The Study of Whether the Single Gene Mutation Affects Cancer Invasiveness and Metastasis by Clinical Data Analysis

Name: Tianxiao Tao

Department: Computational Systems Biology Laboratory

Course: BCMB 4960L

Instructor: Ying Xu

Data: 7/25/2014

Summary

This lab research is designed to explore the main reason to cause the cancer invasiveness and metastasis in the cancer patients' body. We predicted the single gene mutation had a close connection to cancer invasiveness, which meant gene mutation may cause the tumor cancerization and boost the cancer development in the patients.

To figure out the result, the enrichment analysis is run to test for whether any single gene had significant high mutation rate in the any caner type. We used the TCGA somatic mutation data and clinical data of 15 cancer types from GEO database, and then we extracted the somatic mutation data from different pathological state. Next, we analyzed the data in a computational statistical analysis to find out the p value of all the single genes for each cancer type to determine if any single genes are highly relate to the cancer invasiveness.

The result there was no single genes had a significant association to any of two pathological stages when they compared to each other,

furthermore, the result illustrated the single gene mutation had no remarkable effect on the cancer invasiveness and metastasis.

Introduction

Base on the clinical statistic, one out of every two men and one out of every three women will be diagnosed with cancer, and Cancer kills about one American every minute of every day, or about 1,500 people every 24 hours. [1] But despite those huge numbers most individuals do not know what that really means. At the simplest level, cancer or cancer cells are cells that have lost the ability to follow the normal control that the body exerts on all cells. In our body we have billions and billions of cells and they have different functions. It is a very complicated process under incredibly phenomenal control and if something goes wrong and that control is lost and particular cells escape the normal control mechanisms and they continue to grow and they continue to grow and may spread. That is what we call cancer. Those cells together, we would call that a tumor. Specifically cancer is a malignant tumor because not only can it invade into adjacent organs, but unfortunately cancer can spread to other tissues and that can be life threatening.

The main reason to cause cancer is still a puzzle in the medical field, however, there are many predictions are proposed by medical scientists. Gene mutation has been considered to be the primary cause of cancer. [2] Mutations happen often, and the human body is normally able to correct most of these changes. Depending on where in the gene the change occurs, a mutation may be beneficial, harmful, or make no difference at all. Therefore, the likelihood of one mutation leading to cancer is small. Usually, it takes multiple mutations over a lifetime to cause cancer. This is why cancer occurs more often in older people, for whom there have been more opportunities for mutations to build up. Additional DNA damages can arise from exposure to exogenous agents. As one example of an exogenous carcinogeneic agent, tobacco smoke causes increased DNA damage, and these DNA damages likely cause the increase of lung cancer due to smoking.[3] In other examples, UV light from solar radiation causes DNA damage that is important in melanoma,[4] helicobacter pylori infection produces high levels of reactive oxygen species that damage DNA and contributes to gastric cancer,[5] and the Aspergillus metabolite, aflatoxin, is a DNA damaging agent that is causative in liver cancer.[6] DNA damages can also be caused by endogenous (naturally occurring) agents. Katsurano et al. indicated that macrophages and neutrophils in an inflamed colonic epithelium are the source of reactive oxygen species causing the DNA damages that initiate colonic tumorigenesis, [7] and bile acids, at high levels in the colons of humans eating a high fat diet, also cause DNA damage and contribute to colon cancer. [8] Therefore, our hypothesis is some single gene mutations associate to cancer invasiveness and metastasis.

Experimental Methods

To test our hypothesis, we applied TCGA somatic mutation data and clinical data of fifteen cancer types from Gene Expression Omnibus (GEO) database, which included BRCA, CESC, COAD, GBM, KICH, KIRC, KIRP, LUAD, LUSC, OV, PAAD, PRAD, READ, STAD, and THCA.

Next, we divided the clinical data into M (M0 and M1) and N (N0, N1, N2, and N3) stages that related to cancer invasiveness and metastasis, because a cancer tissue in one patient at different progression stages might behave totally different characteristics, and then we extracted the somatic mutation information of samples of different pathological stages.

To avoid some data might lead to poor performance to the result, we constructed a threshold, which was a value that could distinguish each gene with mutations value to be satisfied in a right range. The threshold in this research was determined by using the formula: (#mutations/#all samples)*min {#M0 samples, #M1 samples}>0.8).

To clearly display the multivariate frequency distribution of the amount of mutation and non-mutation gene set, we constructed a contingency table as:

	Gene set	Others	Total
Mutation	S	n-s	n
Non-mutation	length(go_genes)-s	N-length(go_genes)-(n-s)	N-n
Total	length(go_genes)	N-length(go_genes)	N

Table 1: gene set number in variant mutation level contingency table.

Furthermore, fisher exact test which is a statistical significance test used in the analysis of contingency tables, [9] that was applied to test the contingency table of each gene, and then, get a p value for whether the samples of the two different stages have significant different mutation rate.

The distribution of the p values was checked and adjusted by Bonferroni method which is a method used to counteract the problem of multiple comparisons [10] to control false discover rate.

All of above procedures was worked in the R statistical computing environment which is a software environment for statistical computing and graphics. Finally, the p values for each cancer type were plotted as bar graphs.

