

QR_liuzhe.docx

by Zhe LIU

Submission date: 12-May-2021 07:15PM (UTC+0800)

Submission ID: 1357415442

File name: QR_liuzhe.docx (2.9M)

Word count: 5151

Character count: 44101

A exam Research Proposal
April 30, 2021, 10 AM, 458 ST Olin

Zhe LIU (刘哲)

Graduate Student

Department of Biomedical Sciences

Committee Chair:

Committee Members:

1. Title

Apply bioinformatics technologies to study the response of stretching stimuli to skin cells and explain the pathogenesis of complex diseases

Subtitle 1: Molecular dissection of the skin cellular response to stretch

Subtitle 2: The comprehensive identification of AS events that RBPs regulated in the relapse of glioma

Subtitle 3: The MEGENA co-expression network analysis reveals the key CSC regulators of HCC

A. Molecular dissection of the cellular response to stretch.

1. Background and Significance

As the largest organ of mammals, the skin is a mechanical barrier between the body and the outside world. It has a complex tissue structure and multiple physiological functions. The skin is divided into two layers: the epidermis and the dermis. The epidermis is located on the top of the dermis and consists of stratified flat epithelium, thin and contains no blood vessels. The basal layer to the surface can be divided into five layers: the basal layer, spinous layer, granular layer, transparent layer, and stratum corneum. Although researchers have explored some mechanisms of skin to stress, there is no skin response at the molecular level research in complex model organisms *in vivo* at different period of times. There is also no comprehensive description of cell sub-populations different responses to the same mechanism stretch. The stretching can regulate the fate of many different kinds of cell types. For example, shear stress, axial stress, and synergy can affect vascular endothelial cells' stress response through the Ca²⁺ signal pathway. It is known that the Ca²⁺ transfer pathways of the endothelial cell plasma membrane and the Ca²⁺ pool include voltage-gated Ca²⁺ channels, agonist-receptor-gated Ca²⁺ channels, stretch-activated Ca²⁺ channels (Ca²⁺ channels for short) permeable Ca²⁺ channels, Ca²⁺ pumps, and Na⁺-Ca²⁺ converter and so on. The effect of shear stress on the Ca²⁺ signal in the inner wall of the cell is achieved by directly or indirectly affecting some of the Ca²⁺ channels through the shear stress so that Ca²⁺ is released from the extracellular or Ca²⁺ pool to the cytosol, thereby changing the Ca²⁺ concentration. And there is another research about the short-term influence of stem cells (SCs). Biologists study the transient transition of stem cell fate to renewal allows the basal cell population to expand while maintaining differentiation. They conclude that the regulators of the actomyosin cytoskeleton, including formin-like proteins and non-muscle myosin, are essential for skin stretch sensing in the body and are typically mechanically transduced in YAP1 and MAL. The upstream of the child works. It is worth noting that the same signalling pathways are activated during embryonic pancreatic development, indicating that these signalling pathways and transcriptional regulators play a conservative role in the animal's mechanical transduction, embryonic development and adult tissue regeneration. They conclude that short-term mechanical

stretching can induce SC proliferation by activating YAP's main transcription factor, and YAP is a key regulator of epidermal proliferation.

Mining the key TF and regulatory relationship in skin cell are very important for clinical cell regeneration.

Innovation: The importance of skin's response to stretch has been explained by plastic surgery and biologists. However, little research is performed *in vivo*. Cell response to stress is crucial for many cellular functions, yet its molecular mechanisms are not yet fully understood. Previous studies of the cellular stress response were performed on single cultured cell or isolated muscle fibers devoid of cell and/or tissue contexts. Thus, the emerging results were limited to the specific cell types or tissues analyzed and dependent on the growth matrix elasticity.

In the present study, we looked for changes in transcriptome, epigenomics, and functional pathway levels at short and long term in response to stretching of mouse skin cells.

2. Specific aims

As a graduate student in BMS, CityU, my long-term goal is to contribute a novel mechanism of regulatory element - TF in cellular response to stretch in a short period and long period at the single-cell resolution level, separately.

Aim 1: The overview of the shared mechanism of skin's response to stretch.

Previous works summarise the shared mechanism of skin response to stretch, including cell-cell junction, inflammation activities, and stem cell renewal. We now want to qualify the dysfunctional pathways in the long term and short term individually.

Aim 2: The molecular alternations of skin's response to stretch.

Recently ten years, a majority of previous work is about protein alternations. Only recently, two years, researchers are focus on gene expression. Here, we identified differentially expressed genes (DEGs) at the molecular level. And based on the known motifs, we can infer the TF involved in this regulation progress. In addition, we give a motif's footprint to validate our hypothesis.

Aim 3: The cell-cell communication of sub-populations.

Published studies apply one or more cell types. They didn't consider the factors of one cell line contains many cell sub-populations. One cell can release signal molecules, and the other cell can receive this signal from the receptor. Further, we can infer the time trajectory of cell transition.

3. Research design, methods and preliminary data

Methods:

To explore the differential expressed genes between the short time and long time, we utilized the microarray data (AffymetrixHTMG-430) 12-O-tetradecanoylphorbol-13-acetate (TPA) treatment in a short time of 4days with the accession number: GSE126231. T-test and Fold Change are used for identifying the DEGs.

ATAC-seq data was downloaded from the GEO database with the accession number: GSE126734. There are two treatments: control and expansion two days. Firstly, we used the ATACseqQC R package to assess the insert length of Tn5. We can observe the periodically flanking at 100bp (nucleosome-free regions, NFR), 200bp (mononucleosome), 400bp (dinucleosome), 600bp (trinucleosome) and so on. Then,

we infer the TFs that involved in this process. Finally, we can get the motif footprint figures.

GSE146637 downloaded processed scRNA-seq data. We used two independent methods for annotating cell sub-populations: SingleR and SCINA. One is based on the correlation between known transcriptome data and our scRNA-seq data. And the other is based on the expression of known marker genes. In the second method, we provided known marker genes in common cell types information. Finally, we annotate cell types by manual. Seurat R package is used for integrated data and plot the basic figures. Cellphone is used for cell-cell communication analysis. Slingshot R package is used for infer the evolution trajectory.

The shared mechanism of skin's response to stretch.

These shared functional pathways' genes are up-regulated in 4 days after expansion compared to TPA treated, including cell adhesion, cytoskeleton, proliferation, MAPK and so on.

We select 5 adhesion genes: Itga3, Itga6, Itgb1, Parva, and Pxn.

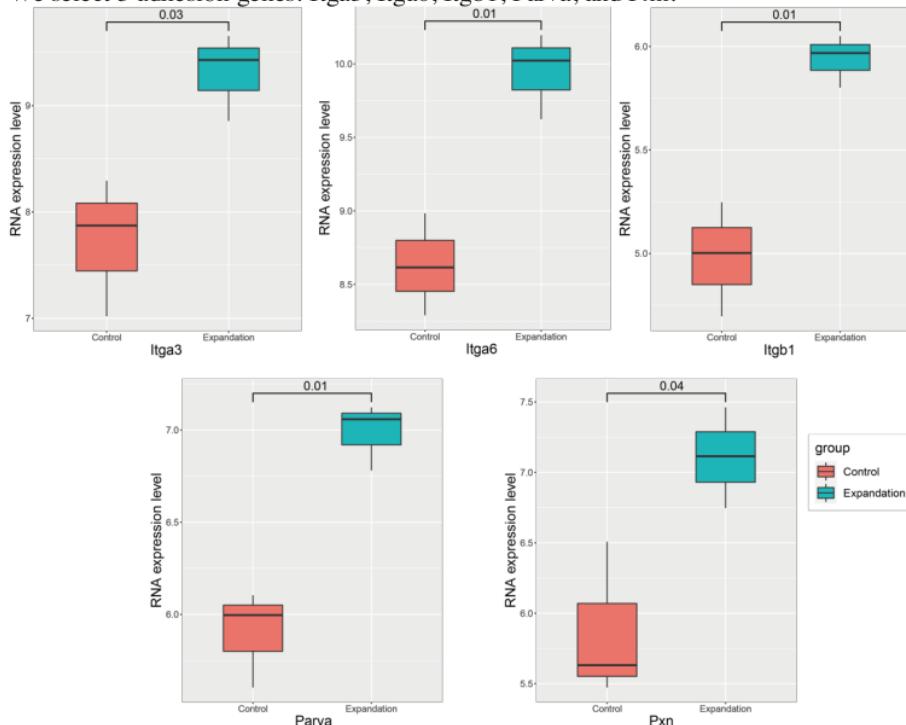


Figure 1. The selected adhesion genes's expression in control, expansion 4 days, and TPA treatments. The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

We select 6 cytoskeleton genes: Actn1, Arpc4, Flna, Myh9, Twf1, and Nav1.

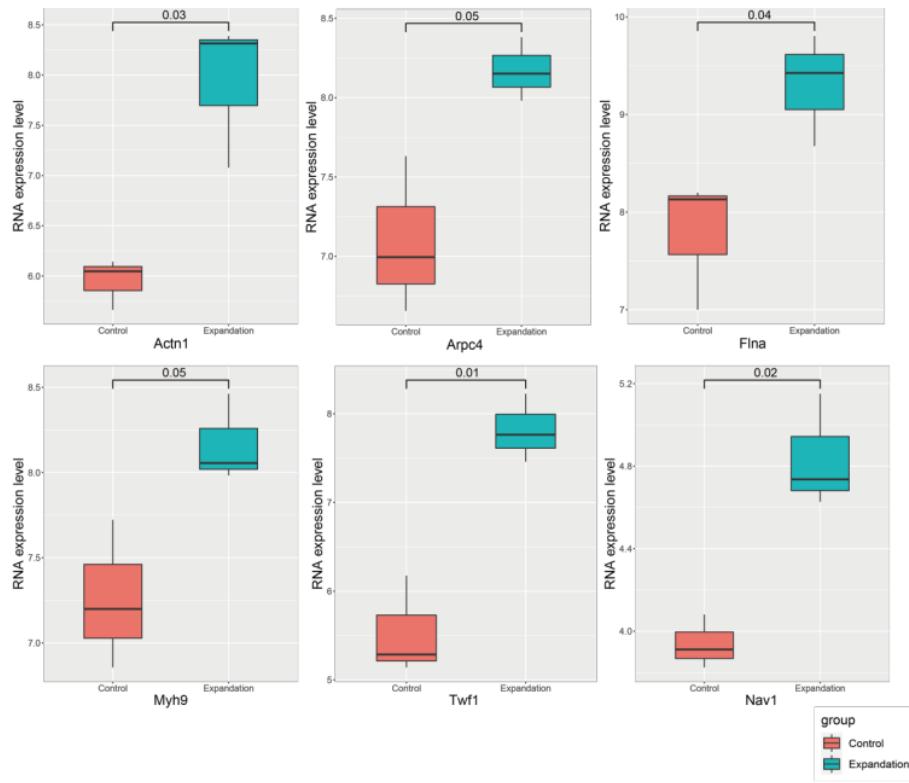


Figure 2. The selected cytoskeleton genes's expression in control, expansion 4 days, and TPA treatments. The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

We select 2 proliferation genes: Rras2, Krt6a, Ccnb2, Ccnb1, and Cdk1.

