Article

Explore the potential mechanism of bufadienolides-like chemicals on breast cancer through Bioinformatics analysis

Yingbo Zhang[1, 4, #], Xiaomin Tang[2, #], Yuxin Pang[2, 3, *], Luqi Huang[3, *], Dan Wang[1, 4], Chao Yuan[1, 4] and Xuan Hu[1, 4]

[1] Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou 571737, P. R. China
[2] Guangdong Pharmaceutical University，Guangzhou 510006, P. R. China;
[3] National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, P. R. China;
[4] Hainan Provincial Engineering Research Center for Blumea Balsamifera, Danzhou 571737, P. R. China.
[#] The author had the same contribution to this work
[*] Correspondence: Yuxin Pang, pyxmarx@126.com; Tel.: +86-898-2330-0268; Luqi Huang, huangluqi01@126.com

Abstract: Bufadienolides-like chemicals, which mostly composed the active ingredient of Chansu, had been widely discovered to possess anti-inflammatory, tumor-suppressing and antipain activity, but the mechanisms of action were not clearly illuminated. In this research, in order to explore the potential mechanism of bufadienolides-like chemicals on breast cancer, a serious of bioinformatics analysis, included (1) differentially expressed genes identification combined with gene set variation analysis, (2) tissue specific co-expression network construction, (3) differentially regulated sub-networks detection with disease phenome, (4) hub gene selection and it's relation to survival probability, and (5) similar small molecule detection were performed with gene expression profiles of bufadienolides-like chemicals. Results indicated bufadienolides-like chemicals had the most same target with valproic, estradiol and etc, could disturbed the pathways in RNA splicing, apoptotic process, cell migration, extracellular matrix organization, adherens junction organization, synaptic transmission, Wnt signaling, AK-STAT signaling, BMP signaling pathway and unfolded protein response, and had the potential ability to be used as anticancer, hormones and vasoprotectives agents.

Keywords: Bufadienolides-like chemicals; Molecular mechanism; Anti-cancer; Bioinformatics

1. Introduction

Despite considerable efforts to the early diagnosis and treatment in the last decade, Breast cancer is still one of the most common malignancies for female worldwide, representing approximately 22% of women's malignancies that pose a threat to women's health [1-4]. In addition to the improvements of early diagnosis, new chemotherapeutic agents and more effective therapies for the treatment become an essential task for improving the mortality of cancer worldwide. Chinese Traditional medicine, had existed and experienced thousands of years of development, and is one of the important sources for antitumor active components screening. Chansu was one of the most famous traditional Chinese medicine, it has been used for centuries in various aspects, such as anaesthesia, antitumor, anti-inflammation and antiarrhythmia [5-8]. The Chansu , mostly come from the glandular secretion dried product of *Bufo gargarizans* Cantor or *B. melanostictus* Schneider [6], and including with several group of mixture, such as the resibufogenin, bufalin, arenobufagin, cinobufagin, bufotoxin, telocinobufagin, bufotaline, cinobufotalin, etc [5-7](Figure 1).

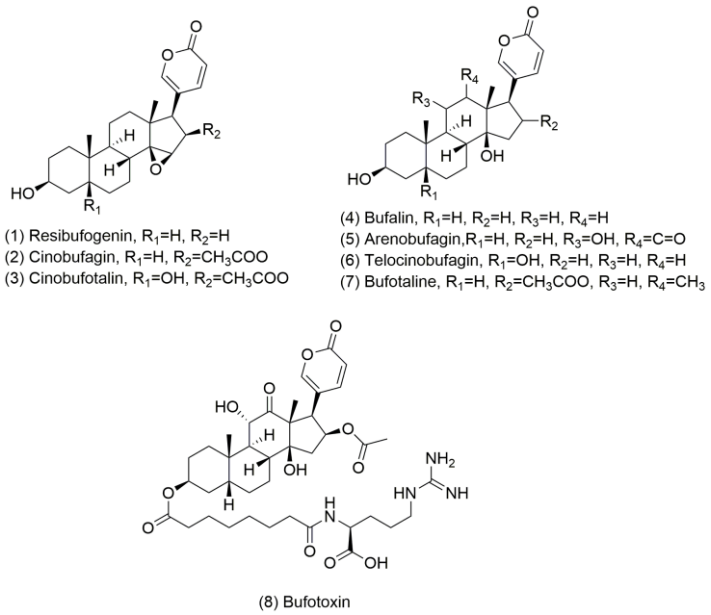


Figure 1. The structural formula of eight bufadienolides-like chemicals

In the last decade, lots of research focused on the pharmacological activities and antitumor activity of bufadienolides-like chemicals. For example, Li et al [9] had reported that cinobufagin has significant cancer-killing capacity for range of cancers, including HCT116 cells, HT29 cells, A431 cells, PC3 cells, A549 cells, and MCF-7 cells, mechanism analysis showed cinobufagin can induced tumor cells apoptosis is likely modulated by the hypoxia-inducing factor-1 alpha subunit (HIF-1α) . Yeh et al [10] and Yu et al [11] had reported that the bufalin and cinobufagin has a potent inhibiting effect on androgen dependent and independent prostate cancer cells, also the same results had been reported by Dong et al [12], Wang et al [13] and Ko et al [14] through HepG2 cells, T24 cells, HeLa cells, and other cells.

These results demonstrate that Chansu is a potent anticancer agent for a range of cancers, but it's potential anticancer mechanisms has been little reported. The increasing public expression profile treat with Chinese Traditional medicine, make it possible to extract the most robust information or potential molecular mechanism from them. In this paper, the gene set variation analysis (GSVA) algorithm [15] was first used for identifying the differentially expressed genes (DEGs) and relative enrichment pathways underlying with eight bufadienolides-like chemicals, and then a serious of bioinformatics analysis, including gene enrichment analysis, tissue specific co-expression network construction, differentially regulated sub-networks detection relate to breast cancer phenome, hub gene selection and it's relation to survival probability, and similar small molecule detection were conducted with the DEGs in the relative enrichment pathways. This work may reveal the potential mechanism of bufadienolides-like chemicals on breast cancer, especially differentially regulated sub-networks relate to breast cancer and hub genes disturbed by bufadienolides-like chemicals, and this work may highlight the potential application of bufadienolides-like chemicals on Breast cancer, especially as a novel agent for cancer therapy.