Results

Figure 1 showed the p value of BRCA N0, N1, N2, and N3 stages when they associated to each other. The results showed the p values for the most genes were 1.0, and only very few gene types had p values rather than 1.0 with a very low frequency in the graph "BRCA_N2_V_N3", "BRCA_N0_V_N2", and "BRCA_N1_V_N2".

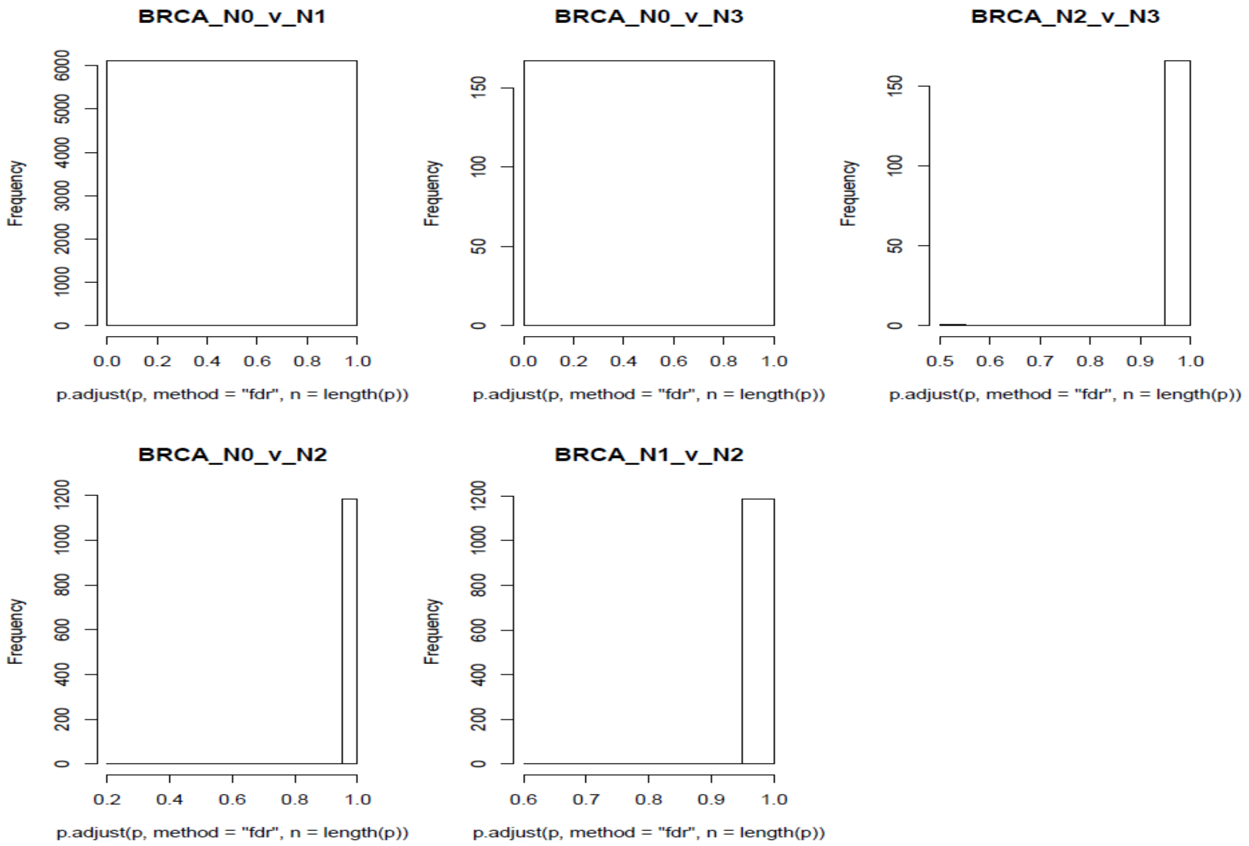

Figure 1. P value vs. gene frequency of BRCA N stages.

Figure 2 showed the p value of CESC N0 and N1 stages when they associated to each other. The result showed the p values for all of the genes were 1.0 with no exception.

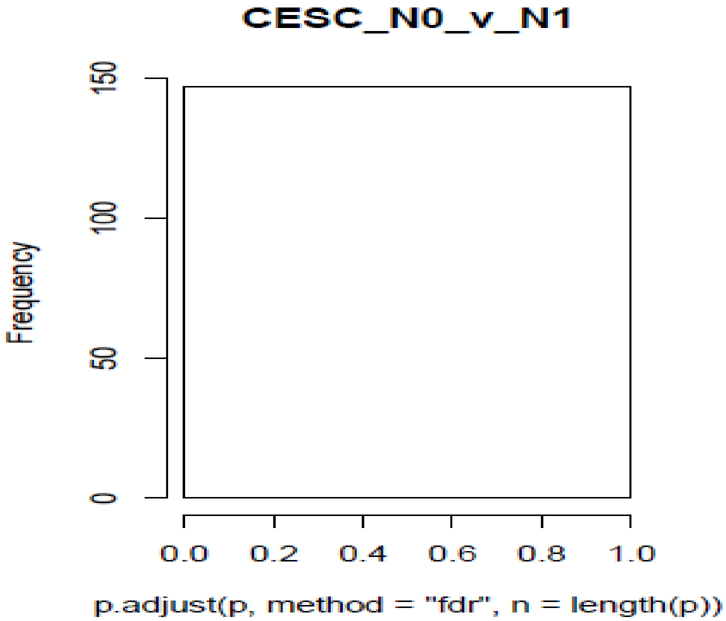

Figure 2. P value vs. gene frequency of CESC N stages.

