IQGAP and HspB8: potent biomarkers in low grade gliomas

Jeremias Ivan1, David Agustriawan2, Stefanus Satrio Hadi Wibowo3, Arli Aditya Parikesit4 and Rizky Nurdiansyah5

Department of Bioinformatics, School of Life Sciences, Indonesia International Institute for Life Sciences
david.agustriawan@i3l.ac.id

Abstract. Low grade gliomas are invasive brain tumors that mostly occur among young adults. Previous studies showed that LGG is characterized by IDH1/2 mutations; however, there were some cases where the cancer patients suffer low mutation rate of those genes. In this study, two genes are proposed as new biomarkers: IQGAP and HspB8. IQGAP has been found to be correlated with various types of cancer, while HspB8 role in LGG has not been determined, according to F-Census database. This study is aimed to identify the expressions of IQGAP and HspB8 genes among LGG patients and assess their potential to be the next biomarkers. The expression data was derived from The Cancer Genome Atlas project and downloaded using TCGA Assembler. Then, IQGAP – miRNA expression correlations and HspB8 expression analysis were conducted using Matlab and Microsoft Excel, respectively. The results were validated using miRTarBase and Firebrowse. The results showed that the strongest IQGAP – miRNA negative correlation was dominated by IQGAP2, peaked at -0.41. HspB8, on the other hand, was found to be highly expressed in LGG. However, due to lack of researches, their roles have not been validated yet. Therefore, the results from this study might become basis for further wet-lab studies.

1. Introduction
Low grade gliomas (LGG) are diffusely infiltrating low- and intermediate-grade gliomas which include World Health Organization (WHO) grade II and III astrocytic tumors, oligodendrogliomas, and diffuse gliomas (oligodendrogliomas) [1][2]. Because of their invasiveness, LGG cannot be completely removed via surgery; the residual tumor may reoccur and become malignant, causing the patients to experience tumor-related complications, even death [2][3]. According to Packer & Stiff [3], there are approximately 2000 adults in United States who are diagnosed by LGG each year, with male to female ratio of 1.58:1 [4]. Moreover, LGG is particularly fatal for young adults, with average survival rate of 7 years [4]. Therefore, it is important to study the pathogenesis of LGG, particularly its biomarker.

Biomarker is measurable sign of biological processes that objectively describes the normal and abnormal states of an organism [5][6]. In terms of diseases, such as cancer, it differentiates the condition between sick and normal patients [7]. As biomarker is disease-specific (i.e. every biomarker is correlated with certain disease), its identification becomes an important aspect in developing personalized medicines [8]. There are various types of biomarker, including gene and microRNA (miRNA) expressions [7]. Previous studies showed that LGG is denoted by mutations on isocitrate dehydrogenase (IDH) 1 and 2 genes [9]. It is acquired in early gliomagenesis and followed by TP53 mutation or 1p/19q loss, causing astrocytic or oligodendrogial phenotype, respectively. However, the frequency of IDH1/2 mutation is yet to be defined, as a low mutation rate of IDH1 was also found among LGG patients [10].
These results show the limitation of proteomics-based studies, particularly with the emergence of transcriptomic [11]. As protein marks the end of central dogma, it undergoes many regulations. For example, there are more than 90,000 individual post-transcriptional modification (PTM) of proteins that have been identified [12]. However, these regulations do not necessarily increase the inferred information [13]. In prokaryotes, the system has become saturated; the regulatory cost becomes so high and limits further genomic and functional regulations [11]. In this case, transcriptomic may become useful; as RNA molecule has less regulatory processes than protein, it is expected to unveil novel information which is failed to be achieved by proteomics, particularly in cancer pathology.

In this study, two possible biomarkers are proposed: IQ Motif Containing GTP-ase Activating Protein (IQGAP) and Heat Shock Protein B8 (HspB8). IQGAPs are conserved family of proteins that regulate diverse cellular processes, such as cytokinesis, cell migration, cell proliferation, and cytoskeletal dynamics [14]. Previous study showed that the overexpression of IQGAPs, particularly IQGAP1, is associated with enhanced tumor proliferation, invasion, and angiogenesis [15]. In order to assess the gene regulation in LGG, transcriptomic approach should become useful. One particular way is by correlating the expression of the genes and each type of human microRNA (miRNA). MiRNA is a small non-coding RNA that generally down-regulates its target gene by binding at the 3’ untranslated region (UTR) of the gene [16]. If a strong negative correlation is obtained from the correlation analysis, there might be a direct correlation between IQGAP and its respective miRNA in LGG; this information can be utilized in further research, particularly in developing transcriptomic-based drugs.

