

RESEARCH ARTICLE

Open Access

# Identification and characterization of alternative exon usage linked glioblastoma multiforme survival

Ahmed Sadeque<sup>1</sup>, Nicola VL Serão<sup>1</sup>, Bruce R Southey<sup>1</sup>, Kristin R Delfino<sup>1</sup> and Sandra L Rodriguez-Zas<sup>1,2,3\*</sup>

## Abstract

**Background:** Alternative exon usage (AEU) is an important component of gene regulation. Exon expression platforms allow the detection of associations between AEU and phenotypes such as cancer. Numerous studies have identified associations between gene expression and the brain cancer glioblastoma multiforme (GBM). The few consistent gene expression biomarkers of GBM that have been reported may be due to the limited consideration of AEU and the analytical approaches used. The objectives of this study were to develop a model that accounts for the variations in expression present between the exons within a gene and to identify AEU biomarkers of GBM survival.

**Methods:** The expression of exons corresponding to 25,403 genes was related to the survival of 250 individuals diagnosed with GBM in a training data set. Genes exhibiting AEU in the training data set were confirmed in an independent validation data set of 78 patients. A hierarchical mixed model that allows the consideration of covariation between exons within a gene and of the effect of the epidemiological characteristics of the patients was developed to identify associations between exon expression and patient survival. This general model describes all three possible scenarios: multi-exon genes with and without AEU, and single-exon genes.

**Results:** AEU associated with GBM survival was identified on 2477 genes (P-value < 5.0E-04 or FDR-adjusted P-value < 0.05). G-protein coupled receptor 98 (*Gpr98*) and epidermal growth factor (*Egf*) were among the genes exhibiting AEU with 30 and 9 exons associated with GBM survival, respectively. Pathways enriched among the AEU genes included focal adhesion, ECM-receptor interaction, ABC transporters and pathways in cancer. In addition, 24 multi-exon genes without AEU and 8 single-exon genes were associated with GBM survival (FDR-adjusted P-value < 0.05).

**Conclusions:** The inferred patterns of AEU were consistent with *in silico* AS models. The hierarchical model used offered a flexible and simple way to interpret and identify associations between survival that accommodates multi-exon genes with or without AEU and single exon genes. Our results indicate that differential expression of AEU could be used as biomarker for GBM and potentially other cancers.

**Keywords:** Alternative Splicing (AS), Alternative Exon Usage (AEU), Glioblastoma Multiforme (GBM), Hierarchical model

\* Correspondence: rodrgz@illinois.edu

<sup>1</sup>Department of Animal Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA

<sup>2</sup>Department of Statistics, University of Illinois Urbana-Champaign, Urbana, IL, USA

Full list of author information is available at the end of the article

## Background

Alternative splicing (AS) is characterized by the formation of different mRNA isoforms as a result of including or excluding different exonic or intronic segments. This process is responsible for generating protein diversity from a finite number of genes [1-3]. Alternative splicing can be divided into three broad categories; intron retention, cryptic splice-site usage, and alternative exon usage (AEU) or exon skipping. Alternative exon usage includes cassette exons, which are discrete exons that can be independently included or excluded, and mutually exclusive splicing, which involves the selection of one from a group of exon variants [2]. Approximately 75% of multi-exon genes exhibit AS in humans [4]. The human genome includes approximately 28,526 annotated genes that express approximately 120,145 transcripts, of which 80,932 are protein coding and 39,213 are non-coding transcripts [5]. The identification of "exon-level" expression profiles and characterization of AS events has become possible with the availability of exon platforms (e.g. GeneChip Exon Array). The brain exhibits particularly high rates of AS [6] and the highest number of AEU events [7]. Regulation of gene expression due to splicing has been associated with cancer. Many AEU events have been associated with disordered cell differentiation and signaling that contribute to stem cell like proliferation of cancer cells [8].

Glioblastoma multiforme (GBM) is an aggressive type of brain cancer and the role of genes and AEU on GBM survival is still not completely understood [9-11]. Most work on AS and GBM studied individual genes or compared AS between GBM and control (e.g. blood) samples. The relationship between AS and the survival of individuals diagnosed with GBM has not been studied. Understanding of the factors influencing survival is particularly important in GBM cases because the median survival after diagnosis is approximately one year [12,13]. Furthermore, several epidemiological factors influence GBM survival including gender, race and treatment [14]. Thus, a more accurate understanding of the relationship between AS and GBM survival must consider epidemiological factors and inter-individual variability.

Several approaches to identify AS events have been proposed. However, most approaches have limitations that can bias the identification and characterization of AEU. For example, Su et al. developed an individual exon approach that does not model the covariation between exons within a gene [15]. Purdom et al. used the residuals from probe level analysis to identify AEU on a per-sample level [16]. The sample-level analysis challenges the detection of AEU events or the identification of common patterns across patients receiving the same treatment or from the same epidemiological strata.

Laderas et al. and Zheng et al. proposed group comparison using linear models to overcome the limitations of the previous approach [2,17]. However, group comparison is not suited to identify AEU associated with other conditions such as survival or time-to-event. In addition, the previous implementation does not account for correlations between exons measured on the same sample. Cline et al. formulated an ANalysis Of Splice VARIation approach that cannot be used in genes that produced more than one splice form [18].

The main goal of this study is to demonstrate an exon-based, gene-centric model to detect AEU events associated with GBM survival. We developed an analytical method that addresses the limitations of previous approaches by modeling the exon-level expression profiles within genes from all samples across all treatments or conditions studied. Our approach accommodates the dependencies between exons within a gene and patient and allows testing the hypothesis of differential exon expression or usage between treatment groups. A unique advantage of our flexible approach is that one model encompasses all scenarios: i) multi-exon genes that have AEU, ii) multi-exon genes that do not have AEU, and iii) single-exon genes. A supporting goal is the three-fold assessment of the approach that encompassed; 1) the use separate training and validation data sets, 2) gene set enrichment and gene functional analyses of the results, and 3) comparison of predicted and reported AS events.

## Methods

### Training data set

Survival, clinical and exon expression information from 250 patients diagnosed with GBM was obtained from The Cancer Genome Atlas repository, May 2011 data freeze (<https://tcga-data.nci.nih.gov/tcga/>). Surgical samples had a minimum of 80% tumor nuclei and maximum of 50% necrosis. The clinical or epidemiological variables considered in the analysis of exon expression included treatment (levels: chemo-radiation-targeted [CRT], chemo-radiation-non targeted [CRnT], radiation [R], other therapies [OTHER], and no therapy [NONE]); racial ethnicity (white Caucasian and others); and gender (male and female). These clinical factors were previously found to be associated with survival [9]. The survival indicator was the time from diagnosis to death, expressed in months.

Exon expression measurements from a frozen sample from each patient were obtained using the Affymetrix-GeneChip<sup>®</sup> Human Exon 1.0 ST Array (Affymetrix, Santa Clara, CA). This platform includes information from 1,432,143 probe sets representing known and predicted exons on both strands of the genome that have been mapped to more than 25,000 genes. Intensity data was log<sub>2</sub> transformed and normalized using quantile

and RMA normalization at the probe level following the procedures described in Beehive (<http://stagbeetle.animal.uiuc.edu/Beehive>). Probes sets within exons were collapsed using a Tukey biweight function that provides an iterative reweighted measure of central tendency. This robust statistic provides a single exon expression that is not heavily influenced by extreme probe expression levels [19].

### Model

Three specifications of this model accommodated three groups of genes: 1) multi-exon genes exhibiting AEU, 2) multi-exon genes with no evidence of AEU, and 3) single-exon genes. The first, second and third model specifications correspond to equations [[1]], [[2]] and [[3]], respectively. Within gene, a general exon expression mixed model was developed to describe the association between and exon expression and GBM survival adjusted for other clinical factors:

$$y_{ijklmn} = \mu + G_i + R_j + T_k + b_2S_l + X_m + b_{3m}(SX)_{lm} + P_n + e_{ijklmn} \quad (1)$$

$$y_{ijklmn} = \mu + G_i + R_j + T_k + b_4S_l + X_m + P_n + e_{ijklmn} \quad (2)$$

$$y_{ijklmn} = \mu + G_i + R_j + T_k + b_5S_l + P_n + e_{ijklmn} \quad (3)$$

where  $y_{ijklmn}$  denotes the expression of the  $m^{\text{th}}$  exon ( $X_m$ ), recorded on the  $n^{\text{th}}$  patient ( $P_n$ ) that has the  $i^{\text{th}}$  gender ( $G_i$ ),  $j^{\text{th}}$  race/ethnicity ( $R_j$ ), and received the  $k^{\text{th}}$  therapy ( $T_k$ ). Survival after diagnosis ( $S_l$ ), expressed in months, was fitted as a covariate. The parameters  $b_2$ ,  $b_4$ , and  $b_5$  describe the overall change in expression per additional survival month across all exons, and  $b_{3m}$  describe the deviation in change in expression per additional survival month for the  $m^{\text{th}}$  exon in multi-exon genes that have AEU. In addition,  $e_{ijklmn}$  is the residual associated with the  $y_{ijklmn}$  observation, and  $SX$  denotes the interaction between survival and exon. In this model the fixed effects were; gender, race/ethnicity, therapy, and survival after diagnosis. The random effects exon, interaction between survival and exon, and patient were assumed to be independent and follow a Gaussian distribution with mean zero and its own variance.

A significant interaction between survival and exon effect constitutes evidence of an AEU scenario and, thus, differential survival across exons (group 1 genes). This model can be used to identify AS biomarkers of GBM survival that exhibit AEU. A significant survival after diagnosis effect together with a non-significant interaction between survival and exon effect constitutes evidence of a general association between gene expression and survival, regardless of exon (group 2 genes). This

result can be used to identify multi-exon biomarkers of GBM survival that do not exhibit AEU.

The specification for the single-exon genes (group 3 genes) is a reduced version of the full multi-exon model that excludes exon and interaction between survival and exon. A significant survival effect is evidence of association between the single-exon gene expression and survival and can be used to identify single-exon biomarkers of GBM survival.

The hierarchical structure of the model used to identify AEU and gene expression associations with survival stems from the presence of two type of descriptive parameters: general or population-level and group-level parameters. There is a general or population-level association between exon expression and survival across all the exons of the gene, and group-level exon-specific deviations from the overall association that reveal alternative exon usage. A second hierarchical structure stems from a population-level exon expression, and patient-specific deviations from the overall expression level. The within-gene analysis supported a gene-centric strategy to uncover expression profiles associated with survival. The analysis of all exon information within a gene-allowed accounting for the covariance between exon expression within a gene and the hierarchical nature of the model allows the inclusion of the covariance between exon-expression within a patient. The analysis of expression data at the exon level permitted the identification of AEU by testing the null hypothesis of no differential association between expression and survival across exons within a gene.

