

# Fonctionnement et recommandations de la plateforme GeT-Biopuces

Laboratoire TBI  
(Toulouse Biotechnology Institute)  
INSA\_Bâtiment 39  
135 Avenue de Ranguel  
31077 Toulouse Cedex 04

**Directrice de la plateforme:**

Marie Ange Teste : [marie-ange.teste@insa-toulouse.fr](mailto:marie-ange.teste@insa-toulouse.fr)

**Responsables séquençage :**

Lidwine Trouilh : [lidwine.trouilh@insa-toulouse.fr](mailto:lidwine.trouilh@insa-toulouse.fr)

Nathalie Marsaud : [nathalie.marsaud@insa-toulouse.fr](mailto:nathalie.marsaud@insa-toulouse.fr)

**Responsables analyses de données :**

Delphine Labourdette : [delphine.labourdette@insa-toulouse.fr](mailto:delphine.labourdette@insa-toulouse.fr)

Etienne Rifa : [etienne.rifa@insa-toulouse.fr](mailto:etienne.rifa@insa-toulouse.fr)

**Responsable R&D**

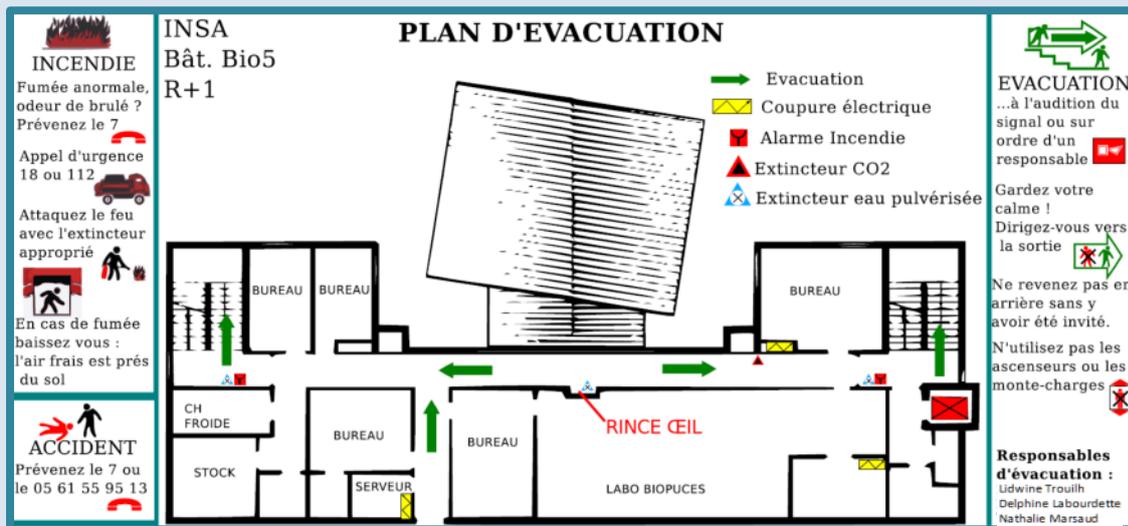
Jean-Luc Parrou : [jean-luc.parrou@insa-toulouse.fr](mailto:jean-luc.parrou@insa-toulouse.fr)

**Mail plateforme :** [biopuces@insa-toulouse.fr](mailto:biopuces@insa-toulouse.fr)

**Téléphone :** +33 (0)5-61-55-96-87

## REMARQUES IMPORTANTES

- Durant le projet et selon leur type, les échantillons seront stockés par la plateforme au congélateur ou au réfrigérateur. **Le demandeur est invité à récupérer ses échantillons dans un délai d'un mois après la fin de la prestation. Passé ce délai, nous ne garantissons plus leur conservation.**
- Nos micro-pipettes sont contrôlées annuellement par une société extérieure. Nos congélateurs sont reliés à un système de suivi de température et la température des réfrigérateurs est relevée hebdomadairement
- A partir du moment où **le contrôle qualité des échantillons est validé et s'il n'y a pas de problème au cours de la prestation**, les expériences sont **généralement réalisées dans un délai de 4 semaines.**
- **La durée de stockage des données informatiques**, scan, analyse d'image, analyses statistiques et analyses de séquençage **est de 1 an (sauf avis contraire de votre part).**
- Les informations relatives à toutes les manipulations faites sur vos échantillons sont archivées dans votre dossier client à la plateforme.
- **Nous vous rappelons que le devis est valable pour les prestations et quantités mentionnées dans celui-ci, toutes les expériences refaites ou supplémentaires feront l'objet d'une nouvelle facturation.**
- **Toute prestation facturée et non réalisée dans un délai de 12 mois ne sera pas remboursée.** Passé ce délai, si vous souhaitez faire la prestation, elle sera réévaluée en fonction des tarifs, consommables et personnels en vigueur et un devis complémentaire sera réalisé.
- Merci de prendre connaissance du plan d'évacuation de la plateforme GeT-Biopuces:



## TABLE DES MATIERES

<b>1. GENERALITES .....</b>	<b>4</b>
<b>2. EXTRACTION DES ADNS LONG FRAGMENT (ADN HMW) .....</b>	<b>4</b>
<b>3. CONTROLE QUALITE DES ARNS ET ADNS .....</b>	<b>5</b>
3.1. ADNs.....	5
3.2. ARNs.....	6
<b>4. SEQUENÇAGE .....</b>	<b>6</b>
4.1. MATERIEL DE DEPART .....	7
4.2. SYNTHESE DES LIBRAIRIES.....	7
4.3. SEQUENÇAGE .....	8
<b>5. ANALYSES.....</b>	<b>9</b>
5.1. CONTROLE QUALITE APRES SEQUENÇAGE .....	9
5.2. ANALYSES STATISTIQUES DES DONNEES ( <i>QUELLE QUE SOIT LEUR ORIGINE</i> ).....	10
5.2.1 <i>RNAseq</i> .....	10
5.2.2 <i>RNAseq sur cellules uniques (single cell)</i> .....	11
5.2.3 <i>Métagénomique ciblée</i> .....	11
<b>6. RENDU DES RESULTATS.....</b>	<b>11</b>
<b>7. REGLES DE CONFIDENTIALITE, DE PROPRIETE INTELLECTUELLE ET DE VALORISATION .....</b>	<b>12</b>

## 1. GENERALITES

La plateforme GeT-Biopuces réalise vos expériences depuis le contrôle qualité des échantillons jusqu'à l'analyse des données. Elle met à votre disposition l'équipement nécessaire pour réaliser vos expériences de séquençage ainsi que le personnel pour analyser vos données quelle que soit leur provenance.

Avant de démarrer un projet, il est nécessaire de contacter soit :

- la responsable de la plateforme Marie Ange Teste ([marie-ange.teste@insa-toulouse.fr](mailto:marie-ange.teste@insa-toulouse.fr))
- les responsables séquençage, Nathalie Marsaud ([nathalie.marsaud@insa-toulouse.fr](mailto:nathalie.marsaud@insa-toulouse.fr)) ou Lidwine Trouilh ([lidwine.trouilh@insa-toulouse.fr](mailto:lidwine.trouilh@insa-toulouse.fr))
- les responsables analyse de données, Delphine Labourdette ([delphine.labourdette@insa-toulouse.fr](mailto:delphine.labourdette@insa-toulouse.fr)) et Etienne Rifa ([etienne.rifa@insa-toulouse.fr](mailto:etienne.rifa@insa-toulouse.fr)) pour les analyses bioinformatiques et statistiques de données de séquençage (métagénomique ciblée, RNAseq, single-cell RNAseq, recherche de variants après séquençage ADN sur les micro-organismes, assemblage de petits génomes) ainsi que les analyses d'enrichissement fonctionnel.

Afin de répondre au mieux à votre demande, nous définissons ensemble le contenu des expériences à réaliser. A l'issue de notre discussion, nous vous communiquons un devis ainsi que la fiche prestation. Cette dernière (accompagnée du bon de commande ou seulement du devis signé pour les prestations internes et sociétés privées) doit être complétée, signée et envoyée à la Plateforme GeT-Biopuces avant le début des expériences. Vous pouvez nous l'envoyer par courrier avec vos échantillons ou par mail après l'avoir scannée.

