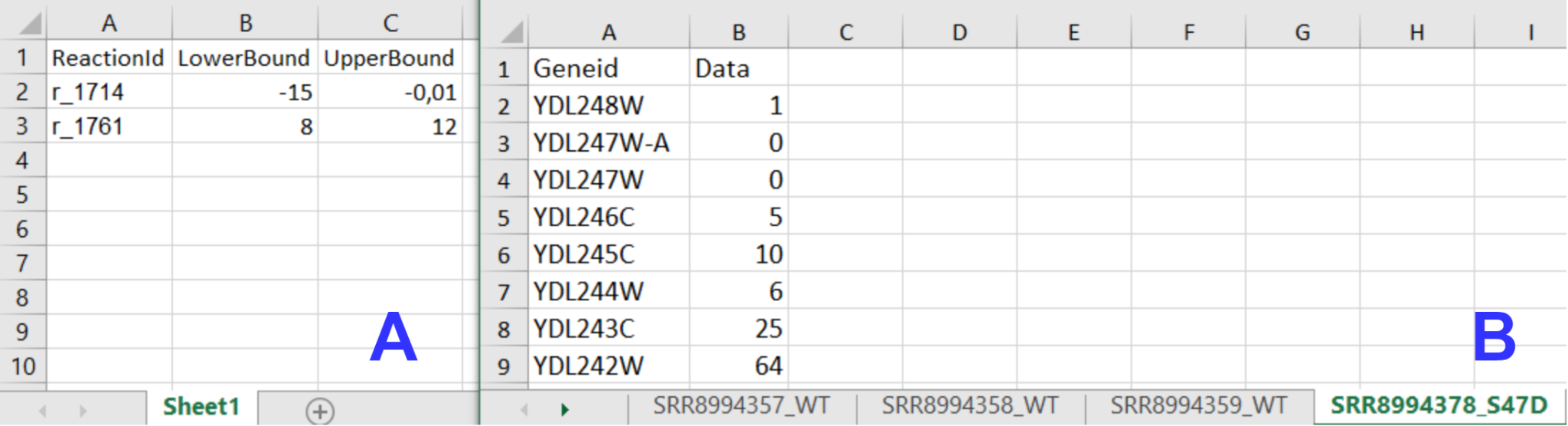
IgemRNA is a library with a graphical user interface written for the MATLAB environment and facilitates some of the Cobra Toolbox 3.0 functionality. IgemRNA performs not only Gene sets enrichment analysis-based functions, but also allows integrate transcriptomics data in metabolic models. Also, IgemRNA allows validate transcriptomics data facilitating interconnectivity of biochemical networks, steady state assumptions, Gene - Protein - Reaction relationship and can use optional medium composition data to create context-specific models.

1. **Folder structure description**

Files are extracted from the archive (https://github.com/BigDataInSilicoBiologyGroup/IgemRNA). The IgemRNA tool consists of four root folders (Data, Scripts and Results non-optimization, Results post-optimization) and an *IgemRNA.m* file which calls the user graphical interface form. Data folder is where the input data files are stored, initially this folder contains the data files used for this demonstration:

1. *MediumData.xlsx* (medium composition data)
2. *Yeast\_8\_4\_0.xls* (the yeast consensus genome-scale model)
3. *TranscriptomicsData.xlsx* (RNA-req measurements) (available <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE130549>)

Transcriptomics data and medium composition data can be provided as an .xls or an .xlsx file and must meet the following format (Fig. 1) where shown columns are provided and named accordingly and sheet names correspond to a phenotype name see 3.2 IgemRNA demonstration.