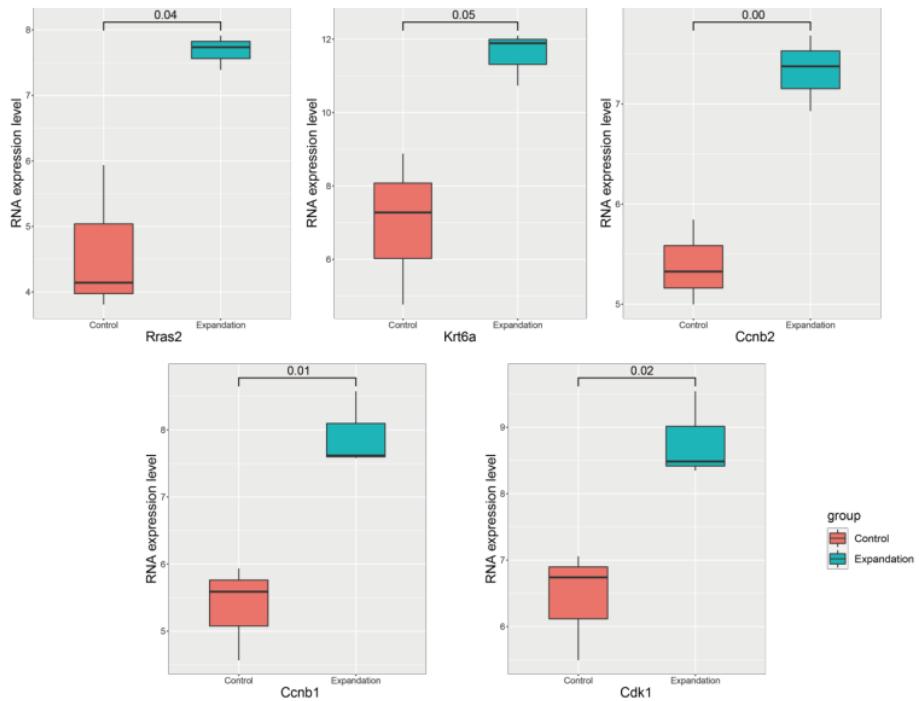


Figure 3. The selected proliferation genes's expression in control and expansion 4 days treatments.
The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

We select 6 immune genes: Tnfaip8l2, Myd88, Il33, Ikbke, Irak1, and Nfkb1.

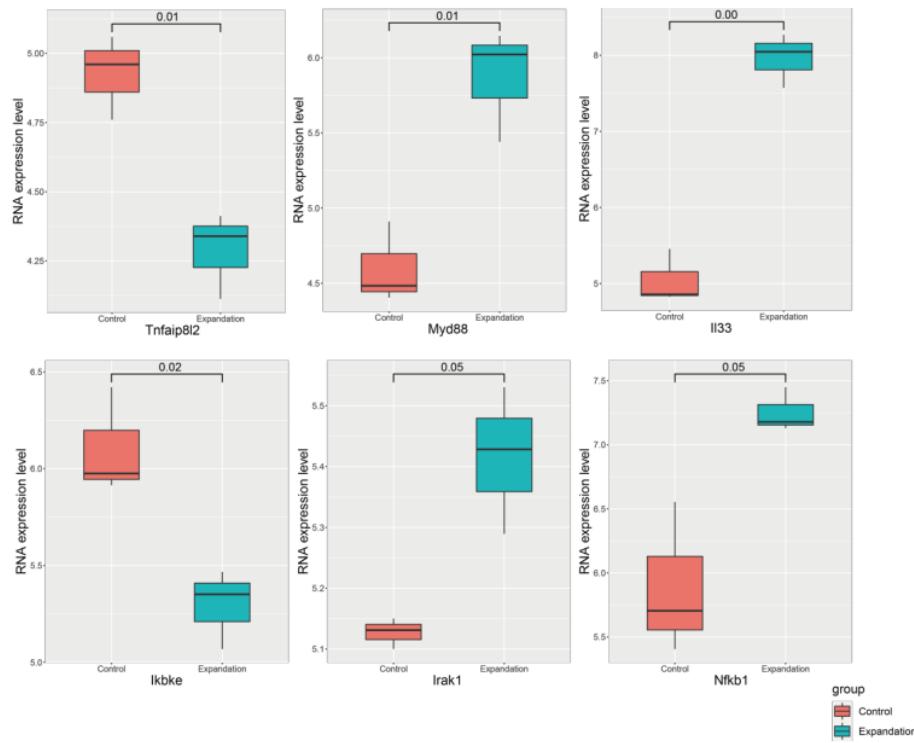


Figure 4. The selected immune pathway genes's expression in control and expansion 4 days treatments. The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

Molecular alternation of stretch induced

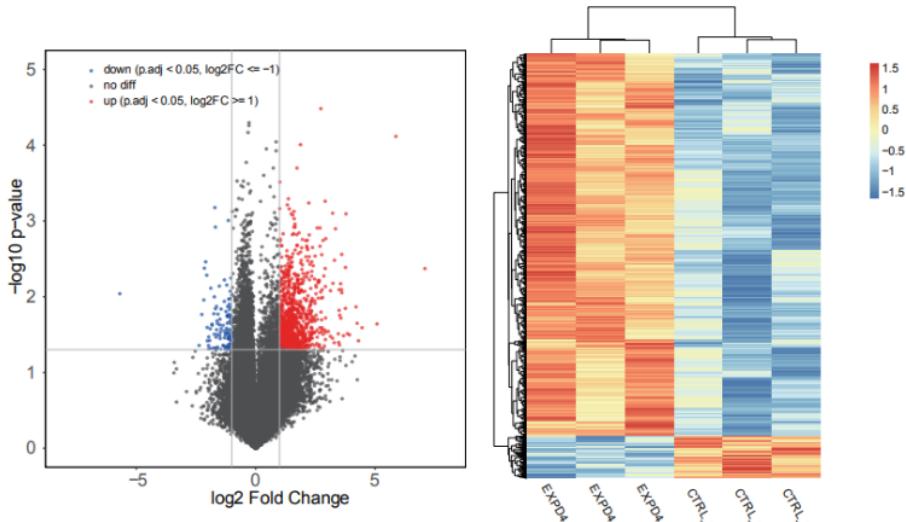


Figure 5. The DEGs between control and expansion 4 days treated mouse skin cells. A. The x-axis indicates the log2FC, while the y-axis indicates the -log10(p-value). B. The heatmap of DEGs across 6 samples. 9

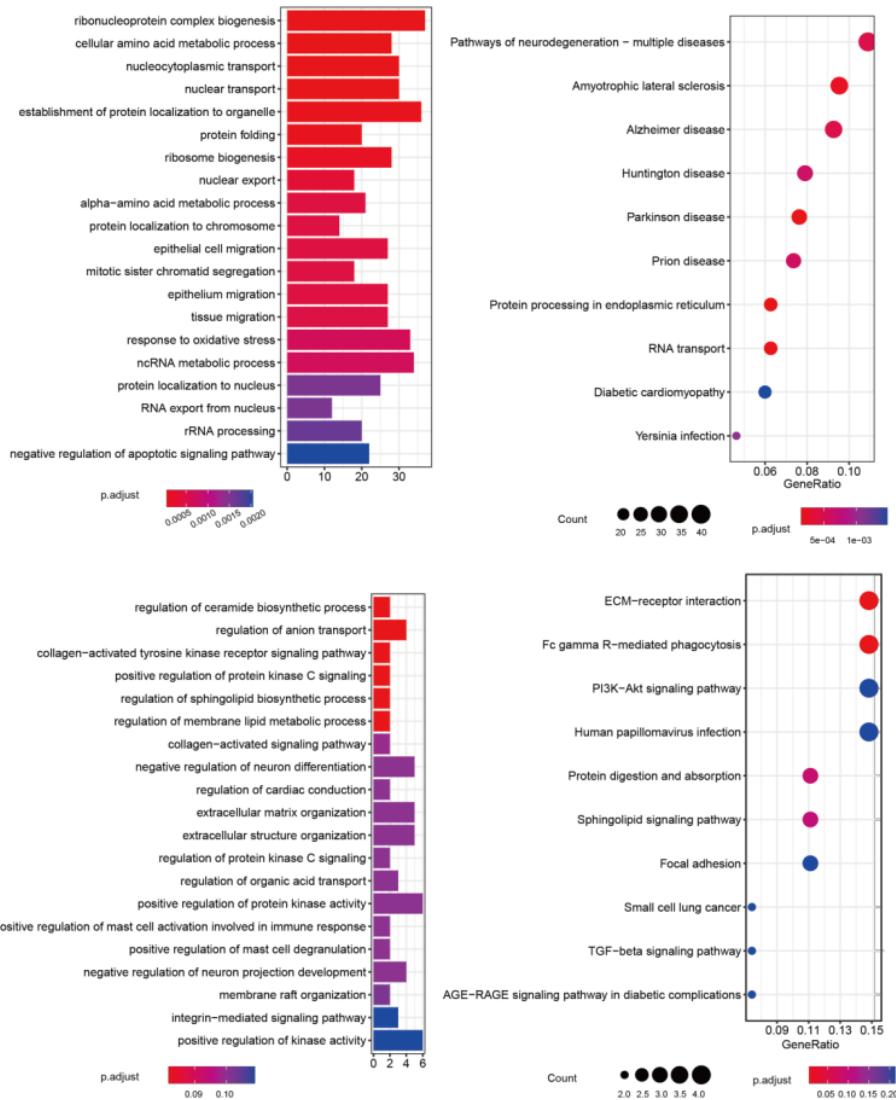


Figure 6. The dysfunctional biological pathways that DEGs involved in. AB. The dysfunctional biological pathways (GO_BP and KEGG pathways) that up-regulated genes in expansion 4 days treatment. CD. The dysfunctional biological pathways (GO_BP and KEGG pathways) that down-regulated genes in expansion 4 days treatment.

The regulatory model of TF

Table 1. The statistics of ATAC-seq data.

Sample	# raw reads	# clean reads	% map reads	# remove duplicates
Ctl_S18	131,087,101	129,279,102	77.86%	93,124,764
Exp-D2	109,031,667	107,629,069	83.28%	63,465,284

According to the results of ATAC-seq, we got 2 significant motifs, and their TFs are: AP1 and TEAD.

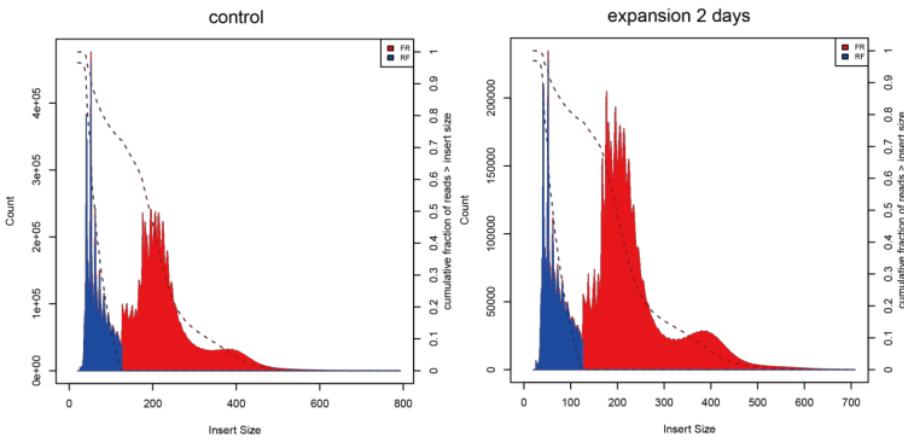


Figure 7. The distribution of Tn5 insert length in control (A) and expansion 2days (B) ATAC-seq data. x-axis indicates the length, and the y-axis indicates the number of counts in mapped data.

Table 2. The TF motifs that enriched in ATAC-seq's peaks.

The first column indicates TFs' symbol, the second column indicates motifs' sequence, and the last column indicates the significant value.

Rank	TF	Motif	pvalue	Target
1	Fos(bZIP)	GGAATGAGTCATCT	1e-1340	17.85%
2	JunB(bZIP)	AAATGAGTCATCT	1e-1315	18.18%
3	AP-1(bZIP)	ATGACTCACTC	1e-1262	19.51%
4	Fosl2(bZIP)	GATGAGTCATCC	1e-1104	12.34%
5	Jun-AP1(bZIP)	GATGAGTCATCC	1e-975	9.80%
6	KLF1(Zf)	GCCCCGCCCC	1e-392	27.30%
7	NFIA	TGCCCAAGT	1e-249	22.64%
8	Elk1(ETS)	TTCCGGTT	1e-157	22.36%
9	p53(p53)	TGTCTGGG	1e-146	4.28%
10	GRHL1	AACATGGTT	1e-129	13.68%
11	CTCF(Zf)	GCCACCTAGTGG	1e-126	1.28%
12	CREB3L4(var.2)	TGACGTCA	1e-104	13.91%
13	NFYC	TGATTAGC	1e-104	25.73%
14	MXI1	ACAATGTCT	1e-80	20.08%

15	RELA		1e-64	10.15%
16	TBX5		1e-60	0.29%
17	BARHL1		1e-56	17.59%
18	NRF(NRF)		1e-55	3.08%
19	POL012.1_TATA- Box		1e-55	0.27%
20	RUNX3		1e-52	0.21%
21	GATA6		1e-52	3.34%
22	Stat3(Stat)		1e-50	5.11%
23	NFYA		1e-41	0.24%
24	TEAD4(TEA)		1e-34	8.63%
25	TEAD1(TEAD)		1e-28	9.41%
26	TEAD3(TEA)		1e-24	10.83%
27	Smad3(MAD)		1e-24	30.93%
28	SOX14		1e-19	2.65%
29	TEAD(TEA)		1e-19	0.10%
30	TEAD2(TEA)		1e-13	5.01%
31	TEAD(TEA)		1e-9	6.00%
32	ZNF384		1e-2	3.10%

We select 2 AP1 regulated genes: Fosl1 and Fos.

