AGREEMENT FOR TRANSCRITOMIC PROJECT

Validation of the agreement document by XX on DD/MM/YYYY

Document established between the collaborator and the transcriptomic platform POPS
This is not a contract

I- GENERAL INFORMATION AND CONDITIONS OF ACCESS

1 - Access

Access to mRNA-seq sequencing for transcriptome analysis projects is possible through partnerships between the Paris-Saclay Plant Transcriptomic Platform (POPS) and other laboratories.
To benefit from this support, interested teams can contact the platform by email.
The platform brings its expertise and know-how on transcriptome analysis and its workforce to this collaboration

A telephone or on-site meeting to define the biological objective(s) and question(s) of the project in order to establish, among other things, the experimental plan will be systematically carried out. Participants in this meeting will be the collaborator(s), the concerned members of the Transcriptome platform and the concerned members of the "Genomic Networks" team.

2 - Order and Payment

A quote can be sent upon request to pops.ips2@universite-paris-saclay.fr.

The collaborator must edit an order form corresponding to the amount of the experiments performed and send it to recettes@ips2.upsaclay.fr. Payment will be made after an invoice has been issued by the IPS2. For more information or if this plan does not match your financial timeline, please contact recettes@ips2.upsaclay.fr.

3 - Sample preparation

It's important to note that many factors influence the level of gene expression in a plant. The control of experimental conditions is therefore crucial if we want to link a difference in expression to the function studied. Thus a control plant compared to a plant that has undergone a specific treatment should be grown in the same nutritional and luminous environment as the latter. For example, a harvest delay during the day will reveal differences due to the circadian expression of many genes. A lack of homogeneity in watering or phytosanitary treatment can be a source of variability unrelated to the process under study. These considerations must be taken into account to ensure the reproducibility of the samples.

3-a - Repetitions

It's essential to distinguish between technical and biological repetition.

Technical repetitions:
The repetitions of a sample are prepared at the same time (sowing, sampling, extraction, etc.).
Allows the observation and quantification of technical biases (technical variability).
Control of the reproducibility of the studies.
Quality control of the data obtained.
The conclusions are only valid for the individual experiment.

Biological repetitions:
The repetitions of a sample are prepared at times delayed in time (sowing, sampling, extraction...) with at least 24 hours of delay (beware of the circadian cycle).
Allows the observation of inter-individual variability.

Validation of the agreement document by XX on DD/MM/YYYY

The conclusions can be generalized to the populations studied.

In any case, it is necessary to plan at least two biological repetitions, *i.e.* 3 times the whole experiment. The objective is to characterize the biological variability between repetitions, and to "eliminate" it in order to identify genes whose difference in expression is related to the only factor studied.

3-b- Quantity and quality of equipment required for experiments

A quantity of **3µg** total RNA with a minimum imperative volume of 15µl per sample is required. Contact the platform if you are unable to obtain this quantity. As RNA purity is one of the most important factors for the success of the experiment, **purification** with the "**Zymo Research** clean and concentrator RNA kit" is required prior to shipment of samples.

For "difficult" samples such as seeds and roots, the addition of PVP is very useful. Please contact us if necessary.

The total RNAs are to be sent on dry ice in the elution solution. Their quality will be estimated on an Agilent Bioanalyzer chip and they will be quantified with Ribogreen after their arrival on the platform.

The shipment of the RNAs will be accompanied by the duly completed information table (on the last page).

RNA samples can be returned to the collaborator upon request and at his/her expense within 6 months after the results are sent by the platform.
After these 6 months, the RNAs will not be kept by the platform, they will be thrown away.

4 - Characteristics of sequencing runs and delays

Sequencing runs are performed on the NextSeq500 (Illumina) of the platform, on the NovaSeq(Illumina) sequencer via the CNS Genomics Institute in Evry.
The number of reads per sample should be adjusted according to your initial biological question.

A deadline for the construction of the libraries and sequencing will be given from the receipt of the RNAs of satisfactory quality and quantity. Depending on the options chosen for bioinformatics and statistical analyses, an additional time will be given. This period will take into account in particular:
- Bioinformatics analyses according to the chosen option
- Statistical analysis: standardization and differential data analysis
- Data integration in CATdb and GEO sending (NCBI).

