**HUMAN TRANSCRIPTOME MAP (HTM)**

TRAM (Transcriptome Mapper) is a software that creates and analyzes quantitative transcriptome maps based on gene expression data from different experiments.

It allows you to import gene expression data recorded in the NCBI **GEO** (Gene Expression Omnibus) and EBI **Array Express** databases in tab-delimited text format and to integrate them with each other by decoding probe set identifiers to gene symbols, normalizing data from multiple platforms using intra-sample and inter-sample normalization (scaled quantile normalization), creating graphical representation of gene expression profile through two ways, "Map mode" and "Cluster mode", and determining the statistical significance of results.

**1. SEARCH IN GEO (NCBI Gene Expression Omnibus)**

Web site <http://apollo11.isto.unibo.it/>

Click on GEO (<https://www.ncbi.nlm.nih.gov/geo/>).

Enter your search terms in "**QUERY**" - "**Datasets**", *eg. if you are interested in experiments on human brain insert “Homo sapiens [ORGANISM] AND brain”*, and click on “**GO”** button.

In "**Filter your results**" (top right corner) choose the item “**Series**”.

*In GEO, the acronym GSE (GEO Series) refers to the experiment or the series, the acronym GSM (GEO Sample) refers to the sample, the acronym GPL (GEO platform) refers to the platform (or slide).*

*Each platform consists of a specific probe set complementary to a gene set, and each probe is assigned a gene identifier (ID).*

Browse the results and read each summary to find the experiment of your interest, *eg. look for those works that analyze gene expression profile in human brain of normal individuals.*

The Organism (Homo sapiens) studied and the type of experiment (Expression profiling by array) performed is shown below the summary.

Once you have identified the first experiment of interest, click on its title to open a page with all the information about the experiment, *eg.* *GSE13214*.

Criteria to be followed before choosing an experiment:

* What is the **sample source**? An organ *in toto* or a section, normal, pathological or treated tissue, cell lines? Furthermore, is it adult, child or fetal tissue?

In GEO the sample source is always indicated, but in some cases it is not the correct one, for example because only a subpart was used.

To be sure of the source sample it is necessary to read the the “Materials and Methods” section of the paper.

If the paper is not published in GEO and is not available in PubMed you need to contact the Corresponding Author.

* What is the **type of platform** used?

Scroll down the GSE page until you find the platform link, *eg.* *GPL1930*. If you want to analyze the gene expression profile you have to be sure that the platform is not an exon array (too many rows) or a different probe array (platform divided into several arrays).

* How many and which **samples** are used in the experiment?

In the GSE page, the “Sample” section is below the platform link. Clicking on “**More...**”, a list of sample accession number will appear, *eg*. *GSM333483, GSM333484, etc*...

You have to identify the GSM of interest, *eg. only controls.*

Make sure that at the bottom of the GSM list there isn’t a message of incomplete list.

In the case of *incomplete list*, a text file is provided in “**Supplementary Files**”, including the expression values of all GSM of that experiment​​, from which it is possible to infer data about the lacking samples.

**1.1 Download the data**

**Create a folder for each experiment (GSE)** found in GEO to save the samples files (GSM) in text format, the platform files (GPL) in text format and the paper (preferably save it with the First Author’s name and the year of publication).

To **download the paper**, go to the GSE page.

At the "Citations" section, click on the PMID number to open PubMed page.

Download and save the paper with the First Author's name and the year of publication, *eg. Silva 2012*.

If the paper has not yet been published in PubMed send an e-mail to the Corresponding Author or any of the Authors mentioned in the GSE page.

To **download the platform**, go to the GSE page, scroll down and click on the platform link.

In the platform page you can find features to be recorded.

The table of the platform is at the bottom of the page.

Click on “**DOWNLOAD FULL TABLE”** button. Wait until the file is completely downloaded.

Choose “**Save As**” from the “**File”** menu of your browser. Choose "**Text**" as the file format.

Save this file in the GSE folder with the name of the platform, *eg. GPL570.txt*.

If you can’t download the platform, but you only see it in html format, click on “**View Full Table**”.

In the “**File**” menu, click on “**Save page as**” and in “**Format**” choose “**Text file**”.