HspB8, on the other hand, is a member of human small heat shock protein family which shares common features with HspBs, such as stress inducibility and chaperon activity [17]. Based on F-Census (http://bioinfo.hrbmu.edu.cn/fcensus/), it was found to be correlated with glioma; however, its role (i.e. oncogene or tumor suppressor gene (TSG)) has not been determined yet. Previous study showed that the expression of TSG was higher in specific tissue where the repression activity is needed [18], resulting an opposite effect for oncogene. As HspB8 was already found to be correlated with glioma, the gene’s function should be able to be determined by calculating and comparing the gene expression among LGG patients. This information can be used to assess the gene regulation and its potential as a target in further drug development.

Currently, there is a lack of researches that discuss about IQGAPs and HspB8 regulations in LGG. This in silico study is aimed to identify the expression of IQGAP with its corresponding microRNA (miRNA) and HspB8 among LGG patients based on TCGA dataset, unveiling the potential of miRNA-regulated IQGAP and HspB8 genes as biomarkers in LGG.

2. Methods

2.1. IQGAP-miRNA interaction

2.1.1. Datasets
The dataset was derived from The Cancer Genome Atlas (TCGA) projects and stored in GDC Data Portal (https://portal.gdc.cancer.gov/). In this analysis, the dataset was TCGA-LGG.

2.1.2. Data pre-processing
First, the metadata of both gene and miRNA expressions datasets were downloaded from the GDC Data Portal. The keywords for retrieving the gene expression data were: Brain (Primary Site), TCGA-LGG (Project ID), HTSeq – FPKM-UQ (Workflow Type), Transcriptome Profiling (Data Category), and Gene Expression Quantification (Data Type). The keywords for retrieving the miRNA expression data were: Brain (Primary Site), TCGA-LGG (Project ID), Transcriptome Profiling (Data Category), and miRNA Expression Quantification (Data Type). The other parameters were left unchecked. Then, the file was converted into CSV file format by using JSON to CSV Converter (https://konklone.io/json/), as a matter of preference.
Next, the metadata were imported into Python 3.6. As every patient is denoted by specific barcode (ID), the first fifteen characters of all TCGA IDs from both metadata, which were patient-specific, were extracted and added into different lists. Then, the lists were matched: IDs that appear in both lists were retrieved. As a result, a list of IDs (patients) whose both gene and miRNA expression data are available was obtained.

These IDs were then inputted into TCGA Assembler 2.0.5 [19][20], an R code implementation for downloading gene and miRNA expression data. In this case, only the first twelve characters of the ID that were inputted, as the program requires. The manual, including all modules and codes, could be downloaded from http://www.compgenome.org/TCGA-Assembler/. After the data were downloaded, they were imported into Microsoft Excel and saved in XLSX format.

2.1.3. Correlation analysis
The gene and miRNA expression files were imported into Matlab R2018a. Then, Spearman correlation test between IQGAP gene and miRNA expressions were conducted. The IQGAP – miRNA interactions with Spearman Rho (R) and significance (p) value lower than -0.2 and 0.05, respectively, were retrieved.

2.1.4. Validation
The top ten IQGAP – miRNA interactions from previous step were validated using miRTarBase. In addition, Firebrowse (http://firebrowse.org/) was utilized to support the results.

2.2. HspB8 expression
2.2.1. Datasets
The datasets were derived from The Cancer Genome Atlas (TCGA) projects and stored in GDC Data Portal. In this analysis, the datasets were TCGA-LGG (lower grade glioma), TCGA-LAML (acute myeloid leukemia), TCGA-CESC (cervical squamous cell carcinoma and endocervical adenocarcinoma), TCGA-BLCA (bladder urothelial carcinoma), TCGA-HNSC (head and neck squamous cell carcinoma), TCGA-STAD (stomach adenocarcinoma), and TCGA-BRCA (breast invasive carcinomas).

