*Proceedings of the 2012 IEEE 12th International Conference on Bioinformatics & Bioengineering (BIBE), Larnaca, Cyprus, 11-13 November 2012*

# Investigating Survival Prognosis of Glioblastoma Using Evolutional Properties of Gene Networks

Alex Upton, Theodoros N Arvanitis
School of Electronic, Electrical and Computer Engineering
University of Birmingham
Birmingham, B15 2TT, England, UK
{axu243, t.arvanitis} @bham.ac.uk

*Abstract*—**In recent years, there has been widespread interest and a large number of publications on the application of graph theory techniques into constructing and analyzing biologically-informed gene networks from cancer cell line data sets. Current research efforts have predominantly looked at an overall static, topological, representation of the network, and have not investigated the application of graph theoretical techniques to evolutionary investigations of cancer. A number of these studies have used graph theory metrics, such as degree, betweenness, and closeness centrality, to identify important hub genes in these networks. However, these have not fully investigated the importance of genes across the different stages of the disease.**

**Previous human glioblastoma publications have identified four subtypes of glioblastoma in adults, based on signature genes. In one such publication, Verhaak et al. found that the subtypes correspond to a narrow median survival range, from 11.3 months for the most aggressive subtype, to 13.1 months for the least aggressive one.**

**In this work, we present an evolutionary graph theory study of glioblastoma based on survival data categorization, confirming genes associated with different survival times identified using established graph theory metrics. The work is extending the application of graph theory approaches to evolutionary studies of cancer cell line data.**

***Keywords-gene network; glioblastoma; graph theory; genomics; cancer evolution, glioblastoma evolution***

## I. INTRODUCTION

Progress in technology, in recent years, allowing measurements of biological data, such as DNA sequencing, protein-protein interactions and gene expression profiles, has led to an increase in the amount of biological data that has become available for further analysis. As a result, tools and techniques used in other disciplines, such as computer science and engineering, have been applied to this data. One particular approach adopted is that of gene networks, which enables gene interactions to be modeled and visualized. Various mathematical approaches have been used to construct these networks; such as correlation networks [1], Bayesian networks [2], and mutual information networks [3]. These networks can then be analyzed using established centrality metrics, such as degree centrality, betweenness centrality, and closeness centrality. However, these approaches tend to focus on 'static', topology-based, representations of data; grouping all the data together and representing it as a single model [4]. In a number of biological scenarios, the evolution of the data is of interest, and an overall static representation may not be informative enough. One such example of this being cancer data sets, where a data set contains information from a number of cancer cell lines at different evolutionary stages of the disease.

To deal with this issue, we propose an evolutionary approach to constructing networks for genomic cancer data sets. A publically available glioblastoma data set from a previously published study is used to illustrate this method [5]. Glioblastoma is the most common primary malignant brain tumour in adults and also one of the most lethal forms of cancer, with a median survival time of only 14 months from the time of first diagnosis [6]. Previous glioblastoma studies, such as that by Verhaak et al, have identified four glioblastoma subtypes, and a number of signature genes that correspond to each subtype [7]. This particular study found that the four subtypes had a narrow median survival range, from 11.3 months for the most lethal subtype, to 13.1 months for the least lethal subtype. Whilst identifying genes that correspond to each subtype, this study did not identify genes that are associated with both poor glioblastoma prognosis, i.e. low survival time, and better glioblastoma prognosis, i.e. longer survival time. In this paper, the proposed methodology identifies genes that are associated with different survival times in glioblastoma cell lines, and can therefore be used as potential prognostic biomarkers of the disease.

## II. METHOD

### A. Categorising the Data Set

The dataset consists of two independent sets of clinical tumour samples, 55 and 65 samples, respectively, obtained at the time of surgery at UCLA [5]. Gene expression profiling of these samples was carried out using Affymetrix high-density oligonucleotide microarrays [8].

