

# Identification of Significant Genes and Pathways Related to Lung Cancer via Statistical Methods

Yuhang Wu

Cranbrook Educational Community, Bloomfield Hills, MI, USA

Email: awutrsxa@outlook.com

**How to cite this paper:** Wu, Y.H. (2018) Identification of Significant Genes and Pathways Related to Lung Cancer via Statistical Methods. *Advances in Bioscience and Biotechnology*, 9, 397-408. <https://doi.org/10.4236/abb.2018.99028>

**Received:** August 2, 2018

**Accepted:** August 27, 2018

**Published:** August 30, 2018

Copyright © 2018 by author and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Cancer genomic research is a relatively new method. It has shown great potential but faces certain challenges. Researchers often have to deal with tens of thousands of genes with a relatively small sample size of patient cases—a dilemma referred to as the “Curse of Dimensionality” [1]—and it makes it hard to learn the data well because of relatively sparse data in high dimensional space. To deal with the dilemma, this study uses p-values of individual genes for pathway enrichment to find statistically significant pathways. The aim of this study is to find significant genes and biological pathways that are related to lung cancer by statistical method and pathway enrichment analysis. Several significant genes, such as WNT2B, VAV2, and significant pathways, such as Metabolism of xenobiotics by cytochrome P450-Homo sapiens (human) and Fatty acid degradation-Homo sapiens (human), are found to be both statistically significant and biological studies supported. Significant genes-including TESK2, C5orf43, and ZSCAN21—and significant pathways such as Pentose and glucuronate interconversions-Homo sapiens (human), are found to be new cancer-related genes and pathways that worth laboratory studies. The idea and method used in this research can be applied to find more significant genes and pathways that worth study experimentally.

## Keywords

Cancer Genomic, Genes and Pathways, Curse of Dimensionality, Biostatistics

## 1. Introduction

In the 21<sup>st</sup> century, cancer research, integrated with biology, genetics, cytology and statistics, continues to be a hot spot. Since last century, many researchers

have been working in this field on clinical observations and theoretical deduction. Among many researches, generic aspects of such, a relatively new method for learning causes and preventions for cancer, have begun to show its potential. The study of cancer genomics can reveal abnormalities in certain genes that drive the development and growth of lung cancer [2]. Over the past decade, large-scale research projects have been launched but faced certain challenges. Acquiring high-quality biological samples needed for genomic studies, managing and analyzing the vast amounts of data involved, and finally having a converged genetic abnormality result all add challenges to genomic method of research [3] [4].

Lung cancer is one of the cancers that caused most deaths around the world each year. About 1 out of 4 cancer deaths are from lung cancer. Each year, more people died of lung cancer than of colon, breast, and prostate combined [5]. Unsurprisingly it has been studied intensively. People have come to accept that smoking is the major cause of lung cancer as early as 1930s [6], but recent studies have shown that genetic aberrations can also raise cancer risks substantially [7], though how specific genes affect lung cancer remains vastly understudied.

In this paper, we focus on lung cancer and gene expressions associated with it. Our dataset is retrieved from GSEA, which collects data in collaboration with National Cancer Institute, National Institutes of Health, National Institute of General Medical Sciences etc. Normalized RNA sequencing data of 868 lung cancer patients is studied statistically with patients' clinical observations. On the basis of clinical observations and gene expressions of lung cancer patients, significant genes and gene pathways are investigated. Two specific clinical data, recurrence free survival (RFS) and whether the patient has a new tumor event after initial treatment (cancer recurrence), are chosen to run regression on. RFS is a continuous variable, and we apply linear regression to identify relevant genes and pathways [8]; cancer recurrence is a binary variable, so logistic regression was applied [9]. The resulting p-values of genes are then analyzed by Gene Set Enrichment Analysis (GSEA), an online pathway enrichment tool, to find statistically significant gene pathways [10].

