

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

Document established between the collaborator and the transcriptomic platform POPS

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 made and sent upon request to pops.ips2@u-psud.fr.

The collaborator must provide an order form corresponding to the amount of the experiments performed and send it to both addresses recettes@ips2.upsaclay.fr and pops.ips2@universite-paris-saclay.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 Replicates

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

Technical replicates:

The replicates of a sample are prepared at the same time (sowing, sampling, extraction, etc.).

They allow the observation and quantification of technical biases (technical variability) and control the reproducibility of studies.

Quality control of the data obtained.

The conclusions are only valid for the individual.

Biological replicates:

The replicates of a sample are prepared (sowing, sampling, extraction...) with at least 24 hours of delay (beware of the circadian cycle).

They allow the observation of inter-individual variability and the conclusions can be generalized to the populations studied.

AGREEMENT FOR TRANSCRIPTOMIC PROJECT

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

In any case, it is necessary to plan at least two biological replicates, i.e. 3 times the whole experiment. The objective is to characterize the biological variability between replicates, and to “remove” it in order to identify genes whose difference in expression is related only to the 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. For "difficult" samples such as seeds and roots, the addition of PVP is very useful. Please contact us if necessary.

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 reception at the platform.

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

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 will be performed** after quality control at reception.

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 NovaSeq (Illumina) sequencers 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

- **B1:** Mapping on reference transcriptome, differential analysis 2 to 2 and contribution to the costs of server maintenance and data storage (project duration + 1 year).
- **B2:** Bioinformatics analyses requiring specific tools or parameters adapted to the biological issue (assembly *de novo* RNAseq, smallRNA processing, detection of new genes, etc.), differential analysis and contribution to the costs of server maintenance and data storage (project duration + 1 year).
- **B3:** Storage of raw data without analysis for 6 months (up to 1 Terra).
- **B4:** Mapping on reference transcriptome and contribution to the costs of server maintenance and data storage (project duration + 1 year).
- **B5:** Data pre-processing and contribution to the costs of server maintenance.
- **B6:** *de novo* assembly and annotation of transcriptome.

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...): protocols from Illumina, Clontech, ...
- Quality control of libraries 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 in condition 1, the average count in condition 2, the log₂-ratio 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 is responsible for downloading the raw data on their own server as soon as possible and within a maximum of 1 month after the sequences are made available.**

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

IPS2 commits to keep the raw data (fastq archiving, not images) 1 year after the data has been made available. 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.* Nucleic Acids Res. 2015** (compatible with the MIAME standard: Brazma *et al.*, 2001. Nat Genet. 29(4):365-71) and transmitted to the NCBI Gene Expression Omnibus (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 management and 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 an industrial 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 TRANSCRIPTOMIC PROJECT

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

The personal information you provide in the prior agreement or submission file for CATdb, the internal DB (excel file) is reserved for the use of POPS and will not be communicated to third parties. However, in these files you may be able to submit personal and professional data concerning technical and scientific actors of the project. This information makes it possible to identify and recognise the authors of scientific work. This personal information is attached to the product data sets and follows the data life cycle. This information will be published in CATdb and NCBI/GEO when the data is automatically available at the time of publication.

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, object and delete information about you. If you wish to exercise this right and obtain information about yourself, please contact us (email from CATdb gnet.db@ips2.upsaclay.fr). If you feel, after contacting us, that your legal rights are not respected, you can send a complaint to a supervisory authority such as the Commission Nationale de l'Informatique et des Libertés (CNIL) by post mail to "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.

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 *et al.* 2002): <http://www.ncbi.nlm.nih.gov/geo/>; accession no. GSEXXXXX and at CATdb (Gagnot *et al.* 2008): <http://tools.ips2.u-psud.fr/CATdb/>; Project: XXXX according to the "Minimum Information About a Microarray Experiment" standards.

IPS2 affiliation:

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.

II- PROJECT DESCRIPTION (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).
4. **Scientific aims (be as accurate as possible including :)** Biological question? Annotation, RNA quantification / Small-RNA, construction of High Density chip? ...

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

5. Experiment design including:

- Number of samples:
- Number of reads per sample:
- Libraries construction protocol:
- Sequencing apparatus: MiniON, NovaSeq, NextSeq500:
- Sequencing Pair End or Single End:
- Sequencing length 75pb, 100bp, 150pb, 300pb:
- "Bioinformatics Analysis / statistics" option selected (see part 4):

Comparisons to be made as part of the analyses:

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

Sending back RNA samples at the end of project: yes/no

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: conditionX_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 | <i>condition_replicate</i> |
| 2) N2_2 | <i>condition_replicate</i> |
| 3) WT_N10_24h_3 | <i>condition_treatment1_treatment2_replicate</i> |
| 4) Mut_light_1 | <i>condition_treatment_replicate</i> |

7. Expected date of samples delivery 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 (DNase ?)