

Expression Differentiation Is Not Helpful in Identifying Prognostic Genes Based on TCGA Datasets

Ning An,¹ Zhuang Yu,¹ and Xue Yang¹

¹Department of Oncology, The Affiliated Hospital of Qingdao University, Qingdao 266003, Shandong Province, China

A routine pipeline seems very common in many cancer studies that expression differentiation might be helpful in identifying prognostic molecules. There also exists a striking unanimity that molecules upregulated in cancer usually shorten survival, while downregulated ones have the opposite effect. In this study, based on the transcriptional profiles of 18 malignancies, cancer and corresponding adjacent normal tissues were used to calculate differential scores. Cox correlation coefficients of global genes were also calculated to denote survival association. The relationship between expression differentiation and survival association has been extensively studied in 18 malignancy types. Contradictory to our stereotypic research pattern, expression differentiation between cancer and adjacent normal tissues was proven irrelevant to corresponding survival correlation. Surprisingly, the more stringent cutoff we used in differentially expressed gene identification, the less prognostic information we would obtain from the collected gene groups. Moreover, the direction of dysregulated genes in cancer was irrelevant to the direction of corresponding survival correlation. Cancer-normal expression differentiation is irrelevant to genes' survival correlation in multiple cancers and, therefore, not helpful in identifying prognostic genes. For future studies, it is more sensible to look into another alternative rather than collect differentially expressed molecules in the initial step.

INTRODUCTION

Due to the rapid development of molecular cancer research, more and more prognostic factors and therapeutic targets have been discovered, greatly aiding the clinical cancer treatment.^{1,2} Carcinogenesis is a highly complicated process, containing substantial and extensive molecular dysregulations. Therefore, understanding the intricate underlying molecular mechanisms is essentially important in introducing prognostic markers and potential therapeutic targets.

However, through extensive literature searching, there seems a routine pipeline of molecular research in cancer, which has been repetitively utilized in many studies. Generally speaking, one molecule (including mRNA, microRNA, long non-coding RNA, protein, etc.) is identified as significantly and differentially expressed in cancer, with the comparison to corresponding normal, precancerous,

or less aggressive tissues. Then, if this molecule is upregulated in cancer, a variety of *in vivo* or *in vitro* experiments are conducted to prove that its overexpression might accelerate the process of carcinogenesis or shorten patient's survival, for instance, by promoting cellular proliferation or distant metastasis, while its suppression would exhibit the opposite effects.^{3,4} If this molecule is downregulated in cancer, the result would be quite possible that this molecule's overexpression might slow down the process or prolong patient's survival.^{5,6} There is nearly no exception in similar cancer studies based on our literature searching. However, during the process of studying colorectal cancer RNA sequencing (RNA-seq) data from The Cancer Genome Atlas (TCGA) database, we found that the overexpression of some genes dramatically upregulated in cancer was significantly associated with a patient's better prognosis. On the contrary, the overexpression of some significantly downregulated ones was statistically associated with poor prognosis. This routine pipeline and our contradictory finding bring up two questions: (1) what is the relationship between expression differentiation in cancer and survival correlation, and (2) is the direction of survival correlation predictable if we already know the direction of expression differentiation? These questions should be carefully addressed, since if this mainstream impression is not as solid as it appears to be, it will probably cause the misleading effect in cancer research, as well as publication bias.

The unanimity of this aforementioned stereotypic research pattern is astonishing. So far, countless studies have been still strictly following this pattern, but these two questions have been seldom asked. Although this stereotypic pattern is not literally announced by any of these researchers, there is still a possibility that the mainstream ideology could drown a different voice. Therefore, in this study, the pan-cancer analysis of transcriptional expression profiles was conducted to investigate the potential inter-relationship between expression differentiation and survival correlation. Hopefully, this study

Received 27 November 2017; accepted 28 February 2018;
<https://doi.org/10.1016/j.omtn.2018.02.013>.

Correspondence: Xue Yang, The Affiliated Hospital of Qingdao University, No. 16 Jiangsu Road, Shinan District, Qingdao 266003, China.

E-mail: yxue0409@outlook.com

Correspondence: Zhuang Yu, The Affiliated Hospital of Qingdao University, No. 16 Jiangsu Road, Shinan District, Qingdao 266003, China.