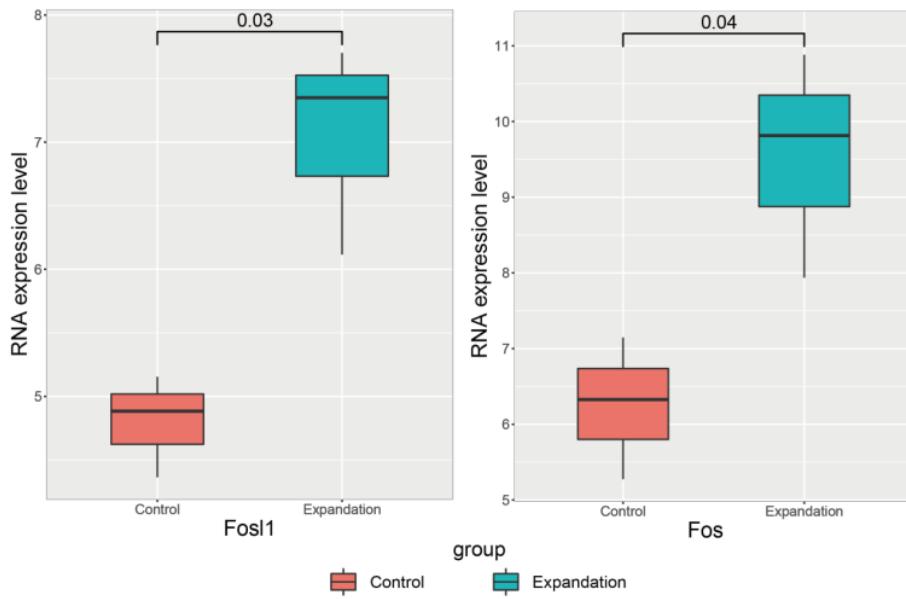


Figure 8. The selected AP1 regulated genes's expression in control, expansion 4 days, and TPA treatments. The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

We select 2 TEAD regulated genes: *Tead1* and *Cyr61*.

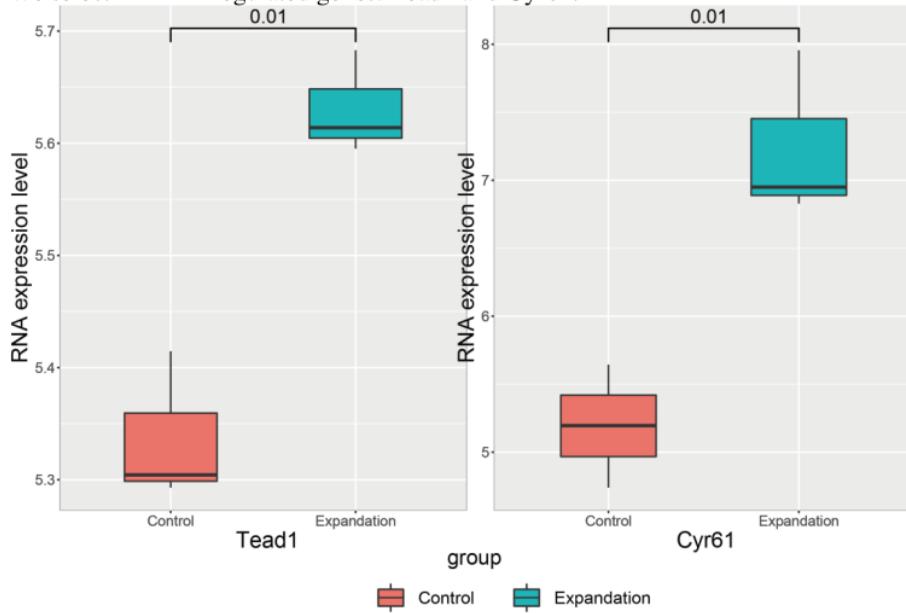


Figure 9. The selected TEAD regulated genes's expression in control, expansion 4 days, and TPA treatments. The x-axis indicates the treatment of the skin cells and the y-axis indicates the normalized RNA expression levels. The value in the single line represents the significance of a difference between groups.

The cell-cell communication.

According to the results of ATAC-seq, we got 2 significant motifs, and their TFs are: AP1 and TEAD.

In the past few years, the focus of developmental research has shifted from embryonic development to the mechanism of tissue homeostasis (tissue homeostasis), and significant progress has been made, which largely relies on cell lineage tracking (Lineage- tracing) and clonal analysis. The emerging cell lineage tracking technology can also exert its unique advantages in these fields. Due to skin's clear tissue anatomical structure and cell biology characteristics can be used as a good tool for this research. The epidermis comprises hair follicles (HFs) and the interfollicular epidermis (IFE) around them. The IFE contains a layer of basal cells (BC) expressing K14/K15 and an upper basal cell layer expressing K1/K10 multi-layer terminal differentiation.

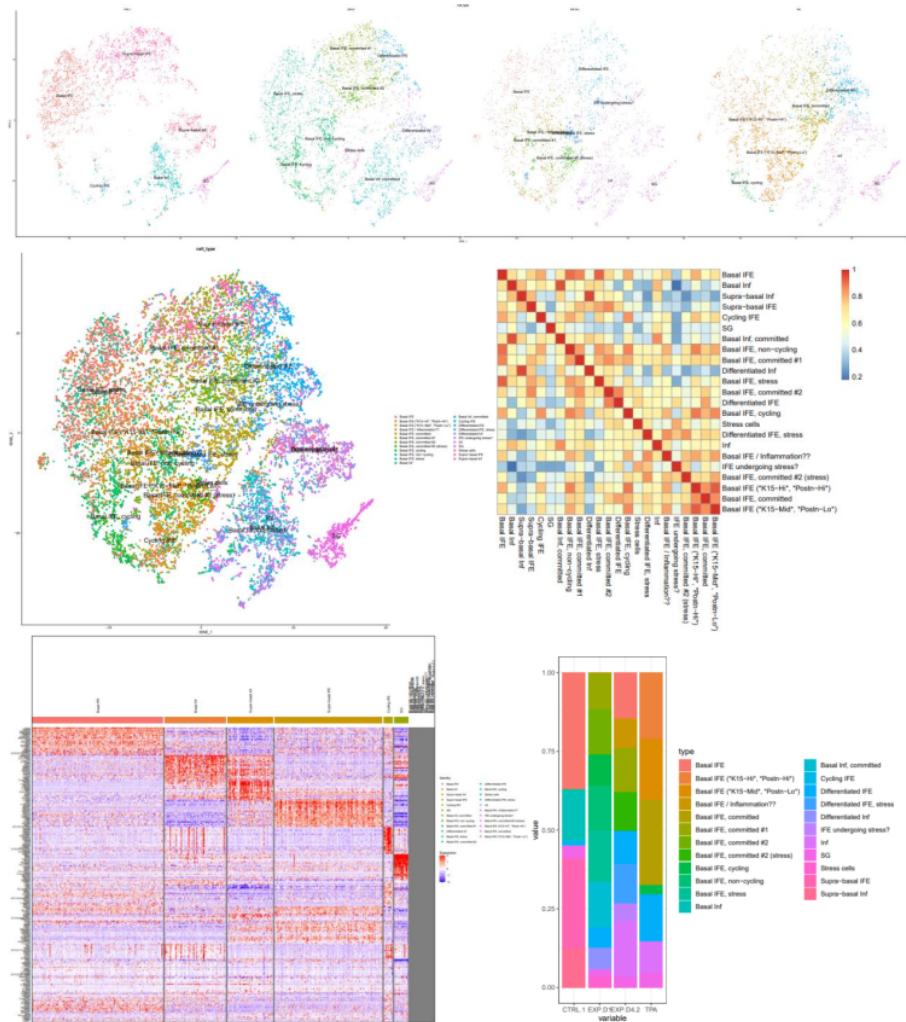


Figure 10. The integrated UMAP visualization of scRNA-seq of control, exp 1, 4 days, and TPA treated skin cells. A. The visualization of integrated scRNA-seq data . B. The heatmap of average top 30 marker genes' expression across 23 cell types . C. The heatmap of top 30 marker genes' expression in 23 cell types . D. The proportion of each cell types across 4 treatments.

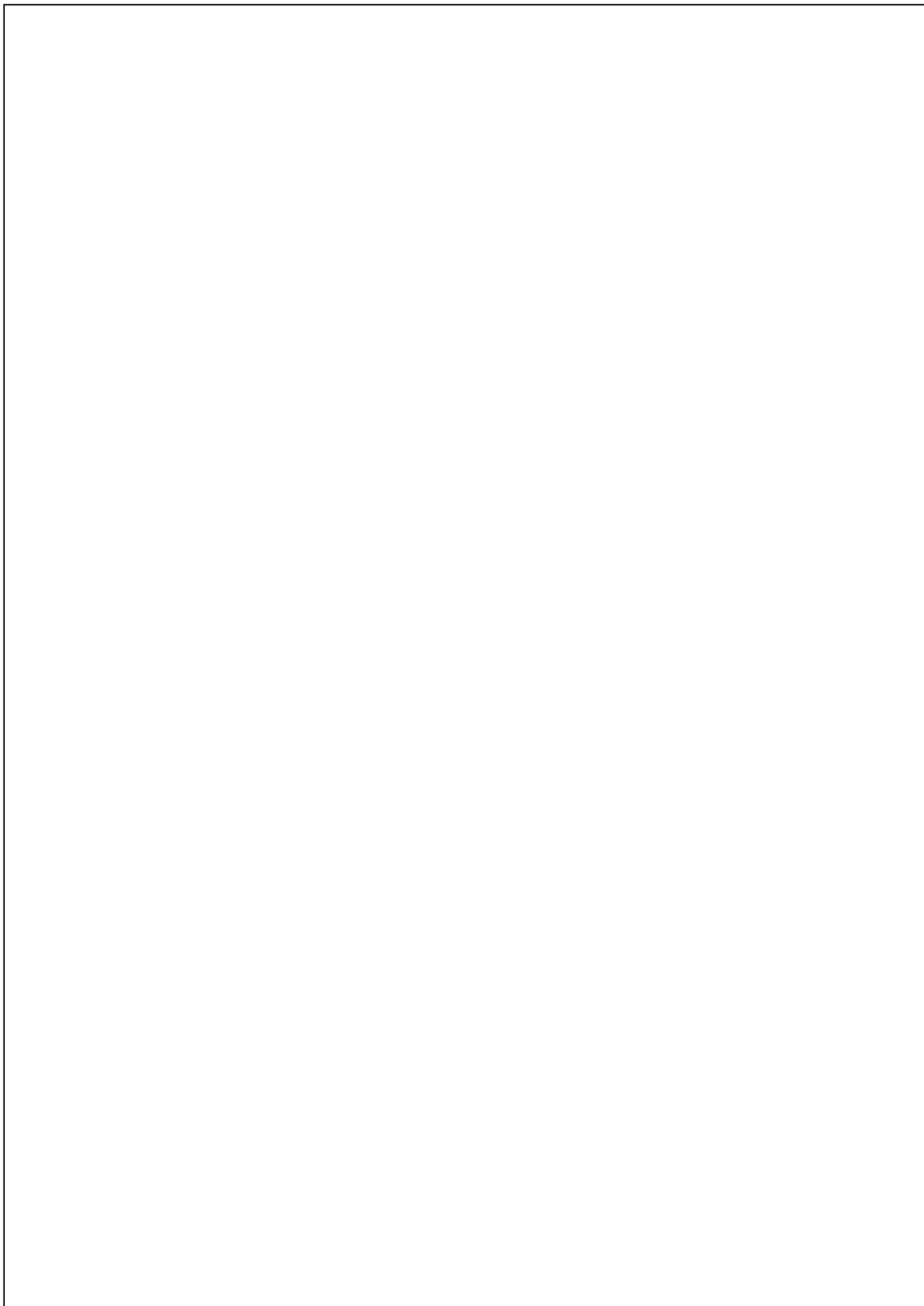


Table 22. The statistics of cell sub-populations.

