

Midterm

- Find one bioinformatic-related paper with your teammate.
- Prepare a summary report that cover at least (1) description of data and biological goal (2) outline 2-4 specific aims and specific analysis plan to achieve these.

Here are some loose guidelines of how you can prepare the report.

1. Data sources:
 - (a) Where did you obtain the data? Are you able to obtain the raw data (e.g. CEL files, fastq or BAM files) or only processed data?
2. Review of the paper:
 - (a) Experimental design: Type of array or sequencing platforms? What kind of samples was applied? Specific condition, mutation or treatments? Biological replicates or experimental replicates?
 - (b) Data preprocessing: How were the data preprocessed? Any normalization and filtering performed?
 - (c) Data analysis: What analyses were performed in the paper?
 - (d) Conclusion: Describe the conclusions in the paper. Any further validation experiments performed? What is the implication and biological importance of the study?
3. Evaluation and re-analysis:
 - (a) Try your best to repeat the analytical procedures in the paper. If the description in the paper is not clear or it involves complicated methods not taught in class, try to use similar analyses.
 - (b) Comment on the pros and cons of their analyses and perform better alternatives based on what you have learned in this class. Do you obtain improved results?
 - (c) Any other further analysis? Report what you did and interpret your result and its implications.
 - (d) It is suggested to collect another array study of the same disease and perform some sort of meta-analysis with your data. If meta-analysis does not improve, discuss the potential reasons of heterogeneity.

Potential (but not limited) choices of analyses to be performed in your project to improve and compare with the paper:

- (a) Preprocessing: probe level analysis, normalization, missing value imputation, gene filtering.
- (b) Detect differentially expressed genes
- (c) Dimension reduction and visualization
- (d) Gene or sample clustering
- (e) Classification analysis
- (f) Enrichment analysis (pathway analysis)
- (g) Other advanced co-regulation analysis or network analysis
- (h) Omics meta-analysis or data integration

Some hints to find data sets:

1. Read through the paper and see if a web address is given to download the data.
2. Check microarray databases: NCBI Gene Expression Omnibus (GEO), Stanford Microarray Database, EBI's ArrayExpress, NCI's caArray, NCBI SRA etc.