Figure 3 showed the p value of COAD N0, N1 and N2 stages when they associated to each other. The results showed the p values for the most genes were 1.0 in each graph, and there were around 4000 genes were 0.85, and only very few gene types had p values rather than 1.0 and 0.85 with a very low frequency in the graph "COAD_N0_V_N1".

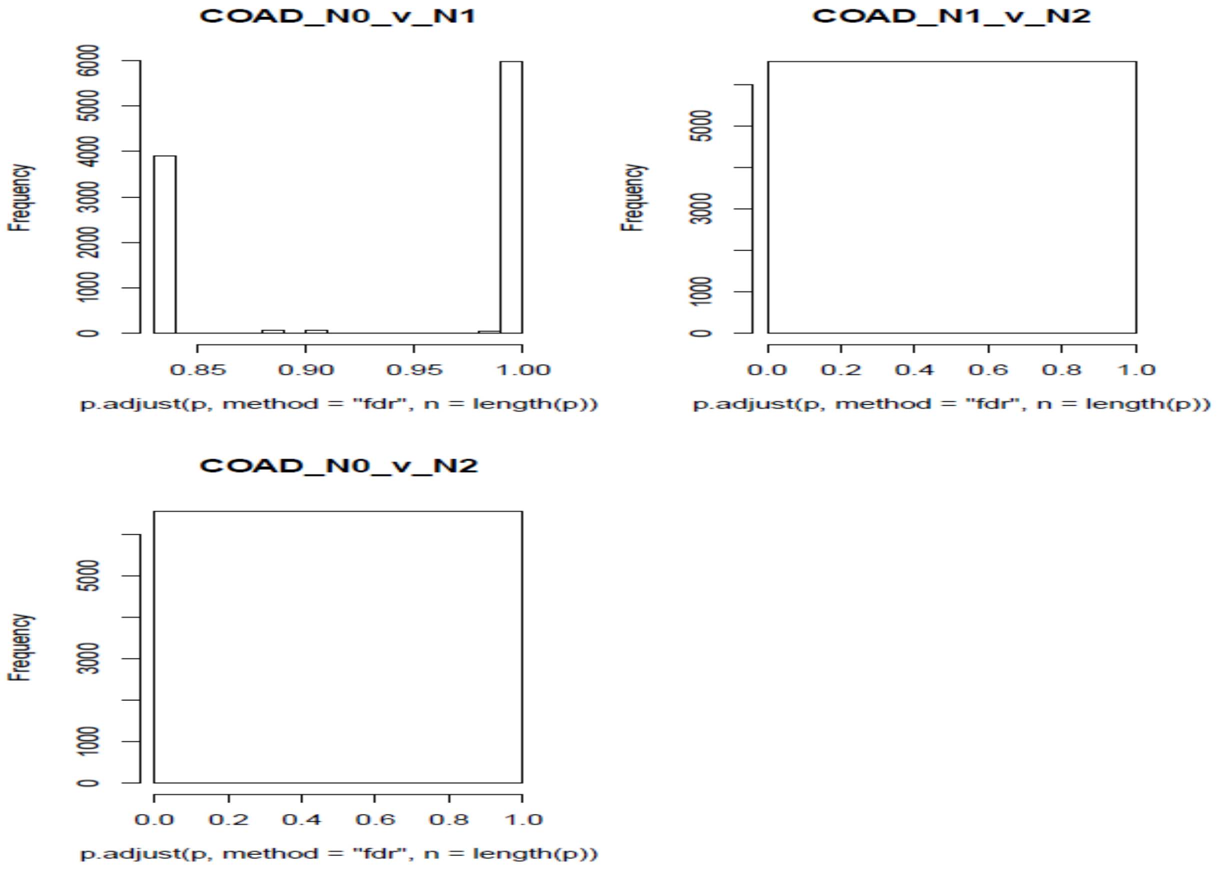

Figure 3. P value vs. gene frequency of COAD N stages.

Figure 4 showed the p value of LUAD N0, N1 and N2 stages when they associated to each other. The result showed the p values for all of the genes were 1.0 with no exception.