Save this file in the GSE folder with the name of the platform.

NOTE: To download the sample data *automatically*, you can use “GEO\_GSM\_Download”, a useful software to automatically download a list of samples from the GEO database.

In order to use this software, you need to have “Python2” or “Python3” applications in the operating system, otherwise you can download it from the site <http://www.python.org/download/>.

You can use “GEO\_GPL\_Download\_2.py” with “Python 2” application.

You can use “GEO\_GPL\_Download\_3.py” with “Python 3” application.

Identify the Platform (GPL) files to be downloaded.

Create and save in the TRAM "Platforms" folder a file named “GPL\_List.txt” in which to write the list of GPL that you want to download:

GPLnnn..GPLnnn (if an interval of consecutive GPL)

GPLnnn  
GPLnnn  
Save in the TRAM "Platforms" folder a copy of “GEO\_Download\_2.py” (for Python 2) or a copy of “GEO\_Download\_3.py” (for Python 3). Double-click the program icon and from “**RUN**” menu select "**Run Module**". Wait until the item "Download completed" appears.

The list of completely downloaded GSM, in text format, will appear in the "Platforms" folder.

Common mistakes:

The platform isn’t in compliance with the search criteria, eg. it is an *exon array* or a *different probe* array;

the platform hasn’t identifiers corresponding to those found in the GSM files;

the platform hasn’t a standard format or it has an atypical number of genes (e.g. <5.000 or> 60.000).

*You should try processing the platform in TRAM (table "Platform", button "Set Up Platform") to verify immediately the compliance of the platform to the minimum criteria (presence of GenBank Accession Numbers and / or gene symbols in usable format).*

You can **download the sample data** in two different ways:

manually;

automatically.  
  
To download the sample data *manually*, go to the GSE page, scroll down to the list of GSM.

For each sample it will open a page with its own features, *eg. man or woman, age, source of the sample, etc...,* and at the bottom of the page there is the data table.

Click on "**View full table**". Wait until the file is completely downloaded.

Choose "**Save As**" from the "**File**" menu of your browser. Choose "**Text**" as file format. Save this file with the name of the sample, *eg. GSM14526.txt*, in the corresponding folder of the GSE.

Common Mistakes:

Saving pages in "Web" or "HTML" format rather than in simple text format.

In this case, you will note e.g. words in bold type in the downloaded file.

Once you have downloaded the first GSM, go to the next one on the list.

Once you have downloaded all the desired GSM, sort files depending on the numbering of the GSM.

NOTE: To download the sample data *automatically*, you can use “GEO\_GSM\_Download”, a useful software to automatically download a list of samples from the GEO database.

Warning: The utility "GEO\_GSM\_Download" is not adequate if in GEO Table "Data Table" there is only the ratio between two channels, while the raw data can be found in "Supplementary Files". In this case, you have to manually download additional data.

In order to use this software, you need to have “Python2” or “Python3” applications in the operating system, otherwise you can download it from the site <http://www.python.org/download/>.

You can use “GEO\_GSM\_Download\_2.py” with “Python 2” application.

You can use “GEO\_GSM\_Download\_3.py” with “Python 3” application.

Identify the GSM files to download.

Create and save in the appropriate GSE folder a file named “GSM\_List.txt” in which to write the list of GSM that you want to download:

GSMnnn...GSMnnn (if an interval of consecutive GSM)

GSMnnn

GSMnnn  
Save in the GSE folder a copy of “GEO\_GSM\_Download\_2.py” (for Python 2) or a copy of “GEO\_GSM\_Download\_3.py” (for Python 3). Double-click the program icon and select "**Run Module**" from “**RUN**” menu.

Wait until the item "Download completed" appears.

The list of completely downloaded GSM, in text format, will appear in the GSE folder.

