

BB2899 2018 Log [Youcheng ZHANG]

Date	Content
Aug 29th	<ol style="list-style-type: none"> 1) To get a brief overview of research project from proposal planning to practical implement 2) To get involved more cutting edge research methodologies 3) To train the ability of handling unexpected situation as well as evaluating the risk to further accomplish the goals
Aug 30th	<ol style="list-style-type: none"> 1) Contacted project supervisor and booked planned meeting 2) Read paper: Mardinoglu, A., Boren, J., Smith, U., Uhlen, M., & Nielsen, J. (2018). Systems biology in hepatology: approaches and applications. <i>Nature Reviews Gastroenterology & Hepatology</i>, 1.
Sept 1st	<ol style="list-style-type: none"> 1) Read paper: Blais, E. M., Rawls, K. D., Dougherty, B. V., Li, Z. I., Kolling, G. L., Ye, P., ... & Papin, J. A. (2017). Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. <i>Nature communications</i>, 8, 14250. 2) Read paper: Mardinoglu, A., Boren, J., Smith, U., Uhlen, M., & Nielsen, J. (2018). Systems biology in hepatology: approaches and applications. <i>Nature Reviews Gastroenterology & Hepatology</i>, 1.
Sept 3rd, 2018	<ol style="list-style-type: none"> 1) Met the project supervisor and discussed the initial plan 2) Received reading tasks
Sept 4th	<ol style="list-style-type: none"> 1) Got basic knowledge about two datasets that will be used in the project: DrugMatrix and Open TG-Gates 2) Read paper: Sutherland, J. J., Jolly, R. A., Goldstein, K. M., & Stevens, J. L. (2016). Assessing concordance of drug-induced transcriptional response in rodent liver and cultured hepatocytes. <i>PLoS computational biology</i>, 12(3), e1004847.
Sept 5th	<ol style="list-style-type: none"> 1) Read paper about preprocessing Affymetrix and Codelink data: Patrick Aboyoun. (2009). Processing Affymetrix Expression Data. Fred Hutchinson Cancer Research Center AND Diez, D., Alvarez, R., & Dopazo, A. (2007). Codelink: an R package for analysis of GE healthcare gene expression bioarrays. <i>Bioinformatics</i>, 23(9), 1168-1169.
Sept 6th	<ol style="list-style-type: none"> 1) Met the supervisor and discussed the further plan
Sept 7th	<ol style="list-style-type: none"> 1) Discuss the initial plan with supervisor and demonstrate the research purpose, research subjects (heart tissue) as well as brief research procedure / pipeline 2) Finish and submit the initial plan: Initial plan Youcheng ZHANG.docx
Sept 9th	<ol style="list-style-type: none"> 1) Finish the Motivation part of the introductory presentation
Sept 10th, 2018	<ol style="list-style-type: none"> 1) Submit the introductory presentation slides
Sept 11th	<ol style="list-style-type: none"> 1) Read articles about DrugMatrix database: Ganter, B., Tugendreich, S., Pearson,

	<p>C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., ... & Breckenridge, N. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. <i>Journal of biotechnology</i>, 119(3), 219-244.</p> <p>2) Summarize the profiles of DrugMatrix into series-matrix-like format: tissue, drugs (compound), dose level, etc..</p> <p>3) Learn R programming to retrieve and preprocess the data on DrugMatrix</p>
Sept 12th	1) Read articles about DrugMatrix database:
Sept 13th	<p>1) Summarize and count the number of arrays, conditions, drugs, tissues for further determining the research subject on the early stage</p> <p>2) Clarify the research subject and condition: heart tissue, valproic acid, repeat dose, ctr/low/high dose level and across all time duration.</p>
Sept 14th	1) Discuss with the supervisor about the further experimental design: Comparison with control of single/repeated dose and control/high dose level to see the partial/whole effect on dose type
Sept 15th	<p>1) Preprocess meta data and count the sample replicates</p> <p>2) Modify the further study control to high dose versus control</p>
Sept 17th, 2018	<p>1) Discuss with the supervisor about the preprocess procedure/pipeline</p> <p>2) Set up weekly study plan</p>
Sept 18th	<p>1) Learn basic R programming</p> <p>2) Study preprocess methods and R packages: <i>codelinks</i></p> <p>3) Revise previous articles about the preprocess procedure</p>
Sept 19th	<p>1) Study preprocess methods and try different normalization approaches: quantile normalization, median normalization, etc.</p> <p>2) Try to write preprocess scripts that can automatically retrieve array expression data (probes) from GEO/ArrayExpress</p>
Sept 22nd	<p>1) Meet with the supervisor</p> <p>2) Skip preprocessing Codelink array data due to no appropriate processing pipeline except the program from the specific company which is not available</p> <p>3) Preprocess Affymetrix array data</p>
Sept 23rd	<p>1) Read online documents about Affymetrix data preprocessing</p> <p>2) Read articles: Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. <i>Nature communications</i>, 5, 3083.</p>
Sept 24th	1) Write python scripts to extract specific drugs with specific tissues, along with particular dose frequency and level: Obtain two metadata tables, one contains all the samples while the other one contains the extracted samples