Une fois reçus, vos échantillons seront contrôlés. Après validation des échantillons, vos expériences seront planifiées puis réalisées et vos résultats vous seront envoyés dans un délai moyen de 1 mois. Ce délai pourra être allongé en fonction du nombre d'échantillons à traiter. Bien entendu, si à une étape (du contrôle qualité à l'analyse des résultats) nous rencontrons des difficultés, nous vous en ferons part et nous déciderons ensemble de la suite à mener.

**Nous ne réalisons pas les extractions des ARNs ni des ADNs, sauf dans le cas des ADN long fragment (HMW = High Molecular Weight) de bactéries, de levures et d'algues.**

## 2. EXTRACTION DES ADNs LONG FRAGMENT (ADN HMW)

La plateforme réalise des extractions d'ADN HMW pour les bactéries Gram+, Gram-, les levures et les algues. La quantité de cellules vous sera donnée en fonction du kit utilisé.

Après avoir prélevé la quantité nécessaire de culture, il est recommandé de centrifuger, éliminer le milieu puis de laver 2-3 fois le culot cellulaire. Après le dernier lavage, il faut enlever au maximum le tampon de lavage, plonger rapidement le culot dans l'azote liquide et le stocker au -80°C ; ou bien stocker les culots directement à -20°C après les lavages.

Nous vous conseillons d'utiliser des tubes de bonne qualité et 'safe lock' afin qu'ils ne se détériorent pas avec l'azote, ni au moment du cassage des cellules. L'envoi des cellules se fait au minimum à -20°C.

### 3. CONTROLE QUALITE DES ARNs ET ADNs

Ils doivent nous être livrés, congelés dans de l'eau.

Les contrôles qualités sont réalisés sur puces RNA 6000 Lab-on-Chip Bioanalyzer de chez Agilent (Electrophorèse capillaire), sur les spectrophotomètres Nanodrop (ND2000/8000) ou le Spectrostar nano (pour les plaques 96 puits) ou sur le Qubit (dosage spécifique des acides nucléiques double brin et simple brin). Ils permettent une validation qualitative et quantitative des ARNs et des ADNs.

#### 3.1. ADNs

Les ADNs sont d'abord contrôlés sur un des spectrophotomètres et/ou au Qubit. Les minimums requis sont :

- un **ratio 260nm/230nm et 260nm/280nm  $\geq$  à 1,8**
- une **concentration minimale d'échantillon** nécessaire pour l'expérience choisie (celle-ci est fournie dans les recommandations d'utilisation du fournisseur du kit).

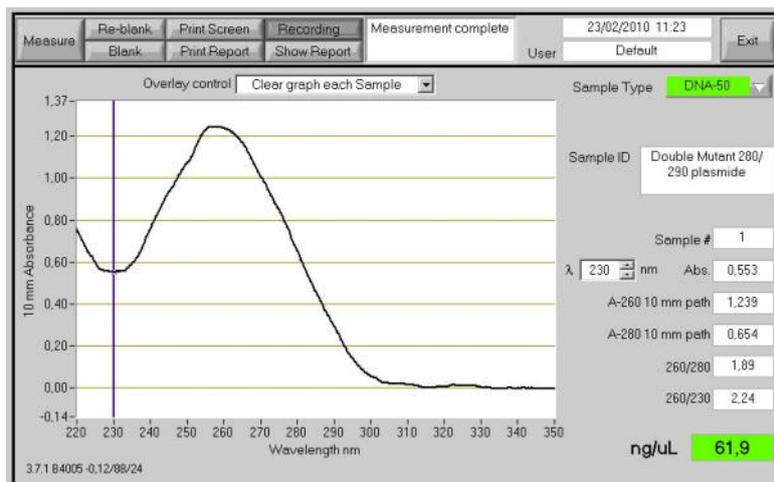


Fig 2: Profil d'ADN correct après lecture au nanodrop

Ensuite, nous demandons à ce qu'une migration des ADNs sur gel d'agarose soit réalisée. Un ADN de qualité correcte et utilisable dans une expérience de séquençage ne doit présenter qu'une seule bande de migration dans les grandes tailles et non un smear.

Au vu de la complexité des ADN destinés au séquençage métagénomique, les critères qualités cités ci-dessus ne s'appliquent pas mais une quantification au spectrophotomètre et/ou au Qubit est nécessaire.

