Journal of Visualized Experiments

Performing Data Mining and Integrative Analysis of Biomarker in Breast Cancer using Multiple Publicly Accessible Databases

--Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE59238R2
Full Title:	Performing Data Mining and Integrative Analysis of Biomarker in Breast Cancer using Multiple Publicly Accessible Databases
Keywords:	Breast cancer, Biomarker, Database, Data mining, Prognosis, Bioinformation
Corresponding Author:	Hao-yu Lin First Affiliated Hospital of Shantou University Medical College Shantou, CHINA
Corresponding Author's Institution:	First Affiliated Hospital of Shantou University Medical College
Corresponding Author E-Mail:	rainlhy@stu.edu.cn
Order of Authors:	Min-na Chen
	De Zeng
	Zhuo-qun Zheng
	Zheng Li
	Jian-le Wu
	Jun-yu Jin
	He-jia Wang
	Cui-zhen Huang
	Hao-yu Lin
Additional Information:	
Question	**Response**
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US$2,400)
Please indicate the **city, state/province, and country** where this article will be **filmed**. Please do not use abbreviations.	Guangzhou,Guangdong, China

TITLE:

Performing Data Mining and Integrative Analysis of Biomarker in Breast Cancer using Multiple Publicly Accessible Databases

AUTHORS AND AFFILIATIONS:

Min-na Chen[1]*, De Zeng[2]*, Zhuo-qun Zheng[3], Zheng Li[3], Jian-le Wu[3], Jun-yu Jin[3], He-jia Wang[3], Cui-zhen Huang[1], Hao-yu Lin[1]

[1]Department of Thyroid and Breast Surgery, the First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, P.R. China

[2]Department of Medical Oncology, Cancer Hospital of Shantou University Medical College, Shantou, PR China

[3]Shantou University Medical College, Shantou, PR China

* These authors contributed equally to this work

Corresponding Author:

Hao-yu Lin (Rainlhy@stu.edu.cn)

Email Addresses of Co-authors:

Min-na Chen (mnchen@stu.edu.cn)
De Zeng (dezeng@stu.edu.cn)
Zhuo-qun Zheng (17zqzheng1@stu.edu.cn)
Zheng Li (17zli2@stu.edu.cn)
Jian-le Wu (14jlwu4@stu.edu.cn)
Jun-yu Jin (17jyjin@stu.edu.cn)
He-jia Wang (16hjwang1@stu.edu.cn)
Cui-zheng Huang (czhuang@stu.edu.cn)

KEYWORDS:

Breast cancer, Biomarker, Database, Data mining, Prognosis, Bioinformation

SUMMARY:

Here, we present a protocol to explore the biomarker and survival predictor of breast cancer based on the comprehensive analysis of pooled clinical datasets derived from a variety of publicly accessible databases, using the strategy of expression, correlation and survival analysis step by step.

ABSTRACT:

In recent years, emerging databases were designed to lower the barriers for approaching the intricate cancer genomic datasets, thereby, facilitating investigators to analyze and interpret genes, samples and clinical data across different types of cancer. Herein, we describe a practical operation procedure, taking ID1 (Inhibitor of DNA binding proteins 1) as an example, to characterize the expression patterns of biomarker and survival predictors of breast cancer based

on pooled clinical datasets derived from online accessible databases, including ONCOMINE, bcGenExMiner v4.0 (Breast cancer gene‐expression miner v4.0), GOBO (Gene expression-based Outcome for Breast cancer Online), HPA (The human protein atlas), and Kaplan-Meier plotter. The analysis began with querying the expression pattern of the gene of interest (e.g., ID1) in cancerous samples vs. normal samples. Then, the correlation analysis between ID1 and clinicopathological characteristics in breast cancer was performed. Next, the expression profiles of ID1 was stratified according to different subgroups. Finally, the association between ID1 expression and survival outcome was analyzed. The operation procedure simplifies the concept to integrate multidimensional data types at the gene level from different databases and test hypotheses regarding recurrence and genomic context of gene alteration events in breast cancer. This method can improve the credibility and representativeness of the conclusions, thereby, present informative perspective on a gene of interest.

INTRODUCTION:

Breast cancer is a heterogeneous disease with diverse prognosis and treatment strategies in different molecular subtypes, in which the pathogenesis and development are probably associated with disparate molecular mechanisms[1-3]. However, identifying a therapeutic target usually takes years, or even decades, from initial discovery in basic research to clinical use[4]. Genome wide application of high-throughput sequencing technology for cancer genome has greatly advanced the process of searching for valuable biomarkers or therapeutic targets [5].

The overwhelming amount of cancer genomics data generated from the large-scale cancer genomics platforms, such as the ICGC (International Cancer Genome Consortium) and TCGA (The Cancer Genome Atlas), is posing a great challenge for researchers to perform data exploration, integration, and analytics, particularly for users lacking intensive training in informatics and computation[6-10]. In recent years, emerging databases, (e.g., ONCOMINE, bcGenExMiner v4.0, and Kaplan-Meier plotter, etc.) were designed and developed to lower the bar for approaching the intricate cancer genomic datasets, thereby, facilitating investigators to analyze and interpret the genes, samples and clinical data across various types of cancer[11]. The goal of this protocol is to describe a research strategy that integrated with multiple levels of gene information from a series of open access databases, which have been widely recognized by a great number of researchers, to identify the potential biomarkers and prognostic factors for breast cancer.

The ONCOMINE database is a web-based data-mining platform with cancer microarray information and is designed to facilitate discovery of novel biomarkers and therapeutic targets[11]. Currently, there are more than 48 million gene expression measurements from 65 gene expression datasets in this database[11,12]. The bcGenExMiner v4.0 (a free tool for non-profit institution), also called breast cancer Gene-Expression Miner, is a user-friendly web-based application comprising DNA microarrays results of 3,414 recovered breast cancer patients and 1,209 experienced a pejorative event[13]. It is designed to improve gene prognostic analysis performance with R statistical software and packages.

The GOBO is a multifunctional user-friendly online tool with microarrays information (e.g., Affymetrix U133A) from a 51-sample breast cancer cell line set and an 1881-sample breast tumor

data set, that allows a wide array of analyses[14]. There are a variety of applications available in the GOBO database, which include rapid analysis of gene expression profiles in different molecular subtypes of breast tumors and cell lines, screening for co-expressed genes for creation of potential metagenes, and correlation analysis between outcome and gene expression levels of single genes, sets of genes, or gene signatures in breast cancer data set[15].

The Human Protein Atlas is an open-access program designed for scientists to explore human proteome, which has already contributed to a large number of publications in the field of human biology and disease. The Human Protein Atlas is recognized as a European core resource for life science community[16,17].

The Kaplan Meier plotter is an online tool integrating gene expression and clinical data simultaneously that allows assessment of the prognostic effect of 54,675 genes based on 10,461 cancer samples, which include 1,065 gastric, 2,437 lung, 1,816 ovarian and 5,143 breast cancer patients with a mean follow-up of 33/49/40/69 months[18]. Information of gene expression, relapse-free survival (RFS) and overall survival (OS) are downloadable from this database[19,20].

