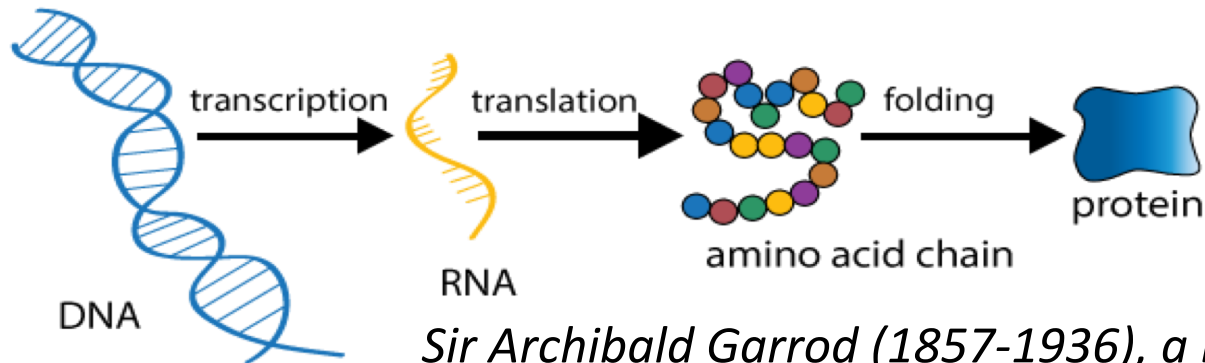


# **Workshop inter-CATI autour des omics**

## **Transcriptomic Resources** **Véronique Brunaud**

# Transcription – Translation

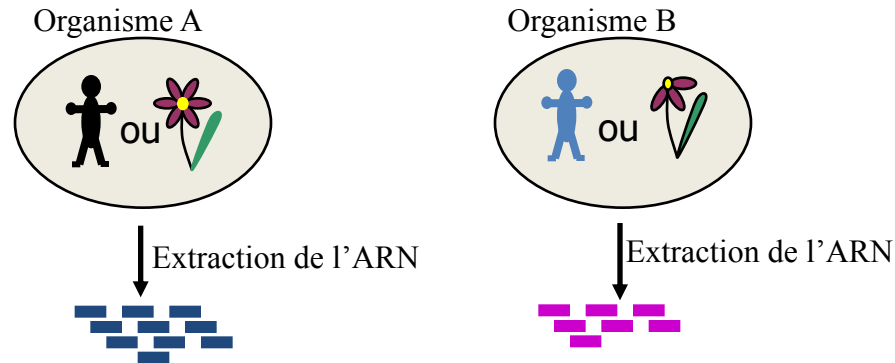


*Sir Archibald Garrod (1857-1936), a British medical doctor, was the first to suggest that genes were connected to enzymes : “one gene encode a single enzyme”*

**Transcriptome** = mesurer les ARNs produits par la plante, dans la plante entière, un organe, ou une cellule à un temp t de son développement et dans des conditions définies.