Pour le séquençage amplicon Miseq, ce dosage est réalisé par nos soins après purification de vos PCR1.

Pour le séquençage long fragment, les ratios 260/280 et 260/230 doivent être compris entre 1.8 et 2.2. Il faut que la différence entre la concentration Nanodrop et la concentration Qubit soit au maximum d'un facteur 2. De plus, il faut qu'un minimum de 60% des fragments ait une taille supérieure à 30kb.

## 3.2. ARNs

En transcriptomique, nous travaillons au minimum avec des triplicats biologiques c'est à dire qu'il faut extraire les ARNs à partir de 3 individus indépendants ou à partir de 3 cultures cellulaires indépendantes. Les ARNs sont contrôlés d'abord sur le Nanodrop.

Les minimums requis sont :

- un **ratio 260nm/230nm et 260nm/280nm  $\geq$  à 1,8**
- une **concentration minimale d'échantillon** nécessaire pour l'expérience choisie (celle-ci est fournie dans les recommandations d'utilisation du fournisseur du kit).

Après mesure au Nanodrop, les ARNs sont contrôlés au Bioanalyzer sur puces Agilent RNA 6000 Lab-on-Chip.

Les minimums requis sont :

- un **ratio 18S/28S  $\geq$  à 1,7 pour les Eucaryotes** ou un **ratio 16S/23S  $\geq$  à 1,2 pour les Procaryotes**
- un **RIN (RNA Integrity Number)  $\geq$  8.5.**

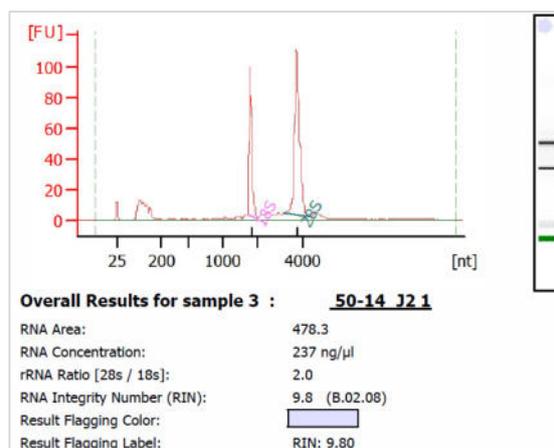


Fig 3: Profil d'ARN correct après lecture au Bioanalyzer

**Un mail vous sera envoyé avec les résultats des QC. Si les ARNs ou ADNs passent nos critères qualité, la prestation est poursuivie. Dans le cas contraire, la décision de continuer ou pas, vous incombe.**

## 4. SEQUENÇAGE

Le séquençage est effectué sur le Ion GeneStudio S5 Prime system (ThermoFisher) selon les spécifications du distributeur. Nous pouvons prendre en charge la ribodéplétion des ARNs ainsi que la synthèse des bibliothèques ARN destinées à du séquençage Illumina.

Nous réalisons également la métagénomique ciblée et du séquençage d'amplicons sur le séquenceur MiSeq Illumina.

Le séquençage long fragment (génome complet, plasmides et fosmidés) est réalisé sur le MinION (Oxford Nanopore) selon les spécifications du distributeur.

#### 4.1. Matériel de départ

Les quantités minimales de matériel au départ pour les applications majeures sont :

- Pour le DNAseq : 10ng à 1µg d'ADN à une concentration entre 1 et 30ng/µl
- Pour le RNAseq : 2-5µg d'ARN total minimum à une concentration minimale 250ng/µL. Ou si vous vous chargez de la sélection des polyA ou de la ribodéplétion: 1 à 500 ng d'ARN-poly(A) ou 10 à 500ng d'ARN déplétés en ARN ribosomaux (concentration minimale de 1-65ng/µL)
- Pour la métagénomique sur S5 : 10µL d'ADN extrait, peu importe la quantité car en métagénomique l'ADN procaryote est souvent plus ou moins contaminé par de l'ADN eucaryote.
- Pour la métagénomique sur le MiSeq : vous réaliserez les PCR1.

Avant de commencer, vous devez nous envoyer les séquences du couple de primers ciblant la région d'intérêt. Nous vous renverrons par retour de mail les séquences auxquelles auront été ajoutés les 1/2 adaptateurs. **Ce sont ces séquences que vous devrez commander** pour réaliser la PCR1 (amplification de la région d'intérêt+1/2 adaptateurs).