2. Results

2.1. Identification of DEGs

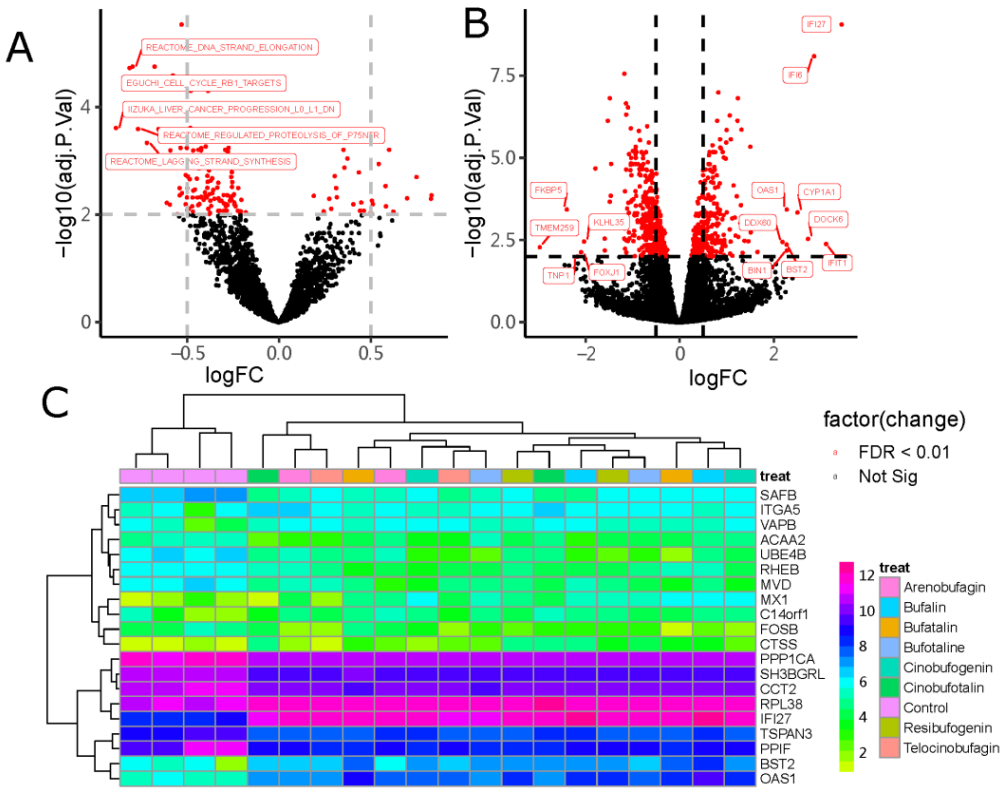


Figure. 2 The DEGs disturbed by bufadienolides-like chemicals through gene set variation analysis (GSVA) algorithm. (A) The differentially expressed gene sets disturbed by bufadienolides-like chemicals (|logFC| ≥ log2(2) and adjPvalue ＜ 0.001). (B) The differentially expressed genes(DEGs) relate to differentially expressed gene sets (|logFC| ≥ log2(2) and adjPvalue ＜ 0.001). (C) The heatmap of top 20 DEGs disturbed by bufadienolides-like chemicals.

Based on the differentially expressed genes analysis associated with gene sets enrichment variation analysis strategy, a total of 80 differentially expressed genes (DEGs) involved in the 44 MSigDB C2 curated gene sets were identified (Fig. 1A and Fig. 1B) , the top 20 DEGs expression heatmap is shown in Fig. 1C. Of which, 38 genes involved in the Singh NFE2L2 targets gene sets, Chang dominant negative gene sets immortalized by HPV31 and Lin silenced gene sets by tumor microenvironment were up-regulated(Table S1 and Table S2), including IF16 (interferon-inducible protein 6), IRF9 (interferon regulatory factor 9), IFIT1 (IFN-induced protein 1 with tetratricopeptide repeats), ISG15 (Interferon-stimulated gene 15), BST2 (bone marrow stromal cell antigen 2), OAS3 (2'-5'-oligoadenylate synthetase 3), OAS1 (2'-5'-oligoadenylate synthetase 1), DDX60 (DEAD box polypeptide 60), CYP1A1 (cytochrome P450 1A1), CEACAM6 (carcinoembryonic antigen-related cell adhesion molecule 6), keratin genes KRT81, and so on.

Among the differentially expressed genes associated with enrichment gene sets, 42 genes involved in the 41 gene sets were down-regulated(Table S1 and Table S2), such the genes involved in Iizuka (Table S1) Liver cancer progression pathway, including PPIF (peptidylprolyl isomerase F), TMED2 (transmembrane trafficking protein 2 with emp24 domain), SAFB (scaffold attachment factor B) , SQLE (squalene epoxidase), PICALM (phosphatidylinositol binding clathrin assembly protein), STIP1 (stress-induced phosphoprotein 1), CYB561 (cytochromes b561) , CCT2 (chaperonin 2β with TCP1 domain), the genes involved Thum systolic heart failure pathway, including CCNG2 (cyclin G2), TMED2 (transmembrane emp24 domain trafficking protein 2), FH (fumarate hydratase), TAF9B (ATA-box binding protein associated factor 9b), CCT2 (chaperonin-containing t-complex

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0629.v1

Peer-reviewed version available at Cancers 2019, 11, 91; doi:10.3390/cancers11010091

polypeptide 1 beta), transmembrane receptor NOTCH2, PICALM (subfamily A (MS4A) and CCNL2 (cyclin L2) , also Reactome DNA strand elongation, Reactome regulated proteolysis of P75NTR, and other gene sets were downregulated with logFC form - 0.89 ~ - 0.27.

In order to obtain a biological interpretation of those genes in GO and KEGG pathway functional groups, GO and KEGG enrichment analysis were performed with clueGO plug [16] in Cystoscape [17]. Results indicated, those genes with up-regulated were rich in the terms of type I interferon signaling response to virus, defense to other organism, regulation of viral genome replication and 2'-5'-oligoadenylate synthetase activity, and those activate may cause by IRF9, IFI6, IFI27, ISG15, IFIT1, OAS1 and OAS3 (Fig3a), also the KEGG pathway enrichment analysis those genes could cause the activate of IFN-induced pathway, type II interferon signaling pathway and regulation of protein ISGylation by ISG15 deconjugating enzyme USP18 pathway (Fig3B). Those genes with down-regulated were rich in the terms of protein kinase complex, transcription factor TFTC complex-1, SAGA- complex and cargo loading into vesicle (Fig3A), further KEGG pathway enrichment analysis those may negative the transport of fringe-modified NOTCH to plasma membrane pathway (Fig3B).

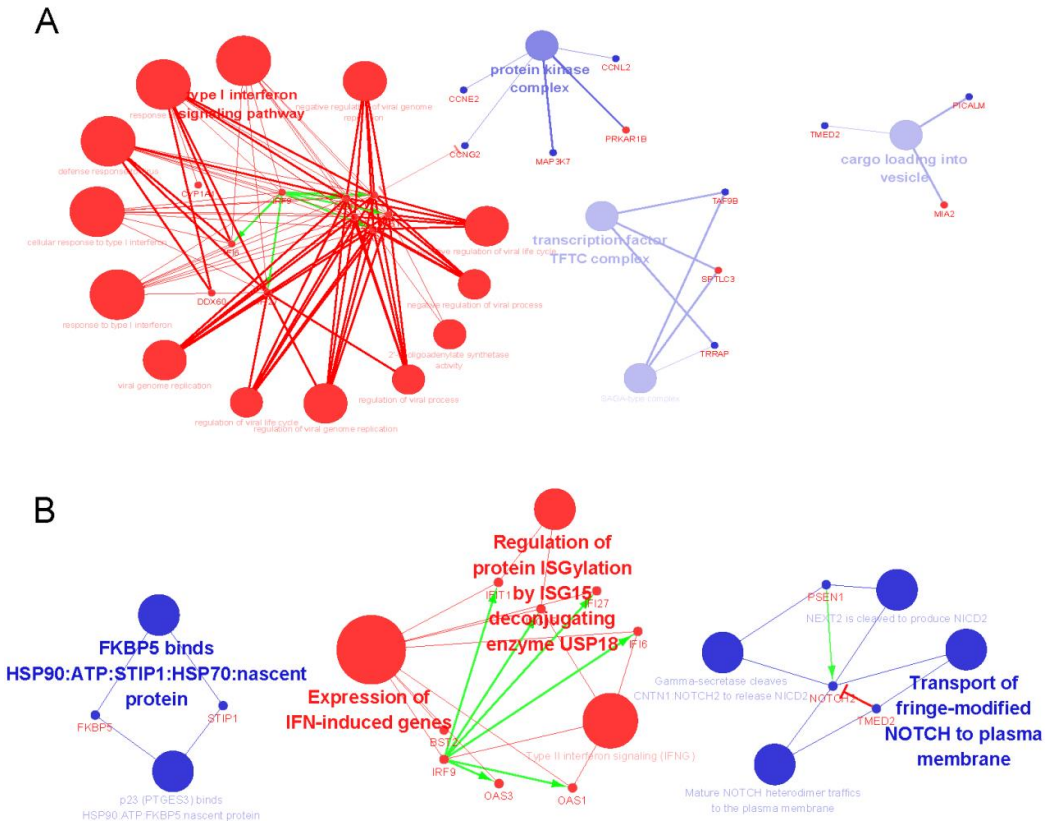


Figure 3. The GO and KEGG enrichment result of DEGs disturbed by bufadienolides-like chemicals. (A) Representative biomolecular network of GO enrichment term, the nodes with red colour and bigger size means the enrichment GO terms with up-regulated genes, the nodes with blue colour and bigger size means the enrichment GO terms with down-regulated genes, the nodes with red colour and smaller size means up-regulated genes, the nodes with blue colour and small size means down-regulated genes, undirected edges means enrichment, the green directed edges means activate from the evidence generated by String database. the red directed edges means suppressive from the evidence generated by String database, (B) Representative biomolecular network of KEGG enrichment term, the nodes and edges also had the same means with Figure 3A.