E-mail: yuzhuang2002@163.com



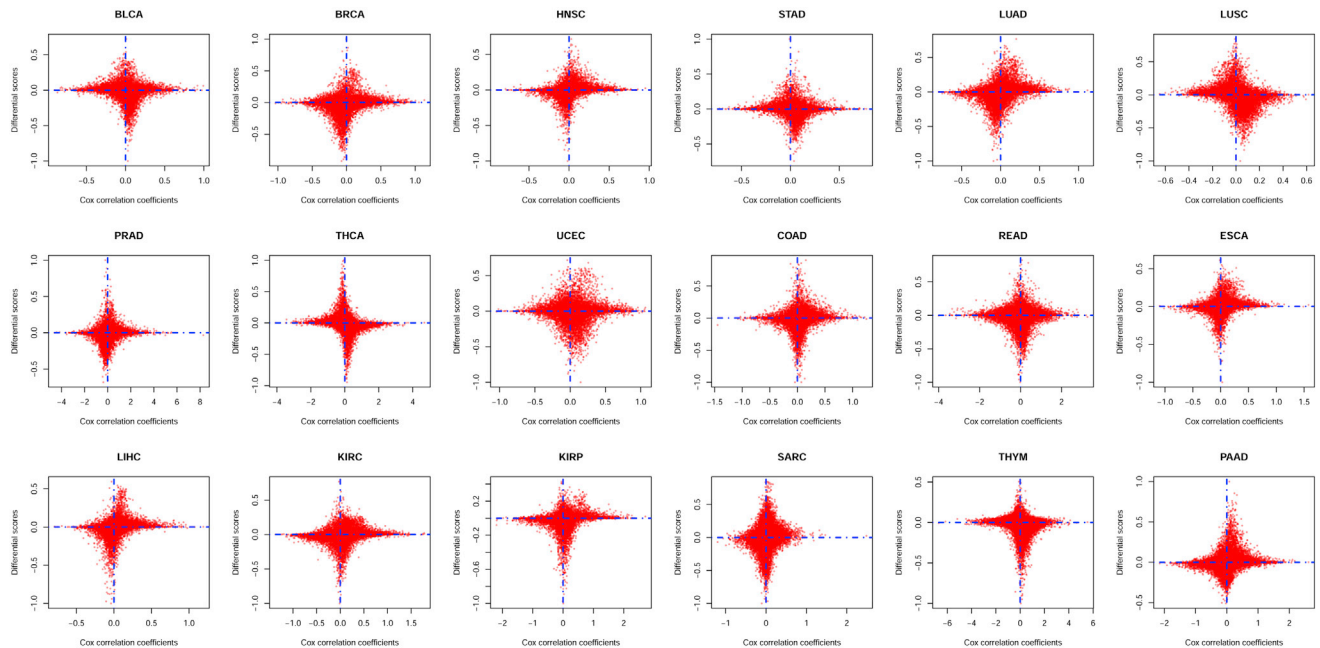


Figure 1. Relationship between Differential Score and Cox Correlation Coefficient of Global Genes

Global genes in the transcriptional profiles of 18 malignancies were projected onto a rectangular coordinate system, in which the x axis represents Cox correlation coefficients and the y axis represents differential score. Abbreviations are as follows: BLCA, bladder urothelial carcinoma; BRCA, breast invasive carcinoma; HNSC, head and neck squamous cell carcinoma; STAD, stomach adenocarcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; PRAD, prostate adenocarcinoma; THCA, thyroid carcinoma; UCEC, uterine corpus endometrioid carcinoma; COAD, colon adenocarcinoma; READ, rectum adenocarcinoma; ESCA, esophageal carcinoma; LIHC, liver hepatocellular carcinoma; KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; SARC, sarcoma; THYM, thymoma; and PAAD, pancreatic adenocarcinoma.

might unshackle cancer researchers from this stereotypic unanimity and let the whole academic community embrace the potentially valuable studies of difference, which are the ones currently swimming against the tide.

RESULTS

Data Retrieval

The transcriptional expression profiles of 18 cancer types were downloaded from TCGA database, containing both cancer samples and normal tissues (Figure S1A). Paired samples (including cancer samples and corresponding adjacent normal tissues) were retrieved to reduce potential individual bias in differentially expressed gene (DEG) identification (Figure S1B). Additionally, clinicopathological factors (including age, gender, stage, race, pT, pN, and pM; Figures S1C–S1I) and overall survival information of cancer patients were also obtained from TCGA database.

Expression Differentiation Was Irrelevant to Survival Status

Differential score (described in Materials and Methods) denoted the status of expression differentiation, and Cox correlation coefficient denoted the interrelation between gene expression and patient's survival status. To sketch the potential relationship between expression differentiation and survival status, we projected all global genes onto a rectangular coordinate system, with the x axis representing

Cox correlation coefficient and the y axis representing differential score in each cancer type (Figure 1). The result indicated that, in all 18 types of malignancies, differential score and Cox correlation coefficients were actually irrelevant, quite contradictory to our stereotypic thinking. Genes with more drastic transcriptional differentiation tended to be concentrated near 0 regarding Cox correlation coefficient, and genes strongly related to patient's overall survival (OS) tended to be equally expressed in cancer and adjacent normal tissues, suggesting identification of differentially expressed genes probably could not be helpful in finding prognostic genes (Figure 1).

Using Stringent Cutoff while DEG Identification Cast away Prognostic Genes

In all 18 types of malignancies, we first normalized differential scores of downregulated genes within the range of $-1-0$ (since all the differential scores of downregulated genes were <0), and we normalized those of upregulated genes within the range of $0-1$, respectively. Then 100 equidistant inter-cutoffs were used to collect genes with more absolute value of differential scores in both downregulated and upregulated gene groups, and, in each selection, the average absolute value of Cox correlation coefficients was also calculated to denote prognostic value of the selected gene groups (Figure 2). The results manifested the general tendency that the more stringent cutoff we used in DEG identification the less prognostic information it left,