Survival data is available for all these samples, allowing categorization of the samples based on survival time. Across

the 120 samples in the dataset, the mean survival time is 447.29 days, with a standard deviation of 426.15 days. It is worth noting that the mean value for survival correlates very closely to the afore-mentioned median survival time of around 14 months, indicating that this is a typically representative glioblastoma dataset. The high value of the standard deviation across the samples, compared to the mean value, shows how the range of values varies greatly. This is shown with the minimum survival time of 7 days, and the highest survival time of 2807 days, giving a range of 2800 days. These high values for both the range and standard deviation give an indication of the potential prognostic differences amongst the samples within the dataset. It also highlights the issue with taking an overall view of the dataset and creating a single network; a vast amount of information about the evolution of the stages of the cancer is potentially lost.

From the data, five categories of samples were created, using survival time as the class discrimination. The five categories are as follows:

- 200 or fewer survival days
- 201- 400 survival days
- 401- 600 survival days
- 601- 800 survival days
- 801 and more survival days

### B. *Constructing the Networks*

For these five categories identified, networks are constructed from the samples belonging to each group. The WGCNA [1] library of functions for the open source statistical analysis programming language R [1, 9] is used to construct a weighted gene correlation network for each category, using a soft threshold power to construct a scale-free network We assume a scale-free character based on previous work. Instead of having a random Poisson distribution of connections, scale-free networks have a power law distribution of connections [10]. In a number of studies, it has been shown that many biological networks have this feature, with a few highly connected nodes, and many low connected nodes [11, 12]. In fact, many biologists would be wary of a gene network that was not scale-free [13].

Once the weighted networks have been constructed, all edges with a weight of less than 0.2 are removed. This is done for two reasons. The first is for biological significance; interactions with a weight of less than 0.2 are unlikely to be biologically meaningful. The second is for computational reasons; calculation of betweenness centrality for each of the original five networks takes around 30 hours, in the network with edges less than 0.2 removed it takes under 30 seconds.

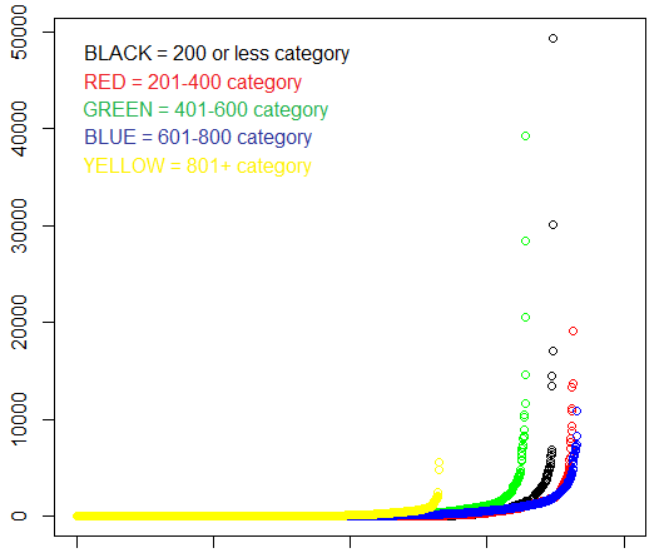
### C. *Analysing the Networks*

The network metrics of betweenness, closeness, and degree centrality are calculated using the igraph library of functions in R [14]. For accurate calculation of metrics of weighted networks, igraph version 0.6 and later is required.

The most commonly applied metric to network analysis is the degree centrality. This metric measures the degree of a node, how many other nodes it is connected to [15]. Extending this for weighted networks, a more useful metric is to measure the sum of the edge weights that a node has to other nodes. As such, this metric will be used, and will be referred to as degree centrality in this paper. Two other commonly used centrality measures are the betweenness centrality, and the closeness centrality. The betweenness centrality measures how many shortest paths in a network pass through a node [16]. This gives a measure of how important a node is in terms of controlling the flow of information in a network. The closeness centrality gives a measure of how close a node is to other nodes in a network, with the distance measured as the shortest path between two nodes [15].

Rankings are assigned to each node in the five networks for each of these metrics, which are then added together, and each gene is then re-ranked based on the total of these scores. This gives the overall ranking score for each node in each category of network. By using the ranking score of each metric, as oppose to the raw score, this ensure that each metric has equal weighting. The figure below shows a plot for the raw total scores for each category. As can be seen from this figure, there is a big difference between the genes with the highest raw scores, and the lowest raw scores. This can in part be attributed to the previously mentioned scale-free behavior; many genes have low degree, thereby having a low raw score for degree, whilst a few have high degree, thereby having high raw score for degree.

