

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

Document établi entre le collaborateur et la plateforme transcriptomique POPS
Ce document n'est pas contractuel.

I- INFORMATIONS GENERALES ET CONDITIONS D'ACCES

1 - Accès

L'accès au séquençage mRNA-seq pour des projets d'analyse du transcriptome est possible dans le cadre de collaborations entre la Plateforme de transcriptOmique Plante de Paris Saclay (POPS) et d'autres laboratoires, dans la mesure des ressources produites à l'IPS2.

Pour bénéficier de cette prise en charge, les équipes intéressées peuvent contacter la plateforme par mail.

La plateforme apporte son expertise et son savoir-faire sur les analyses du transcriptome et sa main d'œuvre dans cette collaboration.

Une réunion en distanciel ou sur site pour définir le(s) objectif(s) et question(s) biologique(s) du projet afin d'établir entre autre le plan d'expérience sera systématiquement réalisée. Les participants à cette réunion seront le(s) collaborateur(s), les membres concernés de plateforme Transcriptome et les membres concernés de l'équipe de « Réseaux Génomiques ».

2 - Bon de commande et Paiement

Un devis pourra être réalisé et envoyé sur simple demande du collaborateur à pops.ips2@universite-paris-saclay.fr.

Le collaborateur devra éditer un bon de commande correspondant au montant du devis et l'envoyer à recettes@ips2.upsaclay.fr et à pops.ips2@universite-paris-saclay.fr. Le paiement sera réalisé après émission d'une facture par l'IPS2. Pour plus d'information ou si ce schéma ne correspondrait pas à votre échéancier financier, veuillez écrire à recettes@ips2.upsaclay.fr.

3 - Préparation des échantillons

Il est important de noter que de nombreux facteurs influent sur le niveau d'expression des gènes d'une plante. Le contrôle des conditions expérimentales est donc crucial si l'on veut relier une différence d'expression à la fonction étudiée. Ainsi une plante témoin comparée à une plante ayant subi un traitement spécifique devra être cultivée dans le même environnement nutritionnel et lumineux que cette dernière. Par exemple, un décalage de récolte en cours de journée révélera des différences dues à l'expression circadienne de nombreux gènes. Un défaut d'homogénéité d'arrosage ou de traitement phytosanitaire peut être source de variabilité sans rapport avec le processus étudié. Ces considérations sont à prendre en compte pour assurer la reproductibilité des échantillonnages.

3-a - Répétitions

Il est essentiel de faire la différence entre une répétition technique et une répétition biologique.

Répétitions techniques :

Les répétitions d'un échantillon sont préparées au même moment (semis, prélèvements, extractions ...).

Permet l'observation et la quantification des biais techniques (variabilité technique).

Contrôle de la reproductibilité des études.

Contrôle de la qualité des données obtenues.

Les conclusions ne sont valables que pour l'individu.

Répétitions biologiques :

Les répétitions d'un échantillon sont préparées à des moments différés dans le temps (semis, prélèvement, extraction ...) avec au minimum 24h de décalage (attention au cycle circadien).

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

Permet l'observation de la variabilité inter-individus.
Les conclusions sont généralisables aux populations étudiées.

En tout état de cause, il est nécessaire **de prévoir au moins deux répétitions biologiques**, c'est-à-dire **3 fois** l'ensemble de l'expérience. L'objectif étant de caractériser la variabilité biologique entre les répétitions, et de « l'éliminer » afin d'identifier les gènes dont la différence d'expression est liée au seul facteur étudié.

3-b - Quantité et qualité du matériel nécessaire aux expériences

Une quantité de **3µg** d'ARN total avec volume minimum impératif de **15µl** par échantillon est requise. Contactez la plateforme si vous n'étiez pas en mesure d'obtenir cette quantité.
Pour les échantillons « difficiles » comme les graines et les racines, l'ajout de PVP au moment du broyage s'avère très utile. Contactez-nous si besoin.

Les ARN totaux sont à envoyer en carboglace dans la solution d'élution. Leur qualité sera estimée sur puce Bioanalyser Agilent et ils seront dosés au « Ribogreen » après leur arrivée sur la plateforme.

L'envoi des ARN est à accompagner du tableau d'informations dûment rempli (en dernière page).

La pureté des ARN est un des facteurs les plus importants pour la réussite de l'expérience, une **purification sera effectuée, suite au contrôle qualité des ARN à réception**, avec le kit « RNA clean and concentrator, Zymo Research ».

Les échantillons d'ARN peuvent être renvoyés au collaborateur sur simple demande et à ses frais dans un délai de 6 mois à compter de l'envoi des résultats par la plateforme.
Passés ces 6 mois, les ARN ne seront plus conservés par la plateforme, ils seront jetés.

4 - Caractéristiques des runs de séquençage et délais

Les runs de séquençage sont réalisés sur le NextSeq500 (Illumina) de la plateforme ou sur séquenceur NovaSeq (Illumina) via l'Institut de Génomique-CNS à Evry.
Le nombre de reads par échantillon est à ajuster en fonction de votre question biologique de départ.

