**Detailed description of our analysis pipeline**

To reproduce all results in this paper, run all scripts in the order in which they are listed in this file.

**Part A**

run: generating\_RNA\_RawCount.R

generates

- a\_results / rnaRawData.Rda

- a\_results / rnaMatrix\_meta.csv

- a\_results / rnaMatrix\_mRNA.csv

- a\_results / rnaMatrix\_RNA.csv

- a\_results / rnaMatrix\_rRNA.csv

- a\_results / rnaMatrix\_tRNA.csv

- a\_results / rnaMatrix.csv

- a\_figures / log2AmountDensity.pdf

run: generating\_Protein\_RawCount.R

generates

- a\_results / proteinMatrix\_w\_NA.csv

- a\_results / proteinMatrix\_wo\_NA.csv

- a\_results / proteinMatrix.csv

run: generating\_Meta\_data.R

generates

- a\_results / sampleSizeDf.csv

- a\_results / metaData.csv

- a\_results / metaProtein.csv

- a\_results / metaRNA.csv

- a\_results / metaRawData.Rda

run: data.normalization.R with parameters.R (RUNS 1 to 15) with given parameters

|  |  |  |  |
| --- | --- | --- | --- |
|  | All Data | Exp | Sta |
| mRNA | 1 | 2 | 3 |
| Protein | 4 | 5 | 6 |
| Int\_mRNA | 7 | 8 | 9 |
| Int\_protein | 10 | 11 | 12 |
| Int\_mrna\_protein | 13 | 14 | 15 |

With parameters

those runs generate

output of run No 1:

- a\_results / metaData\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_noNorm.csv

- a\_results / resDf\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_noNorm.csv

output of run No 2:

- a\_results / metaData\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_noNorm.csv

- a\_results / resDf\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_noNorm.csv

run: qualityControl\_mRNA.R

generates

- qualityControl\_mRNA\_heatMap.pdf

- qualityControl\_mRNA.pdf

run: qualityControl\_protein.R

generates

- a\_figures / qualityControl\_protein.pdf

- a\_figures / qualityControl\_protein\_heatMap.pdf

run: data.normalization.R with parameters.R (RUNS 3 and 4)

|  |  |  |  |
| --- | --- | --- | --- |
|  | **run No** | **3** | **4** |
|  | **saveFiles** | TRUE | TRUE |
|  | **runDeSeqForDifExp** | FALSE | FALSE |
| **DATA FILTERING PARAMETERS** | **dataType** | mrna | protein |
| **badDataSet** | set00 | set00 |
| **referenceParameters** | c("growthPhase","Mg\_mM\_Levels", "Na\_mM\_Levels", "carbonSource", "experiment") | c("growthPhase","Mg\_mM\_Levels", "Na\_mM\_Levels", "carbonSource", "experiment") |
| **referenceLevels** | c("exponential", "baseMg", "baseNa", "glucose", "glucose\_time\_course") | c("exponential", "baseMg", "baseNa", "glucose", "glucose\_time\_course") |
| **experimentVector** | allEx | allEx |
| **carbonSourceVector** | SYAN | SYAN |
| **MgLevelVector** | allMg | allMg |
| **NaLevelVector** | allNa | allNa |
| **growthPhaseVector** | allPhase | allPhase |
| **filterGenes** | c("noMatchFilter") | c("noMatchFilter") |
| **threshold** | NA | NA |
| **roundData** | TRUE | TRUE |
| **sumTechnicalReplicates** | TRUE | TRUE |
| **deSeqSfChoice** | p1Sf | p1Sf |
| **normalizationMethodChoice** | vst | vst |
| **DeSeq2 PARAMETERS** | **test\_for** | not relevant | not relevant |
| **test\_base** | not relevant | not relevant |
| **test\_contrast** | not relevant | not relevant |

output of run No 3:

- a\_results / metaData\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.csv

- a\_results / resDf\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.csv (TableS2 of main paper)

output of run No 4:

- a\_results / metaData\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.csv

- a\_results / resDf\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.csv (TableS3 of main paper)

**PART B**

run: heatMapFigureRNA.R

generates (num rows: 4196)