## 2. Methods

### 2.1. Multivariate Linear Regression

Linear regression is a linear approach for modeling the relationship between a scalar dependent variable and one or more independent variables. Multiple linear regression takes a vector of independent variables and find each independent variables' associations with the dependent variable. Here for each patient, let  $Y$  denotes the dependent variable (RFS) for each patient,  $\beta$  denote a vector of associations between dependent and independent variables, and let  $X$  denotes the vector of independent variable (expressions of genes):

$$X_i = (1, x_{i1}, \dots, x_{ip}) \quad (1)$$

$$\beta = (\beta_0, \beta_1, \dots, \beta_p) \quad (2)$$

So a multiple linear model is fit under the assumption:

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \epsilon \quad (3)$$

Where  $\epsilon$  is natural noise with following distribution:

$$\epsilon \sim N(0, \sigma) \text{ i.i.d.} \quad (4)$$

where  $N$  stands for normal distribution, the distribution that is assumed for real random values, and i.i.d. stands for “independently and identically distributed”, as natural noises have the same probability distribution and are mutually independent.

When  $\beta$  is positive,  $x$  and  $y$  has a positive relationship, and as  $x$  increases  $y$  increases; vice versa. Note that the explanatory variable is also given an intercept  $\beta_0$ .  $\beta$  and  $\beta_0$  are found by Least Square Estimation, rigorously defined by:

$$\min_{\beta} \sum_{i=1}^n (y_i - X_i' \beta)^2 \quad (5)$$

Under such assumption,  $\beta$  is found using the closed form solution:

$$\beta = (X'X)^{-1} X'y \quad (6)$$

Then a linear model is fit to describe the relationship between dependent variable and a vector of independent variables. A p-value is calculated for each  $\beta$  to test its statistical significance. More analysis of p-values will be discussed in Section 2.2.

## 2.2. Multivariate Logistic Regression

Logistic regression is a regression model dealing with categorical dependent variables, which can only take two values, “0” or “1”. It estimates the probability of the binary response based on one or more explanatory variables. There are only two possible outcomes: having ( $y = 1$ ) or not having ( $y = 0$ ) a new tumor. The sum of the two probabilities is assumed to be 1, as stated:

$$p(y_i = 1) = p_i, p(y_i = 0) = 1 - p_i \quad (7)$$

Multiple logistic regression takes a vector of independent variables and find each independent variables' contribution to the classification result of the dependent variable. A linear model could still be used for logistic regression except one problem: the probabilities would be allowed to go smaller than 0 and bigger than 1. So the joint probability takes a logistic transformation:

$$p_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}, \quad (8)$$

where  $X$  denotes the vector of independent variables (expressions of genes) and  $\beta$  denotes the contribution an independent variable makes in determining the classification of the dependent variable:

$$X_i = (1, x_{i1}, \dots, x_{ip}) \quad (9)$$

$$\beta = (\beta_0, \beta_1, \dots, \beta_p) \quad (10)$$

Unlike linear regression, a Least Square Estimation cannot be found in logistic regression, so instead a Maximum Likelihood Estimation is sought. The following likelihood function is maximized:

$$\prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \quad (11)$$

Also unlike linear regression, there is no closed form solution for the optimized  $\beta$ , so the  $\beta$  that maximizes the likelihood function can only be found using iterative algorithm such as the Newton Raphson algorithm [11]. A logistic model is then fit to describe the contribution each independent variable makes in deciding the category of the dependent binary variable. A p-value is then calculated for each  $\beta$  to test its statistical significance. More analysis of p-values will be discussed in Section 2.3.

### 2.3. Correction for Multiple Testing

Multiple testing is when a set of statistical inferences is tested simultaneously. In this study, each gene is given a p-value through multiple testing of all models fit. When a model is fit for each gene, a hypothesis test is performed. Null hypothesis (H0) states that the real coefficient ( $\beta$ ) is zero (the value of this specific independent variable does not have association with the value of the dependent variable), and alternative hypothesis (H1) states the opposite.

From each null hypothesis, an T population distribution is assumed. Then by using a t-test, a p-value—how likely the T statistic, resulted from the alternative hypothesis, is to occur under F distribution—is calculated. In other words, p-value measures the probability of obtaining a result at least as extreme as the sample data, assuming the null hypothesis is true. So when the p-value is small, it is reasonable to reject the null hypothesis. In this study, p-value is used as the metric for selecting top genes and pathways.