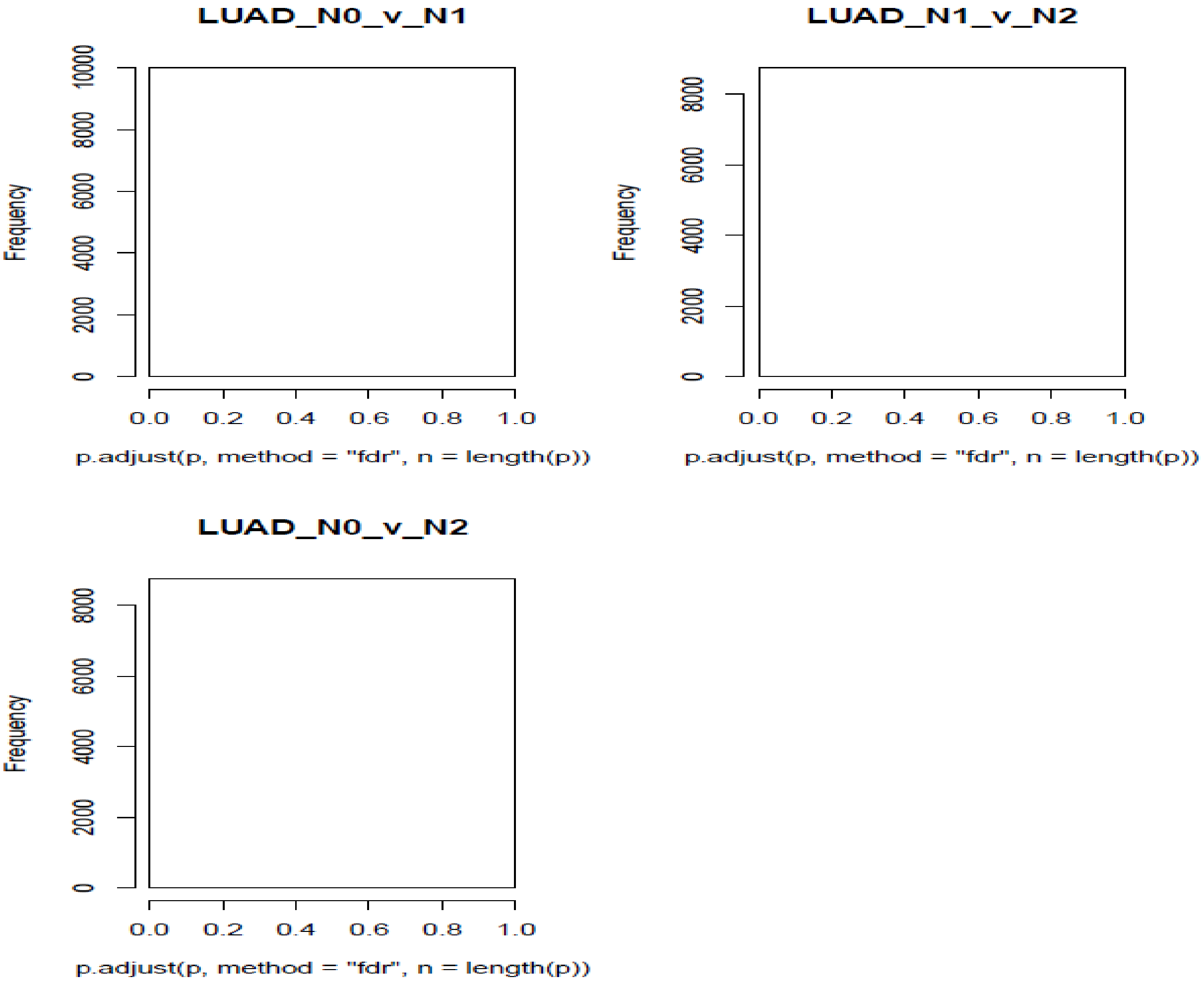

Figure 4. P value vs. gene frequency of LUAD N stages.

Figure 5 showed the p value of LUSC N0, N1 and N2 stages when they associated to each other. The results showed the p values for the most genes were 1.0 in each graph, and only very few gene types had p values rather than 1.0 with a very low frequency in the graph “LUSC_N0_V_N2”.