**Fig. 1.** Input data file structure; A - Medium data file structure;

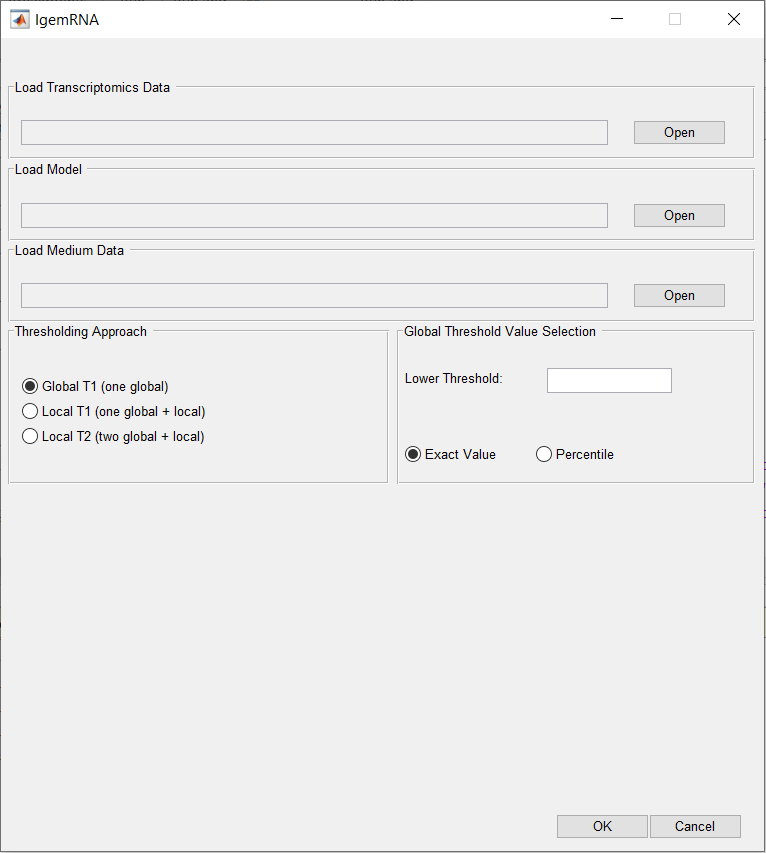
B - Transcriptomics data file structure

The model can be provided in xls, sbml or other formats supported by Cobra Toolbox 3.0.

Scripts folder consists of all the script files that are being executed by the IgemRNA tool according to user's selections in the IgemRNA form as well as the test cases provided in this demonstration.  
The Results non-optimization and Results post-optimization folders are where all the result files are being saved. These folders are initially empty (for more details see section in main publication Materials and Methods 2.2 Tools functionality description).

1. **Starting IgemRNA tool**

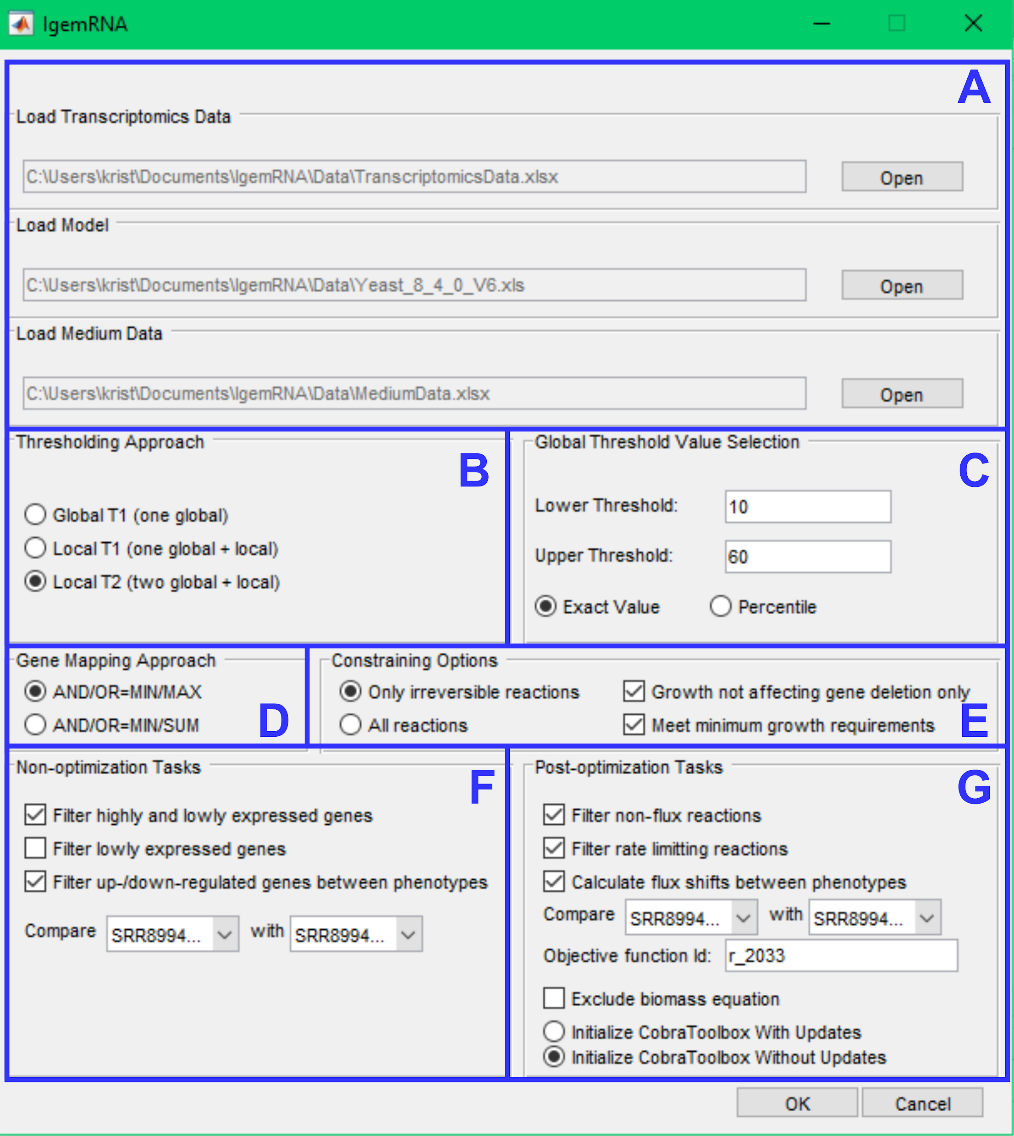
In order to start the IgemRNA tool a user must start the Matlab environment and run the *IgemRNA.m* script located in the root folder. This script opens the graphical user interface of IgemRNA (Fig. 2).



