* **Downloading the file**
* **Edit the excel file**
* Delete the GWEIGHT column
* Delete the EWEIGHT row
* Delete the NAME column
* Change the heading from wt-matA\_(time)min\_nbr029-a format to (time)m (e.g. wt-matA\_0min\_nbr029-a to 0m)
* Change the NAME column heading to Gene Symbol
* Compute the average of the same timepoints in a new excel sheets
* Export a copy of the data into a tab-delimited text format file

(This file should have 12,983 genes)

* **Run ID-sorting-code.py code**
* remove the space in Gene Symbol heading to avoid errors in the results
* make sure to give the correct file paths for code to work

(A file should be created with 5,624 genes with a mix of SGD and systematic names)

* **Download Yeast nomenclature from Yeast mine website (to convert sgd ID’s)**
* Visit [https://yeastmine.yeastgenome.org/yeastmine/bagDetails.do?scope=all&bagName=ALL\_Verified\_Uncharacterized\_Dubious\_ORFs#](https://yeastmine.yeastgenome.org/yeastmine/bagDetails.do?scope=all&bagName=ALL_Verified_Uncharacterized_Dubious_ORFs)
* Click on Export
* Select All columns from options on the left
* De-select Organism short name, Standard name, and name
* Click on Download file

(This gives a tsv file format)

* Save a copy in tab-delimited text format and name it YeastNomenclature.txt
* **Converting SGD ID to systematic ID’s**
* Open the resulting file of ID-sorting-code.py code in excel
* Sort all data from A-Z
* Select all SGD Id’s (there are 721 of them) and copy them in a separate file (tab-delimited text format)
* Run SGD-remover.py code
* Use the resulting file in sgd-to-systematic-ID.py code as the template and then run the code

(The resulting file contains the systematic ID instead of SGD ID’s) [there is fewer genes than original file because the database isn’t complete]

* **Replacing the SGD ID’s not in the database**
* Open the file containing 721 SGD ID’s in excel
* Copy the SGD ID’s from the database to the file containing the SGD ID’s (the file with 721 SGD ID’s)
* Select the SGD ID’s from expression data and the SGD ID’s from yeast database
* Select conditional formatting < Highlight Cells Rules < Duplicate Values < keep default setting and click ok (this will highlight the ID’s that have been replaced by the code)
* The ID’s that are not highlighted are looked up in YeastMine (there are 34 of them)
* Replace their ID’s and add them to the rest of the systematic ID’s
* **Creating a file with systematic names**
* Add the systematic ID’s from the SGD converted file to the rest of the systematic names

(This gives 5590 genes; however, some of them are aliases)

* **Remove the aliases**
* Run the removing\_duplicates.py code to remove aliases

(it doesn’t matter which ID is removed because YEASTRACT can identify the any ID’s including aliases)

* **The final cleaned file contains 5569 genes**
* This was confirmed by running a test to determine all unique expression data on the original data regardless of their ID’s (Meaning there are 5569 genes with unique expression data)
* **The ID’s were then used in YEASTRACT to obtain their standard names**