A peer-reviewed version of this preprint was published in PeerJ on 14 May 2018.

View the peer-reviewed version (peerj.com/articles/4686), which is the preferred citable publication unless you specifically need to cite this preprint.

https://doi.org/10.7717/peerj.4686
Modulation of transcriptional activity in brain lower grade glioma by alternative splicing

Jin Li¹, Yang Wang¹, Xianglian Meng¹, Hong Liang Corresp. ¹

¹ College of Automation, Harbin Engineering University, Harbin, Heilongjiang, China

Corresponding Author: Hong Liang
Email address: lh@hrbeu.edu.cn

Proteins that modify the activity of transcription factor (TF), often called modulators, play a vital role in gene transcriptional regulation. Alternative splicing is a critical step of gene processing and it can modulate gene function by adding or removing certain protein domains, and therefore influences the activity of a protein. The objective of this study is to investigate the role of alternative splicing in modulating the transcriptional regulation in brain lower grade glioma (LGG), especially transcription factor ELK1, which is closely related to various diseases, including Alzheimer’s disease and down syndrome. Results showed that changes in the exon inclusion ratio of proteins APP and STK16 are associated with changes in the expression correlation between ELK1 and its targets. Meanwhile, the structural features of the two modulators are strongly associated with the pathological impact of exon inclusion. Our analysis suggests, protein in different splicing level could play different functions on transcription factors, hence induces multiple genes dysregulation.
Modulation of transcriptional activity in brain lower grade glioma by alternative splicing

Jin Li, Yang Wang, Xianglian Meng, Hong Liang*
College of Automation, Harbin Engineering University, Harbin, Heilongjiang, 150001, China

Corresponding Author:
Hong Liang

Email address: lh@hrbeu.edu.cn

* Corresponding author
lh@hrbeu.edu.cn; Tel: (0451) 82569412; Fax: (0451) 82569412

J Li and Y Wang contributed equally to this work
Abstract

Proteins that modify the activity of transcription factor (TF), often called modulators, play a vital role in gene transcriptional regulation. Alternative splicing is a critical step of gene processing and it can modulate gene function by adding or removing certain protein domains, and therefore influences the activity of a protein. The objective of this study is to investigate the role of alternative splicing in modulating the transcriptional regulation in brain lower grade glioma (LGG), especially transcription factor ELK1, which is closely related to various diseases, including Alzheimer’s disease and down syndrome. Results showed that changes in the exon inclusion ratio of proteins APP and STK16 are associated with changes in the expression correlation between ELK1 and its targets. Meanwhile, the structural features of the two modulators are strongly associated with the pathological impact of exon inclusion. Our analysis suggests, protein in different splicing level could play different functions on transcription factors, hence induces multiple genes dysregulation.

Keywords

alternative splicing, amyloid precursor protein, EST domain-containing protein Elk-1, lower grade glioma, modulator, serine/threonine kinase 16

Introduction

Alternative splicing (AS) is a key regulator of gene expression as it generates numerous transcripts from a single protein-coding gene. In humans, over 95% of multi-exonic protein-coding genes undergo AS(Wang, Sandberg et al. 2008), and AS plays an important role in cellular differentiation and organism development(Castle, Zhang et al. 2008, Wang, Sandberg et al. 2008). As AS affects numerous genes and is highly important for regulating a given gene’s normal
expression and tissue specificity, it is not surprising that changes in AS are frequently associated with human diseases, such as cancers (Kozlovski, Siegfried et al. 2017) and neurodegenerative diseases (Scotti and Swanson 2016). Recent genome-wide analyses of cancer transcriptomes have demonstrated that splicing changes are often global rather than gene specific (Jung, Lee et al. 2015). Undoubtedly, widespread splicing changes, such as altered cassette exon inclusion ratios of proteins, influence the expression of numerous genes and consequently cause aberrant gene regulation.