→ **Comparaisons des transcriptomes**

# Transcriptomic comparisons

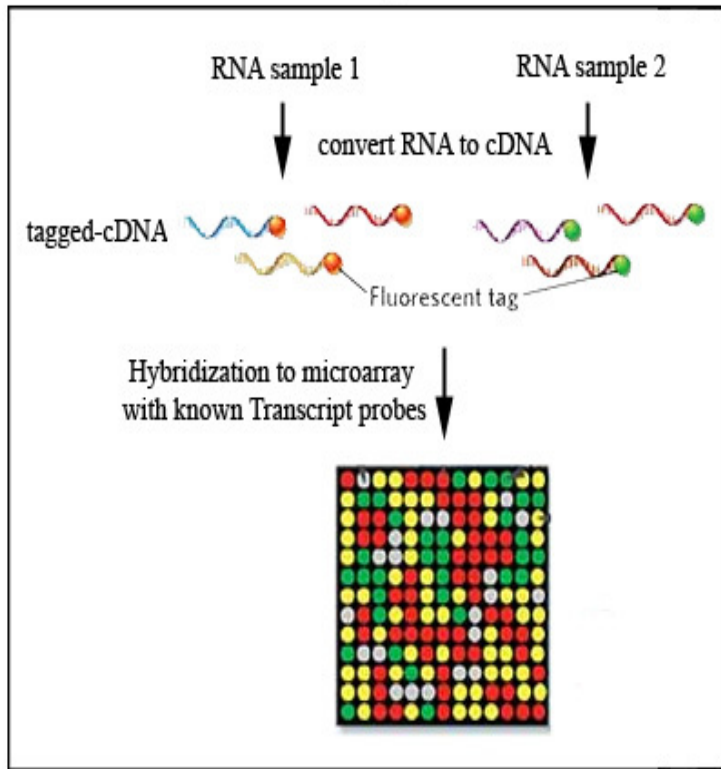


- Comparaison entre des échantillons d'organes différents d'une plante : par exemple fleur et racine.
- Comparaison entre des échantillons de génotypes différents : mutant / sauvage ou l'étude de plusieurs génotypes d'une même espèce.
- Comparaison d'échantillons de plantes cultivées dans 2 conditions : des plantes traitées versus plantes non traitées, effet d'une sécheresse, effet d'un pathogène...
- Comparaison d'échantillons prélevés à des stades de développements différents ou cinétique

....

