**Title:**

Novel isoforms of adhesion G protein coupled receptor B1 (ADGRB1/BAI1) generated from an alternative promoter in intron 17

**Short Title:**

New BAI1 isoforms from alternative promoter

**Authors:**

Rashed Rezwan Parag[1,2*], Takahiro Yamamoto [1*], and Erwin G. Van Meir [1,3,#]

[1] Laboratory of Molecular Neuro-Oncology, Department of Neurosurgery, Heersink School of Medicine, University of Alabama at Birmingham (UAB), Birmingham, AL 35294, USA.

[2] Graduate Biomedical Sciences, University of Alabama at Birmingham (UAB), Birmingham, AL

[3] O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham (UAB), Birmingham, AL 35294, USA.

[*] Equal Contribution

[#] Corresponding author:

Erwin G. Van Meir, Ph.D.

Department of Neurosurgery

The University of Alabama at Birmingham

WTI 520E, 1720 2nd Ave. South, Birmingham, AL 35294

Phone: 205-937-8824

evanmeir@uab.edu

**Keywords:**

Adhesion GPCR, ADGRB1, BAI1, alternative promoter, alternative splicing, transcription variants, protein isoforms.

**Abstract:**

Brain-specific angiogenesis inhibitor 1 (BAI1) belongs to the adhesion G-protein-coupled receptors, which exhibit large multi-domain extracellular N-termini that mediate cell-cell and cell-matrix interactions. To explore for the existence of BAI1 isoforms, we queried genomic datasets for markers of active chromatin and new transcript variants in the *ADGRB1* gene. Two major types of mRNAs were identified in human and mouse brain, those with a start codon in exon 2 encoding a full-length protein of a predicted size of 173.5 kDa and shorter transcripts starting from alternative exons at the intron 17/exon 18 boundary with new or exon 19 start codons, predicting shorter isoforms of 76.9 and 70.8 kDa, respectively. Immunoblots on wild-type and *Adgrb1 exon 2*-deleted mice, reverse transcription PCR and promoter-luciferase reporters confirmed that the shorter isoforms originate from an alternative promoter in intron 17. The shorter BAI1 isoforms lack most of the N-terminus and are very close in structure to the truncated BAI1 isoform generated through GPS processing from the full-length receptor, except that the latter exhibits a 19 amino acid extracellular stalk that can serve as a receptor agonist. Further studies are warranted to compare the functions of these isoforms and examine the distinct roles they play in different tissues and cell types.

## 1. Introduction

Brain-specific angiogenesis inhibitor (BAI1) is a member of the adhesion G-protein-coupled receptors (ADGRs), a class of 33 GPCRs in humans that exhibit long multi-domain extracellular regions mediating cell-cell and cell-matrix interactions.[1, 2] BAI1 is a transmembrane receptor with a predicted size of 173.5 kDa and a modular structure: sequentially from the N-terminus, the extracellular region of BAI1 starts with a signal peptide for intracellular transport, an RGD (Arg-Gly-Asp) integrin-binding motif, and five thrombospondin type 1 repeats (TSRs), that can interact with CD36 on endothelial cells[3], bind to phosphatidylserine on apoptotic cells[4] and RTN4R on neurons.[5] The TSRs domain is followed by a Hormone-binding domain (HBD) of unknown ligand, and a GPCR autoproteolysis-inducing (GAIN) domain that can stimulate autoproteolytic cleavage at an adjacent GPCR proteolysis site (GPS). This cleavage separates the N-terminal extracellular fragment (NTF) from the rest of the receptor, leaving a membrane-bound C-terminal fragment (CTF) with a short (~19 amino acid) extracellular stalk that can then function as an agonist and lead to BAI1 activation.[6] The 7-helical transmembrane (7TM) domain anchors BAI1 to the cell membrane.[7] The first intracellular loop binds to the MDM2 E3-ubiquitin ligase,[8] while the third loop binds $G\alpha_{12/13}$ and activates Rho signaling[9]. Finally, the intracellular C-terminus features a Proline rich region (PRR) that interacts with IRSp53, an adaptor protein that links membrane-bound small G proteins to cytoplasmic effector proteins,[10] a helical domain (HD) that can recruit ELMO/DOCK180 and activate Rac1 signaling,[11] a nuclear localization signal peptide,[12] and ends with amino acids QTEV (Gln-Thr-Glu-Val), which function as a docking site for PDZ domain-containing proteins[10] such as scaffolding protein MAGI-3, PSD-95 and Tiam1/Par3 that activates Rac1 signaling.[9, 13, 14]

BAI1 was initially studied for the anti-angiogenic properties of the TSRs found in its N terminus and its overexpression can inhibit cancer growth.[15] The extracellular region of BAI1 can be cleaved by MMP14, after the first TSR, to release a fragment with a predicted size of 34.7 kDa (vasculostatin-40) and autoproteolytically at the GPS site to release the NTF (also called vasculostatin-120, predicted size: 101.5 kDa), both of which have anti-tumor properties in gliomas.[3, 16] BAI1 expression is elevated in the brain[17] but epigenetically silenced in brain tumors (gliomas and medulloblastomas) and has anti-cancer function by trapping MDM2, which stabilizes the p53 tumor suppressor.[8] BAI1 was also reported to serve as an engulfment receptor for apoptotic

cells and bacteria in macrophages[4] and promote myogenesis, which relates to muscle development and repair [18] through ELMO/Dock180/Rac1 signaling.

BAI1 is highly expressed in glial cells and neurons of the hippocampus, thalamus, amygdala, cortex and striatum[19] and its loss leads to deficits in neurogenesis and brain function. *Adgrb1*$^{exon2-/-}$ mice that lack full-length BAI1 expression have reduced expression of post-synaptic density 95 (PSD-95) and exhibit deficits in spatial learning & memory, and alterations in synaptic plasticity.[13] During brain development, these mice show a decrease in brain weight with reduced neuron density and increased apoptosis in the hippocampus.[17] They also show significant social behavior deficits and increased vulnerability to seizures.[17]

BAI1 is located in the post-synaptic membrane of neurons and its knockdown leads to the immature development of dendritic spines and excitatory synapse formation.[14, 20] BAI1 regulates synaptogenesis by its interaction with Neuroligin-1 (NL-1), a synaptic organizer and through the recruitment of the Par3/Tiam1 polarity complex.[6, 14] In the postsynaptic density (PSD), BAI1 promotes RhoA activation through coupling to $G\alpha_{12/13}$ to maintain synaptic plasticity via microtubules and microfilaments rearrangement.[9] Also, an interaction between BAI1 and BCR (breakpoint cluster region), restricts dendritic growth via RhoA activation.[20] The extracellular domain of BAI1 is involved in promoting synaptogenesis and inhibiting dendrite and axonal growth through a high-affinity interaction between RTN4R and the third TSR domain of BAI1.[5]

