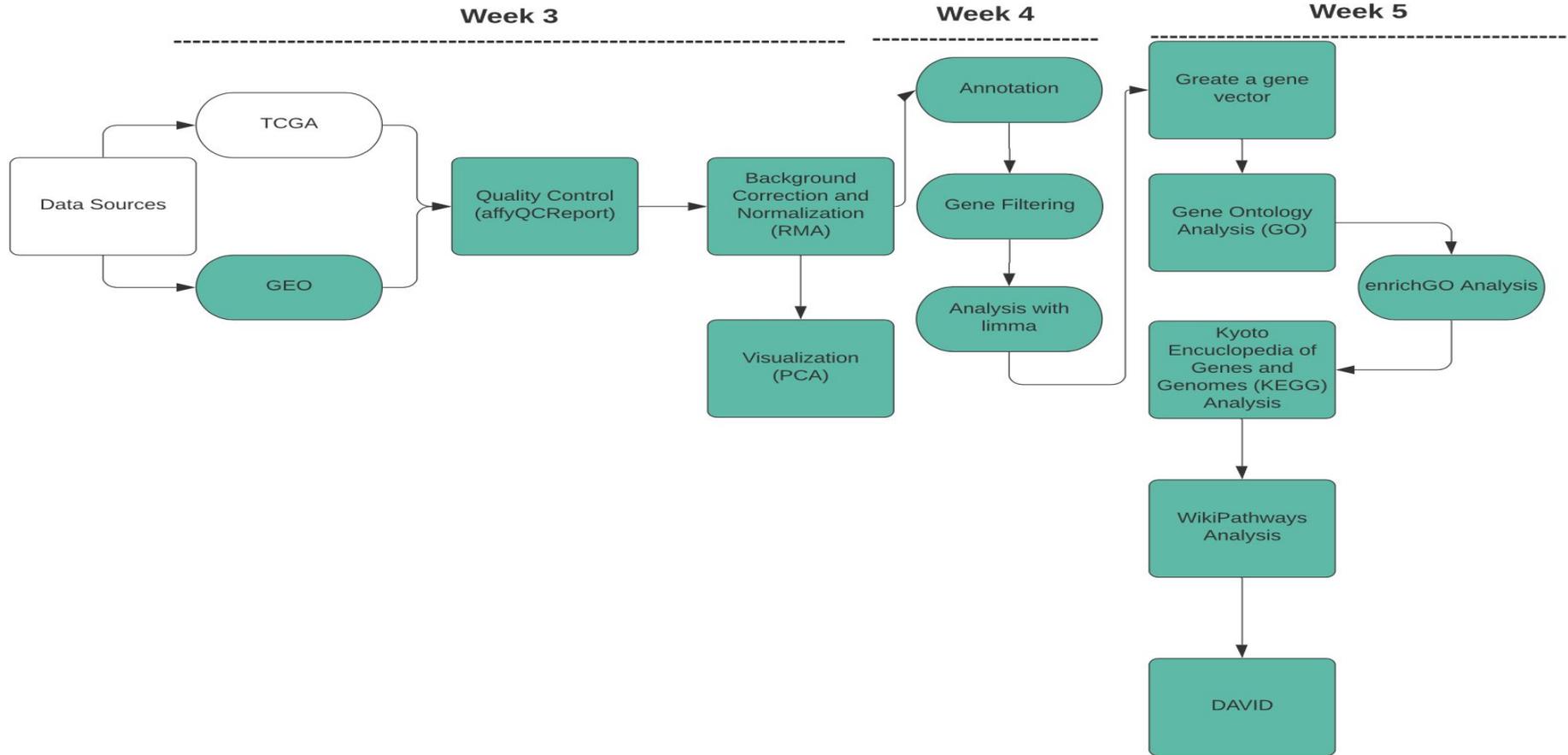


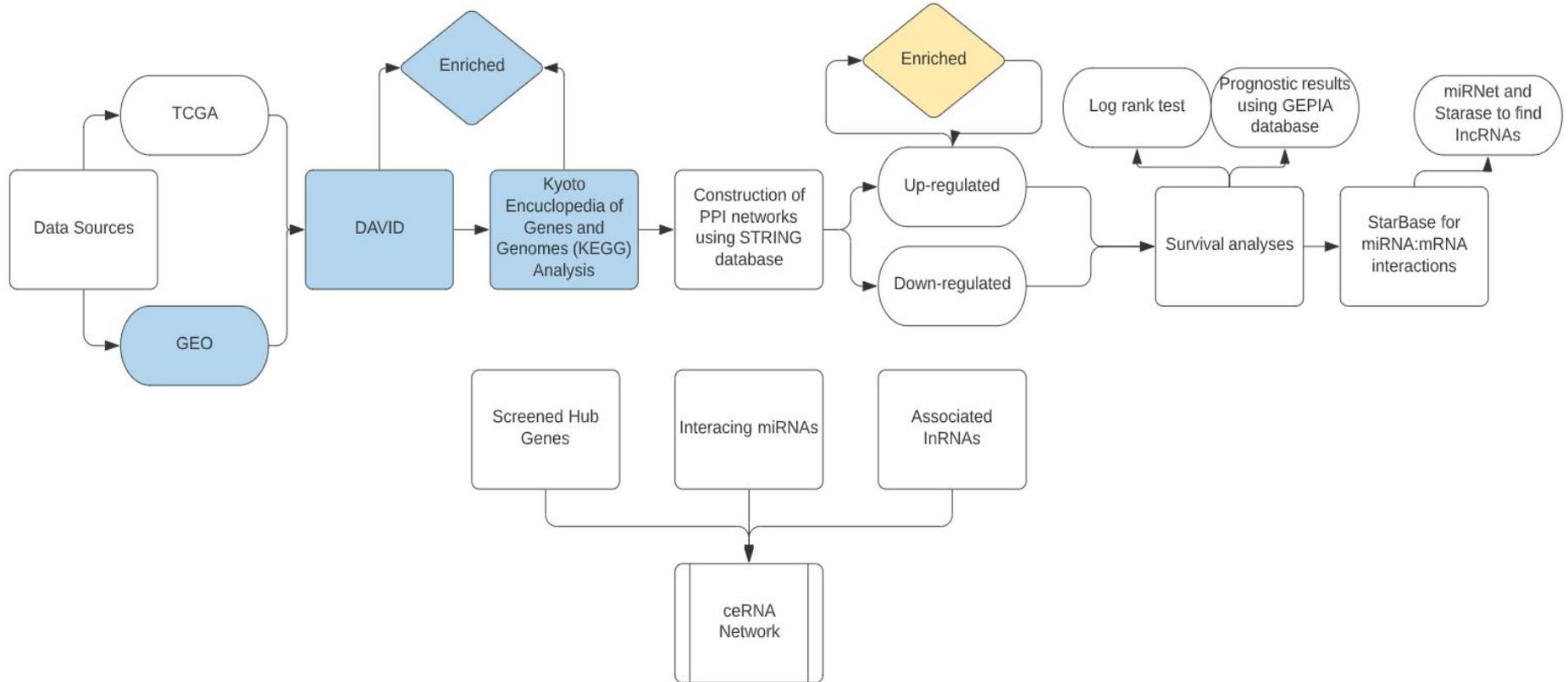
Stem-Away Bioinformatics Internship

By Cesar Juarez

Pipeline of the project



Pipeline of the Paper



What I learned from the pipeline and what is new?

Things I learned

- Quality Control: affyQCReport
- Background Correction and Normalization: rma
- Visualization: PCA plot
- Annotations: acquire gene symbols
- Gene filtering: according to qc and preprocessing
- Analyse with limma: Gene expression differential analysis
- Phenotypic data analysis
- Gene Ontology Analysis (GO): GroupGO(), enrichGO()
- Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis
- WikiPathways analysis
- David
- Network Analysis: STRING

What is new?