2.2. The tissue specific co-expression network and breast cancer associated subnetwork regulated by bufadienolides-like chemicals

It is clear that most of genes exert their function by collaborating with other genes in network which represent rigid molecular machines, cellular structures, or dynamic signaling pathways [18]. In this research, in order to comprehensive understanding the potential function of DEGs involed in Breast cancer, a breast tissue specific co-expression network with DEGs were generated with TCSBN database [19] through NetworkAnalyst web serve [20]. Results indicated the co-expression networks were consisted of 743 nodes and 876 edges (Figure 4 and Table1). Furthermore, a functional enrichment analysis with KEGG pathways revealed that the co-expression networks network with DEGs were enriched in pathways related to tight junction, PPAR signaling pathway, mTOR signaling pathway, influenza A, tuberculosis, N-Glycan biosynthesis, terpenoid backbone biosynthesis, Notch signaling pathway, regulation of cyclin-dependent protein kinase activity and steroid biosynthesis (Table 1). Also the GO BP term enrich analysis, those genes mostly involved in Establishment or maintenance of cell polarity, triglyceride metabolic process, protein targeting to membrane, defense response to virus, tuberculosis, post-translational protein modification, coenzyme biosynthetic process, gamete generation, transcription, DNA-dependent, positive regulation of translation, endoplasmic reticulum unfolded protein response, regulation of cyclin-dependent protein kinase activity, steroid biosynthetic process, regulation of the transcription of DNA-dependent, intra-Golgi vesicle-mediated transport term., and other rigid molecular machines in biological process.

Peer-reviewed version available at Cancers 2019, 11, 91; doi:10.3390/cancers11010091

143 **Table 1**. The tissue specific co-expression network regulated by bufadienolides-like chemicals and it's enrichment with GO and KEGG

Subnetwork Number	Nodes	Edges	Seeds	KEGG Enrichment		GO Enrichment	
				KEGG Pathway	P-value	BP term	P-value
A	492	558	13	Tight junction	4.19E-04	Establishment or maintenance of cell polarity	2.83E-04
B	113	128	3	PPAR signaling pathway	7.75E-06	Triglyceride metabolic process	1.25E-07
C	46	50	2	mTOR signaling pathway	9.62E-03	Protein targeting to membrane	4.93E-67
D	27	86	6	Influenza A	3.04E-10	Defense response to virus	1.24E-22
E	18	17	1	Tuberculosis	2.01E-04	Tuberculosis	2.01E-04
F	11	10	1	N-Glycan biosynthesis	9.19E-03	Post-translational protein modification	6.33E-03
G	6	5	1	Terpenoid backbone biosynthesis	1.72E-04	Coenzyme biosynthetic process	1.55E-05
H	5	4	1	Notch signaling pathway	2.98E-02	Gamete generation	1.34E-02
I	4	3	1	NA	NA	Transcription, DNA-dependent	1.31E-02
J	4	3	1	NA	NA	Positive regulation of translation	1.17E-02
K	4	3	1	NA	NA	Endoplasmic reticulum unfolded protein response	6.51E-03
L	4	3	1	Regulation of cyclin-dependent protein kinase activity	1.24E-02	Regulation of cyclin-dependent protein kinase activity	1.24E-02
M	3	2	1	Steroid biosynthesis	7.68E-03	Steroid biosynthetic process	2.07E-06
N	3	2	1	NA	NA	Regulation of transcription, DNA-dependent	1.84E-02
O	3	2	1	NA	NA	Intra-Golgi vesicle-mediated transport	4.47E-03

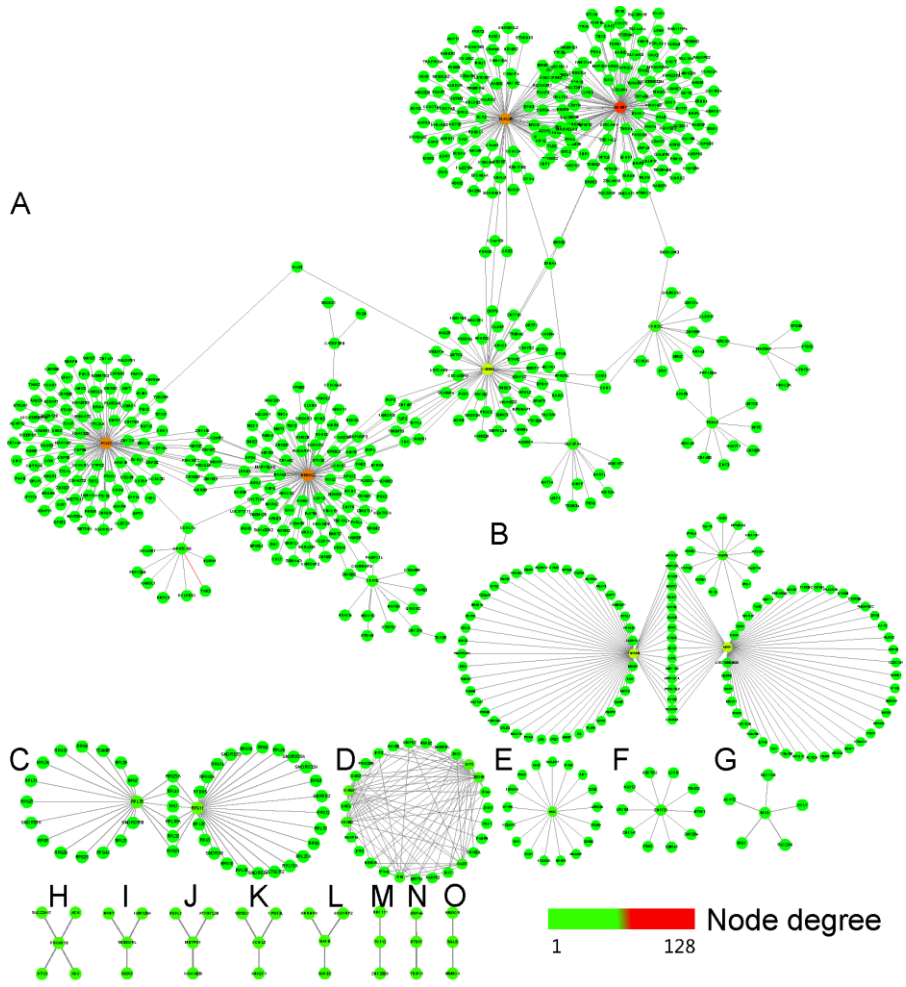


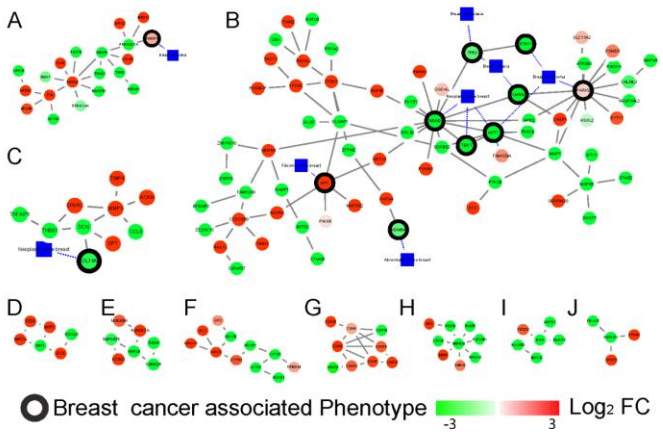
Figure 4. The breast tissue specific co-expression network with DEGs generated by TCSBN database through NetworkAnalyst web serve. (A) ~ (O), the Subnetworks of co-expression network origin from the seeds of DEGs.