	#CTRL	%CTRL	#EXP.D1	%EXP.D1	#EXP.D4	%EXP.D4	#TPA	%TPA	#Total	%Total
Basal IFE	1718	10.32%	0	0.00%	396	2.38%	0	0.00%	2114	12.70%
Basal Inf	760	4.56%	0	0.00%	0	0.00%	0	0.00%	760	4.56%
Supra-basal Inf	578	3.47%	0	0.00%	0	0.00%	0	0.00%	578	3.47%
Supra-basal IFE	1338	8.04%	0	0.00%	0	0.00%	0	0.00%	1338	8.04%
Cycling IFE	85	0.51%	0	0.00%	0	0.00%	0	0.00%	85	0.51%
SG	180	1.08%	123	0.74%	94	0.56%	205	1.23%	602	3.62%
Basal Inf, committed	0	0.00%	721	4.33%	0	0.00%	0	0.00%	721	4.33%
Basal IFE, non-cycling	0	0.00%	700	4.20%	0	0.00%	0	0.00%	700	4.20%
Basal IFE, committed #1	0	0.00%	571	3.43%	383	2.30%	0	0.00%	954	5.73%
Differentiated Inf	0	0.00%	334	2.01%	0	0.00%	0	0.00%	334	2.01%
Basal IFE, stress	0	0.00%	801	4.81%	0	0.00%	0	0.00%	801	4.81%
Basal IFE, committed #2	0	0.00%	712	4.28%	0	0.00%	0	0.00%	712	4.28%
Differentiated IFE	0	0.00%	318	1.91%	288	1.73%	644	3.87%	1250	7.51%
Basal IFE, cycling	0	0.00%	489	2.94%	0	0.00%	137	0.82%	626	3.76%
Stress cells	0	0.00%	165	0.99%	0	0.00%	0	0.00%	165	0.99%
Differentiated IFE, stress	0	0.00%	0	0.00%	343	2.06%	0	0.00%	343	2.06%
Inf	0	0.00%	0	0.00%	477	2.86%	433	2.60%	910	5.47%
Basal IFE / Inflammation	0	0.00%	0	0.00%	252	1.51%	0	0.00%	252	1.51%
IFE undergoing stress	0	0.00%	0	0.00%	152	0.91%	0	0.00%	152	0.91%
Basal IFE, committed #2 (stress)	0	0.00%	0	0.00%	331	1.99%	0	0.00%	331	1.99%
Basal IFE (*K15-Hi*, *Postn-Hi*)	0	0.00%	0	0.00%	0	0.00%	911	5.47%	911	5.47%
Basal IFE, committed Basal IFE (*K15-Mid*, *Postn-Lo*)	0	0.00%	0	0.00%	0	0.00%	1170	7.03%	1170	7.03%
	0	0.00%	0	0.00%	0	0.00%	842	5.06%	842	5.06%

B. The comprehensive identification of AS events that RBPs regulated in the relapse of glioma

1. Background and Significance

Glioma is an intracranial malignant tumour derived from the neuroepithelium, which accounts for about half of all primary intracranial tumours. It is the most common and highly aggressive primary tumour of the central nervous system. The treatment method is an intensive treatment, but the median survival time of the patient is about 15 months, the recurrence rate is extremely high, and the prognosis is inferior. Glioma is characterized by aggressive proliferation, high aggressiveness and high mortality. It can spread to the spinal cord and other parts of the brain, which is a huge challenge in the current treatment process. Genetic changes cause the occurrence of glioma.

Genetically, mutations or deletions of genes such as EGFR, IDH1, PDGFRA, HDM2, PIK3CA, PI3KR1, PTEN, TP53, CDKN2A, NF1, ATRX, and RB1 are known to be related to genetic changes. The occurrence of mass tumours is closely related. At present, the diagnostic methods of glioma mainly rely on histopathological examination, imaging examination and molecular diagnoses, such as isocitrate dehydrogenase (IDH) mutation, chromosome 1p/19q coding deletion, and telomerase reverse transcriptase (TERT). Promoter mutations, and so on.

The treatment of glioma generally adopts traditional treatment methods, such as radiotherapy, chemotherapy and surgical resection. However, due to the high heterogeneity and complexity of gliomas, these treatments have little effect on the survival of patients, and no substantial progress has been made in improving the prognosis of GBM. Temozolomide (TMZ) is an internationally recognized glioma

chemotherapy drug of choice. Its treatment of glioma has a history of more than ten years. Studies have shown that although TMZ can slightly extend the survival time of patients, almost all patients will eventually develop Drug resistance, so the overall curative effect of TMZ is not good, and it is easy to relapse after treatment. After radiotherapy and TMZ chemotherapy, the median recurrence time is generally seven months, and patients die typically after seven months. The current clinical status of glioma treatment is frustrating, and there is an urgent need for a new understanding of the molecular mechanisms of disease progression.

18

Alternative splicing is a way of regulating gene expression. The gene sequence of eukaryotic cells contains introns and exons, which are interspersed with each other. Generally speaking, introns are removed by RNA splices after genes are transcribed into mRNA precursors, and the remaining exons are fragments that can exist in mature mRNA (which will be further translated into proteins). An unspliced RNA contains multiple exons that are cut into different combinations to translate different proteins. This ensures that the exons in the same gene are combined in different forms so that a gene can produce other proteins at different times and in different environments, which increases the complexity or adaptability of the system under physiological conditions. At present, 35% of human genetic diseases may be caused by alternative splicing of RNA, such as retinitis pigmentosa and spinal muscular atrophy. In the human genome, about 95% of multi-exon genes have alternative splicing. It is common for a single gene to produce more than a dozen splicing isoforms through alternative splicing. The most prominent example is the Dscam gene in Drosophila, which has 38,016 potential alternative splicing types. Several typical splicing variants are associated with diseases, such as, SP3B1 in uveal melanoma and U2AF1 in lung cancer.

RNA-Binding Protein (RBP) plays an indispensable role in post-transcriptional regulation. It has a diverse structure of RNA binding domains, so it can participate in regulating multiple transcription mechanisms. RBP is involved in regulating mRNA, and its abnormal regulation is closely related to different kinds of cancer.

Here, we show that many RBPs in tumors have abnormal changes, which affect the translation of mRNA to the protein level and further participate in the occurrence and metastasis of glioma. Our study will help better understand the pathological mechanism of the tumorigenesis and metastasis of glioma.

Innovation: Few studies have been conducted on AS for the recurrence or grade of glioma.

2. Specific aims

My long-term goal as a graduate student in BMS, CityU is to determine some novel AS events in recurrent glioma compared to primary glioma and identify the different regulatory model of RBP.

Aim 1: The overview of the AS events in primary and recurrent glioma.

There are few matched primary and recurrent glioma samples for collection. Thus, few surveys are related to glioma relapse. At the beginning of our analysis, we need to give statistics data for all AS events in primary and recurrent glioma.

Aim 2: The selection of candidate AS events in known oncogenes.

From the above results, we can summarize more AS events in recurrent glioma compared to primary glioma. Then, we identified differential splicing events across primary and recurrent glioma RNA-seq data.

Aim 3: The identification of key RBPs and their regulatory mechanism.

We think the RBP has a differential expression pattern if the RBP involved in the glioma relapse. So we calculate the DEGs and intersect them with RBPs.

3. Research design, methods and preliminary data

In order to explore the differential expressed RBPs between primary and recurrent glioma samples, we utilized the RNA-seq data from the GEO database (GSE139533). Miso and rMAT are used for double-checking our identified AS events by calculating the PSI value.

DESeq2 R packages were used for identifying DEGs across primary and recurrent glioma samples with the criteria: $\log_{2}FC > 1$ and $p\text{-value} < 0.05$. Then, we intersect our DEGs list with the 4,689 known RBPs using combining three papers' supplementary data.

Piranha is used for peak calling of candidate RBP CLIP-seq data.

Table 4. The collection sample statistics data. The number in the cell indicates the number of samples are sequenced.

Accession number	#raw reads*2	#clean reads*2	#map reads	cancer type
GSM4143105	9,122,543	6,946,156	92.10%	primary glioma
GSM4143106	13,580,250	10,597,363	92.66%	primary glioma
GSM4143107	13,681,397	10,818,499	92.53%	recurrent glioma
GSM4143108	13,765,737	10,465,082	91.74%	recurrent glioma

C. The MEGENA co-expression network analysis reveals the key CSC regulators of HCC.

1. Background and Significance

Hepatocellular carcinoma (HCC) is one of the most common tumours in the world, which is the sixth-largest malignant tumour in the world and the third cause of death caused by tumours. Liver cancer is characterized by easily invading the vascular system of the liver. HCC merges Portal vein tumour thrombus (PVTT) affects the prognosis of liver cancer patients, the recurrence of liver cancer, and the choice of treatment methods. Despite continuous eradication or improvement of various treatment techniques, the prognosis of patients with liver cancer is still inferior. The pathogenesis of HCC is complicated. How to monitor the malignant transformation of liver cells or early diagnosis of HCC is still a medical problem.

HCC is prone to invade the portal venous system. 10%-40% of patients have PVTT when they are first diagnosed with liver cancer, and 5.4%-26.0% of patients undergoing liver resection have PVTT; in non-surgical treatment, PVTT occurred in 11.3%-38.0% of the patients, and the incidence rate during the autopsy was 44.0%-62.2%. An important reason for the poor prognosis of liver cancer is the formation of PVTT. Once formed, the disease progresses rapidly, causing portal hypertension, jaundice, ascites, intra-hepatic metastasis. The median survival time of untreated liver cancer with PVTT is only 2.7 months, and the median survival time without PVTT can reach 24 months.

The development of high-throughput sequencing technologies provides new views into mining the genomic, transcriptomic, and epigenomic signatures of various tumours by generating a large number of sequencing reads. That allows us better understand and delineate the mechanism of multiple cancers at the molecular level. The development of the TCGA project provides a rich data-based [19] platform for cancer research. Systems biology, especially network biology methods, has proven to be an

effective method for integrating multiple, large-scale data sets of complex human diseases, especially cancer. Multiscale embedded gene co-expression network analysis (MEGENA) is an efficient and accurate method for extensive data multi-gene analysis. MEGENA has been widely used to identify related clinical models and hub genes in different cancer types. For example, Yun et al., used MEGENA to validate that the expression of six hub genes highly is associated with the progression and prognosis of clear human cell renal cell carcinoma. Based on the gene expression data of breast cancer from GEO and TCGA database, Jin et al. combined with the MEGENA algorithm identified 15 hub genes as candidate breast cancer biomarkers. Wang et al. used MEGENA to analyze the differentially expressed genes (DEGs) of adrenocortical carcinoma. They found four hub genes that may be candidate biomarkers of adrenocortical carcinoma in clinical treatment. To better explore the pathogenesis of HCC, MEGENA is a powerful method to systematically explain the pathogenesis of the HCC.