Lower grade glioma (LGG) is a type of cancer that develops in the glial cells of the brain. Tumors are classified into grades I, II, III or IV based on standards set by the World Health Organization (Ostrom, Gittleman et al. 2013). Regardless of tumor grade, tumors compress normal brain tissue as they grow, frequently causing disabling or fatal effects. The Cancer Genome Atlas (TCGA) consortium has produced a comprehensive somatic landscape of glioblastoma by combining molecular and clinical data, and TCGA has become a valuable resource for studying gene deregulation in LGG.

Modulators are proteins that modify the activity of transcription factors (TFs) and influence the expression of their target genes. Our current knowledge of TF modulation mainly comes from experimental studies that measure the expression levels of a few target genes (Lachmann, Xu et al. 2010). The objective of this study is to explore the role of AS in modulating the transcriptional activities of TFs in LGG. The modulated relationships among TF-modulator-targets are inferred using a known probabilistic model, named GEM (Babur, Demir et al. 2010). EST domain-containing protein Elk-1 (ELK1) is one of the TFs whose regulation activity is most influenced by 162 splicing events, corresponding to 123 AS modulator proteins. Finally, amyloid precursor
protein (APP) and serine/threonine kinase 16 (STK16), modulators whose exon inclusion ratios are associated with the activity of ELK1, are analyzed in detail.

Materials & Methods

Construction of triplets

We implemented a known algorithm named GEM (Babur, Demir et al. 2010) to predict (splicing modulator-TF-target) triplets. There are four input types: gene expression profiles, gene splicing profiles, modulator list and transcription factor-target relations. The modulator hypothesis predicts that the correlation between the expression levels of the TF and the target must change as the splicing level of the modulator changes. The percentage of exon inclusion ratio (PSI) is used to estimate the splicing level of a candidate modulator in LGG. We established a 5% false discovery rate as the threshold to call the triplets.

Data processing and selection

RNA-Seq data were downloaded from the TCGA-LGG data portal as bam files. STAR aligner (version 2.3.0) was used to uniquely align each file uniquely to the hg19 human genome. We kept uniquely aligned reads with a minimum splice junction overhang of five nucleotides using default parameters. Gene expression level is estimated using a tool named NGSUtils (version 0.5.9) (Breese and Liu 2013) with the default parameters for calling gene expression. The splicing level (PSI) is estimated using a probabilistic model called Mixture of Isoforms (MISO) (Katz, Wang et al. 2010). The TF-target relations are derived from the ENCODE (The Encyclopedia of DNA Elements) project. The workflow of data processing and selection is described in Figure 1.

For the candidate modulators, we keep the splicing events where over 95% samples have confidence interval (CI) less than 0.25 and only analyze predicted cassette exons that have at least
10 reads supporting exon inclusion or exclusion in at least one sample. We fill the missing PSI value of a sample with the median PSI value of that splicing event. Finally, AS events were selected based on candidate modulators whose PSI IQR (interquartile range) were larger than 0.1. As the input data require sufficient variability, we filtered out genes whose gene expression coefficient variation (CV) was less than 50% and keep genes in which over 95% of samples had expression values.

Database and related software

The implementation of GEM is available through SourceForge (https://sourceforge.net/projects/modulators). Statistical analysis and data processing were performed using R version 3.0.1 (www.r-project.org). DAVID(Dennis, Sherman et al. 2003) and IPA (Ingenuity Pathway Analysis) were used to perform gene function and pathway analysis. Protein-protein interactions were predicted by the STRING database (http://string-db.org).

Results and Discussion

Global inferring modulators of all TFs

We assume that all TFs have the potential ability to interact with their modulator candidates. Seven hundred and sixty-five AS events were considered as putative modulators, and 173,598 TF-target pairs composed of 74 TFs and 17,425 targets were used to infer modulated triplets. The number of inferred splicing modulators varied across all TFs, and the percent of influenced targets ranged from 0 to 33.5% for each TF (Figure 2).