**Fig.** **2.** IgemRNA start window

1. **File upload**

To access all options in the IgemRNA form, the user must supply input data files Fig. 3 A section. This can be done by pressing the ‘Open’ button in the corresponding file row and finding the necessary files via File Explorer. Transcriptomics data is required to run non-optimization tasks (Fig. 3 F) but an additional model file is necessary to access the post-optimization tasks (Fig. 3 G). Medium composition file is optional if such data is available, the selection of this data file does not extend the form, but specifies the given exchange reaction constraints (upper bounds and lower bounds) on the model for post-optimization tasks analysis. For an organized overview of the analysis and results it is recommended that the necessary data files are located in the Data folder of IgenRNA tool (for more details see main publication section Materials and Methods 2.2 Tools functionality description).



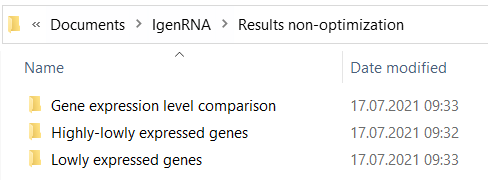
**Fig.** **3.** Full IgemRNA form

1. **Running test cases**

In order to perform test cases provided in this user manual, simply run the provided test case scripts via MATLAB environment having initialized CobraToolbox 3.0 beforehand. Test case script file names are given at the end of each test case section.

1. **Non-optimization tasks**

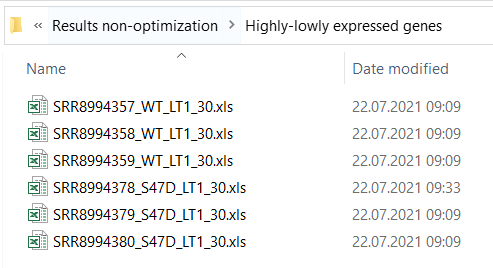
Non-optimization tasks include several transcriptomics data analysis tasks: filter highly and lowly expressed genes, filter lowly expressed genes, filter up/down regulated genes between different phenotypes or data sets. The results for each task are stored in a different folder within the ‘Results non-optimization’ folder: Gene expression level comparison, Highly-lowly expressed genes, Lowly expressed genes (Fig. 4).

****

**Fig.** **4.** Non-optimization results folder

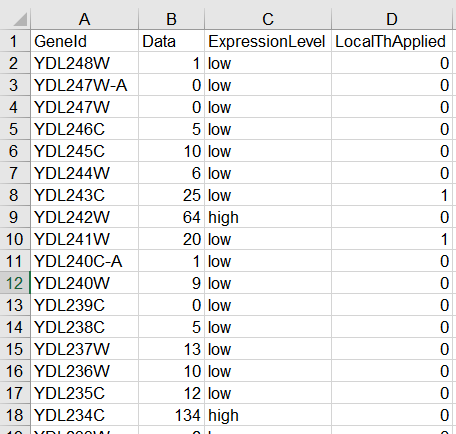
* 1. **Filter highly and lowly expressed genes**

Non-optimization task ‘Filter highly and lowly expressed genes’ generates result excel files for each provided transcriptomics data set. File names are assigned based on the provided dataset and phenotype name (from transcriptomics data), the selected thresholding approach (GT1, LT1, LT2) and provided global thresholds values (Fig. 5).