- TCGA
- PPI networks
- Survival Analysis
- GEPIA and StarBase databases
- MiRNet
- Wilcoxon rank-sum test, a Kruskal-Wallis test, and a paired t-test was used to estimate a potential difference between groups
- Creating a ceRNA network
- DESeq2

What is missing from the paper's pipeline?

- Quality Control
- Normalization Method
- Patient Results

Figure A

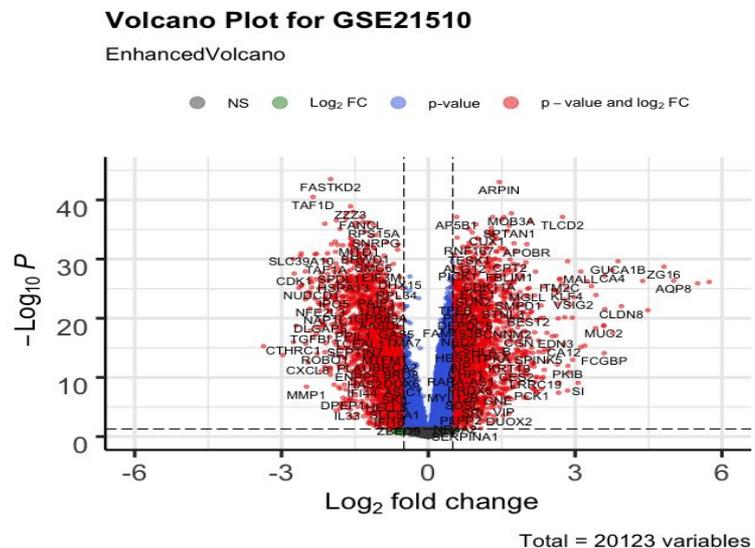


Figure A: The volcano plot shows the distribution of differentially expressed genes using the GSE 21510 dataset.

Figure B

	logFC	AveExpr	t	P.Value	adj.P.Val	B
FASTKD2	-1.9968625	5.739211	-22.02183	1.442517e-48	2.902777e-44	99.95188
ARPIN	1.4502284	6.405963	21.65729	9.548655e-48	9.607380e-44	98.08541
TAF1D	-2.3564843	9.646060	-20.48629	4.633028e-45	3.107681e-41	91.97332
ZZZ3	-1.5888806	9.266400	-19.75414	2.417636e-43	1.216252e-39	88.06155
U2SURP	-1.5401963	10.131803	-19.32824	2.488581e-42	1.001554e-38	85.75427
MOB3A	1.6974121	6.994446	19.19194	5.273355e-42	1.768595e-38	85.01095
TRMT13	-1.6792408	9.088021	-19.15719	6.388350e-42	1.836468e-38	84.82108
AP5B1	0.5797659	4.422189	18.86889	3.155212e-41	7.571469e-38	83.23985
TLCD2	2.7413769	5.844707	18.85455	3.416924e-41	7.571469e-38	83.16095
FANCL	-1.3703037	10.023446	-18.83722	3.762594e-41	7.571469e-38	83.06554