Figure 3 summarizes the number of modulators of 26 TFs whose influence targets over 10%. The number of inferred modulators ranges from 1 to 262. EST domain-containing protein Elk-1 (ELK1) was one of the 26 TFs that had the greatest number of predicted modulators. A total of
262 splicing events corresponding to 187 proteins were identified as ELK1 modulators because their splicing outcome highly correlated with changes in ELK1’s transcriptional activity.

**Gene function analysis of ELK1 modulators in LGG**

ELK1 is a member of the ETS transcription factor family, which is closely related to various diseases, including Alzheimer’s disease, down syndrome and breast cancer, in a dose-dependent manner (Peng, Yang et al. 2017). It can significantly regulate the expression of c-Fos, which is a key gene for cell proliferation and differentiation (Chambard, Lefloch et al. 2007). In this study, we inferred five hundred and forty splicing events as ELK1 modulators.


As many inferred modulators may have similar protein functions or function related, we performed pathway and function enrichment analysis to explore the functions of these modulated genes. We filtered modulators whose influenced targets less than 10% and finally 262 splicing events as modulators corresponding to 129 proteins are remained as ELK1 final modulators. After
removing duplicated gene symbols and unannotated genes, 126 proteins can be mapped to the Ingenuity Knowledge Base that are subject to core analysis.

Results show that over 80% of these splicing proteins related to cancer are enriched and most of the enriched canonical pathways are overlapped with certain genes. As Table 1 summarized, these modulators are enriched in three types of diseases, including neurological disease, organismal injury and abnormalities disease, and cancer, respectively. Molecular and cellular function enrichment analysis showed that over 20% of the modulators were associated with cellular movement (28/123), cellular assembly and organization (32/123), and cellular function and maintenance (26/123); 11% and 8% of the modulators were highly enriched in cell morphology (14/123), and cell-to-cell signaling and interaction process (10/123), respectively. Top5 modulator-enriched pathways (Figure 4B) are highly (p < 0.05) associated with signaling processes, including clathrin-mediated endocytosis signaling, CTLA4 signaling in the cytotoxic T lymphocyte pathway, nNOS signaling in neurons and calcium signaling pathways.

*APP modulates ELK1 transcriptional activity*

Amyloid precursor protein (APP) was one of modulators of interest and its analysis is described in detail here. An interaction between APP and ELK1 is mentioned in the STRING database. Several AS isoforms of APP have been observed in humans. The isoforms range in length from 639 to 770 amino acids, and certain isoforms are preferentially expressed in neurons; changes in the neuronal ratio of these isoforms have been associated with Alzheimer's disease (Matsui, Ingelsson et al. 2007).

One splicing event of APP detected as a modulator was “chr21:27354657:27354790:-@chr21:27372330:27372497:-@chr21:27394156:27394358:-”. Different inclusion ratios of the alternatively spliced exon in APP protein influence 18.6% of the targets of ELK1, and the 7th exon,
which contains a vital domain named BPT/Kunitz inhibitor (BPTI) (residues 291-341), is the alternatively spliced exon. The splice isoforms that contain the BPTI domain possess protease inhibitor activity.

According to GEM algorithm, unmodulated ELK1 activity was classified into three categories according to \( \alpha_f \): activation if positive, inhibition if negative, and inactive if zero. Similarly, by comparing \( \alpha \) and \( \beta \) coefficients, modulators are classified into three classes: enhance, attenuate or invert the activity of the ELK1. Hence, there are six possible categories of action. The APP modulation categories and their interpretations are listed in supplement file Table S1.

As Table 2 summarized, without APP modulation of APP, unmodulated ELK1 inhibits 172 targets, and activates 31 targets. However, when APP interacts with ELK1 as a modulator, the original transcriptional activity of ELK1 becomes different: APP attenuates ELK1 inhibition roles on 164 targets, inverts inhibition activity on 8 targets, and enhances activation on 14 targets. Meanwhile, APP also inverts or attenuate ELK1 activity on 31 targets, including 1 targets activity is inverted and 30 targets activity is attenuated.

