b[1]>b[3])
    print a[1],b[3],b[1],toupper(substr($3,1,1)),"-";
else
    print a[1],b[1],b[3],toupper(substr($3,1,1)),"+"}' mysplicesites
> mysplicesites.ss
```

— Base de données de sites d'éditions connus

RADAR

. Base de données totale

```
$ wget http://lilab.stanford.edu/GokulR/database/Human_AG_all_hg19_v2.txt
```

. Par région

```
$ wget http://lilab.stanford.edu/GokulR/database/Human_AG_Alu_hg19_v2.txt
$ wget http://lilab.stanford.edu/GokulR/database/Human_AG_RepetitiveNonAlu_hg19_v2.txt
$ wget http://lilab.stanford.edu/GokulR/database/Human_AG_NonRepetitive_hg19_v2.txt
```

REDIportal

- . Base de données totale

```
$ wget http://srv00.recas.ba.infn.it/webshare/rediportalDownload/table1_full.txt.gz
```

- . Par région

```
$ gunzip -c REDiportal_full.txt.gz | awk '{if ($7 == "ALU"){print $0}}'
>REDiportal_ALU.txt
$ gunzip -c REDiportal_full.txt.gz | awk '{if ($7 == "NONREP"){print $0}}'
>REDiportal_NONREP.txt
$ gunzip -c REDiportal_full.txt.gz | awk '{if ($7 == "REP"){print $0}}'
>REDiportal_REP.txt
```

- Formatage du fichier

Pour chacun des fichier téléchargés, il faut le formater afin qu'il puisse être utilisé par REDIttools.

```
#EXEMPLE TAB_modification PATH/TO/fichier_entree PATH/OUT/fichier_sortie
$ TAB_modification PATH/TO/REDiportal_full.txt.gz PATH/OUT/REDiportal.tab
```

- Fichier contenant les variants connus

Aller à cette adresse :

<http://genome.ucsc.edu/cgi-bin/hgTables>

Reproduire cette capture d'écran

The screenshot shows the 'Table Browser' interface. At the top, there is a title bar and a detailed description of the tool's purpose. Below this, several configuration options are provided: 'clade' (Mammal), 'genome' (Human), and 'assembly' (Feb. 2009 (GRCh37/hg19)). The 'group' is set to 'Variation' and the 'track' is 'All SNPs(147)'. The 'table' is 'snp147'. The 'region' is set to 'genome'. The 'identifiers' section has 'paste list' and 'upload list' buttons. The 'filter' section has a 'create' button. The 'intersection' section has a 'create' button. The 'correlation' section has a 'create' button. The 'output format' is 'all fields from selected table'. The 'output file' is 'dbSNP.gtf'. The 'file type returned' is 'gzip compressed'. At the bottom, there are 'get output' and 'summary/statistics' buttons, and a link to reset settings.

FIGURE 2 – Procédure pour obtenir les SNPs connus

4 Utilisation du pipeline RNAditing

Le pipeline peut s'utiliser de 2 façons :

- Dans sa totalité en suivant le déroulement schématisé sur la figure 1
- En utilisant et réalisant les analyses des 2 logiciels séparément.

4.1 Utilisation totale du pipeline

Pour une analyse d'échantillons issus de plusieurs populations, il est nécessaire de ranger les échantillons par population. Chaque dossier est nommé par le nom de la population et contient tous les échantillons associés. Enfin, un dossier doit regrouper l'ensemble des dossiers des populations.

Afin de rendre plus lisible l'analyse des résultats un fichier tabulé spécifique doit être donné en entrée au pipeline.

Pour chaque fichier fastq doit être associé le nom de la population (colonne 1), le chemin d'accès complet au fichier (colonne 2) ainsi que le nom de l'échantillon (colonne 3). Ce nom sera utilisé pour nommer l'échantillon dans les résultats. Pour le cas de séquençage paired-end, seul le chemin du fastq_1 est nécessaire.

Sur la figure 3, vous pouvez voir un exemple du fichier demandé.

MN	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/MN/NG-7624_MN_8_lib51967_2845_5_1.paired.fq	8_MN
MN	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/MN/NG-7624_MN_9_lib51968_complete_read_1.paired.fq	9_MN
MN	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/MN/NG-7624_MN_3_lib51966_complete_read_1.paired.fq	3_MN
Me	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Me/NG-7624_NCN_A19_Me_lib51960_complete_read_1.paired.fq	A19_Me
Me	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Me/NG-7624_NCN_A17_Me_lib51956_complete_read_1.paired.fq	A17_Me
Me	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Me/NG-7624_NCN_A21_Me_lib51964_complete_read_1.paired.fq	A21_Me
Me	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Me/NG-7624_NCN_A16_Me_lib51952_complete_read_1.paired.fq	A16_Me
Na	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Na/NG-7624_NCN_A19_Na_lib51958_complete_read_1.paired.fq	A19_Na
Na	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Na/NG-7624_NCN_A21_Na_lib51962_2845_5_1.paired.fq	A21_Na
Na	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Na/NG-7624_NCN_A17_Na_lib51954_complete_read_1.paired.fq	A17_Na
Na	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/Na/NG-7624_NCN_A16_Na_lib51950_complete_read_1.paired.fq	A16_Na
GC	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/GC/NG-7624_NCN_A16_GC_lib51951_complete_read_1.paired.fq	A16_GC
GC	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/GC/NG-7624_NCN_A17_GC_lib51955_complete_read_1.paired.fq	A17_GC
GC	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/GC/NG-7624_NCN_A21_GC_lib51963_complete_read_1.paired.fq	A21_GC
GC	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/GC/NG-7624_NCN_A19_GC_lib51959_complete_read_1.paired.fq	A19_GC
PL	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/PL/NG-7624_NCN_A16_PL_lib51953_complete_read_1.paired.fq	A16_PL
PL	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/PL/NG-7624_NCN_A17_PL_lib51957_complete_read_1.paired.fq	A17_PL
PL	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/PL/NG-7624_NCN_A21_PL_lib51965_2845_4_1.paired.fq	A21_PL
PL	/media/btointf/aeaf2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/Fastq/PL/NG-7624_NCN_A19_PL_lib51961_complete_read_1.paired.fq	A19_PL

FIGURE 3 – Exemple du fichier nécessaire pour le pipeline

Il existe, à l'intérieur de l'archive téléchargée pour l'utilisation de RNAEditor, un fichier configuration.txt. Le nom du fichier ne doit pas être changé. Dans ce fichier doivent être répertoriés les chemins d'accès des différents fichiers utilisés par RNAEditor ainsi que du dossier contenant les exécutables des différents programmes téléchargés (/PATH/TO/miniconda/envs/RNAEditor/bin/).

Ainsi, afin d'utiliser le pipeline voici les commandes à réaliser.