False Discovery Rate adjustment (FDR) of the P-values allowed controlling for multiple testing [20]. In addition, a more stringent P-value threshold was considered for the detection of AEU associations with a significant interaction between survival and exon) than the main survival effect in the multi-exon scenarios. The more stringent P-values required for detection of AEU accounted for the multiple comparisons of the survival-expression associations among potentially numerous exons. A separate FDR adjustment of the P-values from the single-exon analysis was implemented because of the different number of parameters between the multi- and single-exon models. In this study, the significance threshold P-value  $< 5.0E-4$  corresponds to a FDR-adjusted P-value  $< 0.05$  or multi exon genes and to a FDR adjusted P-value  $< 0.1$  for genes with single exon. The mixed effects model was evaluated in a restricted maximum-likelihood framework using the SAS 9.2 MIXED procedure (<http://www.sas.com/>).

Thus, three types of evidence were used to identify AEU: a) significant variations in the associations between exon expression and survival across a gene, b) consistent (over or under-expressed) differential expression in more

than two consecutive exons, and c) a minimum exon differential expression ( $< 0.995$  or  $> 1.005$  fold change / additional survival month). Consistent patterns of expression across consecutive exons were identified using a moving average analysis [21]. A moving average analysis that computes the average expression across multiple exons at a time was used to predict a continuous trajectory of exon expression across the gene. This moving average trend of exon expression across the gene facilitated the identification of consistent changes in the pattern of over or under-expression across the exons within a gene.

Functional and pathway analyses of the genes exhibiting significant evidence ( $P$ -value  $< 5.0E-4$ ) of AEU associated with GBM survival used hypergeometric tests and was implemented in DAVID [22,23]. Gene set enrichment analysis (GSEA) of the association between expression and GBM survival among all the genes studied in the platform followed the approach described by Subramanian et al. [24] implemented in BABELOMICS 4.3 [25]. For this analysis, the association between each gene and survival was characterized by the estimate of change in expression per additional survival month standardized by the standard error of the estimate. The enrichment of Gene Ontology (GO; <http://www.geneontology.org/>) biological processes, molecular functions, and KEGG (<http://www.genome.jp/kegg/pathway.html>) pathways was investigated. Finally,  $P$ -values of the enriched categories were adjusted for multiple testing using the FDR correction.

The exon expression estimates and the moving average trajectory of the estimates across individual genes were aligned to known or predicted alternative transcript variants reported in the AceView database (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>) that are available through the UCSC Genome Browser (<http://genome.ucsc.edu>). This visualization strategy facilitated the interpretation of results and the AS models offered an independent *in silico* confirmation of the AEU events identified.

#### Validation dataset

Genes exhibiting AEU in the training data set were confirmed in an independent set of 78 patients obtained from TCGA (May 2011 data freeze). The reliability of the exon expression profiles associated with survival identified in the training data set was assessed using a two-stage approach. First, the parameter estimates (i.e. changes in exon expression per one additional month of survival) that were obtained from the analysis of the training data set were applied to the covariate information from the patients in the validation data set, and predictions of exon expression were obtained. Second, the predicted exon expression values were compared to the

corresponding observed expression values. The performance of the training estimates was evaluated using the coefficient of determination ( $R^2$ ) that represents the fraction of the total variation of the expression associated with survival, exon and the rest of the model terms [26]. High  $R^2$  on the validation data set based on the training data set estimates indicate the reliability of the exon expression patterns identified.

#### Results and discussion

Expression measurements of 269,951 exons from 25,403 genes were analyzed. Of these, 2,857, 20,288, 1,965 and 293 genes had 1, 2 to 24, 25 to 49, and 50 or more exons, respectively. The number of exons per gene ranged from 1 to 191 and averaged 10.75 exons per gene. Table 1 summarizes the distribution of the 250 and 78 individuals diagnosed with GBM analyzed in the training and validation data sets respectively, across clinical factors, and survival descriptive statistics. The distribution of observations across clinical factors was consistent across data sets.

#### Multi-exon genes exhibiting exon-dependent association with glioblastoma multiforme survival

At a FDR-adjusted  $P$ -value  $< 0.05$  (approximately equivalent to an unadjusted  $P$ -value  $< 5.0E-4$ ) threshold, 2477 multi-exon genes exhibited AEU associated with survival (group 1 genes), 24 multi-exon genes exhibited expression associated with survival albeit no evidence of AEU (group 2 genes), and 8 single-exon genes exhibited expression associated with survival (group 3 genes). A similar number (1,478) of differentially expressed genes was associated short- versus long-term glioblastoma survival [27]. The higher number of associations detected in the present study could be attributed to the analysis of exon-level profiles instead of gene-level profiles. At unadjusted  $P$ -value  $< 1.0E-5$ ,  $P$ -value  $< 1.0E-6$  (approximately equivalent to a FDR-adjusted  $P$ -value  $< 1.0E-2$ ),  $P$ -value  $< 1.0E-7$ ,  $P$ -value  $< 1.0E-8$  (approximately equivalent to a FDR-adjusted  $P$ -value  $< 5.0E-4$ ), the number of genes exhibiting AEU (group 1 genes) were 592, 313, 201, and 129 respectively.

Table 2 summarizes the 36 multi-exon genes that have the most significant ( $P$ -value  $< 0.11$ ) AEU or exon-dependent association with GBM survival (group 1 genes). Additional file 1: Table S1 lists the results for the 129 multi-exon genes that exhibit evidence of AEU at  $P$ -value  $< 1.0E-8$ .

The nature of the association between expression and GBM was characterized by the sign and value of the expression change per additional month in survival. Tables 2, 3 and 4 include a general, exon-independent, gene-wise estimate of expression fold change per additional survival month for completeness. The meaning of

**Table 1 Distribution of patients across clinical factors by data set**

	Training data set		Validation data set		
	Number	Percentage	Number	Percentage	
Patients	250	76.22	78	23.78	
Race <sup>1</sup>					
	Caucasian	222	88.80	71	91.03
	Other	28	11.20	07	8.97
Gender					
	Females	94	37.60	29	37.18
	Males	156	62.40	49	62.82
Therapy <sup>2</sup>					
	R	63	25.20	21	26.92
	CRT	27	10.80	07	8.97
	CRnT	99	39.60	31	39.74
	OTHER	35	14.00	10	12.82
	NONE	26	10.40	09	11.54
Survival (months)	17.46	0.16 - 128	15.02	0.10 - 77.57	

<sup>1</sup>Race: White Caucasian, and all other race-ethnicity groups.

<sup>2</sup>Therapy; R: radiation therapy alone; CRnT: chemotherapy plus radiation and no targeted therapy; CRT: chemotherapy plus radiation and targeted therapy; OTHER: all other therapies; None: no therapy.

this fold change estimate is straightforward for genes in groups 2 and 3 because these genes exhibit a single, general, and exon-independent association with GBM survival. The general fold change estimate for group 1 genes must be considered in the context that these genes have an exon-dependent association with GBM.

The top 36 genes exhibiting significant evidence of AEU had a minimum of 90 exons (Table 2). This result suggests that genes with high number of exons are more likely to experience AEU events that influence GBM survival than genes with few exons. It is unlikely that high number of exons biased the identification of AEU due to the stringent P-value threshold used.

Most of the 36 genes that had significant AEU association with GBM survival have been associated with cancer. Relevant literature references are summarized in Table 2. There were 10 genes including titin (*Ttn*), polycystic kidney disease 1 (*Pkd1*), spectrin repeat containing, nuclear envelope 1 (*Syne1*), small nuclear ribonucleoprotein (*Snrpn*), phosphodiesterase 4D interacting protein (*Pde4dip*), obscurin (*Obscn*), dystonin (*Dst*), microtubule-actin cross-linking factor 1 (*Macf1*), ryanodine receptor 1 (*Ryr1*), and ryanodine receptor 2 (*Ryr2*), that had been previously associated with GBM. Additionally, 13 genes have been previously associated to cancers other than GBM including; Smg-1 homolog (*Smg1*), Nebulin (*Neb*), TBC1 domain family, member 3 (*Tbc1d3*), Anaphase promoting complex subunit 1 (*Anapc1*), Spectrin repeat containing, nuclear envelope 1 (*Syne2*), Neuroblastoma breakpoint family, member 10 (*Nbpf10*), Mucin 19 (*Muc19*), Collagen, type VII, alpha 1 (*Col7a1*), Ubiquitin protein ligase E3 component n-recogin 4 (*Ubr4*), Hemicentin 1 (*Hmcn1*), Collagen, type IV, alpha 5 (*Col4a5*), Ryanodine receptor 3 (*Ryr3*), and G protein-coupled receptor 98 (*Gpr98*).

Previous reports confirmed the overall relationship between the top genes exhibiting AEU and GBM identified in this study. Reports on the overall relationship between genes and GBM are hereby reviewed since the AEU pattern for these genes has not been previously described. The TTN protein is encoded by *Ttn*, and is responsible for the passive elasticity of cells. A mutation resulting in an altered TTN was associated with GBM [28]. *Pkd1* was over-expressed during the progression of low-grade to high-grade gliomas [32]. *Syne1* has been associated with increased GBM survival [9]. Under-expression of *Snrnp* was observed in older GBM patients compared to younger patients [33]. *Pde4dip* is down-regulated in glioma cell lines treated with dB-cAMP that reduces the invasiveness, proliferation and migratory properties of glioma cells and increases the survival of glioma cell lines compared to untreated cell lines [35]. The mutation R4558H in *Obscn* has been associated with GBM [28]. Likewise, a mutation in *Dst* that indirectly regulates the expression of *Otub1* (through regulation of mir-15b has been associated with GBM [44]. Reduced expression of *Macf1* has been observed in glioma cells treated with IL-13 cytotoxin that causes the cells to undergo necrosis. Thus, down-regulation of the expression of *Macf1* was associated with increased GBM survival [53]. *Ryr1* was under-expressed in high-grade gliomas relative to primary (low-grade) gliomas [57]. On the other hand, *Ryr2* was over-expressed in invasive GBM cells compared to normal control cells [51].

