
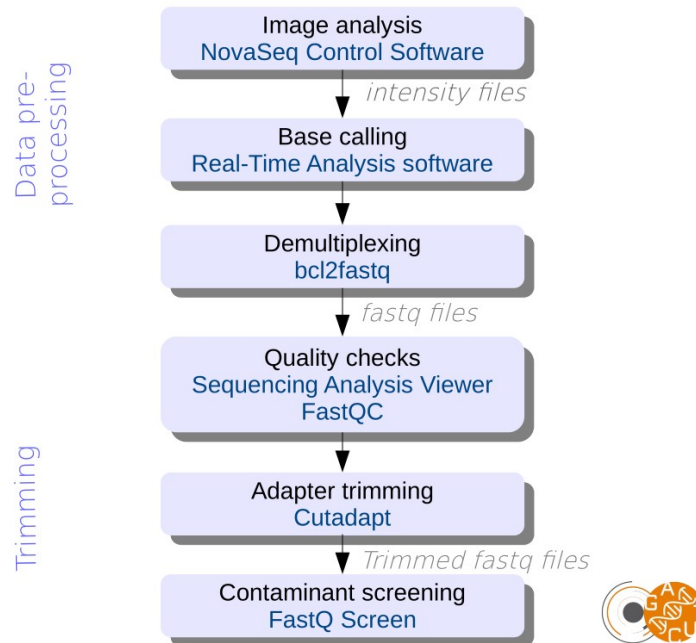


E-EXP-09	Fiche prestation	Date : 28/03/23 Version 9
	<h2>Analyse bioinformatique small RNA-seq</h2>	Page 1/4

Cette technique permet une analyse quantitative et qualitative des micro-ARN (miARN) contenus dans un échantillon.

Prestation proposée

A partir des données brutes de séquençage des échantillons, le plateau technique réalise les étapes décrites dans les paragraphes suivants.



Production des fichiers fastq et démultiplexage des données

Pendant le séquençage, l'analyse d'images et le *base calling* sont réalisés en temps réel par les logiciels embarqués sur le séquenceur. À l'issue du séquençage, le démultiplexage et la production des fichiers fastq est réalisée grâce au logiciel Illumina bcl2fastq.

Contrôle qualité des données

Le contrôle qualité s'appuie sur plusieurs critères :


- validation du run en utilisant une série de critères associés,
- distribution des scores de qualité à chaque cycle,
- distribution des scores moyens de qualité par séquence,
- pourcentage de bases "N" par cycles,
- distribution des bases par cycle.

Nettoyage des séquences d'adaptateur

Les miARN ayant une taille comprise entre 21 et 24 nt, une étape de *trimming* des adaptateurs est nécessaire afin d'enlever les séquences d'adaptateurs lues en 3' des *reads* séquencés. Pour réaliser cette étape, le logiciel [Cutadapt](#) (Martin, 2011) est utilisé.

La recherche de contaminants est réalisée après nettoyage des séquences d'adaptateurs.

Les étapes de contrôle qualité et de nettoyage des séquences d'adaptateurs sont réalisées systématiquement (bioinfo niveau 1). Les étapes suivantes sont réalisées sur demande (bioinfo niveau 2).

E-EXP-09	Fiche prestation	Date : 28/03/23 Version 9
	<h2 style="text-align: center;">Analyse bioinformatique small RNA-seq</h2>	Page 2/4

Quantification des miARN connus

Les séquences obtenues sont alignées sur le génome de référence (la version du génome doit être indiquée par le client). L'annotation des miARN est téléchargée, lorsqu'elle est disponible, sur miRBase. Les reads sont alignés sur le génome de référence à l'aide du logiciel [Bowtie](#) (Langmead *et al.*, 2009), puis la quantification des miARN décrits dans le fichier d'annotation est réalisée à l'aide de [featureCounts](#) (Liao *et al.*, 2014). Cela permet d'obtenir une matrice de comptage pour chaque miARN connu et chaque échantillon séquencé.

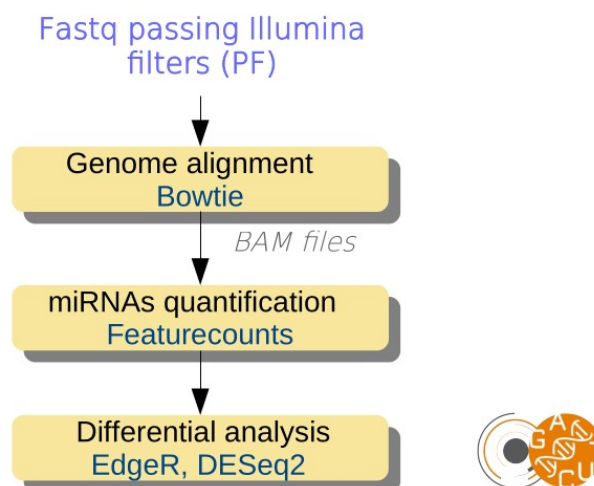
En l'absence d'annotation disponible, il est possible d'utiliser l'annotation d'une espèce proche et d'utiliser le logiciel [Prost!](#) (Desvignes *et al.*, 2019) pour aligner, quantifier et annoter les reads *trimmés*.