FIGURE I TOTAL RAW SCORES OF NETWORK CATEGORIES



### D. Investigating the Top Ranked Genes

Having identified 20 genes of interest in each category based on their rankings, the next step is to see whether any of these genes have previously been identified as being of interest for glioblastoma. There are a large number of publications on glioblastoma, and as such, it would be impossible to consult every single one. Therefore, a number of tools will be used to help with this, namely The Cancer Genome Atlas, the IntOGen browser, and CGPrio [17-19]. Using these tools one can highlight whether a gene has been previously identified as a gene of interest.

## III. RESULTS

Having calculated the ranking for each gene in the five categories of network, the results are presented in the following five sub-sections. Each sub-section presents the top 20 ranked genes for the combined degree, betweenness, and closeness centrality rank in that category. As well as using these metrics, a Spearman Rank correlation coefficient score has been calculated for the common genes in each category of network, to allow further detailed comparison.

### A. 200 or fewer survival days category

Table I below shows the top 20 ranked genes in the 200 or less survival days category. There are a number of results from this list that stand out. The most notable here is SNAP91, which has been identified in a number of studies as a signature gene of the proneural glioblastoma subtype. This is the only signature gene of any glioblastoma subtype that occurs in the top 20 ranking. The 2nd ranked gene in this list, SYT1, is highly ranked as an oncogene by CGPrio, and is also identified in a glioblastoma study by Dong et al as being a candidate gene for the disease [20]. The 9th ranked gene, PRKCZ, has been identified in 3 glioblastoma gene lists by the cancer genome atlas gene checker. PRKCZ has also been show to be crucial to proliferation in glioblastoma cell lines [21]. 8 of the top 20 ranked genes in this category do not appear in the top 20 rankings for any of the other categories, including the top 4 ranked genes. This suggests that these genes identified as being important in this category do not have such an important role in the other categories of network.

In contrast, of the 12 genes in the top 20 list that do appear in top 20 lists for other categories, 3 genes, SNAP91, SYN1, and RGS7 appear in three of the five top 20 lists, including this category. SNAP91 has already been highlighted, however the two other genes have not previously been identified as candidate glioblastoma genes, and this suggests that they may play a role across various stages of the glioblastoma life cycle.

TABLE I TOP 20 RANKED GENES 200 OR LESS SURVIVAL DAYS CATEGORY

| *Gene* | *Overall Rank* | *Gene* | *Overall Rank* |
|---|---|---|---|
| AK5 | 1 | SLC17A7 | 11 |
| SYT1 | 2 | MAST3 | 12 |
| MAP1A | 3 | SYN1 | 13 |
| HSPA12A | 4 | PAK6 | 14 |
| SNAP91 | 5 | VAMP2 | 15 |
| SULT4A1 | 6 | RGS7 | 16 |
| EPB41L1 | 7 | MOAP1 | 17 |
| SCN2A2 | 8 | NAP1L2 | 18 |
| PRKCZ | 9 | NRGN | 19 |
| STXBP1 | 10 | CYFIP2 | 20 |

### B. 201-400 survival days category

Table II below shows the top 20 ranked genes in the 201-400 survival days category. Once again, the presence of SNAP91 amongst the top 20 ranked genes stands out. The second result of note is that the two top ranked genes are both solute carriers; SLC9A6 is a sodium/hydrogen exchanger, and SLC8A2 is involved in sodium/calcium exchange. This result would suggest that the exchange of sodium plays a role in glioblastomas within the 201-400 survival days category, and potentially is an area of interest for glioblastoma studies. 10 of the top 20 ranked genes in this category do not appear in the top 20 rankings for any of the other categories, including the top 2 ranked genes previously discussed. This potentially suggests that the importance of sodium exchange is limited to glioblastomas within this category, and that, as before, there are a number of genes identified as being important in this category that do not have such an important role in other categories.