BAI1 is encoded by the *ADGRB1* (adhesion G protein-coupled receptor B1) gene located on human chromosome 8q24.3, spanning ~95.4 kilobase pairs and featuring 31 exons. Large genes can give rise to multiple protein isoforms with different functions and tissue expression profiles. Such proteomic diversity stems from a variety of mechanisms, including alternative transcription start sites driven by alternative promoters (APs),[21, 22] alternative splicing (AS) of pre-mRNAs[23] alternative polyadenylation[24] and alternative translation initiation.[25] In eukaryotes, these regulatory mechanisms are mainly responsible for structural and functional diversity.[26] In normal physiology, alternative promoters can allow for differential expression in different cell types or tissues and are one of the dominant contributors to proteome diversity.[27] Transcriptional heterogeneity can give rise to protein isoforms with rearranged domains and altered functions[28] as well as totally different proteins with distinct functionality[29]. Alternative promoter usage and

alterations in the splicing machinery can also contribute to disease formation, particularly in cancer.[30, 31]

Little is known about the existence of BAI1 isoforms and the roles they may play in the regulation of the above multiple biological processes, either physiologically or pathologically. Here, we investigated the different transcripts originating from the *ADGRB1* gene and resulting protein isoforms of BAI1 and evidenced new BAI1 isoforms that originate from a heretofore unknown alternative promoter in intron 17 of *ADGRB1*.

## 2. Materials and Methods

The study was conducted in accordance with the Basic & Clinical Pharmacology & Toxicology policy for experimental and clinical studies.[32]

### Cell Culture

Human glioma stem cell (GSC) lines (GSC4, GSC18) were cultured in neural stem cell medium as described.[33] Human cell lines derived from glioblastoma (LN229) and SHH (ONS76 and DAOY) and Group 3 (D556 and D425) medulloblastoma and 293T cells were obtained from ATCC and cultured in DMEM as described.[8] The cells were regularly tested for mycoplasma and their identity was verified by STR profiling.

### Mice

The full-length BAI1 knockout mouse model in which Exon-2 (containing the translation initiation codon) was replaced by homologous recombination with a promoterless *LacZ* (β-galactosidase) gene, a *neo* gene (neomycin resistance gene) followed by a stop codon and poly-A tail has been previously described.[13] Protein extracts from C57BL/6J wild type and *Adgrb1* $^{exon2-/-}$ mouse brain samples were prepared as described.[17] Protocols for in vivo experiments were approved by the Institutional Animal Care and Use Committee (IACUC).

### Cell transfection and Small-interfering (si)-RNA knockdown

Transient transfection of MB cells with pcDNA-BAI1[8] was performed with lipofectamine 3000 (Thermo Fisher Scientific, L3000-015) as prescribed. Briefly, 70-90% confluent cells were cultured in a 6 well plate for transient transfection and siRNA-mediated knockdown. Plasmid DNA (2500 ng/well) diluted in Opti-MEM (Gibco, 31985070), was mixed with enhancer reagent p3000, and lipofectamine 3000. After 10 min incubation, the DNA-lipid complex was added to each well and incubated for 2-3 days. For knock-down experiments, siRNA transient transfection was performed with RNAiMAX (Thermo Fisher Scientific, 13778075) as prescribed. The following siRNAs were used: siBAI1#1 (Ambion, 4392420, ID: s1870) targeting exon 23, and siBAI1#2 (Ambion, AM16708, ID:147207) targeting exon 21 of *ADGRB1* (NM_001702.2).

### Immunoblotting

Western blots were performed on un-boiled protein extracts (see Suppl. Materials and Methods). Antibodies targeting the following proteins were used: BAI1 C-terminal epitope 1537-1567AA (ABCEPTA, AP8170a; 1:1,000), GAPDH (Santa Cruz, sc-47724; 1:1,000), β-Actin (Santa Cruz, sc-69879; 1:1,000). ImageJ 1.54d (NIH) software was used to quantify bands in immunoblots. SnapGene 3.1.4 was used for the prediction of m.w. of different BAI1 isoforms.

**RNA Isolation and quantitative reverse transcription PCR (RT-PCR)**

Total RNA from cells and tissues was isolated using TRIzol reagent (Ambion, 15596018) as prescribed. The isolated RNA was diluted to 50 ng/µl. The RNA was then reverse transcribed into cDNA using iScript Reverse Transcription Supermix (Bio-Rad, 1708841). RT-PCR was performed using gene-specific primers (see Suppl. Materials and Methods) and Taq DNA polymerase (Qiagen, 201203).

**Dual luciferase reporter assay**

The 2210 and 1190 bp long genomic fragments in the 3′ region of human intron 17, were amplified by PCR. Conditions for PCR amplification were: 95°C for 2 minutes, followed by 35 cycles comprising 95°C for 30 seconds, 64°C for 60 seconds and 68°C for 3 minutes, with a final extension at 68°C for 10 minutes. PCR amplified fragments were cloned into the pGL2-basic-firefly luciferase plasmid using the XhoI and HindIII restriction sites. The pGL2-CMV-firefly luciferase plasmid was used as a positive control. LN229 and HEK293T cells were plated in 24-well plates at $1x10^5$ cells/well. At ~80 % confluency, a total of 1µg firefly luciferase plasmids and a transfection control plasmid (pLV-EF1-RenillaLuc plasmid) were transiently transfected into LN229 and HEK293T cells. After 48 hours of transfection, luciferase activity was measured in 20/20n Single Tube Luminometer (Turner BioSystems, 2030-000) using a Dual-Luciferase Reporter Assay System kit (Promega, E1910) as prescribed. Firefly luciferase activity was normalized to the activity of Renilla luciferase. The experiments were performed in three independent biological replicates and reactions of each sample were carried out in triplicate.

**Computational biology analyses**

Normalized mRNA expression levels (nTPM) of *ADGRB1* in different regions of human and mice brain, were analyzed through The Human Protein Atlas (www.proteinatlas.org)[34]. Histone ChIP-seq data from the ENCODE portal (www.encodeproject.org)[35, 36] were queried for histone

modification marks with the following dataset identifiers: ENCSR875PYX for H3K4me3 ChIP-seq, ENCSR532SRK for H3K27ac ChIP-seq, in dorsolateral prefrontal cortex tissue of human; ENCSR258YWW for H3K4me3 ChIP-seq, ENCSR094TTT for H3K27ac ChIP-seq and ENCSR310MLB for ATAC-seq, in forebrain tissue of mice. ATAC-seq data for the caudate region of the human brain (ID: SRX8020343) was extracted from the ChIP-Atlas database (chip-atlas.org)[37]. For the visualization of ChIP-seq and ATAC-seq data, we used Integrative Genomics Viewer (IGV-V2.12.2, Broad Institute). Long-read (LR) RNA seq of breast cancer data was analyzed with Breast Cancer Long Read Transcriptome (brca-isoforms.jax.org).[38] Further analysis of full-length transcript sequencing data of human cortex[39] (SRA: PRJNA664117) and mouse cortex[39] (SRA: PRJNA663877) was visualized by using the UCSC genome browser. Gene promoter candidate regions were predicted with Ensembl Release 109 (useast.ensembl.org).[40] Alternative transcription start sites were predicted through the DataBase of Transcriptional Start Sites (DBTSS; dbtss.hgc.jp). [41, 42] Exon-specific expressions of *ADGRB1* mRNA variants were visualized by the Genotype-Tissue Expression (GTEx) database (www.gtexportal.org/home/).[43] The median read counts per base for each exon were calculated based on a collapsed gene model where all *ADGRB1* mRNA variants are combined into a single transcript by using a heatmap.