Here, we describe a practical operation procedure of using multiple publicly accessible databases to compare, analyze and visualize patterns of alterations in the expression of the gene of interest across multiple cancer studies, with the goal of summarizing the expression profiles, prognostic values and potential biological functions in breast cancer. For example, recent studies have indicated the oncogenic properties of ID proteins in tumors and were associated with malignant features, including cellular transformation, immortalization, enhanced proliferation and metastasis[21-23]. However, each member of the ID family plays distinct roles in different types of solid tumors, and their role in breast cancer remains unclear[24]. In previous studies, explored through this method, we found that ID1 was a meaningful prognostic indicator in breast cancer[25]. Therefore, the protocol will take ID1 as an example to introduce the data mining methods.

The analysis starts from querying the expression pattern of the gene of interest in cancerous samples vs. normal samples in ONCOMINE. Then, the expression correlation of genes of interest in breast cancer was performed using the bc-GenExMiner v4.0, GOBO, and ONCOMINE. Next, the expression profiles of ID1 was stratified according to different subgroups using the above three databases. Finally, the association between ID1 expression and survival out was analyzed using bc-GenExMiner v4.0, the human protein atlas, and Kaplan-Meier plotter. The operation procedure was shown as the flowchart in **Figure 1**.

PROTOCOL

1. Expression Pattern Analysis

1.1 Go to the ONCOMINE web interface[26].

1.2 Obtain the relative expression levels of gene *ID1* in various types of malignancies by typing **ID1** to the **Search Box**.

1.3 Select **Analysis Type** from the **Primary Filters** menu. Then, select **Cancer vs. Normal Analysis**, **Breast Cancer vs. Normal Analysis**.

1.4 Select **Gene Summary View** from the **OTHER VIEWS** menu. Set the threshold of *P*-value at 0.01. Download the figures.

NOTE: The threshold of fold change is 2, as described in the previous study[27].

2. Expression Correlation Analysis

2.1 Go to the bc-GenExMiner v4.0 web interface[28].

2.2 Select **CORRELATION** from the **ANALYSIS** menu, press the **EXHAUSTIVE** button. Type **ID1** to the search box. Press the **Submit** button and the **Start analysis** button.

NOTE: Default setting show expression correlation analysis of all patients, which can be more accurate in different subtypes of breast cancer by pressing the **Molecule subtype** filter.

3. Subgroup Analysis

3.1 Subgroup analysis in bc-GenExMiner v4.0

3.1.1 Go to the bc-GenExMiner v4.0 web interface[28].

3.1.2 Select **EXPRESSION** from the **ANALYSIS** menu, press the **EXHAUSTIVE** button. Type **ID1** to the search box and press the **Submit** button and the **Start analysis** button.

3.1.3 Click the **Nodal status (LN)** and **Scarff Bloom & Richardson grade status (SBR)** thumbnails to view full images. In the SBR images, press the button below to visualize the *P*-values of the figures. Download the figures.

3.2 Subgroup analysis in Gene expression-based Outcome for Breast Cancer Online (GOBO)

3.2.1 Go to the GOBO web interface[14].

3.2.2 Type Gene symbol of interest **ID1** to the screen **upload the gene set**.

3.2.3 Set the search range of **Define gene/probe identifiers** to **Gene Symbol**. Set **All** in **Tumor selection**. Select **Node status** and **Grade stratified** in the **Multivariate parameters**. Other items remain default. Submit the inquiry and download the figures.

4. Survival Analysis

4.1 Survival analysis in bc-GenExMiner v4.0

4.1.1 Go to the bc-GenExMiner v4.0 web interface[28].

4.1.2 Select **PROGNOSTIC** from the **ANALYSIS** menu, press the **EXHAUSTIVE** button. Type **ID1** to the search box and press the **Submit** button and the **Start analysis** button.

4.1.3 In the Exhaustive prognostic analysis, select **Nm, ERm, MR** in the **Population and event criteria** and press the **Submit** button to obtain more information. Press the **Kaplan-Meier curve** thumbnails to export the full graphs.

NOTE: N (+, -, m): nodal status (+: positive, -: negative, m: mixed); ER (+, -, m): oestrogen receptor status (+: positive, -: negative, m: mixed); MR: metastatic relapse

4.2 Survival analysis in The Human Protein Atlas (HPA)

4.2.1 Go to the Human Protein Atlas web interface[29].

4.2.2 Type **ID1** to the search box and click the **Search** button. Select **Pathology** sub-atlas.

NOTE: The mRNA expression levels across the 17 cancer types are shown in the RNA Expression overview section. Every cancer tissue label of the box plot is clickable to access a detailed page providing survival analysis data and RNA expression levels.

4.2.3 Click the label of **Breast Cancer**, then the detailed page to show interactive survival scatter plot and survival analysis. Download the figures.

4.3 Survival analysis in The Kaplan-Meier Plotter Survival

4.3.1 Go to the Kaplan-Meier Plotter web interface[30]. Click **Start KM plotter for breast cancer** in the mRNA gene chip zone.

4.3.2 Type **ID1** to the search bar and select the green item in the candidate menu.

4.3.3 Select **RFS** as survival type and Other items remain default. Click **Draw Kaplan-Meier plot** and download the figures.

NOTE: Settings of the survival types, cutoff types, and follow-up threshold, as well as probe set options, can be changed as required. Subgroup prognostic analysis including ER, PR, HER-2, lymph nodes, grade, Tp53 status, and molecular subtypes can be obtained via changing the setting in the **Restrict analysis to subtypes** box[1]. Likewise, the filter limitation of treatment could be set in **Restrict analysis to selected cohorts'** box.

REPRESENTATIVE RESULTS:

A representative result of data mining and integrative analysis of breast cancer biomarker was performed using ID1, one of the inhibitors of DNA‐binding family members, which have been reported in the previous study [25].

As demonstrated in **Figure 2**, the differences of ID1 mRNA expression between tumor and normal tissues in multiple types of cancer were analyzed using the ONCOMINE database, which contained a total of 445 unique analyses. There were 5 studies which revealed that the mRNA expression level of ID1 was significantly higher in normal tissues than in breast cancer tissues. These data indicated the expression dysregulation of ID1 in breast cancer. **Figure 3** showed the best positive and negative correlative genes of ID1 from the analysis performed in bc-GenExMiner v4.0. To identify the correlation between mRNA expression of ID1 and the clinicopathological parameters of BC patients, bc-GenExMiner v4.0 database was used the analysis. As shown in **Figure 4**, significantly increased mRNA level of ID1 was found in breast cancer patients without lymph node metastasis, as compared to those with lymph node metastasis (P=0.0005). Furthermore, the analysis in GOBO demonstrated that increased mRNA levels of ID1 were correlated to lower tumor grade (**Figure 5**, P<0.00001). These results implied that increased expression of ID1 was linked to lower metastatic potential and lower pathological grade in BC. The analysis from the bc-GenExMiner v4.0 database indicated that higher mRNA level of ID1 was correlated to longer distant metastasis-free survival (DMFS) in breast cancer patients (**Figure 6**, HR=0.82, 95% CI: 0.73‐0.92, P=0.001). Consistently, analysis from The Human Protein Atlas suggested that elevated protein level of ID1 was associated with better survival outcome in breast cancer patients (**Figure 7**, P=0.0389). Survival analysis from the Kaplan-Meier Plotter also showed that higher mRNA level of ID1 expression predicted better recurrence-free survival (RFS) in breast cancer patients (**Figure 8**, HR=0.81, P=0.00023).