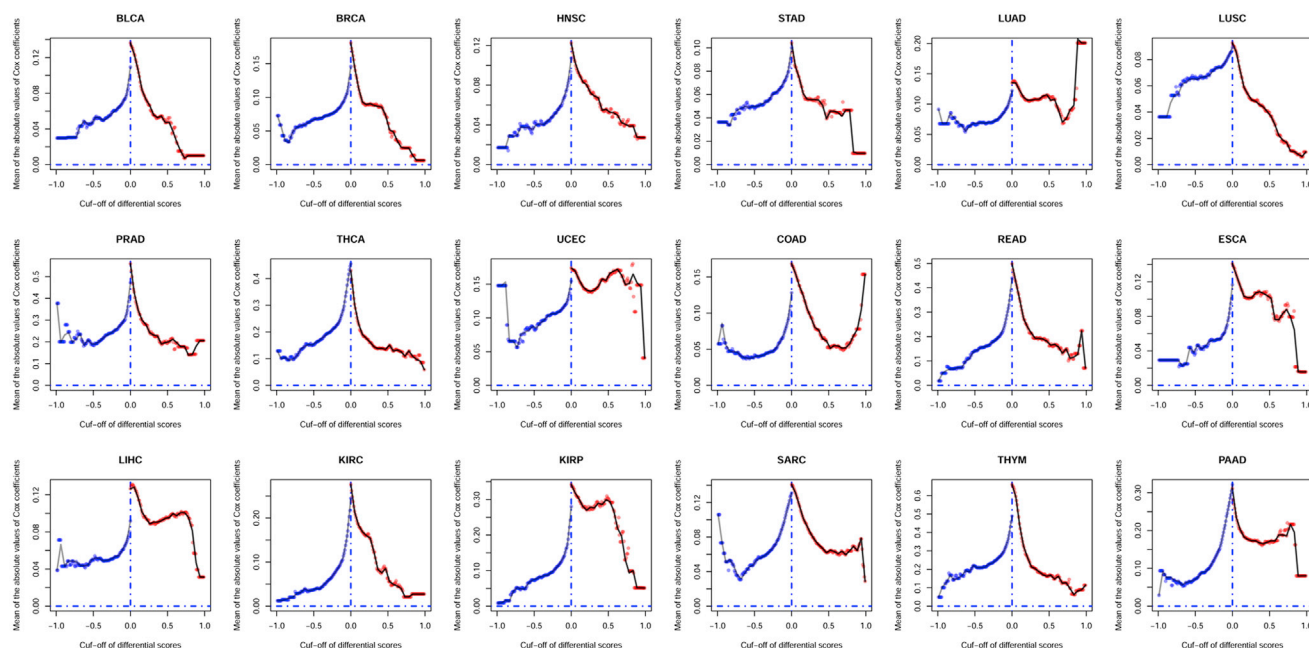


Figure 2. Stringent Cutoff in DEG Identification Reduced Prognostic Value of Selected Genes

Differential scores were normalized within the range of -1 – 1 , respectively. Then 100 equidistant inter-cutoffs were used to collect genes with more absolute value of differential scores in both downregulated (blue dots) and upregulated (red dots) gene groups. The x axis represents the cutoff used in each gene selection, and the y axis represents the mean absolute value of Cox correlation coefficients of the selected genes.

since the average absolute value of Cox correlation coefficients kept decreasing while the absolute value of the differential score cutoff was increasing.

Direction of Expression Differentiation Was Irrelevant to that of Survival Correlation

The top 2,000 downregulated and upregulated genes were collected according to differential score in each cancer type, respectively. The density plots of Cox correlation coefficients are illustrated in Figure 3, with red curves representing upregulated genes and blue curves representing downregulated genes. The result clearly indicates that, under most circumstances, the two groups of top differential genes were similarly distributed, i.e., normal distribution with the mean near 0. In some cancer types, for instance, lung squamous cell carcinoma and pancreatic adenocarcinoma, the two distributions were slightly separated. However, the direction of expression differentiation was irrelevant to that of survival correlation. Since 2,000 was arbitrarily chosen as the number of top DEGs, we also conducted the density analyses with top DEG numbers of 500, 1,000, and 1,500 (see Figures S2–S4), and the results were quite similar, suggesting the distribution of top differential genes was solid.

Expression Differentiation and Survival Correlation of Cancer-Related Genes

The top 100 cancer-related genes with the largest degree were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG)-curated gene list and merged biological network (described

in Materials and Methods). All of these 100 genes were analyzed through paired t test with cancer and corresponding adjacent normal tissues (15 cancer types with ≥ 5 pairs were used in the paired t test) and Cox analysis using cancer samples with OS information. All the p values in t test and Cox analysis were adjusted with the false discovery rate (FDR) method, and an FDR value < 0.001 was regarded as significant (Figure 4). The result indicated that 74% of these cancer-related genes were identified as significantly differentiated in more than half of all cancer types. However, significant cancer genes were quite sparsely dispersed in Cox analysis.

Cancer-Related Genes Showed Distinct Expression Patterns in Cancer and Normal Tissues

To clearly display the clustering distinction between cancer and adjacent normal tissues, 13 types of paired data with ≥ 15 pairs of cancer and corresponding normal samples were used in heatmap construction and principal-component analysis (PCA). These 328 KEGG cancer-related genes could generally cluster cancer and adjacent normal tissues on the basis of expression value in each cancer type, indicating these genes, critical for cancer transformation, underwent substantial expression differentiation in the process of carcinogenesis (Figure 5).