Of the 10 genes that do appear in top 20 ranked lists in other categories, there are again 3 genes that appear in three of the five top 20 lists, including this category. As well as SNAP91, these include PPP1R16B, and MGC8407. Whilst these two genes have not been specifically identified as candidate glioblastoma genes, it should be noted that MGC8407 has been identified as potential cancer gene target by the Broad Institute research group [22].

TABLE II TOP 20 RANKED GENES 201-400 SURVIVAL DAYS CATEGORY

| *Gene* | *Overall Rank* | *Gene* | *Overall Rank* |
|---|---|---|---|
| SLC9A6 | 1 | MGC8407 | 11 |
| SLC8A2 | 2 | DYNC1I1 | 12 |
| INA | 3 | PHYHIP | 13 |
| SNAP91 | 4 | KCNAB2 | 14 |
| PPP1R16B | 5 | NAP1L2 | 15 |
| MEF2C | 6 | MAST3 | 16 |
| WDR7 | 7 | NP25 | 17 |
| CYFIP2 | 8 | KIAA1940 | 18 |
| KIAA0513 | 9 | STXBP1 | 19 |
| PDE2A | 10 | MOAP1 | 20 |

### C. 401-600 survival days category

Table III below shows the top 20 ranked genes in the 401-600 survival days category. The previously mentioned genes MGC8407, SYN1 and RGS7 appear in this list, and are again the only genes that appear in two other top 20 ranked lists, as well as this one. The presence of gene EPHB6 is interesting; as well as being identified in one glioblastoma specific gene list, it has also been identified as being on six other cancer gene lists, such as ovarian cancer and breast cancer, by the cancer genome atlas. SH3GL2 is another glioblastoma gene of interest identified in the glioblastoma

study by Dong et al, and also a 2008 study by Chang [23]. It is also a signature gene of the neural glioblastoma subtype, as are the genes CPNE6, and HPCAL4. It is worth noting the presence of 3 neural subtype signature genes in the top 20 ranked genes for this category.

13 of the 20 genes that appear in this list do not appear on any of the other top 20 genes lists. This is a greater number of unique genes than the previous two lists, suggesting increasingly different network behavior as the survival time increases. It also suggests that genes that are important for the behavior of the network in the two previous categories are not as important for the function of the network in this category, and that different processes are taking place that these genes are not involved in.

TABLE III TOP 20 RANKED GENES 201-400 SURVIVAL DAYS CATEGORY

| *Gene* | *Overall Rank* | *Gene* | *Overall Rank* |
|---|---|---|---|
| RIMS3 | 1 | BZRAP1 | 11 |
| MGC8407 | 2 | SH3GL2 | 12 |
| CA11 | 3 | PCSK2 | 13 |
| CHGB | 4 | PRKAR1B | 14 |
| GLS2 | 5 | HPCAL4 | 15 |
| NELL2 | 6 | MYRIP | 16 |
| EPHB6 | 7 | CPNE6 | 17 |
| INA | 8 | SYN1 | 18 |
| GAD2 | 9 | PAK6 | 19 |
| KIAA1107 | 10 | RGS7 | 20 |

### *D. 601-800 survival days category*

Table IV below shows the top 20 ranked genes in the 601-800 survival days category. The 2nd ranked gene, VAMP2, which is the 15th ranked gene in the lowest survival category network, is ranked by IntOGen as having a high probability of being an oncogene, as is the 5th ranked gene SCAMP5. This would suggest that these genes are highly likely to be involved in interactions with other genes, and the high ranking results here concur with that prediction. The gene PRKCZ is the 20th ranked gene in this list, having previously been highlighted as being fundamental in glioblastoma proliferation in human cell lines.

There are again 13 unique entries on the top 20 ranking list for this category of network, suggesting, as was the case with the last network category, that there is markedly different behavior in the network, compared to the other categories of network. The three genes in this list that also occur in two other lists are PPP1R16B, RGS7, and the proneural signature gene SNAP91.

