IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

Discovery of Novel Genetic Alteration Using Meta-analysis of Colorectal Cancer

Haryeong Eo

Shanghai American School Puxi, 258 Jinfeng Road, Minhang District, Shanghai, China 201107; hre1308@gmail.com

ABSTRACT: More than 5.25 million people worldwide are diagnosed with colorectal cancer (CRC), representing 10% of the global cancer incidence and 9.4% of all cancer-caused deaths. It is common to find genetic and epigenetic alterations in CRC, which are the driving force of tumorigenesis. Therefore, discovering a novel genetic alteration in colorectal cancer can support early diagnosis and finding novel targets for cancer treatment. However, genetic alterations found in colorectal cancer are not fully elucidated. A meta-analysis of colorectal cancer genomics data sets was performed using 12 different studies provided by cBioPortal to identify the novel genetic alteration in colorectal cancer patients. Through cBioPortal analysis, it was hypothesized that chromosome 17q21 amplification is associated with tumor progression. Patient survival analysis was performed through cBioPortal analysis. Also, nine genes located in 17q21 were further analyzed with GeneMania, a web-based program that predicts the function of gene sets. Using cBioPortal analysis, it was discovered that chromosome 17q21 amplification was enriched in deceased patients. Furthermore, through patient survival analysis, amplification of each of the nine genes located in chromosome 17q21 was significantly associated with decreased patient survival rate. Hence, using GeneMania analysis, it was discovered that the gene network of the nine genes was significantly associated with DNA integrity checkpoint function. Through this study, chromosome 17q21 amplification, which may alter the part of DNA integrity in cancer cells, can be used as a biomarker that predicts poor patient survival.

KEYWORDS: Human genetics; Colorectal cancer; Patient survival rate; Chromosome 17; DNA integrity checkpoint.

Introduction

Colorectal cancer (CRC) mostly begins as a polyp, a non-cancerous growth that develops in the colon's inner lining.[1] Polyps are classified as either adenomatous or serrated.[2] Similar to adenomas, serrated polyps, traditional serrated adenomas, and large hyperplastic polyps are associated with an increased risk for CRC.[3] Because sessile serrated polyps (SSPs) are difficult to detect during colonoscopy as they are usually flat, these features make them the precursors for a large proportion of cancers.[4]

Gene mutation found in colorectal cancer affects overall patient survival.[5] For example, approximately half of all colorectal cancers show *TP53*, otherwise known as P53, gene mutations, with higher frequencies observed in the distal colon and rectal tumors.[6] The role of the *TP53* gene is to regulate the cell cycle and apoptosis. Specifically, the P53 protein induces G1 cell-cycle arrest and controls the repairing of the DNA before the cell goes into DNA replication.[7] If the DNA repair is unsuccessful, P53 causes cell death. *TP53* mutation occurs at the time of transition from adenoma to cancer. Several studies attempted to explain the significance of *TP53* mutation in colorectal cancer, with conflicting results. A study concluded that the survival rate for P53 positive patients was far greater than that for P53 negative patients. However, overexpression of P53 in stage three CRC carried a better overall survival in CRC patients.[8]

Gene deletion in colorectal cancer also affects colorectal patients' survival. In chromosome 18, loss of heterozygosity (LOH) in the region of 18q21 is often seen in advanced colorectal cancer. LOH is defined as the loss of one allele at a specific locus.[10] Often, the remaining allele is affected by a deletion mutation or a loss of chromosome from a chromosome pair. Some studies found an inverse relationship between CRC patient survival and 18q LOH. A previous study evaluated the effect of 18q LOH on 532 non-MSI-high, stage I-IV CRC tumors; in patients with non-MSI-high CRC, 18q LOH were not significantly associated with a difference in survival.[11]

The cBio Cancer Genomics Portal (cBioPortal) contains numerous multidimensional cancer genomics data sets.[12] The cBioPortal minimizes the tasks needed to collect data by summarizing complex genomic data from large-scale cancer genomics projects.[13] Through cBioPortal, it is possible to better understand biology and clinical applications.