**Statistical analyses**

GraphPad Prism 9.0 (GraphPad, La Jolla, CA, USA) was used for biostatistics. Data are expressed as mean ± SEM. One-way ANOVA with Tukey's multiple comparisons correction was performed to identify statistical differences. $P < 0.05$ was considered statistically significant.

## 3. Results

### Evaluation of BAI1 protein isoforms in human and mouse

At first, we examined the mRNA expression levels of *ADGRB1* in the Human protein atlas (HPA) with a particular focus on the brain.[34] Analysis of 12 main regions of the human brain evidenced the highest expression in the cerebral cortex, and lowest in the cerebellum and spinal cord (Figure 1A).

Examination of BAI1 protein expression with an anti-C-terminal antibody in normal human brain tissue revealed major isoforms of ~170-200 kDa, representing isoforms of full length BAI1 (FL-BAI1) with variable post-translational modifications, some of which might have a shortened N-terminus due to MMP14 cleavage (Figure 1B). FL-BAI1 expression was strongly reduced in adult glioblastoma and pediatric medulloblastoma cells, as FL-BAI1 is silenced in these cancers.[8, 44, 45] Two additional bands were observed in the brain with m.w of ~70 and 75 kDa and their expression was maintained in the tumor cells. RNA interference with siRNAs targeting *ADGRB1* exons 21 or 23 reduced the expression of both isoforms, confirming they are derived from alternative transcripts and are not the result of non-specific antibody binding (Figure 1C). Transfection of FL-BAI1 cDNA in medulloblastoma cell lines restored FL-BAI1 and evidenced two smaller isoforms of ~70 and 75 kDa, both representing cleavage products of FL-BAI1 (Figure 1D). Altogether these data suggest that short BAI1 isoforms are produced in human brain cells, some the result of alternative transcripts and others from proteolytic cleavage of higher molecular weight BAI1.

As in human, analysis of *Adgrb1* expression in mouse brain showed the highest mRNA levels in the cerebral cortex, moderate levels in the olfactory bulb and lowest in the pituitary (Figure 1E). Examination of BAI1 protein levels in wild type (WT) mice showed predominant bands in the ~160-200 kDa and ~70-75 kDa size ranges, with additional smaller isoforms around ~40 kDa (Figure 1F). A similar analysis of BAI1 isoforms in genetically engineered mice where exon 2 (which contains the ATG in full length *Adgrb1* transcripts) was replaced with a *LacZ* cDNA showed that the ~160-200 kDa isoforms and some of their smaller cleavage products disappeared as expected (Figure 1G). The shorter ~70 and 75 kDa and ~40 kDa BAI1 isoforms were retained in the cerebrum, cerebellum, and olfactory bulb, suggesting that they originate from new transcripts starting downstream of exon 2. Interestingly, the pituitary and heart showed new BAI1 isoforms in the *Adgrb1* $^{exon\ 2-/-}$ mice (Figure 1F, 1G, and figure S1), which will require future

investigation. Overall, these data show strong similarities between human and mouse brain BAI1 isoforms, with the expression of shorter ~70-75 kDa isoforms likely derived from alternative transcripts.

**Analysis of the human *ADGRB1* gene transcriptional diversity**

To identify alternative promoters and potential transcriptional start sites across *ADGRB1*, we queried ChIP-seq results mapping the distribution of histone modification marks across the gene.[46] We further extended this data with ATAC-seq, which identifies regions of open chromatin. In the dorsolateral prefrontal cortex of human brain, ChIP-seq data revealed strong enrichment of H3K4me3 and H3K27ac, two marks of transcriptionally active promoter regions,[47] near exons 1 and 18, with minor peaks in introns 1 and 2 (Figure 2A, Top panel). ATAC-seq peaks (in the caudate region) were also found to overlap with those regions (Figure 2A, Top panel). Ensembl identified a candidate promoter region encompassing exon 1 and part of intron 1 (Figure 2A, Top panel, red rectangle).

To examine how these open/active chromatin regions direct the transcriptional heterogeneity of human *ADGRB1*, we aligned them with transcripts identified by long-read isoform sequencing (Iso-seq) in the cerebral cortex of adult and fetal human brain[39] (Figure 2A). We identified a total of 11 mRNA variants of *ADGRB1* in the human brain (Figure 2A, Middle panel). Three (3) of the novel mRNAs (7/3, 8/22 and 8/23) have new alternative first exons starting in the 3′ end of intron 17. The alternative first exon of transcript 7/3 has a new start codon (ATG) which is in-frame with exon 18 codons of *ADGRB1* and can be translated to generate a 699 AA hBAI1 (human BAI1) isoform (predicted size: 76.9 kDa). The other two (8/22 and 8/23) with an alternative first exon in intron 17 splice to exon 18, but use a translation start codon in exon 19, forming a 644 AA long hBAI1 isoform (predicted size: 70.8 kDa). Shorter transcripts that do not associate with an H3K4me3 mark were also observed: transcript 14.2 starts from an alternative new exon in intron 27 and encodes 448 AA (predicted size: 48.8kDa), and 5 shorter transcripts start near exons 27 (15/2; 16/2) or 28 (18/4, 19/27, 20.2) and will need future study. Long-read (LR) RNA seq of breast tumors also evidenced transcripts originating in intron 17 (Figure S2), showing this is not unique to the brain.[38]

To further interrogate the region near exon 18 for potential transcription start sites (TSS), we queried the DBTSS database.[41, 42] Multiple candidate TSS were identified ~1000 bp upstream of

exon 18 in the adult and fetal normal human brain and in the HEK293 cell line (Figure 2A, Bottom panel).

In addition, to examine whether the longer and shorter *ADGRB1* transcripts lead to different exon usage in different tissues, we queried the Genotype-Tissue Expression database[43] of exon-specific reads in a collapsed gene model (Figure 2B). This evidenced higher median read counts per base after exon 18 in the brain and pituitary, further supporting the existence of transcripts starting around exon 18 in human *ADGRB1*. It also evidences denser reads in exons 29-31, possibly resulting from the short 3′ transcripts (18/14, 19/17, 20/2).

