Workflow for Consensus Modules

1. Formulate research question and prepare data
   1. Determine comparison groups (i.e. compare two viruses across all three days? Compare two viruses on one day?)
   2. Determine which data type (log ratio, expression value) will be used in analysis
   3. Convert everything to an ExprSet data type (Automate)
2. Determine soft power (Automate)

The adjacency matrix will be raised to the soft power to coerce the data into following a power law distribution. This makes biological “sense” since we would expect to see more non-hub genes than hub genes.

* 1. Takes in as input the data as determined from above
  2. Output will be a table of soft power, slope and R2 values and a graph of the softpower values with respect to the slope and R2

1. Calculate Adjacency/Connectivity Matrix (Automate)
2. Determine min connections (Automate)
3. Determine selected genes (Automate)
4. Determine modules
5. Determine consensus modules
6. Determine eigengenes for each module, for each subset
7. Determine genes in each module
   1. Modules with contrasting or similar expression patterns between the two subsets
8. Determine modules of interest
9. GoStat Analysis on genes/modules of interest(Automate)
   1. Are there GO terms that are over-represented in the modules of interest?
      1. Make sure this function is identifier independent



Task List:

1. Look at BIND and see if there is an easy way to curate interactions (Ted)
2. Look at Reactome/BioPAX format and see if there is an easy way to overlay results (Ted)
3. Internal format for all analyses should be Expression Set – look at agilent/Rosetta (Sophia/Armand)
4. Best strategy for mapping external Identifiers?
5. Automate consensus module workflows (Sophia)
6. Automate TFBS analysis (Sophia)