2.2.2. Data pre-processing
The HspB8 gene expression data of patients from seven TCGA projects (sub-point 2.2.1) were retrieved by using the same procedure as sub-point 2.1.2. However, the miRNA expression data of every patient was not downloaded.

2.2.3. Expression analysis
The HspB8 expression of each cancer was statistically analyzed and compared by using built-in functions in Microsoft Excel.

2.2.4. Validation
The results from previous step was compared with the data in Firebrowse.

3. Results and discussion
3.1. IQGAP – miRNA interaction
In TCGA-LGG, there are 529 gene expressions and 530 miRNA expressions data from 511 and 512 patients, respectively. After the gene and miRNA metadata were matched, 522 patient IDs (files) were found in both metadata files. The matching process is essential as the patient must have both gene and miRNA expression data to be included in the correlation analysis. These patient IDs were used as inputs in the TCGA Assembler, which later will download the gene and miRNA expressions of each patient.

Based on the correlation analysis, we found 101 significant negative correlations between miRNA and IQGAP genes. As shown in Table 1, the strongest IQGAP – miRNA correlations were dominated

---

**Figure 1.** Method pipeline. Red: IQGAP-miRNA interaction analysis; blue: HspB8 expression analysis. The pipeline was constructed by utilizing several existing tools: Python, TCGA Assembler (R Studio), Microsoft Excel, and Matlab.
by IQGAP2. However, the correlation values only peaked at -0.41, which was intermediate (i.e. the closer the value to -1, the stronger the correlation). Moreover, those interactions could not be found in miRTarBase, meaning that their relationships have not yet been experimentally validated. In fact, the list of miRNAs in table 1 were found to be correlated with various types of cancer; however, they have different target gene. For example, hsa-mir-128 inhibits the development of glioma by targeting transcription factor Bmi-1 [21] instead of IQGAP. Therefore, the author tried to search the validated IQGAP – miRNA interaction in miRTarBase and analyze its correlation within the LGG dataset.

Table 1. Top 10 IQGAP – miRNA negative correlations in LGG dataset. The table was sorted based on Spearman Rho value (from the biggest to the smallest).

| No. | IQGAP | miRNA         | Rho value | P value |
|-----|-------|---------------|-----------|---------|
| 1   | IQGAP2| hsa-mir-128-1 | -0.41     | <0.0001 |
| 2   | IQGAP1| hsa-mir-9-2   | -0.39     | <0.0001 |
| 3   | IQGAP1| hsa-mir-9-1   | -0.39     | <0.0001 |
| 4   | IQGAP1| hsa-mir-767   | -0.38     | <0.0001 |
| 5   | IQGAP2| hsa-mir-128-2 | -0.37     | <0.0001 |
| 6   | IQGAP2| hsa-mir-3943  | -0.36     | <0.0001 |
| 7   | IQGAP2| hsa-mir-137   | -0.36     | <0.0001 |
| 8   | IQGAP2| hsa-mir-490   | -0.33     | <0.0001 |
| 9   | IQGAP2| hsa-mir-1249  | -0.33     | <0.0001 |
| 10  | IQGAP2| hsa-mir-346   | -0.33     | <0.0001 |

Table 2. Validated IQGAP – miRNA interactions in miRTarBase. The table was sorted based on Spearman Rho value (from the biggest to the smallest).

| No. | IQGAP | miRNA         | Rho value | P value |
|-----|-------|---------------|-----------|---------|
| 1   | IQGAP1| hsa-miR-3074  | 0.32      | <0.0001 |
| 2   | IQGAP1| hsa-miR-215   | 0.26      | <0.0001 |
| 3   | IQGAP1| hsa-miR-324   | 0.21      | <0.0001 |
| 4   | IQGAP2| hsa-miR-548e  | 0.13      | <0.0001 |
| 5   | IQGAP2| hsa-miR-549a  | 0.10      | <0.0001 |
| 6   | IQGAP1| hsa-miR-449b  | 0.09      | <0.0001 |
| 7   | IQGAP1| hsa-miR-506   | -0.09     | <0.0001 |
| 8   | IQGAP1| hsa-mir-124-2 | -0.13     | <0.0001 |
| 9   | IQGAP1| hsa-mir-124-1 | -0.13     | <0.0001 |
| 10  | IQGAP1| hsa-mir-124-3 | -0.13     | <0.0001 |
| 11  | IQGAP3| hsa-miR-1-2   | -0.14     | <0.0001 |