In combination, these analyses support the existence of an alternative promoter region at the intron 17/ exon 18 boundary in the human brain, which is predicted to generate several transcripts that would encode hBAI1 isoforms of 644 and 699 amino acids (predicted sizes: 70.8 and 76.9 kDa, respectively), which may correspond to the ~70 and 75 kDa bands observed by western blotting.

**Analysis of the mouse *Adgrb1* gene transcriptional heterogeneity**

Two main transcriptionally active open chromatin regions are revealed by ATAC-seq and display H3K4me3 and H3K27ac markers: one located near exon 1 and the other straddling the intron 17/exon 18 boundary and both are predicted to be gene promoters by the Ensembl genome browser (Figure 3A, Top panel). DBTSS database analysis identified some putative TSS in a 1,000 bp region near exon 18 in the mouse embryo (Figure 3A, Bottom Panel). A third region upstream of exon 3 also revealed open chromatin but with weak H3K4me3 marks, perhaps representing a weaker promoter as it matches the start of transcript 12/3.

Analysis of the Iso-seq data of mouse cerebral cortex[39] revealed a total of 21 *Adgrb1* transcripts (Figure 3A, middle panel): 5 transcribed from exons 1 or 2 and 14 from the 3′ end of intron 17, matching the two main open/active chromatin areas. Ten of the latter transcripts have new exons in intron 17 that splice to exon 18. Three (42/19, 43/7 and 44/2) have new first exons that are a 5′ extension of exon 18. All of these use a start codon in exon 19. The last transcript (41/5) has an alternative new exon in intron 17 that splices to the middle of exon 18 and uses a new ATG.

To independently validate the existence of these variant mRNA transcripts in the whole mouse brain, we used reverse transcriptase PCR on wild type and *Adgrb1* $^{exon2-/-}$ mice.[13] RT-PCR with a series of primers spanning exons 2 to 31 failed to detect any transcripts between exons 2 and 17 in

exon2-deficient mice (Figure 3B). However, clear mRNA expression was detected from exons 18 to 31, supporting the existence of transcripts initiated from intron 17. To validate the existence of specific transcripts, we used RT-PCR with forward primers starting in the newly predicted exons in intron 17 (new exon 1) and reverse primers in exon 18 or 23. We generated a primer set for new exon 1 that can detect putative transcripts 42/19, 43/7 and 44/2 and were successful in amplifying cDNA in wt and *Adgrb1* $^{exon2-/-}$ mice (Figure 3C). Careful examination of the reading frame of each newly spliced transcript suggests that most of the mRNAs transcribed from the 3′ end of intron 17 have a start codon in exon 19 and are predicted to generate a mBAI1 (mouse BAI1) isoform of 642AA (predicted size: 70.5 kDa). In contrast, transcript 41/5 is translated from a new ATG and generates a 700AA mBAI1 isoform (predicted size: 76.4kDa). We failed in confirming this transcript by RT-PCR in the mouse brain, likely due to the high GC content (89%) in alternative first exon and difficulty in generating unique primers differentiating the alternative exons, so other approaches to confirm the remaining transcripts are warranted.

In conclusion, our data evidence the presence of an alternative promoter in the 3′ region of intron 17 of the *Adgrb1* gene, as for human *ADGRB1*. This mouse promoter can initiate transcription of multiple mRNA transcripts predicted to encode mBAI1 proteins of 642 and 700AA (predicted size: 70.5kDa, 76.4kDa respectively).

**Functional testing of alternative promoter activity in human cells**

To functionally evaluate the presence of an alternative promoter at the 3′ end of intron 17 in human *ADGRB1*, we sub-cloned two overlapping genomic fragments containing the new exon region and 1-2 Kb upstream sequence (1190 and 2210bp) (Figure 4A) into a luciferase expression vector lacking promoter and enhancer regions. Transient transfection in neuronal (293T) and glial (LN229 glioblastoma) cells revealed that the longer fragment elicited robust transcriptional activation of luciferase in both cell lines, while the shorter one showed modest activation in 293T cells and none in LN229 cells (Figure 4 B). These results demonstrate that the 2,210 bp distal region of exon 17 can serve as a gene promoter and contains critical regions for transcriptional activation lacking in the shorter construct.

**The alternative promoter in intron 17 leads to the formation of ~70-75 kDa hBAI1 isoforms**

As proteolytic fragments of full-size BAI1 protein and the shorter BAI1s generated from the heretofore cryptic promoter in intron 17 appear similar in size (~70-77 kDa), we wanted to gain a finer understanding of structural differences at their N-termini. FL-BAI1 human mRNAs have 31 exons and translation starts from exon-2 (Figure 5, Top Panel), generating a predicted protein of 1584 amino acids (173.5 kDa), which migrates on western blots at ~160-200 kDa as it can undergo post-translational modifications and cleavage at the MMP14 site (Figure 1). FL-BAI1 has a signal peptide, RGD motif, TSRs, HBD, and GAIN domain upstream of the GPS region (Figure 5, Middle-Left Panel). Autoproteolysis at the GPS region results in a truncated BAI1 membrane receptor of 658 amino acids (predicted size of 72.2 kDa) with a 19 AA long N-terminal stalk,[6] a 7TM domain, and a C-terminal region (Figure 5, Bottom-Left Panel). The cleaved extracellular N-terminal fragment (NTF) remains non-covalently associated with the 7TM domain of the membrane-associated C-terminal fragment (CTF) till an unknown stimulus can detach it from the receptor and lead to receptor activation by a conformational change, possibly triggered by the remaining stalk acting as a cryptic agonist.[48-50]

Two short BAI1 isoforms are encoded by human and mouse transcripts generated from the intron 17 promoter. The shortest one in human is 644 AA long (theoretical size of 70.9 kDa) and does not undergo additional autoproteolysis as it lacks the GPS site. It starts with MEKAT at its N-terminus, and the remaining 7TM and intracellular motifs are the same as for FL-BAI1 (Figure 5, Bottom-Right Panel). This shorter isoform lacks a signal peptide and cannot form a peptide agonist ("Stachel"), which may affect its cellular localization and signaling function. The second one is 699 AA long and is encoded by three of the novel human mRNAs (7/3 in cerebral cortex and PB.34262.1 and 62.3 in breast tumors) that have a new start codon (ATG) in their alternative first exon that is in-frame with exon 18 codons of full-length *ADGRB1*. The predicted size of this isoform is 76.9 kDa and it features a 60AA long N-terminus (Figure 5 and figure S2) with 7 novel amino acids (white circle in Figure 5 and figure S2) that are absent in FL-BAI1. It has a GPS site but lacks the GAIN domain to induce autoproteolysis.

Altogether, our findings establish the presence of an alternative promoter at the 3′ end of intron-17 of both human and mouse *ADGRB1/Adgrb1* genes, and transcription and translation from this alternative promoter led to the formation of two shorter BAI1s with relatively short N-termini.