#### Functional and pathway analyses of the multi-exon genes exhibiting exon-independent association with glioblastoma multiforme survival

The 2,477 genes exhibiting significant evidence of AEU associated with GBM survival were further investigated

**Table 2 Top 36 multi-exon genes that have significant alternative exon usage associated with glioblastoma multiforme survival**

Gene Symbol	Estimate <sup>1</sup>	SE <sup>2</sup>	P-value AEU <sup>3</sup>	Fold change <sup>4</sup>	Exon Count <sup>5</sup>	Literature <sup>6</sup>
<i>Ttn</i>	0.0007	0.0001	4.2E-38	0.9993	340	[28] <sup>G</sup>
<i>Smg1</i>	0.0017	0.0002	2.0E-24	1.0001	209	[29] <sup>AS</sup>
<i>Neb</i>	0.0007	0.0001	3.2E-21	0.9973	180	[30,31] <sup>G, AS</sup>
<i>Pkd1</i>	0.0010	0.0001	2.0E-19	1.0018	163	[32] <sup>G, AS</sup>
<i>Herc2p2</i>	0.0008	0.0001	2.3E-19	1.0012	163	NA
<i>Syne1</i>	0.0018	0.0002	3.0E-18	0.9984	152	[9] <sup>G</sup>
<i>Snrpn</i>	0.0018	0.0002	3.8E-18	1.0020	151	[33,34] <sup>G, AS</sup>
<i>Pde4dip</i>	0.0016	0.0002	1.3E-17	0.9993	146	[35,36] <sup>G, AS</sup>
<i>Golga8c</i>	0.0031	0.0004	4.2E-17	1.0005	141	NA
<i>Sspo</i>	0.0009	0.0001	1.2E-16	1.0003	137	NA
<i>Ankrd36</i>	0.0026	0.0003	1.3E-16	1.0018	137	NA
<i>Tbc1d3</i>	0.0008	0.0001	2.4E-16	1.0026	135	[37] <sup>C</sup>
<i>Flij45340</i>	0.0018	0.0002	5.5E-16	1.0007	131	NA
<i>Anapc1</i>	0.0009	0.0001	5.8E-15	0.9990	122	[38,39] <sup>C, AS</sup>
<i>Syne2</i>	0.0012	0.0002	6.2E-15	1.0017	115	[40] <sup>C, AS</sup>
<i>Nbpf10</i>	0.0035	0.0005	1.3E-14	0.9992	118	[41] <sup>C, AS</sup>
<i>Muc19</i>	0.0015	0.0002	1.4E-14	1.0001	118	[42] <sup>C, AS</sup>
<i>Obscn</i>	0.0006	0.0001	1.5E-14	0.9999	118	[28,43] <sup>G, AS</sup>
<i>Npip3</i>	0.0019	0.0003	4.1E-14	1.0014	114	NA
<i>Dst</i>	0.0013	0.0002	9.4E-14	0.9997	111	[44,45] <sup>G, AS</sup>
<i>Col7a1</i>	0.0011	0.0001	1.4E-13	1.0001	109	[46,47] <sup>C, AS</sup>
<i>Ubr4</i>	0.0011	0.0001	1.4E-13	0.9994	109	[48,49] <sup>C, AS</sup>
<i>Hmcn1</i>	0.0006	0.0001	2.0E-13	0.9975	109	[50] <sup>C, AS</sup>
<i>Ryr2</i>	0.0011	0.0001	2.7E-13	0.9974	107	[51,52] <sup>G, AS</sup>
<i>Macf1</i>	0.0011	0.0002	3.1E-13	0.9975	106	[53,54] <sup>G, AS</sup>
<i>Mdn1</i>	0.0006	0.0001	3.5E-13	0.9993	106	NA
<i>Col4a5</i>	0.0008	0.0001	3.5E-13	0.9992	106	[55,56] <sup>C, AS</sup>
<i>Ryr1</i>	0.0007	0.0001	4.2E-13	0.9998	105	[31,57,58] <sup>G, AS</sup>
<i>Golga6l5</i>	0.0013	0.0002	5.2E-13	1.0021	104	NA
<i>Ryr3</i>	0.0009	0.0001	1.3E-12	0.9962	102	[59] <sup>C, AS</sup>
<i>Dnah14</i>	0.0007	0.0001	2.0E-12	0.9990	99	NA
<i>Herc2</i>	0.0006	0.0001	3.1E-12	1.0003	97	[60] <sup>AS</sup>
<i>Dnah8</i>	0.0005	0.0001	4.7E-12	0.9997	96	NA
<i>Nomo1</i>	0.0007	0.0001	4.9E-12	0.9996	95	NA
<i>Gpr98</i>	0.0016	0.0002	5.9E-12	0.9948	95	[61] <sup>C, AS</sup>
<i>Golga6a</i>	0.0017	0.0002	7.8E-12	1.0009	93	NA

<sup>1</sup>Estimate: exon-survival interaction variance indicator of alternative exon usage.

<sup>2</sup>SE: standard error of the estimate.

<sup>3</sup>P-value AEU: unadjusted P-value of alternative exon usage or exon-dependent association between expression and glioblastoma multiforme survival.

<sup>4</sup>Fold change: fold change in average exon expression per additional survival month.

<sup>5</sup>Exon Count: number of exons in the gene.

<sup>6</sup>Literature: review of studies that reported associations of the gene with cancers:

<sup>G</sup>: reported association of gene with glioblastoma multiforme,

<sup>C</sup>: reported association of gene with cancer other than glioblastoma multiforme,

<sup>AS</sup>: identification of different variants due to alternative splicing (AS) event,

NA: not available.

**Table 3 Ten most significant KEGG and GO categories enriched among the genes displaying alternative exon usage**

Source	Category	Gene Count <sup>1</sup>	FDR P-value <sup>2</sup>
<b>KEGG Pathway</b>	(hsa04510) focal adhesion	86	3.2E-21
	(hsa04512) ecm-receptor interaction	51	8.5E-20
	(hsa02010) abc transporters	30	2.5E-12
	(hsa04810) regulation of actin cytoskeleton	66	1.7E-07
	(hsa05412) arrhythmogenic right ventricular cardiomyopathy	32	5.9E-06
	(hsa05414) dilated cardiomyopathy	37	1.3E-06
	(hsa04070) phosphatidylinositol signaling system	31	1.2E-05
	(hsa05222) small cell lung cancer	31	3.6E-04
	(hsa05410) hypertrophic cardiomyopathy	32	1.3E-04
	(hsa05200) pathways in cancer	73	3.0E-02
<b>GO Biological Process</b>	(GO:0051056) regulation of small GTPase mediated signal transduction	105	0.055
	(GO:0022610) biological adhesion	197	2.7E-22
	(GO:0007155) cell adhesion	197	2.3E-22
	(GO:0046578) regulation of Ras protein signal transduction	79	5.0E-15
	(GO:0035023) regulation of Rho protein signal transduction	51	1.7E-15
	(GO:0007010) cytoskeleton organization	129	1.3E-15
	(GO:0030029) actin filament-based process	85	2.3E-14
	(GO:0007018) microtubule-based movement	51	2.1E-12
	(GO:0016568) chromatin modification	89	1.9E-12
	(GO:0051276) chromosome organization	132	1.4E-12
<b>GO Molecular Function</b>	(GO:0030554) adenylyl nucleotide binding	451	9.9E-59
	(GO:0005524) ATP binding	433	2.2E-59
	(GO:0032559) adenylyl ribonucleotide binding	437	2.0E-59
	(GO:0001882) nucleoside binding	456	6.3E-58
	(GO:0001883) purine nucleoside binding	451	1.5E-56
	(GO:0017076) purine nucleotide binding	480	5.2E-44
	(GO:0032555) purine ribonucleotide binding	466	2.9E-44
	(GO:0032553) ribonucleotide binding	466	2.9E-44
	(GO:0000166) nucleotide binding	523	7.4E-39
	(GO:0003774) motor activity	86	1.3E-34

<sup>1</sup>Gene Count: number of genes that have significant alternative exon usage within category.

<sup>2</sup>FDR-adjusted P-value: False discovery rate adjusted P-value of the hypergeometric test of category enrichment.

using functional and pathway analyses. At FDR-adjusted P-value < 0.05, 15 KEGG pathways, 87 GO biological processes, and 70 GO molecular functions were enriched. The top 10 pathways, biological processes and molecular functions are summarized in Table 3. Additional file 1: Table S2 lists all KEGG pathways and GO categories with FDR-adjusted P-value < 0.05. Among the 15 pathways significantly enriched, focal adhesion was the most significant pathway encompassing 86 genes. This result was consistent with many reports of the critical role of focal adhesion and gliomas [66-68]. The enrichment of the extracellular matrix- (ECM-) receptor interaction pathway detected in this study has been reported in other

cancers [69,70]. The ATP-binding cassette (ABC) transporter pathway has been associated with gliomas [71]. Our finding of small cell lung carcinoma pathways enrichment associated with GBM was consistent with the multiple studies that have identified commonalities among these cancers [72]. The most enriched biological process among the AEU genes associated with GBM survival included regulation of small GTPase mediated signal transduction (RSGST), and neuron development that has been associated with neuroblastoma [73]. The enrichment of biological adhesion confirms our focal adhesion results. Among the top 70 GO molecular functions significantly enriched were adenylyl ribonucleotide binding, ATP binding,

**Table 4 Top 5 multi-exon genes that have significant exon-independent association with glioblastoma multiforme survival**

Gene Symbol	Estimate <sup>1</sup>	SE <sup>2</sup>	Fold Change <sup>3</sup>	P-value <sup>4</sup>	P-value AEU <sup>5</sup>	Exon Count <sup>6</sup>	Literature <sup>7</sup>
<i>Sirt2</i>	0.0337	0.0092	1.0236	3.2E-04	2.5E-03	17	[62] <sup>G</sup>
<i>Six1</i>	0.0056	0.0015	1.0039	3.3E-04	2.7E-01	5	[63] <sup>C</sup>
<i>Loc100289627</i>	0.0079	0.0022	1.0055	3.8E-04	4.3E-01	2	NA
<i>Sema3e</i>	-0.0256	0.0066	0.9824	1.3E-04	2.4E-03	18	[64] <sup>C</sup>
<i>Golga8j</i>	-0.0536	0.0141	0.9635	1.7E-04	1.1E-03	20	[65] <sup>C</sup>

<sup>1</sup>Estimate: change in average exon expression per additional survival month (in log2 units).

<sup>2</sup>SE: standard error of the estimate.

<sup>3</sup>Fold change: fold change in average exon expression per additional survival month.

<sup>4</sup>P-value: unadjusted P-value of the change in average exon expression per additional survival month.

<sup>5</sup>P-value AEU: non-significant (P-value > 1.0E-03) evidence of alternative exon usage.

<sup>6</sup>Exon Count: number of exons in the gene.

<sup>7</sup>Literature: review of studies that reported associations of the gene with cancers:

<sup>G</sup>: reported association of gene with glioblastoma multiforme,

<sup>C</sup>: reported association of gene with cancer other than glioblastoma multiforme,

NA: not available.

nucleotide binding and helicase activity. These related nucleotide binding functions have been associated with GBM [74].

#### Multi-exon genes exhibiting exon-independent association with glioblastoma multiforme survival

At unadjusted P-value < 5.0E-4, 24 multi-exon genes exhibited exon-independent association with GBM survival (group 2 genes). In other words, there was no evidence of AEU in these genes because the expressions of all the exons were consistently associated with GBM survival and a single general or overall association between the gene and survival can be identified. Table 4 lists the top five multi-exon genes that have the most significant exon-independent association with GBM survival. Additional file 1: Table S3 lists the results for the 24 multi-exon genes exhibiting expression associated with survival albeit no evidence of AEU at P-value < 5.0E-4.