In the case of multiple testing, a general rule of 0.05 as cutoff for p-values for statistical significance is not sufficient. Consider a large test number such as 10,000. If 0.05 is used as cutoff, it is expected that 500 true hypothesis to be rejected even though all null hypotheses are true. Also, because p-value cannot be larger than 1, using 0.05 cutoff would yield many “statistical significant” cases just by chance. As a result, an accommodated cutoff would be applied for multiple testing.

There are multiple ways to calculate the cutoff for p-value. For example, Bonferroni cutoff is a rigid correction for multiple comparison that controls the overall probability of a single false discovery. It adjusts the significance cutoff to  $\alpha/m$  where  $\alpha$  is the normal 0.05 cutoff and  $m$  is the number of hypothesis. In this study, a Bonferroni cutoff would be extremely small because there are tens of thousands of genes, thus an  $m$  value of tens of thousands [12].

Another correction cutoff is False Discovery Rate (FDR), designed to control the expected proportion of rejected null hypotheses that are false. It is less strin-

gent than the Bonferroni cutoff [13]. According to Benjamini-Hochberg (BH)'s method, the cutoff is calculated as:

$$p_{BH} = \max \left\{ P_i : P_i \leq \alpha \frac{i}{m} \right\}$$

Both corrections are used in this study, and the results will be discussed in Section 3.

## 2.4. Pathway Enrichment Analysis

With the complexity of traits of genes, it is sometimes difficult to interpret the p-values of all genes. Therefore, gene set pathway enrichment analysis is also used. It is useful for finding out if the differentially expressed genes are associated with a certain biological process or molecular function, so that an association between cancer and certain gene pathways can be drawn. Here, a list of genes and their p-values is analyzed using Gene Set Enrichment Analysis (GSEA) through KEGG pathway.

## 3. Results and Discussions

The variables of interest in this study, RFS and Cancer Recurrence, are both of great importance in cancer study. RFS, as known as DFS (disease-free survival), is the length of time after primary treatment for a cancer ends that the patient survives without any signs or symptoms of that cancer [14]. Cancer Recurrence is the primary way of judging whether a treatment works. Both variables provide a way to examine how well a treatment works for the patient. Specific to this study, these two variables provide a standard to determine whether certain differently expressed genes or certain pathways have high correlation with whether a treatment is useful.

### 3.1. Gene Level Results

The sorted p-values (ascending order) for both regressions are plotted to help visualization. There were three to four thousand p-values that passed the rule-of-thumb 0.05 cutoff in both regressions, so we proceed to use more rigorous corrections of cutoff: FDR corrections and Bonferroni cutoff.

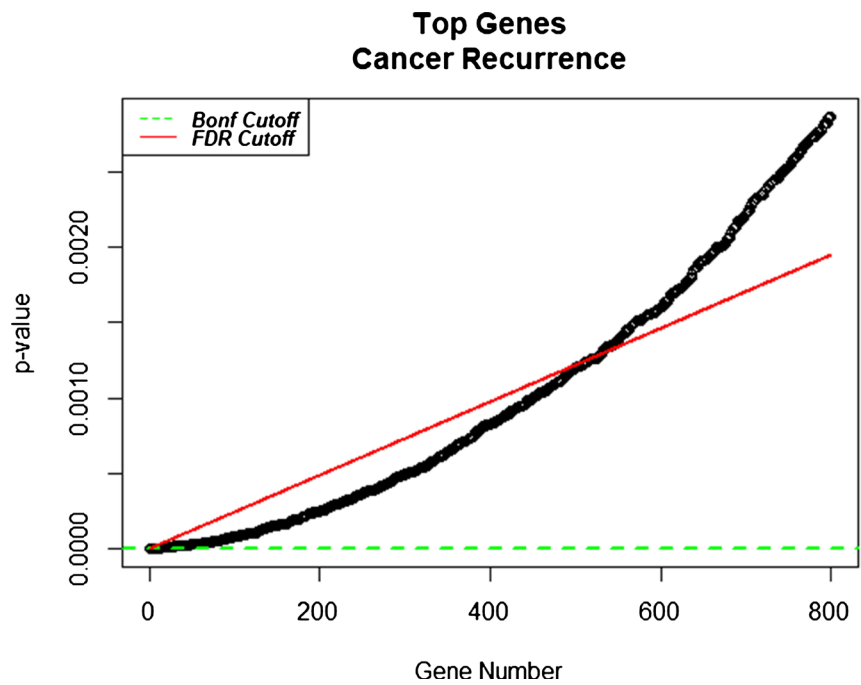
#### 3.1.1. Logistic Regression

The largest Pi found in FDR correction is that of 528<sup>th</sup> p-values: 1.280e-3; the Bonferroni cut off is 2.435e-6 (**Figure 1**).