# Micro-Puces & RNA-Seq

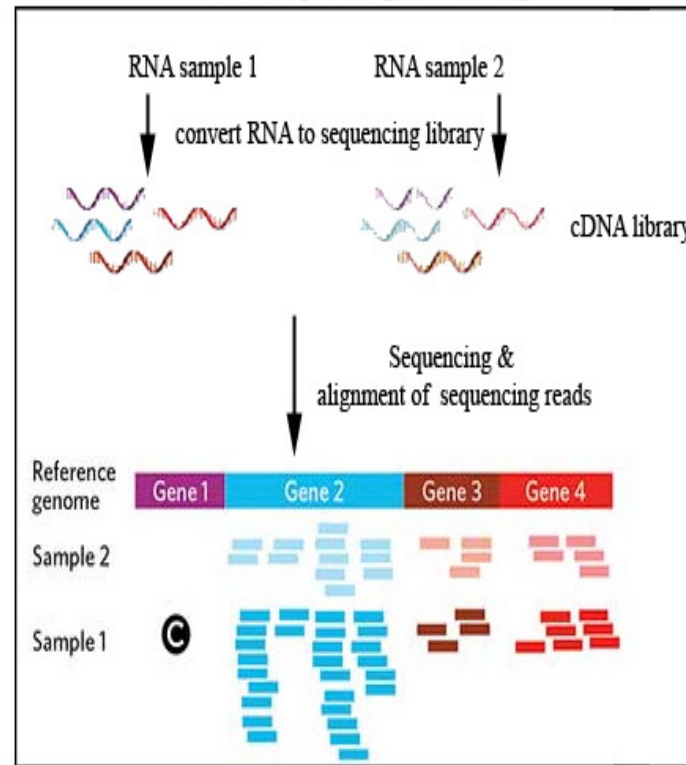
## Microarray



Hybridation

Puces d'ADN: 1990

## RNA Sequencing (RNA-Seq)



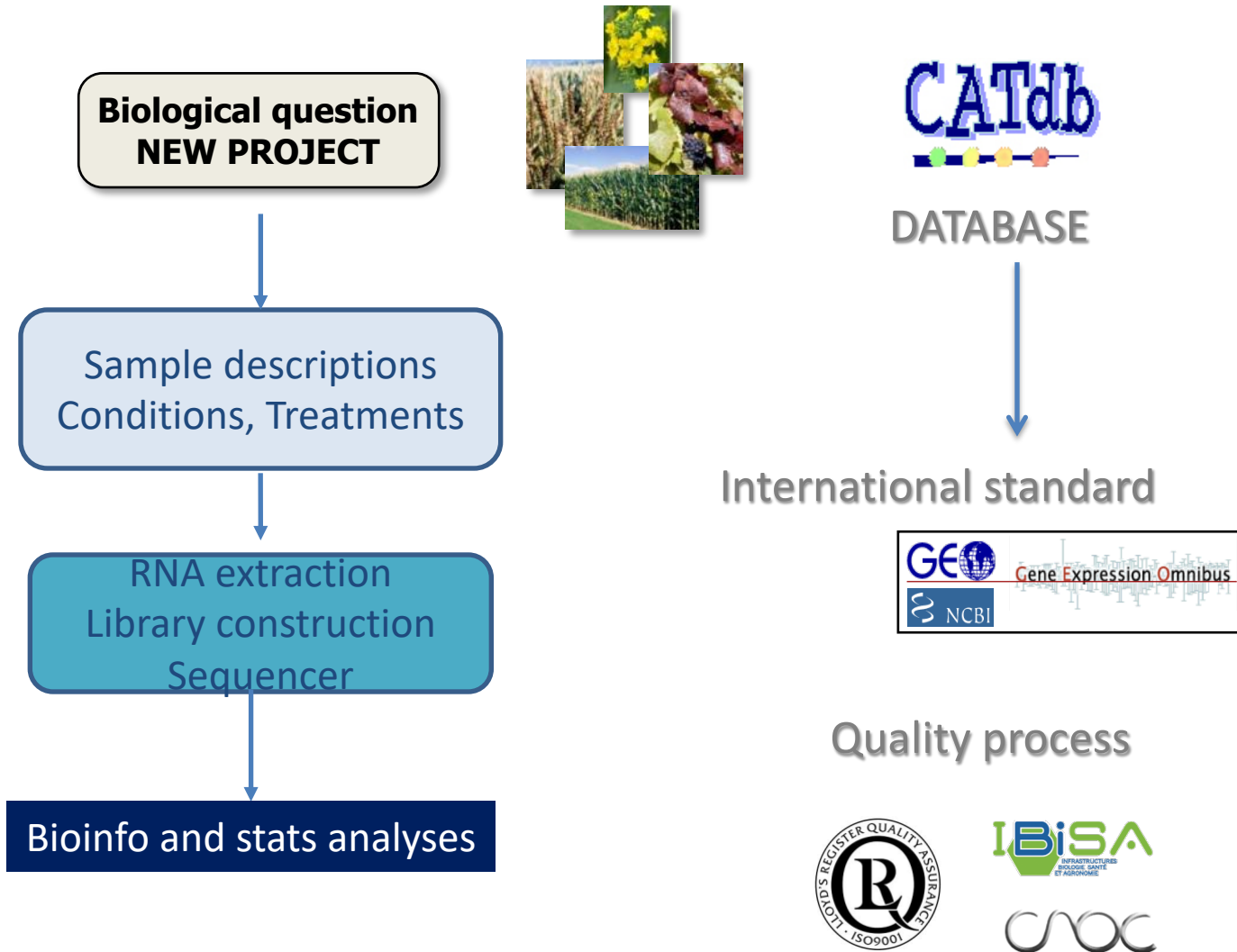
Séquençage

RNA-Seq: 2006

Différences majeures entre RNA-Seq et puces à ADN

- Pas d'obligation d'utiliser une annotation
- Pas besoin d'avoir toutes les comparaisons dès le début (design complet en théorie !)

# Project management for RNA-Seq



# CATdb Information system

PowerAMC - [MPD\_MPD\_CATdb\_2, DiagrammePhysique\_1 - C:\Documents and Settings\tamby\Mes documents\URGV\MPD\_CATdb2.mpd]

ESpace de travail  
MPD\_CATdb\_2  
DiagrammePhysique\_1  
Tables  
Références  
MPD\_CATdb\_1

File Edit View History Bookmarks Tools Help

http://ohno.evry.inra.fr/cgi-bin/projects/transcriptome/Edit.pl

DataBase Google Dictionnaire anglais adas - ADAS BiblioVIE - Portail d' PubMed Central Hom... TAIR (The Arabidopsi...