DISCUSSION

Since carcinogenesis is a highly complicated biological process, studies of underlying molecular mechanisms in various types of cancers have always been the highlighted spot in cancer research. The precious

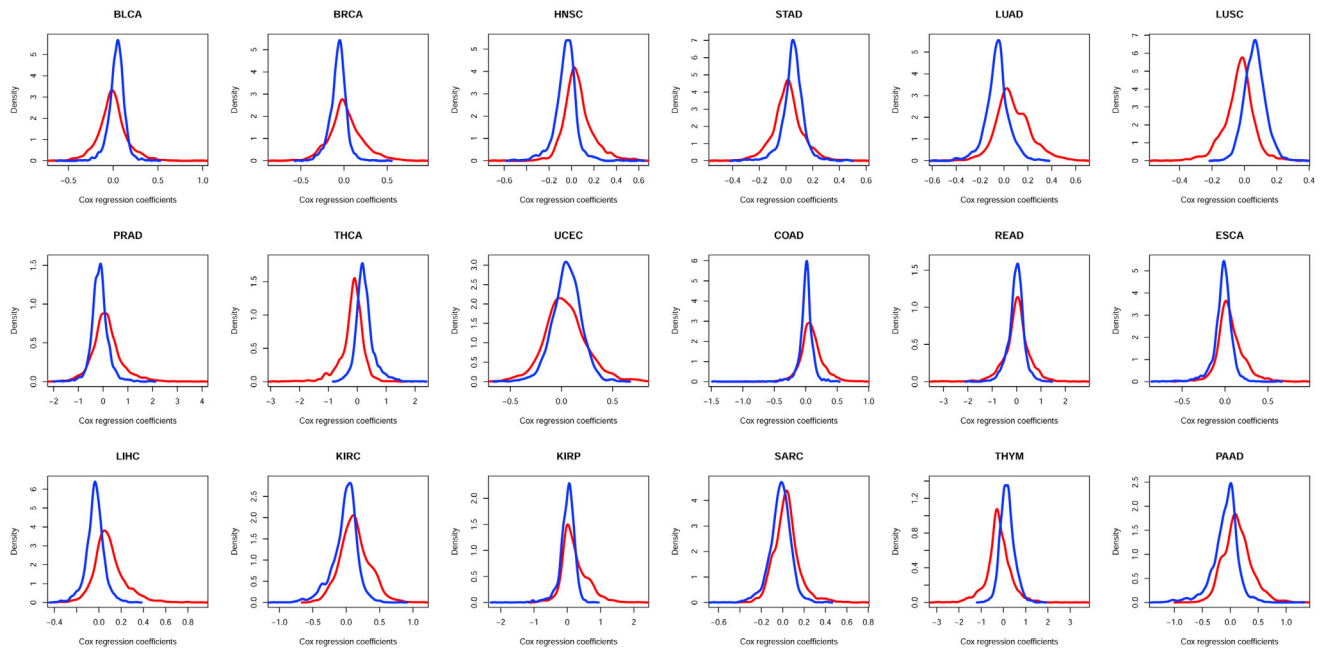


Figure 3. Density Plot of the Top 2,000 Upregulated and Top 2,000 Downregulated Genes on the Basis of Cox Correlation Coefficients

The red curve represents the density plot of upregulated genes, and the blue curve represents that of downregulated genes.

knowledge we obtained from previous cancer studies undoubtedly provided us with the effective therapeutic methods to treat cancer patients. For instance, emerging molecularly targeted drugs ushered us into a brand new era of clinical oncology.^{7–10} However, the aforementioned routine pipeline of molecular cancer research seems to accommodate itself very well under almost all kinds of circumstances. The striking unanimity of cancer-related studies gives us the impression that expression differentiation between cancer and adjacent normal tissues might be helpful in the identification of prognostic molecules; molecules upregulated in cancer might only shorten patient's survival, while downregulated ones might have the opposite effect. This impression is certainly not explicitly explained by any researchers, but the mainstream voice strongly advocates this research pattern, since nearly no exception has been found. Is it the truth or a stereotypic thinking caused by simply following the steps of academic ancestors? It is a good question needing exhaustive investigation, since, if it is not the truth, this research pattern could probably hinder the development of the whole cancer research community. Moreover, the prosperity of molecular oncology and rapid development of biological technology probably make further molecular mechanism studies much easier. It is quite convenient to just switch to another new molecule while using the safest stereotypic pattern of research in order to receive the maximal approval of the whole cancer research community. Therefore, a question occurred to us as to what the potential inter-relationship between molecular differentiation and its influence on patient's survival really is. Now, it is time to change.

As for the first question we proposed previously, we tried to establish the potential linkage between cancer-normal expression differ-

entiation and survival correlation. In numerous cancer studies, the identification of differential expression is the first step in the identification of prognostic indicators. For instance, *AHNAK2* expression was identified as upregulated in clear cell renal cell carcinoma based on the comparison between cancer and adjacent normal tissues, leading to further exploration of its prognostic relevance.¹¹ The under-expression of long non-coding RNA *LINC00261* was also first identified in gastric cancer when comparing with normal adjacent tissues, and it was then proven associated with poor prognosis by promoting distant metastasis.¹² Admittedly, the whole story sounds quite logical and plausible. Molecules undergoing substantial differentiation probably play important roles during carcinogenesis, and, furthermore, they influence a cancer patient's survival accordingly.

TCGA database is an immeasurable source of knowledge launched in 2005, which provides publicly available cancer genomic datasets.¹³ With the aid of TCGA database, we successfully retrieved the RNA sequencing data of 18 types of malignancies, and we tried to establish the linkage between cancer-normal expression differentiation and survival correlation. Paired samples were extracted in all cancer types in order to reduce individual bias. However, the result of our study is quite contradictory to the common impression we received, that is, DEG identification might not be helpful in finding survival-related genes. Actually, the more stringent cutoff we used in DEG identification, the less prognostic information we got (Figures 1 and 2). Moreover, this stereotypic pattern probably misleads cancer researchers into a gloomy detour during the process of discovering prognostic molecules. Therefore, it is more reasonable

