1. General Information
   1. Project Title: Analysis of gene expression changes in response to EGF and TGF-b
   2. Project Start Date: September 23,2013
   3. Project leader:Bryan Linggi
   4. Data analysis leader:Bryan Linggi
   5. Location of project files: WE25473/Documents-1/R\_onAir/ANOVA\_LDRD
2. Pre-experiment Information
   1. Experiment type:

New Data Analyses

Public Data Analyses

* + 1. Gene Expression

Microarray

RNA-Seq

RT-PCR

Other:

* + 1. Epigenetic analysis

Methyl-Seq

ChIP

* + 1. Protein expression

Mass-spectrometry

Other:

* 1. Experiment Design
     1. Experimental Hypothesis: EGF and TGF-b synergize to regulate gene expression of some target genes and not others
     2. Cell Type(s):HMEC, MDA-MB-231
     3. Conditions:combination of egf, tgfb, with pathway inhibitors
        1. Controls

Positive:EGF alone on IL8 promoter

Expected result:activation

Negative:no ligand

Expected result:basal level expression

* 1. Power analysis
     + 1. Replicates needed for desired power:not calculated ahead of time
          1. Type of replicate:at least duplicate. Experiment repeated multiple times

1. Output result (check all required):
2. Powerpoint presentation of highlights
   * + 1. Excel table of raw data
3. Excel table of normalized data
4. Excel table with p-values and fold changes
   * + - 1. p-value cutoff:
         2. multiple hypothesis correction? y/n
         3. fold change cutoff:
5. Graph of data summary
   * + - 1. Details:
6. All analyses files (will be archived even if not returned to user)
7. VI. Preprocessing Steps:

(Mod from Faraway 2002)

Set up Git locally and on GitHub. Record details of Github locally):

Github address: https://github.com/bedward1/LDRD\_Anova.git

Data collection (export to 'diagnostics' folder)

how are data collected (random sampling?):

is there non-response?

are there missing values?

how are data coded?

how are qualitative variables represented?

Do I need to convert them to factors?

yes

no

if yes, describe steps

What are units of measurement?

Can I get rid of extraneous digits to save space?

yes

no

If yes, how is it done?

1. Initial data Analysis (export to 'diagnostics' folder, see location from )
   1. Print out summaries of data
   2. look for coding errors (NA converted to 0s)

b. Make box plots and examine variance

c. Make scatter plots for correlation analysis

d. check for missing values

use boxplot or summaries (See Farraway PRA pdf)

e. recode for all factors

f. perform anova

g. output to single file for different primers