A Novel Association of EGFR Gene Alteration with Decreased Glioblastoma Patient Survival Rate

June Hyun
Harrow School, 5 High St, Harrow, Middlesex, HA1 3HP, United Kingdom; junehyun2024@gmail.com
Mentor: Professor Woo Rin Lee

ABSTRACT: Glioblastoma multiforme (GBM) is the most aggressive and prevalent adult central nervous system (CNS) tumor. Epidermal growth factor receptor (EGFR) alterations are prognostic and diagnostic markers of GBM. Our study aims to investigate the association of EGFR alteration with changes in GBM patients' overall survivability. cBioPortal genomic data analysis was used to retrieve genomic data and clinical attributes from 1,122 GBM patients. We performed a comparative study on EGFR-altered and non-altered patient groups, mutation and amplification frequency analysis, EGFR mutation analysis, and clinical attribute analysis. The mutation and amplification frequency analysis showed that only the EGFR gene had higher amplification and mutation counts in the deceased group, indicating a notable association between EGFR alteration and decreased survivability. Mutation analysis identified the mechanism by which EGFR mutation altered signal transduction and cell proliferation, leading to uncontrolled cell division. A clinical attribute analysis on the relationship between TERT expression, EGFR alteration, and chromosome 7 gain/chromosome 10 loss (Chr 7 gain/Chr 10 loss) suggested the viability of targeting EGFR signaling for GBM therapy. Genomic data analysis revealed a strong association between EGFR gene alteration and decreased GBM patient survival rate. Thus, potential EGFR molecular alteration is a field of therapeutic value.

KEYWORDS: Biomedical Engineering, Cancer Biology; Genetics; EGFR; Patient Survival; Glioblastoma.

Introduction

Glioblastoma multiforme (GBM) is a grade IV astrocytoma that accounts for 47.7% of all central nervous system (CNS) tumor diagnoses; over 12,000 Americans receive a GBM diagnosis annually.1 90% of GBMs occur de novo in the brain, while the remaining 10% develop from low-grade astrocytoma. Malignant cells transfer to adjacent brain cells, contributing to the tumor's rapid growth. Web-like vein synthesis and the sheer size of the mass often cause peritumoral edema, intracranial hypertension, and tumor-induced epilepsy. Thus, the median survival time is 15 months, and the 5-year survival is less than 5%.2 Although surgical removal of over 85% of the tumor does prolong survival, radio-chemo-therapy-resistant GBM relapse is common. IDH1 mutation, MGMT methylation status, TERT promoter mutation, and EGFR mutation are all prognostic and diagnostic markers of GBM.3 Therefore, our research aims to describe the relationship between EGFR mutation and overall patient survivability to address the therapeutic value of the molecular alteration.

The accumulation of genetic and epigenetic alterations causes cancer development.4 DNA mutations, rearrangements, deletion, and amplification are well-known abnormalities in cancer cells. Recent advancements in sequencing technology allow the study of the human genome in nucleotide resolution. Identifying the genetic differences that may cause cancer is possible by sequencing DNA or RNA. By sequencing DNA or RNA, identifying the genetic differences that may cause cancer is possible. This approach also quantifies the activity of genes encoded in human DNA to understand the proteins involved in cancer cell characteristics, such as uncontrolled cell growth.5

When cancer-causing genetic alterations are identified, we can better understand the molecular basis of cancer development. Therefore, clinical data that describes each cancer patient and their genomic data are essential.6 Recently, cBioPortal, the open-source cancer genomic database with patient clinical data, was shared with researchers worldwide. As cBioPortal provides a significant portion of worldwide cancer genomic data, it opens new opportunities for discovering gene alterations that cause cancer development.7