The heterogeneity of tumours and the dry characteristics of cancer stem cells (CSCs) are one of the main reasons why tumours are challenging to cure, and they are also the main obstacles to cancer treatment research. **Table 5** lists some common tumours: cancer stem cell markers and signalling pathways. CSCs cells have the following characteristics:

1. *In vitro* culture can form neurosphere-like cells
2. Expression of stem cell marker genes SOX2 and CD133
3. Self-renewal and proliferation ability
4. After orthotopic transplantation, it can be Form a tumour similar in nature to the original tumour

It is speculated that CSCs are closely related to the occurrence and development of diseases, recurrence and treatment resistance, and even play a decisive role. Molecular and genomic heterogeneity and the persistence of sub-populations of cancer cells with stem characteristics after radiotherapy and chemotherapy are considered the main reasons for treatment resistance and related poor prognosis.

5

Table 5. The CSC markers and signaling pathways of common tumors.

Cancer Type	CSC markers	Main stemness pathways
Lung cancer	CD44+ CD24-, CD133+	Wnt/β-catenin , SCF-c-kit
Liver cancer	CD133+, CD13+, CD45-CD90+	Wnt/β-catenin , Notch
Esophageal cancer	ALDH1A1+	Pten/PI3K/Akt
Breast cancer	ALDH1+ CD44+ CD24-	JAK2/STAT3 , Pten
Stomach cancer	CD44+, CD133+, ALDH1+	Hedgehog
Pancreatic cancer	CD133+ CXCR4+	Hedgehog
Colorectal cancer	Lgr5+	Wnt/β-catenin , Yap1
Kidney cancer	CD133+, CD105+	Hedgehog
Prostate cancer	CD44+	Hedgehog
Bladder cancer	CD47+, ABCG2+	Hedgehog , Wnt/β-catenin
Cervical cancer	CD44+, CD49+	Notch
Leukemia	CD133+, CD34+/CD38-	Hedgehog , Wnt/β-catenin
Brain cancer	CD133+, CD44	Notch , Pten

Recently, accumulated data have shown the potential value of new blood-borne markers such as circulating blood tumor cells, key signaling molecules in related pathways, carcinoembryonic-specific proteins, long non-coding RNAs, and microRNAs in the diagnosis of liver cancer. This article focus on the gene co-expression network in the tumorigenesis and metastasis of HCC.

2. Specific aims

Aim 1: The construction of MEGENA network.

In order to identify the gene-gene interactions at the systematical biology, we need to construct two networks. The first network describe the interact between modules. We can identify hub module. And the second network describe the gene-gene interactions of each module. We can identify hub genes.

Aim 2: The determination of gene sets of CSC.

We identify candidate module based on the association with these genes in CSC.

Aim 3: The determination of hub genes.

The gene plays an important role in the biological process if the gene is located upstream of the regulatory relationship. We want to find the important biomarkers instead of passenger biomarkers.

3. Research design, methods and preliminary data

Firstly, the Pearson correlation coefficient (PCC) score is calculated for each gene pair. And the gene pair is sorted by the PCC value from big to small. Then, the network is embedded into the planar filtered networks (PFN). Finally, the shortest distance, local path index, and whole module are all used for iteration.

To optimize the modules that related to stem cells, we download 24 embryonic stem cells (ESCs) and 11 mesenchymal stem cells (MSCs) gene list from GSEA MSigDB database. Hypergeometric distribution test is applied for enrichment analysis of top 20 significant modules and each gene list.

The randomForest R package is used for constructing the regulatory network. There are two parameters in our algorithm: 1. The number of decision tree (10,000), and 2. The number of candidate input genes that can be selected for each node (the square root of the numbers of all input genes).

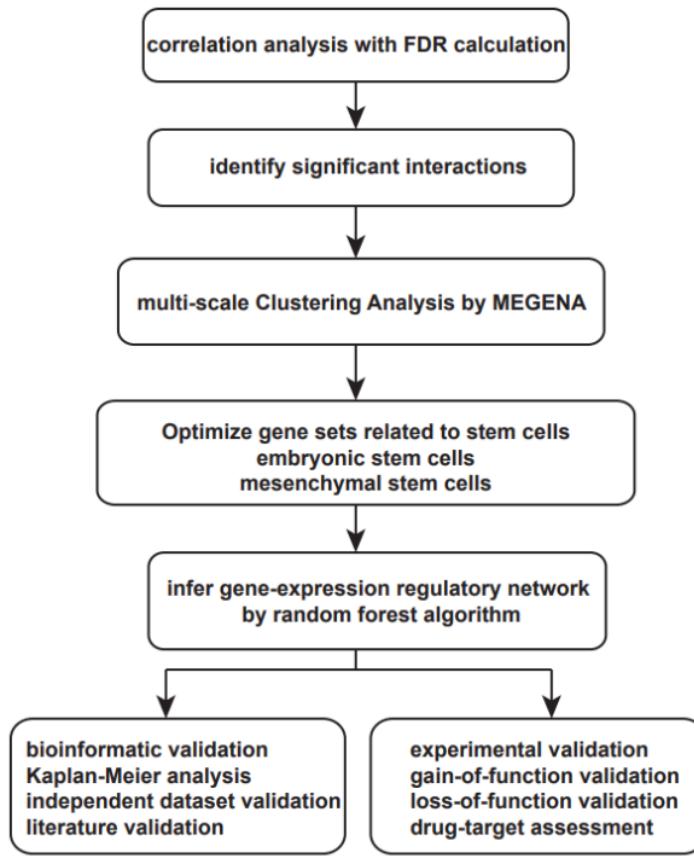


Figure 11. The workflow of the MEGENA network analysis of HCC RNA-seq data. The MEGENA algorithm first calculates the correlation between any two genes, and sorts the gene pairs according to the size of the correlation; then embeds them into the topological network through the planar maximum filtered graph (PMFG) algorithm to construct a planar filter network (PFN), and then perform multiple iterations on the original PFNs to obtain a more accurate classification by the shortest path distance, local path index and overall modularity. The enrichment analysis was performed on the modules obtained by the MEGENA algorithm and the gene sets related to the two kinds of stem cells(ESCs and MSCs) downloaded from the MSigDB data. Finally, our hub genes was validated by bioinformatics (survival analysis, an independent dataset, and literature) and experiments (knock-down, over-expression, MGMT dysfunction experiments in U87 cell line) technologies.



Figure 12. The module hierarchy. The node in the network indicates the module, and the edge indicates clustering relationship between two modules. The size of node represents the number of genes in the module multiple the scale factor of the module.

Table 6. The characterise of the top 20 significant modules

module	size	hub genes
M1	1634	TBX3(51),RNF43(40),DYNC1II(38),FAM124A(37),DUXA(36),ACS BG1(34),AQP9(32),ABCB5(31),LAMA3(31),MLANA(30),NMNAT2 (30),FBXO31(30),NKD1(30),GSTM5(28),TREX2(28),WNT6(28),DC LK2(27),CLSTN2(27),B4GALT6(27),CTSK(26),IL13(25),IGFBP5(25 ,SPARCL1(25),ZNF503(24),ADIPOQ(23),CHRD1(23),TRIM63(23 ,HLF(23),IGFN1(22),POPDC2(22),ATP8B4(21),C16orf89(21),DAPL 1(21),ITPR2(21),LEP(20),DCLK1(20),AVIL(20),OR7C1(20),SYT14(20), ARHGAP28(19),CRTAC1(19),ANK2(19),LRRC39(19),LMO3(19 ,CRISP2(19),ABL1(19),SERF2(19),PMP22(19),FABP3(18),CDHR1(18), MAMDC2(18),IGFBP6(18),CPEB1(17),HHATL(17),CHRM3(17 ,COX4I1(17),PSAP(17),HSPB2(16),AEBP1(16),GLUL(16),COX7B(1 6),MERTK(16),GCK(16),SP5(16),PTCH2(15),CMA1(15),TRDN(15), PLN(15),MEIS3(15),IGFBP7(15),NOTUM(15),FAM8A1(15),MYL9(14), FABP4(14),NDUFB7(14),MYBPHL(14),AP1S2(14),ESRRG(14), ARHGAP1(14),IRS1(14),NUTF2(14),CD63(14),FAM168B(14),TIMP 2(14),AZU1(13),CASQ1(13),JPH2(13),GPNMB(13),EXOC6B(13),C

DC14A(13),PCYT1B(13),AXIN2(13),MXRA8(13),HEPACAM(13),D
 ISC1(13),TECR(13),TWIST1(13),TCEAL5(13),TSPAN5(13)
 TPX2(79),ADAMTS14(63),TOP2A(42),CACNG7(38),FABP7(36),O
 AS2(35),SASH3(35),ACTB(32),BMP5(31),CLEC14A(31),MCM6(31)
 ,TYROBP(31),YWHAQ(31),FERMT3(29),CD53(27),KIF2C(27),RPN
 2(26),LCK(25),CYBB(25),LAPTM5(25),CDH5(25),CCNB1(25),MFS
 D10(25),SPARC(24),ACTR2(24),FCER2(23),ARHGAP30(23),NUSA
 P1(23),CYYR1(23),KPNB1(23),THBS2(23),IL16(22),TMSB15A(22),
 HCLS1(22),CD93(22),NEK2(22),FOXM1(22),NCAPD2(22),CD5(21)
 ,JKZF1(21),CSF1R(21),LGALS1(21),KIFC1(21),LMNB1(21),CRISP
 LD1(20),CORO1A(20),SPI1(20),PTN(20),KLHDC7B(20),RACGAP1
 (20),ESAM(20),ACTG1(20),S100A11(20),GPR84(19),PLAU(19),DD
 X11(19),PTHLH(19),ADAM33(19),APCDD1L(19),CDC20(19),MS4
 A4A(19),PSMB9(19),RUNX2(19),CDK1(19),PCNA(19),RCC2(19),T
 SHZ3(19),HS3ST3A1(18),SLC35E4(18),OLFML2B(18),CD3E(18),A
 IF1(18),TCF19(18),CLTC(18),UBE2C(18),DAPP1(17),LOXL1(17),G
 PSM3(17),MCM2(17),COL4A2(17),TAP1(17),MYH9(17),CLEC11A(16),MMP2(16),DIO3(16),COL1A2(16),SELPLG(16),MELK(16),C1Q
 BP(16),BPTF(16),TPM4(16),CANT1(16),AP1G1(16),UBE2L6(16),B
 3GALT2(15),HLA-
 DPB1(15),DSE(15),FOXH1(15),ANTXR1(15),RANBP17(15),NDC80
 (15),KHDRBS1(15),DBN1(15),HLA-E(15),CALR(15),DHX38(15),CDC42(15),PLCD4(15),DYNCL1I2(15),CTBP1(15),MBTPS1(15),FCRL3(14),GLI3(14),COL1A1(14),DOCK2(14),MS4A6A(14),GCNT1(14),GMFG(14),TUBA1B(14),MCM3(14),NOTCH3(14),PTTG1(14),RAB31(14),DNAJC8(14),RSAD2(14),EFTUD2(14),UBE2Z(14),SET(14),OLAH(14),SEC61A1(14),ATXN1L(14),RBP2(14)
 DHX9(45),KIAA0319(44),ANXA8(40),ATP1A1(35),HNRNPA1(35),HNRNPU(35),LCA5(30),PDGFD(30),GPR50(29),HDGFL1(26),ILF3(26),PTBP1(26),TOMM20(26),KCNH3(25),KHSRP(25),HNF1B(25),ITGB1(25),KLK6(24),EIF4G2(24),VTCN1(24),WDR91(24),CITED1(23),CTTNBP2(23),SLC4A2(23),NEURL3(22),PMEP1(22),AQP1(21),COL11A1(20),CFTR(20),RHOV(20),CCT6A(20),HNRNPK(20),HNRNPUL1(20),RAPS(19),SCD5(19),TPR(19),CHST3(18),FAM117B(18),SFRP5(18),TGFB2(18),ADSL(18),IGFL1(17),SLC34A2(17),CTTNBP2NL(17),DCDC2(17),NEO1(17),BAZ1B(17),EIF4H(17),SUCNR1(17),GALNT6(16),ATP8A2(16),MITF(16),IAPP(16),RAB36(16),ANXA4(16),CUL4A(16),FGFR3(16),LARP1(16),XRCC6(16),DPPA2(15),HUNK(15),PKHD1(15),ITIH5(15),BAZ2A(15),REPIN1(15),TCF12(15),ZMIZ1(15),PHOX2A(14),B3GNT7(14),RPL6(14),TNF(14),DRG1(14),SPNS2(14),COPB1(14),TMEM127(14),ZDHHC1(14),ARL14(13),CEACAM7(13),KCNJ4(13),TMEM72(13),SCTR(13),GRM8(13),RPL3(13),EEF1A1(13),CBX3(13),KDM3B(13),AHCTF1(13),EEF2(13),ZNF317(13),RPL38(66),MRPL27(60),GALR2(50),ZBTB8OS(46),MRPL52(44),ZNF774(43),ZNF19(39),C7orf61(39),RPL24(39),MRPL22(38),RPL37A(37),SCAMP3(37),ZNF492(36),NDUFA7(36),SEC24C(36),NCAN(35),MRPL47(35),SDHAF2(35),SNRPF(35),BLOC1S1(35),ZNF778(33),ARID3A(32),FBL(31),RPL32(31),TXNRD1(31),ZNF20(30),SNRPD2(30),FNBP4(29),VPS52(29),SHARPIN(28),RPL36A(28),RPL31(28),SNRPG(28),TKT(28),RPL37(27),BRPF3(27),PI4KB(27),LSM3(27),HPSE2(26),ZNF519(26),ZNF625(26),ZNF658(26),RBM39(26),SF3B4(26),GPS1(26),RPS27A(25),NCL(25),SMARCC2(25),ZNF423(24),ZNF442(24),SNORD91B(24),TIA1(24),GEMIN7(23),RPS3(23),HSPB