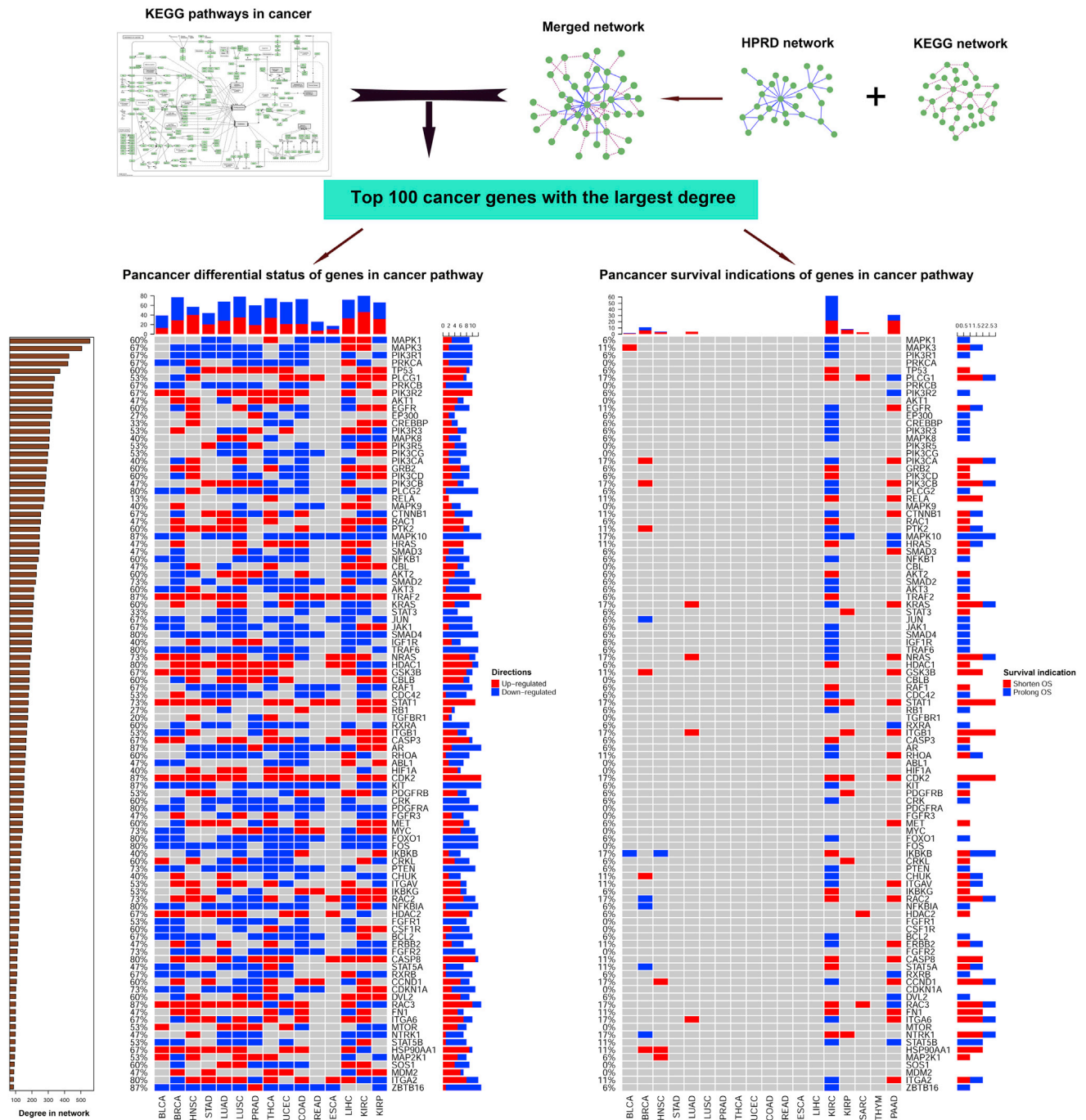


Figure 4. Most of the Top 100 Cancer-Related Genes Were Significant in Differential Analysis but Insignificant in Cox Analysis

KEGG cancer-related genes were downloaded from the KEGG database, and the merged biological network was established through merging the HPRD and KEGG networks. The top 100 cancer-related genes were retrieved according to node degree in the merged biological network. Pan-cancer differential status and survival indication of these genes were illustrated in two heatmaps. Red entries represent significantly upregulated in cancer, or shortening patient's OS, while blue entries represent significantly downregulated in cancer or prolonging OS (FDR < 0.001).

to identify prognostic molecules from all the global genes, rather than starting from DEGs, in case of missing potentially informative ones.

As for the second question, in almost all of these similar studies, upregulated molecules in cancer might always shorten patient's survival, while downregulated molecules might always prolong