**1.2 Recording of data**

Create a spreadsheet (e.g., Excel) table with as many columns as are the information to be recorded about the experiments, *eg. a table with 14 columns using the following fields:*

(PROGRESSIVE NUMBER) - GSE - GPL - NUMBER OF SAMPLES - GSM - SEX - AGE - SOURCE - ARRAY TITLE - VALUE TYPE - SAMPLE ROWS - PLATFORM ROWS - REFERENCE - PMID

Fill in the table as the example:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GSE** | **GPL** | **SAMPLES** | **GSM** | **SEX** | **AGE** | **SOURCE** | **ARRAY TITLE** | **VALUE TYPE** | **SAMPLE ROWS** | **PLATFORM ROWS** | **REFERENCE** | **PMID** |
| GSE5390 | GPL96 | 6  (CONTROL) | GSM123273 | M | 59 | BRAIN | [HG-U133A] | RMA | 22283 | 22283 | LOCKSTONE 2007 | [17950572](https://www.ncbi.nlm.nih.gov/pubmed/17950572) |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |  |

For each sample, insert the accession numbers of the GSM, GSE, and of the corresponding GPL.

In the GSM page are usually given the **age and sex** of the individual doning the sample and the **sample source**.

**The array title** is indicated in the GPL page.

At the bottom of GSM page, above the data table, it is indicated the **value type** (the nature of the number).

At the bottom of the GSM page, under the data table, are indicated the **sample rows**.

At the bottom of the GPL page, under the platform table, it is indicated the number of the **platform rows**.

In the GSE page are indicated the **Author** surname**,** the **year** of publication and the **PMID** of the paper.

**1.3 Common issues**

**a) The expression values** ​​can be **linear** or **logarithmic**.

If expression values are logarithmic, it must be specified which logarithmic base is used because TRAM will request it during the "importing files" step.

Values are **linear** ​​when linear value, MAS 5 (Affymetrix Microarray Suite 5 method) or GCOS are indicated in "Value Type".

Values are **log2** when log2 value or RMA (robust multi-array average) is indicated.

In most cases logarithmic values ​​are log2, but you always need to check the paper.

When the identification of the value type ​​used is not immediate you have to see if it is specified in the “Materials and Methods” section of paper or you have to ask theAauthor via e-mail.

*The user could distinguish linear ​​from logarithmic values because linear values are presented as very different from each other even as order of magnitude, instead the logarithmic values as small numbers, typically decimal and close together.*

*In the case of logarithmic numbers, if you can’t distinguish the type of logarithmic base used you will need to ask the author.*

**b) The GSM data are raw data**, in other words the intensity values of the spots ​​ without the background (*signal coming from the area around the spot*) subtraction.

However, download the data tables in text format.

TRAM executes the background intensity subtraction from the spot intensity.

Before starting the processing, check in the tables if the position of the columns including the intensities of the spots and the intensities of the background is the same in all samples. Otherwise, always download the data table of each GSM in text format, open it with a spreadsheet (e.g., Excel) and select the columns of your interest, then, the column with the gene ID, the column with the intensity of the spots and the column with the intensity of the background.

Choose the columns with the median values​​.

Copy and paste these columns into a new worksheet and save the new file in tab-delimited text format.  
Once it’s been done, you can import the files into TRAM.

**c) The values ​​are obtained from the ratio between the intensities of two compared channels**, *eg. log ratio Ch1/Ch2.* The channel in which the signals of interest, *e.g. Ch1: brain, Ch2: pool of tissues,* are located is specified on the GSM page.

To distinguish the signals from the two channels, two fluorophores that emit at different wavelengths, Cy3 and Cy5, are used. The Cy3 emits at 532 nm, the Cy5 at 633 nm. Identify which fluorophore is used for the channel of interest.

Download the data table in text format.

Import the GSM files in TRAM. The software asks the user to select the column containing the IDs and the column containing the **intensities** of the channel of your interest. Always choose the column that shows the medians of the intensities.

If the **background** has not been subtracted, also select the column containing the medians of the background. TRAM will perform the subtraction of the background from the signal intensity.

You must be sure that the selected columns are in the same position in all GSM files, otherwise open the GSM.txt files with a spreadsheet (e.g., Excel), select the column containing the ID and the column containing the intensities coming from the channel of interest. Always choose the column that shows the **medians** of the intensities.

If the background has not been subtracted, also select the column containing the **medians** of the background. Copy and paste the selection into a new Excel worksheet. In a new column enter the subtraction operation between the median intensities of the spots and those of the background. Save the table in text format.

