


E-EXP-21	Fiche prestation	Date : 24/02/2020 Version 1
	Analyse de cribles CRISPR/Cas9	Page 1/3

Les cribles génétiques CRISPR/Cas9 permettent d'introduire des mutations dans l'ensemble des gènes connus et ainsi d'étudier leurs fonctions. La représentation des guides ARN (*single guide RNA* - sgRNA) est déterminée par séquençage haut-débit. Une analyse bioinformatique permet ensuite de quantifier les sgRNA, puis de déterminer leur enrichissement (dans le cas de cribles positifs) ou leur appauvrissement (cibles négatifs).

Prestation proposée

A partir des données brutes de séquençage des échantillons d'ARN d'intérêt, le plateau technique réalise les étapes de :

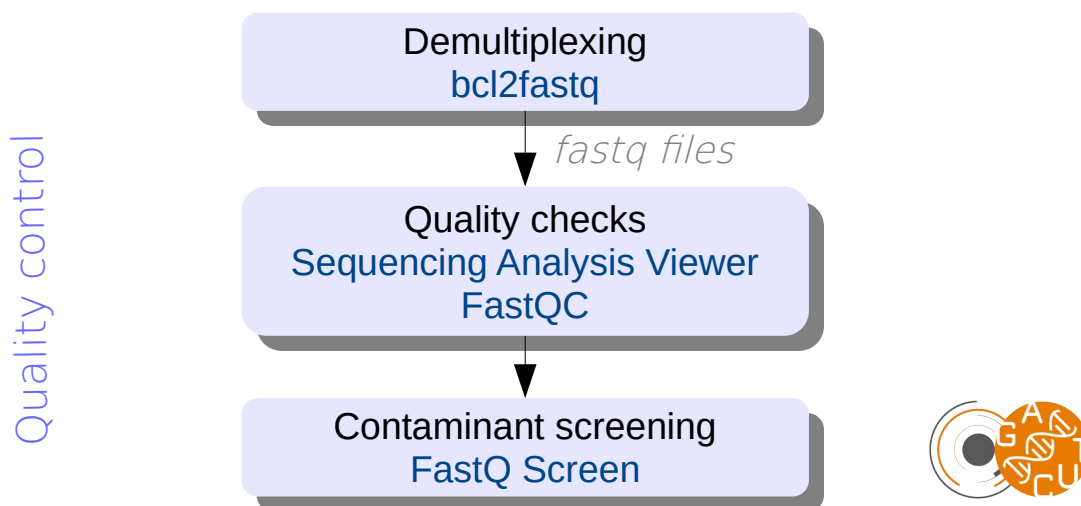
Production des fichiers fastq et démultiplexage des données

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
- recherche de contaminants / adaptateurs




Cette étape est réalisée systématiquement (bioinfo niveau 1). Les étapes suivantes sont réalisées sur demande (bioinfo niveau 2).

Nettoyage des séquences

Les guides ARN (sgRNA) sont généralement des séquences d'une vingtaine de nucléotides. La première étape de l'analyse consiste donc à identifier et supprimer (*trimmer*) les séquences des primers et des éventuelles bases décalées (*stagers*) qui sont lues autour de la séquence du sgRNA dans les reads séquencés. Cette étape est réalisée avec le logiciel [Cutadapt](#) (Martin, 2011).

Alignement et quantification

E-EXP-21	Fiche prestation	Date : 24/02/2020 Version 1
 MGX	Analyse de cribles CRISPR/Cas9	Page 2/3