****

**Fig. 5.** Highly-lowly expressed genes folder

Each excel file contains one sheet with the list of genes provided by transcriptomics data and 4 columns: GeneId, Data (the expression value), ExpressionLevel and LocalThApplied. The ExpressionLevel column contains the expression levels determined based on the chosen thresholding approach, provided global and for thresholding approaches (LT1 and LT2) calculated local thresholds. Column LocalThApplied displays whether a local threshold for a specific gene was applied (Fig. 6).

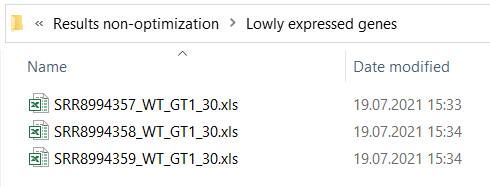
****

**Fig. 6.** Filter highly and lowly expressed genes result file (thresholding approach LT1)

To perform this test case run the file *TestCase\_determineGeneActivity.m* in the ‘Scripts’ folder of IgemRNA tool.

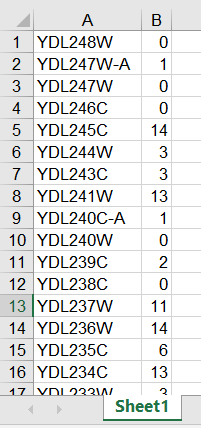
* 1. **Filter lowly expressed genes**

Non-optimization task ‘Filter lowly expressed genes’ generates separate excel result files for each dataset provided in transcriptomics data file. These result files contain filtered gene lists including genes that are below a given threshold based on the selected thresholding approach. File names include dataset and phenotype name (from transcriptomics data file), thresholding approach (GT1, LT1, LT2) name and provided global threshold values (Fig. 7).

****

**Fig.** **7.** Non-optimization results folder

The result for lowly expressed genes coincides with the provided transcriptomics data format. Each file consists of two columns, gene identificator and the expression level, but only those genes that are below a given threshold (depending on which thresholding approach is applied) are listed in the result files. The test case provided for this task shows genes with expression level below 30 using the Global T1 approach (Fig. 8).

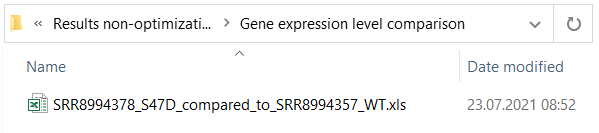


**Fig. 8**. Lowly expressed genes result file

To perform this test case run the file *TestCase\_filterLowlyExpressedGenes.m* in the ‘Scripts’ folder of IgemRNA tool.

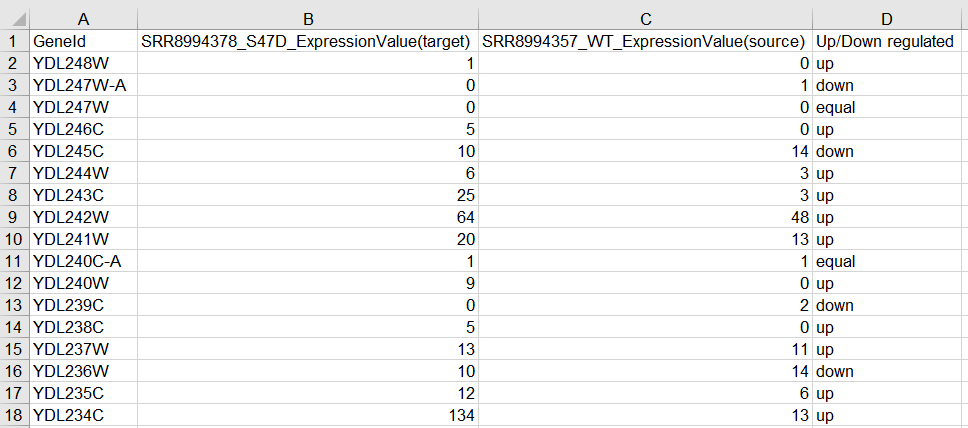
* 1. **Filter up/down regulated genes between phenotypes**

Non-optimization task ‘Filter up/down regulated genes between phenotypes’ generates result excel files in the ‘Gene expression level comparison’ folder. Result file names contain dataset and phenotype names for both transcriptomics datasets that have been compared (Fig. 9).