FIGURE LEGENDS:

Figure 1. Overview of exploring the expression patterns and prognostic values of distinct breast cancer biomarkers and online databases selection. Systematic analysis of distinct breast cancer biomarkers was performed step by step in a variety of databases. First, the expression pattern of the gene of interest in cancerous samples vs. normal samples. Then, the expression correlation of genes of interest in breast cancer was performed. Next, the expression profiles of ID1 was stratified according to different. Finally, the association between ID1 expression and survival out was analyzed.

Figure 2. The mRNA expression pattern of the ID1 in different types of human cancer. The mRNA expression of ID1 analyzed with the ONCOMINE database. The graphic demonstrated the numbers of datasets with statistically significant mRNA overexpression (red) or downregulated expression (blue) of the target gene. The number in each cell represented the number of analyses that meet the threshold within those analysis and cancer types. The gene rank was analyzed by percentile of the target gene in the top of all genes measured in each research. Cell color was determined by the best gene rank percentile for the analyses within the cell. The P-value was set

up at 0.01 and fold‑change was defined as 2, as shown in the red frame. This figure has been modified from the previous study[25].

Figure 3. Gene correlation analysis of ID1 in bc-GenExMiner v4.0. The mRNA expression correlation of ID1 and relevant genes in 5, 696 breast cancer patients within 36 studies analyzed in bcGenExMiner v4.0. This figure has been modified from the previous study [25].

Figure 4. The relationship between ID1 expression and lymph node metastasis status. The mRNA expression level of ID1 in 4, 307 breast cancer patients with different lymph node (LN) status analyzed in bcGenExMiner v4.0. This figure has been modified from the previous study [25].

Figure 5. The relationship between the gene expression level of ID1 and tumor grade. The mRNA expression level of ID1 in breast cancer patients with different pathological grade was analyzed in GOBO. The global significant difference between groups was assessed to generate *P*‑values and $P<0.05$ was considered to indicate a statistically significant difference. 1,2,3 in x-axis stand for sub-groups of patients in different pathological grade 1, grade 2, grade 3. This figure has been modified from the previous study [25].

Figure 6. The prognostic values of ID1 for distant metastasis-free survival in breast cancer patients. The association between ID1 mRNA levels and distant metastasis-free survival estimates was analyzed in bcGenExMiner v4.0. This figure has been modified from the previous study [25].

Figure 7. The survival probability of ID1 in breast cancer patients. Impact of ID1 protein level for the survival of patients with breast cancer was analyzed in the human protein atlas (HPA). This figure has been modified from the previous study [25].

Figure 8. The prognostic values of ID1 in breast cancer according to recurrent-free survival (RFS). Different ID1 mRNA level in all 3, 951 breast cancer patients analyzed in Kaplan-Meier plotter. This figure has been modified from the previous study [25].

DISCUSSION:

Comprehensive analysis of public databases may indicate the underlying function of the gene of interest and reveal the potential link between this gene and clinicopathological parameters in specific cancer[27,31]. The exploration and analysis based on one single database might provide limited or isolated perspectives due to the potential selection bias, or in a certain extent, possibly due to the variety of data quality, including data collection and the analytical algorithm of the database[19]. The most important step of this protocol is to select the appropriate databases, which should be widely recognized by a larger number of scientists with adequate representativeness. The investigator should use multiple databases to test the hypothesis and corroborate the results derived from different databases, rather than use a single database.

The protocol described here is an investigator friendly operation procedure. The advantage of this method is that it allows for the rapid visualization and interpretation of a gene's potential

role in breast cancer. Moreover, all the results obtained through this procedure can be immediately tested and repeated by simply querying the corresponding websites. The limitation of this method is that the conclusions which come from the comprehensive analysis of the databases may not exactly reflect the actual function or relationship in the clinical setting. This could stem from the systematical bias of the database, and in some case, possibly due to inadequate sample size[32,33]. Using more than one database to query the same research question could mutually confirm the results and increase the credibility of the conclusion[34]. It is strongly recommended to use samples from the investigator's institution to verify the results, or if feasible, to perform related basic experiments to test the results.

More and more online cancer genomics or proteomics databases will be available and accessible for researchers [35,36]. The protocol might provide an efficient and economical method for the researcher to identify a potential target gene and the associated signaling pathway through in-depth analysis of online databases and by using genomics, transcriptomics, and epigenomics approach.

ACKNOWLEDGMENTS:

This work was partly supported by the Natural Science Foundation of Guangdong Province, China (No. 2018A030313562), the Teaching Reform Project of Guangdong Clinical Teaching Base (NO. 2016JDB092), National Natural Science Foundation of China (81600358), and Youth Innovative Talent Project of Colleges and Universities in Guangdong Province, China (NO. 2017KQNCX073)

DISCLOSURES:

The authors have nothing to disclose

REFERENCES:

1 van 't Veer, L. J. et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* **415** (6871), 530-536 (2002).

2 Loi, S. et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *Journal of Clinical Oncology.* **25** (10), 1239-1246 (2007).

3 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature.* **490** (7418), 61-70 (2012).

4 Emerson, J. W., Dolled-Filhart, M., Harris, L., Rimm, D. L. & Tuck, D. P. Quantitative assessment of tissue biomarkers and construction of a model to predict outcome in breast cancer using multiple imputation. *Cancer Informatics.* **7**, 29-40 (2009).

5 Yu, H. et al. Integrative genomic and transcriptomic analysis for pinpointing recurrent alterations of plant homeodomain genes and their clinical significance in breast cancer. *Oncotarget.* **8** (8), 13099-13115 (2017).

6 He, W. et al. TCGA datasetbased construction and integrated analysis of aberrantly expressed long noncoding RNA mediated competing endogenous RNA network in gastric cancer. *Oncology Reports.* 10.3892/or.2018.6720 (2018).