Un délai pour la construction des bibliothèques et le séquençage vous sera donné à partir de la réception des ARN de qualité et quantité satisfaisante. Suivant les options choisies pour les analyses bioinformatiques et statistiques un délai supplémentaire sera donné. Ce délai tiendra compte notamment :

- Des analyses bioinformatiques suivant l'option choisie
- Des analyses statistiques : normalisation et analyse différentielle des données
- De l'intégration des données dans CATdb et envoi GEO (NCBI).

Options d'analyses

- **B1** : Mapping sur transcriptome de référence, analyse différentielle 2 à 2 et participation aux frais de maintenance des serveurs et de stockage des données (durée du projet + 1 an).
- **B2** : Analyses bioinformatiques nécessitant des outils ou paramètres particuliers adaptés à la question biologique (assemblage de novo RNAseq, traitement des smallRNA, détection de nouveaux gènes ...), analyse différentielle et participation aux frais de maintenance des serveurs et de stockage des données (durée du projet + 1 an).
- **B3** : Stockage des données brutes sans analyses pendant 6 mois (dans la limite de 1 Terra).

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

- **B4** : Mapping sur transcriptome de référence et participation aux frais de maintenance des serveurs et de stockage des données (durée du projet + 1 an).
- **B5** : preprocessing des données et participation aux frais de maintenance des serveurs
- **B6** : Assemblage et annotation de transcriptome de novo

5 - Processus opérationnel

- Contrôle qualité des ARN (Bioanalyser Agilent) et quantification (Ribogreen).
- Construction des bibliothèques (RNA-seq, Small-RNA, directionnel-RNA-seq, UltraLow ...): protocoles Illumina, Clontech, ...
- Contrôle qualité des banques sur puce Bioanalyzer (Agilent)
- Séquençage ou NextSeq500 ou sur NovaSeq
- Assemblage si nécessaire
- Contigs si nécessaire
- Mapping
- Comptage
- Analyse différentielle avec DiCoExpress

Après analyse statistique des résultats bruts, une liste de gènes par comparaison est produite sous forme de fichier Excel. Elle comprend le comptage moyen de la condition 1, le comptage moyen de la condition 2, le ratio et une p_value brute et ajustée pour permettre un contrôle des faux-positifs.

6 - Format d'échange des données

L'ensemble des résultats (comptages, fichier Excel, ACP ...) sera envoyé par mail avec des liens de téléchargement.

Les données brutes (fastq), contigs (si réalisés), seront disponibles en chargement via un cloud ou un site sécurisé. **Le laboratoire partenaire s'engage à rapatrier les données brutes sur leur propre serveur dès la réception dans un délai d'1 mois maximum après la mise à disposition des séquences.**

Le stockage des données et les analyses bioinformatiques et statistiques sont faites à l'IPS2.

L'IPS2 s'engage à garder les données brutes (archivage des fastq, pas les images) 1 an après la mise à disposition des données au collaborateur. Passé ce délai, les données seront détruites.

7 - Bases de données

Il est prévu que les résultats des expériences soient intégrés dans la base de données **CATdb, Gagnot et al. Nucleic Acids Res. 2008 and Zaag et al. NAR 2015** (compatible avec le standard MIAME : Brazma *et al*, 2001. *Nat Genet.* 29(4):365-71) et transmis à la base de données **Geonibus** (GEO) du NCBI. GEO délivrera un numéro d'accès conseillé pour toute publication des résultats transcriptome.

Pour ce faire la plateforme vous fera parvenir un fichier de soumission pour recueillir les informations nécessaires à ces soumissions (conditions de culture, traitement, ...).

Attention, si vous ne comptez pas publier toutes les données en même temps, remplissez 2 fichiers différents pour avoir 2 numéros d'accès (si besoin nous contacter pour plus d'information).

Seuls les projets pour lesquels nous réalisons les analyses (Option B1 ou B2) seront soumis dans les 2 bases de données citées ci-dessus.

8 - Publication et gestion des données

Les données seront rendues publiques **2 ans** après la fin du projet. Il y a néanmoins des exceptions qui seront discutées au cas par cas :



Concevoir en amont pour analyser en aval

ENTENTE PREALABLE AU PROJET DE TRANSCRIPTOMIQUE

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

- 1) si le projet est dans le cadre d'un partenariat avec un industriel
- 2) si le projet est dans le cadre d'un projet ANR/KBBE; les résultats sont mis dans le public 1 an après la fin du projet en lui-même seulement.
- 3) si les résultats transcriptome sont en cours de publication ou de valorisation pour dépôt de Brevet.

Les informations personnelles que vous fournissez dans l'entente préalable ou fichier de soumission pour CATdb, BD interne (fichier excel) sont réservées à l'usage de POPS et ne seront pas communiquées à des tiers. En revanche, dans ces fichiers vous pourrez éventuellement soumettre des données personnelles et professionnelles concernant des acteurs techniques et scientifiques du projet. Ces informations permettent d'identifier et de reconnaître les auteurs des travaux scientifiques. Ces informations personnelles sont jointes aux ensembles de données produits et suivent le cycle de vie des données. Ces informations seront publiées dans CATdb et NCBI/GEO quand les données deviendront automatiquement publiques lors de la soumission de publication.