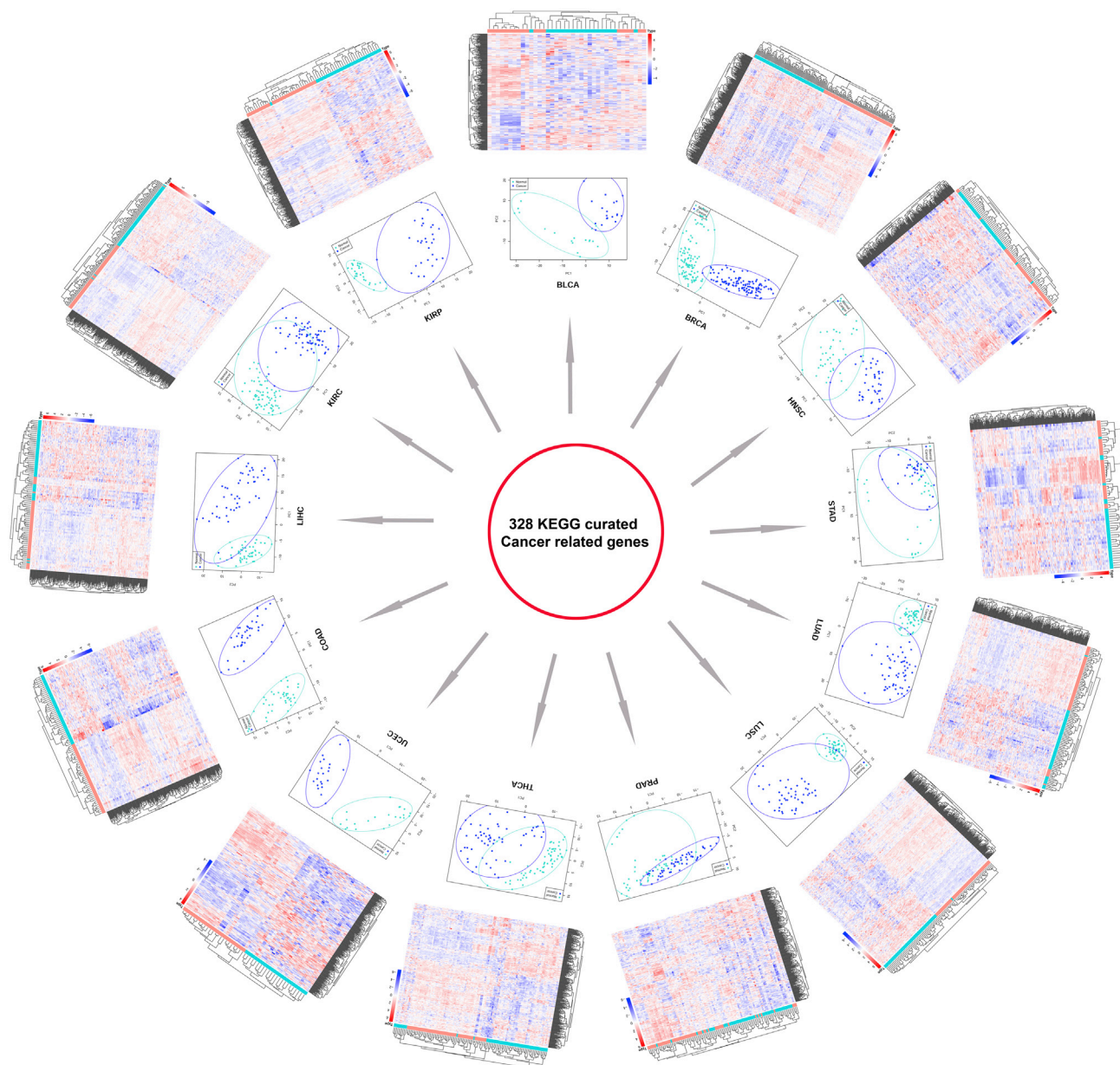


Figure 5. Heatmap Construction and PCA Using KEGG Cancer-Related Genes

328 cancer-related genes were used in heatmap construction (outer layer) and PCA (inner layer) in 13 types of cancers (containing ≥ 15 pairs of samples). In heatmaps, rows represent cancer-related genes and columns represent samples. Both rows and columns were clustered using an unsupervised clustering algorithm. In both heatmaps and PCAs, cancer and adjacent normal tissues could be generally clustered into two groups.

patient's survival.^{14–18} It is very temptingly plausible since bad ones do bad things and good ones do good deeds. However, the result of our study indicated that the direction of DEG's survival correlation was actually independent of its expression differentiation (Figure 3), that is, overexpression of upregulated genes could shorten or prolong patient's survival randomly, and vice versa. However, the striking accordance between the directions of expression differentiation and survival correlations is astoundingly

common based on our literature searching. Additionally, this contradiction against the stereotypic pattern seemed to not only apply for epithelium-derived cancers but also probably for mesenchyme-derived sarcoma. The stereotypic thinking might prevent us from clearly understanding the underlying mechanisms during carcinogenesis, since it probably muffles the voice of difference, which is absolutely harmful for the whole cancer research community.

Identification of prognostic molecules in cancer is certainly a greatly important issue in molecular oncology. In many cancer studies, finding DEGs between cancer and adjacent normal tissues is the first step in finding prognostic genes, and the DEGs upregulated in cancer tend to shorten patient's survival, while downregulated ones tend to prolong it. However, in the present study, the result of this pan-cancer analysis indicated that this methodology we commonly take for granted is at least controversial. The majority of genes with prognostic value are actually not differentially expressed between cancer and normal tissues, and the direction of this expression differentiation is irrelevant to that of survival correlation. Finding prognostic genes among DEGs might miss considerable survival information in the first place. Thus, the validity of this research methodology is questionable so far, and future cancer research should find another alternative.

Furthermore, according to our research, expression differentiation actually seems to be helpful in collecting genes with essential importance in cancer transformation, rather than those with much prognostic information (Figure 4). Most of KEGG-curated cancer genes were significantly dysregulated in the pan-cancer analysis, suggesting DEG identification might help in narrowing down the genes essential during carcinogenesis. However, could this phenomenon also be caused by this stereotypic research pattern existing decades ago while discovering cancer-related genes? It is hard to tell whether this is true. However, there is a big possibility that most of cancer-related genes were probably biasedly identified toward differentially expressed ones, regarding the long history of expression differentiation between cancer and normal tissues. What we do know now from our pan-cancer study is, on one hand, DEG identification between cancer and adjacent normal tissues is probably not a good start for prognostic molecule identification; on the other hand, a DEG's differential direction could not predict its survival correlation.