## 4. Discussion

Class B human GPCR are comprised of 33 ADGR cell surface receptors with very long N-termini that enable interactions with neighboring cells and the extracellular environment.[51] While all ADGRs share a GAIN domain and GPS, their N-termini are highly variable and harbor different structural domains organized in a modular fashion, allowing them to interact with a variety of binding partners. Hence, the extracellular domains of ADGRs can mediate individual interactions with specific binding partners to carry out distinct functions or serve as a scaffold for the assembly of multiple partner proteins into a complex. ADGR receptors are encoded by large genes that can express a variety of transcripts encoding multiple protein isoforms, leading to variation in the functional domains they carry, thus impacting overall protein function.[23] A comprehensive understanding of ADGR isoforms made in different tissue types is important to unravel their function in physiology and disease.

The purpose of this study was to explore the proteomic diversity of BAI1 and start to catalog the different protein isoforms observed by immunoblotting and decipher their molecular origin. Full-length BAI1 (1584 amino acids; predicted size of 173.5kDa) migrates at an apparent size of ~160-200 kDa on immunoblots. Identifying individual ADGR isoforms on Western blots can be challenging due to the extensive post-translational modifications that can occur, particularly glycosylation. Also, excessive loading of SDS on the helical hydrophobic 7-transmembrane segments can create shifts between expected and observed protein sizes.[52, 53]

Further size variation can stem from proteolytic cleavage and prior studies had demonstrated that BAI1 can undergo two main proteolytic events.[16, 54] Cleavage between TSR1 and TSR2 by cancer-associated protease matrix metalloproteinase 14 (MMP14) yields an anti-angiogenic N-terminal (Vasculostatin-40; predicted size of 34.7/35.0 kDa for human/mouse) which migrates at ~37-40 kDa and a remaining truncated membrane receptor (predicted size of 138.8 kDa), which migrates at ~160 kDa. Autoproteolytic cleavage at the GPS site generates an anti-angiogenic and anti-tumorigenic extracellular NTF (Vasculostatin-120; predicted size of 101.3/101.5kDa for human/mouse) that migrates at ~120 kDa and the remainder membrane-bound CTF containing a “Stachel” peptide, 7TM domain, and C-terminal domain (predicted size of 72.2/71.8 kDa in human/mouse) that migrates at ~75 kDa. Thus, through two proteolytic cleavage events 5 different isoforms of BAI1 can be generated: three that act at the cell surface and two in the cell environment

where they might engage different targets as they will diffuse differentially due to their size and domain adhesiveness. They can be easily detected upon cDNA transfection in BAI1 silent cells using N- and C-terminal antibodies in conditioned media and cell extracts. Engineered point mutations can help define the cleavage site location and the role of surrounding amino acids. For example, we previously showed that an S927A substitution abrogated, while a phospho-mimetic S927D substitution increased the cleavage of BAI1 in glioma cells.[16, 54] An initial immunoblotting survey of the brain and other organ extracts revealed much more complexity and suggested that further isoforms are present in different tissues. The brain showed the expected full length BAI1 at ~ 160-200 kDa and major isoforms at ~70-75 kDa and smaller ones at ~35-45 kDa. Weaker bands were also observed, and some might be non-specific signals, while others are likely different isoforms expressed at higher levels in other tissues. To distinguish isoforms resulting from proteolytic cleavage of full length BAI1 versus those generated from potential alternative promoters we also examined the organs of mice genetically engineered with a deletion of exon 2 where the translational start site of the full length BAI1 is located. These data evidenced that major brain isoforms of ~70-75 kDa and ~35-45 kDa were still present, showing they did not originate from proteolytic cleavage of FL-BAI1, complicating the interpretation of bands on Western blots. Unexpectedly, we also found isoforms that were only expressed in exon 2-deleted mice in cerebrum (~60-65 kDa), pituitary (~100 kDa) and heart (~160 kDa and ~65 kDa). These may derive from other cryptic promoters and/or splicing events that are inactive in wt mice, and induced upon FL-BAI1 loss, suggesting a potential regulatory network.

To further explore the origin of the BAI1 brain isoforms and determine the potential location of alternative transcription start sites and alternative splicing events, we queried long-read RNAseq data and overlayed it with information on active chromatin on the mouse and human *ADGRB1* genes. These analyses support the existence of an alternative promoter region (~2.5 kb) at the intron 17/exon 18 boundary and transcription start site prediction software and RNAseq reads supports the existence of a cluster of transcription start sites in that region, yielding a series of mRNAs starting with new untranslated alternative exons (5' UTRs) that splice with exon 18. To independently validate promoter activity of this intronic region, we confirmed the existence of one of the mouse transcripts by reverse transcription on exon 2-deficient mouse brain and showed that a genomic fragment of 2.2 kb just upstream of human exon 18 was sufficient to direct transcription of a reporter gene in both glial and neuronal cells. Further work is warranted to define which

transcription factors are essential to activate this promoter region in different brain regions and other tissues.

Examination of the splicing pattern of the three human transcripts starting in intron 17 showed that they splice to exon 18 and retain exons 19-31. Other shorter transcripts starting with exons 25, 27 or 29 were also present. Some of these transcripts undergo alternative splicing with the addition of a new exon in intron 27 that adds 45AA (4.6kDa) and will generate further diversity. In mouse, 14 transcripts start from new exons in intron 17, some of which carry a new exon in intron 24, or exon rearrangements downstream of exon 26. The short transcript variants starting with exons 25-29 were not found in mouse. Whether this represents a biological difference between the two species or results from the different techniques used in transcript mapping awaits further clarification. The abundance and functions of these BAI1 isoforms in diverse cells and tissue types will require further study.

The structure of the two human BAI1 isoforms produced from the alternative promoter in intron 17 are very similar to that of BAI1 resulting from proteolytic cleavage at the GPS site (Figure 5) with slight variation in the size of the small N-terminal extracellular stalk. Autocatalytic proteolysis of FL-BAI1 induced by the GAIN domain leaves a 19AA stub that is part of a cryptic peptide (STFAILAQLSADANMEKAT) that can serve as an agonist to activate the receptor.[6] One short isoform of hBAI1 (predicted size: 70.8 kDa) is translated from an in-frame ATG in exon 19 and has a minimal N-terminus (MEKAT), and is expected to lack agonistic autoactivation capability. The second type of short isoform (predicted size: 76.9 kDa) is translated from a start codon in intron-17 that is in frame with exon-18 reading frame (Supplementary figure S2) yielding a BAI1 isoform with a 60AA N-terminus. This isoform retains the GPS site, but whether it can undergo autoproteolysis is uncertain as it lacks the GAIN domain. It is tempting to speculate as to the function of these new isoforms. They might lack the receptor activity that is dependent upon the conformational change induced by the cryptic agonist but may retain other functions such as the ability to serve as a docking station for PDZ-containing proteins. Since they have a full 7-TM, they can likely bind the cleaved extracellular NTF of FL-BAI1 (Vstat120), which can repress the CTF through non-covalent association.[54-56] In situations where both isoforms are co-expressed, the shorter BAI1 might regulate the activity of FL-BAI1 by serving as a sink for Vstat-120 and thereby limiting the release of this anti-angiogenic molecule in the extracellular milieu. As cleaved CTFs