The top 10 genes are shown in **Table 1** as following (Note that there are 528 genes that passed the FDR corrections, and p-values passed Bonferroni cutoff are highlighted):

After studying existing literatures, we find some medical and laboratory evidence for correlation between some genes above and lung cancer.

WNT2B, the very top gene on the list, is a protein that encodes a member of the wingless-type MMTV integration site (WNT) family of highly conserved,



**Figure 1.** P-value-gene number of cancer recurrence.

**Table 1.** Top 10 genes and its P-value.

Top 10 genes	Table Column Head	
	Gene Name	P-value
1	WNT2B	4.395e-7
2	OPN3	8.234e-7
3	TESK2	1.344e-6
4	C5orf43	1.889e-6
5	LOC541473	2.779e-6
6	STK11	3.000e-6
7	PTPRS	4.010e-6
8	KRT15	4.521e-6
9	RDH14	4.842e-6
10	LOC221442	4.955e-6

secreted signaling factors. WNT family members function in a variety of developmental processes including regulation of cell growth and differentiation [15]. A study of WNT2B expression in human cancer suggested that WNT2B (and mostly only WNT2B but not WNT2 or WNT2B, which also belong to WNT Family), was differently expressed in breast cancer, teratocarcinoma, and gastric cancer [16]. Another study that intended to find effects of WNT2B in GLUT1 overexpressing also concluded that WNT2B plays a role in tumorigenesis and chemotherapy resistance [17]. Also mentioned by this study was that GLUT1 overexpression has been detected in several malignances including breast cancer

and lung cancer. Researches have also concluded that WNT2B is oncogenic and is especially related to non-small cell lung cancer [18].

OPN3 is a protein coding gene, a member of the guanine nucleotide-binding (G protein)-coupled receptor superfamily [19]. A study to find correlation between OPN3 and 5-fluorouracil treatment found that therapeutic strategies targeting OPN3 may improve chemotherapy function [20]. Another study pointed directly to the role OPN3 regulatory sequences play in oncogenesis and drug resistance [21].

Other genes on the list, such as TESK2 and C5orf43, have not been studied in laboratory or medical researches thoroughly but have been connected to cancer indirectly in other existing literatures. Genes further down the lists are also worthy of studying.

### 3.1.2. Linear Regression

The largest Pi found in FDR correction is that of 374<sup>th</sup> p-values:  $9.106e-4$ ; the Bonferroni cut off is  $2.435e-6$  (Figure 2).

Top 10 genes are shown in Table 2 as following (Note that there are 374 genes that passed the FDR corrections, and p-values passed Bonferroni cutoff are highlighted):

After studying existing researches, we find some medical and laboratory evidence for correlation between some genes above and lung cancer.

LSM2 is a Protein Coding Gene, a member of LSm family of RNA-binding proteins. It binds specifically to the 3-terminal U-tract of U6 snRNA, and it may have a correlation with pre-mRNA splicing [22]. LSm family has been found to have a strong impact on breast cancer [23], and it has also been reported that overexpression of LSm1 exhibits in many lung cancer cases [24].