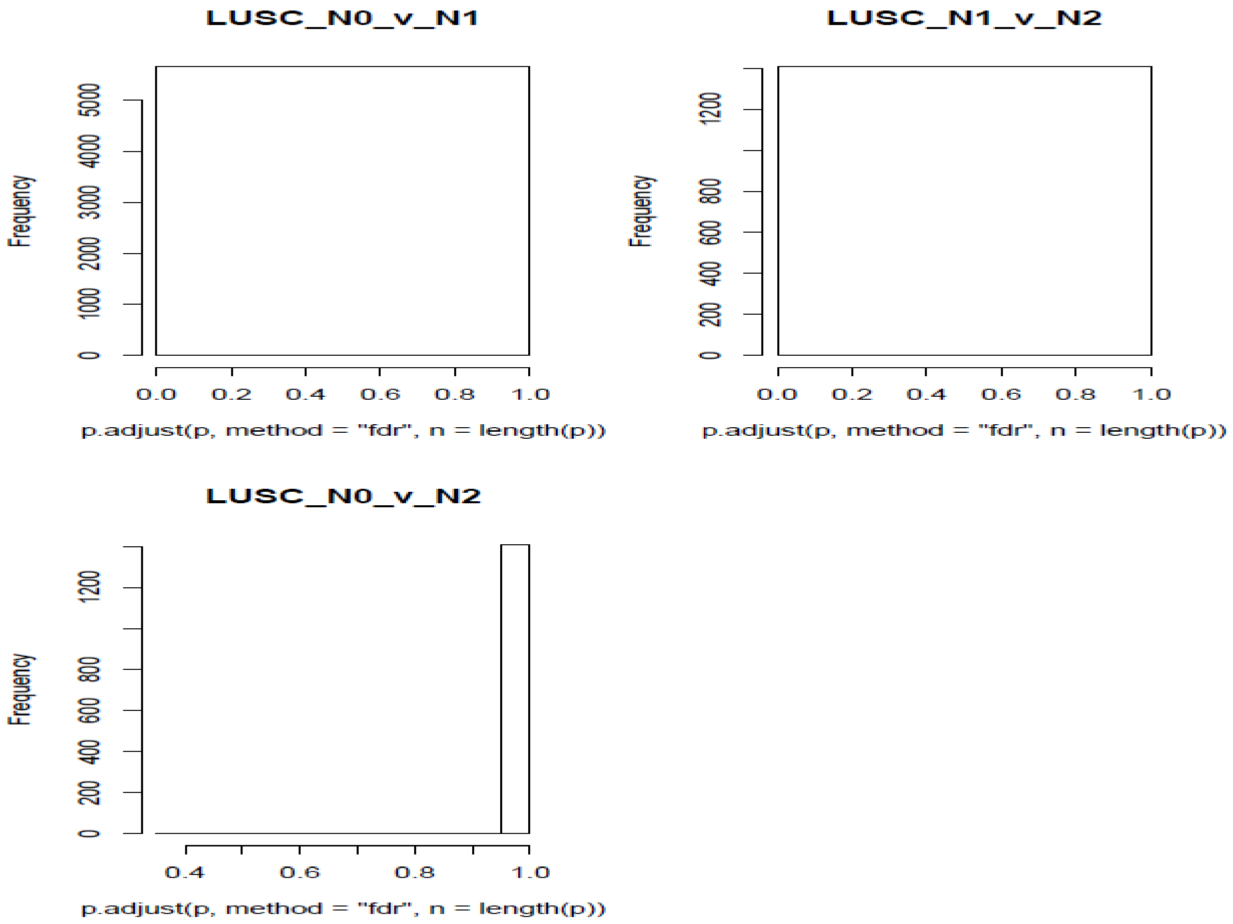

Figure 5. P value vs. gene frequency of LUSC N stages.

Figure 6 showed the p value of PAAD N0 and N1 stages when they associated to each other. The result showed the p values for all of the genes were 1.0 with no exception.

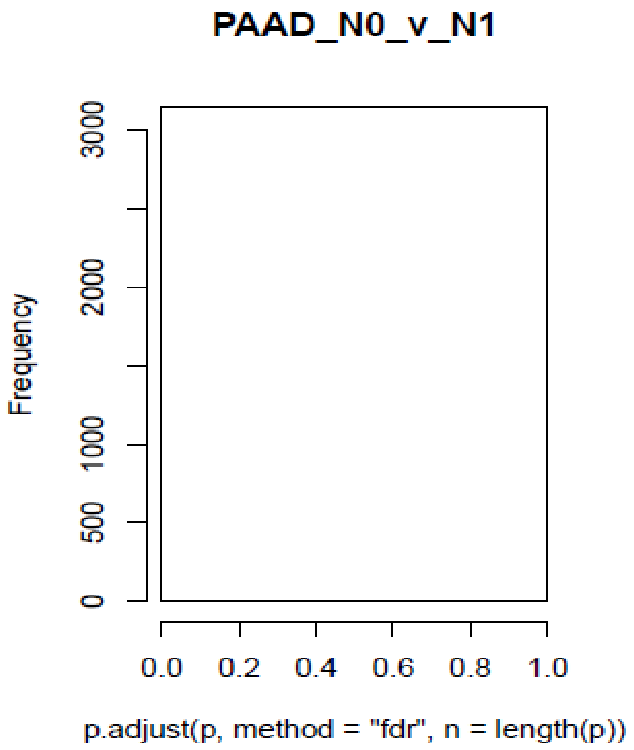

Figure 6. P value vs. gene frequency of PAAD N stages.

Figure 7 showed the p value of PAAD N0 and N1 stages when they associated to each other. The result showed the p value of over 20 genes were 0.85, and less than 5 genes were 0.90, and between 5 to 10 genes were 1.00.

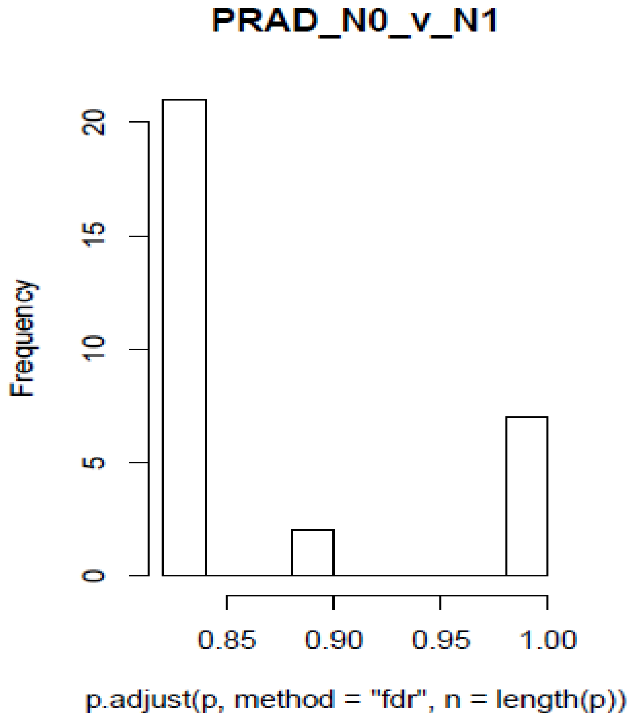

Figure 7. P value vs. gene frequency of PRAD N stages.

Figure 8 showed the p value of READ N0, N1 and N2 stages when they associated to each other. The results showed the p values for the most genes were 1.0 in each graph, and there were 300 genes were 0.95 in the graph “READ_N0_V_N2”.