Prédiction de nouveaux miARN


Après alignement des séquences sur le génome de référence, il est possible d'utiliser l'outil [miRDeep2](#) (Friedländer *et al.*, 2012) pour prédire de nouveaux miARN, qui ne seraient pas décrits dans le fichier d'annotations. Pour ce faire, miRDeep2 calcule la stabilité énergétique des structures secondaires de précurseurs de miARN prédites en fonction des résultats de l'alignement.

Analyse différentielle

Pour plus de détail, voir la fiche prestation E-EXP-11 « Analyse bioinformatique RNA-seq ».



Pour chaque package, un fichier est généré avec la liste des miARN différentiellement exprimés. Ce fichier contient pour chaque miARN: un *fold-change* (log du rapport des deux moyennes observées entre les deux conditions), une *p-value* et une *p-value* ajustée après correction des tests multiples.

E-EXP-09	Fiche prestation	Date : 28/03/23 Version 9
	Analyse bioinformatique small RNA-seq	Page 3/4

Prestations complémentaires

Des analyses complémentaires peuvent être effectuées selon la demande du client (bioinfo niveau 3).

Prédiction de cible ARNm

Pour des expériences menées sur des échantillons humains, de souris ou de rat, il est possible de rechercher les cibles potentielles des miARN différentiellement exprimés à l'aide de la base de données [miRGate](#).

Annotation fonctionnelle

L'étape d'enrichissement fonctionnel consiste en l'analyse statistiques de termes *Gene Ontology* associés aux ARNm interagissant avec des miARN différentiellement exprimés. Les annotations reposant sur des prédictions d'interactions doivent permettre d'orienter les futures recherches expérimentales associées au projet.

Matériel initial

Les données nécessaires à l'analyse bio-informatique sont directement issues d'un des séquenceurs du plateau technique.

En revanche, ce type d'analyse n'est envisageable que sur des espèces pour lesquelles il existe un génome de référence, et éventuellement une annotation (position des gènes). Dans le cas où il n'y a pas d'annotation pour l'espèce, seule la prédiction de nouveaux miARN pourra être envisagée.

Restitution des résultats

A l'issue des analyses, plusieurs fichiers sont disponibles :

- Un rapport d'analyse au format PDF disponible depuis le logiciel de gestion de projet redmine

Pour chaque échantillon (*) :

- *.fastq.gz : fichier texte (compressé) contenant les séquences nucléotidiques ainsi que les scores de qualité correspondants

Pour chaque échantillon (*), en cas d'alignement :


- *.bam : le format bam est la version binarisée (compressée) du format sam. Ce fichier contient les résultats de l'alignement, et permet notamment leur visualisation avec un logiciel comme IGV.
- *.bam.bai : fichiers d'index allant de paire avec les fichiers *.bam. Cette paire de fichier vous permet de visualiser les alignements sur IGV.

En cas d'alignement et de quantification avec Prost! :

- Fichier excel prost_output.xlsx : résultats fournis par le logiciel. En particulier, l'onglet by_annotation contient la matrice de comptages par miARN annoté pour les différents échantillons séquencés.

Pour chaque échantillon (*), en cas de prédiction de nouveaux miARN :

- *.arf : format d'alignement spécifique à l'outil miRDeep2.
- *.bed : fichiers bed contenant des informations sur les miARN prédits.
- *.html : fichier à ouvrir avec votre navigateur web. Contient des tableaux sur les résultats.

E-EXP-09	Fiche prestation	Date : 28/03/23 Version 9
	Analyse bioinformatique small RNA-seq	Page 4/4

Pour chaque comparaison, en cas d'analyse différentielle :

- fichiers excel contenant les miARNs présentant une expression différentielle
- comptages bruts et comptages normalisés
- images des MA-plots et des nuages de points entre paires d'échantillons
- images des volcano-plots des gènes différentiellement exprimés entre conditions

Pour chaque comparaison, en cas de prédiction de cibles:

- *p-value0_05_miRNAs_miRGate.xls : fichiers contenant la liste des miARN et les ARNm avec lesquels ils interagissent.
- *p-value0_05_miRNAs_miRGate_1_5.xls : fichiers contenant la liste des miARN et les ARNm avec lesquels ils interagissent avec un seuil de score d'interaction supérieur à 1,5.
- Liste des termes GO statistiquement surreprésentés.

Pour chaque comparaison, en cas d'analyse Gene Ontology :

- fichiers html contenant les termes significativement enrichis
- fichiers csv contenant les termes significativement enrichis
- images des sous-graph GO des termes significativement enrichis

Les fichiers volumineux (notamment les fichiers fastq.gz et fichiers bam) sont rendus disponibles sur le serveur sftp du plateau, à partir de la mise en ligne du rapport sur le gestionnaire de projet. Les fichiers peu volumineux (comme les résultats des analyses différentiels), sont mis en ligne sur le gestionnaire de projet. Le serveur sftp est accessible par identifiant et mot de passe, fournis avec le rapport d'analyse.

Durée de conservation des données

Les documents mis en ligne sur le gestionnaire de projet n'ont pour l'instant aucune limite de validité.

En revanche, les fichiers qui sont mis en ligne sur le serveur sftp n'y sont hébergés que pour une durée de 10 jours à compter de l'édition du rapport de résultat.

Le dépôt des données brutes dans une base de données peut être demandé avant publication ; nous ne nous engageons pas à conserver ces données.