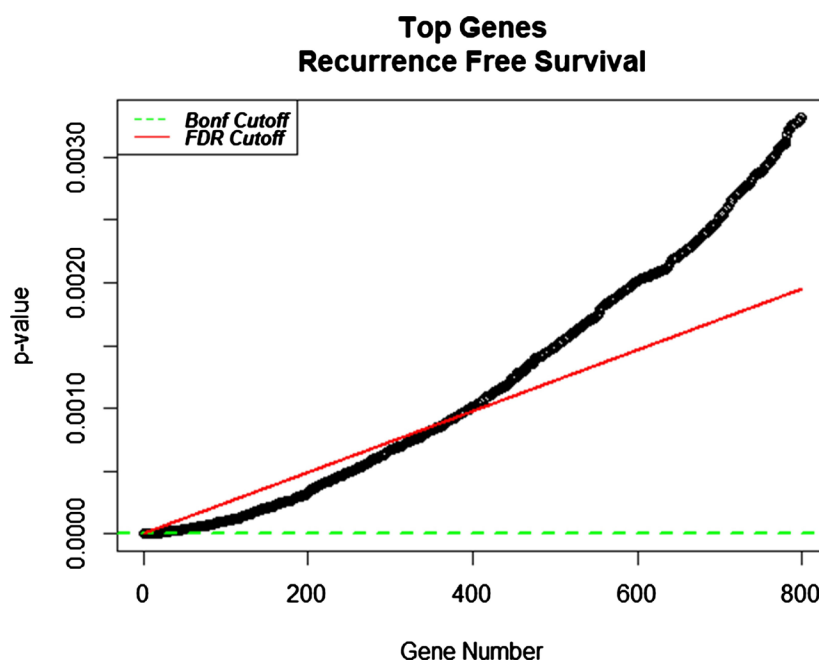


Figure 2. P-value-gene number of RFS.

**Table 2.** TOP 10 Genes and its P-value.

Top 10 genes	Table Column Head	
	Gene Name	P-value
1	NACC2	6.653e-7
2	ZSCAN21	8.739e-7
3	LSM2	1.474e-6
4	VAV2	1.717e-6
5	NCOR2	2.656e-6
6	NCOR1	2.861e-6
7	BUD31	2.913e-6
8	PMEPA1	3.306e-6
9	SFRS3	3.726e-6
10	C15orf28	4.993e-6

VAV2 is a Protein Coding Gene, a member of VAV guanine nucleotide exchange factor family of oncogenes. Its transcripts were found in most tissues, and it plays an important role in angiogenesis [25]. A research of a well-known cancer-related protein suggests that VAV2 facilitates cancer cell motility and a molecular level. Vimentin is an intermediate filament protein whose expression correlates with cancer diseases. In exploring the molecular role of vimentin in cancer cell motility, they find that VAV2 localizes to vimentin-positive focal adhesions in lung cancer cells. Based on their study, the group built a model of vimentin-VAV2 pathway and proposed it as a potential novel regulator of lung cancer motility [26].

Both NCOR1 and NCOR2 are Protein Coding Genes that belong to family of nuclear receptor corepressors. Both of them encode proteins that mediate transcriptional repression by impeding the access of basal transcription factors [27] [28]. In a study of transcription factor corepressors, the group investigated the roles of these corepressors in cancers, and pointed NCOR1 and NCOR2 as examples of how features of transcriptional rigidity affect cancer cells [29].

BUD31 is a Protein Coding Gene that is associated with transcriptional factor activity and sequence-specific DNA binding [30]. MYC overexpression is one of the most common drivers of human cancer, but it is so far recalcitrant to therapeutic inhibition. To tackle this problem, Hsu and her team conducted a research and found that spliceosome could be a vulnerability of MYC. BUD31, as identified by the team, is a component of core spliceosome required for its assembly and catalytic activity, thus a MYC-synthetic lethal gene in human cells [31].

Interestingly, top two genes of the regression result cannot be related to many existing researches. In fact, few researches have been focused on the correlation between these genes and cancer whatsoever (besides NACC2 was distantly mentioned to be associated with bladder cancer survival [32]). According to the sig-



nificance of genes following them on the list, these two top genes, NACC2 and ZSCAN21, probably should be attempted more specifically to find correlation with cancer.

### 3.2. Pathway Enrichment Results

Results of analyses of the two variables appear to be very similar in terms of the top pathways and genes. Five pathways appear in the top six of both results. (Table 3).

After studying existing researches, we find some medical evidence for correlation between some pathways above and cancer.

Drug metabolism—cytochrome P450-Homo sapiens (human), id hsa00982, belongs to class of Xenobiotics biodegradation, Metabolism [33]. A 1991 research related drug metabolism with cancer predictions, and proposed to use this pathway of genes to predict individual risk of cancer [34].