****

**Fig. 9.** Up/Down regulated genes in comparison to another phenotype

These result excel data files contain a full gene list from the target dataset and the corresponding genes that are found in the source dataset (Fig. 10. A column). Expression values for both target and source dataset are displayed (Fig. 10. B, C columns) as well as the determined up/down regulation status (Fig. 10. D column).

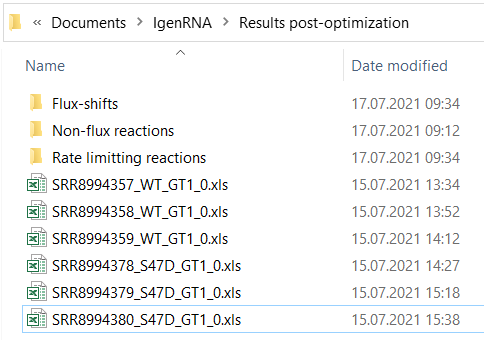
****

**Fig. 10.** Up/Down regulated genes in comparison to another phenotype result file

To perform this test case run the file *TestCase\_compareGeneExpressionLevels.m* in the ‘Scripts’ folder of IgemRNA tool.

1. **Post-optimization tasks**

Context-specific models generated by IgemRNA post-optimization tasks as well as the results of the analysis performed on these models are saved in the ‘Results post-optimization’ folder of IgemRNA tool (Fig. 11). The post-optimization tasks are saved in the folders with the corresponding name: Flux-shifts, Non-flux reactions and Rate limiting reactions (for more details see section Materials and Methods 2.2 Tools functionality description).

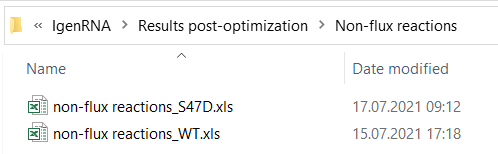


**Fig. 11.** Results folder after post-optimization task execution

To generate context-specific models used for these test cases run the file *TestCase\_generateContextSpecificModels.m* in the ‘Scripts’ folder of IgemRNA tool. Since this script takes a long time to execute, the context-specific model files used for this demonstration have already been provided in the ‘Results post-optimization’ folder.

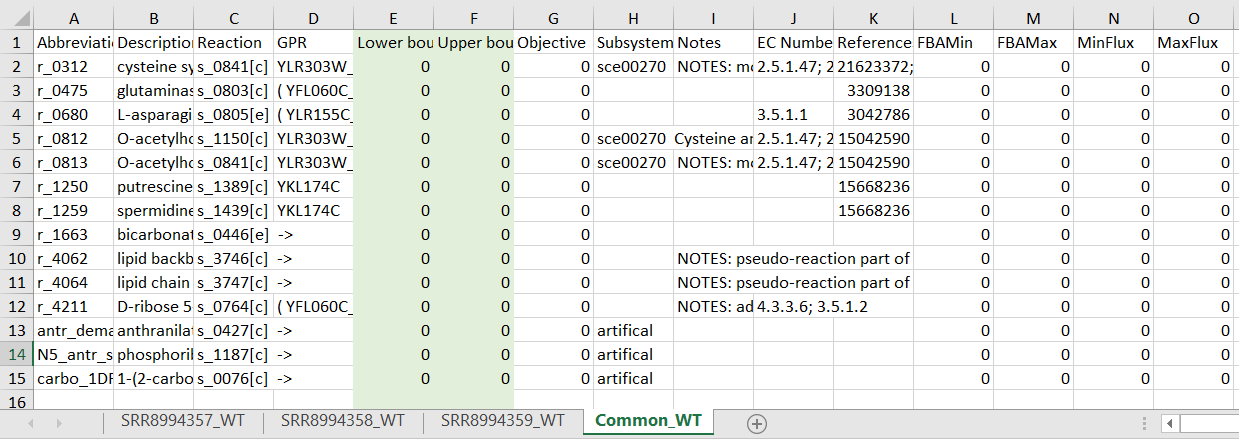
1. **Filter non-flux reactions**

Post-optimization task ‘filter non-flux reactions’ performs an analysis on the created context-specific models of the same phenotype, the name of the phenotype is included in the result file name (Fig. 12). This analysis filters those reactions that carry no flux.