In miRTarBase, there are 12 entries for human IQGAP1, more than IQGAP2 and IQGAP3 (8 and 1 entries, respectively). It is reasonable as the majority of studies focus on IQGAP1 instead of the other two [15]. Based on table 2, there are 11 interactions of IQGAP – miRNA that were found to be significant (p < 0.05) in LGG dataset. However, the correlation strength is much weaker than those in table 1. Moreover, each of them was not found to be correlated with LGG, meaning that there have been no studies that identify these interactions in LGG. As a result, the listed IQGAP – miRNA interaction could not be further analyzed.
Based on Firebrowse, the expressions of IQGAP1, IQGAP2, and IQGAP3 in LGG are the second, fourth, and fourth lowest among all types of cancer, respectively (figure 2). Previous study showed the expression of IQGAP1 is positively-correlated with the malignancy of LGG [22]; its interaction with CDC42, a Rho family member, improves human glioma cell proliferation and migration ability. Moreover, other TCGA-based study found that there is an overexpression of IQGAP2 in LGG; the fold changes between normal and cancer patients for astrocytoma, oligodendroglioma, and oligoastrocytoma were 3.55, 2.54, and 2.80, respectively [23]. However, the role IQGAP3 in LGG (i.e. overexpressed or underexpressed) has not yet been determined.

3.2. HspB8 expression
In TCGA, the number of files/patients of each project ID was different. There were 151/151, 309/304, 433/408, 546/501, 407/380, 529/511, 1222/1092 data for LAML, CESC, BLCA, HNSC, STAD, LGG, and BRCA projects, respectively. Figure 3 was constructed based on 75 LAML, 300 CESC, 400 BLCA, 400 HNSC, 400 STAD, 400 LGG, and 400 BRCA data. The lower limit of the data (75) decreased after the gene and miRNA metadata were matched, while the upper limit was set to be 400 to narrow the range of the value.

As shown in figure 3 and table 3, the expression of HspB8 in LGG are generally higher than other six types of cancer. Even though the maximum value is found in bladder cancer (36656.17), LGG has the highest average (5320.03). This result is validated by using Firebrowse (figure 4), showing that the HspB8 expression in LGG is the fifth highest among all types of cancer; it differs 0.6 and 0.3 from KIRC (Kidney Renal Clear Cell Carcinoma) in terms of first quartile and median, respectively, while the third quartile is the same. It might infer that HspB8 is overexpressed in LGG, thus making the gene a potent biomarker in LGG.

In estrogen receptor positive (ER+) breast cancer, HspB8/Hsp22 was found to be induced by estrogen, resulting an overexpression of Hsp22 protein which contributes to the progression of the cancer [24]. However, other study showed that the Hsp22 (HspB8/Hsp11) exhibited anti-proliferative property in human glioblastoma cells [25]; the knockdown of Hsp22 (HspB8/Hsp11) increased the expression of Sam68 (Src-associated protein in mitosis 68 kDa) and enhanced the proliferation of glioblastoma cell. The result of this study is contrary with the proposed hypothesis, as HspB8 turns to be a tumor suppressor gene (TSG) instead of oncogene despite of its high expression in LGG. In order to validate the HspB8 role in LGG, further study is required to be conducted.

The importance of Hsp(s) family in cancer development was described by Chatterjee & Burns [26]. In eukaryote, Hsp primarily acts as molecular chaperone, facilitating and maintaining the folding of protein. Cancer cell, on the other hand, often consists of misfolded oncoprotein, thus requiring Hsp to do the correction. As a result, Hsp is found to be highly expressed in cancer. One of the most studied Hsp is Hsp90, which is correlated with several types of cancer, such as lung [27] and medulloblastoma [28]. The inhibition of this particular Hsp produced pre-clinically promising results [26], meaning that comprehensive Hsp study, including HspB8, is essential to uncover potent therapeutic treatment for cancer.