7 Liu, J. et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell.* **173** (2), 400-416 e411 (2018).

8 Esgueva, R. et al. Next-generation prostate cancer biobanking: toward a processing protocol amenable for the International Cancer Genome Consortium. *Diagnostic Molecular Pathology.* **21** (2), 61-68 (2012).
9 Joly, Y., Dove, E. S., Knoppers, B. M., Bobrow, M. & Chalmers, D. Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). *PLoS Computational Biology.* **8** (7), e1002549 (2012).
10 Zhang, J. et al. International Cancer Genome Consortium Data Portal--a one-stop shop for cancer genomics data. *Database (Oxford).* **2011** bar026 (2011).
11 Rhodes, D. R. et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia.* **6** (1), 1-6 (2004).
12 Rhodes, D. R. et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia.* **9** (2), 166-180 (2007).
13 Jezequel, P. et al. bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Research and Treatment.* **131** (3), 765-775 (2012).
14 http://co.bmc.lu.se/gobo/gsa.plb (2018).
15 Ringner, M., Fredlund, E., Hakkinen, J., Borg, A. & Staaf, J. GOBO: gene expression-based outcome for breast cancer online. *PLoS One.* **6** (3), e17911 (2011).
16 Ponten, F., Jirstrom, K. & Uhlen, M. The Human Protein Atlas--a tool for pathology. *Journal of Pathology.* **216** (4), 387-393 (2008).
17 Ponten, F., Schwenk, J. M., Asplund, A. & Edqvist, P. H. The Human Protein Atlas as a proteomic resource for biomarker discovery. *Journal of Internal Medicine.* **270** (5), 428-446 (2011).
18 Gyorffy, B. 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 Research and Treatment.* **123** (3), 725-731 (2010).
19 Stevinson, C. & Lawlor, D. A. Searching multiple databases for systematic reviews: added value or diminishing returns? *Complementary Therapies in Medicine.* **12** (4), 228-232 (2004).
20 Yin, J. et al. Integrating multiple genome annotation databases improves the interpretation of microarray gene expression data. *BMC Genomics.* **11** 50 (2010).
21 Patel, D., Morton, D. J., Carey, J., Havrda, M. C. & Chaudhary, J. Inhibitor of differentiation 4 (ID4): From development to cancer. *Biochimica et Biophysica Acta.* **1855** (1), 92-103 (2015).
22 Kamalian, L. et al. Increased expression of Id family proteins in small cell lung cancer and its prognostic significance. *Clinical Cancer Research.* **14** (8), 2318-2325 (2008).
23 Cruz-Rodriguez, N. et al. High expression of ID family and IGJ genes signature as predictor of low induction treatment response and worst survival in adult Hispanic patients with B-acute lymphoblastic leukemia. *Journal of Experimental and Clinical Cancer Research.* **35**, 64 (2016).
24 Yang, H. Y. et al. Expression and prognostic value of Id protein family in human breast carcinoma. *Oncology Reports.* **23** (2), 321-328 (2010).
25 Zhou, X. L. et al. Prognostic values of the inhibitor of DNAbinding family members in breast

cancer. *Oncology Reports.* **40** (4), 1897-1906 (2018).
26 www.oncomine.org (2018).
27 Lin, H. Y., Zeng, Liang, Y. K., Wei, X. L. & Chen, C. F. GATA3 and TRPS1 are distinct biomarkers and prognostic factors in breast cancer: database mining for GATA family members in malignancies. *Oncotarget.* **8** (21), 34750-34761 (2017).
28 http://bcgenex.centregauducheau.fr/BCGEM/GEM-requete.php (2018).
29 https://www.proteinatlas.org (2018).
30 http://kmplot.com/analysis (2018).
31 Zhu, Y. F. & Dong, M. Expression of TUSC3 and its prognostic significance in colorectal cancer. *Pathology-Research and Practice.* **214** (9), 1497-1503 (2018).
32 Nelson, J. C. et al. Validation sampling can reduce bias in health care database studies: an illustration using influenza vaccination effectiveness. *Journal of Clinical Epidemiology.* **66** (8 Suppl), S110-121 (2013).
33 Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics.* **24** (19), 2200-2208 (2008).
34 Yang, C. et al. Understanding genetic toxicity through data mining: the process of building knowledge by integrating multiple genetic toxicity databases. *Toxicology Mechanisms and Methods.* **18** (2-3), 277-295 (2008).
35 Cannata, N., Merelli, E. & Altman, R. B. Time to organize the bioinformatics resourceome. *PLoS Computational Biology.* **1** (7), e76 (2005).
36 Wren, J. D. & Bateman, A. Databases, data tombs and dust in the wind. *Bioinformatics.* **24** (19), 2127-2128 (2008).