Epidermal growth factor receptor (EGFR) is a transmembrane glycoprotein receptor tyrosine kinase (RTK) responsible for cell proliferation, differentiation, and survival.8 Seven ligands, EGF, TGFα, AREG, EREG, BTC, HBEGF, and EPGN, bind to the extracellular domain. The ligand interaction with the receptor homodimerizes or heterodimerizes EGFR with a protein of the ErbB (RTK) family.9 The dimerization enables the intracellular c-terminus of the EGFR to autophosphorylate, triggering many different canonical and non-canonical intracellular cascades. Point mutation, amplification, and up-regulation of the EGFR gene are hallmarks of tumorigenesis and are often used in diagnosing non-small cell lung cancer (NSCLC) and anal cancer.10 EGFR-positive adenocarcinoma of the lung has a median overall survival of 36.9 months, while EGFR-negative adenocarcinoma has a median overall survival of 55.4 months. Loss of regulation of EGFR is present in 50% of GBM patients. Therefore, investigating the relationship between EGFR gene up-regulation and patient survival is vital to identify a new EGFR-targeting treatment pathway in GBM.11
Since gene copy number alteration and mutation may cause the loss of regulation of EGFR, we investigated the gene using the genomic data of 1,122 GBM patients in this study. The low survivability of GBM and common relapses necessitate investigations to discover novel genetic associations with GBM. However, research has yet to thoroughly investigate the relationship between patients’ survival rate and EGFR gene mutation in GBM patients. Therefore, we hypothesized that genetic alteration of EGFR may affect the GBM patient survival rate.

Methods

**cBioPortal genomic data analysis:**

We used the cBioPortal database was used to retrieve genomic data and clinical attributes from 1,122 GBM patients and performed two comparative survival analyses. The first analysis divided the patients into two groups - EGFR-amplified and EGFR-non-amplified patients. The second analysis divided the patients into EGFR-mutated and EGFR-non-mutated groups. Then, we analyzed the overall patients’ survival with statistical analysis by calculating the Log-rank p-value. **Analyzing EGFR amplification and mutation frequency:**

We performed gene amplification and mutation enrichment analysis using cBioPortal. After collecting the genomic and survival data of 1,122 GBM patients, we divided the patients into living and deceased groups. First, enrichment analysis ranked the most amplified and mutated genes in 1,122 GBM patients. Then, we performed statistical analysis to calculate the p-value using Wilcoxon Test.

**EGFR mutation analysis:**

The cBioPortal mutation analysis function plotted the illustration of EGFR mutation numbers and positions. The x-axis represented the position of the amino acid sequence of EGFR and the position of the mutations within the protein domain; the y-axis represented the number of mutations found in GBM patients.

**GBM patient’s clinical attribute analysis:**

cBioPortal is a web-based tool that allows users to analyze and visualize the clinical attributes of cancer patients. Using cBioPortal, we analyzed the clinical attributes of 1,122 GBM patients within two groups: patients with or without EGFR alterations. In addition, cBioPortal generates a visual representation of the data, such as a bar chart or scatter plot.

Results

**Figure 1:** EGFR amplified and mutated GBM patients show significantly lower survival rates than EGFR non-amplified and non-mutated GBM patients. (A) A graph indicating the probability of overall survival (y-axis) and overall survived month (x-axis) of two groups: EGFR amplified patients and EGFR non-amplified patients. (B) A graph indicating the probability of overall survival (y-axis) and overall survived month (x-axis) of two groups: EGFR mutated patients and EGFR non-mutated patients. The p-value (log-rank test) and median overall survival (95% CI) are indicated in the graph.

We analyzed EGFR amplification and EGFR mutation against the patient’s overall survival months to investigate the genetic association between the EGFR gene and GBM survivability. We used the cBioPortal database to retrieve genomic data from 1,122 GBM patients and performed two comparative survival analyses. The first analysis divided the patient group into EGFR-amplified and EGFR-non-amplified patients; the second analysis divided the patient group into EGFR-mutated and EGFR-non-mutated patients. The overall survival analysis result indicates that EGFR-amplified patients have a significantly decreased survival rate than EGFR-non-amplified patients (Figure 1A). The median overall survival in months is 14.30 for EGFR-amplified patients; the median overall survival in months is 36.80 for EGFR-non-amplified patients (Figure 1A). The difference in the median overall survival, 22.50 months, suggests EGFR amplification’s role in poor prognosis in GBM patients (Figure 1A). The second patient’s comprehensive survival analysis shows that EGFR mutation decreases the survival rate compared to EGFR non-mutated patients (Figure 1B). EGFR-mutated patients’ median overall survival in months is 13.80, while EGFR-non-mutated patients’ median overall survival in months is 44.40 (Figure 1B). The even more significant difference between the median overall survival, 30.60 months, suggests that EGFR mutation influences GBM patient prognosis negatively (Figure 1B). EGFR amplification and EGFR mutation are significant factors that affect GBM patient survival rates. However, since EGFR mutation has a more considerable difference between median overall survival in months than EGFR amplification, the EGFR gene mutation might be a more significant factor in overall survival.