Uniré de Recherche en Génomique végétale  
URGV

**CATDB DATABASE**  
~Edit~  
-User:brnaud-

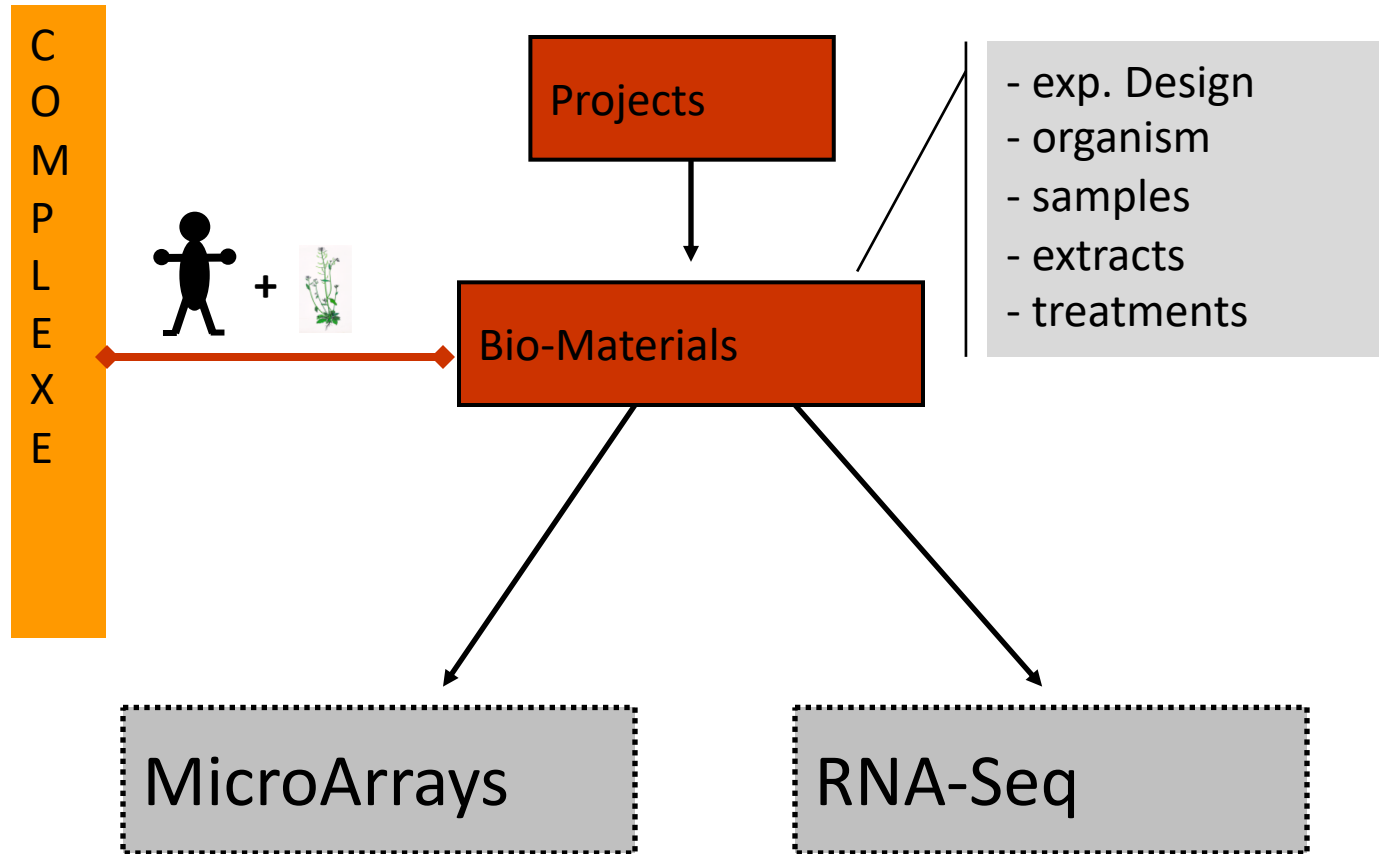
Project/Experiment: Choose the project\*: CAT-seq NEW  
Choose the experiment\*: BF\_vs\_F NEW