Based on the novel differentially regulated sub-networks detection tool, PhenomeScape [21], which could combined the fold changes of genes into the knowledge of networks and disease phenotypes, and then a series of differentially regulated sub-networks associated with phenotypes were identified with random walk algorithm. In this research, 7 phenotypes relate to breast cancer were choosed as the seed phenotypes, subsequently a total of 23 differentially regulated sub-networks enriched in breast cancer phenotype related subnetwork were identified (Table 2). The Sub-networks distributed by bufadienolides-like chemicals included RNA splicing (p-value=2.00E-03), apoptotic process (p-value=2.00E-03), extracellular matrix organization (p-value=1.00E-03), canonical Wnt signaling pathway (p-value=2.20E-02), synaptic transmission (p-value=1.40E-02), negative regulation of JAK-STAT cascade (p-value=4.20E-02), adherens junction organization (p-value=3.80E-02), BMP signaling pathway (p-value=4.10E-02) , negative regulation of cell migration (p-value=1.30E-02), activation of signaling protein activity involved in unfolded protein response (p-value=1.90E-02) (Fig. 4). The subnetwork A (Fig. 5A), relate to the RNA splicing function, was the first identified dysregulation subnetwork, it could observe the genes involved in mRNA splicing spliceosome were down regulated, included the serine and arginine rich splicing factor members SRSF4, SRSF5, SRSF6 and peroxisome proliferator activated receptor gamma coactivator PPARGC1A. The apoptotic process (Fig. 5B), also could been dysregulated by bufadienolides-like chemicals, and this dysregulation were performed with the increase expression of SYT11, PARK2, PYHIN1, APC, RNF40, SERPINB3, TIAM2, ITSN1, SH3GL2, CASP1, GATA4, ITSN2 and PDE4DIP . Several cancer signaling pathway included Wnt signaling pathway, JAK-STAT signaling pathway and BMP signaling pathway also could had been dysregulated by bufadienolides-like chemicals (Fig. 5D, 5F and 5H), this may gave further evidence of bufadienolides-like chemicals could increase the apoptotic process through a series of pathways or regulation network. The Subnetwork C (Fig. 5C), mostly

related to the extracellular matrix organization were upregulated, included the genes TIMP4, MMP3, SPARC, DPT and ACAN, also in this subnetwork the genes referred to the regulation of cell migration were downregulated, included the genes TNFAIP6, DCN, SPARC, THBS1 and CCL8, this means the increase of extracellular matrix may hindered the migration of tumor, also the negative of synaptic transmission, adherens junction organization and regulation of cell migration could find in subnetwork E, G and I (Fig. 5E, 5G and 5I). Several metabolic process also had been discovered, included the drug metabolic process, xenobiotic metabolic process, oligosaccharide metabolic process and etc. All other PhenomeScape networks can be found in Supplementary Fig. 1.

Table 2 Summary of differentially regulated sub-networks disturbed by bufadienolides-like chemicals

Subnetwork number	No. of nodes	GO-BP	Empirical P-value
A	21	RNA splicing	2.00E-03
B	73	apoptotic process	2.00E-03
C	11	extracellular matrix organization	1.00E-03
D	6	canonical Wnt signaling pathway	2.20E-02
E	7	synaptic transmission	1.40E-02
F	11	negative regulation of JAK-STAT cascade	4.20E-02
G	9	adherens junction organization	3.80E-02
H	9	BMP signaling pathway	4.10E-02
I	6	negative regulation of cell migration	1.30E-02
J	4	activation of signaling protein activity involved in unfolded protein response	1.90E-02
K	12	drug metabolic process	1.20E-02
L	6	negative regulation of lipid storage	4.50E-02
M	6	xenobiotic metabolic process	1.70E-02
N	8	relaxation of cardiac muscle	4.80E-02
O	5	very long-chain fatty acid metabolic process	1.70E-02
P	4	oligosaccharide metabolic process	3.10E-02
Q	4	collagen catabolic process	2.50E-02
R	4	response to cocaine	2.70E-02
S	4	behavioral response to nicotine	4.20E-02



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0629.v1

Peer-reviewed version available at Cancers 2019, 11, 91; doi:10.3390/cancers11010091

Figure 5. The differentially expressed networks regulated by bufadienolides-like chemicals, and generated by PhenomeScape plug. Sub-networks linked to breast cancer, RNA splicing(2.00E-03)(A), apoptotic process(2.00E-03)(B), extracellular matrix organization(1.00E-03)(C), canonical Wnt signaling pathway(2.20E-02)(D), synaptic transmission(1.40E-02)(E), negative regulation of JAK-STAT cascade(4.20E-02)(F), adherens junction organization(3.80E-02)(G), BMP signaling pathway(4.10E-02) (H) , negative regulation of cell migration(1.30E-02) (I), activation of signaling protein activity involved in unfolded protein response(1.90E-02) (J). The fold change of the proteins is shown by the node colour and breast cancer associated phenotype annotated proteins used to generate the sub-networks are shown with a black border.

Hub genes, mostly the highly connected nodes in network, were identified by node degree and MCC algorithm with Cytoscape plugin cytoHubba [22]. Based on the threshold of degree (degree > 5) and MCC algorithm, 10 genes with MCC scores ranged from 126~953 were identified as hub genes (Figure 6A, Figure 6B). Among the 10 hub genes, included 3 2'-5'-oligoadenylate synthetase genes OAS1, OAS2 and OAS3, included 5 interferon-induced genes ISG15, IFIT1, IFI6, IFI44 and IFIL44L, also two other genes included the kelch-like family member 35 (KLHL35) and Golgi Membrane Protein 1 (GOLM1) were selected as the hub genes. Further investigated with TCGA [23] and Kaplan-Meier database [24] indicates, 10 hub genes except KLHL35, were increased both in treat with bufadienolides-like chemicals and TCGA Breast cancer sample (Figure 6C) , 6 hub genes, included IFIT1, ISG15, IFI6, GOLM5, KLHL35 and OAS2 were associated the total survival probability in Breast cancer patients (Figure 6D). Further analyzed with the correlation between the hub genes and total survival time in Breast cancer, the high expression of GOLM5, KLHL35 and OAS2 were associated with better survival probability.