Among the 24 multi-exon genes that were associated with GBM survival on a general, exon-independent manner, the five genes that have the lower AEU evidence (AEU unadjusted P-value > 1.0E-3, approximately FDR adjusted P-value > 0.1) are listed in Table 4. The expression of three of these genes increased with increasing survival. Noteworthy was the low number of exons in these genes, relative to the higher number of exons in genes exhibiting evidence of AEU.

Four of five multi-exon genes have been associated to different cancers in studies listed in Table 4, and the remaining gene (LOC100289627) is similar to Guanine nucleotide-binding protein subunit beta-2-like 1. Sir-tuin2 (*Sirt2*) has been associated with GBM while the other three genes golgin subfamily A member 8J (*Golga8j*), semaphorin 3E (*Sema3e*) and SIX homeobox 1 (*SIX1*) were associated with other cancers. Under-expression of *Sirt2* has been reported in glioma cells relative to control cells [62]. This result is also consistent

with our findings that higher levels of *Sirt2* were associated with higher GBM survival *Golga8j* has been associated with pancreatic cancer and the trend is consistent with our finding of lower GBM survival with higher expression levels of this gene [65]. *Sema3e* promotes invasiveness and metastatic ability of the cancerous cells [64]. *Sema3e* is associated with many cancers like prostate cancer colon cancer and lung adenocarcinoma [75]. This result is consistent with our findings that higher levels of *Sema3e* were associated with lower GBM survival. The gene *Six1* is associated with lower survival in cancerous cells [63]. This result is inconsistent with our results showing an increase in *Six1* expression associated with an increase in GBM survival.

#### Single-exon genes associated with glioblastoma multiforme survival

Eight single-exon genes were associated with GBM survival (group 3 genes) at unadjusted P-value < 5.0E-4 (Table 5). Among these, three genes had a negative relationship such that lower expression levels were associated with higher survival. Four members of the family of small nucleolar RNA CD box (*Snord*) genes were associated with GBM survival, and three had a positive association such that higher expression levels were associated with higher survival. *Snord* are a type of small nucleolar RNA (SnoRNA) that guides the methylation of rRNAs and snRNAs. These snoRNAs can target other RNAs and are associated with carcinogenesis. Reduced and dysregulated expression of snoRNAs have been associated with progression of many human malignancies [78]. Along with their loss in brain tumorigenesis, snoRNAs have been also linked to other cancers such as prostate, breast and lung cancer [76,78]. In this study, a positive association between the levels of H1 histone family member 0 (*H1f0*) and GBM survival was identified. The expression of *H1f0* was high in breast



**Table 5 Single-exon genes associated with glioblastoma multiforme survival**

Gene symbol	Estimate <sup>1</sup>	SE <sup>2</sup>	Fold Change <sup>3</sup>	P-value <sup>4</sup>	Literature <sup>5</sup>
<i>Hist1h1t</i>	0.0118	0.0024	1.0082	2.5E-06	NA
<i>Snord116-11</i>	0.0101	0.0025	1.0070	9.7E-05	[76] <sup>C</sup>
<i>Loc729852</i>	-0.0074	0.0018	0.9949	5.8E-05	NA
<i>Snord123</i>	-0.0087	0.0025	0.9940	4.8E-04	[76] <sup>C</sup>
<i>Snord104</i>	0.0067	0.0019	1.0047	4.1E-04	[76] <sup>C</sup>
<i>Dkfp77911853</i>	-0.0083	0.0023	0.9943	3.9E-04	NA
<i>H1f0</i>	0.0062	0.0017	1.0043	2.3E-04	[77] <sup>C</sup>
<i>Snord28</i>	0.0166	0.0044	1.0116	1.8E-04	[76] <sup>C</sup>

<sup>1</sup>Estimate: change in gene expression per additional survival month (in log2 units).

<sup>2</sup>SE: standard error of the estimate.

<sup>3</sup>Fold change: fold change in gene expression per additional survival month.

<sup>4</sup>P-value: unadjusted P-value of the change in average exon expression per additional survival month;

<sup>5</sup>Literature: review of studies that reported associations of the gene with cancers:

<sup>C</sup>: reported association of gene with cancers other than glioblastoma multiforme,

NA: not available.

tumor cells, and decreased when the breast tumor cell lines were reverted-back into normal ME carcinoma cells [77].

#### Gene set enrichment analyses of all genes in consideration of their association with glioblastoma multiforme survival

Gene set enrichment analysis considered the level and sign of association between the expression of all the genes studied and GBM survival. At FDR-adjusted P-value < 0.05, 94 KEGG pathways, 402 GO biological processes, and 203 GO molecular functions were enriched. Results from the top 10 most significant pathways, biological processes and molecular functions are summarized in Tables 6, 7 and 8. Additional file 1: Table S4, S5 and S6 lists all biological processes,

molecular functions and pathways respectively that have FDR-adjusted P-value < 0.05. Pathways and GO categories are characterized in GSEA by the number of genes that have a positive or negative association between expression and GBM survival, by the log odds ratio indicating whether the category is more enriched among the genes that have a positive or negative association and the corresponding P-value.

Noteworthy was that all top ten results had negative log odds ratio indicating that the categories were more enriched among the genes that have a negative association between expression and survival relative to the enrichment among the genes that have a positive association between expression and GBM survival. Negative log<sub>e</sub> odds ratio indicates that the enrichment was higher among the genes with negative association

**Table 6 Ten most significant GO biological processes from the gene set enrichment analysis of the genome**

GO Identifier	GO Biological Process	Over-Expressed Gene <sup>1</sup>	Under-Expressed Genes <sup>2</sup>	Log Odds Ratio <sup>3</sup>	FDR P-value <sup>4</sup>
GO:0046907	intracellular transport	357	560	-0.7338	3.79E-24
GO:0034613	cellular protein localization	245	433	-0.8490	4.78E-24
GO:0043067	regulation of programmed cell death	351	490	-0.6110	1.68E-15
GO:0016192	vesicle-mediated transport	271	400	-0.6639	1.16E-14
GO:0006629	lipid metabolic process	424	538	-0.5148	1.30E-12
GO:0044265	cellular macromolecule catabolic process	373	485	-0.5379	2.10E-12
GO:0044255	cellular lipid metabolic process	346	457	-0.5528	2.41E-12
GO:0050793	regulation of developmental process	442	549	-0.4932	4.27E-12
GO:0007049	cell cycle	418	522	-0.4978	1.11E-11
GO:0009966	regulation of signal transduction	414	509	-0.4812	9.93E-11

<sup>1</sup>Over-Expressed Genes: number of genes that have a positive association between expression and glioblastoma multiforme survival.

<sup>2</sup>Under-Expressed Genes: number of genes that have a negative association between expression and glioblastoma multiforme survival.

<sup>3</sup>Log Odds Ratio: indicates whether the category is more enriched among the genes that have a positive association between expression and survival relative to the enrichment among the genes that have a negative association between expression and glioblastoma survival (positive log<sub>e</sub> odds ratio) or vice versa (negative log<sub>e</sub> odds ratio). Extreme values indicate higher difference in the enrichment percentages between the positive and negative association groups meanwhile values close to zero indicate similar enrichment percentages between positive and negative association groups.

<sup>4</sup>FDR-adjusted P-value: False discovery rate adjusted P-value of the log odds ratio test.

**Table 7 Ten most significant GO molecular functions from the gene set enrichment analysis**

GO Identifier	GO Molecular Function	Over-Expressed Gene <sup>1</sup>	Under-Expressed Genes <sup>2</sup>	Log Odds Ratio <sup>3</sup>	FDR P-value <sup>4</sup>
GO:0000287	magnesium ion binding	196	300	-0.6962	3.23E-11
GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus containing anhydrides	419	521	-0.4933	5.21E-11
GO:0016462	pyrophosphatase activity	417	520	-0.4962	5.21E-11
GO:0016817	hydrolase activity, acting on acid anhydrides	428	527	-0.4834	9.62E-11
GO:0016773	phosphotransferase activity, alcohol group as acceptor	393	475	-0.4624	5.64E-09
GO:0016301	kinase activity	421	501	-0.4473	5.64E-09
GO:0016788	hydrolase activity, acting on ester bonds	349	429	-0.4781	9.84E-09
GO:0003723	RNA binding	357	437	-0.4741	9.84E-09
GO:0030695	GTPase regulator activity	193	260	-0.5651	4.10E-07
GO:0016874	ligase activity	205	272	-0.5501	4.10E-07

<sup>1</sup>Over-Expressed Genes: number of genes that have a positive association between expression and glioblastoma multiforme survival.

<sup>2</sup>Under-Expressed Genes: number of genes that have a negative association between expression and glioblastoma multiforme survival.

<sup>3</sup>Log Odds Ratio: indicates whether the category is more enriched among the genes that have a positive association between expression and survival relative to the enrichment among the genes that have a negative association between expression and glioblastoma survival (positive log<sub>e</sub> odds ratio) or vice versa (negative log<sub>e</sub> odds ratio). Extreme values indicate higher difference in the enrichment percentages between the positive and negative association groups meanwhile values close to zero indicate similar enrichment percentages between positive and negative association groups.

<sup>4</sup>FDR-adjusted P-value: False discovery rate adjusted P-value of the log odds ratio test.

with GBM survival. Positive log odds ratios were observed for less significant (P-value < 0.05) pathways and categories. The more extreme log odds ratios observed in the GSEA of KEGG pathways indicate higher difference between the enrichment percentages in the positive and negative association groups meanwhile values close to zero in the GSEA of GO categories indicate lower differences in the enrichment percentages between positive and negative association groups.

Among the most differentially enriched pathways (Table 6) were the pancreatic and non-small cell lung cancer pathways. Additional pathways identified in this study that have been associated with gliomas include glycolysis/gluconeogenesis [79] and oxidative phosphorylation [80]. Among the top enriched GO biological processes, lipid metabolism and cell cycle have been associated with glioma [81,82]. Likewise, the GO molecular functions of hydrolase and ligase activity have been linked to glioma [83,84].

**Table 8 Ten most significant KEGG pathways from the gene set enrichment analysis of the genome**

KEGG Identifier	KEGG Pathway	Over-Expressed Gene <sup>1</sup>	Under-Expressed Genes <sup>2</sup>	Log Odds Ratio <sup>3</sup>	FDR P-value <sup>4</sup>
hsa03010	ribosome	119	16	-2.4779	9.7E-10
hsa00010	glycolysis / gluconeogenesis	57	27	-1.1614	3.6E-04
hsa00190	oxidative phosphorylation	103	39	-0.9392	3.6E-04
hsa05212	pancreatic cancer	54	45	-0.9460	4.7E-04
hsa05130	pathogenic escherichia coli infection	44	41	-1.0575	4.7E-04
hsa00240	pyrimidine metabolism	42	78	-0.8800	5.0E-04
hsa03050	proteasome	33	32	-1.0965	7.2E-04
hsa00280	valine, leucine and isoleucine degradation	20	48	-1.1353	8.5E-04
hsa04662	b cell receptor signaling pathway	34	65	-0.9084	8.5E-04
hsa05223	non-small cell lung cancer	25	52	-0.9922	9.0E-04

<sup>1</sup>Over-Expressed Genes: number of genes that have a positive association between expression and glioblastoma multiforme survival.