Metabolism of xenobiotics by cytochrome P450-Homo sapiens (human), id hsa00980, belongs to class of Xenobiotics biodegradation, Metabolism [33]. A research to find a new way of cancer diagnosis found about how a specific expression of cytochrome P450 is related to multiple cancers including lung cancer. The team found that a cloned dioxin-inducible form of the cytochrome P450 family, CYP1B1, expresses at a high frequency in a wide range of cancers, so a test on this pathway can provide a new way of diagnosis [35].

Fatty acid degradation—Homo sapiens (human), id hsa00071, belongs to class of Lipid metabolism, Metabolism [33]. A study of cancer from lipid perspective suggests that through control of fatty acid pathway, cancer cell proliferation can be controlled since it requires fatty acids for synthesis of membranes and signaling molecules [36].

The other two of the five pathways on the lists, though appeared top, have not been researched for cancer association. There is a great chance that new perspectives of cancer will be found through these two pathways.

## 4. Discussion

By now, extensive researches have been conducted about prevention, diagnosis and therapy of lung cancers. Genetics provide a wide range of opportunities for cancer research, and so far the potential still exists for much further studies. In

**Table 3.** Analysis of the two variables in terms of the top pathways and genes.

	Name	#Gene	FDR (Whether new tumor occurred)	FDR (recurrence free survival)
hsa00040	Pentose and glucuronate interconversions—Homo sapiens (human)	31	0.242	0.18
hsa00053	Ascorbate and aldarate metabolism—Homo sapiens (human)	25	0.288	0.328
hsa00982	Drug metabolism—cytochrome P450-Homo sapiens (human)	66	0.339	0.296
hsa00980	Metabolism of xenobiotics by cytochrome P450-Homo sapiens (human)	70	0.393	0.423
hsa00071	Fatty acid degradation—Homo sapiens (human)	42	0.413	0.475

this study, some significant genes and pathways related to lung cancer are pointed out with existing medical proof. On the other hand, some genes or pathways that appeared to be equally or even more significant have not been studied yet. For example, TESK2, C5orf43, NACC2, and ZSCAN21 should be attempted to find specific associations with lung cancer as many genes further down the list have been proved to be cancer-related. Also, the top two pathways of enrichment results, hsa00040, Pentose and glucuronate interconversions—Homo sapiens (human), and hsa00053, Ascorbate and aldarate metabolism—Homo sapiens (human), should be further investigated biologically given the following pathways' proven correlations with cancer.

In this paper, an important method is pathway enrichment analysis, through which pathways related to lung cancer can be studied. However, compared to that with single genes, association of pathways with cancers has been studied relatively less intensively. Fewer literatures can be found to support the statistical result of this paper in pathway session, whereas many resources can support the findings at gene level and it was only due to the amount of time that more were not investigated. It is understood that pathway is a far more sophisticated system than a single gene, so this result was more or less expected.

The two major statistical methods used in this paper, linear regression and logistic regression, seemed to be doing a great job as they pointed to some significant results that were supported by laboratorial studies. More complicated statistical method could probably be used, and better biological interpretations could probably be added to draw more comprehensive connections among top genes. Overall, this is one of the many steps that human take to understand and finally defeat cancer. As more advanced developments occur in statistics, biology and genetics, a world without cancer approaches steadily.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Keogh, E. and Mueen, A. (2011) Curse of Dimensionality. *Encyclopedia of Machine Learning*, Springer US, 257-258.
- [2] Hu, Z., Chen, J., Tian, T., Zhou, X., Gu, H., Xu, L., Chen, Y, *et al.* (2008) Genetic Variants of miRNA Sequences and Non-Small Cell Lung Cancer Survival. *The Journal of Clinical Investigation*, **118**, 2600-2608.
- [3] Cancer Genomics Research (2016) National Cancer Institute. <https://www.cancer.gov/research/areas/genomics>
- [4] Detterbeck, F.C., Boffa, D.J., Tanoue, L.T. and Wilson, L.D. (2010) Details and Difficulties regarding the New Lung Cancer Staging System. *Chest*, **137**, 1172-1180. <https://doi.org/10.1378/chest.09-2626>
- [5] The American Cancer Society Medical and Editorial Content Team (2017) Key Statistics for Lung Cancer. American Cancer Society.

