# MED283/BNFO286 Project

**Overview**

This project will require you apply network methods to genome wide association studies (GWAS). In groups of 2 or 3, create a network based method (either of your own design or based on a method discussed in class) to re-prioritize GWAS results from a large schizophrenia cohort. Your goal is to analyze the GWAS data to return a ranked list of disease-associated genes, with the goal of increasing the power of GWAS without significantly increasing the false discovery rate. All methods will be evaluated by how well they predict genes linked to disease in a second large cohort. This second dataset will be used by the instructors to rank all methods and will not be revealed to the students until well after the end of class.

**Existing methods**
- **Network propagation**
- **GWAB**
- **NetWAS**

**Data**

Training data consists of association statistics from a Schizophrenia GWAS of 36,989 cases and 113,075 controls produced by the Psychiatric Genomics Consortium (PMID:25056061). We are providing SNP-level and gene-level values of association with disease.

*SNP-level summary stats (hg18 coordinates):*
TritonED -> Content -> Project -> "SNP Level Summary Statistics"

*Gene-level summary stats:*
TritonED -> Content -> Project -> "Gene Level Summary Statistics"

We performed the SNP to Gene assignment, although you are welcome to use a custom function make this mapping and provide it in your code.  We took the smallest p-value within 10 kilobases of each gene body and assigned it to the gene. Unless you provide us with a different way of mapping the SNPs to genes, this is the mapping that will be used on the test data as well.  We will provide the original SNP summary statistics, the hg18 genome mapping, the python code that does the mapping, and the Gene-wise p-values for the training data set.

Although you are afforded complete flexibility with respect to the networks you use, we recommend the following NDEx network, which is a composite of several large global network compendia:
http://www.ndexbio.org/#/network/f93f402c-86d4-11e7-a10d-0ac135e8bacf

**Result evaluation**

You will be expected to provide a new ranked list of genes based your method. This list will then be compared to another GWAS result. The goal is to improve over simply ranking genes by the original GWAS results with the help of biological networks.

**Grading**
- 50%. Assignment completion: viable project turned in and code runs
- 20%. Approach: solid rationale and logic, correctness of implementation
- 20% Presentation clarity: Oral and written, including e-notebook
- 10%. Competitive performance in GWAS
- (+10%). Innovation and novelty

**Deliverables**
- Class presentation in **Week 10**
- Final report due on **June 15** (last day of finals week)
  - The report should include a brief section on each of the following:
    - Motivation of applying network methods to GWAS
    - Network methods and rationale
    - Results
    - Discussion
  - Link to GitHub code