**d) The GSM data are available as image files**, *eg. CEL files*. The operator should convert them into files that can be imported in TRAM.

Use **AltAnalyze**, a software freely downloadable that converts CEL files into text files with the numerical values ​​of expression in log2.

Create 2 folders: one in which to save the file.CEL and another empty where AltAnalyze will save the file.txt. Open AltAnalyze and execute commands.

In cases where there are no data tables for each GSM, look for the supplementary files.

In these there should be a file containing a matrix of processed data of all GSMs.

Download the file, extract the zip-compressed file and open it with a spreadsheet (e.g., Excel). A series of columns will appear, one with the IDs and the other with the expression values ​​for each GSM.

Select the ID-column and the intensities column of each individual GSM, copy and paste them into a new worksheet and save in text format.

**e) Some values in the GSM file are recorded using the word "null".**

The operator should delete them using a word processor.

**2. SEARCH IN EBI ARRAY EXPRESS**

Go to <http://www.ebi.ac.uk/arrayexpress/>

Click on “**Experiment**”.

At the top right insert a keyword, *eg. Brain*, and click on “**Search**”button.

In “**Filter experiment**” choose the organism, *eg. Homo sapiens*, filter for “**all arrays**”, as experiment type choose “**RNA assay**” by “**Array assay**”, and click on “**Filter**”button.

You will see a page with a series of experiments, each with its own accession number.

Before choosing the experiment you need to click on the accession number and read the summary in “**Description**”.

Always check that data correspond to your search criteria, then the organism in “**Organism**” and type of platform in “**Array**”.

Once identified the first experiment, proceed with the same method of data recording used for the research in GEO. Complete the recording table inserting the necessary data in the specific fields.

The Array Express accession number corresponds to the GSE in GEO.

In the experiment page at the item “**Array**”, the features of the platform are specified, and there is the file to be downloaded in text format (the platforms are always in format *.adf.txt*, *eg. A-MEXP-131.adf.txt*).

In “**Citations**”, the paper title, the author and the PMID are indicated.

In “**Samples**”, click on “**Click for detailed sample information and links to data**”.

A page will open with the experiment samples and the main features of each of them.

Identify those of interest.

On the right of the page, in the column “**Raw**”, there are data files of each sample; in the adjacent column-“**Processed**”, there are files with processed data matrix, which can be used in the absence of files in “**Raw**”. In these cases, download the processed matrix from which to extract the data you need.

In the event of lacking files in both “**Raw**” and “**Processed**”, go back to the home page of the experiment, under the item “**Files**” you may find supplementary files from which to obtain the gene expression data.

Each data table must contain a column with the ID and a column with the gene expression values​​.

Open the processed matrix with a spreadsheet (e.g., Excel), copy and paste the columns you need in an new worksheet, save this data in text format.

**3. Importing data in tram**

To download the software go to the website <http://apollo11.isto.unibo.it/software/>, click on the TRAM folder, click on TRAM 2017 and then click on the compatible version with your operating system, *eg.* TRAM\_Mac\_OS\_X /.

For data processing, you can also use TRAM.zip.

Extract TRAM from the zip file.

In <http://apollo11.isto.unibo.it/software/>, click TRAM, then click on TRAM\_2023 and then on Docs /**,** where you find a detailed guide to be able to use the software (TRAM\_Guide\_1.3.1.html).

*Before using the software check in* ***System Preferences*** *that the* ***number format is UK****.*

Copy and paste the software in the folder of each experiment.

Open TRAM.

Sub-folders will appear including one called “**Series**”.

Copy and paste all GSM files downloaded from that experiment in “**Series**”.

Rename the GSM files with the letter “S” followed by the progressive number*, eg. S1.txt, S2.txt, etc*..

To rename files more quickly you could use “**NameChanger**” (a software utility for the Mac).

Click on “**NameChanger**”, paste the names of the GSM files, choose “**Sequence**” from the pulldown menu at the top.

Assign the letter “S” to the file list, insert “**1**” in “**Number of digits**” and reinsert “**1**” in “**Starting at**”.

Click “**Replace entire name**”, and then click on “**Rename**”.

Select “**Append**” from the pulldown menu and in the adjacent space insert “.txt”.