- <https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/key-statistics.html>
- [6] Hoffman, F.L. (1931) Cancer and Smoking Habits. *Annals of Surgery*, **93**, 50. <https://doi.org/10.1097/0000658-193101000-00009>
- [7] Hopkin, M. (n.d.) Genetic Link for Lung Cancer Identified. *Nature*. <https://doi.org/10.1038/news.2008.733>
- [8] Weisberg, S. (2005) Applied Linear Regression. Vol. 528, John Wiley & Sons. <https://doi.org/10.1002/0471704091>
- [9] Hosmer Jr., D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. Vol. 398, John Wiley & Sons. <https://doi.org/10.1002/9781118548387>
- [10] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Mesirov, J.P., *et al.* (2005) Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles. *Proceedings of the National Academy of Sciences*, **102**, 15545-15550. <https://doi.org/10.1073/pnas.0506580102>
- [11] Hastie, T. and Tibshirani, R. (1990) Generalized Additive Models. John Wiley & Sons, Inc.
- [12] Armstrong, R.A. (2014) When to Use the Bonferroni Correction. *Ophthalmic and Physiological Optics*, **34**, 502-508. <https://doi.org/10.1111/opo.12131>
- [13] Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.
- [14] National Cancer Institute Dictionary of Cancer Terms (2015) National Cancer Institute. <https://www.cancer.gov/publications/dictionaries>
- [15] Katoh, M. (2005) WNT2B: Comparative Integromics and Clinical Applications. *International Journal of Molecular Medicine*, **16**, 1103-1108. <https://doi.org/10.3892/ijmm.16.6.1103>
- [16] Katoh, M. (2001) Differential Regulation of WNT2 and WNT2B Expression in Human Cancer. *International Journal of Molecular Medicine*, **8**, 657-660.
- [17] Li, S.J., Yang, X.N. and Qian, H.Y. (2015) Antitumor Effects of WNT2B Silencing in GLUT1 Overexpressing Cisplatin Resistant Head and Neck Squamous Cell Carcinoma. *American Journal of Cancer Research*, **5**, 300-308.
- [18] Halford, S., Bellingham, J., Ocaka, L., Fox, M., Johnson, S., Foster, R.G. and Hunt, D.M. (2001) Assignment1 of Panopsin (OPN3) to Human Chromosome Band 1q43 by *in Situ* Hybridization and Somatic Cell Hybrids. *Cytogenetic and Genome Research*, **95**, 234-235. <https://doi.org/10.1159/000059351>
- [19] Stewart, D.J. (2014) Wnt Signaling Pathway in Non-Small Cell Lung Cancer. *Journal of the National Cancer Institute*, **106**, djt356. <https://doi.org/10.1093/jnci/djt356>
- [20] Jiao, J., Hong, S., Zhang, J., Ma, L., Sun, Y., Zhang, D., Zhu, C., *et al.* (2012) Opsin3 Sensitizes Hepatocellular Carcinoma Cells to 5-fluorouracil Treatment by Regulating the Apoptotic Pathway. *Cancer Letters*, **320**, 96-103. <https://doi.org/10.1016/j.canlet.2012.01.035>
- [21] Al-Obaide, M.A.I., Abdel-Salam, A.S.G., Alobydi, H. and Srivenugopal, K.S. (2016) Bioinformatic Analysis of Human OPN3 Alternative Promoters Associated with Cancer. *Applied Mathematics*, **10**, 1223-1229. <https://doi.org/10.18576/amis/100402>
- [22] Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Lührmann, R. (1999) A Doughnut-Shaped Heteromer of Human Sm-Like Proteins Binds to the 3'-end of U6 snRNA, Thereby Facilitating U4/U6 Duplex Formation *in Vitro*. *The EMBO Journal*, **18**, 5789-5802. <https://doi.org/10.1093/emboj/18.20.5789>