We randomly selected four targets (ANKRD34A, DDX27, DVL3 and HEATR1) of ELK1 to explore the different activities of ELK1 under the modulation of differential inclusion levels of APP protein. Ideally, the inclusion level of the splicing modulator and expression of ELK1 should have high variance and low correlation in the samples. We divided rank-ordered PSI values of APP splicing modulators, extracting ELK1 and its target samples that were consistent with APP splicing modulator samples in upper/lower tertile, and estimated the differences in correlation between ELK1 and its target using Spearman’s correlation.

Figure 5A shows the examples of APP-modulated ELK1 target genes and the corresponding action modes. As shown in Figure 5A, when the exon inclusion level of APP was in lower tertile,
an increase in the gene expression level of ELK1 resulted in a significant increase in the gene expression of its target ANKRD34A. Spearman’s correlation of gene expression between ELK1 and ANKRD34A was 0.71 (p < 2.2e-16), which means that in this condition, ELK1 plays an enhancement role on its target. However, when the PSI value of the APP modulator is in upper tertile, the correlation decreased into 0.30 (p = 0.0085). For the other two targets, DDX27 and DVL3, the correlations changed from -0.47 and -0.53, respectively, to non-significant (p > 0.1).

For these three cases, the APP modulator attenuates the activation of ELK1. The opposite modulation occurs on target HEATR1. When the exon is spliced out of the protein, ELK1 negative regulates the expression of HEATR1 with a correlation as -0.38 (p = 0.0005); however, when the exon is excluded from the mature mRNA, the APP modulator inverts the activation of ELK1 on its target with a correlation of 0.70 (p < 2.2-16).

We evaluated the exon’s impact on APP protein using ExonImpact (Li, Feng et al. 2017). The results showed that the alternatively spliced exons of APP protein that we detected have a high probability (0.57 and 0.48) of being associated with disease. This result indicates that changes in the inclusion or exclusion level of spliced exons can lead to significant changes with respect to APP protein function.

Figure 5B visualized the global effect of changing the inclusion ratios of alternately spliced exons in APP and influences on the relationship between ELK1 and its target. The two groups of samples are selected based on the 7th exon inclusion ratio of APP. The high and low inclusion groups contain samples with the top and bottom 30% of PSI values. The correlation patterns between ELK1 and its targets in the two groups are different, clearly showing that different splicing levels of APP can modulate the transcriptional activity of ELK1.

STK16 modulates ELK1 transcriptional activity
The AS event for protein serine/threonine kinase 16 (STK16) is another interesting modulator that we identified. Inferred STK16 modulated triplets and their modulation categories are listed in Table S2. STK16 is a membrane-associated protein kinase that phosphorylates on serine and threonine residues. An interaction between STK16 and ELK1 is inferred from the biochemical effect of one protein upon another in the BioGrid database. The alternatively spliced exon that acts as a modulator of STK16 is the 4th exon and it locates in a region that encodes for a kinase domain named Pkinase that is associated with the protein’s proton acceptor.

Figure 6A shows the modulating effect of STK16 on ELK1, and TMEM60 is one of targets we randomly detected. The samples in the two groups are selected using the same method mentioned above. A negative correlation (-0.37, p = 0.0004) is only shown when the exon is included in the final product. The exon’s impact in protein function analysis (Li, Feng et al. 2017) shows that this alternatively spliced exon has a high disease probability of 0.67, which indicates that changes in the exon inclusion or exclusion ratio might cause a gain or loss in protein function. We know that the specific alternatively spliced exon of STK16 encodes a kinase domain, so it is not surprising that the loss of this exon will cause a change in protein function and may ultimately influence numerous normal gene functions.

Figure 6B shows the global modulating effect of STK16 on ELK1. The low and high inclusion groups contain samples with the top and bottom 30% of PSI values, which indicate exon exclusion and inclusion in the final protein. A positive correlation between ELK1 and its targets is clearly shown when the exon is excluded, whereas this correlation becomes negative when the exon is included. This result suggests that the 4th cassette exon in STK16 is important to final protein
function and that changes in the splicing level of STK16 are associated with the differential transcriptional activity of ELK1.