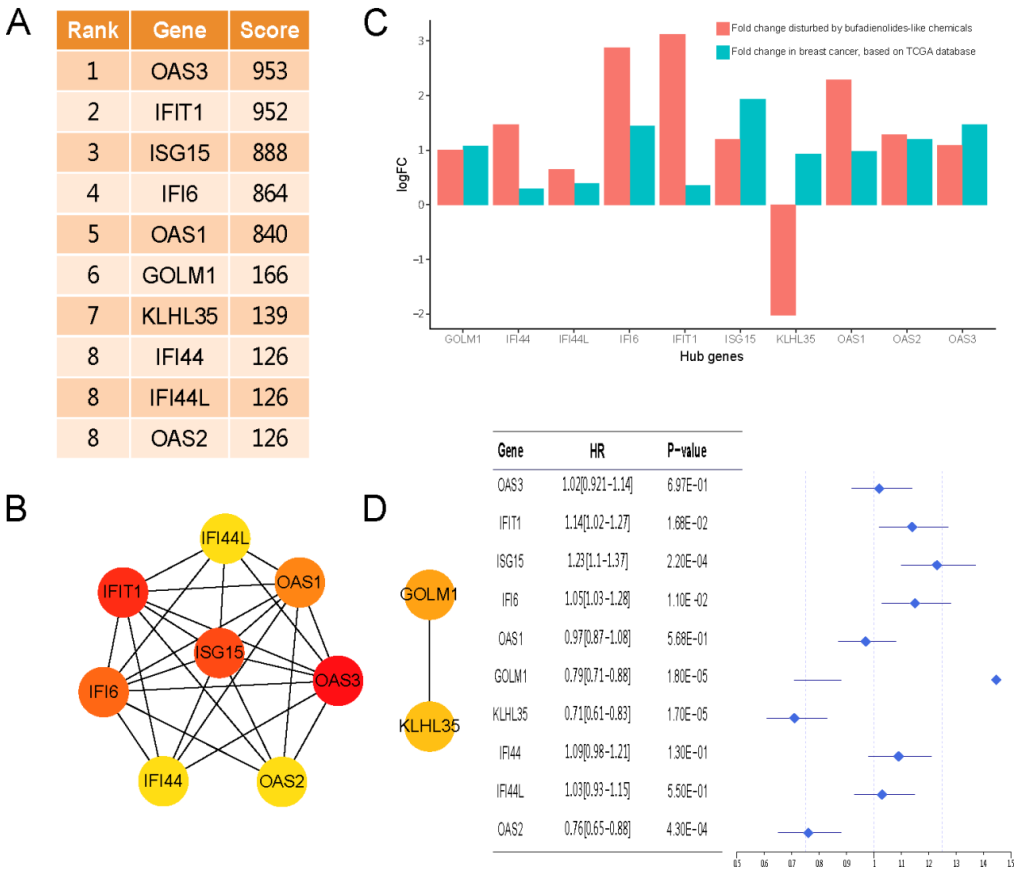


Figure. 6 The 10 hub genes and it's correlation with total survival probability in Breast cancer. (A) The 10 hub genes and it's MCC score. B) The network of hub genes. C) The expression correlation with breast cancer, validated by TCGA database. D) The total survival probability correlation with breast cancer, validated by Kaplan-Meier (KM) plotter database.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0629.v1

Peer-reviewed version available at Cancers 2019, 11, 91; doi:10.3390/cancers11010091

2.3. Similar small molecule detection

Detect the similar small molecule with Comparative Toxicogenomics Database (CTD) [25] and connectivity map (CMAP2) [26, 27] database will provide a better understanding the molecular mechanism of bufadienolides-like chemicals, and its potential value of novel agent for cancer therapy. Based on the results with detecting CTD Database, valproic, cyclospoine and estradiol had the most same target with bufadienolides-like chemicals (Figure 7). Valproic, a histone deacetylase inhibitor, which once had been widely used as antiepileptic, and recently also had been proved with anti-cancer activity in vitro/vivo model [28] [29]. Estradiol, a sex hormones with anticancer activity, and also widely used for the treatment of Breast cancer, especially the postmenopausal women [30-32].

Based on the results from connectivity map (CMAP2) database (Table2) [26, 27], the type of V03AF, G03GB, C05AX and C05CX were the top matching drugs with bufadienolides-like chemicals, V03AF, one type of detoxifying agents for antineoplastic treatment, had an opposing effect on expression of the bufadienolides-like chemicals, this result indicated the evidence of bufadienolides-like chemicals had an potential value of novel agent for cancer therapy. G03GB, one type of sex hormones and modulators of the genial system, had the mostly same expression profile with bufadienolides-like chemicals, this means the bufadienolides-like chemicals had the same use of estradiol, epimestrol, cyclofenil in Breast cancer. C05AX and C05CX, two types of vasoprotectives agents, this means the bufadienolides-like chemicals also had the potential use of vasoprotectives-like drugs.

From the evidence from detecting the similar small molecule with CTD database and CMAP2 database, indicated bufadienolides-like chemicals were one kinds of steroids with the same physiological activity as estradiol and G03GB (ATC code), had the potential value for cancer, especially the Breast cancer.

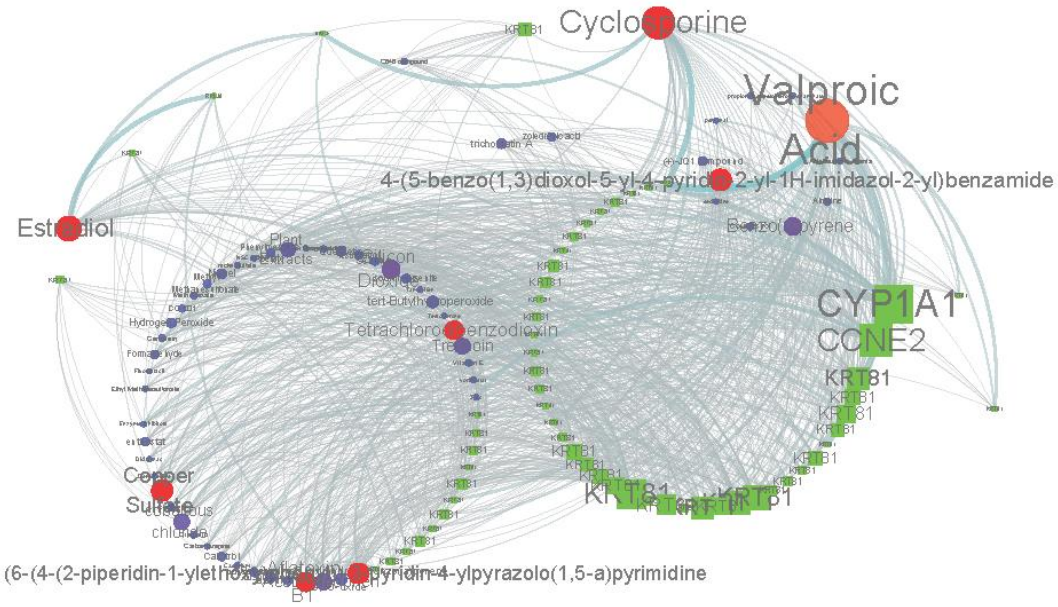


Figure 7. Chemicals-gene interaction network for the DEGs disturbed by bufadienolides-like chemicals. Square nodes represent for the DEGs. Circle nodes represent for the chemicals predicted by Comparative Toxicogenomics Database. The size of nodes represent for the degree. Circle nodes with red represent the Similar small molecule predicted by degree (degree ≥ 30)

Table 3. Top 20 CMAP hits correlated with bufadienolides-like chemicals treatment