**Figure 2:** EGFR gene amplification event frequency is the highest of the ten genes, with the largest disparity in amplification frequency between living and deceased patient groups in GBM. Living (n=472) and deceased (n=260) patients were analyzed. P < 0.05(*)

Since gene amplification occurs in many genes in GBM patients, we analyzed the frequency of gene amplification in the whole genome of living and deceased GBM patients. We found ten genes with the most significant difference (represented by an asterisk) in amplification frequency between living and deceased patient groups (Figure 2). The top ten amplified genes were: EGFR, SEC61G, LANCL2, VOPP1, VSTM2A, FKBP9P1, AGAP2, CDK4, MARCHF9,
and TSPAN31 (Figure 2). All ten genes showed significantly higher amplification frequency in the deceased group (Figure 2). The EGFR gene showed the highest amplification frequency in the deceased group and the largest difference between amplification frequency in living and deceased groups (Figure 2). Since the deceased group had a higher amplification frequency than the living group in all ten genes, the entire genome of the deceased group may have low stability. Furthermore, the high amplification frequency of EGFR in the deceased group suggests that the amplification of EGFR may contribute to a lower survival rate of GBM patients.

![Figure 3: EGFR mutation frequency is significantly higher in the deceased than in the living patient group. Living (n=477) and deceased (n=268) patients were analyzed P < 0.05 (9)](image)

Mutation occurs throughout the genome of a GBM patient; thus, we analyzed the mutation frequency of the whole genome of living and deceased patients. We found the top ten genes with the most difference in mutation frequency between living and deceased groups. Only four genes, IDH1, ATRX, EGFR, and CIC, showed a significant difference in mutation frequency (Figure 3). Of the four genes, IDH1, ATRX, and CIC showed a higher mutation frequency in the living group, while EGFR displayed a higher mutation frequency in the deceased group (Figure 3). The significantly higher mutation frequency in the living group for the IDH1, ATRX, and CIC genes suggests that a high mutation frequency in the living group may not be associated with a lower patient survival rate. Since EGFR is the only gene with a significant mutation frequency in the deceased group, a high mutation frequency of EGFR may lower the survival rate of GBM (Figure 3).

**Figure 4: Representation of mutation position in amino acid of EGFR in GBM patients.** The x-axis represents the amino acids 0-1210 of the EGFR amino acid sequence; the y-axis represents the number of mutations found in each EGFR amino acid position.

The EGFR gene is significantly mutated in deceased patients. Thus, we analyzed the position of EGFR mutation in the EGFR amino acid sequence to find important mutation positions associated with lower survival in GBM patients. We found five separate domains in the EGFR amino acid sequence: Receptor L domain (57-167 aa), Furin-like cysteine-rich region (185-338 aa), Receptor L domain (361-480 aa), Growth factor receptor domain IV (505-636 aa), and Protein tyrosine kinase (713-965 aa) (Figure 4). The number of EGFR mutations was highest in the Furin-like cysteine-rich region (Figure 4). The growth factor receptor domain IV had the second highest number of EGFR mutations out of the domains (Figure 4). Previous studies confirm that EGFR mutations at alanine 289 (EGFR A289D/T/V) of the Furin-like cysteine-rich region are associated with a significant reduction in the overall survival of GBM patients. Therefore, the high number and frequency of EGFR mutations in the Furin-like cysteine-rich region and the growth factor receptor domain IV suggest that mutations in the two domains may be significantly associated with lower survival rates in GBM patients.

**Figure 5: The EGFR altered patient group significantly increased TERT expression level and Chr 7 gain/Chr 10 loss.** (A) The graph shows the TERT expression level (log 2) of EGFR-altered and EGFR-unaltered patient groups (Wilcoxon Test, p < 0.01). (B) The graph shows the percentage of samples with Chr 7 gain/Chr 10 loss of EGFR-altered and EGFR-unaltered patient groups (Wilcoxon Test, p < 0.01).