TABLE IV TOP 20 RANKED GENES 601-800 SURVIVAL DAYS CATEGORY

| *Gene* | *Overall Rank* | *Gene* | *Overall Rank* |
|---|---|---|---|
| TUBB | 1 | SNAP91 | 11 |
| VAMP2 | 2 | ATP6V1G2 | 12 |
| HLF | 3 | RGS7 | 13 |
| ARHGEF9 | 4 | GFOD1 | 14 |
| SCAMP5 | 5 | PDE2A | 15 |
| PPP3CB | 6 | ATP2B2 | 16 |
| SNPH | 7 | CHGA | 17 |
| PPP1R16B | 8 | EPB49 | 18 |
| GOT1 | 9 | S100A1 | 19 |
| IQSEC3 | 10 | PRKCZ | 20 |

### *E. 801+ survival days category*

Table V below shows the top 20 ranked genes in the 801+ survival days category. The presence of GABRD and GABRA1 as the top two ranked genes is immediately noticeable, suggesting that the GABR area is of interest. In fact, whilst these two genes are not signature genes of any glioblastoma subtype, the gene GABR2 is a proneural signature gene, as identified by Verhaak et al [7]. The 9th ranked gene, KALRN, appears in 3 glioblastoma gene lists in the cancer genome atlas gene ranker, as well as appearing in 3 other cancer gene related lists.

As with the two previous categories, there are 13 unique entries on the top 20 ranking list for this network. There are also 3 genes in the list below that also appear on two other lists, these genes are MGC8407, SYN1, and PPP1R16B.

TABLE V TOP 20 RANKED GENES 801+ SURVIVAL DAYS CATEGORY

| *Gene* | *Overall Rank* | *Gene* | *Overall Rank* |
|---|---|---|---|
| GABRD | 1 | GLS2 | 11 |
| GABRA1 | 2 | CACNG3 | 12 |
| EPB41L1 | 3 | SLC12A5 | 13 |
| BSN | 4 | KIAA1107 | 14 |
| MGC8407 | 5 | PNOC | 15 |
| DRD1IP | 6 | ARK5 | 16 |
| NY-REN-7 | 7 | SYN2 | 17 |
| SYN1 | 8 | CABP1 | 18 |
| HAPIP | 9 | GOT1 | 19 |
| PPP1R16B | 10 | SNCB | 20 |

### *F. Spearman Rank Correlation Coefficient Scores*

Having calculated the overall ranks for each gene in the five categories of network, the next step is to calculate the correlation between the common genes in the different categories of network. Previously, it had been noted that a number of genes occur in more than one top 20 ranked list. As such, extending this to calculate the correlation between all the common genes in the categories gives a good way of measuring how similar the networks are in terms of ranking the nodes by the centrality measures. Table VI below shows the Spearman Rank correlation coefficient scores for all pairs of networks. The high score for correlation between the two lowest survival categories, 0.75, and the relatively low scores for the highest survival category with all the other network categories stand out.

TABLE VI SPEARMAN RANK CORRELATION COEFFICIENT BETWEEN THE DIFFERENT CATEGORIES OF NETWORK

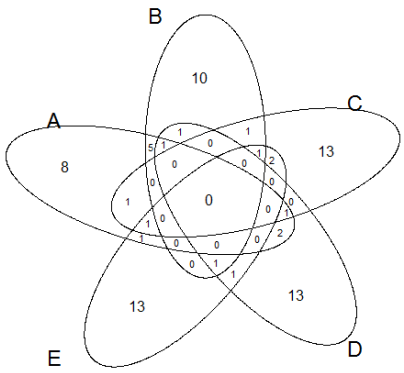
| *Category* | *200* | *201-400* | *401-600* | *601-800* | *801+* |
|---|---|---|---|---|---|
| ***200*** | 1 | 0.753619 | 0.695625 | 0.66201 | 0.525223 |
| ***201-400*** | 0.753619 | 1 | 0.673188 | 0.693234 | 0.562439 |
| ***401-600*** | 0.695625 | 0.673188 | 1 | 0.581407 | 0.514119 |
| ***601-800*** | 0.66201 | 0.693234 | 0.581407 | 1 | 0.509671 |
| ***801+*** | 0.525223 | 0.562439 | 0.514119 | 0.509671 | 1 |

## IV. CONCLUSIONS

There are two main areas of this work that conclusions can be drawn from. The first area to concentrate on is the gene rankings in each category. There are 57 unique entries in the five top 20 ranked gene lists. This represents quite a high proportion of genes being unique to one evolutionary category, and also that there are more unique entries in the top 20 ranked lists than common ones. This high number of unique genes suggests that the five categories of network are very different. This is especially relevant if previous static studies are considered that grouped all the samples in a data set together and constructed one network to represent their behavior, it is quite clear from this study how different the behavior of sample at different evolutionary stages is.