**Conclusions**

We globally dissected the role of AS in regulating the transcriptional activity of TFs in LGG using TCGA-LGG data. ELK1 was one of TFs with the greatest number of inferred modulators, e.g., APP and STK16. The results show that changes in the AS of APP and STK16 proteins are associated with changes in the transcriptional activity of ELK1 and that the structural features of the two proteins are strongly associated with the pathological impact of exon inclusion. The presented results provide important insights on the modulating role of AS on transcription regulation in LGG.
Acknowledgments

We appreciate detailed suggestions from anonymous reviewers who significantly helped us improve the early version of this manuscript.

Competing interests

The authors declare that they have no competing interests.
References


Additional files

Additional file 1: Table S1. Inferred APP modulated triplets and their modulation categories.

Additional file 2: Table S2. Inferred STK16 modulated triplets and their modulation categories.
Table 1 (on next page)

ELK1 modulators protein function and disease enrichment (p<0.001)

Each number in the table indicates the account of ELK1 modulators enriched in specific function or disease. The statistical threshold is p<0.001.
### Molecular and Cellular Functions

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular assembly and organization</td>
<td>32</td>
</tr>
<tr>
<td>Cellular function and maintenance</td>
<td>26</td>
</tr>
<tr>
<td>Cell-to-cell signaling and interaction</td>
<td>10</td>
</tr>
<tr>
<td>Cellular movement</td>
<td>28</td>
</tr>
<tr>
<td>Cell morphology</td>
<td>14</td>
</tr>
</tbody>
</table>

### Diseases and Disorders

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurological disease</td>
<td>37</td>
</tr>
<tr>
<td>Organismal injury and abnormalities disease</td>
<td>113</td>
</tr>
<tr>
<td>Cancer</td>
<td>110</td>
</tr>
</tbody>
</table>
Table 2 (on next page)

Interpretation of the categories of APP modulation, and the inequality constraints that the category should satisfy

Each number in the table indicates the number of triplets in each classification. ‘+’ and ‘-’ signs in the columns indicate significantly positive and negative values, respectively.
<table>
<thead>
<tr>
<th>Modulation classification</th>
<th>Explanation</th>
<th># triplets</th>
<th>γ</th>
<th>α (_f)</th>
<th>β (_f)</th>
<th>β (_m)</th>
<th>α (_f) + β (_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attenuates inhibition</td>
<td>F, alone, inhibits T – M attenuates F activity</td>
<td>164</td>
<td>+</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enhances inhibition</td>
<td>Modulated F inhibits T</td>
<td>0</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inverts inhibition</td>
<td>F, alone, inhibits T - M inverts F activity</td>
<td>8</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Inverts activation</td>
<td>F, alone, activates T – M inverts F activity</td>
<td>1</td>
<td>-</td>
<td>+</td>
<td></td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enhances activation</td>
<td>Modulated F activates T</td>
<td>14</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Attenuates activation</td>
<td>F, alone, activates T – M attenuates F activity</td>
<td>30</td>
<td>-</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

'+' and '-' signs in the columns indicate significantly positive and negative values, respectively.
Figure 1 (on next page)

Workflow for data processing and selection.

The whole workflow including three parts: obtain the transcriptional profile, expression and splicing calling and construct the modulated triplets.
The effect of transcription factors activity regulated by splice modulator proteins.

Each row represents a candidate modulator and each column indicates a transcription factor. The color much darker means a much higher percent targets of TF is influenced.
Figure 3 (on next page)

Summarized counts of inferred modulators of TFs.