```
$ source activate RNAEditor
$
# Exemple
$ RNAditing -A -a -c PATH/TO/configuration.txt -f PATH/TO/DOSSIER_FASTQ/
-s PATH/TO/fichier_echantillon.txt -o /PATH/TO/DOSSIER_BAM/ -p 0.75 -e paired
```

Ainsi, l'option :

- -c : renseigne le chemin du fichier configuration.txt nécessaire à RNAEditor.
- -f : renseigne le dossier contenant les sous-dossiers de populations.
- -s : permet de donner le chemin d'accès du fichier d'échantillons.
- -o : renseigne le chemin du dossier contenant les fichiers BAM qui seront créés par RNAEditor et utilisés ensuite par REDIttools.
- -p : pourcentage minimal de présence d'un site dans une population pour être considéré comme représentatif de celle-ci.

Exemple 0.75 : dans une population de 4 échantillons, le site doit être retrouvé chez au moins 3 échantillons pour représenter la population.

-
- -e : Pour renseigner le pipeline d'un séquençage paired-end ou single-end (mot-clés : single / paired)

Chaque fichier de chacun des dossiers va tout d'abord subir un contrôle qualité et va ensuite être traité par le logiciel RNAEditor. Ainsi, des résultats individuels répertoriant les sites d'édérations de l'ARN prédits dans chacun des échantillons vont être obtenus.

Afin de connaître les sites représentatifs d'une population, des étapes de comparaisons entre échantillons de même population sont réalisées suivies d'étapes d'annotation, de récupération de gènes édités et de création de graphiques.

Un message diffusé sur la sortie standard indique la suite de la procédure à adopter (figure 4).



```
##### Prediction ok: #####
# Path to RNAEditor results directory: /media/bioinf/faef2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/RNAEditor/ #
# Path to BAM files: /media/bioinf/faef2f7b-527b-4b8a-859e-72ad73bd98bc/TEST/BAM/ #
#### To continue the analysis please make the following command lines ####
$ source deactivate
$ source activate REDIttools
$ python RNAAditing.py -E -a -i /media/bioinf/faef2f7b-527b-4b8a-859e-72ad73bd98bc/TEST/BAM/ -c /media/bioinf/faef2f7b-527b-4b8a-859e-72ad73bd98bc/New_fastq/paired/RNAEditor/ [options]
```

FIGURE 4 – Message de sortie à la fin de l'étape de prédiction

Afin de continuer l'analyse, il faut maintenant réaliser les lignes de commandes suivantes.