In this way there will be all the GSM files renamed as file “S” with its own serial number and in text format in the folder “**Series**”.

TRAM uses linear expression values ​​and not logarithmic, but if necessary it can convert them into linear values ​​using the option “**Help with data**”in the Home page of the software.

Double-click the TRAM icon which is located in the folder “**Series**”.

Click “**Help with data**” and in the new page that will open click on “**Data batch file processing**”.

TRAM will import the S1.txt, S2.txt... files.

You have to indicate which columns correspond to the gene ID and to the expression value.

Check if there is a correspondence between the gene IDs used for the samples and the ones used in the platform.

Then go to the top right corner of the screen and click “**Continue**”.

Indicate to the software the name of the platform used in the experiment.

After a few minutes TRAM will have processed the “S#.txt” files in the folder “**Series**”, creating a series of data uniformly processed and renamed as “P”.

Open each "P.txt" file and check in the gene symbols column that there are no strange characters instead of the official gene symbol.

**3.1 File Subdivision in the folders "batch import a" and "batch import b"**

The software was also designed to make comparisons between two distinct biological conditions.

The user can then download the data of each condition and sort them into two different batches, A and B.

*E.g. the objective is to compare the gene expression between a normal tissue (brain) and a pool of other normal tissues.*

You need to find in GEO and Array Express the experiments on brain and on more tissues.

The user will create the pool of tissues to use as Batch B.

To create the pool of tissues you have to consider only experiments on organs in toto (eg.: "Adrenal Gland", not "Adrenal Gland Cortex").

Download samples from the experiments, sorting them for tissue (a folder for each tissue) and process them with TRAM.

Create two folders, one named "Batch Import A" and the other "Batch Import B".

Copy the “P” files obtained from GSE on the brain in the "Batch Import A" folder, renaming them as “A” followed by a serial number.

Copy the “P” files obtained from GSE on the tissues of pool in the "Batch Import B" folder, renaming them as files “B” followed by a serial number.

To rename files more quickly use “**NameChanger**” as described above. You can rename the samples with “NameChanger” in the same way to rename the GSM files in “S” files.

Add a new column to the spreadsheet table of the experiment, recording and inserting in this the correspondence between the GSM files and files A or B.

**WORKING PROTOCOL**

**I. Search data in GEO**

1. Web site: <https://www.ncbi.nlm.nih.gov/geo/>

2. In "**QUERY**"– "**Datasets**"insert search terms and click on "**GO**".

3. In "**Filter your results**"choose the "**Series**" item.

4. Choose the experiment and click on the title.

**I.I Download data**

5. Create a folder with the name of the experiment (GSE) where you will save

the paper, the samples and the platform.

6. In the GSE page click on the PMID, which refers to PubMed.

7. Download and save the paper with the name of the first author and the year of publication.

8. In the GSE page, click on the platform (GPL) link.

Common mistakes:

The platform isn’t in compliance with the search criteria, eg. it is an *exon array* or *different probes*;

the platform hasn’t identifiers corresponding to those found in the files GSM;

the platform hasn’t a standard format or it has an atypical number of genes (eg. <5.000 or> 60.000).

*You should try processing the platform in TRAM (Table "Platform", button "Set Up Platform")*

*to verify immediately the compliance of the platform to the minimum criteria*

*(presence of GenBank Accession Numbers and / or gene symbols in usable format).*

9. At the bottom of the page, click "**Download Full table**".

10. Wait until the page is completely downloaded.

11. Choose “**Save file**” from the "**File**" menu.

12. In "**Format**" choose "**text file**".

NOTE: Steps 8.-12. may be replaced (provided that alert at point. 8 about common mistakes has been considered) by using the utility “GEO PlatformDownload” to download the sample data *automatically* (section 1.1 above).

13. Save the file using the platform name, *eg. GPL96.txt,* in the corresponding GSE folder.

14. Identify the GSM files to download (manually or automatically).

14.1 Download *manually* GSM files.

14.1.1 Go to the GSE page, scrolldown to the GSM list.

14.1.2 Click on each GSM file.

14.1.3 Click on “**View full table**” at the bottom of the GSM page.

14.1.4 Wait until thepage is completely downloaded.

14.1.5 Choose "**Save as**" from the browser "**File**" menu, and choose "**Text**" as file format.

14.1.6 Save using the sample name, *eg. GSM 14526.txt*, and in the corresponding GSE folder.

Common Mistakes:

Save pages in "Web" or "HTML" format or in no simple text format.