Analysis options

- **B0** : Storage of raw data without analysis for 6 months (up to 1 Terabyte).
- **B1** : Mapping on reference transcriptome, differential analysis and contribution to the costs of server maintenance and data storage (project duration + 1 year).
- **B2** : Bioinformatics analysis of small RNA, differential analysis and contribution to the costs of server maintenance and data storage (project duration + 1 year).
- **B3** : *De novo* transcriptome assembly and annotation (including nanopore) and single-cell sample.

Validation of the agreement document by XX on DD/MM/YYYY

5 - Operational process

- RNA quality control (Bioanalyzer Agilent) and quantification (Ribogreen).
- Libraries construction (RNA-seq, Small-RNA, directional-RNA-seq, UltraLow...) : Illumina, Clontech … protocols
- Quality control of banks on chip Bioanalyzer (Agilent)
- Sequencing on NextSeq500 or NovaSeq
- Assembly if necessary
- Contigs if necessary
- Mapping
- Counting
- Differential analysis with DiCoExpress

After statistical analysis of the raw results, a list of genes by comparison is produced as an Excel file. It includes the average count of condition 1, the average count of condition 2, the logratio and a raw and adjusted p_value to allow false positives to be controlled.

6 - Data exchange format

All the results (counts, Excel file, ACP...) will be sent via Renater.

The raw data (fastq), contigs (if realized), will be available for loading via a cloud or a secure site. **The partner laboratory undertakes to repatriate the raw data on their own server as soon as it receives it within a maximum of 1 month after the sequences are made available.**

Data storage, bioinformatics and statistical analyses are carried out at IPS2.

The IPS2 will store the raw data (fastq files, not images) for 1 year after the data have been made available to the collaborator. After this period, the data will be destroyed.

7 - Databases

It is expected that the results of the experiments will be integrated into the **CATdb database, Gagnot *et al.* Nucleic Acids Res. 2008 and Zaag *et al.* NAR 2015** (compatible with the MIAME standard: Brazma *et al.,* 2001. Nat Genet. 29(4):365-71) and transmitted to the NCBI **Geomnibus** (GEO) database. GEO will issue an accession number recommended for any publication of transcriptome results.

To do this, the platform will send you a submission file to collect the information necessary for these submissions (cultivation conditions, treatment, etc.).

Attention, if you do not intend to publish all the data at the same time, fill in 2 different files to have 2 accession numbers (if necessary contact us for more information).

Only the projects for which we carry out the analyses (Option B1 or B2) will be submitted in the 2 databases mentioned above.

8 - Data release

The data will be made public **2 years** after the end of the project. There are, however, exceptions that will be discussed on a case-by-case basis:

1) if the project is in partnership with a private company

2) if the project is part of an ANR/KBBE project; the results are made available to the public only 1 year after the end of the project itself.

3) if the transcriptome results are being published or valued for patent filing.

AGREEMENT FOR TRANSCRITOMIC PROJECT

Validation of the agreement document by XX on DD/MM/YYYY

The personal information that you provide in this agreement or submission file for CATdb, internal BD (excel file) is reserved for the use of POPS and will not be communicated to third parties. On the other hand, in these files you can possibly submit personal and professional data concerning technical and scientific actors of the project. This information makes it possible to identify and recognize the authors of scientific work. This personal information is attached to the product datasets and follows the data lifecycle. This information will be published in CATdb and NCBI / GEO when the data automatically becomes public upon publication submission.