There are 14 genes that appear in two lists. 9 of these appear in the first two categories, suggesting that these two categories are the most similar based on the top ranking genes. These two categories are the two with the lowest survival days, so a reasonable presumption would be to suggest that these 9 genes play a role in the later stages of glioblastoma. It is also worth noting the very high correlation that these two categories have for common genes, 0.75, which is the highest correlation for any two categories. This again suggests that these two categories are similar. 5 genes appear in three lists, including SNAP91. Their presence in lists across the evolutionary stages would suggest they are involved in processes that are common throughout the glioblastoma life cycle, and that they do not play such an important role in the specific processes related to the evolution of glioblastoma. The Venn diagram below shows the overlap of genes between the different network categories, note that as shown by this diagram, there are no genes that appear either in four of five lists.

FIGURE II VENN DIAGRAM SHOWING GENE OVERLAP BETWEEN CATEGORIES




The second area of focus is the biological relevance of these results. The identification of genes previously identified in a number of publications using a solely graph theory approach shows the biological relevance of this approach. Genes such as SNAP91, EPHB6, PRKCZ, CPNE6, HPCAL4 and SH3GL2 have been identified without any prior biological knowledge or bias. A number of glioblastoma signature genes were identified across the evolutionary categories using the metrics; however there was no correlation between the identification of these as being high ranking and the evolutionary category. This corresponds to the study by Verhaak et al that showed the four glioblastoma subtypes had a narrow survival range, as from this study, it cannot be concluded that one glioblastoma subtype can be significantly associated with any of the categories of network. The findings of this study suggest that high network centrality scores for specific genes may be a better indicator of glioblastoma survival time, than subtype classification.

To conclude, this work has identified different genes as being of interest at different evolutionary stages of the glioblastoma disease progression, based on their network metrics. It has shown that in this study, these rankings give a better guide to glioblastoma survival time, than glioblastoma subtype categorization. Our work has built upon previous work, investigating the static networks of disease, and in the future, it is hoped that this method could be applied to other diseases and could be refined further, to aid in the discovery of potential candidate genes of interest in the progression, and hence prognosis, of cancer disease.

## ACKNOWLEDGMENT

The authors would like to thank Steve Horvath for his permission to use the glioblastoma dataset. The work has been partially funded by the EPSRC.

## REFERENCES

[1] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics,* vol. 9, p. 559, 2008.

[2] A. Djebbari and J. Quackenbush, "Seeded Bayesian Networks: Constructing genetic networks from microarray data," BioMed Central Ltd., 2008.

[3] A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Favera, and A. Califano, "ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context," *BMC Bioinformatics,* vol. 7, p. S7, 2006.

[4] A. Platzer, P. Perco, A. Lukas, and B. Mayer, "Characterization of protein-interaction networks in tumors," *BMC Bioinformatics,* vol. 8, p. 224, 2007.

[5] S. Horvath, B. Zhang, M. Carlson, K. Lu, S. Zhu, R. Felciano, M. Laurance, W. Zhao, Q. Shu, Y. Lee, A. Scheck, L. Liau, H. Wu, D. Geschwind, P. Febbo, H. Kornblum, T. Cloughesy, S. Nelson, and P. Mischel, "Analysis of oncogenic signaling networks in Glioblastoma identifies ASPM as a novel molecular target," *Proc Natl Acad Sci USA,* vol. 103, pp. 17402 - 17407, 2006.

[6] E. G. Van Meir, C. G. Hadjipanayis, A. D. Norden, H.-K. Shu, P. Y. Wen, and J. J. Olson, "Exciting New Advances in Neuro-Oncology: The Avenue to a Cure for Malignant Glioma," *CA: A Cancer Journal for Clinicians,* vol. 60, pp. 166-193.