Rank	ATC code	Mean score	enrichment	p-value	specificity
1	V03AF	-0.471	-0.71	4.45E-03	3.82E-02
2	G03GB	0.449	0.655	3.29E-02	7.47E-02
3	C05AX	0.41	0.689	1.95E-02	4.76E-02
4	C05CX	0.41	0.689	1.95E-02	4.76E-02
5	D07XC	-0.372	-0.661	1.44E-03	8.10E-03
6	N05BE	-0.359	-0.719	1.26E-02	1.22E-02
7	C08EA	0.292	0.539	1.87E-02	1.45E-01
8	N05AC	0.259	0.365	2.32E-03	3.90E-01
9	D06BB	-0.252	-0.405	9.39E-03	1.44E-01
10	D06BX	-0.249	-0.72	3.74E-03	1.38E-02
11	N02BB	0.244	0.404	2.71E-03	1.75E-02
12	N02CX	0.189	0.481	3.16E-02	4.43E-02
13	A07EA	-0.186	-0.343	6.96E-03	2.55E-02
14	S02BA	-0.167	-0.383	5.03E-03	1.31E-02
15	B01AC	0.152	0.243	2.71E-02	1.19E-01
16	S03BA	-0.144	-0.366	2.02E-02	4.80E-02
17	R03BA	-0.141	-0.29	1.19E-02	4.00E-02
18	S01CB	-0.136	-0.326	1.21E-02	2.61E-02
19	R01AD	-0.113	-0.266	4.30E-03	4.83E-02
20	C07AA	-0.109	-0.262	1.14E-02	2.22E-01

3. Discussion

Recently, the gene expression profile technology, included the microarray and RNA-seq, had been widely to detect the potential mechanism of chemicals, but an central problem still perplex the researchers on pharmacology and biology, that is the chemicals how to disturb pathways and phenotypes through gene and its co-expression network. In this research, with use of the bioinformatics tools, especially the differentially regulated sub-networks detection tools PhenomeScape [21], comparative toxicogenomics database [25] and connectivity map [26, 27] database, revealed several differentially regulated sub-networks treat with bufadienolides-like chemicals, also the hub genes in co-expression network and its relation to survival probability of breast cancer, similar small molecule detection and other results may highlight the potential molecular mechanism and application of bufadienolides-like chemicals on cancer, especially as a novel agent for Breast cancer.

First, during the process of differentially expressed genes identification, in contrast to use the conventional method of differentially expressed genes selection with significance in statistics, a non-parametric unsupervised method of gene set variation analysis were used for differentially expressed genes identification. Results indicated a total of 80 differentially expressed genes (DEGs) involved in the 44 MSigDB C2 curated gene sets were identified (Fig. 1A and Fig. 1B). Further analysis with enrichment of GO and KEGG pathway, we found the genes with up-regulated most rich in interferon signaling response to virus, defense to other organism, regulation of viral genome replication and 2'-5'-oligoadenylate synthetase activity, KEGG pathway enrichment analysis showed those genes could cause the activate of IFN-induced pathway, type II interferon signaling pathway and regulation of protein ISGylation. But the genes with down-regulated were rich in the terms of protein kinase complex, transcription factor TFTC complex-1, SAGA- complex and cargo loading into vesicle, KEGG pathway enrichment analysis showed those genes may involve in negative the transport of fringe-modified NOTCH to plasma membrane pathway. Compare the method with statistical significance, those differentially expressed genes in gene set variation maybe much less, but with more same participate in same pathway or biology function, also the same results had been proved by the examples of GSVA package [15].

Second, during the process of co-expression network reconstruction and dysregulated sub-networks detection, a novel plug of PhenomeScape was used, which could combine the data of gene expression into the knowledge of protein–protein interaction networks and disease phenotype [21]. During the analysis with damaged osteoarthritic cartilage gene expression profile, several significant sub-networks related to damaged osteoarthritic cartilage were identified, including mitotic cell cycle, Wnt signalling, apoptosis and matrix organisation [33, 34]. In this research, with PhenomeScape tool [21], a total of 23 differentially regulated sub-networks were identified, and 10 sub-networks had been proved to relate to breast cancer by evidence, included RNA splicing, apoptotic process, cell migration, extracellular matrix organization, adherens junction organization, synaptic transmission and so on.

Third, during the process of similar small molecule detection, Comparative Toxicogenomics Database (CTD) [25] and connectivity map (CMAP2) [26, 27] database were used. Results indicated bufadienolides-like chemicals had the same effect with valproic and estradiol, valproic, a histone deacetylase inhibitor, it had been proved to inhibit proliferation through Wnt/β catenin signalling activation. The estradiol, also had been proved to with anticancer activity, especially the postmenopausal women. Also the evidence form connectivity map database indicated bufadienolides-like chemicals had the potential ability to be used as anticancer, hormones and vasoprotectives agents.

During the hub gene selection and it's relation to survival probability indicated 10 hub genes except KLHL35, were increased both Breast cancer and samples treat with bufadienolides-like chemicals, further analysis with relation to total survival probability, 6 hub genes, included IFIT1, ISG15, IFI6, GOLM5, KLHL35 and OAS2 were associated the total survival time and high expression of GOLM5, KLHL35 and OAS2 were associated with better survival probability.

4. Materials and Methods

4.1 Microarray data information

The gene expression profiles of GSE85871 (https://www.ncbi.nlm.nih.gov/gds/), which is an gene expression profile treat with 102 Chinese traditional medicine, and it was based on Affymetrix GPL571 platform (Affymetrix Human Genome U133A 2.0 Array), was submitted by Lv et al [35].

In this study, the raw data of 4 controls and 14 samples treat with bufadienolides-like chemicals, including resibufogenin, bufalin, arenobufagin, cinobufagin, bufotoxin, telocinobufagin, bufotaline and cinobufotali, were downloaded from GEO database through GEOquery [36] packages in R3.5.1 [37] environment.

296 **4.2 Identification of DEGs associated with relative enrichment pathways**

297 In order to obtain the biological interpretation of differentially expressed genes（DEGs）
298 disturbed by bufadienolides-like chemicals, a novel R package GSVA [15] was employed, which
299 allows the assessment of the DEGs underlying pathway activity variation by transforming the gene
300 expression profile into the prior knowledge of gene set. In accordance with MIAME standards [38,
301 39], the differentially expressed genes (DEGs) disturbed by bufadienolides-like chemicals were
302 identified by a serious of standard flow with R environment. First, the quality assessments,
303 background correction and normalization were preprocessed and normalization with affy [40] and
304 gcrma [41] packages. Then, the batch effects were examined and removed out with combat and sva
305 functions in SVA package [42]. Subsequently, an non-specific probes filtering step were carried out
306 with nsFilter function in the genefilter package [43], the quality control probes of Affymetrix,
307 probesets without Entrez ID annotation, probesets whose associated Entrez ID is duplicated in the
308 annotation and the top 20% with smaller variability were removed. Finally, the GSVA [15],
309 GSEABase [44], limma [45] package and c2BroadSets from Molecular Signatures Database
310 (MSigDB) [46, 47] were used for the selection for DEGs with relative enrichment pathways.
311 During the process of DEGs selection with relative enrichment sets, the gene expression profile
312 was first transformed into the prior knowledge gene set of c2BroadSets and the enrich gene sets
313 were selection with the screening criteria of FDR < 0.01. Then the different genes enrichment in the
314 c2BroadSets gene sets were selected with limma [45] package, and the screening criteria were set
315 with FDR < 0.01 and |logFC| > 1, and those DEGs associated with relative enrichment pathways
316 were used for further analyzed and validated
317 During the process of DEGs identification, the Biobase [48]package and GSVAdata [49] package
318 were also applied. The results were visualized with ggplot2 [50], ggpubr [51], pheatmap [52]and
319 cowplot [53] package.

320 **4.3 Gene enrichment analysis**

321 On the bias of the DEGs selection associated with relative enrichment sets, in order to obtain a
322 comprehensive understanding of those genes involved in the prior knowledge of gene sets, GO and
323 KEGG enrichment analysis were performed with clueGO plug [16] in Cystoscape [17]. The
324 significantly enrich GO terms and KEGG pathways were calculated by the hypergeometric test [54],
325 and cut-off criteria was set as FDR < 0.05. Another statistical parameter of Kappa Score were set as
326 middle stringency, its means the terms in network were combined with middle related terms as
327 based on their overlapping genes. The min percentage and min genes enriched in GO terms or
328 KEGG pathways were set as 1.0% and 2, also the term fusion was chosen. Other options, including
329 the statistical options, reference options, grouping options and visual options were set with default
330 setting.