In accordance with the European Regulation on the protection of personal data (European Regulation 2016/679), you and the actors have the right to access, rectify, oppose and delete information concerning you. If you wish to exercise this right and obtain information about yourself, please contact us (catdb email gnet.db@ips2.upsaclay.fr). If you believe, after contacting us, that your IT rights and freedoms are not respected, you can send a complaint to a supervisory authority such as the National Commission for Computing and Liberties (CNIL) by mail " Commission Nationale de l'Informatique et des Libertés - 3 Place de Fontenoy - TSA 80715 - 75334 PARIS CEDEX 07 or online http://www.cnil.fr/

9 - Publication of results

These are scientific collaborations between IPS2 and the partner, in which the platform provides its expertise. Only the cost of consumables is covered by the partner laboratory. As such, a member of the Transcriptome platform and a member of the IPS2 "Genomic Networks" team will be co-authors of the first publication in which transcriptome data will be presented/used. The same agreement will be applied for the filing of Patents at the initiative of the collaborator and in which the transcriptome results will be used. However, the Intellectual property deriving from these patents remains entirely with the collaborator.

You will also be asked to cite in the data description text, the CATdb database (for example: "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.

IPS2 affiliation:

1. Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris-Saclay, CNRS, INRAE, Université Evry, Gif sur Yvette, 91190, France.
2. Institute of Plant Sciences Paris-Saclay (IPS2), Université Paris Cité, CNRS, INRAE, Gif sur Yvette, 91190, France.

II- PROJECT DESIGN (required)

We ask you to provide the following information:

1 - Title of project

2 -Name and address of project manager

3 - Name and address of the person responsible for monitoring the project in relationship with POPS platform (if different from the project manager)

AGREEMENT FOR TRANSCRITOMIC PROJECT

Validation of the agreement document by XX on DD/MM/YYYY

4 - Scientific aims: be as accurate as possible including
Biological question?
Annotation, RNA quantification / Small-RNA, construction of High Density chip? ….

5 - Experiment design including:

Number of samples:
Number of reads per sample:
Protocol of sequencing library construction:

- TruSeq mRNA Ligation Prep
- NEXTflex Small RNA
- UltraLow SMARTER +TruSeq
- TruSeq Stranded RNA with Ribo-Zero Plant
- QuantSeq 3'
- BRB-seq
- ribo bacterien
- Nanopore polyA V14
- Nanopore polyA V14 with capture
- Nanopore rRNA v14
- Nanopore non polyA V14
- Nanopore non polyA V14 with capture
- Nanopore non polyA V14 sitools
- Single-cell 10X Genomics

Sequencing machine: ☐ P2 (ONT), ☐ NovaSeq (Illumina), ☐ NextSeq500 (Illumina)

Sequencing: ☐ Pair End, ☐ Single End

Sequencing length: ☐ 75pb, ☐ 100bp, ☐ 150pb, ☐ 300pb

"Bioinformatics Analysis / statistics" option selected (see I-part 4):
☐ B0, ☐ B1, ☐ B2, ☐ B3, ☐ B4

Comparisons to be made as part of the analyses:
Describe the factors and modalities (for example: factor genotype: genotype1, genotype2, genotype3; factor stress: control, stress1, stress2)

Is there a reference genome or UniGene set (transcriptome) available? yes/no
If so, which one?

Do you wish to get RNA samples back at the end of project? ☐ yes, ☐ no
If not, do you authorize POPS to use them for testing, or as a positive control? ☐ yes, ☐ no

AGREEMENT FOR TRANSCRITOMIC PROJECT

Validation of the agreement document by XX on DD/MM/YYYY

6 - Number of libraries - description of samples per run (organ, stage of sampling according to Boyes *et al.* Plant Cell 2001, treatment ...)

The nomenclature to be followed for the names of the samples is as follows*: condition**X**_**Y***

- ***X**: There is at least one condition (example 1: genotype and example 2: treatment) or more conditions (example 3: genotype_treatment_treatment2 and example 4: genotype_treatment 1), in this case they will be separated by a_.*
- ***Y**: replicate number, separated by a_.*

Examples:

1) XP17_1 condition_replicate
2) N2_2 condition_replicate
3) WT_N10_24h_3 condition_treatment1_treatment2_replicate
4) Mut_light_1 condition_treatment_replicate

7 - Expected date of delivery of samples to IPS2

8 - Table to join to the RNA samples when sending:

Tube name	Sample name with POPS nomenclature (conditionX _Y)	Concentration µg/µl	RNA extraction method