Genomics and bioinformatics at Georgia Tech

Comparative analysis from gene expression revealed the function of ZC3H18 and CDK12 in ovarian cancer

NOV 4

Posted by **xhang7gatechedu**

Group Member: Shen-Yi Cheng, Shuting Lin, Xin Hang, Winnie Zheng, Shuheng Gan, Jie Zhou

Introduction

In today's United States, cancer already becomes the second leading cause of death. Based on the research conducted by the Centers for Disease Control and Prevention (CDC), 25% of deaths in the United States is because of cancer. Among those different types of cancers, ovarian cancer has a relatively high death rate in women [1]. Therefore, conducting some researches about ovarian cancer is very significant and important for society. Previous studies reported that two breast cancer genes (BRCA1 and BRCA2) account for 10% ~ 15% of all ovarian cancers [2]. However, gene regulation in ovarian cancer is still unclear. In our research, we chose two RNA-seq datasets of ovarian cancer cells with separately one gene was silenced. These datasets are available with accession numbers GSE136533 and GSE138288 in the GEO database. Using the siRNA transfection method, the target gene ZC3H18 was depleted in the former dataset (GSE136533) and CDK12 in the later dataset (GSE138288). Each dataset includes 3 samples of human ovarian cancer cell line OVCAR-8, one of which is the control sample with luciferase siRNA transfection and the rest are two independent treated samples with target gene depleted. Each sample contains three technical replicates. The sequencing platform is Illumina Hiseq 2500. We plan to compare the gene expression profiles of two datasets and find their unique and common gene clusters and pathways. The results will help understand what roles ZC3H18 and CDK12 gene are playing in the ovarian cancer cell.

According to the previous researches, we can find that the main function of CDK12 is to regulate mRNA splicing, 3′ end processing, pre-replication complex assembly, and genomic stability during embryonic development [3]. This work revealed that the inhibition of CDK12 already becomes an effective method to inhibit tumor growth and in ovarian cancer, the genomics alterations in CDK12 have been found [3]. However, the main function of ZC3H18 on cancer cells has not been detected yet. Therefore, in our research, we will perform the comparison between two datasets and explore the possible function of ZC3H18 based on the known function of the CDK12 gene in the ovarian cancer cells.

Data overview：

GSE136533 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE136533)

GSE138288 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138288)

Methods

workflow

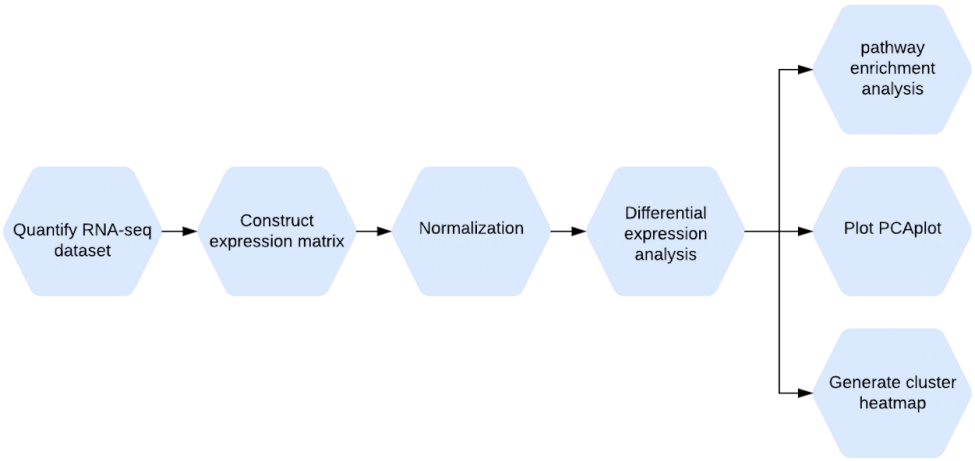


figure 1. Workflow of the project

Qualification

We utilized salmon to quantify the expression of transcripts for our RNA-seq datasets, GSE136533 and GSE138288. The reference transcriptome we used to help this analysis was Homo sapiens.GRCh38 gained from Ensembl.