In this study, the dataset was limited to the one that is available in TCGA. There was no available data for normal patients; the author could not compare the gene and miRNA expressions between cancer and normal patients. Moreover, the choice of cancer types in HspB8 expression analysis (figure 3) was completely random; it would be better to conduct pre-studies before choosing the cancer types to strengthen the result.

4. Conclusions
One hundred one significant miRNA-regulated IQGAP were found in LGG, with the strongest negative correlation reached intermediate value, -0.41. On the other hand, HspB8 was found to be highly expressed in LGG than other types of cancer (mean = 5320.03). These results provided basis of novel potent LGG biomarkers for further wet lab studies.
Figure 2. The expression of IQGAP genes among all types of cancer. (A) IQGAP1 (B) IQGAP2 (C) IQGAP3. The data is TCGA-based and sorted from the highest (left to right). The gene expression in LGG is denoted by green arrow. The legend of the scatter plot is located at the top-right of the graph. The boxplots were taken from Firebrowse.
**Figure 3.** The expression of HspB8 among seven types of cancer. The data was taken from TCGA. The legend of the scatter plot is located at the bottom of the graph. LAML: Acute Myeloid Leukemia; CESC: Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma; BLCA: Bladder Urothelial Carcinoma; HNSC: Head and Neck Squamous Cell Carcinoma; STAD: Stomach Adenocarcinoma; LGG: Low-Grade Glioma; BRCA: Breast Invasive Carcinoma.

**Table 3.** The statistical result of Figure 3. The data was taken from TCGA. The statistics were calculated by using Ms. Excel.

|        | BLCA     | BRCA     | CESC     | HNSC   | LAML    | LGG     | STAD    |
|--------|----------|----------|----------|--------|---------|---------|---------|
| Max.   | 36656.17 | 22274.46 | 6148.85  | 27521.56 | 9.05   | 34413.73 | 19877.17 |
| Min.   | 8.89     | 13.40    | 23.68    | 22.81  | 0       | 48.90   | 13.61   |
| Average| 1838.61  | 1306.27  | 901.23   | 2365.85 | 0.35   | 5320.03  | 1270.88 |
Figure 4. The expression of HspB8 among all types of cancer. The data is TCGA-based and sorted from the highest (left to right). The gene expression in LGG is denoted by green arrow. The legend of the scatter plot is located at the top-right of the graph. The boxplots were taken from Firebrowse.

5. Author contributions
JI prepared the Python code and executed the protocols. SSWH helped in executing the protocols. DA prepared the Matlab code and supervised the research. RN and AAP reviewed the manuscript.

6. Acknowledgement
The authors would like to thanks Research and Community Engagements Institute (LPPM) and Bioinformatics Department of i3L for funding and reviewing this research.