Figure 1

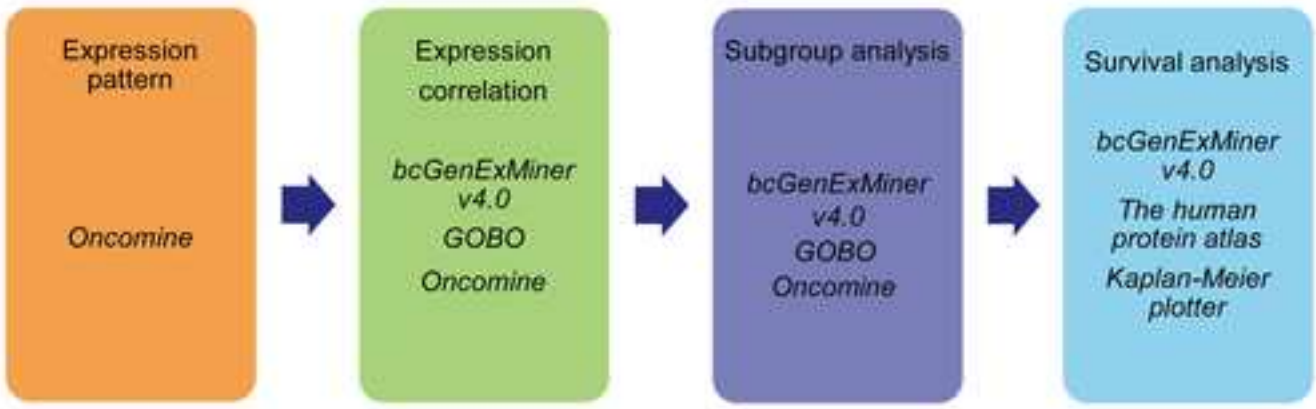


Figure 2

Click here to access/download;Figure;fig1-2018-9-19.tif

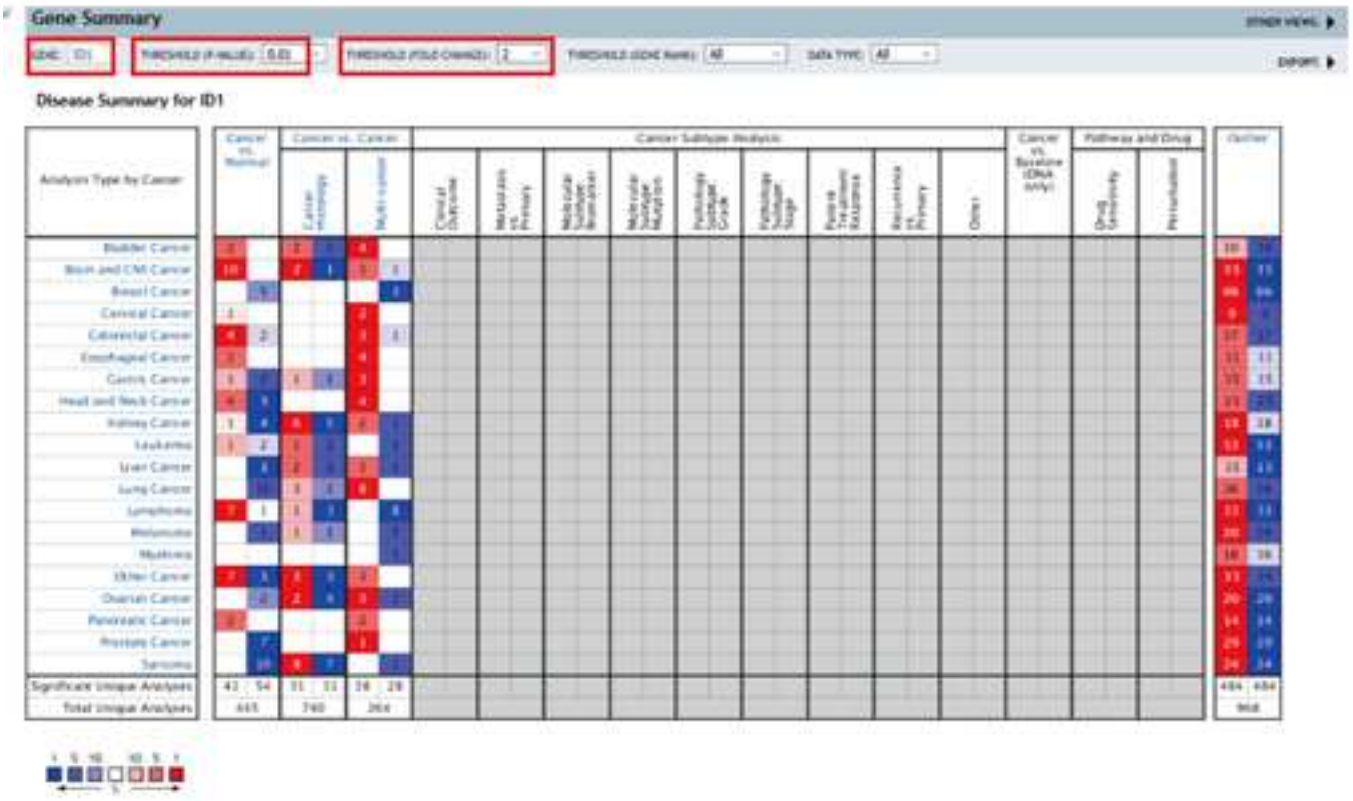


Figure3

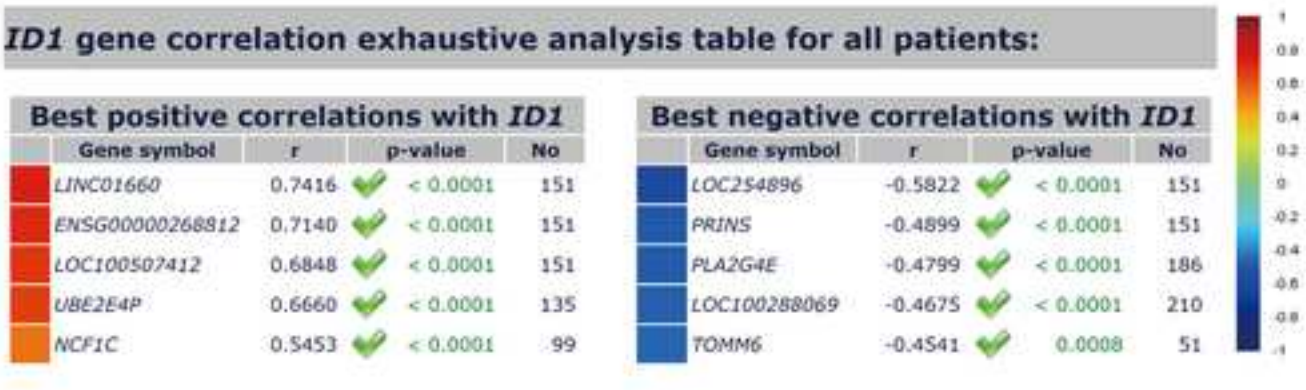


Figure 4

Click here to access/download;Figure;fig2-2018-9-18.tif