- b\_figures/ heatMap\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.png

- b\_figures/ mRNAHeatmap.pdf (fig03 of main paper)

- b\_results / treeData\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.RData

run: heatMapFigureProtein.R (fig04 of main paper)

generates (num rows: 4196)

- b\_figures/ heatMap\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.png

- b\_figures/ proteinHeatmap.pdf

- b\_results / treeData\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.RData

run: distance\_table\_mrna.R

uses

 - b\_results / treeData\_protein\_trT\_set00\_StcYtcNasAgrNgrMgh\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.RData

 - b\_results / treeData\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst.RData

generates

- b\_results / clustering\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst\_cophenetic.csv (table01 A of main paper & TableS4 of main paper)

run: distance\_table\_protein.R

generates

- b\_results / clustering\_mrna\_trT\_set00\_StcAllEx\_SYAN\_baseMgAllMg\_baseNaAllNa\_ExpAllPhase\_noMatchFilter\_p1Sf\_vst\_cophenetic.csv (table01 B of main paper & TableS5 of main paper)

**PART C**

run: data\_normalization\_DeSeq2\_multiple.R

This will make 24 different DeSeq2 runs

run data source Phase Test base\_condition test\_for control

1 mrna exp carbon source glucose glycerol batchNumber

2 mrna exp carbon source glucose gluconate batchNumber

3 mrna exp carbon source glucose lactate batchNumber

4 protein exp carbon source glucose glycerol batchNumber

5 protein exp carbon source glucose gluconate batchNumber

6 protein exp carbon source glucose lactate batchNumber

7 mrna sta carbon source glucose glycerol batchNumber

8 mrna sta carbon source glucose gluconate batchNumber

9 mrna sta carbon source glucose lactate batchNumber

10 protein sta carbon source glucose glycerol batchNumber

11 protein sta carbon source glucose gluconate batchNumber

12 protein sta carbon source glucose lactate batchNumber

13 mrna exp Mg\_mM\_Levels base Mg low Mg batchNumber

14 mrna exp Mg\_mM\_Levels base Mg high Mg batchNumber

15 protein exp Mg\_mM\_Levels base Mg low Mg batchNumber

16 protein exp Mg\_mM\_Levels base Mg high Mg batchNumber

17 mrna sta Mg\_mM\_Levels base Mg low Mg batchNumber

18 mrna sta Mg\_mM\_Levels base Mg high Mg batchNumber

19 protein sta Mg\_mM\_Levels base Mg low Mg batchNumber

20 protein sta Mg\_mM\_Levels base Mg high Mg batchNumber

21 mrna exp Na\_mM\_Levels base Na high Na batchNumber

22 protein exp Na\_mM\_Levels base Na high Na batchNumber

23 mrna sta Na\_mM\_Levels base Na high Na batchNumber

24 protein sta Na\_mM\_Levels base Na high Na batchNumber

each run will generate 4 different csv files under c\_results folder

1. c\_results \ resDf\_... .csv (results of DeSeq2)

2. c\_results \ metaData\_... .csv (associated meta data file)

3. c\_results \ genes\_P0.05Fold2\_... .csv (plain enriched gene names with p<0.05 and fold change >2)

4. c\_results \ ez\_... .csv (corresponding ez names for DAVID)