Conformément au Règlement européen sur la protection des données personnelles (Règlement européen 2016/679), vous et les acteurs disposez d'un droit d'accès, de rectification, d'opposition et de suppression des informations vous concernant. Si vous souhaitez exercer ce droit et obtenir des informations vous concernant, veuillez nous contacter (mail de catdb gnet.db@ips2.upsaclay.fr). Si vous estimez, après nous avoir contacté, que vos droits informatiques et libertés ne sont pas respectés, vous pouvez adresser une réclamation à une autorité de contrôle telle que la Commission Nationale de l'Informatique et des Libertés (CNIL) par courrier " Commission Nationale de l'Informatique et des Libertés - 3 Place de Fontenoy – TSA 80715 – 75334 PARIS CEDEX 07 ou en ligne <http://www.cnil.fr/>

9 - Publication des résultats

Il s'agit de collaborations scientifiques entre l'IPS2 et le partenaire, dans laquelle la plateforme apporte son expertise. Seul le coût des consommables est pris en charge par le laboratoire partenaire. A ce titre, un membre de la plateforme Transcriptome et un membre de l'équipe de « Réseaux Génomiques » de l'IPS2 seront cosignataires de la première publication dans laquelle les données de transcriptome seront présentées/exploitées. La même convention sera appliquée pour le dépôt de Brevet à l'initiative du collaborateur et dans lequel les résultats transcriptome seront exploités.

Il vous sera également demandé de citer dans le texte de description des données, la base de données CATdb (par exemple : "Microarray data from this article were deposited at Gene Expression Omnibus (Edgard 2002): <http://www.ncbi.nlm.nih.gov/geo/>; accession no. GSEXXXXX and at CATdb (Gagnot 2007): <http://urgv.evry.inra.fr/CATdb/>; Project: XXXX according to the "Minimum Information About a Microarray Experiment" standards.").

Affiliation IPS2 :

1. Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91405, Orsay, France.
2. Université Paris Cité, Institute of Plant Sciences Paris-Saclay (IPS2), F-91405 Orsay, France.

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

II- DESCRIPTION DU PROJET

Nous vous demandons de décrire votre projet en complétant ces rubriques avant de nous renvoyer ce document :

1 - Titre du projet

2 - Nom et coordonnées du responsable du projet

3 - Nom et coordonnées de la personne chargée du suivi du projet en liaison avec la plateforme POPS (si différente du responsable du projet) :

4 - Objectifs scientifiques (à renseigner les plus précisément possible et notamment)

Question biologique :

Annotation, quantification RNA/Small-RNA, ... :

5 - Plan d'expérience avec notamment

Nombre d'échantillons :

Nombre de reads par échantillon :

Protocole de construction de banques :

Séquenceur : MiniON, NovaSeq, NextSeq500:

Séquençage Pair End ou Single End :

Longueur de séquençage 75pb, 100bp, 150pb, 300pb :

Option « Analyses Bioinformatique/statistiques » choisie (voir partie 4) :

Comparaisons à réaliser dans le cadre des analyses : décrire les facteurs et leurs modalités (exemple :

facteur génotype : génotype1, génotype2, génotype3, facteur stress : contrôle, stress1, stress2)

Y a-t-il un génome de référence ou un set UniGene (transcriptome) disponible : oui/non :

Si oui, lequel ?

Renvoi des ARN à la fin du projet : oui/non

6 - Nombre de bibliothèques – description des échantillons (organe, stade de prélèvement selon Boyes *et al.* . *Plant Cell* 2001, traitement...)

La nomenclature à suivre pour les noms des échantillons est la suivante : conditionX_Y

- **X** : Il y a au minimum 1 condition (exemple 1 : génotype et exemple 2 : traitement) ou plusieurs conditions (exemple 3 : génotype_traitement1_traitement2 et exemple 4 : génotype_traitement1, dans ce cas ils seront séparés par un _).
- **Y** : numéro de réplica, séparé des conditions par un _.

Exemples:

- | | |
|------------------------|---|
| 1) XP17_1 | condition_réplikat |
| 2) N2_2 | condition_réplikat |
| 3) WT_N10_24h_3 | condition_traitement1_traitement2_réplikat |
| 4) Mut_light_1 | condition_traitement_réplikat |



Concevoir en amont pour analyser en aval

ENTENTE PREALABLE AU PROJET DE TRANSCRIPTOMIQUE

Validation du document d'entente préalable par XX le JJ/MM/AAAA.

7 - Date prévue de fourniture des échantillons à l'IPS2

8 - Tableau à joindre aux échantillons d'ARN lors de l'envoi

Nom sur les tubes	Nom nomenclature PF (conditionX_Y)	Concentration en $\mu\text{g}/\mu\text{l}$	Méthode d'extraction