Part of this experiment: Choose a step of the experiment :  
Sample Source  
Treatment  
Extract  
Extract Pool  
Labelled Protocol  
Array Hybridization  
Scan  
Normalization  
Swap  
ReplicateAffy  
Export to GEO  
Download result files

Define a new design : NEW

Done

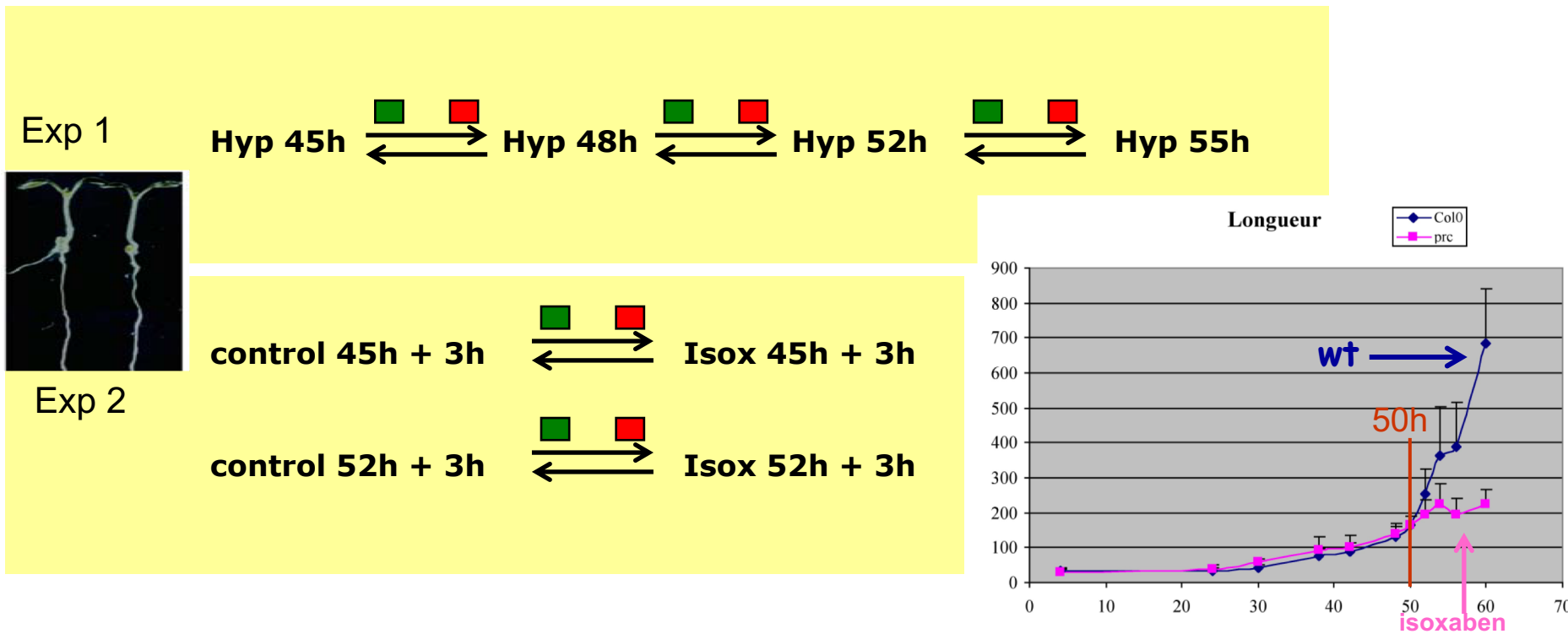
# Experiment and Sample description



# Plan d'expérience : exemple

## RA3-4 (INRA2003-04) Elongation cellulaire, paroi primaire

**Echantillons** : *A. thaliana*, hypocotyles poussés à l'obscurité. Ceux traités à l'isoxabène sont récoltés 3h après le traitement