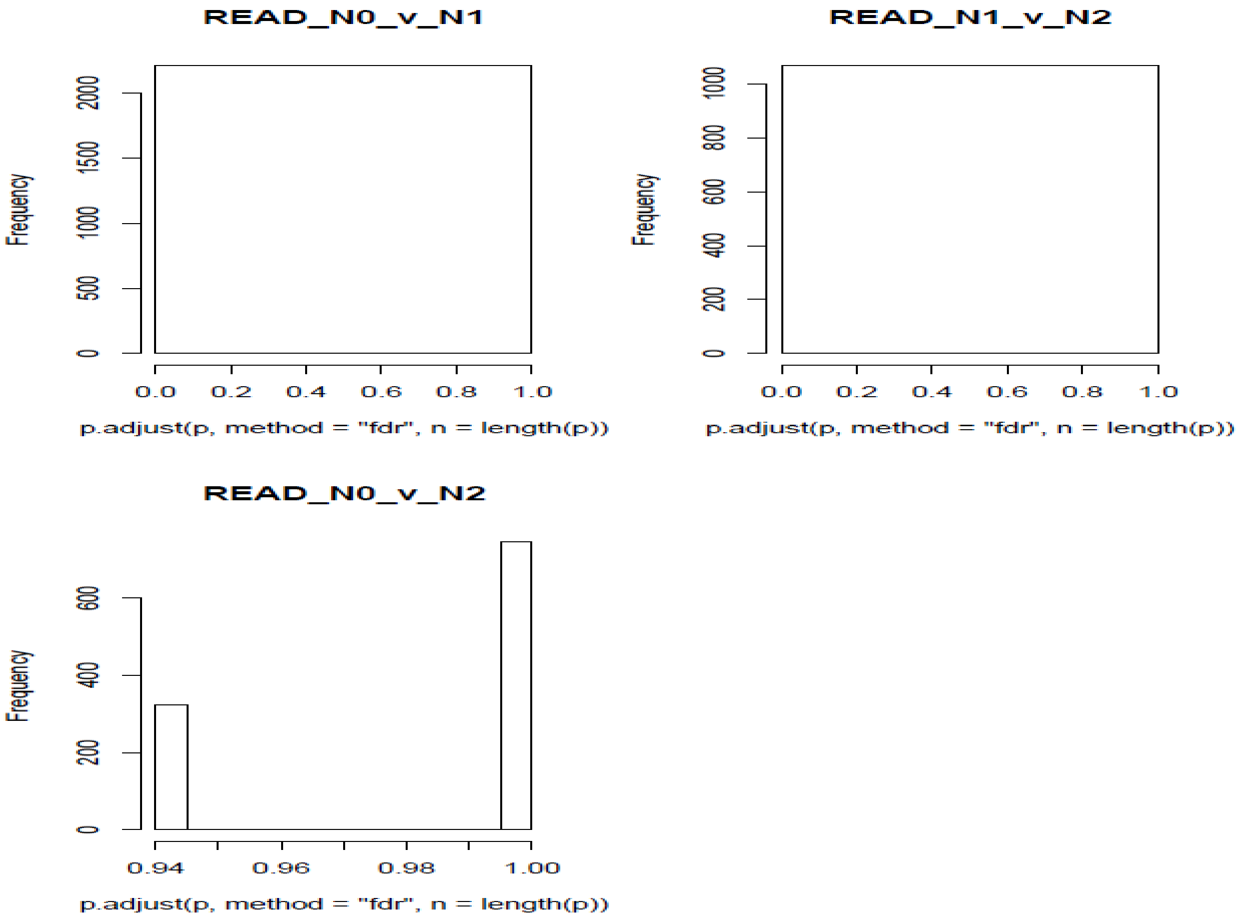

Figure 8. P value vs. gene frequency of READ N stages.

Figure 9 showed the p value of STAD N0, N1, N2 and N3 stages when they associated to each other. The result showed the p values for all of the genes were 1.0 with no exception.