In this case, you will note eg. words in bold type in the downloaded file.

Once the first GSM is downloaded, go to the next one on the list.

Once you have downloaded all GSM, sort files respecting the numbering of the GSM.

14.2 Download *automatically* the GSM files

14.2.1 Create and save in the GSE folder a file named “GSM\_List.txt” in which to write

the list of GSM files you want to download and then:

GSMnnn...GSMnnn (if an interval of consecutive GSM)

GSMnnn  
 GSMnnn

14.2.2 Save in the folder a copy of “GEO\_Download\_2.py” (for Python 2)

or a copy of “GEO\_Download\_3.py” (for Python 3).

Note: Using the utility "GEO\_Download" is not adequate if in Table "Data Table" there is only the ratio between two channels, while the raw data can be found in "Supplementary Files". In this case, you have to manually download additional data.

14.2.3 Double-click the program icon and from the “**RUN**” menu select "**Run Module**". 14.2.4 Wait until the message "Download completed" appears.

14.2.5 The list of completely downloaded GSM files in text format will be in the GSE folder.

15. Create a working folder to save all GSE found in GEO and Array Express.

16. Respect the serial numeration of the GSE both in the folder and in the table of data recording.

**I.II. Data Recording**

17. Create an excel table to record the information about the experiment using the following fields:

GSE - GPL - NUMBER OF SAMPLES - GSM - SEX - AGE - SOURCE - ARRAY TITLE - VALUE TYPE - SAMPLE ROWS - PLATFORM ROWS - REFERENCE - PMID - NOTE

For each sample, insert the number of the GSM, GSE, and of the GPL.

In the GSM page are usually given the **age and sex** of the individual doning the sample

and the **sample source**.

**The array title** is indicated in the GPL page.

At the bottom of GSM page, above the data table, is indicated the **value-type**.

At the bottom of the GSM page, under the data table, are indicated the **sample rows**.

At the bottom of the GPL page, under the platform table, are indicated the **platform rows**.

In the GSE page are indicated **the author, the year of publication and the PMID of the paper.**

**I.III. Data importing and processing in tram**

18. Copy an empty version of TRAM in the GSE folder.

19. Click on TRAM.

20. Open the “**Series**” folder.

21. Copy and paste GSM files in the “**Series**” folder.

22. Rename GSM files in “**S**” (e.g., using “**NameChanger**”).

23. Click on TRAM icon in the “**Series**” folder.

24. Click on “**Help With Data**” button.

25. Click on “**Data File Batch Processing**” button.

26. When the software requests to specify the column, click on “**OK**”.

27. Indicate the ID column and the expression value column.

28. When TRAM requests the platform, indicate the GPL name if available.

29. Wait while TRAM processes the data.

30. Verify that the number of processed GSM corresponds to the number of GSM present in GEO for that GSE.

31. Create the “Batch Import A” and “Batch Import B” folders.

32. Copy and paste in these folders the corresponding “P” files and rename them as “A” or as “B”.

**II. Search in Array Express**

1. Web site <http://www.ebi.ac.uk/arrayexpress/>

2. Click on "**Experiment**".

3. At the bottom insert a key word, *eg. Brain,* and click on “**Search**” button.

4. In "**Filter experiment**” choose the organism*, eg. Homo sapiens*, filter by “**all arrays**”,

as experiment type choose “**RNA assay**” by “**Array assay**” and click on “**Filter**”.

5. Click on the experiment title.

**II.I Download data**

6. Create a folder in which to save data of each experiment.

7. In the experiment page, in “**Citation(s)**” click on the link to download the paper,

or use the contact e-mail provided in “**Contact**” to request it.

8. Download and save the paper with the first author's name and the year of publication.

9. In the experiment page, at the “**Array**” item click on the platform title.

10. Download the platform file (text format) at the bottom of the page.

11. In the experiment page, at the “Samples” item click on:

**Click for detailed sample information and links to data.**

12. Identify the samples of interest.

13. Download files under the column “**Raw**” (text format).

In the absence of “**Raw**” files, download “**Processed**” files.