Figure B: The top 10 differentially expressed genes of GSE21510

Figure C

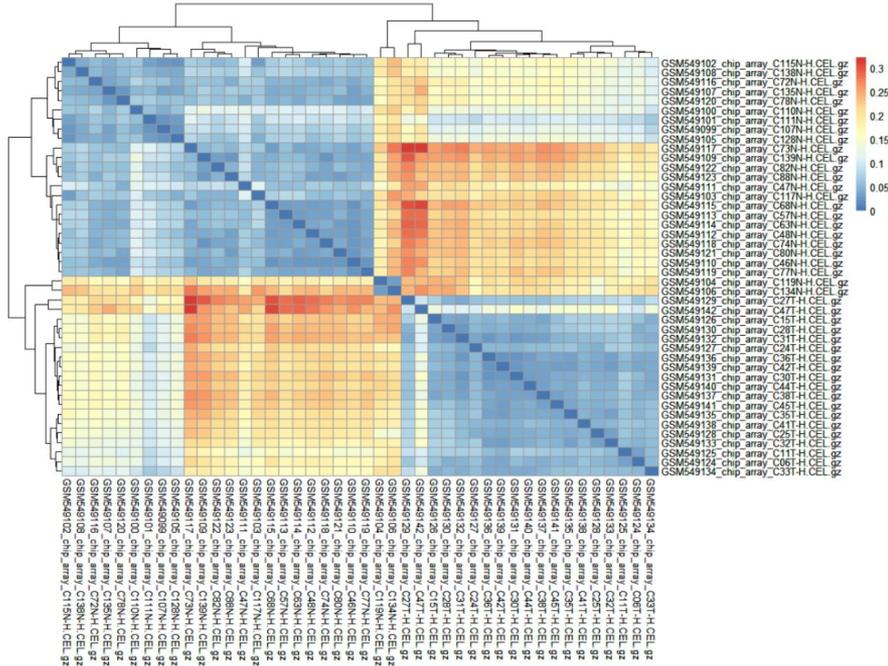
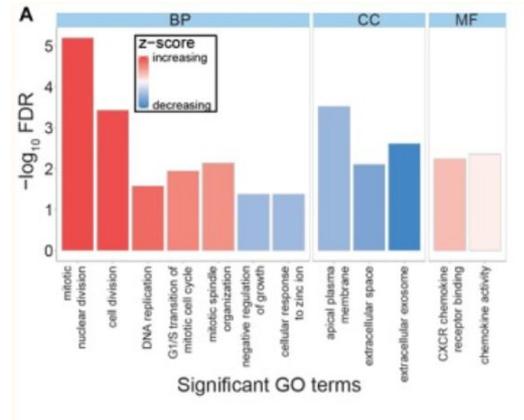


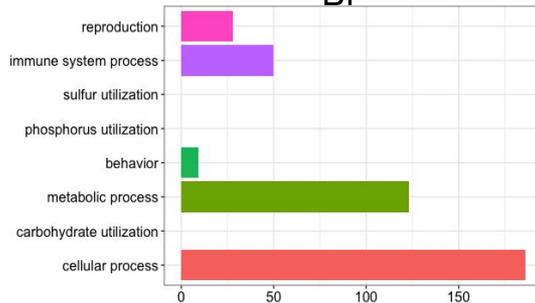
Figure C: The heatmap shows the detailed expression patterns of these genes in tumor and normal samples

Gene Ontology (GO) Analysis

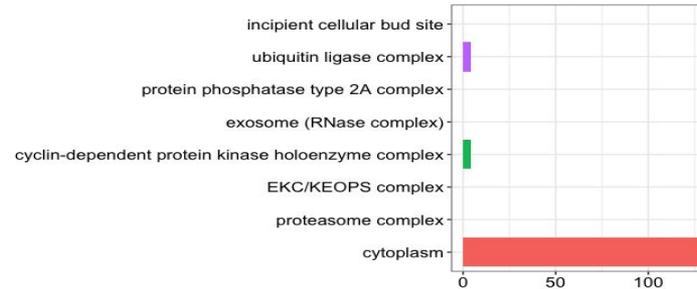


A) The distribution of significantly enriched Gene Ontology (GO) terms. BP, biological function; CC, cellular component; MF, molecular function

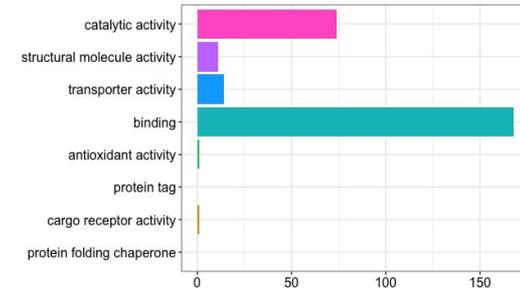
GSE21510
BP



GSE21510
CC



GSE21510
MF



What I learned from this project

Technical Skills:

- STEM-AWAY forums
- Asana
- GitHub
- GEO database website
- How to read a scientific paper efficiently
- R and RStudio (loading packages, coding, making graphs and plots, QC, Normalization (RMA), Visualization (PCA), Annotation, Gene Filtering, Limma, gene vectors, Gene Ontology Analysis (GO), Kyoto Encyclopedia Genes of Genomes (KEGG) Analysis, Enrich, WikiPathways Analysis, DAVID)

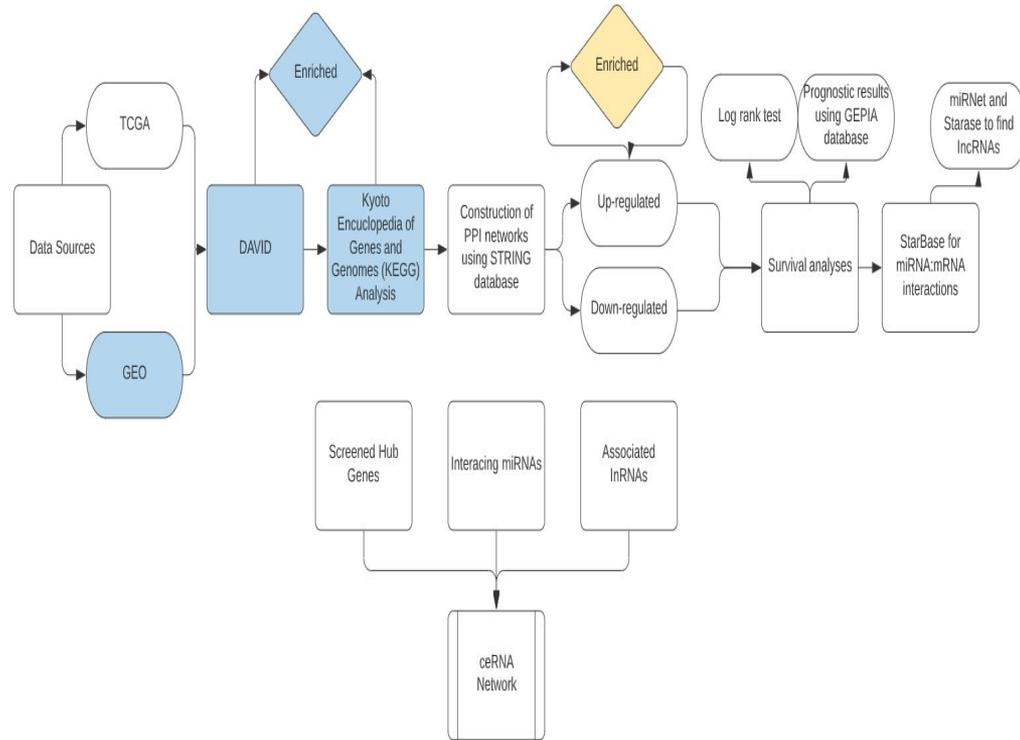
Soft Skills:

- Communication with team and subteams
- Time-Management
- Organization
- Team Work

Challenges

- Learning how to use R and RStudio
- Understanding the deliverables and the project pipeline
- Meeting with subteams since we all had different time zones
- Being able to balance work with the internship

Future Projects



- Survival Analysis
- Starbase, miRNet databases
- Creating a ceRNA network
- Work with a different type of cancer or disease
- Analyze data with DESeq2

3 achievement highlights

1. Working with a dataset in R and being able to make plots, graphs and tables.
2. Finished the deliverables on time.
3. Doing a self-assessment weekly so I understand what I can do to improve, have goals, and remembering the things that I went through for that week.

One thing that you learned in this session that has made you more excited about the field of Bioinformatics

One of the things that I learned from this project is how to use R and Rstudio. Since I am currently working at an immuno-oncology company, knowing how to use the programs will help me with analyzing data related to the company's research.

Thank you