Sept 25th	<ol style="list-style-type: none"> 1) Write python scripts 2) Read online documents: Introduction To Bioconductor Robert Gentleman Sandrine Dudoit Denise Scholtens Computational and Statistical Aspects of Microarray Analysis University of Milan May 26-30, 2003 3) Read online documents: Preprocessing Affymetrix Data Educational Materials ©2005 R. Irizarry and R. Gentleman Modified 21 November, 2009, M. Morgan
Sept 26th	<ol style="list-style-type: none"> 1) Learn R programming about basic syntax
Sept 28th	<ol style="list-style-type: none"> 1) Read articles about microarray data: Chapter 11 An Introduction to Microarray Data Analysis M. Madan Babu 2) Read online document: A Tutorial Review of Microarray Data Analysis Alex Sánchez and M. Carme Ruíz de Villa, July 7, 2008
Sept 30th	<ol style="list-style-type: none"> 1) Read online document about microarray data: A Tutorial Review of Microarray Data Analysis Alex Sánchez and M. Carme Ruíz de Villa, July 7, 2008
Oct 1st	<ol style="list-style-type: none"> 1) Submit assignment: Describe a project leader in your field
Oct 2nd	<ol style="list-style-type: none"> 1) Submit the 1st updated plan: Initial plan Youcheng ZHANG.docx
Oct 3rd	<ol style="list-style-type: none"> 1) Read article: Mardinoglu, A., Agren, R., Kampf, C., Asplund, A., Uhlen, M., & Nielsen, J. (2014). Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. <i>Nature communications</i>, 5, 3083. 1) Read article: Lee, S., Zhang, C., Liu, Z., Klevstig, M., Mukhopadhyay, B., Bergentall, M., ... & Deshmukh, S. (2017). Network analyses identify liver- specific targets for treating liver diseases. <i>Molecular systems biology</i>, 13(8), 938.
Oct 10th	<ol style="list-style-type: none"> 1) Meet with the supervisor for further discussion 2) Use built-in methods to preprocess data 3) Compare various built-in bg-correction, normalization, PM correction, summarization methods
Oct 11th	<ol style="list-style-type: none"> 1) Write custom methods for preprocessing data (MAD, 'huber') 2) Use custom methods in preprocessing
Oct 13th	<ol style="list-style-type: none"> 1) Establish normalization pipeline 2) Finish data preprocessing with R
Oct 15th	<ol style="list-style-type: none"> 1) Submit the report outline: Assignment_ReportOutline.docx
Oct 16th	<ol style="list-style-type: none"> 1) Perform the quality control on raw and preprocessed data and output the corresponding reports with outliers detections. 2) Annotate probeID with gene symbol from Rat db and remove non-annotated probeID as well as control probeID 3) Summarize and output an expression level table with columns (sample name) and rows (gene name)
Oct 25th	<ol style="list-style-type: none"> 1) Optimize the R scripts to easily retrieve and preprocess data from one