GeneMania provides hypotheses about gene function by showing a list of genes and categorizing genes based on their functions.[14] After receiving comprehensive gene lists, GeneMania groups the genes based on their function, followed by genomics and proteomics data. GeneMania determines whether the functional genomic dataset follows its predictive value during this process. GeneMania also predicts gene function. Using a single gene, GeneMania searches for several genes with the same function based on its interactions with other genes.[14] Overall, GeneMania allows researchers to analyze genes more efficiently and intuitively.

DOI: 10.36838/v5i1.7

After analyzing colorectal cancer genomics data sets consisting of 12 different studies via cBioPortal, novel genetic alterations were found in colorectal cancer patients. Using the given data, it was hypothesized that 17q21 amplification is highly associated with tumor progression, affecting the patient survival rate. In addition, using GeneMania, nine genes located in 17q21 that play a role in the function of gene sets were further identified.

Methods

Analyzing genomic alteration on colorectal patients' genomic data using cBioPortal:

Through cBioPortal, 4341 colorectal patients (4488 samples) were analyzed to find a novel genetic alteration associated with patients' survival. After the patient samples were divided into two groups: living group (n= 1275) and deceased group (n= 273), amplified genes enriched in the deceased group were found. The gene location, percentage of alteration in each group, log-ratio between living and deceased group, p-value, and q-value was analyzed using cBioPortal. The log-rank statistical test was used to calculate the p-value.

Gene network analysis using the GeneMania program:

Gene network analysis was performed with eleven genes found in the cBioPortal (*GJD3, CCR7, TOP2A, CDC6, IGFBP4, WIPF2, KRT222, SMARCE1, RARA*). GeneMania predicts the function of listed genes and their genetic network, including co-expression, physical interaction, shared domains, and biological pathways. GeneMania analyzes the gene lists and prioritizes the genes for functional assays. It finds functionally similar genes within the genomics and proteomics data that have been previously published. Since hundreds of millions of interactions had been collected by the database from GEO, BioGRID, IRefIndex, and I2D, the interaction databases were used to predict the gene function.

Patient survival analysis using cBioPortal:

cBioPortal for cancer genomics provides visualization analysis of overall patient survival status. The overall survival of patient groups between the gene amplified group (*TOP2A and CDC6*) and the non-amplified group was analyzed. Kaplan-Mier analysis and log-rank test were performed to calculate the p-value. The median survival month in each group was also investigated.

Results and Discussion

Table 1: The amplified genes located on 17q21.2 enriched in the deceased colorectal cancer patient's group.

Gene	Cytoband	LIVING (n = 1275)	DECEASED (n = 273)	Log Ratio	p-Value	q-Value	Enriched in
GJD3	17q21.2	19 (1.49%)	15 (5.49%)	-1.88	2.57E-04	5.97E-03	DECEASED
CCR7	17q21.2	17 (1.33%)	12 (4.82%)	-1.85	1.06E-03	0.0149	DECEASED
TOP2A	17q21.2	19 (1.49%)	13 (4.76%)	-1.68	1.72E-03	0.0209	DECEASED
CDC6	17q21.2	19 (1.49%)	12 (4.82%)	-1.69	2.13E-03	0.0239	DECEASED
IGFBP4	17q21.2	19 (1.49%)	12 (4.82%)	-1.69	2.13E-03	0.0239	DECEASED
WIPF2	17q21.2	19 (1.49%)	12 (4.82%)	-1.69	2.13E-03	0.0239	DECEASED
KRT222	17q21.2	14 (1.10%)	10 (4.02%)	-1.87	2.65E-03	0.0289	DECEASED
SMARCE1	17q21.2	15 (1.18%)	10 (4.02%)	-1.77	3.79E-03	0.0358	DECEASED
RARA	17q21.2	28 (1.27%)	26 (2.73%)	-1.1	3.84E-03	0.0359	DECEASED

Gene amplification is when there is an increase in the copy number of DNA present in a specific region of the chromosome or an increase in the RNA and protein made from that gene. Cancer cells often produce multiple copies of genes, and some of the amplified genes can cause cancer cells to grow faster or become resistant to anticancer drugs. 4341 colorectal patients (4488 samples) were analyzed to find a novel genetic alteration associated with patients' survival through the cBioPortal database. First, the patient samples were divided into two groups: living group (n = 1275) and deceased group (n = 273). It was found that chromosome position 17q21.2 amplification is enriched in the decreased patient group (Table 1). In total, nine amplified genes are located in 17q21.2: *GJD3, CCR7, TOP2A, CDC6, IGFBP4, WIPF2, KRT222, SMARCE1*, and *RARA*. Overall, it was found that the novel amplification of nine genes in colorectal cancer patient samples enriched in decreased patient groups.