can associate with NTFs released from other ADGRs,[56] one may even speculate that shorter BAI1 isoforms may trap NTF of BAI2, BAI3 or other ADGRs and in this way activate them. Such heterotypic NTF-CTF associations may also have new signaling functions that await further discovery. At first, it will be important to examine whether the new ~70-75 kDa BAI1 isoforms are expressed at the cell surface as they lack a signal peptide. This may not be required, as a prior study showed that a BAI1 deletion mutant lacking the entire N-terminus including the "Stachel" trafficked to the plasma membrane and exhibited robust signaling activity.[50] In this context it is useful to mention that the BAI1 C-terminus harbors a domain (AA 1489-1506) predicted to function as a nuclear localization sequence (NLS).[12] It will be of particular interest to determine whether any of the newly identified BAI1 isoforms carrying this NLS preferentially traffic to the nucleus where they might exert novel functions.

Finally, this study further highlights the importance of fully characterizing the impact of gene knockdown/knockout strategies in ADGR research, so as to fully understand their biological impact. For example, our prior studies have found that *Adgrb1* $^{exon2\ -/-}$ mice have decreased expression of post-synaptic protein 95 (PSD95) expression, a scaffolding protein important for dendritic spines/synapse formation, and exhibit severe deficits in synaptic plasticity and hippocampus-dependent spatial learning and memory.[13] We can now conclude that these functions are dependent upon the expression of full length BAI1 and are not rescued by the shorter isoforms. Yet, the dendritic arborization and spine density of these mice appeared similar to wt mice, contrasting with other studies targeting BAI1 acutely in neurons with shRNAs where clear spine deficits were observed.[57] While potential reasons underlying these divergent observations have already been proposed,[57] this study suggests that the contribution of shorter BAI1 isoforms to dendritic spine maintenance may also be worth considering. Our data show that the expression of several isoforms is permanently altered in *Adgrb1* $^{exon\ 2\ -/-}$ mice that lack FL-BAI1, suggesting a compensation effect. These isoforms might directly sustain spine development or do so indirectly. For example, if some can bind to the NTF of BAI3, one might hypothesize that they might indirectly rescue spine deficits by enhancing activation of this closely related receptor that also regulates spinogenesis.[5] It will be of interest to examine whether acute transient knockdown of BAI1 with shRNAs also leads to a similar reprogramming of isoform expression.

G-protein-coupled receptors (GPCRs) are the largest family of membrane-bound receptors and are the targets of ~35% of clinically-approved drugs.[58] BAI1 is a member of class B GPCRs that are involved in multiple physiological processes including brain development, inflammation, phagocytosis, and diseases such as neurological disorders and tumorigenesis.[10, 17, 59] A full characterization of all isoforms of BAI1 will foster a better understanding of this interesting receptor's role in health and disease and evaluate its potential as a drug target. Identifying disease-associated isoforms and distinguishing uniqueness in their structures will be critical to target the relevant ones and avoid collateral side effects on others.

## Acknowledgements

This work was supported in part by grants from the NIH (R01 NS096236, CA235162, and NS117666 to EGVM and P30 CA13148 to the O'Neal Comprehensive Cancer Center) and in part by JSPS Overseas Research Fellowships to TY. We thank Dr. Hiroaki Wakimoto (Harvard University, Cambridge, MA) for providing the glioma stem cells, Drs. Antonio Di Stasi, Erin Eun-Young Ahn and Virginea De Araujo Farias for experimental advice and/or proofreading.

## Competing Interests

The authors declare no conflict of interest.

## Author Contributions

E.G.V.M. conceived the project. R.R.P. and T.Y. performed most of the experiments. R.R.P., T.Y. and E.G.V.M. wrote the manuscript. All authors provided advice and comments on the manuscript.

## Data availability

Data available on request.

## Conflict of Interest Statement

The authors declare no conflict of interest. EGVM is a founder and shareholder of OncoSpherix, Inc. (not related to the current study)

## References


1. Nishimori H, Shiratsuchi T, Urano T, Kimura Y, Kiyono K, Tatsumi K, et al. A novel brain-specific p53-target gene, BAI1, containing thrombospondin type 1 repeats inhibits experimental angiogenesis. Oncogene. 1997;15(18):2145-50.10.1038/sj.onc.1201542
2. Fredriksson R, Lagerstrom MC, Lundin LG, Schioth HB. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. Mol Pharmacol. 2003;63(6):1256-72.10.1124/mol.63.6.1256
3. Kaur B, Cork SM, Sandberg EM, Devi NS, Zhang Z, Klenotic PA, et al. Vasculostatin inhibits intracranial glioma growth and negatively regulates in vivo angiogenesis through a CD36-dependent mechanism. Cancer Res. 2009;69(3):1212-20.10.1158/0008-5472.CAN-08-1166
4. Park D, Tosello-Trampont AC, Elliott MR, Lu M, Haney LB, Ma Z, et al. BAI1 is an engulfment receptor for apoptotic cells upstream of the ELMO/Dock180/Rac module. Nature. 2007;450(7168):430-4.10.1038/nature06329
5. Wang J, Miao Y, Wicklein R, Sun Z, Wang J, Jude KM, et al. RTN4/NoGo-receptor binding to BAI adhesion-GPCRs regulates neuronal development. Cell. 2021;184(24):5869-85 e25.10.1016/j.cell.2021.10.016
6. Tu YK, Duman JG, Tolias KF. The Adhesion-GPCR BAI1 Promotes Excitatory Synaptogenesis by Coordinating Bidirectional Trans-synaptic Signaling. J Neurosci. 2018;38(39):8388-406.10.1523/JNEUROSCI.3461-17.2018
7. Harmar AJ. Family-B G-protein-coupled receptors. Genome Biol. 2001;2(12):REVIEWS3013.10.1186/gb-2001-2-12-reviews3013
8. Zhu D, Osuka S, Zhang Z, Reichert ZR, Yang L, Kanemura Y, et al. BAI1 Suppresses Medulloblastoma Formation by Protecting p53 from Mdm2-Mediated Degradation. Cancer Cell. 2018;33(6):1004-16 e5.10.1016/j.ccell.2018.05.006
9. Stephenson JR, Paavola KJ, Schaefer SA, Kaur B, Van Meir EG, Hall RA. Brain-specific angiogenesis inhibitor-1 signaling, regulation, and enrichment in the postsynaptic density. J Biol Chem. 2013;288(31):22248-56.10.1074/jbc.M113.489757
10. Moon SY, Shin SA, Oh YS, Park HH, Lee CS. Understanding the Role of the BAI Subfamily of Adhesion G Protein-Coupled Receptors (GPCRs) in Pathological and Physiological Conditions. Genes (Basel). 2018;9(12).10.3390/genes9120597
11. Das S, Owen KA, Ly KT, Park D, Black SG, Wilson JM, et al. Brain angiogenesis inhibitor 1 (BAI1) is a pattern recognition receptor that mediates macrophage binding and engulfment of Gram-negative bacteria. Proc Natl Acad Sci U S A. 2011;108(5):2136-41.10.1073/pnas.1014775108
12. Knapp B, Roedig J, Boldt K, Krzysko J, Horn N, Ueffing M, et al. Affinity proteomics identifies novel functional modules related to adhesion GPCRs. Ann N Y Acad Sci. 2019;1456(1):144-67.10.1111/nyas.14220
13. Zhu D, Li C, Swanson AM, Villalba RM, Guo J, Zhang Z, et al. BAI1 regulates spatial learning and synaptic plasticity in the hippocampus. J Clin Invest. 2015;125(4):1497-508.10.1172/JCI74603
14. Duman JG, Tzeng CP, Tu YK, Munjal T, Schwechter B, Ho TS, et al. The adhesion-GPCR BAI1 regulates synaptogenesis by controlling the recruitment of the Par3/Tiam1 polarity complex to synaptic sites. J Neurosci. 2013;33(16):6964-78.10.1523/JNEUROSCI.3978-12.2013
15. Good DJ, Polverini PJ, Rastinejad F, Le Beau MM, Lemons RS, Frazier WA, et al. A tumor suppressor-dependent inhibitor of angiogenesis is immunologically and functionally