Normalization & Expression matrix construction

For the two datasets, we calculated normalization factors to align samples of the gene expression matrix, and then, we estimated common negative binomial dispersion by conditional maximum likelihood. In the following analysis, we estimated tagwise dispersion values by an empirical Bayes method based on weighted conditional maximum likelihood. After that, we exacted tests for differences between two groups of negative-binomial counts. To control the false discovery rates (FDR), we applied the Benjamini-Hochberg method on the p-values. After that, we generated a list of significantly differentially expressed genes with FDR adjusted p-values of <0.01(p-values cutoff by 0.01) and absolute log2-fold changes(logFC) between each pair of samples > 2(absolute logFC cutoff by 2).

Heatmap & PCA plot generation

We plot the clustering heatmap by the distance matrix from normalization results to show the similarities and dissimilarities between samples. Then we plot PCA plot based on the result from which can be used to figure out what produces the differences among clusters.

Differential expression analysis

For the qualification results we obtained from salmon, we used edgeR to create differential expressed genes then used biomaRt to convert the multiple Ensembl IDs into respective gene symbols.

Pathway enrichment analysis

We used the DAVID to analyze the differentially expressed genes by KEGG pathway maps.

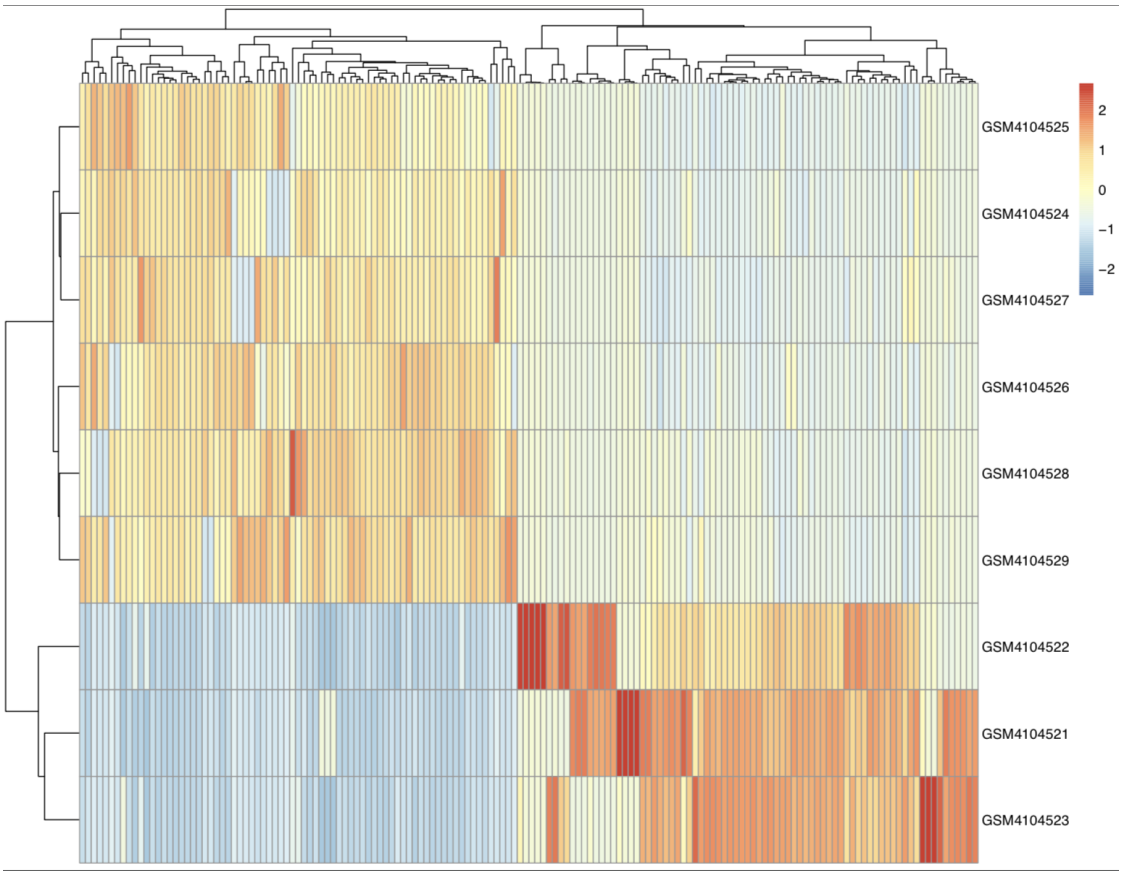
Results

Differential expression results

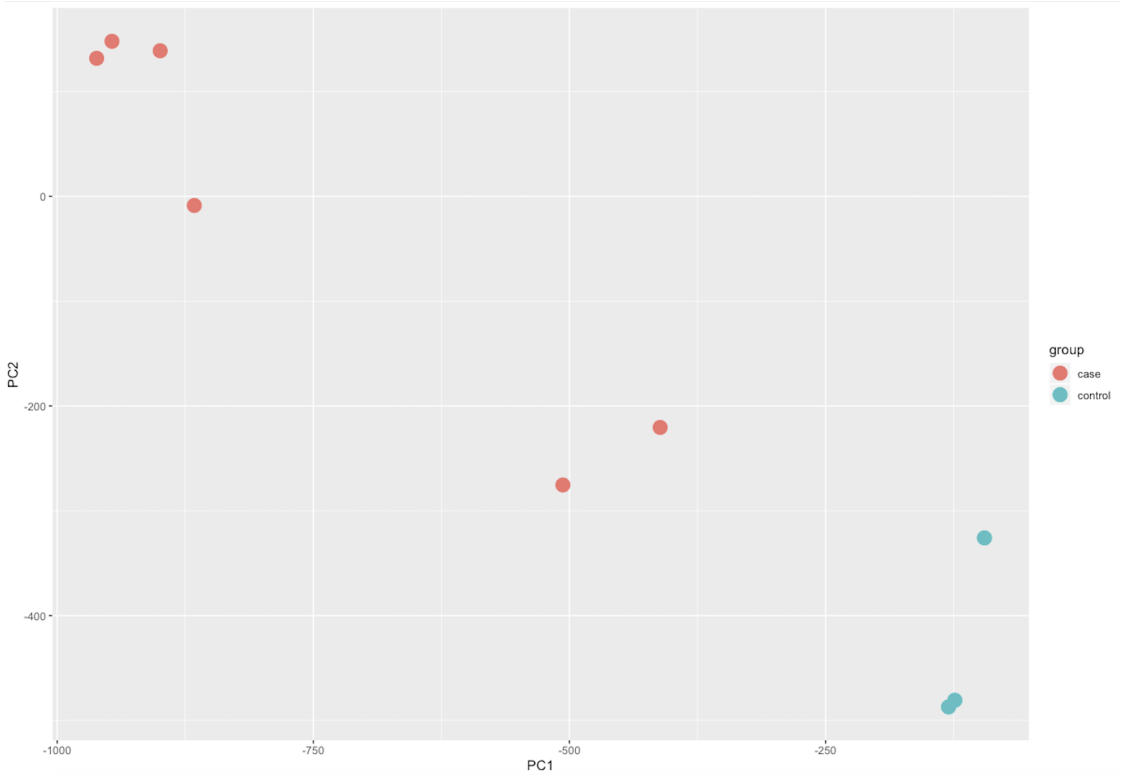
For GSE136553, we obtained 471 differentially expressed genes, but 6 genes cannot find their gene symbols. For GSE138288, the same method was applied and 150 genes left after deleting 4 genes without gene symbols. Detailed differential expression results are shown in the table below.

	Total genes	Up-regulated genes	Down-regulated genes
GSE136533	465	169	296
GSE138288	150	74	76

Clustering & pathway analysis of GSE138288



We plot this clustering heatmap by the distance matrix from differential expression analysis results to show the dissimilarities between treatement and control samples. In this figure, the red part means the higher gene expression, and the blue part means the lower gene expression. It is clear that the samples are clustered together by control and treatment separately.