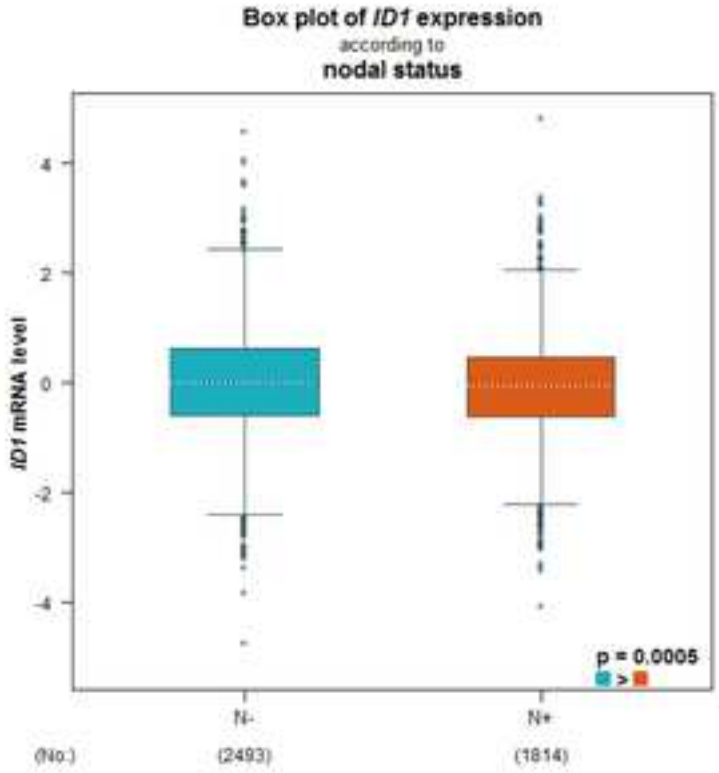


Figure 5

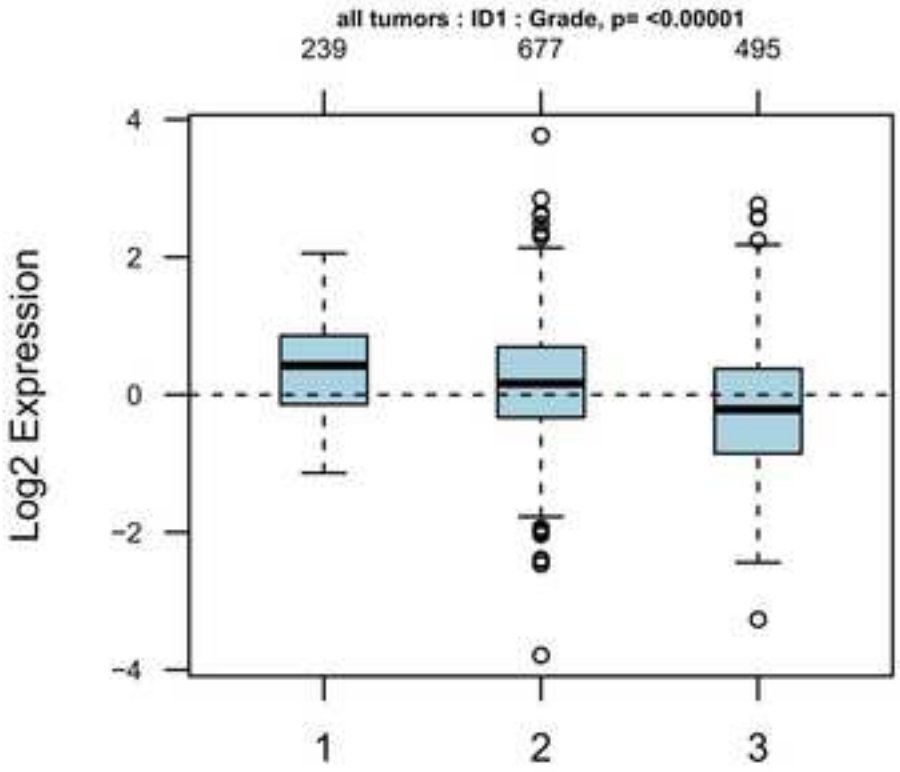


Figure 6

Click here to access/download;Figure;fig3-2018-9-18.tif

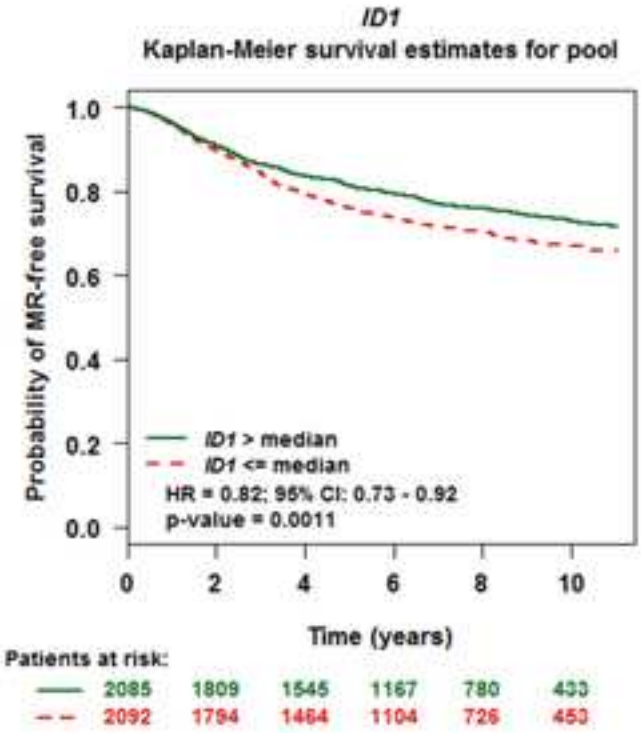


Figure 7

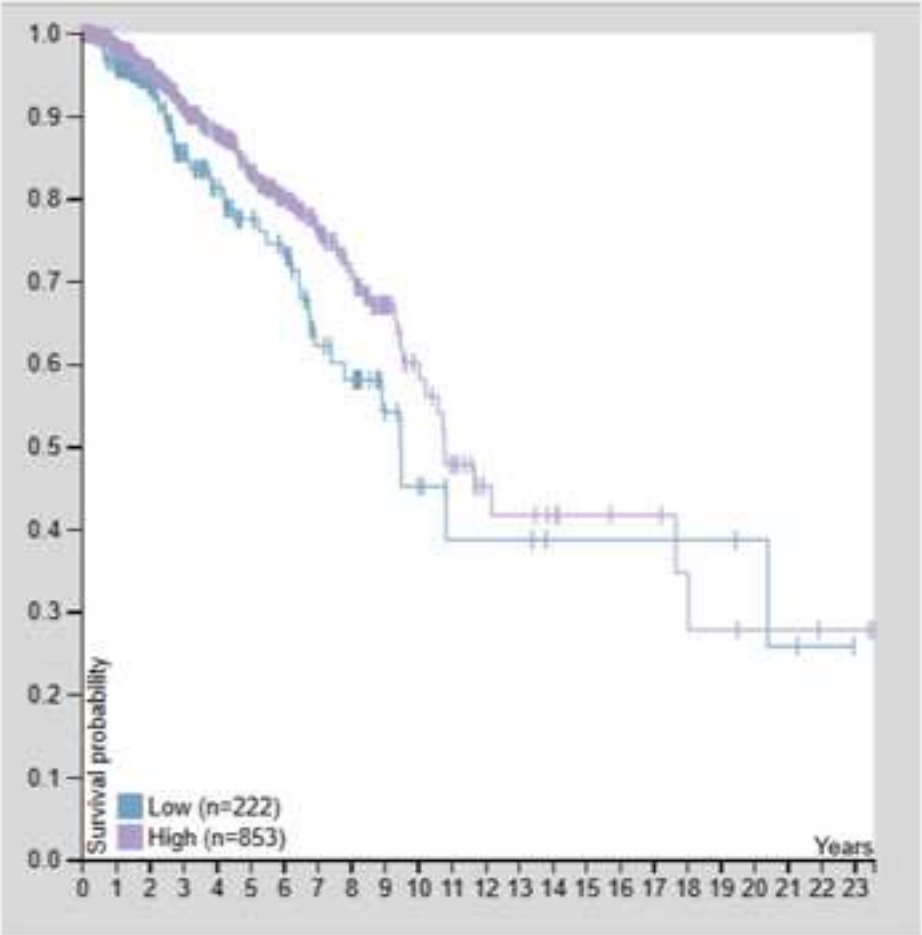
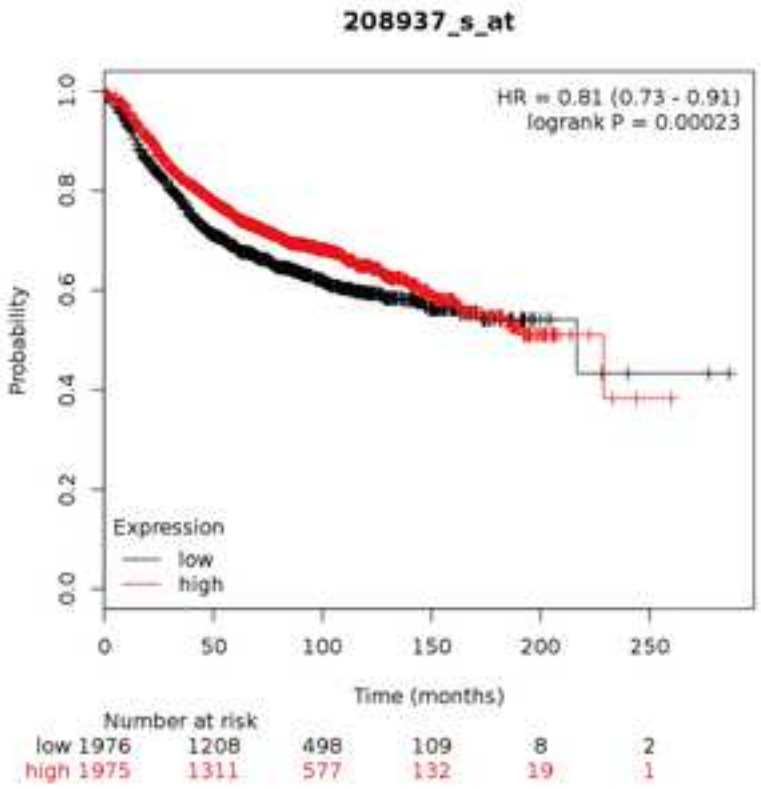