Vous pourrez déposer votre PCR1 sur gel afin de vérifier qu'elle a bien fonctionné avant de nous envoyer vos plaques PCR en colis froid. Nous avons besoin de 50µl de produit de PCR1 à une concentration minimale de 20ng/µl (sauf pour les contrôles négatifs) pour pouvoir réaliser la PCR2 (intégration de l'autre 1/2 adaptateur avec index) puis le séquençage.

#### 4.2. Synthèse des librairies

Après contrôle qualité des échantillons, les étapes précédant le séquençage sont :

- La synthèse de librairie avec fixation des adaptateurs de séquençage et du barcode (dans le cas du multiplexage) sur les fragments d'ADN ou d'ARN rétro-transcrits.
- L'amplification de la librairie à séquencer, les PCR2 (ou le pool de librairies à séquencer dans le cas de multiplexage) sur des billes par PCR en émulsion (S5), ou directement sur la Flowcell (Illumina)
- L'addition des barcodes et le pool sur les FlowCells MinION

Chaque étape est validée par un contrôle qualité au « bioanalyzer » sur puces Agilent, par mesure au spectrophotomètre et/ ou au Qubit afin de déterminer la qualité, la taille des fragments et la quantité de matériel produite.

## 4.3. Séquençage

**Sur le séquenceur Ion GeneStudio S5 Prime system**, la puce de séquençage est chargée avec la polymérase et les billes « actives ». Nous disposons de trois formats de puces permettant de générer plus ou moins de séquences en 'lecture simple' (single read) en 200 bp.

	Données	Lectures	Taille des lectures
Ion 510 Chip	0.3-0.5 Gb	2-3 M	200 bp
Ion 520 Chip	0.6-1 Gb	4-6 M	200 bp
Ion 530 Chip	3-4 Gb	15-20 M	200 bp
Ion 540 Chip	10-15 Gb	60-80 M	200 bp
Ion 550 Chip	40-50 Gb	100-130 M	200 bp

Fig 4: Types de puces de séquençage sur le S5

La puce est chargée dans le Ion GeneStudio S5 Prime system pour être séquençée avec le programme approprié à l'application demandée.

A la fin du séquençage, un rapport du run permettant de vérifier le bon déroulement du séquençage est édité au format .pdf.

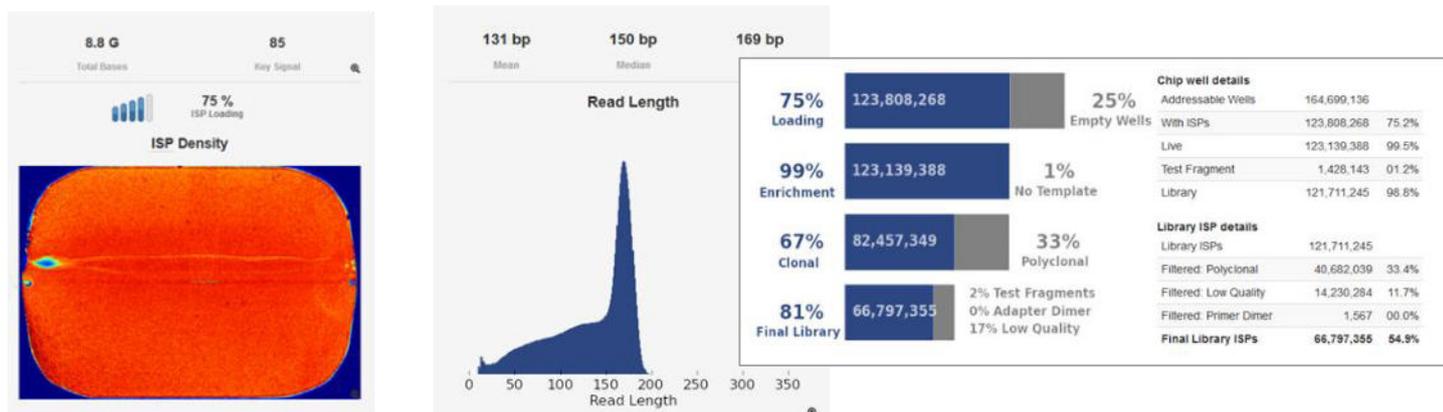


Fig 5: Puce chargée, taille des fragments séquençés, contrôle qualité du séquençage (de gauche à droite)

**Sur le séquenceur Miseq**, nous faisons uniquement du séquençage d'amplicons de toute taille et de tout type. Les puces utilisées sont des V2 ou V3 et produisent plus ou moins de séquences en fonction de leur format.