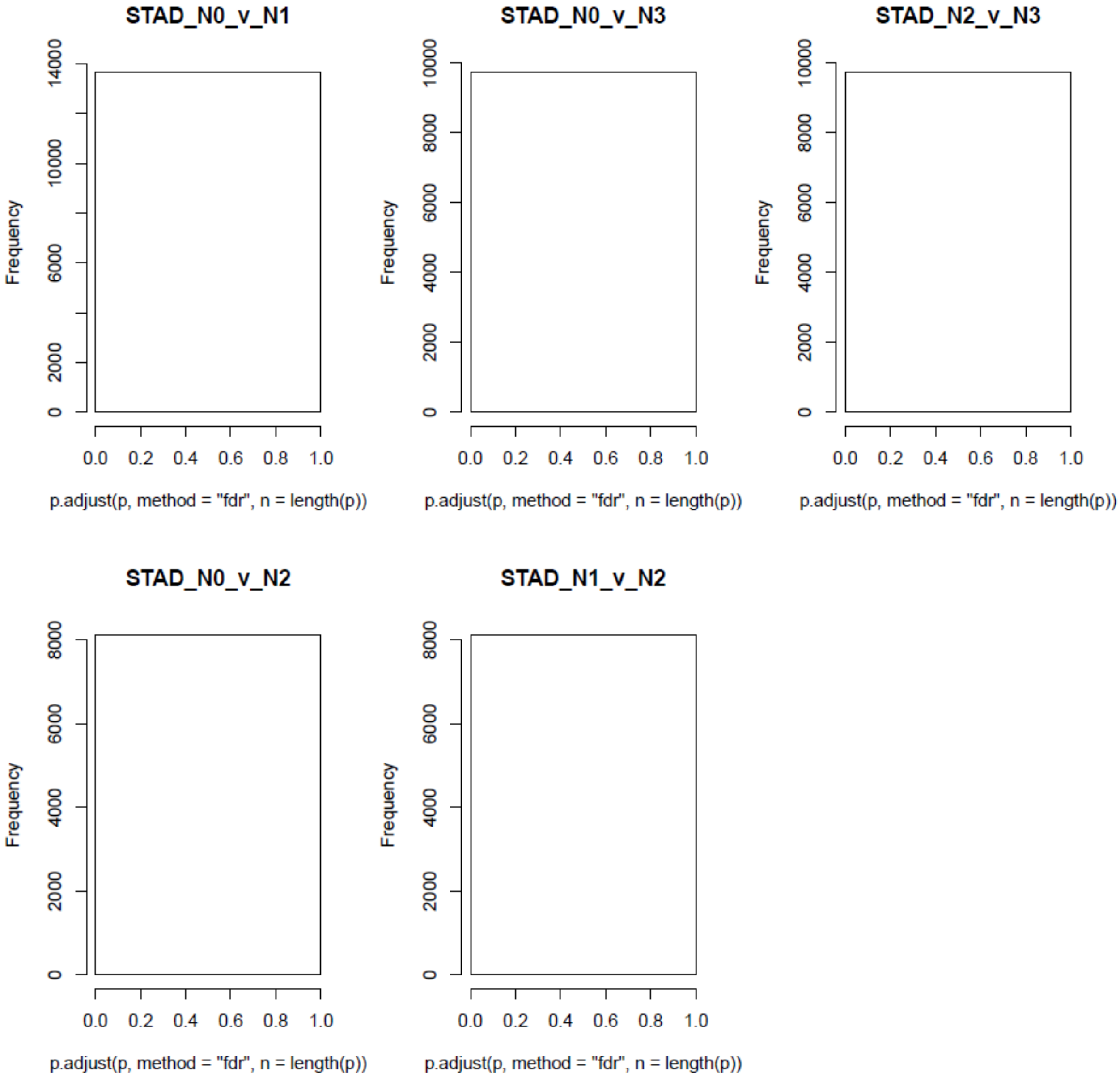

Figure 9. P value vs. gene frequency of STAD N stages.

Figure 10 showed the p values of BRCA, READ, KIRC, COAD, STAD, and LUAD M stage. The results showed the p values for the most genes were 1.0 in each graph, and only very few gene types had p values rather than 1.0 with a very low frequency in the READ M stage.