11(23),POLR2H(22),ISG20L2(22),POLR2J(22),PRCC(22),PYGO2(22),RPL35(22),PSMA2(22),CNIH4(22),EIF3A(22),EP300(22),ARID1A(22),CRH(21),CA3(21),ZNF486(21),RPL18(21),RPL27A(21),SNHG6(21),RPL35A(21),SEC61G(21),DYNLRB1(21),PSMC5(21),SSB(21),NHP2(21),PRPF31(21),PCBP1(21),CHST13(21),UBE2J2(21),FGF17(20),RPL34(20),NUDT1(20),SCRIB(20),ADNP(20),ARPC3(20),NEDD8(20),UBL4A(20),ZNF713(19),ENY2(19),CCT4(19),NOP58(19),NOSIP(19),ANAPC11(19),PEX2(19),STX8(19),SNRPD1(19),ZNHIT1(19),CDC20B(18),CAGE1(18),RPL14(18),RPL23(18),NME1(18),ARMC1(18),SALL3(18),SETDB1(18),MEA1(18),HGS(18),TRIM28(18),UQCR11(18),SEMA4G(18),TPRKB(18),TOM1L1(18),ZP4(17),ZPB P2(17),RPS21(17),MT1IP(17),NDUFA2(17),EXOSC4(17),AURKAIP1(17),NAA38(17),BCCIP(17),CLK2(17),VPS28(17),SNRPE(17),SNR PC(17),FAM122B(17),EHMT2(17),MRPL45(17),RAN(17),PIR(17),PSMD10(17),MED11(17),RPS8(17),PABPC1L(17),HCFC1(17),ZNF256(17),TIMM9(17),UBAP2L(17),SNRNP200(17),TIMM50(17),NCCR P1(16),RBPJL(16),HHLA2(16),TKTL1(16),POLR2K(16),RPS24(16),ZNF90(16),RPS10(16),NDUFA9(16),HAX1(16),MRPS7(16),CSNK2B(16),CCT7(16),PPP2R5D(16),PSMA4(16),ASXL1(16),NDUFB8(16),TTC1(16),HIC2(16),SF3B5(16),GOPC(16),ZNF770(16),MAK16(15),ZNF223(15),ZNF235(15),CYHR1(15),MRPL13(15),ZNF549(15),BOP1(15),COPS6(15),NOP56(15),SYMPK(15),AGAP6(15),CPNE3(15),RP9P(15),RPS18(15),PABPC1(15),DNTTIP2(15),ANAPC7(15),SSNA1(15),DHX16(15),AARS2(15),MRPS33(15),REXO4(15),UBR5(15),TRAF7(15),LUC7L3(15),CREBF(15),ZCCHC3(15),ZNF606(15),TT C13(15),SERINC1(15),ZNF732(14),HOXD13(14),PIK3CG(14),PUF60(14),DUS1L(14),HSF1(14),PFDN4(14),NDUFB1(14),PTRH2(14),MPDU1(14),CHCHD1(14),ATP6AP1(14),RPS23(14),RPS29(14),FIS1(14),FARSB(14),MRPS21(14),VBP1(14),TATDN1(14),SS18L2(14),ZF551(14),CHCHD5(14),NSMCE1(14),TPD52L2(14),NUP43(14),IK(14),LSM14B(14),UBE2F(14),DPH2(14),ANKRD17(14),PSMD13(14),HNF1A(14),GSDMB(14),RPF1(14),VPS72(14),TRIM71(14),YWHA Z(14),MATN4(13),DMWD(13),RPS12(13),BNIPL(13),LSM14A(13),NUCD1(13),UBA2(13),DENR(13),SNRBP2(13),NDUFA13(13),COX16(13),KATNA1(13),NDUFA11(13),PFDN5(13),POGZ(13),AKR1B10(13),DPM2(13),UBE2A(13),ZNF444(13),SSU72(13),WDR46(13),EXOSC1(13),MCTS1(13),PCYT2(13),PPIH(13),MED6(13),CPSF3(13),HRAS(13),PIH1D1(13),NDUFS6(13),TIMM8B(13),WAC(13),NMT1(13),ATXN2L(13),BTAF1(13),SAP30BP(13),NOP16(13),SMG7(13),PPAN(13),MIIP(13),RFNG(13),RPS27(13),PSMA5(13),SRRM2(13),YTHDF1(13),SDF4(13),ZNF687(13),GRIK1(47),KIF26A(43),MIA(41),KLK10(39),CDC42BPG(37),PFKP(30),TMC4(29),NKAIN4(28),ITGB4(28),SEL1L3(28),FAM167A(27),SPINT2(25),SPINT1(25),PLAC4(24),ELOVL7(24),TMEM132A(24),SLC16A14(24),MUCL1(23),AGRN(23),PTPRR(22),SFMBT2(22),DS G2(22),THSD4(22),SLC25A15(22),BIRC7(21),ATP1A3(21),SCEL(20),MAPRE1(20),ALDOA(20),CTNND1(20),IQGAP1(19),MSLN(18),DDR1(18),PDE3A(18),CHMP4B(18),TMPRSS4(18),CAPN9(17),C3orf70(17),PAX2(17),DCST2(17),OTOP3(17),ARG2(17),EPCAM(17),AKAP12(17),SYNGR3(17),VSIG1(16),C2orf15(16),LGALS9B(16),MOS(16),APLP1(16),PDE1C(16),ESRP1(16),CHRNBB3(15),MSTN(15),BCL2L15(15),C2CD4C(15),GALNT3(15),C9orf57(15),IRF4(15),SHANK2(15),MGAT5(15),SULT2B1(15),SERHL2(15),TMC5(15),ERN2(14),FAM83E(14),DPP10(14),GPRC5A(14),CKMT1B(14),SOX1(14),AHNAK2(14),ATP8A1(14),BHLHA15(14),PCDHGB5(14),RAB25(

		14),NCEH1(14),MUC12(14),SOD3(14),PLEKHB2(14),SFTA2(14),Z DHHC5(14),VTRNA1- 2(14),WNT1(14),ZPLD1(14),AGR3(13),SLC22A8(13),HSPA6(13),S YT13(13),UNC5D(13),EPHB4(13),GANAB(13) GLYATL1(115),AADAT(66),GYS2(63),DMGDH(61),SLC27A5(52), ETFDH(47),LCAT(45),ALDH6A1(43),LDHD(42),CSRNP1(41),SOR D(41),PEX11G(40),ABAT(40),PLK3(39),EHHADH(39),GLYAT(35), SARDH(35),ASPDH(31),ANGPTL6(29),PCK2(28),COLEC10(26),M PDZ(26),MOGAT2(25),DAO(25),SGMS2(25),ZFP36(25),SCP2(24),C PT2(23),HINT2(23),FCN2(22),DHODH(22),ACAA1(22),GNE(22),S ERTAD1(22),GHR(21),EPHA2(21),PHLDA1(21),HMGCL(21),SOCS 3(21),GLYCTK(21),CLEC4G(20),DCAF1I(20),SAR1B(20),PINK1(2 0),ZFAND5(20),CAT(19),RALGDS(19),NUBP2(19),CHMP7(19),SL C10A1(19),FNDC5(19),KLF15(19),TTC36(18),C8A(18),SEC14L2(18)),HPX(18),MRPL54(18),NNMT(18),IVD(17),AOX1(17),STEAP3(17 ,GRHPR(16),METTL7A(16),XDH(16),CYP4V2(15),C1RL(15),FAH D2A(15),PFKFB1(15),CYP3A4(15),HIBADH(15),C9(15),OIT3(15),H AGH(15),SLC39A9(15),PIGV(15),SLC35D1(15),C19orf38(14),CT47 B1(14),MTHFD1(14),DHR4L2(14),GPD1(14),FOXO1(14),GCDH(1 4),ACACB(14),SPRYD4(14),SPOCK3(14),CD302(14),SLC6A1(14),T TC37(14),SLC25A46(14),DCN(13),HSD17B8(13),EPHX2(13),CDC1 4B(13),ASB15(13),FAM149A(13),KLF9(13),ETS2(13),PAFAH2(13), AASS(13),PJA2(13),GBP7(13),NAT2(13) FAM124A(37),DUXA(36),ACSBG1(34),ABCB5(31),MLANA(30),N MNAT2(30),GSTM5(28),DCLK2(27),CLSTN2(27),B4GALT6(27),C TSK(26),WNT6(26),IL13(25),IGFBP5(25),ADIPOQ(23),CHRDL1(23 ,TRIM63(23),IGFN1(22),POPDC2(22),ATP8B4(21),C16orf89(21),D APL1(21),LEP(20),DCLK1(20),AVIL(20),OR7C1(20),SYT14(20),AR HGAP28(19),CRTAC1(19),ANK2(19),LRRC39(19),LMO3(19),CRIS P2(19),ABL1(19),SERF2(19),PMP22(19),FABP3(18),IGFBP6(18),CP EB1(17),HHATL(17),MAMDC2(17),CHRM3(17),COX4I1(17),PSAP (17),HSPB2(16),AEBP1(16),COX7B(16),GCK(16),PTCH2(15),CMA 1(15),TRDN(15),PLN(15),IGFBP7(15),MYL9(14),FABP4(14),MEIS3 (14),NDUFB7(14),MYBPHL(14),AP1S2(14),ARHGAP1(14),NUTF2(1 4),CD63(14),FAM168B(14),TIMP2(14)
M6	1807	TREX2(28),CDHR1(18),ESRRG(14),DISC1(13) TBX3(51),RNF43(40),DYNCI1I(38),AQP9(32),LAMA3(31),FBXO3 1(30),NKD1(30),SPARCL1(25),ZNF503(24),HLF(23),ITPR2(21),GL UL(16),SP5(16),NOTUM(15),MERTK(15),FAM8A1(15),IRS1(14),A XIN2(13),HEPACAM(13),TSPAN5(13) ADAMTS14(63),CACNG7(38),FABP7(36),OAS2(35),SASH3(35),A CTB(32),BMP5(31),CLEC14A(31),TYROBP(31),YWHAQ(31),FER MT3(29),CD53(27),RPN2(26),LCK(25),CYBB(25),LAPTM5(25),CD H5(25),MFSD10(25),SPARC(24),ACTR2(24),FCER2(23),ARHGAP3 0(23),CYYR1(23),THBS2(23),IL16(22),TMSB15A(22),HCLS1(22),C D93(22),CD5(21),IKZF1(21),CSF1R(21),LGALS1(21),CORO1A(20), SPI1(20),PTN(20),KLHDC7B(20),ESAM(20),ACTG1(20),S100A11(20), GPR84(19),CRISPLD1(19),PLAU(19),DDX11(19),PTH1LH(19),A DAM33(19),APCDD1L(19),MS4A4A(19),PSMB9(19),RUNX2(19),T SHZ3(19),HS3ST3A1(18),SLC35E4(18),OLFML2B(18),CD3E(18),A IF1(18),DAPP1(17),LOXL1(17),GPSM3(17),COL4A2(17),TAP1(17), MYH9(17),CLEC11A(16),MMP2(16),DIO3(16),COL1A2(16),SELPL G(16),C1QBP(16),TPM4(16),AP1G1(16),UBE2L6(16),B3GALT2(15 ,HLA-
M7	1057	
M8	197	
M9	380	
M10	1728	