331

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 November 2018 doi:10.20944/preprints201811.0629.v1

Peer-reviewed version available at Cancers 2019, 11, 91; doi:10.3390/cancers11010091

4.4 Gene co-expression network analysis and disease phenotype association

In order to comprehensive understanding the potential mechanism of DEGs in involed in Breast cancer, co-expression network analysis, phenome association and survival correlation analysis were investigated with NetworkAnalyst database [20] and PhenomeScape plug [21] in Cystoscape [17], also other plugs and databases including the cytoHubba [22] , TCSBN database [19], TCGA database [23] and Kaplan-Meier (KM) plotter database [24] and Phenomiser [55] web tool were also used for hub genes selection and survival correlation analysis. First, the breast mammary tissue specific co-expression networks were investigated with TCSBN database through NetworkAnalyst web server, also the GO and KEGG enrichment terms of networks were also investigated with NetworkAnalyst web server. Subsequently, the differentially regulated sub-networks enriched in genes associated with breast cancer phenotype were identified by random sampling (10,000 sub-networks) methods with PhenomeScape plug and Phenomiser web tool. First, through the search with Phenomiser web tool and the manual of UberPheno ontology [55], six phenotypes (Table 4) were chosen as the breast cancer association phenotype, and then with the parameters of maximum initial sub-network size of 7 and an empirical P-value threshold of 0.05 were used for filtering the differentially regulated sub-networks enriched in genes associated with breast cancer phenotype.

Table 4. UberPheno phenotype terms selected for differentially regulated sub-network detection in co-expression network

Phenotype ID	Phenotype Description
HP:0100783	Breast aplasia
HP:0100013	Neoplasm of the breast
HP:0003002	Breast carcionma
HP:0003187	Breast hypoplasia
HP:0000769	Abnormality of the breast
HP:0010619	Fibroma of the breast

Hub genes, highly interconnected with nodes in network, have been considered functionally significant in network. In our study, the top 10 hub genes were defined by node degree and MCC algorithm in Cytoscape plugin cytoHubba [22]. Reference the previously described workflow of selection the essential proteins from the yeast protein interaction network with MCC algorithm [22]. First, the degrees of nodes were computed by NetworkAnalyzer [56] in Cytoscape. Then the node with degree greater than a threshold were choosed as potential candidate Hub genes, and the threshold is the maximum integer as $2 \times \sum_{v \in V, Deg(v) > t} Deg(v) > \sum_{v \in V,} Deg(v)$, where v is the collection of nodes within the network V, *Deg*(*v*) is the degree of node v. Last, the top 10 hub genes were ranked by MCC algorithm in cytoHubba plugin. Hub genes common in breast tissue co-expression networks were chosen as the candidates to be further analyzed and validated with

TCGA [23] and Kaplan-Meier (KM) plotter database (http://kmplot.com/analysis/) [24].

4.5 Similar small molecule detection

In order to detect the similar small molecule with bufadienolides-like chemicals, the DEGs with up or down were respectively submitted to the comparative toxicogenomics database (CTD) [25] and connectivity map (CMAP2, http://www.broadinstitute.org/cMAP/) database [26, 27]. During the process of detection similar small molecule with CTD Database, the threshold of degree in the degree filter network was set as 10. During the process of detection similar small molecule with connectivity map database, the enrichment score and p-value of were choose as similarity index between the gene expression profile of the query signature and that of chemicals in CMAP2.

Also the potential toxicity same as bufadienolides-like chemicals were also detected by CEBS database (https://manticore.niehs.nih.gov/cebssearch/) [57], but there is no evidence to prove the bufadienolides-like chemicals with obvious toxicity.

5. Conclusions

In this research, with a serious of bioinformatics analysis, we take notice the bufadienolides-like chemicals may perform anticancer activity through RNA splicing, apoptotic process, cell migration, extracellular matrix organization, adherens junction organization, synaptic transmission, Wnt signaling, AK-STAT signaling, BMP signaling pathway and unfolded protein response, and those may highlight the potential molecular mechanism of bufadienolides-like chemicals on Breast cancer, but still there are several problem had better solution, the toxicity of bufadienolides-like chemicals, especially the cardiotoxicity, which had been widely observe from clinic. The second problem, the difference of potential molecular mechanism among bufadienolides-like chemicals also had been clear illuminated in this research.

Supplementary Materials: The following are available online, Figure S1: Other differentially expressed networks regulated by bufadienolides-like chemic, Table S1: The DEGs disturbed by bufadienolides-like chemicals, Table S2: The different gene sets disturbed by bufadienolides-like chemicals.

Author Contributions: Y. P., L. H., and Y. Z. designed the flow of this analysis; D. W., and X. H. collected the gene expression profile information; X. T., and Y. Z. performed the analysis; Y. Z., and C. Y. Discussed the results. Y. Z., and X. T., wrote the manuscript. All of the authors made important suggestions to the manuscript and reviewed and approved the manuscript.

Funding: This work was supported by grants from National Natural Science Foundation of China (No. 81374065 and No. 81403035), and Basal Research Fund of Central Public-interest Scientific Institution (No. 1630032015039).

Acknowledgments: We thank Doctor Fulai Yu and Xiaolu Chen, for the advice and review of the manuscript.

Conflicts of Interest: The following authors report no conflicts of interest.

References

1. Mai, F.T. and H.A. Omar, *Immunotherapy, an evolving approach for the management of triple negative breast cancer: Converting non-responders to responders.* Critical Reviews in Oncology/hematology, 2018. **122**.
2. Zhou, H., et al., *RING1 and YY1 binding protein suppresses breast cancer growth and metastasis.* International Journal of Oncology, 2016. **49**(6): p. 2442.
3. Pan, Z., et al., *SATB1 is Correlated with Progression and Metastasis of Breast Cancers: A Meta-Analysis.* Cellular Physiology & Biochemistry, 2016. **38**(5): p. 1975-1983.