TERT expression increase led to poor overall survival in GBM patients in previous studies. Therefore, we analyzed TERT expression in EGFR-altered and unaltered patient groups to investigate whether EGFR alteration affected TERT expression. In addition, we also analyzed the percentage of Chr 7 gain/Chr 10 loss in EGFR-altered and unaltered patient groups to study the genomic instability of GBM patients with EGFR alteration. We found that the EGFR-altered group (median = 4.39) showed significantly increased TERT expression levels than the EGFR-unaltered group (median = 1) (Figure 5A). Furthermore, the EGFR-altered group showed significantly higher Chr 7 gain/Chr 10 loss (78.87%) than the EGFR-unaltered group (17.82%) (Figure 5B). In conclusion, the high level of TERT expression in the EGFR-altered group suggests a significant association between EGFR alteration and increased TERT expression that may cause a poor survival rate in GBM patients. Also, increased Chr 7 gain/Chr 10 loss in the EGFR-altered group further suggests an association between EGFR alteration and increased genomic instability, influencing overall survivability.

**Discussion**

This paper is the first study to investigate the association of EGFR gene amplification and mutation with overall survivability in GBM patients. Figure 1 shows the comparison of survival months of EGFR-amplified, non-amplified, mutated,
and non-mutated patients. The graph shows an apparent reduction in survival for EGFR-amplified and mutated patients, highlighting a significant association between amplification, mutation, and the deceased patient group. Furthermore, EGFR-mutated and non-mutated groups have a larger difference in median overall survivability than EGFR-amplified and non-amplified groups, suggesting that EGFR mutation, rather than amplification, may more significantly affect patient survivability in GBM. The strong association between EGFR gene status and patient survival suggests that the gene status may be used as a new biomarker factor to consider when predicting GBM patient prognosis, improving the prediction accuracy, and providing patients and relatives with a confident disclosure.

Previous studies show similar EGFR gene relationships with patient survivability in NSCLC. The likelihood of EGFR amplification increases in advanced clinical stages of lung cancer. Furthermore, EGFR amplification increases with EGFR tyrosine kinase inhibitor (TKI) resistance. As TKI is a common drug used to treat NSCLC, TKI resistance is associated with poor outcomes in NSCLC patients. Therefore, EGFR amplification is significantly associated with low survivability of patients, which agrees with our study's result from Figure 1.

Past studies investigating EGFR in GBM corroborate the result. EGFR amplification observed in this paper may lead to EGFR overexpression in approximately 50-60% of GBM tumors. Furthermore, EGFR mutations lie primarily in the extracellular domain. Most mutations in Figure 4 lie between amino acids 0-600 aa in the extracellular domain showing consistency with the previous study.

We analyzed the gene amplification and mutation frequency in the whole genome of living and deceased GBM patients to further investigate the association of EGFR amplification and mutation with decreased patient survivability. Many genes, including EGFR, displayed significantly higher amplification frequency in deceased patient groups than in living patient groups, suggesting that genetic instability of the whole genome contributes to the lower survivability of patients with EGFR amplification observed in previous analyses. On the other hand, mutation frequency analysis identified four genes (IDH1, ATRX, EGFR, and CIC) with significant differences between living and deceased group mutation frequency. IDH1, ATRX, and CIC displayed higher mutation frequency in living groups, suggesting that mutations in these genes do not lower the survivability of GBM patients. However, EGFR had a higher mutation frequency in the deceased group. Therefore, the analysis found that EGFR is the only gene in the GBM patient’s whole genome with a higher amplification and mutation frequency which may lower survivability, highlighting the importance of further research on the role of EGFR in GBM.

This study analyzed the position of mutations in the EGFR amino acid sequence to identify mutations that significantly affect survivability. We categorized five distinct domains in the sequence, of which three (furin-like cysteine-rich region, growth factor receptor domain IV, and receptor L domain) contained many EGFR mutations. Furin-like domain facilitates receptor aggregation for signal transduction by RTKs. A mutation in the furin-like domain may enhance the internal signaling, promoting uncontrolled cell growth. Furthermore, previous studies indicate that mutation in the furin-like domain correlates with less effective targeted therapy and potential drug resistance, decreasing survivability in GBM patients.