Suite au séquençage, un contrôle qualité du bon déroulement du run est réalisé en interne sur le logiciel SAV (Sequencing Analysis Viewer).



Fig 6: Analyse sur SAV d'un séquençage amplicon Miseq en 2x250pb

**Sur le séquenceur MinION**, nous séquençons l'ADN de petits génomes (bactéries, levures), ainsi que les plasmides et les fosmidés (séquence complète).



Fig 7: Visualisation sur Minknow des pores actifs pendant le séquençage

**Enfin, le séquençage PacBio, NovaSeq ou NextSeq Illumina** est réalisé par une de nos plateformes partenaires.

## 5. ANALYSES Contrôle qualité après séquençage

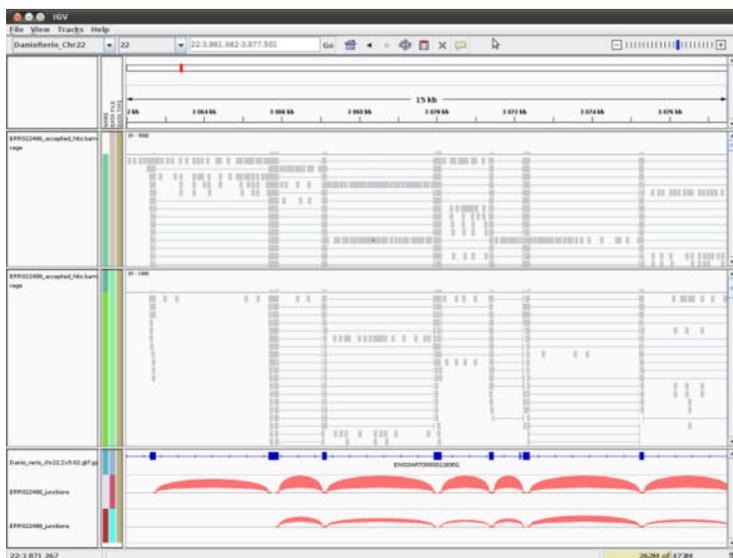
Un contrôle qualité est effectué après le séquençage.

Pour le séquençage S5, NovaSeq, NextSeq, PacBio et MinION, les fichiers FastQC vous sont envoyés et vos données peuvent être récupérées sur un disque dur (non fourni par la plateforme). Les données sont conservées un an sur la plateforme.

Suite au séquençage Miseq, l'ensemble des données est ensuite rendu accessible sur NG6

(<https://ng6.toulouse.inra.fr/>) via un login/mdp.

Analyse bioinformatique des données de séquençage (quelle que soit leur origine)  
 Quelle que soit la provenance de vos données de séquençage (GeT-Biopuces ou organisme extérieur), l'analyse des données est faite à façon.



La prestation peut comprendre l'analyse qualité des données, le mapping et la reconstruction de transcrits sur un génome de référence ainsi qu'une mesure « brute » de l'expression au niveau transcrit/gène/exon selon la demande, la recherche de SNP sur un génome ou des régions ciblées.

Fig 8: Exemple de résultats d'analyse bioinformatique de RNAseq

Les alignements sont remis au format bam (fichiers compressés), le listing des transcrits/gènes au format gtf (fichiers textes), les quantifications au format texte et les analyses de variants au format vcf.

Dans le cas du séquençage long fragment, nous effectuons un assemblage des séquences et dans certains cas, une annotation automatique. Les résultats sont envoyés au format fasta.

## 5.2. Analyses statistiques des données (quelle que soit leur origine)

### 5.2.1 RNAseq

Quelle que soit la provenance de vos données RNAseq (GeT-Biopuces ou organisme extérieur), l'analyse statistique des données est réalisée par à l'aide du logiciel R et des packages Bioconductor. Le résultat de l'analyse est remis sous la forme d'un rapport contenant le protocole expérimental, les contrôles qualités avant et après normalisation, une explication sur le choix de la normalisation effectuée, la liste des gènes différentiellement exprimés ou non ainsi que les graphiques de l'analyse exploratoire (Analyse en Composante Principale et Heatmap).