**II.II Data Recording**

14. Create an Excel table to record the information about the experiment using the following fields:

GSE - GPL - NUMBER OF SAMPLES - GSM - SEX - AGE - SOURCE - ARRAY TITLE - VALUE TYPE - SAMPLE ROWS - PLATFORM ROWS - REFERENCE - PMID - NOTE

For each sample, enter the name used by ArrayExpress instead of GSM, GSE, and of the GPL.

In the GSM page are usually given the **age and sex** of the individual doning the sample

and the **sample source**.

In the platform page is shown **the name of the platform**.

In the file of each sample present in "**RAW**" is indicated **the type of expression values**​​.

In the experiment page, in "**Citation (s)**" are given the author, the year of publication and the PMID.

**II.III Data importing and processing in tram**

15. Copy an empty version TRAM in the experiment folder.

16. Click on TRAM.

17. Open the “**Series**”folder.

18. Copy and paste samples files in “**Series**”.

19. Rename the files names with “**NameChanger**” in “**S**” files.

20. Click on the TRAM icon under the "**Series**" folder.

21. Click on “**Help with data**”.

22. Click on “**Data File Batch Processing**”.

23. When TRAM requests to indicate columns, click on “**OK**”.

24. Indicate the ID column and the expression value column.

25. When TRAM requests the platform, indicate the **Array** name.

26. Wait while TRAM processes data.

27. Create “**Batch Import A**” and “**Batch Import B**” folders

28. Copy and paste in these folders the corresponding “P” files and rename them as “A” or as “B”.

(folder "**A**": A1.txt, A2.txt...; folder "**B**": B1.txt, B2.txt...).

**TRAM Basic Final Checklist - Before you start your analysis**

- Are you running "TRAM" from your local hard drive? Avoid remote/network drives.

**Set Up**

- Is the "Chromosome" table filled in with the correct chromosome data?

- Is the "Gene" table filled in?

- Is the "Aliases" table (accessed from the "Gene" table) filled in?

**Expression Data Import**

- Are the data files numbered as A1.txt, A2.txt ... in the "Batch Folder A"?

- Are the data files numbered as B1.txt, B2.txt ... in the "Batch Folder B"?

- Are the numbers in the data file in UK format? (Decimal separator is a full stop mark, i.e. ".").

- Is your computer operating system set to use the same standard (English) for numbers?

- Are the numbers in the data files linearized? (Not logarithmic).

If you use "Platform" table to resolve gene identifiers recorded in the expression data files:

- Is each data file labeled by the name of the correct, related Platform in the third column of the first row?

- Is each Platform, recorded by any data file, actually and entirely loaded in the "Platform" table?

**Analysis**

- In the "Map" mode setting: have you set the "Number of genes in the Window" parameter consistently with the "Window" size? E.g., for the analysis of human genes set "1" if the window is 25,000 bp.

**Results**

In the "Map" layout:

- Have you regenerated the visualization of over/-underexpressed or significant Segments clicking on the related buttons when shifting from "Genome median based" layouts to "Chromosome median base layouts"? (e.g., from "Map A" layout to "Map (chr) A" layout).

**Suggested Readings**

TRAM (Transcriptome Mapper): database-driven creation and analysis of transcriptome maps from multiple sources

Lenzi et al., *BMC Genomics*, 2011

<https://www.ncbi.nlm.nih.gov/pubmed/21333005>

Integrated Transcriptome Map Highlights Structural and Functional Aspects of the Normal Human Heart

Caracausi et al., *J Cell Physiol*, 2017

<https://www.ncbi.nlm.nih.gov/pubmed/27345625>

A molecular view of the normal human thyroid structure and function reconstructed from its reference transcriptome map

Vitale et al., *BMC Genomics*, 2017

<https://www.ncbi.nlm.nih.gov/pubmed/28923001>

Integrated Quantitative Transcriptome Maps of Human Trisomy 21 Tissues and Cells

Pelleri et al., *Front Genet*, 2018

<https://www.ncbi.nlm.nih.gov/pubmed/29740474>