Growth factor receptor domain IV facilitates the dimerization of EGFR when EGF is received, which regulates phosphorylation and subsequent signal transduction for cell growth. Ligand stimulation in mutated growth factor receptor domain IV results in the formation of a tyrosine-phosphorylated, disulfide-bonded dimer without EGF. Therefore, the mutation causes EGFR to trigger cell growth without EGF binding, leading to unregulated cell growth in GBM.

The analysis of mutation positions in the EGFR amino acid sequence identified that most mutations in the receptor L domain occurred in the L1 subdomain, located near the N-terminus and responsible for the binding of ligands. The L1 subdomain causes a conformational change in EGFR when a ligand binds to the receptor, influencing the intracellular domain to trigger downstream signaling pathways that promote cell growth and division. Therefore, a mutation in L1 can disrupt ligand binding to EGFR or enhance the signaling pathways, promoting cancer growth. Furthermore, previous studies suggest a link between L1 subdomain mutation and cancer drug resistance. The analysis highlights the multitude of EGFR mutations that affect signal transduction triggering cell growth and division, suggesting that EGFR may be an area of interest for therapeutic research for GBM drugs.

We investigated EGFR alteration association with TERT expression and Chr 7 gain/Chr 10 loss to compare the clinical attributes of GBM patients with and without EGFR alteration. The previous study indicated that EGFR signaling induces TERT expression by transcription factors such as c-Myc, Sp1, and Ets-2. Since TERT is often increased in cancer cells that activate the infinite cell division, targeting EGFR signaling is considered an effective therapeutic strategy for lung cancer. 21 Chr 7 gain/Chr 10 loss is also often associated with EGFR gene mutation in lung cancer. Previous studies indicate that these mutations may activate EGFR signal pathways, contributing to cancer development and growth. Our study shows a similar relationship between EGFR, TERT, and Chr 7 gain/ Chr 10 loss exists in GBM, demonstrating the importance of further investigation into the targeting of EGFR alteration for novel therapeutic pathways.

**Conclusion**

In this study, we found that EGFR amplification and mutation significantly reduce GBM patient survival rate and that mutation might significantly influence overall survivability. EGFR amplification had the highest frequency in the deceased group out of the top ten genes, with the largest disparity in frequency between living and deceased groups, suggesting that EGFR amplification may affect survivability. Out of four significant mutations, only the EGFR mutation had a significant mutation frequency in the deceased group; thus, a higher EGFR mutation frequency may lower the survival rate. When we further analyzed the EGFR mutation position, the high mutation frequency was detected in the furin-like cysteine-
rich region. Growth factor receptor domain IV suggests a significant association between lower survivability and the two domains. We also analyzed the clinical attributes associated with EGFR-altered patient groups. We found that high levels of TERT expression and Chr 7 gain/Chr 10 loss in EGFR-altered groups show that alteration in EGFR may lower survivability in GBM patients and increase genomic instability. 

This study analyzed patient genomic data associated with decreased survival rates from one published database. Therefore, we must conduct in vitro experiments to further analyze the effect of EGFR amplification and mutation on cancer cell function, such as the speed of cancer cell division, migration, and invasion. Also, we should perform an in vivo mouse experiment or xenograft to verify the results, indicating that EGFR amplification and mutation decrease the patient’s survival rate. In addition, many other genes, such as SEC61G and LANCL2, have high amplification frequencies in the deceased group. Thus, we must analyze other amplified genes in the deceased group indicated in Figure 2 to discover any interactions between genes that lead to decreased survival rates in GBM patients.

**Acknowledgments**

I want to thank Professor Woo Rin Lee at the University of Suwon for acting as a mentor and remotely guiding my research. Furthermore, I would like to show appreciation for Dr. Luca Passamonti at Biogen, who inspired my passion for neurology and degenerative diseases.

**References**


**Author**

June Hyun is a junior at Harrow School in the United Kingdom. As the Scientific Society President, he is involved in various school volunteering and research activities. His interest in neurology, neurodegenerative diseases, and robotic neurorehabilitation drive his aspiration to become a physician-researcher.