<sup>2</sup>Under-Expressed Genes: number of genes that have a negative association between expression and glioblastoma multiforme survival.

<sup>3</sup>Log Odds Ratio: indicates whether the category is more enriched among the genes that have a positive association between expression and survival relative to the enrichment among the genes that have a negative association between expression and glioblastoma survival (positive log<sub>e</sub> odds ratio) or vice versa (negative log<sub>e</sub> odds ratio). Extreme values indicate higher difference in the enrichment percentages between the positive and negative association groups meanwhile values close to zero indicate similar enrichment percentages between positive and negative association groups.

<sup>4</sup>FDR-adjusted P-value: False discovery rate adjusted P-value of the log odds ratio test.

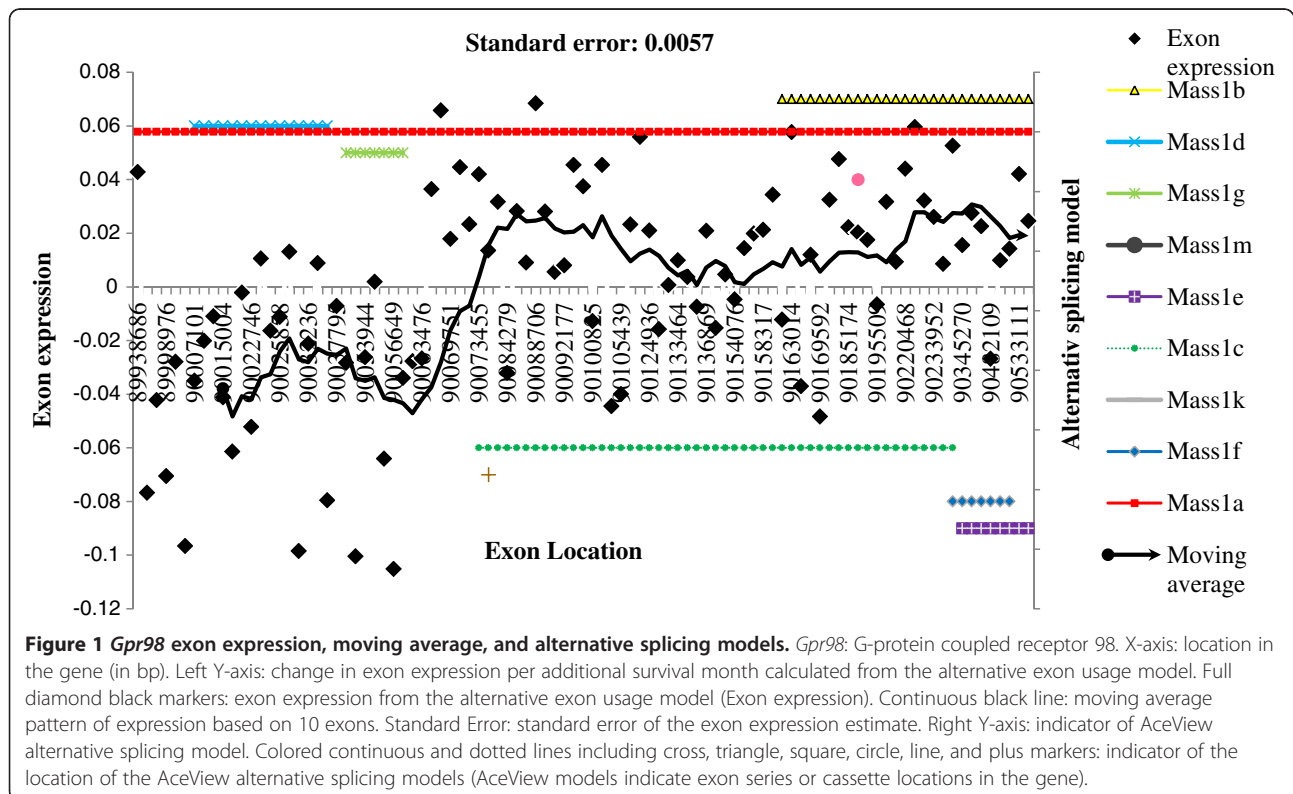
### Demonstration of alternative exon usage

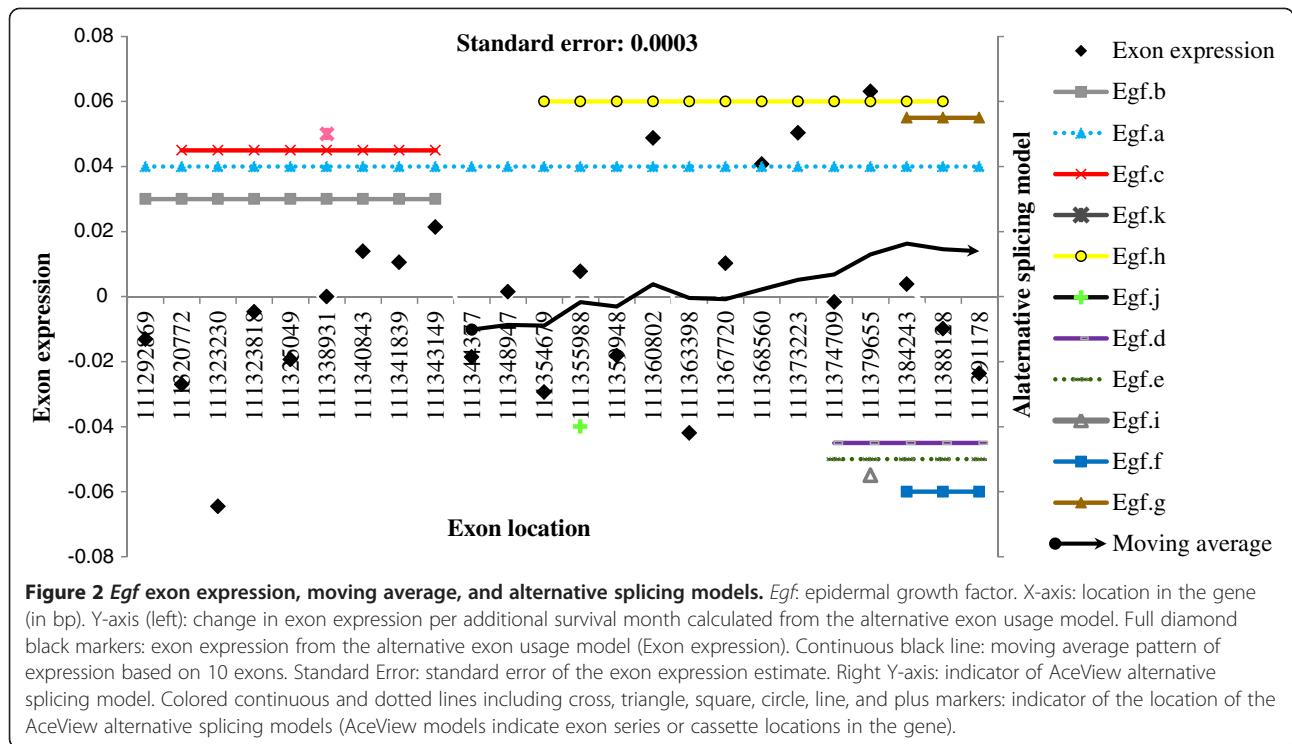
The identification of patterns of differential exon expression across a gene and comparison against predicted AS models helped to confirm associations between AS and survival. Figures 1 and 2 depict patterns of exon expression associated with GBM survival and reported AS gene models for two genes that exhibited significant AEU associated with GBM survival. The patterns of two other genes that exhibited AEU are presented in Additional file 2: Figures S1 and S2. The two genes depicted in Figures 1 and 2 are G-protein coupled receptor 98 (*Gpr98*) and epidermal growth factor (*Egf*), respectively: The two genes depicted in the Additional file 2: Figures S1 and S2 are anaphase promoting complex subunit 1 (*Anapc1*) and HECT domain and RLD domain containing E3 ubiquitin protein ligase 2 (*Herc2*), respectively. The parallel alignment of estimated exon expression resulting from our analysis, the moving average trend, and the AS prediction from AceView offered *in silico* verification of the identified AEU [2]. The AS models are denoted by lines parallel to the x-axis and identify the corresponding exons. However, no expression values should be assigned to the AS model lines and experimental confirmation of the AEU cases identified in this study is necessary.

*Gpr98* is located on chromosome 5 and is highly expressed in the central nervous system (CNS) [85]. This gene has been associated with Usher syndrome

characterized by hearing loss and progressive vision loss and Familial Febrile seizures [86,87]. *Gpr98* has been linked to cancer [88] and smaller variants of *Gpr98*, produced due to AS, have been associated with increased survival against lymphoblastic leukemia [61]. *Gpr98* exhibited AEU in this study and the expression of approximately 30 exons (out of 90 exons) exhibited significant association with GBM survival (Figure 1). Several over-expressed exons detected by our model are consistent with AS gene models including *Mass1b*, *Mass1f*, *Mass1e*, and *Mass1c*. Conversely, some under-expressed exons identified in our study are supported by gene models including *Mass1d* and *Mass1g*. These results are consistent with previous studies that indicated association of smaller transcripts of *Gpr98* with cancer survival by inducing apoptosis in cancerous cells [61]. In agreement with our GO analyses, *Gpr98* is affiliated to the enriched GO biological processes of cell adhesion, neuron development and sensory perception of mechanical stimulus. Additionally, *Gpr98* has the GO molecular function of cytoskeletal protein binding and ion binding. For *Gpr98*, the relative difference in  $R^2$  between the training and validation data sets was 18.1%. The Pearson correlation of the exon-survival associations between the training and validation data sets was 83.2%.

*Egf* is located on HAS 4 and over-expression of *Egf* has been associated with tumor progression and lower GBM survival [89]. *Egf* exhibited AEU and of the 24





exons analyzed, nine exons had significant associations with GBM survival. Several over-expressed exons detected in our analysis correspond to AS gene models including *jdec03* and *hdec03* (Figure 2). In accord with the pathway and functional analyses, *Egf* is part of many enriched KEGG pathways including focal adhesion, regulation of actin cytoskeleton, and cancer pathways. For *Egf*, the relative difference in  $R^2$  between the training and validation data sets was 24.9%. The Pearson correlation of the exon-survival associations between the training and validation data sets was 76.6%.