Based on the result from PCA analysis, we subtract the PC1 and PC2 to plot the PCA plot. This plot can be used to figure out what produces the differences among clusters. It is clear that, relatively, the case groups cluster with each other and the control groups cluster with each other and they are mainly separated by the PC1.

Functional Annotation Chart

Help and Manual

Current Gene List: List_2
Current Background: Homo sapiens
368 DAVID IDs
Options

Rerun Using Options | Create Sublist

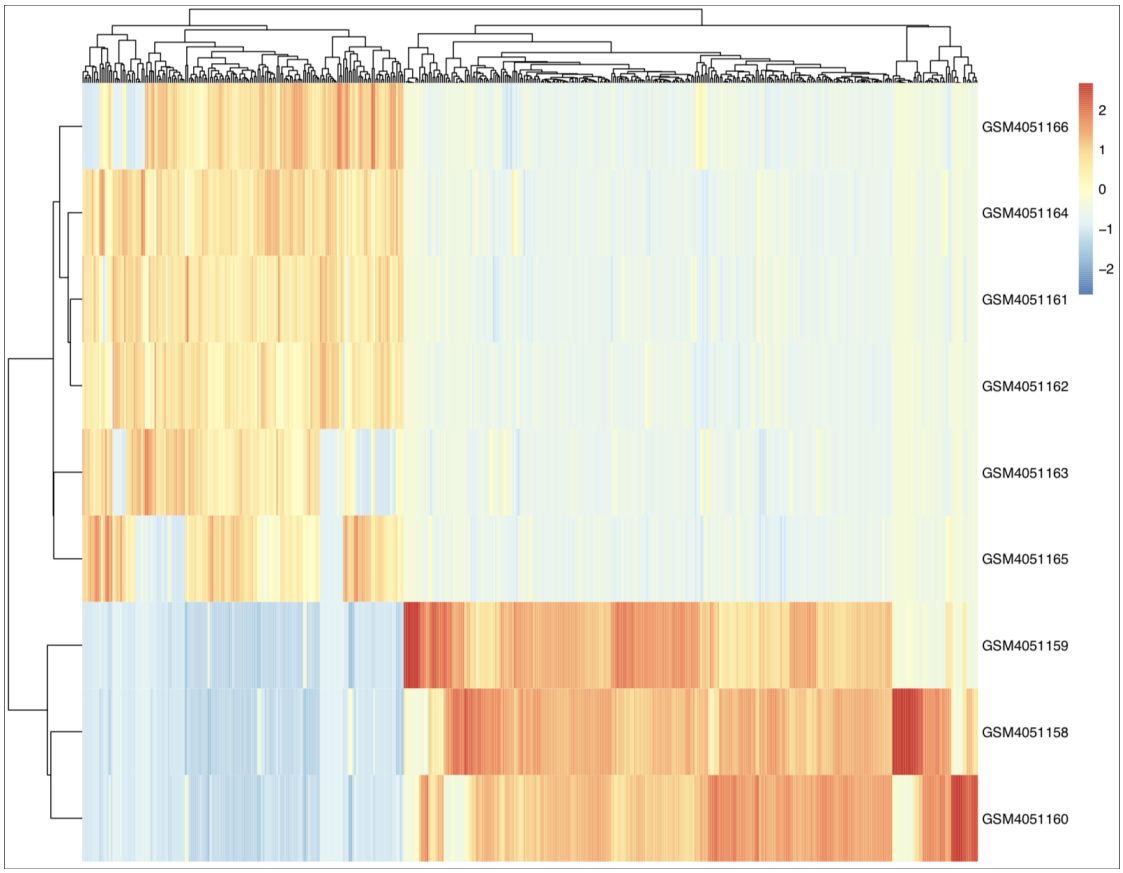
9 chart records Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	KEGG_PATHWAY	Ribosome	RT		10	2.7	2.0E-3	3.4E-1
	KEGG_PATHWAY	Endocytosis	RT		12	3.3	1.1E-2	6.9E-1
	KEGG_PATHWAY	Metabolic pathways	RT		36	9.8	2.5E-2	8.2E-1
	KEGG_PATHWAY	Peroxisome	RT		6	1.6	2.9E-2	7.7E-1
	KEGG_PATHWAY	p53 signaling pathway	RT		5	1.4	5.0E-2	8.8E-1
	KEGG_PATHWAY	Viral carcinogenesis	RT		9	2.4	6.3E-2	8.9E-1
	KEGG_PATHWAY	Insulin signaling pathway	RT		7	1.9	6.7E-2	8.7E-1
	KEGG_PATHWAY	Arginine and proline metabolism	RT		4	1.1	8.5E-2	9.0E-1
	KEGG_PATHWAY	Fanconi anemia pathway	RT		4	1.1	9.7E-2	9.0E-1