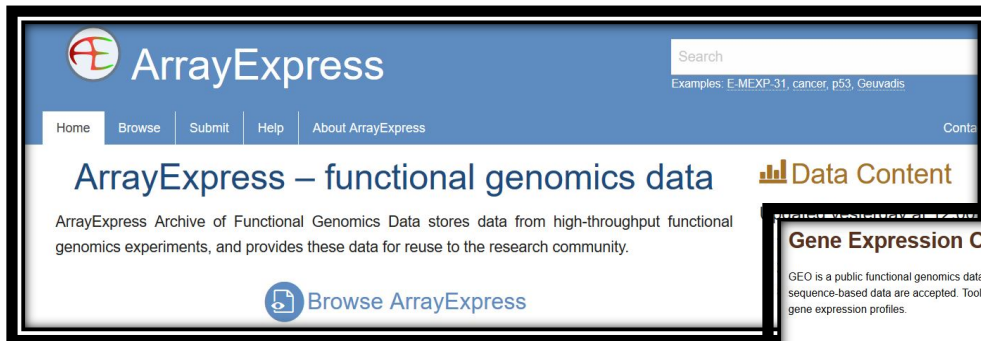


# CATdb Information system

<b>Experiment types (*)</b>	dormant vs non dormant seed embryo dose response ecotype comparison gain of function epimutation gene knock in (transgenic)	<b>Other</b> <input type="text"/>			
<b>Experiment factors (*)</b>	sources of nitrogen species storage protocol strain stress	<b>Other</b> <input type="text"/>			
<b>Description (*)</b>	Overexpressing ZmMYB31 transgenic lines vs transgene-negative sibling plants				
<b>Analysis type (*)</b>	<input type="radio"/> Arrays <input type="radio"/> ChIP-chip <input checked="" type="radio"/> RNA-Seq <input type="radio"/> RIP-Seq				
<b>Array/Seq Platform (*)</b>	<input type="radio"/> CATMA <input type="radio"/> Affymetrix <input type="radio"/> NimbleGen <input type="radio"/> Agilent <input checked="" type="radio"/> Illumina <input type="radio"/> Ion Torrent	<b>Other</b> <input type="text"/>			
<b>Organism list (*)</b> <small>(Gender species)</small>	Triticum urartu x Aegilops tauschii Vitis vinifera Zea mays	<input type="text"/> <i>N.B.: new organism name is automatically checked at <a href="#">NCBI Taxonomy Browser</a> to comply to GEO</i>			
<b>Ecotype list</b>	--- 163av 41528 population	<input type="text"/> <small>(ecotype or sub-species)</small>			
<b>Genotype (*)</b>	A188	<small>(mutant = 'yes')</small>			
<b>Mutant characteristics:</b>	<small>(for NON mutant organism, let the 3 following fields in BLANK)</small>				
type	overexpression	agent	-----	factor	
<b>Mutant gene ID(s)</b>	MYB31-GRMZM2G050305	<small>(e.g.: AT1G12980, separated by comma for multiple IDs)</small>			
<b>Growth conditions (*)</b> <b>(Media/Hygrometry/Temperature/Light)*</b>	PhenoArch Platform				

# International repository and standards

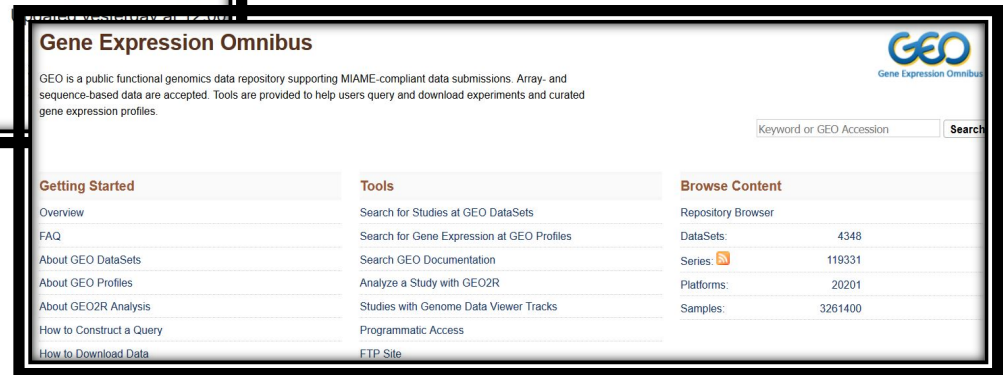
Resources	Experiments/Samples	Raw data (fastq)
NCBI	GEO (~120000, > 3 million)	SRA
EBI	ArrayExpress (~73000, > 2 million)	ENA
DDBJ		DRA