De plus, les graphiques sont remis au format png ou pdf, les données brutes et les données normalisées, ainsi que la liste des gènes différentiellement exprimés au format Excel et pdf.

### 5.2.2 RNAseq sur cellules uniques (single cell)

L'analyse de données single cell issues de bibliothèques Chromium 10X, sont démultiplexées et alignées à l'aide de Cell Ranger de 10X Genomics qui génère les matrices correspondantes. Les analyses sont ensuite effectuées sous R à l'aide du package et des options les plus appropriées selon la question biologique et le plan expérimental (majoritairement avec le package Seurat : réduction de dimensions, clustering, identification de gènes marqueurs...).

Les données peuvent être issues d'expressions single cell RNA-seq et ATAC-seq, séquencées sous Illumina. Les données sont au format h5 ou mtx pour les matrices Cell Ranger, au format txt pour le reste des résultats. Les graphiques issus des analyses statistiques sont au format png ou pdf.

### 5.2.3 Métagénomique ciblée

Concernant l'étude des communautés microbiennes d'environnements complexes, l'analyse de données de métagénomiques ciblée sur amplicon (16S procaryotes, 18S/ITS eucaryotes) est réalisée à l'aide du package R rANOMALY (<https://f1000research.com/articles/10-7>). Ce workflow nous permet de faire le traitement des données brutes et la génération de la table d'abondance des variants de séquences grâce au package DADA2.

Après assignation taxonomique des variants de séquences, les données sont agglomérées en un objet phyloseq et différentes fonctions de rANOMALY produisent les analyses statistiques et figures spécifiques à ce type de données : diagrammes/histogrammes des communautés microbiennes, analyses alpha et beta diversité, analyses différentielles, taxons partagés entre conditions. L'ensemble des résultats est rendu sous la forme d'un rapport Rmarkdown accompagné de l'archive contenant les données brutes et les analyses.

## 6. RENDU DES RESULTATS

Les résultats d'analyse ou les données brutes sont mis à disposition via FileSender. Toutes les informations nécessaires ainsi que l'adresse de l'accès à vos données vous seront envoyées par mail.

Pour information, les fichiers sont seulement téléchargeables pendant une durée limitée, à compter de leur mise à disposition à l'URL indiquée.

**Attention, il arrive que malgré de bons contrôles qualités, les analyses ne permettent pas d'obtenir de liste de gènes différentiellement exprimés. Ceci ne remet pas en cause la qualité des résultats obtenus mais cela peut être le reflet d'une réalité biologique des échantillons.**

## 7. REGLES DE CONFIDENTIALITE, DE PROPRIETE INTELLECTUELLE ET DE

### VALORISATION

Dans le cadre d'une prestation de service, vous avez la propriété exclusive de vos résultats ainsi que leur maîtrise et leur valorisation. **Cependant, pour la visibilité de la plateforme, nous vous sommes reconnaissants de nommer le personnel qui s'est chargé de votre prestation parmi les auteurs de toute publication. A minima la plateforme GeT- Biopuces de la Génopole Toulouse Midi-Pyrénées doit être citée dans les remerciements des articles, posters et travaux de thèse qui découlent de l'utilisation de ses ressources. Dans un cas comme dans l'autre, merci de bien vouloir nous en informer.**

L'adresse à utiliser pour l'ensemble des publications est :

**Plateforme GeT-Biopuces, TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.**

Une copie des papiers présentant les travaux citant la plateforme sera appréciée.

Afin de protéger la confidentialité de vos projets, dès la première étape de discussion, nous sommes en mesure avec l'aide des juristes du SAIC de l'INSA de signer des accords de confidentialité.

De plus le SAIC négocie et exécute les accords et conventions à caractère industriel et commercial, en particulier, les contrats d'essais, de recherche, d'études, d'analyses, de conseils et d'expertises que nous pourrions effectuer pour le compte de tiers. Les entreprises peuvent bénéficier du « crédit d'impôt recherche »