MATERIALS AND METHODS

Data Preprocessing and Normalization

The RNA sequencing level 3 data (raw counts) of 18 types of malignancies (including bladder urothelial carcinoma, breast invasive carcinoma, head and neck squamous cell carcinoma, stomach adenocarcinoma, lung adenocarcinoma, lung squamous cell carcinoma, prostate adenocarcinoma, thyroid carcinoma, uterine corpus endometrioid carcinoma, colon adenocarcinoma, rectum adenocarcinoma, esophageal carcinoma, liver hepatocellular carcinoma, kidney renal clear cell carcinoma, kidney renal papillary cell carcinoma, sarcoma, thymoma, and pancreatic adenocarcinoma) were retrieved from TCGA database. Although sarcoma is actually not within the scope of cancer, we still included this tumor to see whether the pattern would also apply for mesenchyme-derived malignancies. For each cancer type, the genes with raw read counts <5 in more than half of all samples (including cancer and normal tissues) were eliminated from further analysis, and then the read counts of all entries were log₂ transformed to approximate normal distribution.

Calculation of Differential Score with Paired TCGA Samples

Due to the limited sample numbers, we considered both fold change and the p value of Wilcoxon rank-sum and signed-rank test (Wilcox test) to calculate differential score. For each cancer type, if the paired RNA sequencing data contained <3 pairs of cancer and adjacent normal samples, the differential score of a particular gene *i* was simply the log₂ of fold change (FC_i) in cancer comparing adjacent normal tissues, since the exact p value of the Wilcox test could be hardly calculated; if the data contained ≥ 3 pairs of samples, differential score D_i for gene *i* was calculated as follows:

$$D_i = \log_2(FC_i) * (-\log_{10}(P_i)), \quad (\text{Equation 1})$$

in which, P_i represented the p value of the Wilcox test.

Collection of Cancer-Related Genes

First, the merged a priori knowledge-based biological network was established. The protein-protein interaction network was downloaded from the Human Protein Reference Database (HPRD), and the KEGG network was constructed with Bioconductor package KEGGgraph. Therefore, the gene regulatory network was established by merging the HPRD and KEGG networks, including 10,340 nodes and 60,642 edges after eliminating self-loops and duplicated edges; this method has been used in our previous studies.^{19–21} The 328 genes within the KEGG pathways in cancer (KEGG: hsa05200) were downloaded from the KEGG website (<http://www.kegg.jp/>), of which the node degree number in the merged network was used to quantify the importance in the process of cancer transformation.

Heatmap Construction and PCA Using KEGG Cancer-Related Genes

The heatmaps of 328 KEGG cancer-related genes were constructed in each paired cancer data. Rows represent cancer-related genes while columns represent samples. Both rows and columns were clustered using an unsupervised clustering algorithm. As for a given cancer type, the first and second principal components (PC1 and PC2) were calculated through PCA of gene expression values across corresponding samples. As PC1 and PC2 capture the majority of total variance, they represent a summary measure for the overall expression status of all the samples. In this manner, all the samples were projected onto the rectangular coordinate system with the PC1 and PC2 as x and y axes, respectively, and the spatial distance denoted the overall distinction on the basis of gene expression values.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and can be found with this article online at <https://doi.org/10.1016/j.omtn.2018.02.013>.

AUTHOR CONTRIBUTIONS

N.A., Z.Y., and X.Y. participated in the study design and writing the paper. N.A. carried out the sample selection, algorithm construction, and data analysis. All authors have read and approved the manuscript

and its contents and are aware of responsibilities connected to authorship.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The present study was funded by a grant from the Natural Science Foundation of Shan Dong province, China (ZR201702170463).