The screenshot shows the ArrayExpress website homepage. At the top left is the ArrayExpress logo. To its right is a search bar with the text "Search" and "Examples: E-MEXP.31, cancer, p53, Geuvadis". Below the search bar is a navigation menu with links for "Home", "Browse", "Submit", "Help", "About ArrayExpress", and "Contact". The main heading is "ArrayExpress – functional genomics data" followed by a "Data Content" icon. Below this is a paragraph: "ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community." At the bottom left is a "Browse ArrayExpress" button with a magnifying glass icon.



Collaboration between the 3 databases



The screenshot shows the Gene Expression Omnibus (GEO) website homepage. At the top right is the GEO logo with the text "Gene Expression Omnibus". Below the logo is a search bar with the text "Keyword or GEO Accession" and a "Search" button. The main heading is "Gene Expression Omnibus". Below this is a paragraph: "GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles." Below the paragraph are three columns of links: "Getting Started" (Overview, FAQ, About GEO DataSets, About GEO Profiles, About GEO2R Analysis, How to Construct a Query, How to Download Data), "Tools" (Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, Analyze a Study with GEO2R, Studies with Genome Data Viewer Tracks, Programmatic Access, FTP Site), and "Browse Content" (Repository Browser, DataSets: 4348, Series: 119331, Platforms: 20201, Samples: 3261400).



**The International Nucleotide Sequence Database Collaboration (INSDC)**

# International repository and standards

## MIAME and MINSEQE guidelines

The **MIAME** (Minimum Information About a Microarray Experiment) and **MINSEQE** (Minimum Information About a Next-generation Sequencing Experiment) guidelines outline the minimum information that should be included when describing a microarray or sequencing study. Many journals and funding agencies require microarray data to comply with MIAME and MINSEQE standards.

**MIAME compliance is not related to the submission format or route, but rather to the content provided**

GEO deposit procedures enable and encourage submitters to supply MIAME and MINSEQE compliant data. All GEO submission procedures are designed to closely follow the MIAME and MINSEQE checklists; if you provide all requested information, your submission will be compliant.

The six most critical elements contributing towards MIAME are:

- Raw data for each assay (e.g., CEL or FASTQ files)
- Final processed (normalized) data for the set of assays in the study (e.g., the gene expression data count matrix used to draw the conclusions in the study)
- Essential sample annotation (e.g., tissue, sex and age) and the experimental factors and their values (e.g., compound and dose in a dose response study)
- Experimental design including sample data relationships (e.g., which raw data file relates to which sample, which assays are technical, which are biological replicates)
- Sufficient annotation of the array or sequence features examined (e.g., gene identifiers, genomic coordinates)
- Essential laboratory and data processing protocols (e.g., what normalization method has been used to obtain the final processed data)

# International repository and standards



Home	About Us	Mission	Projects	Conferences	
------	----------	---------	----------	-------------	--

## **MINSEQE**

### **Minimum Information about a high-throughput SEQuencing Experiment**

**MINSEQE** describes the **Minimum Information about a high-throughput nucleotide SEQuencing Experiment** that is needed to enable the unambiguous interpretation and facilitate reproduction of the results of the experiment. By analogy to the [MIAME](#) guidelines for microarray experiments, adherence to the MINSEQE guidelines will improve integration of multiple experiments across different modalities, thereby maximising the value of high-throughput research. [[MINSEQE version 1.0](#) (pdf), FGED Society (June 2012)]