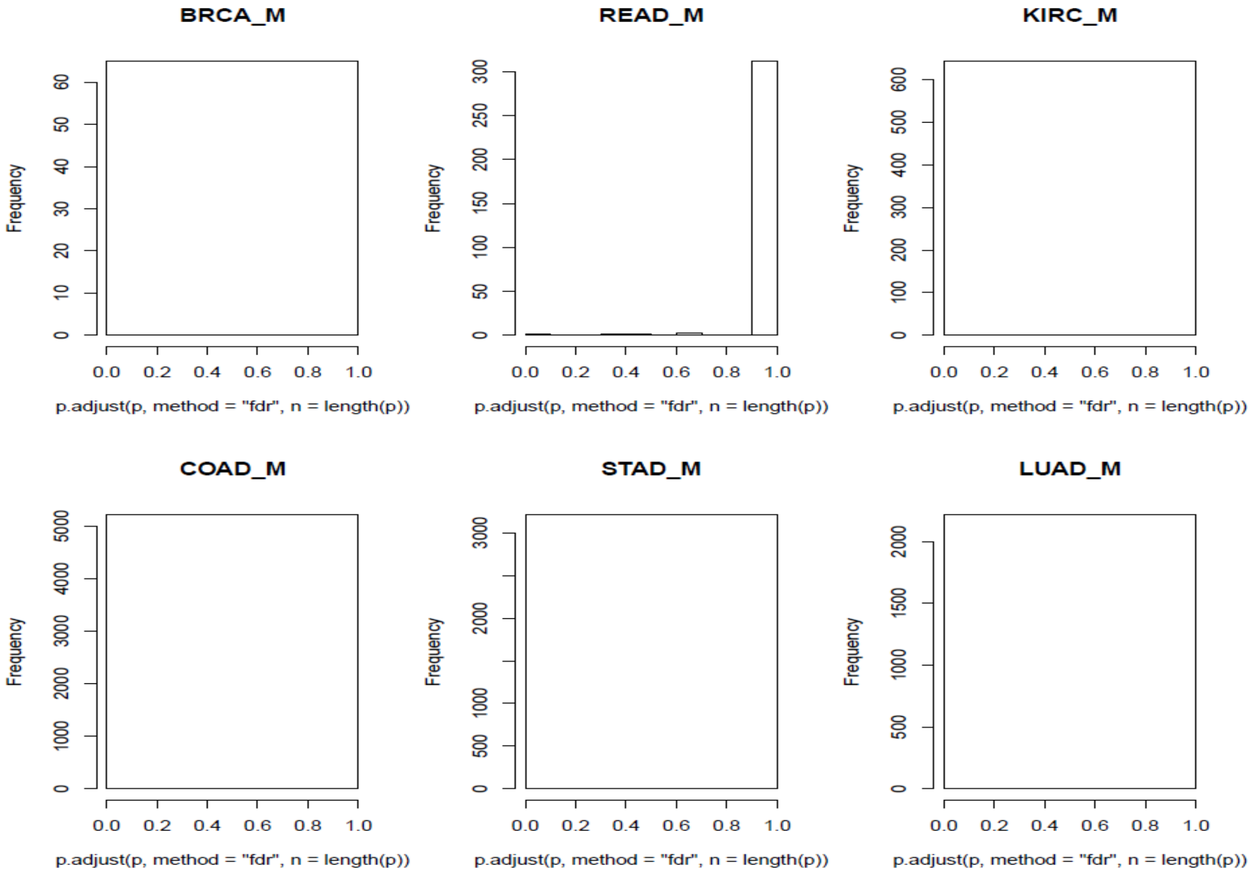

Figure 10. P value vs. gene frequency of BRCA, READ, KIRC, COAD, STAD, and LUAD M stage.

Discussion

Based on the results which were showed in each figure, the p value for most genes were 1.0, and only very few genes had p values rather than 1.0 with a low frequency in all of the cancer types that we discussed in the research. In addition, there was no genes showed a significant low p value (less than 0.05), that meant there was no single genes had a significant association to any of two pathological stages when they compared to each other, furthermore, the result illustrated the single gene mutation had no remarkable effect on the cancer invasiveness and metastasis. Therefore, the results rejected our presumption that some single gene mutations associate to cancer invasiveness and metastasis.

The p values that rather than 1.0 might be caused by some experimental error, such as misreading, the sample size was not big enough, or due to patients' special body conditions. However, those particular genes should not make any differences to the final result.

Despite this research showed there was no single gene mutation that was related to cancer invasiveness and metastasis, gene mutation still might be a part of reason to cause cancer. Due to genes could depress or activate each other, the group gene mutations might have an effect to cancer invasiveness and metastasis, and this project could be our future research interests.

In the future, doctors hope to learn more about the role of genetic changes in the development of cancer, which may lead to improvements in finding and treating cancer, as well as predicting a person's risk of cancer. The main reason to cause cancer may be founded in the future, all in all, the main purpose of our research is to let more people know about cancer and have a healthy life habit to live away from cancer.

Reference

1. Sasieni PD, Shelton J, Ormiston-Smith N, et al. What is the lifetime risk of developing cancer?: The effect of adjusting for multiple primaries. Br J Cancer, 2011. 105(3): p. 460-5.
2. Bernstein C, Prasad AR, Nfonsam V, Bernstein H. (2013). DNA Damage, DNA Repair and Cancer, New Research Directions in DNA Repair, Prof. Clark Chen (Ed.), ISBN 978-953-51-1114-6, InTech, http://www.intechopen.com/books/new-research-directions-in-dna-repair/dna-damage-dna-repair-and-cancer
3. Cunningham, F.H.; Fiebelkorn, S.; Johnson, M.; Meredith, C. (2011). "A novel application of the Margin of Exposure approach: Segregation of tobacco smoke toxicants". *Food and Chemical Toxicology* 49 (11): 2921–2933
4. Kanavy, Holly E.; Gerstenblith, Meg R. (2011). "Ultraviolet Radiation and Melanoma". *Seminars in Cutaneous Medicine and Surgery* 30 (4): 222–228.

5. Handa, Osamu; Naito, Yuji; Yoshikawa, Toshikazu (2011). "Redox biology and gastric carcinogenesis: The role of Helicobacter pylori". *Redox Report* 16 (1): 1–7.
6. Smela, ME; Hamm, ML; Henderson, PT; Harris, CM; Harris, TM; Essigmann, JM (2002). "The aflatoxin B(1) formamidopyrimidine adduct plays a major role in causing the types of mutations observed in human hepatocellular carcinoma". *Proceedings of the National Academy of Sciences of the United States of America* 99 (10): 6655–60.
7. Katsurano, M; Niwa, T; Yasui, Y; Shigematsu, Y; Yamashita, S; Takeshima, H; Lee, M S; Kim, Y-J; Tanaka, T; Ushijima, T (2011). "Early-stage formation of an epigenetic field defect in a mouse colitis model, and non-essential roles of T- and B-cells in DNA methylation induction". *Oncogene* 31 (3): 342–351.
8. Bernstein, Carol; Holubec, Hana; Bhattacharyya, Achyut K.; Nguyen, Huy; Payne, Claire M.; Zaitlin, Beryl; Bernstein, Harris (2011). "Carcinogenicity of deoxycholate, a secondary bile acid". *Archives of Toxicology* 85 (8): 863–71.

9. Fisher, R. A. (1922). "On the interpretation of χ2 from contingency tables, and the calculation of P". *Journal of the Royal Statistical Society* 85 (1): 87–94.

10. Holm, S. (1979). "A simple sequentially rejective multiple test procedure". *Scandinavian Journal of Statistics* 6 (2): 65–70.