**File Index**

**Stage 1– find database:**

Files:

How\_much\_data\_on\_each\_ortholog\_panther

Purpose:

How many ortholog of each type does this database have?

For each ortholog pair- how much data do we have for it?

**Stage 2– assemble database from FIT data:**

Files:

* wrapper\_calc\_comp\_files.r
* job\_calc\_comp\_files\_panther.r
* job\_calc\_compfiles\_RNAseq\_panther.r
* wrapper\_zscores\_FC\_allOrthologs.r
* job\_zscores\_FC\_allOrthologs.r
* merge\_all\_comp\_files.r

Purpose:

Create following table for each ortholog pair in the panther database.

Comments:

Database table is same as in FIT, but FIT database only has 1:1 orthologs in the database. Also the ortholog pairs in FIT are from the MGI database while mine are from Panther database.

**Stage 3– Correlation analysis:**

Files:

* correlations\_analysis.Rmd

Purpose:

Choose 1 pair of orthologs pair group that can represent the whole group. The pairs will be tested by their correlation coefficient.

Comments:

Outputs:

* Correlation Coefficients Between All Pairs In Each Ortholog Group. sorted by max(r), distance between 2 best orthologs.
* Table showing how many human genes have good candidates, for different thresholds.

**Stage 4– Lasso\_regression:**

Files:

* wrapper\_bootstraping\_lasso.r
* job\_bootstraping\_lasso.r

Purpose of this file:

for each ortholog couple perform lasso regression 100 times (bootstrapping).

**Stage 5– lasso regression on whole group:**

Files:

* wrapper\_lasso\_on\_whole\_group.r
* job\_lasso\_on\_whole\_group.R
* job\_merge\_betas\_trained\_as\_couples.r
* job\_merge\_betas\_trained\_as\_groups.r
* predictions\_lasso\_on\_whole\_groups.R

Purpose of this file:

For each ortholog group perform lasso regression considering all mouse genes together.

Comments:

Wrapper runs all 3 job files. Read comments to run them in time!

Predictions script also uses the lasso slopes calculated for each couple seperatly (for the weighted average model).