indistinguishable from a fragment of thrombospondin. Proc Natl Acad Sci U S A. 1990;87(17):6624-8.10.1073/pnas.87.17.6624
16. Cork SM, Kaur B, Devi NS, Cooper L, Saltz JH, Sandberg EM, et al. A proprotein convertase/MMP-14 proteolytic cascade releases a novel 40 kDa vasculostatin from tumor suppressor BAI1. Oncogene. 2012;31(50):5144-52.10.1038/onc.2012.1
17. Shiu FH, Wong JC, Yamamoto T, Lala T, Purcell RH, Owino S, et al. Mice lacking full length Adgrb1 (Bai1) exhibit social deficits, increased seizure susceptibility, and altered brain development. Exp Neurol. 2022;351:113994.10.1016/j.expneurol.2022.113994
18. Hochreiter-Hufford AE, Lee CS, Kinchen JM, Sokolowski JD, Arandjelovic S, Call JA, et al. Phosphatidylserine receptor BAI1 and apoptotic cells as new promoters of myoblast fusion. Nature. 2013;497(7448):263-7.10.1038/nature12135
19. Sokolowski JD, Nobles SL, Heffron DS, Park D, Ravichandran KS, Mandell JW. Brain-specific angiogenesis inhibitor-1 expression in astrocytes and neurons: implications for its dual function as an apoptotic engulfment receptor. Brain Behav Immun. 2011;25(5):915-21.10.1016/j.bbi.2010.09.021
20. Duman JG, Mulherkar S, Tu YK, Erikson KC, Tzeng CP, Mavratsas VC, et al. The adhesion-GPCR BAI1 shapes dendritic arbors via Bcr-mediated RhoA activation causing late growth arrest. Elife. 2019;8.10.7554/eLife.47566
21. Landry JR, Mager DL, Wilhelm BT. Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet. 2003;19(11):640-8.10.1016/j.tig.2003.09.014
22. de Klerk E, t Hoen PA. Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. Trends Genet. 2015;31(3):128-39.10.1016/j.tig.2015.01.001
23. Knierim AB, Rothe J, Cakir MV, Lede V, Wilde C, Liebscher I, et al. Genetic basis of functional variability in adhesion G protein-coupled receptors. Sci Rep. 2019;9(1):11036.10.1038/s41598-019-46265-x
24. Arora A, Goering R, Lo HYG, Lo J, Moffatt C, Taliaferro JM. The Role of Alternative Polyadenylation in the Regulation of Subcellular RNA Localization. Front Genet. 2021;12:818668.10.3389/fgene.2021.818668
25. Kochetov AV. Alternative translation start sites and hidden coding potential of eukaryotic mRNAs. Bioessays. 2008;30(7):683-91.10.1002/bies.20771
26. Wen H, Chen W, Chen Y, Wei G, Ni T. Integrative analysis of Iso-Seq and RNA-seq reveals dynamic changes of alternative promoter, alternative splicing and alternative polyadenylation during Angiotensin II-induced senescence in rat primary aortic endothelial cells. Front Genet. 2023;14:1064624.10.3389/fgene.2023.1064624
27. Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, et al. Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 2006;16(1):55-65.10.1101/gr.4039406
28. Pelechano V, Wei W, Steinmetz LM. Extensive transcriptional heterogeneity revealed by isoform profiling. Nature. 2013;497(7447):127-31.10.1038/nature12121
29. Chin L, Pomerantz J, DePinho RA. The INK4a/ARF tumor suppressor: one gene--two products--two pathways. Trends Biochem Sci. 1998;23(8):291-6.10.1016/s0968-0004(98)01236-5
30. Li TW, Ting JH, Yokoyama NN, Bernstein A, van de Wetering M, Waterman ML. Wnt activation and alternative promoter repression of LEF1 in colon cancer. Mol Cell Biol. 2006;26(14):5284-99.10.1128/MCB.00105-06