[7] R. G. W. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou, and D. N. Hayes, "Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1," *Cancer cell,* vol. 17, pp. 98-110, 2010.

[8] R. A. Irizarry, B. M. Bolstad, F. Collin, L. M. Cope, B. Hobbs, and T. P. Speed, "Summaries of Affymetrix GeneChip probe level data," *Nucleic Acids Research,* vol. 31, p. e15, February 15, 2003 2003.

[9] R. Ihaka and R. Gentleman, "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics,* vol. 5, pp. 299-314, 1996.

[10] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review,* vol. 45, pp. 167-256, 2003.

[11] S. Bergmann, J. Ihmels, and N. Barkai, "Similarities and Differences in Genome-Wide Expression Data of Six Organisms," *PLoS Biol,* vol. 2, p. e9, 2003.

[12] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi, "Hierarchical organization of modularity in metabolic networks," *Science,* vol. 297, pp. 1551 - 5, 2002.

[13] B. Zhang and S. Horvath, "A General Framework for Weighted Gene Co-expression Network Analysis," *Stat Appl Genet Mol Biol,* vol. 4, p. Article 17, 2005.

[14] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal Complex Systems,* 2006.

[15] L. Freeman, "Centrality in social networks: Conceptual clarification," *Social Networks,* vol. 1, pp. 215 - 239, 1978.

[16] L. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry,* vol. 40, pp. 35-41, 1977.

[17] "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," *Nature,* vol. 455, pp. 1061-1068, 2008.

[18] G. Gundem, C. Perez-Llamas, A. Jene-Sanz, A. Kedzierska, A. Islam, J. Deu-Pons, S. Furney, and N. Lopez-Bigas, "IntOGen: Integration and data-mining of multidimensional oncogenomic data," *Nature Methods,* 2010.

[19] S. J. Furney, B. Calvo, P. LarraÃ±aga, J. A. Lozano, and N. Lopez-Bigas, "Prioritization of candidate cancer genes - an aid to oncogenomic studies," *Nucleic Acids Research,* vol. 36, p. e115, October 1, 2008 2008.

[20] H. Dong, L. Luo, S. Hong, H. Siu, Y. Xiao, L. Jin, R. Chen, and M. Xiong, "Integrated analysis of mutations, miRNA and mRNA expression in glioblastoma," *BMC Systems Biology,* vol. 4, p. 163, 2010.

[21] A. M. Donson, A. Banerjee, F. Gamboni-Robertson, J. M. Fleitz, and N. K. Foreman, "Protein Kinase C ζ Isoform is Critical for Proliferation in Human Glioblastoma Cell Lines," *Journal of Neuro-Oncology,* vol. 47, pp. 109-115, 2000.

[22] R. K. Thomas, A. C. Baker, R. M. DeBiasi, W. Winckler, T. LaFramboise, W. M. Lin, M. Wang, W. Feng, T. Zander, L. E. MacConaill, J. C. Lee, R. Nicoletti, C. Hatton, M. Goyette, L. Girard, K. Majmudar, L. Ziaugra, K.-K. Wong, S. Gabriel, R. Beroukhim, M. Peyton, J. Barretina, A. Dutt, C. Emery, H. Greulich, K. Shah, H. Sasaki, A. Gazdar, J. Minna, S. A. Armstrong, I. K. Mellinghoff, F. S. Hodi, G. Dranoff, P. S. Mischel, T. F. Cloughesy, S. F. Nelson, L. M. Liau, K. Mertz, M. A. Rubin, H. Moch, M. Loda, W. Catalona, J. Fletcher, S. Signoretti, F. Kaye, K. C. Anderson, G. D. Demetri, R. Dummer, S. Wagner, M. Herlyn, W. R. Sellers, M. Meyerson, and L. A. Garraway, "High-throughput oncogene mutation profiling in human cancer," *Nat Genet,* vol. 39, pp. 347-351, 2007.

[23] C.-L. Chang, "Genome-Wide Oligonucleotide Microarray Analysis of Gene-Expression Profiles of Taiwanese Patients with Anaplastic Astrocytoma and Glioblastoma Multiforme," *Journal of Biomolecular Screening,* vol. 13, pp. 912-921, October 1, 2008 2008.