**Intensive integration using Web APIs**

A recent paper (DOI: 10.1371/journal.pone.0108567) executes a meta-analysis of a few thousand published co-expressed gene sets from Arabidopsis.  They break these co-expression sets into ~20 sub-networks of <200 genes each, that they find consistently co-expressed with one another.  Assume that you want to take the next step in their analysis, and see if there is already information linking these predicted sub-sets into known regulatory networks.  One step in this analysis would be to determine if the co-expressed genes are known to bind to one another.

Using the co-expressed gene list from Table S2 of the supplementary data from their analysis (I have extracted the data as text for you on the course Moodle → a list of AGI Locus Codes.  The link to the text file is below):

* use a combination of any or all of:  dbFetch, Togo REST API, EBI’s PSICQUIC REST API, DDBJ KEGG REST, and/or the Gene Ontology
* Find all protein-protein interaction networks that involve members of that gene list
* Determine which members of the gene list interact ***with each other***.

USE COMMON SENSE FILTERS IN YOUR CODE! (e.g. for species, and for quality!!!).

Note:  here is where you can get the current status of all PSICQUIC services:

<http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS>

This page tells you the base URL for each database's REST API.  Read the documentation for how to construct a PSICQUIC REST URL (link is here: https://psicquic.github.io/PsicquicSpec\_1\_3\_Rest.html)   I suggest that you use the BAR database from UToronto (it will return matches using AGI Locus Codes)

TASKS:

* Create an “InteractionNetwork” Object to contain the members of each network
* Annotate it with any KEGG Pathways the interaction network members are part of
  + both KEGG ID and Pathway Name
* Annotate it with the GO Terms associated with the total of all genes in the network
  + BUT ONLY FROM THE biological\_process part of the GO Ontology!
  + Both GO:ID and GO Term Name
* Create a report of which members of the gene list interact with one another, together with the KEGG/GO functional annotations of those interacting members.

BONUS MARKS (to get +2% up to a perfect score)

+1% if you create a ‘uso-general’ annotation object that can hold any functional annotation

+1% if you also annotate with the AraCyc pathways (see linked file on Moodle) --> The AraCyc bonus task is not available this year, because the download website is no longer available.

**Hints:  You need to make "smart" decisions...**

**- A interacts with B interacts with A creates an infinite loop.**

**- A interacts with B interacts with C interacts with D interacts with E interacts with F... how deep should you go?**

**- A interacts with B interacts with C interacts with A is \*a single network\***