Figure 8

Click here to access/download;Figure;fig6-2018-9-18.tif



Name of Material/ Equipment	Company	Catalog Number	Comments/Description
A personal computer or computing device with an Internet browser with Javascript enabled	Microsoft	051690762553	We support and test the following browsers: Google Chrome, Firefox 3.0 and above, Safari, and Internet Explorer 9.0 and above
Adobe Flash player	Adobe Systems Inc.	It can be freely downloaded from http://get.adobe.com/flashplayer/.	This browser plug-in is required for visualizing networks on the network analysis tab.
Chrome Broswer	Google Inc.	It can be freely downloaded from https://www.google.cn/chrome/	This is necessary for viewing PDF files including the Pathology Reports and many of the downloadable files.
Java Runtime Environment	Oracle Corporation	It can be downloaded from http://www.java.com/getjava/.	This is necessary for viewing the Pathology Reports and for viewing many of the downloadable files.
Office 365 ProPlus for Faculty	Microsoft	2003BFFD8117EA68	

Vectr Online	Vectr Labs Inc.	It can be freely used from https://vectr.com/ new	This is necessary for visualizing and editing many of the downloadable files and pictures.

JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: A protocol to perform data mining and integrative analysis of biomarker in breast cancer using multiple publicly accessible databases

Author(s): Min-na Chen, De Zeng, Zhuo-qun Zheng, Zheng LI, Jian-le Wu, Jun-yu Jin, He-jia Wang, Cui-zheng Huang, Hao-yu Lin

Item 1 (check one box): The Author elects to have the Materials be made available (as described at http://www.jove.com/author) via: ☑ Standard Access ☐ Open Access

Item 2 (check one box):

☑ The Author is NOT a United States government employee.

☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.

☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. Defined Terms. As used in this Article and Video License Agreement, the following terms shall have the following meanings: **"Agreement"** means this Article and Video License Agreement; **"Article"** means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; **"Author"** means the author who is a signatory to this Agreement; **"Collective Work"** means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; **"CRC License"** means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode; **"Derivative Work"** means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; **"Institution"** means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; **"JoVE"** means MyJove Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments;* **"Materials"** means the Article and / or the Video; **"Parties"** means the Author and JoVE; **"Video"** means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. Background. The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. Grant of Rights in Article. In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4** and **7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the "Open Access" box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

jove JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

jove
JOURNAL OF VISUALIZED EXPERIMENTS
1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. Indemnification. The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in *JoVE* or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. Fees. To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. Transfer, Governing Law. This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to me one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Hao-yu Lin
Department:	Department of Thyroid and Breast Surgery
Institution:	The First Affiliated Hospital of Shantou University Medical College
Article Title:	A protocol to perform data mining and integrative analysis of biomarker in breast cancer using multiple publicly accessible databases
Signature:	Hao-yu Lin
Date:	2013/10/2

Please submit a signed and dated copy of this license by one of the following three methods:

1) Upload a scanned copy of the document as a pfd on the JoVE submission site;
2) Fax the document to +1.866.381.2236;
3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Dear Dr. Vineeta Bajaj

Review Editor

JoVE

Manuscript：JoVE59238

Title: “A protocol to perform data mining and integrative analysis of biomarker in breast cancer using multiple publicly accessible databases”

Thank you very much for providing us the opportunity to revise our manuscript. We have carefully read through the editors and reviewers’ comments and revised the manuscript according to their suggestions.

We have uploaded a copy after revisions, with tracked changes highlighted with red color. The below shows our responses to editors’ and reviewers’ comments point-by-point. We hope you will find that the revisions have significantly improved the quality of the manuscript. We believe that this latest version will meet the standards for publication in *JoVE* and will satisfy the readers of your journal.

I look forward to hearing from you soon.

Best regards,

Corresponding author:

Haoyu Lin, MD, PhD,

Department of Breast and Thyroid Surgery, the First Affiliated Hospital of Shantou University Medical College, 57 Changping Road, Shantou, China. rainlhy@stu.edu.cn

EDITORIAL CORRECTIONS REQUIRED (COMPULSORY - these must be addressed in addition to the reviewers' comments):

Editorial comments:

Changes to be made by the author(s) regarding the manuscript:

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

Reply: Thank you for your suggestion. We have read through the whole manuscript

and corrected the spelling or grammar errors

2. Please revise lines 207-210, 211-214 to avoid previously published text.

Reply: Thank you for your comment. We have revised in the manuscript to avoid duplication from previously published text.

3. Please obtain explicit copyright permission to reuse any figure/table from a previous publication. Explicit permission can be expressed in the form of a letter from the editor or a link to the editorial policy that allows re-prints. Please upload this information as a .doc or .docx file to your Editorial Manager account. The Figure/Table must be cited appropriately in the Figure Legend, i.e. “This figure/table has been modified from [citation].”

Reply: Thank you. We have asked for the copyright permission of reusing some figures from our previous study published in Oncology Reports. Attached is the explicit permission from Oncology Reports.

4. Please provide an institutional email address for each author.

Reply: Thank you for your reminding. Institutional email address for each author have added.

5. Keywords: Please remove commercial language.

Reply: We have removed the Keywords “ONCOMINE, bcGenExMiner v4.0, GOBO,

HPA, KM plotter" according to your suggestion.

6. Please define all abbreviations before use.

Reply: Thank you for your comment. We have defined all abbreviations in the manuscript.

7. Please revise the protocol text to avoid the use of any personal pronouns (e.g., "we", "you", "our" etc.).

Reply: Thank you for your comment. We have corrected in the manuscript according to your suggestion.

8. Please revise the protocol to contain only action items that direct the reader to do something (e.g., "Do this," "Ensure that," etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as "could be," "should be," and "would be" throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a "Note." Please include all safety procedures and use of hoods, etc. However, notes should be used sparingly, and actions should be described in the imperative tense wherever possible. Please move the discussion about the protocol to the Discussion.

Reply: Thank you for your suggestion. We have read through the protocol section and make appropriate revise according to your requirement.

9. Lines 100-107: The Protocol should contain only action items that direct the reader to do something. Please move this section to the Introduction.

Reply: Thank you for your comment. We have moved this section to the Introduction according to your suggestion.

10. Please note that your protocol will be used to generate the script for the video and must contain everything that you would like shown in the video. Software must have a GUI (graphical user interface) and software steps must be more explicitly explained ('click', 'select', etc.). Please add more specific details (e.g. button clicks for software actions, numerical values for settings, etc.) to your protocol steps. There should be enough detail in each step to supplement the actions seen in the video so that viewers can easily replicate the protocol. Some examples:

Reply: Thank you for your comment. We have provided more detail in the protocol section.

11. 2.2, 3.3, 3.7, etc.: What does “type free text” mean?

Reply: Thank you for your comment. We have change the statement to be “Type Gene symbol of interest ‘ID1’”

12. 2.6, 3.5: Please describe how.