7. References
[1] Cancer Genome Atlas Research 2015 Network Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas N Engl J Med 372(26) 1–18
[2] Louis DN, Perry A, Reifenberger G, Deimling A Von, Figarella D, Webster B, et al. 2016 The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary Acta Neuropathol 131(6) 803–20
[3] Packer RJ, Schiff D 2012 Neuro-oncology 1st ed. (Chichester, West Sussex: Wiley-Blackwell)
[4] Claus EB, Walsh KM, Wiencke JK, Molinaro AM, Wiemels JL, Schildkraut JM, et al. 2015 Survival and low-grade glioma: the emergence of genetic information Neurosurg Focus 38(1) 1–10
[5] Strimbu K, Tavel JA 2010 What are biomarkers? Curr Opin HIV AIDS 5(6) 463–6
[6] Goossens N, Nakagawa S, Sun X, Hoshida Y 2015 Cancer biomarker discovery and validation Transl Cancer Res 4(3) 256–69
[7] Henry NL, Hayes DF 2012 Cancer biomarkers Mol Oncol 6(2) 140–6
[8] Ziegler A, Koch A, Kroekenberger K 2012 Personalized medicine using DNA biomarkers: a review Hum Genet 131(10) 1627–38
[9] Olar A, Prabhu SS 2013 Molecular Classification of Diffuse Gliomas Oncology [Internet] 27(6) Available from: http://www.cancernetwork.com/brain-tumors/molecular-classification-diffuse-gliomas
[10] Ellezam B, Theeler BJ, Walbert T, Mammoser AG, Horbinski C, Arie BKK, et al. 2012 Low rate of R132H IDH1 mutation in infratentorial and spinal cord grade II and III diffuse gliomas Acta Neuropathol 124(3) 449–51
[11] Mattick JS 2004 Legislative Activity: RNA regulation: a new genetics? Nat Rev Genet 5(4) 316–23
[12] Khoury GA, Baliban RC, Floudas CA 2011 Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database Sci Rep 1(90) 1–5
[13] Amaral P, Mattick J 2008 Noncoding RNA in development Mammalian Genome 19(7-8) 454-92
[14] Hedman AC, Smith JM, Sacks DB 2015 The biology of IQGAP proteins: beyond the cytoskeleton EMBO Rep 16(4) 1–20
[15] White CD, Brown MD, Sacks DB 2009 IQGAPs in cancer: A family of scaffold proteins underlying tumorigenesis FEBS Lett 583(12) 1817–24
[16] Kehl T, Backes C, Kern F, Fehlmann T, Ludwig N, Meese E, et al. 2017 About miRNAs, miRNA seeds, target genes and target pathways Oncotarget 8(63) 107167-75
[17] Carra S, Brunsting JF, Lambert H, Landry J, Kampinga HH 2009 HspB8 Participates in Protein Quality Control by a Non-chaperone-like Mechanism That Requires J Biol Chem 284(9) 5523–32
[18] Muir B, Nunney L 2015 The expression of tumour suppressors and proto-oncogenes in tissues susceptible to their hereditary cancers British Journal of Cancer 113(2) 345-53
[19] Wei L, Jin Z, Yang S, Xu Y, Zhu Y, Ji Y 2017 TCGA-Assembler 2: Software Pipeline for Retrieval and Processing of TCGA / CPTAC Data Bioinformatics 34(9) 1615–7
[20] Zhu Y, Qiu P, Ji Y 2014 TCGA-Assembler : open-source software for retrieving and processing TCGA data *Nature Methods* **11**(6) 599–600

[21] Li M, Fu W, Wo L, Shu X, Liu F, Li C 2013 miR-128 and its target genes in tumorigenesis and metastasis *Exp Cell Res* **319**(20) 3059–64

[22] Cui X, Song L, Bai Y, Wang Y, Wang B, Wang W 2017 Elevated IQGAP1 and CDC42 levels correlate with tumor malignancy of human glioma *Oncol Rep* **37**(2) 768–76

[23] Kumar D, Hassan K, Pattnaik N, Mohapatra N, Dixit M 2017 Reduced expression of IQGAP2 and higher expression of IQGAP3 correlates with poor prognosis in cancers *PLoS One* **12**(10) 1–23

[24] Sun X, Fontaine J, Bartl I, Behnam B, Welsh MJ, Sun X, et al. 2007 Induction of Hsp22 (HspB8) by estrogen and the metalloestrogen cadmium in estrogen receptor – positive breast cancer cells *Cell Stress Chaperones* **12**(4) 307–19

[25] Modem S, Chinnakannu K, Bai U, Reddy GP-V, Reddy TR 2011 Hsp22 (HspB8/H11) Knockdown Induces Sam68 Expression and Stimulates Proliferation of Glioblastoma Cells. *J Cell Physiol* **226**(11) 2747–51

[26] Chatterjee S 2017 Targeting Heat Shock Proteins in Cancer : A Promising Therapeutic Approach *Int J Mol Sci* **18**(9) 1–39

[27] Cohen V, Esfahani K 2016 HSP90 as a novel molecular target in non-small-cell lung cancer *Lung Cancer Targets Ther* **7** 11–7

[28] Alexiou GA, Vartholomatos G, Stefanaki K, Paterei A, Dova L, Karamoutsios A, et al. 2013 Expression of heat shock proteins in medulloblastoma *J Neurosurg Pediatr* **12**(5) 452–7