!Sample_source_name_ch[n]	1 per channel	any	Briefly identify the biological material and the experimental variable(s), e.g., vastus lateralis muscle, exercised, 60 min.
!Sample_organism_ch[n]	1 or more	use standard <a href="#">NCBI Taxonomy</a> nomenclature	Identify the organism(s) from which the biological material was derived.
!Sample_characteristics_ch[n]	1 or more	'Tag: Value' format	Describe all available characteristics of the biological source, including factors not necessarily under investigation. Provide in 'Tag: Value' format, where 'Tag' is a type of characteristic (e.g. "gender", "strain", "tissue", "developmental stage", "tumor stage", etc), and 'Value' is the value for each tag (e.g. "female", "129SV", "brain", "embryo", etc). Include as many characteristics fields as necessary to thoroughly describe your Samples.
!Sample_biomaterial_provider_ch[n]	0 or more	any	Specify the name of the company, laboratory or person that provided the biological material.
!Sample_treatment_protocol_ch[n]	0 or more	any	Describe any treatments applied to the biological material prior to extract preparation. You can include as much text as you need to thoroughly describe the protocol; it is strongly recommended that complete protocol descriptions are provided within your submission.
!Sample_growth_protocol_ch[n]	0 or more	any	Describe the conditions that were used to grow or maintain organisms or cells prior to extract preparation. You can include as much text as you need to thoroughly describe the protocol; it is strongly recommended that complete protocol descriptions are provided within your submission.
!Sample_molecule_ch[n]	1 per channel	total RNA, polyA RNA, cytoplasmic RNA, nuclear RNA, genomic DNA, protein, or other	Specify the type of molecule that was extracted from the biological material.

# Limites par rapport aux normes ou manques de normes

## Usage coté données Transcriptomes

**Types de données : description des exp, description des échantillons, données générées (RAW data and analyzed)**

BD internationales (GEO/ArrayExpress) et normes (Minimum d'informations MIAME, MINSEQE)

Données: raw data dans BD internationales, RNA-Seq analyses: tables de comptages par gènes ou liste de gènes exprimés différemment.


**Normes permettent :**

- Description des échantillons avec des consignes minimum
- Liste des champs obligatoires
- Plus riche dans catdb mais normal puisque sert de LIMS à la PF  
ex: protocoles pour toutes les étapes, cahier de laboratoire

**Pb rencontrés :**

- Pas de mots clés imposés pour décrire le type de comparaisons ou d'expériences.
- Des ontologies pourraient être imposées ex: Plant Ontology  
ex: dans CATdb 33 tissus ou organes contre 2021 termes de la plante ontologie
- Versioning et normalisation des noms de gènes, très dépendant de l'espèce (vérification Taxonomique mais pas des listes de gènes, juste donner la référence ou source)

# Plant Ontology



## Ontology Lookup Service

Home | Ontologies | Documentation | About

GO

GO

Reset tree

Show all siblings

- continuant
- occurrent
- plant anatomical entity
  - plant anatomical space
  - plant structure**
    - cardinal part of multi-tissue plant structure
    - collective plant structure
    - embryo plant structure
    - in vitro plant structure
    - multi-tissue plant structure
    - plant cell
    - plant ovary
    - portion of plant tissue
    - rhizoid
    - trichome
    - trichome apex
    - trichome tip
    - whole plant
  - portion of plant substance
- plant structure development stage
- root

**Cell Culture-Based Biosynthesis** NCIT NCIT:C112922

**Cell Culture System** NCIT NCIT:C19147

**Cell Culture Procedure** NCIT NCIT:C116004

**Cell Culture** NCIT NCIT:C16396

**Cell Culture Pooling** NCIT NCIT:C112952

Search OLS for **cell cultur**

**aleurone grain** GO GO:0033095

**aleurone layer** BTO BTO:0000057

**aleurone layer morphology trait** TO TO:0000942

**aleurone layer appearance** TO TO:0000944

**aleurone layer color** TO TO:0000943

Search OLS for **aleurone**