**Fig.** **12.** Non-flux reactions result folder

Each result excel file contains a list for each transcriptomics dataset of reactions that carry no flux in the result context-specific model created by integration of the supplied transcriptomics data into the provided model. An additional sheet for all the common non-flux reactions of the same phenotype is also provided in the sheet ‘Common\_(phenotype name)’ (Fig. 13).

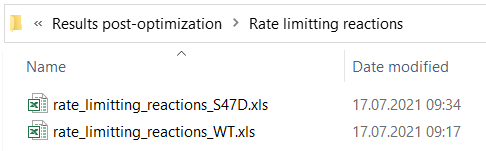


**Fig.** **13.** Wild type non-flux reaction task result

To perform this test case run the file *TestCase\_NonFluxReactions.m* in the ‘Scripts’ folder of IgemRNA tool.

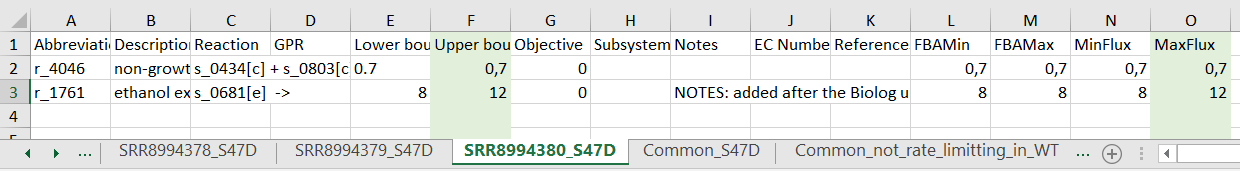
1. **Filter rate limiting reactions**

Post-optimization task ‘filter rate limiting reactions’ performs analysis on the generated context-specific models of the same phenotype, the phenotype name is included in the result files (Fig. 14).

****

**Fig. 14.** Rate limiting reactions result folder

Each result file contains sheets for each provided transcriptomics dataset of the same phenotype that has been integrated in the supplied model creating context-specific models. An analysis on these context-specific models have been performed in order to filter reactions that have reached the maximal flux (MaxFlux calculated by FVA) based on the upper bound set according to transcriptomics data and GPR associations. An additional sheet for common rate limiting reactions has also been added to the result file where rate limiting reactions that are present in all datasets are listed (Fig. 15).

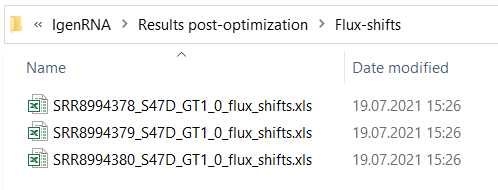
****

**Fig. 15.** S47D phenotype rate limiting reaction task result

To perform this test case run the file *TestCase\_RateLimitingReactions.m* in the ‘Scripts’ folder of IgemRNA tool.