```
$ source deactivate
$ source activate REDIttools
$
# Exemple
$ RNAAditing -E -a -i PATH/TO/BAM_folder -d PATH/TO/fichier_database
-r PATH/TO/Genome_ref.fasta -o PATH/TO/DOSSIER_RESULTATS/ -p 0.75
-c /PATH/TO/DOSSIER_RESULTAT_RNAEDITOR/
```

Ainsi, l'option :

- -i : chemin du dossier contenant les fichiers BAM de chacun des échantillons
- -r : chemin du génome de référence au format fasta
- -d : chemin du fichier contenant les sites présents dans les bases de données
- -o : renseigne le chemin du dossier contenant les futurs résultats
- -p : pourcentage minimal
- -c : chemin du dossier contenant les résultats de RNAEditor

Des options supplémentaires sont disponibles :

- -v : chemin d'accès du fichier dbSNP.
- -s : chemin d'accès du fichier contenant les sites d'épissage.

Le logiciel REDIttools va rechercher les sites de RNA Editing connus dans chacun de nos échantillons de façon individuelle. Ensuite, une comparaison des sites entre échantillons d'une même population va être réalisée afin de garder uniquement les sites représentatifs de la population. Enfin des étapes d'annotation de récupération de gènes édités et de création de graphiques vont finir cette partie d'analyse.

Les résultats du logiciel REDIttools sont disponibles à l'endroit spécifié par l'utilisateur (option -o).

La dernière grande étape consiste à récupérer les sites d'édérations communs aux deux logiciels. Celle-ci se fait automatiquement.

Un fichier par population est donc créé recensant tous les sites d'édérations de l'ARN représentatifs d'une population et retrouvés dans les bases de données.

Cependant, afin de ne pas perdre d'informations sont aussi recensés les sites d'édérations de l'ARN

considérés comme de novo. Des étapes d’annotations des sites de RNA Editing sont également réalisées tout comme la récupération de gènes édités et la création de graphiques.

4.2 Utilisation séparée des logiciels

Il est aussi possible d’utiliser les logiciels séparément.

Pour ce faire, je vous renvoie vers les différentes aides disponibles :

— Pour RNAEditor

```
#####
# RNAditing: RNA Editing sites research and prediction on RNA-Seq data #
# VERSION: 1.0.0 #
# Author: Benjamin Delaune, Bioinformatic Student, Team 11 CRCINA #
#####

USAGE: RNAditing --RnaEditor [-i or -m or -a or -n] [options]
Options:

-i or --GUI                Use this option alone for use the RNAEditor software with its interface.

-a or --all                Use this option for use RNAEditor in commande line and make a comparison by population

    -c or --configure      Use this option to give to RNAEditor the configuration file path (needed)
    -f or --fastq          Use for specify the FASTQ folder used (needed)
    -p or --percentage      Minimum percentage of a site for it to be considered representative of a population : 0-1 (needed)
    -s or --sample         Used for give the samples file path (needed)
    -e or --end            Paired-end or single-end sequencing. Write paired or single (needed)

-n or --analyse            Use this option for use RNAEditor in commande line (needed)

    -c or --configure      Use this option to give to RNAEditor the configuration file path (needed)
    -f or --fastq          Use for specify the FASTQ folder used (needed)
    -s or --sample         Used for give the samples file path (needed)
    -e or --end            Paired-end or single-end sequencing. Write paired or single (needed)

-m or --merge              Use for make a comparison by population (MN_3, MN_8, MN_9 for example)

    -f or --fastq          Use for specify the FASTQ folder used (needed)
    -p or --percentage      Minimum percentage of a site for it to be considered representative of a population : 0-1 (needed)
```

FIGURE 5 – Aide d’utilisation de RNAEditor

— Pour REDIttools

```
#####
# RNAditing: RNA Editing sites research and prediction on RNA-Seq data #
# VERSION: 1.0.0 #
# Author: Benjamin Delaune, Bioinformatic Student, Team 11 CRCINA #
#####

*** Options must be given ***

USAGE: RNAditing --ESknowns [-m or -a or -n] [options]
Options:

-a or --all                Use this option for use REDIttools in commande line and make a comparison by population

    -d or --Database       Database with editing sites already known (needed)
    -i or --Bam            Path to directory containing BAMs to analyzed (needed)
    -r or --Reference       Path to reference fasta file (needed)
    -s or --Splice         Path to splice file (splicites.ss)
    -v or --variant        Path to variant file (dbSNP)
    -o or --output         Path to results folder (needed)
    -p or --percentage      Minimum percentage of a site for it to be considered representative of a population : 0-1 (needed)
    -c or --software_comp  Path to RNAEditor results (use if you want to make a softwares results comparison)

-n or --analyse            Use this option for use REDIttools in commande line

    -d or --Database       Database with editing sites already known (needed)
    -i or --Bam            Path to directory containing BAMs to analyzed (needed)
    -r or --Reference       Path to reference fasta file (needed)
    -s or --Splice         Path to splice file (splicites.ss)
    -v or --variant        Path to variant file (dbSNP)
    -o or --output         Path to results folder (needed)

-m or --merge              Use for make a comparison by population (For MN_3, MN_8, MN_9 for example)

    -d or --Database       Database with editing sites already known (needed)
    -o or --output         Path to results folder (needed)
    -p or --percentage      Minimum percentage of a site for it to be considered representative of a population : 0-1 (needed)

## For the comparison with RNAEditor add :

-c or --software_comp      Path to RNAEditor results (needed)
```

FIGURE 6 – Aide d’utilisation de REDIttools