Reply: Download the figures and arrange with Adobe Illustrator

13. 3.9: Please specify the value selected.

Reply: We have descripted in the manuscript,

14. 5.6: Please describe how to revise with image processing software.

Reply: After download the figures, we arrange with them with Adobe Illustrator, for example, to use the same font for all figures for consistency.

15. Please combine some of the shorter Protocol steps so that individual steps contain 2-3 actions and maximum of 4 sentences per step.

Reply: We have changed in the manuscript according to your suggestion.

16. Please include single-line spaces between all paragraphs, headings, steps, etc.

Reply: Thank you for your suggestion. make appropriate revise according to your suggestion.

17. After you have made all the recommended changes to your protocol (listed above), please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive story of the Protocol.

Reply: We have changed in the manuscript according to your suggestion.

18. Please highlight complete sentences (not parts of sentences). Please ensure that

the highlighted part of the step includes at least one action that is written in imperative tense.

Reply: We have made proper highlight in the manuscript according to your suggestion.

19. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.

Reply: We have made proper highlight in the manuscript according to your suggestion.

20. References: Please do not abbreviate journal titles.

Reply: Thank you for your comment. We have changed the references according to your suggestion.

21. Table of Materials: Please sort the items in alphabetical order according to the Name of Material/Equipment.

Reply: Thank you for your comment. We have changed the Table of Materials according to your requirment.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

the issue is interesting, however I have some concerns

Major Concerns:

-English is unacceptable

Reply: Thank you for your suggestion. We have read through the whole manuscript to correct the spelling or grammar errors.

-The text is often confused

Reply: Thank you for your suggestion. We have made proper change to make it more clearly and readable.

-Authors should better describe the involved databases, in particular what kind of data are available in each of them

Reply: Thank you for your comment. We have provided more details about those involved databases in the revised manuscript.

-The discussion section is confused and repetitive; in this section, authors should primarily examine and discuss the results obtained with their sample gene, ID1, in breast cancer

Reply: Thank you for your suggestion. We have revised the discussion section to make it more logically according to your suggestion.

Reviewer #2:

Manuscript Summary:

The article: "A protocol to perform data mining and integrative analysis of biomarker in breast cancer using multiple publicly accessible databases" describes the procedure to analyze biomarker potential and survival predictivity of genes based on their expression using different on-line databases. In particular, the authors analyze the expression pattern of ID1 in normal vs breast tumor tissue, they also search for correlations between expression and clinicopathological characteristics. The article is informative and the procedure for analyzing each database is well described. Below are some suggestions the authors could include to the manuscript.

Minor Concerns:

1-For the ONCOMINE protocol the authors mention to analyze the data using appropriate software, maybe mentioning some examples of such software could be useful.

Reply: Thank you for your suggestion. We have described the exact software we used.

2-The on-line tools mentioned in the paper are specific for breast cancer except for oncomine and Kaplan Meier plotter, maybe mentioning that there are other sites such as UCSC Xena for other types of cancer could also be informative for researchers working in other areas.

Reply: Thank you for your suggestion. We chose breast cancer in the study to test the feasibility of this method. One of the important reasons is that breast cancer is a heterogeneous disease. The diagnosis, treatment and prognosis of different molecular subtypes of breast cancer may vary. Therefore, it is particularly important to find efficient on-line tool or database with information to reflect the disparate subtypes of breast cancer. It is undeniable that there may be some reference value for other cancers.

3-It would be important to mention in the discussion section that for this type of in silico approach it is important to previously now the nature of the data i.e. does the data come from RNA seq or microarray technology? is the data normalized? How? This information is important for subsequent statistical analysis.

Reply: Thank you for your comment. We have described the sources of data in each database in both Introductions and Discussion sections. In the algorithm of each database, there is a corresponding method to standardize the data, but for the comparison among different databases, this standardization is insufficient, which is also the limitation of the method we described in Discussion. However, through analysis through multiple databases, which could be mutually validated or

corroborated at different levels, and more accurately showing a trend or tendency and providing clue for further in-depth research. And the preliminary conclusion definitely needs further basic or clinical trials to verify it.

4-There are some grammatical errors in the manuscript, this should be corrected.

Reply: Thank you for your suggestion. We have read through the whole manuscript and corrected the spelling or grammar errors

Dear Dr. Vineeta Bajaj

Review Editor

JoVE

Manuscript：JoVE59238

Title: “A protocol to perform data mining and integrative analysis of biomarker in breast cancer using multiple publicly accessible databases”

Thank you very much for carefully review of our manuscript. We have carefully read through your comments and revised the manuscript accordingly.

We have uploaded a copy after revisions, with tracked changes highlighted with red color. The below shows our responses to editors’ comments point-by-point. We hope you will find that the revisions have significantly improved the quality of the manuscript. We hope that this latest version will meet the standards for publication in *JoVE* and will satisfy the readers of your journal.

I look forward to hearing from you soon.

Best regards,

Corresponding author:

Haoyu Lin, MD, PhD,

Department of Breast and Thyroid Surgery, the First Affiliated Hospital of Shantou

University Medical College, 57 Changping Road, Shantou, China. rainlhy@stu.edu.cn

Editorial comments:

1. The editor has formatted the manuscript to match the journal's style. Please retain the same.

Reply: Thank you for your comment. We have formatted the manuscript to match the journal's style.

2. Please address specific question marked in the manuscript.

Reply: Thank you for your comment. We have made proper response to the questions marked in the manuscript.

3. Please proofread the manuscript carefully as there are still many grammatical and spelling errors which makes it difficult to follow.

Reply: Thank you for your suggestion. We have read through the whole manuscript and corrected the spelling or grammar errors.

4. Please rewrite lines 79-80, 86-93, 95-103, 105-111 as these matches with previously published literature.

Reply: Thank you for your comment. We have revised in the manuscript to avoid duplication from previously published text.

5. For the protocol section, please write as if you are describing someone how to perform your experiment. Please provide the details in a stepwise manner and use imperative tense throughout.

Reply: Thank you for your suggestion. We have revised in the manuscript.

6. Once done please ensure that the highlight is no more than 2.75 pages including headings and spacings.

Reply: We have changed in the manuscript according to your suggestion.

7. Please revise the table of materials to include all the software used for the experiment.

The table should include the name, company, and catalog number of all relevant materials in separate columns in an xls/xlsx file.

Reply: Thank you for your comment. We have provided all the name, company of the relative software used for the experiment, for some software which is free and open assess, we have provided the website to download,

Copyright permission of reusing figures from Oncology Reports

January 8, 2019

Dear Dr Hao-yu Lin,

Oncol Rep. 2018 Oct;40(4):1897-1906.

Title: Prognostic values of the inhibitor of DNA-binding family members in breast cancer by Zhou *et al*
Figures

Thank you for your inquiry and for your interest in our Journals. Permission is hereby granted for the reproduction of the figures which were published in Oncol Rep. 2018 Oct;40(4):1897-1906, for the purpose outlined in your message.

The original work should be cited.

Please do not hesitate to contact us for anything further.

Yours sincerely,
Spandidos Publications