4. Xu, H., et al., *CD44 correlates with clinicopathological characteristics and is upregulated by EGFR in breast cancer.* International Journal of Oncology, 2016. **49**(4): p. 1343-1350.
5. Akiko, E., et al., *Inhibitory effects of bufadienolides on interleukin-6 in MH-60 cells.* Journal of Natural Products, 2004. **67**(12): p. 2070-2072.
6. Hong, Z., ., K. Chan, ., and H.W. Yeung, *Simultaneous determination of bufadienolides in the traditional Chinese medicine preparation, liu-shen-wan, by liquid chromatography.* Journal of Pharmacy & Pharmacology, 2011. **44**(12): p. 1023-1026.
7. Tian-Jie, et al., *Efficacy and safety of gemcitabine-oxaliplatin combined with huachansu in patients with advanced gallbladder carcinoma.* World Journal of Gastroenterology, 2008. **14**(33): p. 5210-5216.
8. Wang, J., et al., *Involvement of caspase-3 activity and survivin downregulation in cinobufocini-induced apoptosis in A 549 cells.* Experimental Biology & Medicine, 2009. **234**(5): p. 566-572.
9. Chun, L., et al., *The mechanisms of chansu in inducing efficient apoptosis in colon cancer cells.* Evidence-Based Complementray and Alternative Medicine,2013,(2013-5-30), 2013. **2013**(12): p. 849054.
10. Jiun-Yih, Y., et al., *Effects of bufalin and cinobufagin on the proliferation of androgen dependent and independent prostate cancer cells.* Prostate, 2010. **54**(2): p. 112-124.
11. Yu, C., et al., *Apoptotic signaling in bufalin- and cinobufagin-treated androgen-dependent and -independent human prostate cancer cells.* Cancer Science, 2010. **99**(12): p. 2467-2476.
12. Q., D.Y., et al., *Effect of cinobufagin on nuclear factor-kappa B pathway in HepG2 cells.* ournal of Southern Medical University, 2010. **30**(01): p. 137-139+142.
13. Ko, W.S., et al., *Induction of apoptosis by Chansu, a traditional Chinese medicine, in human bladder carcinoma T24 cells.* Oncology Reports, 2005. **14**(2): p. 475-480.
14. Wang, L., et al., *Pilot study on the mechanisms of growth inhibitory effect of cinobufagin on HeLa cells.* Chinese Journal of Oncology, 2005. **27**(12): p. 717.
15. Haenzelmann and Sonja, *GSVA: gene set variation analysis for microarray and RNA-Seq data.* Bmc Bioinformatics, 2013. **14**(1): p. 7-7.
16. Bindea, G., et al., *ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks.* Bioinformatics, 2009. **25**(8): p. 1091-1093.
17. Su, G., et al., *Biological network exploration with cytoscape 3.* Curr Protoc Bioinformatics, 2014. **47**: p. 8.13.1.
18. Barabási, A.L. and Z.N. Oltvai, *Network biology: Understanding the cell's functional organization.* Nature Reviews Genetics, 2004. **5**(2): p. 101- 113.
19. Lee, S., et al., *TCSBN: a database of tissue and cancer specific biological networks.* Nucleic Acids Research, 2017. **46**(Database issue): p. gkx994.
20. Xia, J., E.E. Gill, and R.E.W. Hancock, *NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data.* Nature Protocols, 2015. **10**(6): p. 823-844.
21. Soul, J., et al., *PhenomeScape: a cytoscape app to identify differentially regulated sub-networks using known disease associations.* Bioinformatics, 2016. **32**(24): p. 3847–3849.
22. Chin, C.H., et al., *cytoHubba: identifying hub objects and sub-networks from complex interactome.* Bmc Systems Biology, 2014. **8 Suppl 4**(S4): p. S11.
23. Kosinski M, B.P., *RTCGA: The Cancer Genome Atlas Data Integration. R package version 1.12.0.* 2018.
24. Balazs, G., et al., *An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients.* Breast Cancer Res Treat, 2010. **123**(3): p. 725-731.
25. Davis A. P, G.C.J., Johnson R. J., Sciaky D., King B. L., McMorran R., Wiegers J., Wiegers T. C., Mattingly C. J. , *The Comparative Toxicogenomics Database: update 2017.* Nucleic Acids Res. , 2017. **45**(D1): p. D972-D978.

26. Justin, L., et al., *The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease.* Science, 2006. **313**(5795): p. 1929-1935.
27. Subramanian, A., et al., *A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles.* Cell, 2017. **171**(6): p. 1437-1452.
28. Minegaki T., S.A., Mori M., Tsuji S., Yamamoto S., Watanabe A., Tsuzuki T., Tsunoda T., Yamamoto A., Tsujimoto M., Nishiguchi K., *Histone deacetylase inhibitors sensitize 5-fluorouracil-resistant MDA-MB-468 breast cancer cells to 5-fluorouracil.* Oncology letters, 2018. **16**(5): p. 6.
29. Riva G., C.C., Bazzoni R., Cadamuro M., Negroni C., Butta V., Strazzabosco M., Dalprà L., Lavitrano M., Bentivegna A., *Valproic Acid Inhibits Proliferation and Reduces Invasiveness in Glioma Stem Cells Through Wnt/β Catenin Signalling Activation.* Genes (Basel), 2018. **9**(11): p. E522.
30. Bennink, H.J.T.C., et al., *The use of high-dose estrogens for the treatment of breast cancer.* Maturitas, 2017. **95**: p. 11-23.
31. Hankinson, S.E., et al., *Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women.* J Natl Cancer Inst, 1998. **91**(8): p. 1292-1299.
32. Mdc, B.O., et al., *Use of anastrozole in the chemoprevention and treatment of breast cancer: A literature review.* Revista Da Associacao Medica Brasileira, 2017. **63**(4): p. 371.
33. Soul J., H.T.E., Boot-Handford R. P., Schwartz J. M., *PhenomeExpress: A refined network analysis of expression datasets by inclusion of known disease phenotype.* Scientific Reports, 2015(5): p. 8117.
34. Dunn S. L., S.J., Anand S., Schwartz J. M., Boot-Handford R. P., Hardingham T. E., *Gene expression changes in damaged osteoarthritic cartilage identify a signature of non-chondrogenic and mechanical responses.* Osteoarthritis & Cartilage, 2016. **24**(8): p. 9.
35. Chao, L., et al., *The gene expression profiles in response to 102 traditional Chinese medicine (TCM) components: a general template for research on TCMs.* Sci Rep, 2017. **7**(1): p. 352.
36. Sean, D. and P.S. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.* Bioinformatics, 2007. **23**(14): p. 1846-7.
37. Team, R.C., *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* . 2018.
38. A., B., *Minimum Information About a Microarray Experiment (MIAME)--successes, failures, challenges.* Scientific World Journal., 2009. **9**: p. 3.
39. Dondrup, M., et al., *EMMA 2 – A MAGE-compliant system for the collaborative analysis and integration of microarray data.* Bmc Bioinformatics, 2009. **10**(1): p. 50.
40. Gautier, L., B.M. Cope LBolstad, and R.A. Irizarry, *affy - analysis of Affymetrix GeneChip data at the probe level.* Bioinformatics, 2004. **20**(3): p. 307-315.
41. Gharaibeh, R.Z., A.A. Fodor, and C.J. Gibas, *Background correction using dinucleotide affinities improves the performance of GCRMA.* Bmc Bioinformatics, 2008. **9**(1): p. 452-452.
42. Leek, J.T., et al., *The sva package for removing batch effects and other unwanted variation in high-throughput experiments.* Bioinformatics, 2012. **28**(6): p. 882-883.
43. Gentleman R, C.V., Huber W, Hahne F, *genefilter: genefilter: methods for filtering genes from high-throughput experiments. R package version 1.64.0.* 2018.
44. Martin Morgan, S.F.a.R.G., *GSEABase: Gene set enrichment data structures and methods. R package version 1.44.0.* 2018.
45. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies.* Nucleic Acids Research, 2015. **43**(7): p. e47.

46. Arthur, L., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739.
47. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
48. Wolfgang, H., et al., *Orchestrating high-throughput genomic analysis with Bioconductor.* Nature Methods, 2015. **12**(2): p. 115-21.
49. Castelo, R., *GSVAdata: Data employed in the vignette of the GSVA package.* 2018.
50. Wickham., H., *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York., 2016.
51. Kassambara, A., *ggpubr: 'ggplot2' Based Publication Ready Plots.* 2018.
52. Kolde, R., *pheatmap: Pretty Heatmaps.* 2018.
53. Wilke, C.O., *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.* 2018.
54. Berkopec, A., *HyperQuick algorithm for discrete hypergeometric distribution.* Journal of Discrete Algorithms, 2007. **5**(2): p. 341-3472004.
55. Sebastian, K.H., et al., *Clinical diagnostics in human genetics with semantic similarity searches in ontologies.* American Journal of Human Genetics, 2009. **85**(4): p. 457-464.
56. Yassen, A., et al., *Computing topological parameters of biological networks.* Bioinformatics, 2008. **24**(2): p. 282-4.
57. Lea, I.A., et al., *CEBS: a comprehensive annotated database of toxicological data.* Nucleic Acids Research, 2017. **45**(D1): p. D964.