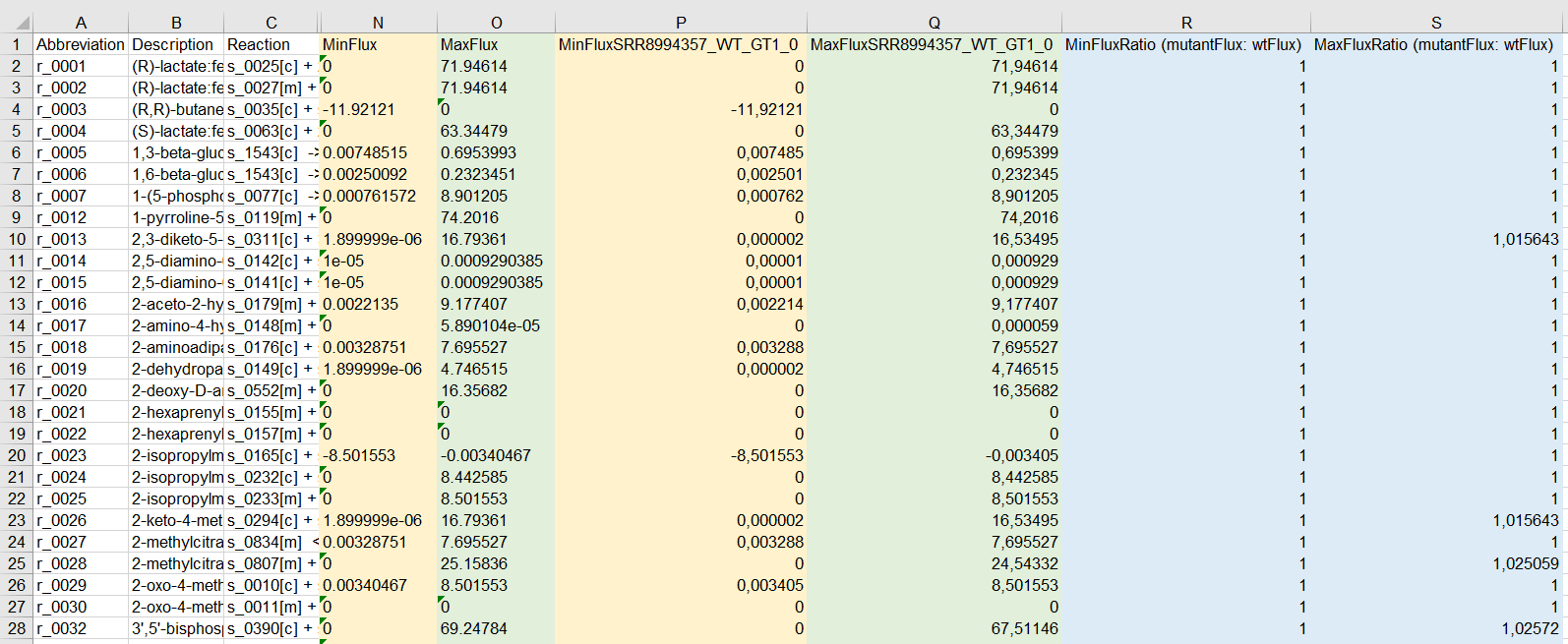
1. **Flux shifts calculation between different phenotypes**

Post-optimization task ‘calculate flux shifts between phenotypes’ compares flux values (calculated by FVA on the context-specific models) between two different phenotypes. In this demonstration flux shifts analysis task was performed on the S47D phenotype datasets using global T1 thresholding approach with the lower global threshold value of 0, these phenotypes were compared to the wild type dataset SRR8994357\_WT of the same thresholding approach and threshold values. Result file names include the dataset and phenotype name (provided in transcriptomics data file), used thresholding approach and global threshold values (Fig. 16).



**Fig. 16.** Flux-shifts result folder

Each result file contains a full reaction list that corresponds to the ‘Reaction List’ sheet in the provided model file as well as additional columns for the calculation results: MinFlux and MaxFlux values (phenotype that is compared, Fig. 17. N, O columns), MinFlux/MaxFlux(dataset name)\_(phenotype name)\_(thresholding approach)\_(global threshold values) phenotype that is used for comparison (Fig. 17. P, Q columns) and the MinFlux/MaxFlux ratio between these two phenotypes (Fig. 17. R, S columns).



**Fig. 17.** Flux-shifts between

To perform this test case run the file *TestCase\_fluxShifts.m* in the ‘Scripts’ folder of IgemRNA tool.

Most genome scale metabolic models use biomass objective function to simulate biomass accumulation rates, but in many cases such, *S. cerevisiae* Yeast\_8\_4 version models have specific wild type function. Optimizing different mutant strains, deletions and specific omics data integration (like transcriptomics data) on the models, can make infeasible optimization solutions although experimental conditions show the opposite. IgemRNA has functionality to remove biomass objective function from model and allow apply flux distribution with transcriptome and/or medium data sets and analyze results.

1. **Nomenclature of file names**

IgemRNA also provides standardized output file naming for easier filtering of analysis datasets (Table 1).

Table. 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Dataset and Phenotype Name** | **Thresholding Approach** | **Global Threshold Values** |
| Source | Sheet name in transcriptomics data file (Fig. 1. B) | Based on the selected thresholding approach in the IgemRNA form section B (Fig. 3.) | Based on the provided global threshold values in the IgemRNA form section C (Fig. 3.) |
| Example/Possible values | SRR8994357\_WT | 1. GT1 (Global T1) 2. LT1 (Local T1) 3. LT2 (Local T2) | 1. 30 (GT1, LT1) 2. 30\_100 (LT2) |