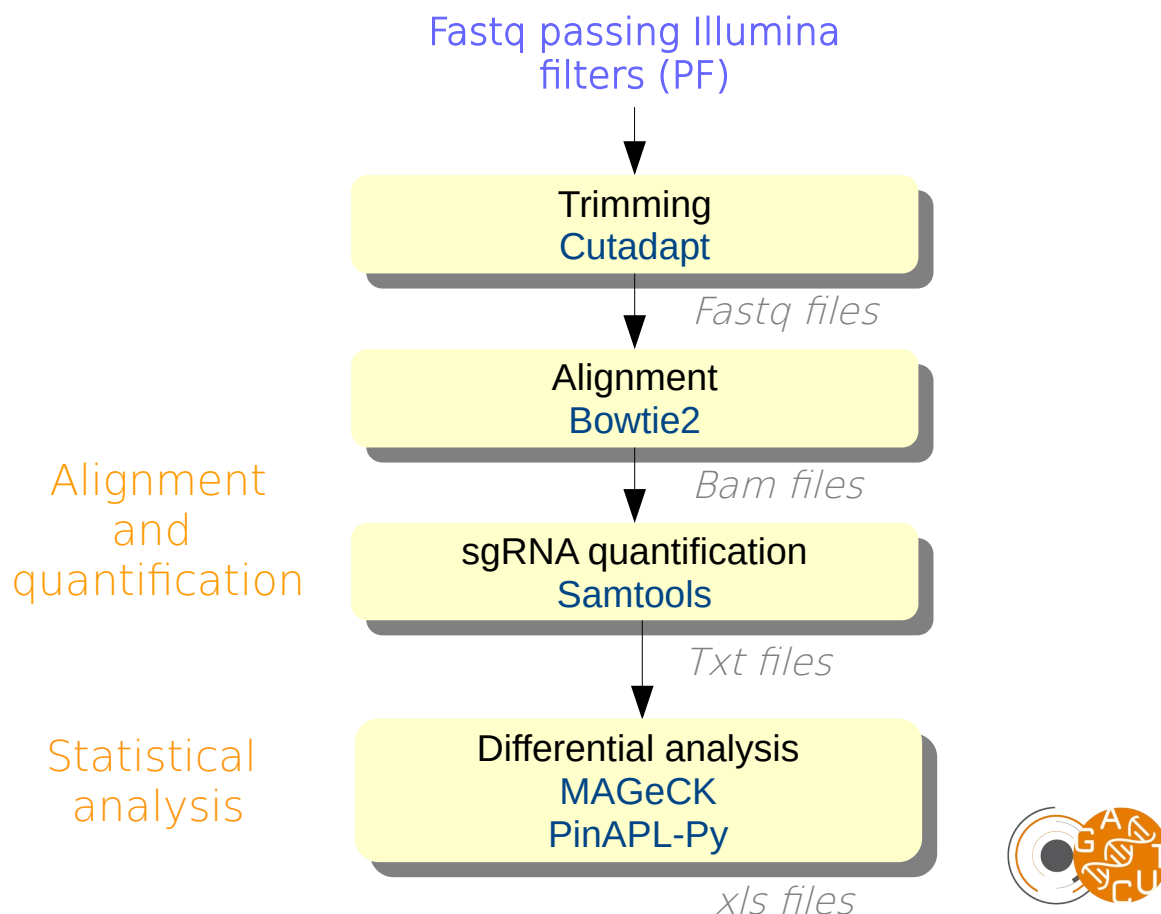
Les séquences obtenues sont alignées sur les séquences de la banque de référence, afin d'identifier chaque sgRNA séquencé. Cette étape est réalisée avec le logiciel [Bowtie2](#) (Langmead & Salzberg, 2012).

Le nombre de reads s'alignant sur chaque sgRNA de la référence est ensuite quantifié avec l'outil [Samtools](#) idxstats (Li *et al.*, 2009).

Analyse statistique

En présence de réplicats, il est possible de réaliser une analyse statistique afin de déterminer les sgRNA enrichis ou appauvris dans une condition par rapport à une autre.

Pour cela, les logiciels [MAGeCK](#) (Li *et al.*, 2014) ou [PinAPL-Py](#) (Spahn *et al.*, 2017) peuvent être utilisés. Les comptages par sgRNA sont modélisés selon une loi binomiale négative puis un test statistique permet d'identifier les sgRNA qui ont des comptages significativement différents entre les deux conditions. Un test supplémentaire est ensuite réalisé pour identifier les gènes significatifs, c'est-à-dire, pour résumer, ceux qui ont plusieurs sgRNA significatifs.




Matériel initial

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

Restitution des résultats

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

E-EXP-21	Fiche prestation	Date : 24/02/2020 Version 1
	Analyse de cribles CRISPR/Cas9	Page 3/3

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

Pour chaque échantillon (*) :

- *.fastq : fichier texte contenant les séquences nucléotidiques ainsi que les scores de qualité correspondants
- *.cutadapt.fastq : fichier texte contenant les séquences nucléotidiques ainsi que les scores de qualité correspondants, après *trimming*

En cas d'alignement et de quantification des sgRNA :

- *.bam : le format bam est la version binarisée (compressée) du format sam. Ce fichier contient les résultats de l'alignement
- *counts.txt : comptages bruts par sgRNA.

En cas d'analyse différentielle, un dossier par comparaison contenant les fichiers résultats des analyses :

- fichiers tabulés avec les sgRNA ayant une expression différentielle, et leur p-value associée
- comptages bruts et comptages normalisés
- fichiers tabulés avec les gènes ayant une expression différentielle, et leur p-value associée.

L'ensemble de ces fichiers est disponible sur le serveur SFTP du plateau, à partir de la mise en ligne du rapport sur le gestionnaire de projet. Ce serveur est accessible par login et mot de passe, fournis avec le rapport d'analyse.

Durée de conservation des données

Les login et mot de passe pour accéder au gestionnaire de projet et les documents qui y sont mis en ligne 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.