### Validation

The  $R^2$  is the percentage of the variation of the observations explained by the exon-based, gene-centric approach. Simply put, the  $R^2$  is an indicator of the correlation between exon expression and patient survival, adjusting for therapy, ethnicity, gender, among all terms accounted for in our model). The median  $R^2$  in the training and validation data sets were 75.7% and 63.4% for the multi- and single-exon genes significantly associated with survival. The relative difference in  $R^2$  between the training and validation data sets was 16.2%. The small drop in median  $R^2$  between the training and validating data set is a first, global indicator of the similar exon-survival relationship identified in both independent data sets. A difference between training and validating data is expected due to simple sampling effects such as between-subject variation.

Additional insight into the validation of associations detected in the training data sets was gained from the study of the correlation of the exons-survival association (e.g. estimated solution or slope) between training and validation data sets. For the multi- and single-exon genes significantly associated with survival, the median Pearson and Spearman correlations of the exon-survival associations between the training and validation data sets were 89.7% and 85.9%, respectively. The high correlation of the exon-survival associations relative to the drop in  $R^2$  model fit between training and validation data sets suggests that the exon associations with survival detected are more consistent or have lower noise than the other model terms including race, gender and therapy.

### Conclusions

In conclusion, AEU is a complex process and, thus, the detection and characterization of AEU associated with survival is challenging. The hierarchical model developed in this study allowed the simultaneous detection of differential expression of exons within a gene and differentially expressed genes associated with survival. From a total of 25,403 genes investigated, 2,477 multi-exon and 13 single exon genes were associated with GBM. Most of the significant genes detected by the model have been previously associated to GBM (27.78%) or other types of cancer (36.11%). This suggests that differential expression associated with AEU could be used as biomarkers

for GBM and potentially other cancers. The AEU events detected for several genes (*Egf*, *Herc2*, *Gpr98* and *Anapc1*) were consistent with AS models in AceView.

The approach used to identify alternative exon usage and gene expression associated with survival adjusted for race, gender, and therapy differences among the patients analyzed. Thus, prognostic biomarkers of glioblastoma survival were identified. Stratified analyses of the patients by age, race, therapy, and gender or evaluation of potential interactions between exons and clinical factors could uncover predictive biomarkers and offer additional insights into the alternative exon usage associations that can lead to more personalized treatments and predictive tools.

Extensions to the hierarchical model proposed in this study to identify AEU can be considered. First, the model can incorporate information of the mapping of the exons to the gene. In addition, the distance between the exons can be accommodated on the variance-covariance matrix. This would allow modeling of potentially higher dependencies between proximal exons relative to distant exons. Second, the model can incorporate information on different splicing scenarios [2]. The hierarchical model can be applied to other cancer types and to indicators other than survival.

In this study, the vast majority of the exons within a gene mapped to one strand and few exons mapped to the other strand. Thus, AEU was studied among the exons that mapped to the most frequent strand. When sufficient information on both strands within a gene is available, our model allows the consideration of information across strands. This model would allow the study of sense-antisense gene overlap and its impact on AS and regulation of gene expression following the work of Sorana Morrissy et al [90]. Their work suggested an anti-sense transcription-mediated mechanism of splicing regulation in human cells.

A simple yet comprehensive analytical strategy for *in silico* identification of survival-associated alternative exon usage and general gene expression was demonstrated. The findings from this strategy and stringent biological and statistical thresholds were validated on an independent group of patients. Our approach can be used as a first step in the identification of cancer molecular biomarkers. A subsequent step is the experimental validation of the identified alternative exon usage and patterns of association between exons series or cassettes and survival. Experimental confirmation can be obtained from exon expression studies of the proliferation and survival of glioblastoma cell lines [91] or from studies of primary glioblastoma sphere cultures (gliospheres), an established *in vitro* model for cancer stem cell expansion [92]. Furthermore, the proposed analytical strategy can be applied to next-generation sequencing data, allowing a

thorough investigation of the expression pattern associated with cancer survival and other complex phenotypes.

Further validation of alternative exon usage biomarkers can be carried out using Reverse Transcriptase-quantitative Polymerase Chain Reaction (RT-qPCR) assays or RNA-Seq technologies as new samples become available. Confirmation of results using RNA-Seq offers various advantages. Unlike exon arrays that require probe design and annotations, RNA-Seq can detect both known and previously unreported alternative splicing events and yet to be annotated transcripts. RNA-Seq has substantially better coverage for differentially expressed genes compared to arrays. The enhanced exon coverage and increased sensitivity to detect alternative splicing sites and differentially expressed exons constitutes a robust tool that can substantially enhance the understanding of alternative exon usage associated with complex phenotypes. The clinical benefit of alternative exon usage and associated exon cassettes or transcripts will be most valuable in cases when these biomarkers provide a significant improvement in the precision to predict survival over routinely available clinical tests and overall gene expression-based biomarkers. As a result, the impact of the superior biomarkers will likely be greatest in diseases with short average post-diagnostic survival such as glioblastoma multiforme. Likewise, alternative exon usage-based biomarkers have the potential to be helpful to predict phenotypes not accurately predicted by general gene-expression profiles. For several cancer types the recurrence of metastasis is the most compelling assessment of the efficacy of therapy. In these cases, accurate and replicable exon-based prognostic tools offer the most advantage and can complement available clinical tests.

## Additional files

**Additional file 1: Table S1.** Lists the results for the 129 multi-exon genes that exhibit evidence of AEU at P-value < 1.0E-8. Lists the results for the 129 multi-exon genes that exhibit evidence of AEU at P-value < 1.0E-8. Table S2: Significant KEGG and GO categories enriched among the genes displaying alternative exon usage. Lists all KEGG pathways, GO Biological Processes and GO Molecular Function categories at FDR-adjusted P-value < 0.05. Table S3: Multi-exon genes that have significant exon-independent association with glioblastoma multiforme survival. Lists the results for the 24 multi-exon genes exhibiting expression associated with survival albeit no evidence of AEU at P-value < 5.0E-4. Table S4: Significant GO biological processes (levels 3-6) from the gene set enrichment analysis of the genome. Lists all biological processes at FDR-adjusted P-value < 0.05. Table S5: Significant GO molecular functions (levels 3-6) from the gene set enrichment analysis of the genome. Lists all molecular functions at FDR-adjusted P-value < 0.05. Table S6: Significant KEGG pathways (levels 3-6) from the gene set enrichment analysis of the genome. Lists all pathways at FDR-adjusted P-value < 0.05.

**Additional file 2: Figure S1.** Anapc1 exon expression, moving average, and alternative splicing models. Depicts the alternative exon expression, moving average and alternative splicing models for anaphase promoting

complex subunit 1. Figure S2. Herc2 exon expression, moving average, and alternative splicing models. Depicts the alternative exon expression, moving average and alternative splicing models for HECT domain and RLD domain containing E3 ubiquitin protein ligase 2.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

AS obtained the exon expression data set, compiled information, performed the statistical analysis contributed to the interpretation of results, and drafted the manuscript. NVLS developed the clinical classifications, developed statistical routines, and contributed to the interpretation of the results. BRS wrote code to implement the statistical model and contributed to the manuscript. KRD contributed with the analysis. SRZ obtained funding for the study, participated in its conception, coordination, interpretation of results, and helped write the manuscript. All authors have read and approved the final version of this manuscript.

#### Acknowledgements

The support of NCI (Grant Number: 1R03CA143975), NIH/NIDA (Grant Numbers: R21DA027548 and P30DA018310), TCGA Data Access Committee project #1347, and Biosciences Department COMSATS Institute of Information Technology, and the reviewer's suggestions are greatly appreciated.

#### Author details

<sup>1</sup>Department of Animal Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA. <sup>2</sup>Department of Statistics, University of Illinois Urbana-Champaign, Urbana, IL, USA. <sup>3</sup>Institute of Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA.

Received: 6 July 2012 Accepted: 27 November 2012

Published: 4 December 2012

#### References

- Barash Y, Blencowe BJ, Frey BJ: **Model-based detection of alternative splicing signals.** *Bioinformatics* 2010, **26**(12):i325–i333.
- Laderas TG, Walter NA, Mooney M, Vartanian K, Darakjian P, Buck K, Harrington CA, Belknap J, Hitzemann R, McWeeney SK: **Computational detection of alternative exon usage.** *Front Neurosci* 2011, **5**:69.
- Sakabe NJ, Vbranovski MD, de Souza SJ: **A bioinformatics analysis of alternative exon usage in human genes coding for extracellular matrix proteins.** *Genet Mol Res* 2004, **3**(4):532–544.
- Johnson JM, Castle J, Garrett-Engel P, Kan Z, Loerch PM, Armour CD, Santos R, Schadt EE, Stoughton R, Shoemaker DD: **Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays.** *Science* 2003, **302**(5653):2141–2144.
- Pal S, Gupta R, Davuluri RV: **Alternative transcription and alternative splicing in cancer.** *Pharmacol Ther* 2012, **136**(3):283–294.
- Ramskold D, Wang ET, Burge CB, Sandberg R: **An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data.** *PLoS Comput Biol* 2009, **5**(12):e1000598.
- Yeo G, Holste D, Kreiman G, Burge CB: **Variation in alternative splicing across human tissues.** *Genome Biol* 2004, **5**(10):R74.
- Cheung HC, Baggerly KA, Tsavachidis S, Bachinski LL, Neubauer VL, Nixon TJ, Aldape KD, Cote GJ, Krahe R: **Global analysis of aberrant pre-mRNA splicing in glioblastoma using exon expression arrays.** *BMC Genomics* 2008, **9**:216.
- Serao NV, Delfino KR, Southey BR, Beever JE, Rodriguez-Zas SL: **Cell cycle and aging, morphogenesis, and response to stimuli genes are individualized biomarkers of glioblastoma progression and survival.** *BMC Med Genomics* 2011, **4**:49.
- Delfino KR, Serao NV, Southey BR, Rodriguez-Zas SL: **Therapy-, gender- and race-specific microRNA markers, target genes and networks related to glioblastoma recurrence and survival.** *CANCER GENOMICS PROTEOMICS* 2011, **8**(4):173–183.
- Lo HW, Zhu H, Cao X, Aldrich A, Ali-Osman F: **A novel splice variant of GLI1 that promotes glioblastoma cell migration and invasion.** *Cancer Res* 2009, **69**(17):6790–6798.
- Johnson DR, O'Neill BP: **Glioblastoma survival in the United States before and during the temozolomide era.** *J Neurooncol* 2012, **107**(2):359–364.
- Krex D, Klink B, Hartmann C, von Deimling A, Pietsch T, Simon M, Sabel M, Steinbach JP, Heese O, Reifenberger G, Weller M, Schackert G, German Glioma Network: **Long-term survival with glioblastoma multiforme.** *Brain* 2007, **130**(Pt 10):2596–2606.
- Lamborn KR, Chang SM, Prados MD: **Prognostic factors for survival of patients with glioblastoma: recursive partitioning analysis.** *Neuro Oncol* 2004, **6**(3):227–235.
- Su WL, Modrek B, GuhaThakurta D, Edwards S, Shah JK, Kulkarni AV, Russell A, Schadt EE, Johnson JM, Castle JC: **Exon and junction microarrays detect widespread mouse strain- and sex-bias expression differences.** *BMC Genomics* 2008, **9**:273.
- Purdom E, Simpson KM, Robinson MD, Conboy JG, Lapuk AV, Speed TP: **FIRMA: a method for detection of alternative splicing from exon array data.** *Bioinformatics* 2008, **24**(15):1707–1714.
- Zheng H, Hang X, Zhu J, Qian M, Qu W, Zhang C, Deng M: **REMAS: a new regression model to identify alternative splicing events from exon array data.** *BMC Bioinforma* 2009, **10**(Suppl 1):S18.
- Cline MS, Blume J, Cawley S, Clark TA, Hu JS, Lu G, Salomonis N, Wang H, Williams A: **ANOSVA: a statistical method for detecting splice variation from expression data.** *Bioinformatics* 2005, **21**(Suppl 1):i107–i115.
- Maronna RA, Martin RD, Yohai VJ: *Times Series, in Robust Statistics: Theory and Methods.* Ltd, Chichester, UK: J. Wiley & Sons; 2006.
- Klipper-Aurbach Y, Wasserman M, Braunspeigel-Weintrob N, Borstein D, Peleg S, Assa S, Karp M, Benjamini Y, Hochberg Y, Laron Z: **Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus.** *Med Hypotheses* 1995, **45**(5):486–490.
- Ji H, Wong WH: **TileMap: create chromosomal map of tiling array hybridizations.** *Bioinformatics* 2005, **21**(18):3629–3636.
- da Huang W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1–13.
- da Huang W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44–57.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**(43):15545–15550.
- Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, Conesa A, Tarraga J, Pascual-Montano A, Nogales-Cadenas R, Santoyo J, Garcia F, Marba M, Montaner D, Dopazo J: **Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling.** *Nucleic Acids Res* 2010, **38**(Web Server issue):W210–W213.
- Olson KM, Vanraden PM, Tooker ME, Cooper TA: **Differences among methods to validate genomic evaluations for dairy cattle.** *J Dairy Sci* 2011, **94**(5):2613–2620.
- Marko NF, Toms SA, Barnett GH, Weil R: **Genomic expression patterns distinguish long-term from short-term glioblastoma survivors: a preliminary feasibility study.** *Genomics* 2008, **91**(5):395–406.
- Balakrishnan A, Bleeker FE, Lamba S, Rodolfo M, Daniotti M, Scarpa A, van Tilborg AA, Leenstra S, Zanon C, Bardelli A: **Novel somatic and germline mutations in cancer candidate genes in glioblastoma, melanoma, and pancreatic carcinoma.** *Cancer Res* 2007, **67**(8):3545–3550.
- Gewandter JS, Bambara RA, O'Reilly MA: **The RNA surveillance protein SMG1 activates p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA.** *Cell Cycle* 2011, **10**(15):2561–2567.
- Donner K, Sandbacka M, Lehtokari VL, Wallgren-Pettersson C, Pelin K: **Complete genomic structure of the human nebulin gene and identification of alternatively spliced transcripts.** *Eur J Hum Genet* 2004, **12**(9):744–751.
- Difilippantonio S, Chen Y, Pietas A, Schluns K, Pacyna-Gengelbach M, Deutschmann N, Padilla-Nash HM, Ried T, Petersen I: **Gene expression profiles in human non-small and small-cell lung cancers.** *Eur J Cancer* 2003, **39**(13):1936–1947.