run: data\_normalization\_DeSeq2\_growthRate\_multiple.R

This will make 24 different DeSeq2 runs with additional growth rate control

run data source Phase Test base\_condition test\_for control

1 mrna exp carbon source glucose glycerol batchNumber + growthRate

2 mrna exp carbon source glucose gluconate batchNumber + growthRate

3 mrna exp carbon source glucose lactate batchNumber + growthRate

4 protein exp carbon source glucose glycerol batchNumber + growthRate

5 protein exp carbon source glucose gluconate batchNumber + growthRate

6 protein exp carbon source glucose lactate batchNumber + growthRate

7 mrna sta carbon source glucose glycerol batchNumber + growthRate

8 mrna sta carbon source glucose gluconate batchNumber + growthRate

9 mrna sta carbon source glucose lactate batchNumber + growthRate

10 protein sta carbon source glucose glycerol batchNumber + growthRate

11 protein sta carbon source glucose gluconate batchNumber + growthRate

12 protein sta carbon source glucose lactate batchNumber + growthRate

13 mrna exp Mg\_mM\_Levels base Mg low Mg batchNumber + growthRate

14 mrna exp Mg\_mM\_Levels base Mg high Mg batchNumber + growthRate

15 protein exp Mg\_mM\_Levels base Mg low Mg batchNumber + growthRate

16 protein exp Mg\_mM\_Levels base Mg high Mg batchNumber + growthRate

17 mrna sta Mg\_mM\_Levels base Mg low Mg batchNumber + growthRate

18 mrna sta Mg\_mM\_Levels base Mg high Mg batchNumber + growthRate

19 protein sta Mg\_mM\_Levels base Mg low Mg batchNumber + growthRate

20 protein sta Mg\_mM\_Levels base Mg high Mg batchNumber + growthRate

21 mrna exp Na\_mM\_Levels base Na high Na batchNumber + growthRate

22 protein exp Na\_mM\_Levels base Na high Na batchNumber + growthRate

23 mrna sta Na\_mM\_Levels base Na high Na batchNumber + growthRate

24 protein sta Na\_mM\_Levels base Na high Na batchNumber + growthRate

each run will generate 4 different csv files under c\_results folder

1. c\_results \ resDf\_... .csv (results of DeSeq2)

2. c\_results \ metaData\_... .csv (associated meta data file)

3. c\_results \ genes\_P0.05Fold2\_... .csv (plain enriched gene names with p<0.05 and fold change >2)

4. c\_results \ ez\_... .csv (corresponding ez names for DAVID)

\*\* the resDf\*.\* files generated by "data\_normalization\_DeSeq2\_multiple.R" are copied to c\_results \ DeSeq2\_diffGene\_batch\_Results\

\*\* the resDf\*.\* files generated by "data\_normalization\_DeSeq2\_growthRate\_multiple.R" are copied to c\_results \ DeSeq2\_diffGene\_batchGrowth\_Results\

run: difExpGene\_batch\_Figures.R

generate

- c\_figures \ exp\_mrna\_batch\_venn.jpeg

- c\_figures \ exp\_protein\_batch\_venn.jpeg

- c\_figures \ sta\_mrna\_batch\_venn.jpeg

- c\_figures \ sta\_protein\_batch\_venn.jpeg

- c\_figures \ venn\_batch.png

- c\_figures \ difExpressedGenesBatch\_mrna.pdf

- c\_figures \ difExpressedGenesBatch\_protein.pdf

- c\_figures \ difExpressedGenesBatch.pdf (fig05 of main paper)

- c\_figures \ venn\_batch.pdf (fig06 of main paper)

run: difExpGene\_batchGrowth\_Figures.R

generate

- c\_figures \ exp\_mrna\_batchGrowth\_venn.jpeg

- c\_figures \ exp\_protein\_batchGrowth\_venn.jpeg

- c\_figures \ sta\_mrna\_batchGrowth\_venn.jpeg

- c\_figures \ sta\_protein\_batchGrowth\_venn.jpeg

- c\_figures \ venn\_batchGrowth.png

- c\_figures \ difExpressedGenesBatchGrowth\_mrna.pdf

- c\_figures \ difExpressedGenesBatchGrowth\_protein.pdf

- c\_figures \ difExpressedGenesBatchGrowth.pdf

- c\_figures \ venn\_batchGrowth.pdf

run: combinedDataFrameForDeSeq2Results.R

generate

- c\_results \ combinedDifferentiallyExpressedGenes\_DeSeq.csv  (all significantly altered genes from all 48 tests) (TableS7 of main paper)

- c\_results \ combinedOutputDF\_DeSeq.csv (all genes from all 48 tests) (TableS6 of main paper)

run: change\_with\_growth\_Figure.R

generates

-c\_figures \ difference\_rtw\_GrowthControl.pdf (the figure shows the change between the results of DeSeq2 with controlled and not controlled growth)

the DAVIDWebInterfaces should run through command line and before running them java should be setup properly

The david web services need to locate location of local java library manually. So this segment of scripts might need to change in different computers

dyn.load('/Library/Java/JavaVirtualMachines/jdk1.8.0\_25.jdk/Contents/Home/jre/lib/server/libjvm.dylib')

run the R script

Rscript DAVIDWebInterface\_batch\_multiple.R

 and

Rscript DAVIDWebInterface\_batchGrowth\_multiple.R

these 2 scripts generate results from DAVID WEB SERVICES for each one of given 24 x 2 conditions they generate 3 results.