- [23] Streicher, K.L., Yang, Z.Q., Draghici, S. and Ethier, S.P. (2007) Transforming Function of the LSM1 Oncogene in Human Breast Cancers with the 8p11-12 Amplicon. *Oncogene*, **26**, 2104-2114. <https://doi.org/10.1038/sj.onc.1210002>
- [24] Watson, P.M., Miller, S.W., Fraig, M., Cole, D.J., Watson, D.K. and Boylan, A.M. (2008) CaSm (LSm-1) Overexpression in Lung Cancer and Mesothelioma Is Required for Transformed Phenotypes. *American Journal of Respiratory Cell and Molecular Biology*, **38**, 671-678. <https://doi.org/10.1165/rcmb.2007-0205OC>
- [25] Henske, E.P., Short, M.P., Jozwiak, S., Bovey, C.M., Ramlakhan, S., Haines, J.L. and Kwiatkowski, D.J. (1995) Identification of VAV2 on 9q34 and Its Exclusion as the Tuberous Sclerosis Gene TSC1. *Annals of Human Genetics*, **59**, 25-37. <https://doi.org/10.1111/j.1469-1809.1995.tb01603.x>
- [26] Havel, L.S., Kline, E.R., Salgueiro, A.M. and Marcus, A.I. (2015) Vimentin Regulates Lung Cancer Cell Adhesion through a VAV2-Rac1 Pathway to Control Focal Adhesion Kinase Activity. *Oncogene*, **34**, 1979-1990. <https://doi.org/10.1038/onc.2014.123>
- [27] Hörlein, A.J., Näär, A.M., Heinzl, T., Torchia, J., Gloss, B., Kurokawa, R., Rosenfeld, M.G., *et al.* (1995) Ligand-Independent Repression by the Thyroid Hormone Receptor Mediated by a Nuclear Receptor Co-Repressor. *Nature*, **377**, 397. <https://doi.org/10.1038/377397a0>
- [28] Chen, J.D. and Evans, R.M. (1995) A Transcriptional Co-Repressor That Interacts with Nuclear Hormone Receptors. *Nature*, **377**, 454-457. <https://doi.org/10.1038/377454a0>
- [29] Battaglia, S., Maguire, O. and Campbell, M.J. (2010) Transcription Factor Co-Repressors in Cancer Biology: Roles and Targeting. *International Journal of Cancer*, **126**, 2511-2519. <https://doi.org/10.1002/ijc.25181>
- [30] Hla, T., Jackson, A.Q., Appleby, S.B. and Maciag, T. (1995) Characterization of edg-2, a Human Homologue of the Xenopus Maternal Transcript G10 from Endothelial Cells. *Biochimica et Biophysica Acta (BBA)—Gene Structure and Expression*, **1260**, 227-229. [https://doi.org/10.1016/0167-4781\(94\)00219-S](https://doi.org/10.1016/0167-4781(94)00219-S)
- [31] Hsu, T.Y.T., Simon, L.M., Neill, N.J., Marcotte, R., Sayad, A., Bland, C.S., Karlin, K.L., *et al.* (2015) The Spliceosome Is a Therapeutic Vulnerability in MYC-Driven Cancer. *Nature*, **525**, 384-388. <https://doi.org/10.1038/nature14985>
- [32] Shivakumar, M., Lee, Y., Bang, L., Garg, T., Sohn, K.A. and Kim, D. (2017) Identification of Epigenetic Interactions between miRNA and DNA Methylation Associated with Gene Expression as Potential Prognostic Markers in Bladder Cancer. *BMC Medical Genomics*, **10**, 30. <https://doi.org/10.1186/s12920-017-0269-y>
- [33] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Research*, **45**, D353-D361. <https://doi.org/10.1093/nar/gkw1092>
- [34] Nebert, D.W. (1991) Role of Genetics and Drug Metabolism in Human Cancer Risk. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **247**, 267-281. [https://doi.org/10.1016/0027-5107\(91\)90022-G](https://doi.org/10.1016/0027-5107(91)90022-G)
- [35] Murray, G.I., Taylor, M.C., McFadyen, M.C., McKay, J.A., Greenlee, W.F., Burke, M.D. and Melvin, W.T. (1997) Tumor-Specific Expression of Cytochrome P450 CYP1B1. *Cancer Research*, **57**, 3026-3031.
- [36] Currie, E., Schulze, A., Zechner, R., Walther, T.C. and Farese, R.V. (2013) Cellular Fatty Acid Metabolism and Cancer. *Cell Metabolism*, **18**, 153-161. <http://cancergenome.nih.gov/>  
<https://doi.org/10.1016/j.cmet.2013.05.017>