Previous studies showed that 17q21 amplification was detected in gastric cancer and breast cancers.[15,16] The comparative analysis of DNA copy number and microarray in gastric cancer shows that the 17q12-q21 region is amplified and many genes in this region are overexpressed. A breast cancer study indicated that HER2/NEU amplification (both positioned on 17q21) is responsible for the development of Trastuzumab, one of the first immunotherapeutic drugs for the successful treatment of breast cancers.[17] In conclusion, the amplification in region 17q21 not only caused colorectal cancer but also gastric cancer and breast cancer. This shows how 17q21 plays a critical role in cancer progression.

Table 2: The result of functional prediction of nine amplified genes with extended genes that are functionally similar using GeneMANIA.

Function	P-value	Coverage
DNA integrity checkpoint	3.33e-1	4/147
Negative regulation of epithelial cell proliferation	3.33e-1	3/54
Hormone receptor binding	3.33e-1	4/148
Mitotic cell cycle checkpoint	3.33e-1	4/136
DNA replication checkpoint	3.68e-1	2/10
Insulin-like growth factor binding	5.30e-1	2/13
Cell cycle checkpoint	5.75e-1	4/201
Regulation of chromosome organization	6.13e-1	4/239
Positive regulation of DNA-templated transcription, initiation	6.13e-1	2/23
DNA recombination	6.13e-1	4/229

GeneMANIA analysis was performed to find a novel function of nine amplified genes on colorectal cancer progression. It was found that DNA integrity checkpoint, negative regulation of epithelial cell proliferation, hormone receptor binding, and mitotic cell cycle checkpoint were determined to be the most significant functions of the nine amplified genes (p-value= 3.33e-1) (Table 2). Since the maintenance of genomic integrity is important in normal cell growth and development, gene alteration on DNA integrity checkpoints is found in many cancer cells. In addition, DNA integrity checkpoints provide cells with time to repair damaged DNA, but cancer-initiating cells have lost DNA repair or cell-cycle checkpoints. In conclusion, four genes among nine amplified genes were significantly associated with the function of the DNA integrity checkpoint meaning alteration in this func-

DOI: 10.36838/v5i1.7

function may be linked to the poor survival rate of colorectal patients.

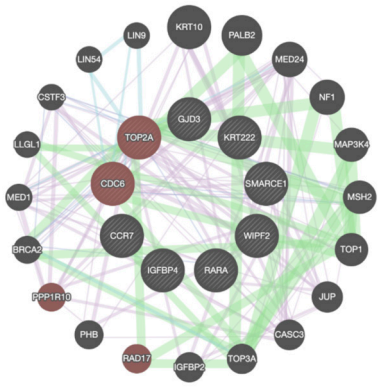

Figure 1: Gene network analysis of ten DEGs, located in the center as slash patterned. The gene sets associated with ten DEGs are located around the ten DEGs. The four red genes are associated with the DNA integrity checkpoint. The green line indicates the physical interaction between the genes. The purple line indicates the co-expressed genes. The blue line indicates the co-localized genes.

Cancer-associated necrosis produces more abnormal DNA fragments than apoptosis in human cells. Therefore, this study focused on the DNA integrity checkpoints because DNA integrity, which relates to the uneven copy number of DNA fragments, is highly related to cancer. According to the model, *TOP2A* and CDC6 are among the nine amplified genes, and *PPP1R10* and *RAD17* are functionally related to the nine amplified genes. In conclusion, amplification of *TOP2A* and *CDC6* and the alteration of relative genes (*PPP1R10* and *RAD17*) may affect DNA integrity checkpoints, which shortens the survival rate of colorectal cancer patients (Figure 1).