So total of (2 (batch, batch&growth) x  24 (individual test) x (3 figure for each test: kegg, mf\_n, mf\_o)) =  144 - >133 .csv files

If there is no significantly altered pathway/gene then there is no corresponding data frame

- 1. enhanced KEGG pathways (New version of DAVID)

- 2. enhanced GO annotations related with molecular function (New version of DAVID)

- 3. enhanced GO annotations related with molecular function (Old version of DAVID)

run: change\_with\_growth\_Figure.R

which compare the results of DeSeq2 with control of Batch effects with control of batch effects and growth rate. It generates a figure that represents the change between differentially expressed genes between the DeSeq2 runs and produces DAVID results by looking at some of those differences.

produce

- c\_figures \ difference\_rtw\_GrowthControl.pdf (figS34 of main paper)

- c\_results \ changed\_protein\_carbonSource\_ExpSta.csv (Proteins significantly altered under control of "batch + growth rate" but not under control of "batch" alone associated with carbon source changes for both exponential and stationary phase.) (Supplementary Table S9)

- c\_results \ changed\_DAVID\_P05.csv (Significantly altered KEGG pathways and molecular functions due to DAVID Web Service for proteins significantly altered under control of "batch + growth rate" but not under control of "batch" alone associated with carbon source changes for either exponential or stationary phases.) (Supplementary Table S10)

they represent the changed pathways and molecular functions between control for growth rate and not control for growth rate for carbon source and for proteins.

**PART D**

the part aims to generate figures related with pathways.

there are two single test run files -- One does NOT need to run those two files-- (One for KEGG pathways and the other is for GO Annotations related with Molecular Function)

1. DESeq\_David\_MF.R

2. DESeq\_David\_kegg.R

Then there are corresponding codes for generating multiple kegg and go annotation figures

1. DESeq\_David\_MF\_multiple.R  (figS02 - figS17 of main paper)

2. DESeq\_David\_kegg\_multiple.R (figS18 - figS33 of main paper)

Those codes generate

 - 2 figures for each condition, one is with title other is without

 - a corresponding data frame for each DAVID RUN that shows the significantly changing pathways and related conditions

 - and a combined figure

run: combinedDataFrameForDAVIDResults.R

combines all DAVID results in a single file

generates

-d\_results \ combinedResultList\_DAVID.csv (TableS8 of main paper)

run: summaryTableFigures.R

generates tables that show what are the mostly affected pathways under different conditions

SOME outputs

- d\_figures \ resultTable\_kegg.png (fig07 of main paper)

- d\_figures \ resultTable\_mf.png (figS01 of main paper)

run: combinedMainTextFigure.R

generate table figures for selected conditions to put inside main text

generates

- d\_figures \ combined.pdf (fig08 of main paper)

**PART E**

part aims to deal with flux data

run: FluxDoublingTimeSaltStressAnalyze.R

generates

- e\_results \ flux\_doublingTime\_fits\_sep.csv (TableS12 of main paper)

- e\_results \ flux\_doublingTime\_fits\_tog.csv (TableS13 of main paper)

- e\_figures \ Exp\_flux\_vs\_doub\_sep.pdf

- e\_figures \ Exp\_flux\_vs\_doub\_tog.pdf (fig09 of main paper)

- e\_figures \ GLYfromSER.pdf

run: FluxSaltStressAnalyze.R

generates

- e\_results \ flux\_data.csv (TableS11 of main paper)

- e\_results \ flux\_p\_values.csv

- e\_figures \ Exp.pdf (figS35 of main paper)

- e\_figures \ Sta.pdf

- e\_figures \ SaltStressPlots.pdf