MAJOR TO DO:

**Reviewer 1 -**

1. Extended comparison to alternative methods:

a. Compare to kNN, factor analysis (or N/A?).

b. Prediction accuracy vs. gene abundance (DONE).

2. Robustness wrt annotation:

a. Impact of false positive gene sets and/or missing markers?

b. Impact of program size, variance in training population, mean expression? (DONE)

c. Prediction accuracy of single genes vs. number of programs they’re in? (DONE)

3. Algorithmic details:

a. Run time?

b. Measure of confidence? (DONE)

4. Limitations in datasets with single-gene perturbations:

c. Can we detect single/small number of gene changes?

**Reviewer 3 -**

1. Rephrase, “reconstruct”

2. Accuracy vs. #-samples in training.

**Reviewer 4 -**

1. Robustness analysis of pathways like reviewer 1. Random pathway performance.

2. Report other measures of reconstruction performance. e.g. RMSE.

3. Clarify usage of intra-submission accuracy.

4. How good is the measured gene as a proxy for program expression? See reviewer 1 baselines.

**Reviewer 5 -**

None major.

**Expts to do**

**Expt 1: Increasingly random gene-sets -** Take gene sets, randomly replace X% of the members in each set. Look at avg PCC vs % randomly replaced. This answers 1.2.a and 4.1.

**Expt 2: Error analysis -** Predict prediction error (PCC and RMSE) as a function of gene/gene-set abundance, variance, number of gene-set’s they’re in. Answers 1.1.b, 1.2.b-c, and 4.2. (Analysis Done)

**Expt 3: Credible intervals -** Build credible intervals and calculate how often measured abundance falls within CI in a cross validation setting. For a 95% CI it should fall within 95% of the time (DONE).

**Expt 4: Accuracy vs training sample size -** Hold out most recent 10% of samples and see how performance improves with increasing number of submissions used during training.