32. Ma Y, Yuan RQ, Fan S, Hu C, Goldberg ID, Latorra JJ, Rosen EM: **Identification of genes that modulate sensitivity of U373MG glioblastoma cells to cis-platinum.** *Anticancer Drugs* 2006, **17**(7):733-751.
33. Korshunov A, Sycheva R, Golanov A: **Genetically distinct and clinically relevant subtypes of glioblastoma defined by array-based comparative genomic hybridization (array-CGH).** *Acta Neuropathol* 2006, **111**(5):465-474.
34. Ditttrich B, Buiting K, Korn B, Rickard S, Buxton J, Saitoh S, Nicholls RD, Poustka A, Winterpacht A, Zabel B, Horsthemke B: **Imprint switching on human chromosome 15 may involve alternative transcripts of the SNRPN gene.** *Nat Genet* 1996, **14**(2):163-170.
35. Moreno MJ, Ball M, Andrade MF, McDermid A, Stanimirovic DB: **Insulin-like growth factor binding protein-4 (IGFBP-4) is a novel anti-angiogenic and anti-tumorigenic mediator secreted by dibutyryl cyclic AMP (dB-cAMP)-differentiated glioblastoma cells.** *Glia* 2006, **53**(8):845-857.
36. 't Hoen PA, Hirsch M, de Meijer EJ, de Menezes RX, van Ommen GJ, den Dunnen JT: **mRNA degradation controls differentiation state-dependent differences in transcript and splice variant abundance.** *Nucleic Acids Res* 2011, **39**(2):556-566.
37. Hodzic D, Kong C, Wainszelbaum MJ, Charron AJ, Su X, Stahl PD: **TBC1D3, a hominoid oncoprotein, is encoded by a cluster of paralogues located on chromosome 17q12.** *Genomics* 2006, **88**(6):731-736.
38. He ML, Chen Y, Chen Q, He Y, Zhao J, Wang J, Yang H, Kung HF: **Multiple gene dysfunctions lead to high cancer-susceptibility: evidences from a whole-exome sequencing study.** *Am J Cancer Res* 2011, **1**(4):562-573.
39. Gamsiz ED, Ouyang Q, Schmidt M, Nagpal S, Morrow EM: **Genome-wide transcriptome analysis in murine neural retina using high-throughput RNA sequencing.** *Genomics* 2012, **99**(1):44-51.
40. Venables JP, Klinck R, Bramard A, Inkel L, Dufresne-Martin G, Koh C, Gervais-Bird J, Lapointe E, Froehlich U, Durand M, Gendron D, Brosseau JP, Thibault P, Lucier JF, Tremblay K, Prinos P, Wellinger RJ, Chabot B, Rancourt C, Elela SA: **Identification of alternative splicing markers for breast cancer.** *Cancer Res* 2008, **68**(22):9525-9531.
41. Vandepoele K, Andries V, Van Roy N, Staes K, Vandesompele J, Laureys G, De Smet E, Berx G, Speleman F, van Roy F: **A constitutional translocation t(1;17)(p36.2;q11.2) in a neuroblastoma patient disrupts the human NBPF1 and ACCN1 genes.** *PLoS One* 2008, **3**(5):e2207.
42. Das B, Cash MN, Hand AR, Shivazad A, Culp DJ: **Expression of Muc19/SmgG gene products during murine sublingual gland development: cytodifferentiation and maturation of salivary mucous cells.** *J Histochem Cytochem* 2009, **57**(4):383-396.
43. Kim JH, Park BL, Pasaje CF, Kim Y, Bae JS, Park JS, Uh ST, Kim YH, Kim MK, Choi IS, Cho SH, Choi BW, Koh I, Park CS, Shin HD: **Contribution of the OBSCN Nonsynonymous Variants to Aspirin Exacerbated Respiratory Disease Susceptibility in Korean Population.** *DNA Cell Biol* 2012, **31**(6):1001-1009.
44. Dong H, Luo L, Hong S, Siu H, Xiao Y, Jin L, Chen R, Xiong M: **Integrated analysis of mutations, miRNA and mRNA expression in glioblastoma.** *BMC Syst Biol* 2010, **4**:163.
45. Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiu L, Ares M Jr, Black DL: **A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons.** *Genes Dev* 2007, **21**(13):1636-1652.
46. Kita Y, Mimori K, Tanaka F, Matsumoto T, Haraguchi N, Ishikawa K, Matsuzaki S, Fukuyoshi Y, Inoue H, Natsugoe S, Aikou T, Mori M: **Clinical significance of LAMB3 and COL7A1 mRNA in esophageal squamous cell carcinoma.** *Eur J Surg Oncol* 2009, **35**(1):52-58.
47. Wessagowit V, Nalla VK, Rogan PK, McGrath JA: **Normal and abnormal mechanisms of gene splicing and relevance to inherited skin diseases.** *J Dermatol Sci* 2005, **40**(2):73-84.
48. Dalglish GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR, Futreal PA: **Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes.** *Nature* 2010, **463**(7279):360-363.
49. Fehlbaum-Beurdeley P, Jarrige-Le Prado AC, Pallares D, Carriere J, Guihal C, Soucaille C, Rouet F, Drouin D, Sol O, Jordan H, Wu D, Lei L, Einstein R, Schweighoffer F, Bracco L: **Toward an Alzheimer's disease diagnosis via high-resolution blood gene expression.** *Alzheimers Dement* 2010, **6**(1):25-38.
50. Arne G, Kristiansson E, Nerman O, Kindblom LG, Ahlman H, Nilsson B, Nilsson O: **Expression profiling of GIST: CD133 is associated with KIT exon 11 mutations, gastric location and poor prognosis.** *Int J Cancer* 2011, **129**(5):1149-1161.
51. Hoelzinger DB, Mariani L, Weis J, Woyke T, Berens TJ, McDonough WS, Sloan A, Coons SW, Berens ME: **Gene expression profile of glioblastoma multiforme invasive phenotype points to new therapeutic targets.** *Neoplasia* 2005, **7**(1):7-16.
52. George CH, Rogers SA, Bertrand BM, Tunwell RE, Thomas NL, Steele DS, Cox EV, Pepper C, Hazeel CJ, Claycomb WC, Lai FA: **Alternative splicing of ryanodine receptors modulates cardiomyocyte Ca2+ signaling and susceptibility to apoptosis.** *Circ Res* 2007, **100**(6):874-883.
53. Han J, Yang L, Puri RK: **Analysis of target genes induced by IL-13 cytotoxin in human glioblastoma cells.** *J Neurooncol* 2005, **72**(1):35-46.
54. Misquitta-Ali CM, Cheng E, O'Hanlon D, Liu N, McGlade CJ, Tsao MS, Blencowe BJ: **Global profiling and molecular characterization of alternative splicing events misregulated in lung cancer.** *Mol Cell Biol* 2011, **31**(1):138-150.
55. Gorlov IP, Byun J, Gorlova OY, Aparicio AM, Efstathiou E, Logothetis CJ: **Candidate pathways and genes for prostate cancer: a meta-analysis of gene expression data.** *BMC Med Genomics* 2009, **2**:48.
56. Lemmink HH, Kluijtmans LA, Brunner HG, Schroder CH, Knebelmann B, Jelinkova E, van Oost BA, Monnens LA, Smeets HJ: **Aberrant splicing of the COL4A5 gene in patients with Alport syndrome.** *Hum Mol Genet* 1994, **3**(2):317-322.
57. van den Boom J, Wolter M, Kuick R, Miskic DE, Youkilis AS, Wechsler DS, Sommer C, Reifenberger G, Hanash SM: **Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction.** *Am J Pathol* 2003, **163**(3):1033-1043.
58. Kimura T, Lueck JD, Harvey PJ, Pace SM, Ikemoto N, Casarotto MG, Dirksen RT, Dulhunty AF: **Alternative splicing of RyR1 alters the efficacy of skeletal EC coupling.** *Cell Calcium* 2009, **45**(3):264-274.
59. Zhang L, Liu Y, Song F, Zheng H, Hu L, Lu H, Liu P, Hao X, Zhang W, Chen K: **Functional SNP in the microRNA-367 binding site in the 3'UTR of the calcium channel ryanodine receptor gene 3 (RYR3) affects breast cancer risk and calcification.** *Proc Natl Acad Sci U S A* 2011, **108**(33):13653-13658.
60. Hochrainer K, Mayer H, Baranyi U, Binder B, Lipp J, Kroismayr R: **The human HERC family of ubiquitin ligases: novel members, genomic organization, expression profiling, and evolutionary aspects.** *Genomics* 2005, **85**(2):153-164.
61. Rainer J, Lelong J, Bindreither D, Mantinger C, Ploner C, Geley S, Kofler R: **Research resource: transcriptional response to glucocorticoids in childhood acute lymphoblastic leukemia.** *Mol Endocrinol* 2012, **26**(1):178-193.
62. Haigis MC, Sinclair DA: **Mammalian sirtuins: biological insights and disease relevance.** *Annu Rev Pathol* 2010, **5**:253-295.
63. Qamar L, Deitsch E, Patrick AN, Post MD, Spillman MA, Iwanaga R, Thorburn A, Ford HL, Behbakht K: **Specificity and prognostic validation of a polyclonal antibody to detect Six1 homeoprotein in ovarian cancer.** *Gynecol Oncol* 2012, **125**(2):451-457.
64. Casazza A, Kigel B, Maione F, Capparuccia L, Kessler O, Giraudo E, Mazzone M, Neufeld G, Tamagnone L: **Tumour growth inhibition and anti-metastatic activity of a mutated furin-resistant Semaphorin 3E isoform.** *EMBO Mol Med* 2012, **4**(3):234-250.
65. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, Hong SM, Fu B, Lin MT, Calhoun ES, Kamiyama M, Walter K, Nikolskaya T, Nikolsky Y, Hartigan J, Smith DR, Hidalgo M, Leach SD, Klein AP, Jaffee EM, Goggins M, Maitra A, Iacobuzio-Donahue C, Eshleman JR, Kern SE, Hruban RH, Karchin R, Papadopoulos N, Parmigiani G, Vogelstein B, Velculescu VE, Kinzler KW: **Core signaling pathways in human pancreatic cancers revealed by global genomic analyses.** *Science* 2008, **321**(5897):1801-1806.