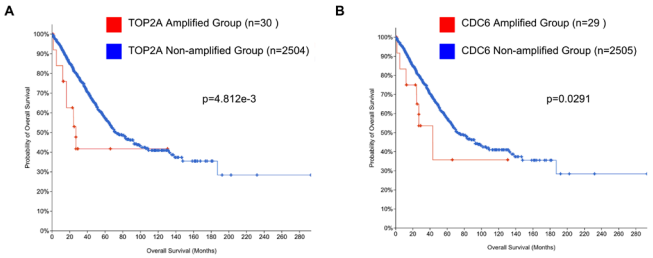

Figure 2: The analysis of overall survival of *TOP2A* and CDC6 amplified patient group compared to *TOP2A* and CDC6 non-amplified patient group. (A) *TOP2A* amplified group (n=30) showed a lower survival rate compared to *TOP2A* non-amplified group (n=2504) (p= 4.812e-3). (B) CDC6 amplified group (n=29) showed lower survival rate compared to CDC6 non-amplified group (n=2505). (p= 0.0291) Log-rank test was used to calculate the p-value.

To validate the effect of *TOP2A* and *CDC6* amplification on colorectal cancer patients' survival, the overall patient survival rate was further analyzed with the cBioPortal database. The patients were divided into two groups: the groups with amplified genes and the groups without amplified genes. A graph in Figure 2 shows the probability of overall survival (Y-axis) and overall survival months (X-axis). It was found that the median survival months of *TOP2A* amplified patients (27 months) was significantly lower than those of *TOP2A* non-amplified patients (71.93 months). Similarly, the median survival months of *CDC6* amplified patients (43.17 months) was significantly lower than the median survival months of *CDC6* non-amplified patients (72.47 months). In conclusion, it was speculated that since *TOP2A* and *CDC6* are associated with the function of DNA integrity checkpoint, the lower survival rate in amplified groups may have abnormal DNA structure with severe DNA damage.

Conclusions

To reiterate the findings shown in Figure 2, two genetic alterations *TOP2* and *CDC6* both significantly decreased the patient survival rate. Using the given information, when analyzing genomics for colorectal cancer patients in the future, it is possible to more accurately diagnose and predict the survival rate of the patients. Furthermore, it was found that 17q21 amplification affected DNA integrity checkpoints the most. With a better understanding of the impact of amplified genes, ways to recover the affected DNA integrity checkpoint can be found, possibly supporting the development of a novel treatment for colorectal cancer. Lastly, a considerably large sample size of 4448 patients was analyzed. Therefore, the results from this study provide a more accurate interpretation of the impact of 17q21 amplification, further reducing the margin of error when treating colorectal patients in general.

However, the study is limited in the methods used, as only data analysis was performed. The media used to perform meta-analysis has the potential for publication bias, skewed data, and difficulties in combining studies that may have differences in population, interventions, etc. Furthermore, the study is limited in scope, as only gene copy alterations, disregarding other genetic alterations, such as mutation and fusion genes, were focused on. Therefore, cancer cell experiments to not only validate the real implications of 17q21 amplification on cancer cell development can be performed but it is also possible to find potential treatments for colorectal cancer. In addition, cBioPortal can be used to further analyze other genetic variations which could have played a significant role in tumor progression.

Acknowledgments

I would like to acknowledge and thank my research mentor Dr. Lee, who provided constructive feedback and support. His guidance helped me overcome some of the difficulties I faced while researching. I would also like to give special thanks to my parents for encouraging me to take this opportunity to research more about cancer. Throughout this project, they warmly listened to every progress I made and asked questions that helped me to make a better research paper.

References

1. Bond, J. H. Colon Polyps and Cancer. Endoscopy 2003, 35 (1), 27–35.
2. Alecu, M.; Simion, L.; Straja, N.; Brătucu, E. Multiple Polyps and Colorectal Cancer. *Chirurgia (Bucur)* **2014**, 109 (3), 342–346.
3. Demetriades, H.; Kanellos, I.; Blouhos, K.; Tsachalis, T.; Vasiliadis, K.; Pramateftakis, M. G.; Betsis, D. Synchronous Polyps in Patients with Colorectal Cancer. *Tech. Coloproctol.* **2004**, 8 Suppl 1, s72-5.
4. Farris, A. B.; Misdraji, J.; Srivastava, A.; Muzikansky, A.; Deshpande, V.; Lauwers, G. Y.; Mino-Kenudson, M. Sessile Serrated Adenoma: Challenging Discrimination from Other Serrated Colonic Polyps. *Am. J. Surg. Pathol.* **2008**, 32 (1), 30–35.
5. Cavagnari, M. A. V.; Silva, T. D.; Pereira, M. A. H.; Sauer, L. J.; Shigueoka, D.; Saad, S. S.; Barão, K.; Ribeiro, C. C. D.; Forones, N.