REFERENCES

1. Das, V., Kalita, J., and Pal, M. (2017). Predictive and prognostic biomarkers in colorectal cancer: A systematic review of recent advances and challenges. *Biomed. Pharmacother.* *87*, 8–19.
2. Rai, V., Abdo, J., Alsuwaidan, A.N., Agrawal, S., Sharma, P., and Agrawal, D.K. (2018). Cellular and molecular targets for the immunotherapy of hepatocellular carcinoma. *Mol. Cell. Biochem.* *437*, 13–36.
3. Gibney, G.T., Aziz, S.A., Camp, R.L., Conrad, P., Schwartz, B.E., Chen, C.R., Kelly, W.K., and Kluger, H.M. (2013). c-Met is a prognostic marker and potential therapeutic target in clear cell renal cell carcinoma. *Ann. Oncol.* *24*, 343–349.
4. Yu, G., Herazo-Maya, J.D., Nukui, T., Romkes, M., Parwani, A., Juan-Guardela, B.M., Robertson, J., Gauldie, J., Siegfried, J.M., Kaminski, N., and Kass, D.J. (2014). Matrix metalloproteinase-19 promotes metastatic behavior in vitro and is associated with increased mortality in non-small cell lung cancer. *Am. J. Respir. Crit. Care Med.* *190*, 780–790.
5. Leclerc, B.G., Charlebois, R., Chouinard, G., Allard, B., Pommey, S., Saad, F., and Stagg, J. (2016). CD73 Expression Is an Independent Prognostic Factor in Prostate Cancer. *Clin. Cancer Res.* *22*, 158–166.
6. Jin, H., Wang, C., Jin, G., Ruan, H., Gu, D., Wei, L., Wang, H., Wang, N., Arunachalam, E., Zhang, Y., et al. (2017). Regulator of Calcineurin 1 Gene Isoform 4, Down-regulated in Hepatocellular Carcinoma, Prevents Proliferation, Migration, and Invasive Activity of Cancer Cells and Metastasis of Orthotopic Tumors by Inhibiting Nuclear Translocation of NFAT1. *Gastroenterology* *153*, 799–811.e33.
7. Peters, S., Camidge, D.R., Shaw, A.T., Gadgeel, S., Ahn, J.S., Kim, D.W., Ou, S.I., Pérol, M., Dziadziuszko, R., Rosell, R., et al.; ALEX Trial Investigators (2017). Alectinib versus Crizotinib in Untreated ALK-Positive Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* *377*, 829–838.
8. Davies, M.A., Saig, P., Robert, C., Grob, J.J., Flaherty, K.T., Arance, A., Chiarion-Sileni, V., Thomas, L., Lesimple, T., Mortier, L., et al. (2017). Dabrafenib plus trametinib in patients with BRAF^{V600}-mutant melanoma brain metastases (COMBI-MB): a multicentre, multicohort, open-label, phase 2 trial. *Lancet Oncol.* *18*, 863–873.
9. Vermorken, J.B., Mesia, R., Rivera, F., Remenar, E., Kawecki, A., Rottey, S., Erfan, J., Zabolotnyy, D., Kienzer, H.R., Cupissol, D., et al. (2008). Platinum-based chemotherapy plus cetuximab in head and neck cancer. *N. Engl. J. Med.* *359*, 1116–1127.
10. Bellmunt, J., de Wit, R., Vaughn, D.J., Fradet, Y., Lee, J.L., Fong, L., Vogelzang, N.J., Climent, M.A., Petrylak, D.P., Choueiri, T.K., et al.; KEYNOTE-045 Investigators (2017). Pembrolizumab as Second-Line Therapy for Advanced Urothelial Carcinoma. *N. Engl. J. Med.* *376*, 1015–1026.
11. Wang, M., Li, X., Zhang, J., Yang, Q., Chen, W., Jin, W., Huang, Y.R., Yang, R., and Gao, W.Q. (2017). AHNAK2 is a Novel Prognostic Marker and Oncogenic Protein for Clear Cell Renal Cell Carcinoma. *Theranostics* *7*, 1100–1113.
12. Fan, Y., Wang, Y.F., Su, H.F., Fang, N., Zou, C., Li, W.F., and Fei, Z.H. (2018). Retraction Note to: Decreased expression of the long noncoding RNA LINC00261 indicate poor prognosis in gastric cancer and suppress gastric cancer metastasis by affecting the epithelial-mesenchymal transition. *J. Hematol. Oncol.* *11*, 2.
13. Tomczak, K., Czerwińska, P., and Wizniewski, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. (Pozn.)* *19* (1A), A68–A77.
14. Zhou, Z., Ji, Z., Wang, Y., Li, J., Cao, H., Zhu, H.H., and Gao, W.Q. (2014). TRIM59 is up-regulated in gastric tumors, promoting ubiquitination and degradation of p53. *Gastroenterology* *147*, 1043–1054.
15. Nguyen, L.H., Robinton, D.A., Seligson, M.T., Wu, L., Li, L., Rakheja, D., Comerford, S.A., Ramezani, S., Sun, X., Parikh, M.S., et al. (2014). Lin28b is sufficient to drive liver cancer and necessary for its maintenance in murine models. *Cancer Cell* *26*, 248–261.
16. Cui, B.G., Campagne, A., Bell, G.W., Lembo, A., Orso, F., Lien, E.C., Bhasin, M.K., Raimo, M., Hanson, S.E., Marusyk, A., et al. (2014). MSC-regulated microRNAs converge on the transcription factor FOXP2 and promote breast cancer metastasis. *Cell Stem Cell* *15*, 762–774.
17. Chen, J., Rajasekaran, M., Xia, H., Zhang, X., Kong, S.N., Sekar, K., Seshachalam, V.P., Deivasigamani, A., Goh, B.K., Ooi, L.L., et al. (2016). The microtubule-associated protein PRC1 promotes early recurrence of hepatocellular carcinoma in association with the Wnt/β-catenin signalling pathway. *Gut* *65*, 1522–1534.
18. Chang, Y.W., Chiu, C.F., Lee, K.Y., Hong, C.C., Wang, Y.Y., Cheng, C.C., Jan, Y.H., Huang, M.S., Hsiao, M., Ma, J.T., and Su, J.L. (2015). CARMA3 Represses Metastasis Suppressor NME2 to Promote Lung Cancer Stemness and Metastasis. *Am. J. Respir. Crit. Care Med.* *192*, 64–75.
19. Feng, L., Tong, R., Liu, X., Zhang, K., Wang, G., Zhang, L., An, N., and Cheng, S. (2016). A network-based method for identifying prognostic gene modules in lung squamous carcinoma. *Oncotarget* *7*, 18006–18020.
20. An, N., Yang, X., Zhang, Y., Shi, X., Yu, X., Cheng, S., Zhang, K., and Wang, G. (2016). Cell cycle related genes up-regulated in human colorectal development predict the overall survival of late-stage colorectal cancer patients. *Mol. Biosyst.* *12*, 541–552.
21. An, N., Yang, X., Cheng, S., Wang, G., and Zhang, K. (2015). Developmental genes significantly afflicted by aberrant promoter methylation and somatic mutation predict overall survival of late-stage colorectal cancer. *Sci. Rep.* *5*, 18616.