66. Francescone R, Scully S, Bentley B, Yan W, Taylor SL, Oh D, Moral L, Shao R: **Glioblastoma-derived tumor cells induce vasculogenic mimicry through Flk-1 activation.** *J Biol Chem* 2012, **287**(29):24821–24831.
67. Garzon-Muvdi T, Schiapparelli P, ap Rhys C, Guerrero-Cazares H, Smith C, Kim DH, Kone L, Farber H, Lee DY, An SS, Levchenko A, Quinones-Hinojosa A: **Regulation of brain tumor dispersal by NKCC1 through a novel role in focal adhesion regulation.** *PLoS Biol* 2012, **10**(5):e1001320.
68. Swiatek-Machado K, Mieczkowski J, Ellert-Miklaszewska A, Swierk P, Fokt I, Szymanski S, Skora S, Szeja W, Gryniewicz G, Lesyng B, Priebe W, Kaminska B: **Novel small molecular inhibitors disrupt the JAK/STAT3 and FAK signaling pathways and exhibit a potent antitumor activity in glioma cells.** *Cancer Biol Ther* 2012, **13**(8):657–670.
69. Jiang ZQ, Gui SB, Zhang YZ: **Differential gene expression by fiber-optic beadarray and pathway in adrenocorticotrophin-secreting pituitary adenomas.** *Chin Med J (Engl)* 2010, **123**(23):3455–3461.
70. Qiu X, Guo S, Wu H, Chen J, Zhou Q: **Identification of Wnt pathway, uPA, PAI-1, MT1-MMP, S100A4 and CXCR4 associated with enhanced metastasis of human large cell lung cancer by DNA microarray.** *Minerva Med* 2012, **103**(3):151–164.
71. Kievit FM, Wang FY, Fang C, Mok H, Wang K, Silber JR, Ellenbogen RG, Zhang M: **Doxorubicin loaded iron oxide nanoparticles overcome multidrug resistance in cancer in vitro.** *J Control Release* 2011, **152**(1):76–83.
72. Zheng M, Morgan-Lappe SE, Yang J, Bockbrader KM, Pamarthy D, Thomas D, Fesik SW, Sun Y: **Growth inhibition and radiosensitization of glioblastoma and lung cancer cells by small interfering RNA silencing of tumor necrosis factor receptor-associated factor 2.** *Cancer Res* 2008, **68**(18):7570–7578.
73. Serra A, Haberle B, Konig IR, Kappler R, Suttorp M, Schackert HK, Roesner D, Fitze G: **Rare occurrence of PHOX2b mutations in sporadic neuroblastomas.** *J Pediatr Hematol Oncol* 2008, **30**(10):728–732.
74. Lymbouridou R, Soufla G, Chatzinikola AM, Vakis A, Spandidos DA: **Down-regulation of K-ras and H-ras in human brain gliomas.** *Eur J Cancer* 2009, **45**(7):1294–1303.
75. Blanc V, Naricula J, Munson P, Freeman A, Klocker H, Masters J, Williamson M: **A role for class 3 semaphorins in prostate cancer.** *Prostate* 2011, **71**(6):649–658.
76. Liao J, Yu L, Mei Y, Guarnera M, Shen J, Li R, Liu Z, Jiang F: **Small nucleolar RNA signatures as biomarkers for non-small-cell lung cancer.** *Mol Cancer* 2010, **9**:198.
77. Klein A, Guhl E, Zollinger R, Tzeng YJ, Wessel R, Hummel M, Graessmann M, Graessmann A: **Gene expression profiling: cell cycle deregulation and aneuploidy do not cause breast cancer formation in WAP-SVT/t transgenic animals.** *J Mol Med (Berl)* 2005, **83**(5):362–376.
78. Askarian-Amiri ME, Crawford J, French JD, Smart CE, Smith MA, Clark MB, Ru K, Mercer TR, Thompson ER, Lakhani SR, Vargas AC, Campbell IG, Brown MA, Dinger ME, Mattick JS: **SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer.** *RNA* 2011, **17**(5):878–891.
79. Oudard S, Boitier E, Miccoli L, Rousset S, Dutrillaux B, Poupon MF: **Gliomas are driven by glycolysis: putative roles of hexokinase, oxidative phosphorylation and mitochondrial ultrastructure.** *Anticancer Res* 1997, **17**(3C):1903–1911.
80. Clarion L, Schindler M, de Weille J, Lolmede K, Laroche-Clary A, Uro-Coste E, Robert J, Mersel M, Bakalara N: **7beta-Hydroxycholesterol-induced energy stress leads to sequential opposing signaling responses and to death of C6 glioblastoma cells.** *Biochem Pharmacol* 2012, **83**(1):37–46.
81. Laurenti G, Benedetti E, D'Angelo B, Cristiano L, Cinque B, Raysi S, Alecci M, Ceru MP, Cifone MG, Galzio R, Giordano A, Cimini A: **Hypoxia induces peroxisome proliferator-activated receptor alpha (PPARalpha) and lipid metabolism peroxisomal enzymes in human glioblastoma cells.** *J Cell Biochem* 2011, **112**(12):3891–3901.
82. Liu S, Yin F, Fan W, Wang S, Guo XR, Zhang JN, Tian Z, Fan M: **Over-expression of BMPR-IB reduces the malignancy of glioblastoma cells by upregulation of p21 and p27Kip1.** *J Exp Clin Cancer Res* 2012, **31**(1):52.
83. Lenman A, Fowler CJ: **Interaction of ligands for the peroxisome proliferator-activated receptor gamma with the endocannabinoid system.** *Br J Pharmacol* 2007, **151**(8):1343–1351.
84. Jiang X, Xing H, Kim TM, Jung Y, Huang W, Yang HW, Song S, Park PJ, Carroll RS, Johnson MD: **Numb regulates glioma stem cell fate and growth by altering epidermal growth factor receptor and skp1-cullin-f-box ubiquitin ligase activity.** *Stem Cells* 2012, **30**(7):1313–1326.
85. Piro RM, Molineris I, Ala U, Di Cunto F: **Evaluation of candidate genes from orphan FEB and GEFS+ loci by analysis of human brain gene expression atlases.** *PLoS One* 2011, **6**(8):e23149.
86. Millan JM, Aller E, Jaijo T, Blanco-Kelly F, Gimenez-Pardo A, Ayuso C: **An update on the genetics of usher syndrome.** *J Ophthalmol* 2011, **2011**:417217.
87. Nakayama J, Fu YH, Clark AM, Nakahara S, Hamano K, Iwasaki N, Matsui A, Arinami T, Ptacek LJ: **A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures.** *Ann Neurol* 2002, **52**(5):654–657.
88. Nagayama K, Kohno T, Sato M, Arai Y, Minna JD, Yokota J: **Homozygous deletion scanning of the lung cancer genome at a 100-kb resolution.** *Genes Chromosomes Cancer* 2007, **46**(11):1000–1010.
89. Sjostrom S, Andersson U, Liu Y, Brannstrom T, Broholm H, Johansen C, Collatz-Lai H, Henriksson R, Bondy M, Melin B: **Genetic variations in EGF and EGFR and glioblastoma outcome.** *Neuro Oncol* 2010, **12**(8):815–821.
90. Morrissy AS, Griffith M, Marra MA: **Extensive relationship between antisense transcription and alternative splicing in the human genome.** *Genome Res* 2011, **21**(8):1203–1212.
91. Poch E, Minambres R, Mocholi E, Ivorra C, Perez-Arago A, Guerri C, Perez-Roger I, Guasch RM: **RhoE interferes with Rb inactivation and regulates the proliferation and survival of the U87 human glioblastoma cell line.** *Exp Cell Res* 2007, **313**(4):719–731.
92. Janiszewska M, Suva ML, Riggi N, Houtkooper RH, Auwerx J, Clement-Schatlo V, Radovanovic I, Rheinbay E, Provero P, Stamenkovic I: **Imp2 controls oxidative phosphorylation and is crucial for preserving glioblastoma cancer stem cells.** *Genes Dev* 2012, **26**(17):1926–1944.

doi:10.1186/1755-8794-5-59

Cite this article as: Sadeque et al.: Identification and characterization of alternative exon usage linked glioblastoma multiforme survival. *BMC Medical Genomics* 2012 5:59.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