M11	626	DPB1(15),DSE(15),ANTXR1(15),RANBP17(15),DBN1(15),HLA-E(15),CALR(15),DHX38(15),CDC42(15),PLCD4(15),DYNCL1I2(15),CTBP1(15),MBTPS1(15),FCRL3(14),GLI3(14),COL1A1(14),DOCK2(14),MS4A6A(14),GCNT1(14),GMFG(14),NOTCH3(14),RAB31(14),DNAJC8(14),RSAD2(14),OLAH(14),SEC61A1(14),ATXN1L(14),RB P2(14) TPX2(79),TOP2A(42),MCM6(31),KIF2C(27),CCNB1(25),NUSAP1(23),KPNB1(23),NEK2(22),FOXM1(22),NCAPD2(22),KIFC1(21),LMNB1(21),RACGAP1(20),CDC20(19),CDK1(19),PCNA(19),RCC2(19),TCF19(18),CLTC(18),UBE2C(18),MCM2(17),MELK(16),BPTF(16),CANT1(16),FOXH1(15),NDC80(15),KHDRBS1(15),TUBA1B(14),MCAM3(14),PTTG1(14),EFTUD2(14),UBE2Z(14),SET(14),KIAA0319(44),ANXA8(40),ATP1A1(35),LCA5(30),PDGFD(30),KC NH3(25),HNF1B(25),ITGB1(25),VTCN1(24),WDR91(24),CITED1(23),CTTNBP2(23),SLC4A2(23),NEURL3(22),PMEP A1(22),AQP1(21),CFTR(20),RHOV(20),RAPSN(19),SCD5(19),CHST3(18),FAM117B(18),SFRP5(18),TGFB2(18),IGFL1(17),SLC34A2(17),CTTNBP2NL(17),DCDC2(17),NEO1(17),BAZ1B(17),EIF4H(17),SUCNR1(17),ATP8A2(16),MITF(16),IAPP(16),RAB36(16),ANXA4(16),CUL4A(16),FGFR3(16),DPPA2(15),HUNK(15),PKHD1(15),ITIH5(15),REPIN1(15),TCF12(15),ZMZ1(15),PHOX2A(14),B3GNT7(14),TNF(14),SPNS2(14),TMEM127(14),ZDHHC1(14),ARL14(13),CEACAM7(13),TMEM72(13),SCTR(13),GRM8(13),EEF2(13),ZNF317(13)
M12	887	KIAA0319(44),ANXA8(40),ATP1A1(35),LCA5(30),PDGFD(30),KCNH3(25),HNF1B(25),ITGB1(25),VTCN1(24),WDR91(24),CITED1(23),CTTNBP2(23),SLC4A2(23),NEURL3(22),PMEP A1(22),AQP1(21),CFTR(20),RHOV(20),RAPSN(19),SCD5(19),CHST3(18),FAM117B(18),SFRP5(18),TGFB2(18),IGFL1(17),SLC34A2(17),CTTNBP2NL(17),DCDC2(17),NEO1(17),BAZ1B(17),EIF4H(17),SUCNR1(17),ATP8A2(16),MITF(16),IAPP(16),RAB36(16),ANXA4(16),CUL4A(16),FGFR3(16),DPPA2(15),HUNK(15),PKHD1(15),ITIH5(15),REPIN1(15),TCF12(15),ZMZ1(15),PHOX2A(14),B3GNT7(14),TNF(14),SPNS2(14),TMEM127(14),ZDHHC1(14),ARL14(13),CEACAM7(13),TMEM72(13),SCTR(13),GRM8(13),EEF2(13),ZNF317(13)
M13	122	GPR50(29),HDGFL1(26),KLK6(24),COL11A1(20),GALNT6(14),DHX9(45),HNRNPU(35),HNRNPA1(34),JLF3(26),PTBP1(26),TOMM20(26),KHSRP(25),EIF4G2(24),CCT6A(20),HNRNPK(20),HNRNPU(20),TPR(19),ADSL(18),LARP1(16),XRCC6(16),BAZ2A(15),RP L6(14),DRG1(14),COPB1(14),KCNJ4(13),CBX3(13),KDM3B(13),HCTF1(13)
M14	407	GALR2(50),ZNF774(43),ZNF19(39),ZNF492(36),NCAN(35),ZNF778(33),FBL(31),TXNRD1(31),ZNF20(30),SNRPD2(30),TKT(28),HPS E2(26),ZNF519(26),ZNF625(26),ZNF658(26),NCL(25),ZNF423(24),ZNF442(24),GEMIN7(23),CRH(21),CA3(21),ZNF486(21),NHP2(21),UBE2J2(21),ZNF713(19),RPS3(19),CDC20B(18),TRIM28(18),ZP4(17),ZPBP2(17),RPS21(17),MT1IP(17),AURKAIP1(17),RAN(17),PIR(17),RPS8(17),TIMM50(17),NCCR P1(16),RBPJL(16),HHLA2(16),TKTL1(16),ZNF90(16),MAK16(15),ZNF223(15),ZNF235(15),ZNF549(15),NOP56(15),SYMPK(15),ZCCHC3(15),ZNF732(14),HOXD13(14),ZNF551(14),TPD52L2(14),IK(14),LSM14B(14),DPH2(14),MATN4(13),DMWD(13),RPS12(13),BNIPL(13),LSM14A(13),UBA2(13),AKR1B10(13),SSU72(13),PPIH(13),MIIP(13),YTHDF1(13),SDF4(13),RPL38(66),MRPL27(60),ZBTB8OS(46),MRPL52(44),C7orf61(39),RPL24(39),MRPL22(38),RPL37A(37),NDUFA7(36),SEC24C(36),MRPL47(35),SDHAF2(35),SNRPF(35),BLOC1S1(35),ARID3A(32),RPL32(31),VPS52(29),SHARPIN(28),RPL36A(28),RPL31(28),SNRPG(28),RPL37(27),BRPF3(27),LSM3(27),GPS1(26),RPS27A(25),SMARCC2(25),SNORD91B(24),HSPB11(23),POLR2H(22),POLR2J(22),RPL35(22),PSMA2(22),EIF3A(22),EP300(22),ARID1A(22),SNHG6(21),RPL35A(21),SEC61G(21),DYNLRB1(21),PSMC5(21),SSB(21),PRPF31(21),PCBP1(21),CHST13(21),FGF17(20),RPL27A(20),RPL34(20),NUDT1(20),SCRIB(20),ARPC3(20),NEDD8(20),UBL4A(20),ENY2(19),CCT4(19),NOP58(19),NOSIP(19),ANAPC11(19),PEX2(19),STX8(19),SNRPD1(19),ZNHIT1(19),CAGE1(18),RPL14(18),RPL23(18),NME1(18),ARMC1(18),MEA1(18),HGS(18),UQCR11(18),SEMA4G(18),
M15	1141	
M16	2609	

		TPRKB(18),TOM1L1(18),RPL18(17),NDUFA2(17),EXOSC4(17),NAA38(17),BCCIP(17),VPS28(17),SNRPC(17),FAM122B(17),EHMT2(17),MRPL45(17),PSMD10(17),MED11(17),HCFC1(17),TIMM9(17),SNRNP200(17),POLR2K(16),RPS24(16),RPS10(16),NDUFA9(16),MRPS7(16),CSNK2B(16),CCT7(16),PPP2R5D(16),PSMA4(16),NDUFB8(16),TTC1(16),HIC2(16),SF3B5(16),GOPC(16),ZNF770(16),CYHR1(15),MRPL13(15),BOP1(15),COPS6(15),CPNE3(15),RP9P(15),RPS18(15),PABPC1(15),DNTTIP2(15),ANAPC7(15),SSNA1(15),DHX16(15),AARS2(15),MRPS33(15),REXO4(15),UBR5(15),SERINC1(15),PIK3CG(14),PUF60(14),DUS1L(14),HSF1(14),PFDN4(14),NDUFB1(14),PTRH2(14),MPDU1(14),CHCHD1(14),ATP6AP1(14),RPS23(14),RPS29(14),FIS1(14),SNRPE(14),FARSB(14),VBP1(14),TATDN1(14),SS18L2(14),CHCHD5(14),NSMCE1(14),NUP43(14),UBE2F(14),ANKRD17(14),PSMD13(14),HNF1A(14),RPF1(14),TRIM71(14),YWHAZ(14),NUDCD1(13),DENR(13),SNRPB2(13),NDUFA13(13),COX16(13),KATNA1(13),NDUFA11(13),PFDN5(13),DPM2(13),UBE2A(13),ZNF444(13),WDR46(13),EXOSC1(13),MCTS1(13),PCYT2(13),MED6(13),CPSF3(13),HRAS(13),PIH1D1(13),NDUFS6(13),TIMM8B(13),WAC(13),NMT1(13),SAP30BP(13),NOP16(13),PPAN(13),RFNG(13),RPS27(13),PSMA5(13)
M17	507	SCAMP3(37),FNBP4(29),PI4KB(27),RBM39(26),SF3B4(26),TIA1(24),ISG20L2(22),PRCC(22),PYGO2(22),CNIH4(22),ADNP(20),SALL3(18),SETDB1(18),CLK2(17),PABPC1L(17),ZNF256(17),UBAP2L(17),HAX1(16),ASXL1(16),AGAP6(15),TRAF7(15),LUC7L3(15),CREBFZ(15),ZNF606(15),TTC13(15),MRPS21(14),GSDMB(14),VPS72(14)
M18	366	GRIK1(47),MIA(41),NKAIN4(28),PLAC4(24),MUCL1(23),PTPRR(22),BIRC7(21),SCEL(20),MSLN(18),TMPRSS4(18),C3orf70(17),PAX2(17),DCST2(17),LGALS9B(16),BCL2L15(15),GALNT3(15),ERN2(14),FAM83E(14),GPRC5A(14),CKMT1B(14),ATP8A1(14),MUC12(14),SOD3(14),SFTA2(14),VTRNA1-2(14)
M19	895	KIF26A(43),KLK10(39),CDC42BPG(37),PFKP(30),TMC4(29),ITGB4(28),SEL1L3(28),FAM167A(27),SPINT2(25),SPINT1(25),ELOVL7(24),TMEM132A(24),SLC16A14(24),AGRN(23),SFMBT2(22),DSG2(22),THSD4(22),SLC25A15(22),ATP1A3(21),MAPRE1(20),ALDOA(20),CTNND1(20),IQGAP1(19),DDR1(18),PDE3A(18),CHMP4B(18),CAPN9(17),OTOP3(17),ARG2(17),EPCAM(17),AKAP12(17),SYNR3(17),VSIG1(16),C2orf15(16),MOS(16),APLP1(16),PDE1C(16),ESRP1(16),CHRNB3(15),MSTN(15),C2CD4C(15),C9orf57(15),IRF4(15),SHANK2(15),MGAT5(15),SULT2B1(15),SERHL2(15),TMC5(15),DPP10(14),SOX1(14),AHNAK2(14),BHLHA15(14),PCDHGB5(14),RAB25(14),NCEH1(14),PLEKHB2(14),ZDHHC5(14),WNT1(14),ZPLD1(14),SLC22A8(13),SYT13(13),UNC5D(13),GANAB(13)
M20	76	COLEC10(22),CLEC4G(19),ANGPTL6(18),FCN2(17),OIT3(15)