	tissue
Oct 27th	<ol style="list-style-type: none"> 1) Optimize the R scripts to automatically retrieve and preprocess data from all three tissues 2) Learn R package LIMMA (DE analysis) and TopGO (GSEA)
Oct 30th	<ol style="list-style-type: none"> 1) Read online articles: https://f1000research.com/articles/5-1384/v2 2) Learn R package reactomeRA (network analysis)
Nov 5th	<ol style="list-style-type: none"> 1) Meet with supervisors and fix the following plan 2) Run LIMMA, TopGO, Rx package for differential expression analysis, gene set enrichment analysis and pathway enrichment analysis on pre-test data
Nov 12th	<ol style="list-style-type: none"> 1) Modify scripts for retrieving each compound for preprocessing and selecting the significant genes with same logFC changing direction 2) Select the same responding direction genes in all three tissues (liver, kidney and hepatocyte) 3) Select the other genes as the background genes for gene set enrichment analysis
Nov 13th	<ol style="list-style-type: none"> 1) Discuss with the supervisor about gene set enrichment analysis 2) Learn GSEA R package piano
Nov 15h	<ol style="list-style-type: none"> 1) Output and save tables with logFC and adj.P.value on columns and genes on rows for each drug in three different tissues
Nov 18th	<ol style="list-style-type: none"> 1) Perform piano package for GSEA on one drug as pre-test 2) Find solutions for the situations that only few genes (less than 10) meet the criterion (both significant and same direction in three tissues) which raise error during performing gene set enrichment analysis
Nov 21st	<ol style="list-style-type: none"> 1) Meet with supervisor and discuss the GSEA problems 2) Modify the scripts and output the logFC and P value instead of adj.P.value for proper input using piano package
Nov 23rd	<ol style="list-style-type: none"> 1) Output two tables for differential expression genes and GO biological processes
Nov 25th	<ol style="list-style-type: none"> 1) Output two tables for differential expression genes and GO biological processes
Nov 27th	<ol style="list-style-type: none"> 1) Start writing report: Introduction part 2) Prepare software and programs for metabolic modeling on UPPMAX 3) Study MatLab, CPLEX, Raven
Nov 28th	<ol style="list-style-type: none"> 1) Optimize results from differential expression analysis and GSEA 2) Expand the analysis from 3 tissues to ≥ 2 tissues to increase the number

	of genes and GO findings
Nov 29th	<ol style="list-style-type: none"> 1) Write report: Methodology part 2) Install and pre-set the MatLab and Raven 3) Apply for license for CPLEX
Nov 30th	<ol style="list-style-type: none"> 1) Write report: Discussion part
Dec 1st	<ol style="list-style-type: none"> 1) Output another two detailed tables that contain information of differential genes and GO terms for each tissues 2) Output figures based on these two tables
Dec 2nd	<ol style="list-style-type: none"> 1) Learn R package for plotting: ggplot2
Dec 3rd	<ol style="list-style-type: none"> 1) Discuss with the supervisor for report details 2) Fix the bug in Gene set enrichment analysis step 3) Tune the parameters in GSEA such as the number of permutation
Dec 5th	<ol style="list-style-type: none"> 1) Set up the Github repository for storing all the raw data: https://github.com/YouchengZHANG/ProjectDrugRepurposing 2) Organize the Github structure
Dec 9th	<ol style="list-style-type: none"> 1) Write 1st version report 2) Modify Figures 3) Prepare for the final presentation
Dec 12th	<ol style="list-style-type: none"> 1) Double check working scripts 2) Investigate other selected compounds and compare the results in DrugBank database 3) Prepare for the final presentation
Dec 13th	<ol style="list-style-type: none"> 1) Final presentation 2) Further check the final scripts and reports
Dec 14th	<ol style="list-style-type: none"> 1) Submit the 1st version report 2) Discuss the further plan
Jan 2nd	<ol style="list-style-type: none"> 1) Revise the final report 2) Finish the DISCUSSION part
Jan 4th	<ol style="list-style-type: none"> 1) Double check the final report 2) Check the grammar and do proof-reading
Jan 5th	<ol style="list-style-type: none"> 1) Submit the final report 2) All the scripts and results are uploaded to Github: https://github.com/YouchengZHANG/ProjectDrugRepurposing