The c-axis represents the transcription factor list, and the y-axis represents the counts of inferred modulators. The number on each TF indicates the number of modulators of each TF that influence more than 10% of its targets.
Number of modulators influence targets over 10%

<table>
<thead>
<tr>
<th>Transcription factors</th>
<th>Number of modulators</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2F4</td>
<td>2</td>
</tr>
<tr>
<td>GABPA</td>
<td>2</td>
</tr>
<tr>
<td>ZNF143</td>
<td>15</td>
</tr>
<tr>
<td>ELK1</td>
<td>262</td>
</tr>
<tr>
<td>SPI1</td>
<td>2</td>
</tr>
<tr>
<td>TFAP2C</td>
<td>13</td>
</tr>
<tr>
<td>BHLHE40</td>
<td>16</td>
</tr>
<tr>
<td>TCF12</td>
<td>5</td>
</tr>
<tr>
<td>PBX3</td>
<td>1</td>
</tr>
<tr>
<td>RFX5</td>
<td>2</td>
</tr>
<tr>
<td>ZEB1</td>
<td>22</td>
</tr>
<tr>
<td>GTF3C2</td>
<td>19</td>
</tr>
<tr>
<td>ZNF217</td>
<td>1</td>
</tr>
</tbody>
</table>

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.26482v1 | CC BY 4.0 Open Access | rec: 2 Feb 2018, publ: 2 Feb 2018
Statistical analysis of the modulators of ELK1.

(A) Distribution of the number of ELK1 modulators. The x-axis represents the percentage of targets influenced by modulator proteins. The y-axis indicates the percent of modulators of ELK1. The number noted on each column indicates the percent of modulators in each classification. (B) IPA of ELK1 modulators that influenced over 10% of its targets. The x-axis is the -log10 transformed p of each enriched pathway (y-axis).
(A) Percent of splicing modulators (%)

- >30%: 0.93%
- 20%-30%: 8.15%
- 15%-20%: 10.37%
- 10%-15%: 10.56%
- 5%-10%: 18.52%
- <5%: 51.48%

Percent of targets influenced by splicing modulators

(B) Enriched pathways

- Glutamate Receptor Signaling: -log10(p-value) = 1.34
- Inositol Pyrophosphates Biosynthesis: -log10(p-value) = 1.38
- nNOS Signaling in Neurons: -log10(p-value) = 1.49
- Virus Entry via Endocytic Pathways: -log10(p-value) = 1.61
- Calcium Signaling: -log10(p-value) = 1.69
- Epithelial Adherens Junction Signaling: -log10(p-value) = 1.94
- Lipid Antigen Presentation by CD1: -log10(p-value) = 2.03
- CTLA4 Signaling in Cytotoxic T Lymphocytes: -log10(p-value) = 2.52
- Clathrin-mediated Endocytosis Signaling: -log10(p-value) = 5.5
Figure 5 (on next page)

APP is a modulator that influences the activity of ELK1 in LGG.

(A) Examples of different correlations between ELK1 and its targets under the modulation of APP with differential splicing levels. (B) Visualization of how APP regulates the stability of ELK1 protein. Gene expression profiles are displayed with genes in rows and samples in columns. Expression values of each gene are rank transformed, median centered and rescaled between [-0.5, 0.5]. Samples were partitioned based on the alternatively spliced exon inclusion level of APP and sorted by the expression levels of ELK1 within each partition.
Splicing PSI low
ELK1

chr21:27394156:27394358: -
@chr21:27372330:27372497: -
@chr21:27354657:27354790: -

ANKRD3A
DDX27
DVL3
HEATR1

Splicing PSI high
ELK1

chr21:27394156:27394358: -
@chr21:27372330:27372497: -
@chr21:27354657:27354790: -

ANKRD3A
DDX27
DVL3
HEATR1
STK16 is a modulator that affects the transcriptional activity of ELK1 in LGG.

(A) Examples of differential regulation activities of ELK1 on its target under the modulation of STK16 with differential splicing levels. The spliced exon is excluded in the final production of STK16 in the first scenario, and the exon is included in the final production of STK16 in the second scenario. (B) STK16 regulates the stability of ELK1. See Figure 3B for interpretation of this graph.
(A) r = 0.16, p-value = 0.21

(B) Splicing PSI low

chr2:220111379:220111598:+
@chr2:220111835:220111968:+
@chr2:220112137:220112257:+

Splicing PSI high

r = -0.37, p-value = 0.0004