Table 7. Embryonic stem cell-related geneset from MSigDB

2	gene set	collection	genes
	BENPORATH_ES_1	c2	379
	BENPORATH_ES_2	c2	40
	BENPORATH_ES_CORE_NINE	c2	9
	BENPORATH_ES_CORE_NINE_CORRELATED	c2	100
17	BENPORATH_ES_WITH_H3K27ME3	c2	1114
	MEISSNER_ES_ICP_WITH_H3K4ME3	c2	31
	MEISSNER_ES_ICP_WITH_H3K4ME3_AND_H3K27ME3	c2	14

2	MIKKELSEN_ES_HCP_WITH_H3_UNMETHYLATED	c2	15	66
	MIKKELSEN_ES_HCP_WITH_H3K27ME3	c2		41
	MIKKELSEN_ES_ICP_WITH_H3K27ME3	c2		43
	MIKKELSEN_ES_ICP_WITH_H3K4ME3	c2		733
	MIKKELSEN_ES_ICP_WITH_H3K4ME3_AND_H3K27ME3	c2		136
	MIKKELSEN_ES_LCP_WITH_H3K27ME3	c2		14
	MIKKELSEN_ES_LCP_WITH_H3K4ME3	c2		145
	MIKKELSEN_ES_LCP_WITH_H3K4ME3_AND_H3K27ME3	c2		7
	TESAR_ALK_AND_JAK_TARGETS_MOUSE_ES_D4_DN	c2		6
	TESAR_ALK_AND_JAK_TARGETS_MOUSE_ES_D4_UP	c2		5
	TESAR_ALK_TARGETS_HUMAN_ES_4D_DN	c2		6
	TESAR_ALK_TARGETS_HUMAN_ES_5D_DN	c2		7
	TESAR_ALK_TARGETS_HUMAN_ES_5D_UP	c2		5
	TESAR_JAK_TARGETS_MOUSE_ES_D3_DN	c2		8
	WP_LET7_INHIBITION_C1_ES_CELL_REPROGRAMMING	c2		15
	BHATTACHARYA_EMBRYONIC_STEM_CELL	c2		84
	WONG_EMBRYONIC_STEM_CELL_CORE	c2		335

Table 8. Mesenchymal stem cells-related geneset from MSigDB

gene set	collection	genes
GOBP_MESENCHYMAL_STEM_CELL_DIFFERENTIATION	c5	9
GOBP_MESENCHYMAL_STEM_CELL_MAINTENANCE_INVOLVED_IN_NEPHRON_MORPHOGENESIS	c5	5
GOBP_MESENCHYMAL_STEM_CELL_PROLIFERATION	c5	8
GOBP_REGULATION_OF_MESENCHYMAL_STEM_CELL_DIFFERENTIATION	c5	6
BOQUEST_STEM_CELL_CULTURED_VS_FRESH_DN	c2	31
BOQUEST_STEM_CELL_CULTURED_VS_FRESH_UP	c2	427
BOQUEST_STEM_CELL_DN	c2	218
BOQUEST_STEM_CELL_UP	c2	261
IZADPANAH_STEM_CELL_ADIPOSE_VS_BONE_DN	c2	107
IZADPANAH_STEM_CELL_ADIPOSE_VS_BONE_UP	c2	128
GOBP_POSITIVE_REGULATION_OF_MESENCHYMAL_STEM_CELL_PROLIFERATION	c5	5

Table 9. Identify modules related to stem cells

Module	gene set	pvalue	categories	rank
M11	WONG_EMBRYONIC_STEM_CELL_CORE	1.50E-36	ESC	1
M10	BOQUEST_STEM_CELL_DN	4.97E-16	MSC	2
M2	BOQUEST_STEM_CELL_DN	2.29E-10	MSC	3
M11	BHATTACHARYA_EMBRYONIC_STEM_CELL	1.53E-08	ESC	4
M2	WONG_EMBRYONIC_STEM_CELL_CORE	1.49E-07	ESC	5
M10	BOQUEST_STEM_CELL_UP	3.95E-07	MSC	6
M16	WONG_EMBRYONIC_STEM_CELL_CORE	1.03E-05	ESC	7
M11	JENPORATH_ES_CORE_NINE_CORRELATED	9.55E-05	ESC	8
M10	BOQUEST_STEM_CELL_CULTURED_VS_FRE_SH_UP	2.90E-04	MSC	9
M7	BOQUEST_STEM_CELL_CULTURED_VS_FRE_SH_UP	3.23E-04	MSC	10
M4	WONG_EMBRYONIC_STEM_CELL_CORE	3.44E-04	ESC	11
M20	BOQUEST_STEM_CELL_CULTURED_VS_FRE	7.96E-04	MSC	12

M20	S ₂₀ DN BOQUEST_STEM_CELL_UP	6 8.78E-04	16 MSC	13
M7	IZADPANAH_STEM_CELLADIPOSE_VS_BON E_DN	1.09E-03	MSC	14
M3	GOBP_MESENCHYMAL_STEM_CELL_DIFFER ENTIATION ①	1.40E-03	MSC	15
M2	BHATTACHARYA_EMBRYONIC_STEM_CELL	1.62E-03	ESC	16
M2	BOQUEST_STEM_CELL_UP	1.87E-03	MSC	17
M12	GOBP_MESENCHYMAL_STEM_CELL_DIFFER ENTIATION	2.16E-03	MSC	18
M10	IZADPANAH_STEM_CELLADIPOSE_VS_BON E_UP	2.98E-03	MSC	19
M8	MIKKELSEN_ES_HCP_WITH_H3K27ME3	3.50E-03	ESC	20
M2	BENPORATH_ES_1	4.42E-03	ESC	21
M14	① WONG_EMBRYONIC_STEM_CELL_CORE	5.11E-03	ESC	22
M18	BOQUEST_STEM_CELL_CULTURED_VS_FRE SH_UP	5.52E-03	MSC	23
M2	BOQUEST_STEM_CELL_CULTURED_VS_FRE SH_UP	6.15E-03	MSC	24
M2	BENPORATH_ES_CORE_NINE_CORRELATED	1.14E-02	ESC	25
M7	BOQUEST_STEM_CELL_CULTURED_VS_FRE SH_DN	1.15E-02	MSC	26
M3	GOBP_REGULATION_OF_MESENCHYMAL_S TEM_CELL_DIFFERENTIATION	2.07E-02	MSC	27
M10	IZADPANAH_STEM_CELLADIPOSE_VS_BON E_DN	2.66E-02	MSC	28
M15	BENPORATH_ES_CORE_NINE_CORRELATED	3.05E-02	ESC	29
M12	BOQUEST_STEM_CELL_UP	3.6E-02	MSC	30
M20	BOQUEST_STEM_CELL_DN	4.03E-02	MSC	31
M11	BENPORATH_ES_2	4.37E-02	ESC	32
M8	IZADPANAH_STEM_CELLADIPOSE_VS_BON E_UP	4.73E-02	MSC	33
M10	MIKKELSEN_ES_LCP_WITH_H3K4ME3	4.84E-02	ESC	34

To optimize the modules that related to stem cells, we download 24 embryonic stem cells (ESCs) and 11 mesenchymal stem cells (MSCs) gene list from GSEA MSigDB database. Hypergeometric distribution test is applied for enrichment analysis of top 20 significant modules and each gene list.

The randomForest R package is used for constructing the regulatory network. There are two parameters in our algorithm: 1. The number of decision tree (10,000), and 2. The number of candidate input genes that can be selected for each node (the square root of the numbers of all input genes).

Bibliography

- Ovariectomy in female C57BL/6 mice leads to a dynamic alteration of the gut microbiota associated with obesity and intestinal inflammation. Hui Zhao, Qi Wang, Liqiu Hu, Shaojun Xing, Hui Gong, **Zhe Liu**, Panpan Qin, Jihui Du, Wen Ai, Songlin Peng, Yifan Li. European Journal of Epidemiology. (reback to the journal)

2. Comprehensive identification of key genes and latent pathways of thyroid cancer by integrated bioinformatics analysis. Lei Wang, Liu Xudong, Yu Zhang, **Zhe Liu**, Ling Xu, Ying Ma, Bohang Zhang, Linlin Zhao, Meng Wu, Jinjin Xie, Xinqiang Song, Yuxuan Chen, Dayun Feng, Zhiguo Feng. Frontiers in Oncology. (Under the 4th reviewer)
3. Identification of Hub Genes in Colorectal Cancer by integrating WGCNA and Clinical information from TCGA. Yu Zhang, Jia Luo, **Zhe Liu**, Xudong Liu, Ying Ma, Bohang Zhang, Yuxuan Chen, Xiaofei Li, Zhiguo Feng, Ningning Yang, Dayun Feng, Xinqiang Song. Frontiers in Oncology. (revision of the research manuscript to the company)

QR_liuzhe.docx

ORIGINALITY REPORT



PRIMARY SOURCES

1	funcpred.com Internet Source	1 %
2	software.broadinstitute.org Internet Source	1 %
3	www.ncbi.nlm.nih.gov Internet Source	1 %
4	Noam Zuela-Sopilniak, Jehudith Dorfman, Yosef Gruenbaum. "Molecular dissection of the cellular response to mechanical stress", Cold Spring Harbor Laboratory, 2019 Publication	1 %
5	Submitted to University of Hong Kong Student Paper	1 %
6	mcf.gsfc.nasa.gov Internet Source	<1 %
7	www.nature.com Internet Source	<1 %
8	Submitted to University of Bristol Student Paper	<1 %

- 9 www.hindawi.com Internet Source <1 %
- 10 Submitted to Griffith University Student Paper <1 %
- 11 Muhammad Jamal, Abdul Saboor Khan, Hina Iqbal Bangash, Tian Xie et al. "Integrated computational analysis revealed hub genes in lung adenocarcinoma", Research Square, 2021 <1 %
Publication
- 12 François Cardarelli. "Encyclopaedia of Scientific Units, Weights and Measures", Springer Science and Business Media LLC, 2003 <1 %
Publication
- 13 dipot.ulb.ac.be Internet Source <1 %
- 14 Z. Wang, K. Sakthivel, X. Jin, K. Kim. "3D bioprinting composite tissue", Elsevier BV, 2018 <1 %
Publication
- 15 manualzz.com Internet Source <1 %
- 16 S. SALTISSI, A. CROWTHER, C. BYRNE, S. CLARKE, B. S. JENKINS, M. M. WEBB-PEPLOE. "The effects of oral digoxin therapy in primary <1 %

mitral leaflet prolapse", European Heart Journal, 1983

Publication

-
- 17 pathways.mcdb.ucla.edu <1 %
Internet Source
-
- 18 Submitted to Citrus Community College <1 %
Student Paper
-
- 19 "Gravity, Geoid and Marine Geodesy", <1 %
Springer Nature, 1997
Publication
-
- 20 L Cantini, G Bertoli, C Cava, T Dubois, A <1 %
Zinovyev, M Caselle, I Castiglioni, E Barillot, L
Martignetti. "Identification of microRNA
clusters cooperatively acting on Epithelial to
Mesenchymal Transition in Triple Negative
Breast Cancer", Cold Spring Harbor
Laboratory, 2018
Publication
-
- 21 August Yue Huang, Pengpeng Li, Rachel E. <1 %
Rodin, Sonia N. Kim et al. "Parallel RNA and
DNA analysis after Deep-sequencing (PRDD-
seq) reveals cell type specific lineage patterns
in human brain", Cold Spring Harbor
Laboratory, 2020
Publication
-
- 22 Mariaceleste Aragona, Alejandro Sifrim, Milan <1 %
Malfait, Yura Song et al. "Mechanisms of

stretch-mediated skin expansion at single-cell resolution", Nature, 2020

Publication

- 23 Xiaohui Zhao, Chao Jiang, Rui Xu, Qingnan Liu, Guanglin Liu, Yan Zhang. "TRIP6 enhances stemness property of breast cancer cells through activation of Wnt/β-catenin", Cancer Cell International, 2020 <1 %
- Publication
-
- 24 Ying-Jian Song, Juan Tan, Xin-Huai Gao, Li-Xin Wang. "Integrated analysis reveals key genes with prognostic value in lung adenocarcinoma", Cancer Management and Research, 2018 <1 %
- Publication
-

Exclude quotes

On

Exclude bibliography

On

Exclude matches

< 5 words