M. Impact of Genetic Mutations and Nutritional Status on the Survival of Patients with Colorectal Cancer. *BMC Cancer* **2019**, 19 (1), 644.
6. Iacopetta, B. TP53 Mutation in Colorectal Cancer. *Hum. Mutat.* **2003**, 21 (3), 271–276.
7. Guimaraes, D. P.; Hainaut, P. TP53: A Key Gene in Human Cancer. *Biochimie* **2002**, 84 (1), 83–93.
8. Williams, D. S.; Mouradov, D.; Browne, C.; Palmieri, M.; Elliott, M. J.; Nightingale, R.; Fang, C. G.; Li, R.; Mariadason, J. M.; Faragher, I.; et al. Overexpression of TP53 Protein Is Associated with the Lack of Adjuvant Chemotherapy Benefit in Patients with Stage III Colorectal Cancer. *Mod. Pathol.* **2020**, 33 (3), 483–495.
9. Armaghany, T.; Wilson, J. D.; Chu, Q.; Mills, G. Genetic Alterations in Colorectal Cancer. *Gastrointest. Cancer Res.* **2012**, 5 (1), 19–27.
10. Chang, S.-C.; Lin, J.-K.; Lin, T.-C.; Liang, W.-Y. Loss of Heterozygosity: An Independent Prognostic Factor of Colorectal Cancer. World J. *Gastroenterol.* 2005, 11 (6), 778–784..
11. Ogino, S.; Nosho, K.; Irahara, N.; Shima, K.; Baba, Y.; Kirkner, G. J.; Meyerhardt, J. A.; Fuchs, C. S. Prognostic Significance and Molecular Associations of 18q Loss of Heterozygosity: A Cohort Study of Microsatellite Stable Colorectal Cancers. *J. Clin. Oncol.* **2009**, 27 (27), 4591–4598.
12. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B. E.; Sumer, S. O.; Aksoy, B. A.; Jacobsen, A.; Byrne, C. J.; Heuer, M. L.; Larsson, E.; *et al*. The CBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2012**, 2 (5), 401–404.
13. Gao, J.; Aksoy, B. A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S. O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; *et al*. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the CBioPortal. *Sci. Signal*. **2013**, 6 (269), pl1.
14. Warde-Farley, D.; Donaldson, S. L.; Comes, O.; Zuberi, K.; Badrawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C. T.; *et al*. The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res*. **2010**, 38 (Web Server issue), W214-20.
15. Tanner, M.; Hollmén, M.; Junttila, T. T.; Kapanen, A. I.; Tommola, S.; Soini, Y.; Helin, H.; Salo, J.; Joensuu, H.; Sihvo, E.; et al. Amplification of HER-2 in Gastric Carcinoma: Association with Topoisomerase IIalpha Gene Amplification, Intestinal Type, Poor Prognosis and Sensitivity to Trastuzumab. *Ann. Oncol*. **2005**, 16 (2), 273–278.
16. McDonald, S. L.; Stevenson, D. A. J.; Moir, S. E.; Hutcheon, A. W. ; Haites, N. E.; Heys, S. D.; Schofield, A. C. Genomic Changes Identified by Comparative Genomic Hybridisation in Docetaxel-Resistant Breast Cancer Cell Lines. *Eur. J. Cancer* **2005**, 41 (7), 10 86–1094.
17. Negri, T.; Tarantino, E.; Orsenigo, M.; Reid, J. F.; Gariboldi, M.; Zambetti, M.; Pierotti, M. A.; Pilotti, S. Chromosome Band 17q21 in Breast Cancer: Significant Association between Beclin 1 Loss and HER2/NEU Amplification. *Genes Chromosomes Cancer* **2010**, 49 (10), 901–909.

Author

Haryeong Eo is a junior at Shanghai American School Puxi. She is interested in studying biochemical engineering and genetics. She hopes this research paper will open a wider door in preparing herself to study more sophisticated biology topics in the future.

DOI: 10.36838/v5i1.7