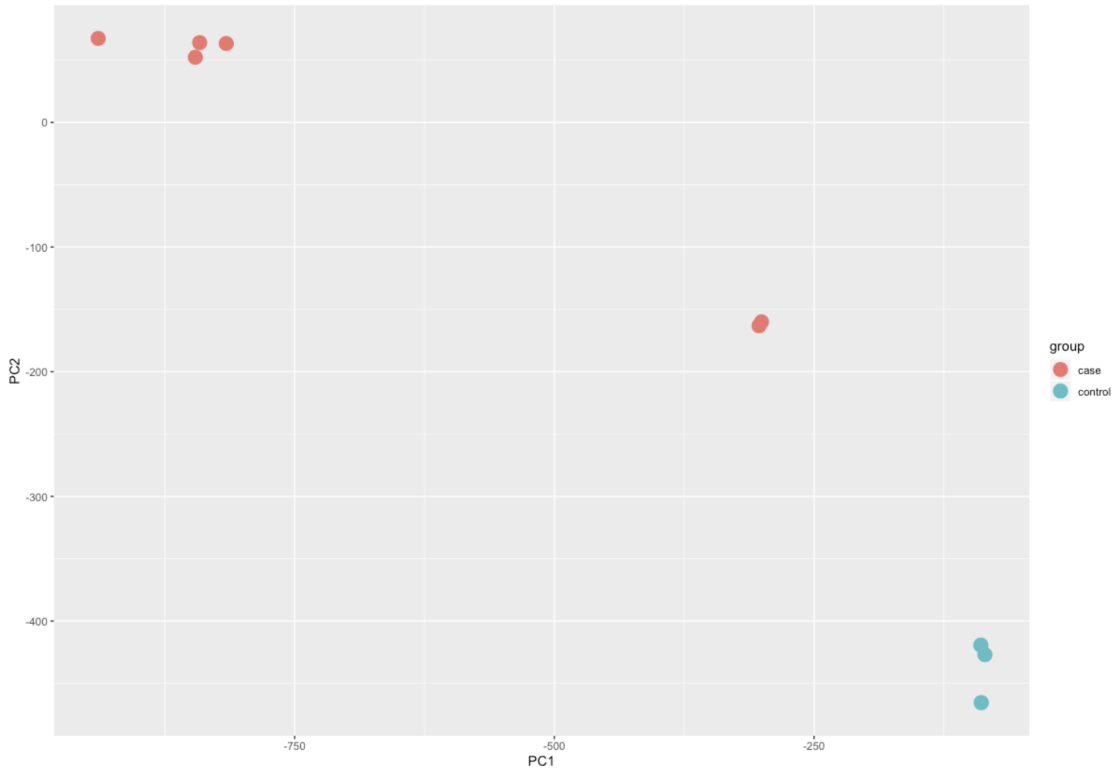
Figure 4. Pathway analysis of GSE138288

In GSE138288 datasets, we obtain 9 pathways. They are ribosome, endocytosis, metabolic pathways, peroxisome, p53 signaling pathway, viral carcinogenesis, insulin signaling pathway, arginine and proline metabolism, Fanconi anemia pathway. Among these different pathways, the p53 signaling pathway plays an important role in the inhibition of angiogenesis which performs as an anti-cancer promotion agent.

Clustering & pathway analysis of GSE136533



We plot this clustering heatmap by the distance matrix from differential expression analysis results to show the dissimilarities between control and treatement samples. In this figure, the red part means the higher gene expression, and the blue part means the lower gene expression. It is clear that the samples are clustered together by control and treatement separatly.



Based on the distance matrix from salmon results, we subtract the PC1 and PC2 to plot the PCA plot. This plot can be used to figure out what produces the differences among clusters. It is clear that, relatively, the case groups cluster with each other and the control groups cluster with each other and they are mainly separated by the PC1. Comparing with GSE138288, GSE136533's control groups are relatively more separated by PC2.

Functional Annotation Chart

Help and Manual

Current Gene List: List_1
Current Background: Homo sapiens
110 DAVID IDs
Options

Rerun Using Options | Create Sublist

3 chart records

Download File

Sublist	Category	Term	RT	Genes	Count	%	P-Value	Benjamini
	KEGG_PATHWAY	Viral carcinogenesis	RT		6	5.5	6.3E-3	5.8E-1
	KEGG_PATHWAY	Epstein-Barr virus infection	RT		4	3.6	3.3E-2	9.0E-1
	KEGG_PATHWAY	Peroxisome	RT		3	2.7	8.4E-2	9.8E-1

Figure 7. Pathway analysis of GSE136533

In GSE136533 datasets, we obtain 3 pathways. They are Viral carcinogenesis, Epstein-Barr virus infection, and peroxisome.

Disscussion

In the previous study, CDK12 has already varified to play a positive role in inhibiting tumor growth and the recent published material has revealed that *ZC3H18*, a gene located in a chromosomal region frequently deleted in HGSOC, is essential for HR (Homologous recombination), which is a high-fidelity DNA repair mechanism that requires the sequential activities of a series of proteins, including BRCA1 and BRCA2 tumor suppressors, and that ZC3H18 depletion sensitizes ovarian cancer cells to platinum agents and PARP inhibitors.

Through series analysis, we got the differential expressed genes and conducted pathway enrichment analysis, PCA analysis, generated clustering heatmap. The results were good for each of the datasets. The control group and treatment group were separated during our project. The pathway analysis shows that both two genes are related to the tumor which corresponds to the previous research.

Reference

[1] USCS Data Visualizations – CDC. (2016). Retrieved October 10, 2019, from https://gis.cdc.gov/Cancer/USCS/DataViz.html.https://gis.cdc.gov/Cancer/USCS/DataViz.htm (https://gis.cdc.gov/Cancer/USCS/DataViz.html.https://gis.cdc.gov/Cancer/USCS/DataViz.html).

[2] https://ocrahope.org/patients/about-ovarian-cancer/risk-factors/ (https://ocrahope.org/patients/about-ovarian-cancer/risk-factors/)

[3] Lui GYL, Grandori C, Kemp CJ. CDK12: an emerging therapeutic target for cancer. Journal of clinical pathology. 2018; 71:957–62.

[4] Kanakkanthara, Arun, et al. "ZC3H18 specifically binds and activates the BRCA1 promoter to facilitate homologous recombination in ovarian cancer." *Nature communications* 10.1 (2019): 1-15.

[5] Joshi, Poorval M., et al. "Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly (ADP-ribose) polymerase inhibitors." *Journal of Biological Chemistry* 289.13 (2014): 9247-9253.

[6] Marshall, Catherine H., et al. "CDK12 inactivation across solid tumors: an actionable genetic subtype." *Oncoscience* 6.5-6 (2019): 312.

[7] De Picciotto, Nicolas, et al. "Ovarian cancer: status of homologous recombination pathway as a predictor of drug response." *Critical reviews in oncology/hematology* 101 (2016): 50-59.

Posted on November 4, 2019, in Uncategorized. Bookmark the permalink. Leave a comment.

- **Leave a comment**

- ## Comments 0