31. Murray-Zmijewski F, Lane DP, Bourdon JC. p53/p63/p73 isoforms: an orchestra of isoforms to harmonise cell differentiation and response to stress. Cell Death Differ. 2006;13(6):962-72.10.1038/sj.cdd.4401914
32. Tveden-Nyborg P, Bergmann TK, Jessen N, Simonsen U, Lykkesfeldt J. BCPT policy for experimental and clinical studies. Basic Clin Pharmacol Toxicol. 2021;128(1):4-8.10.1111/bcpt.13492
33. Jahan N, Lee JM, Shah K, Wakimoto H. Therapeutic targeting of chemoresistant and recurrent glioblastoma stem cells with a proapoptotic variant of oncolytic herpes simplex virus. Int J Cancer. 2017;141(8):1671-81.10.1002/ijc.30811
34. Sjostedt E, Zhong W, Fagerberg L, Karlsson M, Mitsios N, Adori C, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science. 2020;367(6482).10.1126/science.aay5947
35. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.10.1038/nature11247
36. Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48(D1):D882-D9.10.1093/nar/gkz1062
37. Zou Z, Ohta T, Miura F, Oki S. ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. Nucleic Acids Res. 2022;50(W1):W175-W82.10.1093/nar/gkac199
38. Veiga DFT, Nesta A, Zhao Y, Deslattes Mays A, Huynh R, Rossi R, et al. A comprehensive long-read isoform analysis platform and sequencing resource for breast cancer. Sci Adv. 2022;8(3):eabg6711.10.1126/sciadv.abg6711
39. Leung SK, Jeffries AR, Castanho I, Jordan BT, Moore K, Davies JP, et al. Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. Cell Rep. 2021;37(7):110022.10.1016/j.celrep.2021.110022
40. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988-D95.10.1093/nar/gkab1049
41. Suzuki A, Kawano S, Mitsuyama T, Suyama M, Kanai Y, Shirahige K, et al. DBTSS/DBKERO for integrated analysis of transcriptional regulation. Nucleic Acids Res. 2018;46(D1):D229-D38.10.1093/nar/gkx1001
42. Suzuki A, Wakaguri H, Yamashita R, Kawano S, Tsuchihara K, Sugano S, et al. DBTSS as an integrative platform for transcriptome, epigenome and genome sequence variation data. Nucleic Acids Res. 2015;43(Database issue):D87-91.10.1093/nar/gku1080
43. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369(6509):1318-30.10.1126/science.aaz1776
44. Kaur B, Brat DJ, Calkins CC, Van Meir EG. Brain angiogenesis inhibitor 1 is differentially expressed in normal brain and glioblastoma independently of p53 expression. Am J Pathol. 2003;162(1):19-27.10.1016/S0002-9440(10)63794-7
45. Zhu D, Hunter SB, Vertino PM, Van Meir EG. Overexpression of MBD2 in glioblastoma maintains epigenetic silencing and inhibits the antiangiogenic function of the tumor suppressor gene BAI1. Cancer Res. 2011;71(17):5859-70.10.1158/0008-5472.CAN-11-1157
46. Bentsen M, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, et al. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat Commun. 2020;11(1):4267.10.1038/s41467-020-18035-1

47. Gates LA, Foulds CE, O'Malley BW. Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. Trends Biochem Sci. 2017;42(12):977-89.10.1016/j.tibs.2017.10.004
48. Krasnoperov V, Lu Y, Buryanovsky L, Neubert TA, Ichtchenko K, Petrenko AG. Post-translational proteolytic processing of the calcium-independent receptor of alpha-latrotoxin (CIRL), a natural chimera of the cell adhesion protein and the G protein-coupled receptor. Role of the G protein-coupled receptor proteolysis site (GPS) motif. J Biol Chem. 2002;277(48):46518-26.10.1074/jbc.M206415200
49. Cork SM, Van Meir EG. Emerging roles for the BAI1 protein family in the regulation of phagocytosis, synaptogenesis, neurovasculature, and tumor development. J Mol Med (Berl). 2011;89(8):743-52.10.1007/s00109-011-0759-x
50. Kishore A, Purcell RH, Nassiri-Toosi Z, Hall RA. Stalk-dependent and Stalk-independent Signaling by the Adhesion G Protein-coupled Receptors GPR56 (ADGRG1) and BAI1 (ADGRB1). J Biol Chem. 2016;291(7):3385-94.10.1074/jbc.M115.689349
51. Hamann J, Aust G, Arac D, Engel FB, Formstone C, Fredriksson R, et al. International Union of Basic and Clinical Pharmacology. XCIV. Adhesion G protein-coupled receptors. Pharmacol Rev. 2015;67(2):338-67.10.1124/pr.114.009647
52. Frenster JD, Stephan G, Ravn-Boess N, Bready D, Wilcox J, Kieslich B, et al. Functional impact of intramolecular cleavage and dissociation of adhesion G protein-coupled receptor GPR133 (ADGRD1) on canonical signaling. J Biol Chem. 2021;296:100798.10.1016/j.jbc.2021.100798
53. Rath A, Glibowicka M, Nadeau VG, Chen G, Deber CM. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci U S A. 2009;106(6):1760-5.10.1073/pnas.0813167106
54. Kaur B, Brat DJ, Devi NS, Van Meir EG. Vasculostatin, a proteolytic fragment of brain angiogenesis inhibitor 1, is an antiangiogenic and antitumorigenic factor. Oncogene. 2005;24(22):3632-42.10.1038/sj.onc.1208317
55. Koh JT, Kook H, Kee HJ, Seo YW, Jeong BC, Lee JH, et al. Extracellular fragment of brain-specific angiogenesis inhibitor 1 suppresses endothelial cell proliferation by blocking alphavbeta5 integrin. Exp Cell Res. 2004;294(1):172-84.10.1016/j.yexcr.2003.11.008
56. Silva JP, Lelianova V, Hopkins C, Volynski KE, Ushkaryov Y. Functional cross-interaction of the fragments produced by the cleavage of distinct adhesion G-protein-coupled receptors. J Biol Chem. 2009;284(10):6495-506.10.1074/jbc.M806979200
57. Duman JG, Tu YK, Tolias KF. Emerging Roles of BAI Adhesion-GPCRs in Synapse Development and Plasticity. Neural Plast. 2016;2016:8301737.10.1155/2016/8301737
58. Sriram K, Insel PA. G Protein-Coupled Receptors as Targets for Approved Drugs: How Many Targets and How Many Drugs? Mol Pharmacol. 2018;93(4):251-8.10.1124/mol.117.111062
59. Mathema VB, Na-Bangchang K. Regulatory roles of brain-specific angiogenesis inhibitor 1(BAI1) protein in inflammation, tumorigenesis and phagocytosis: A brief review. Crit Rev Oncol Hematol. 2017;111:81-6.10.1016/j.critrevonc.2017.01.006

## Figure legends

**FIGURE 1: Assessment of *ADGRB1* mRNA variants and BAI1 protein isoforms in human and mouse tissues.**

**A)** Normalized mRNA expression levels (nTPM) of *ADGRB1* in 12 regions of human brain. Source: The Human Protein Atlas (www.proteinatlas.org).[34]

**B)** Western blot analysis of BAI1 in normal human brain, human medulloblastoma (D556, ONS76, DAOY, and D425) and glioblastoma (LN229) cell lines and human glioma stem cell lines (GSC4, GSC18) using a C-terminal antibody (epitope 1537-1567aa). Long exposure times were used for all cell lines as expression is lower than in normal human brain.

**C)** BAI1 protein level after siRNA mediated knockdown of *ADGRB1* in DAOY cell line. si BAI1 #1 targets exon 23 and si BAI1 #2 targets exon 21 (NM_001702.2). Bar graphs represent relative fold change of BAI1 level (normalized to GAPDH) for ~175 kDa and ~75 kDa bands.

**D)** Transient transfection of *ADGRB1* cDNA in medulloblastoma cell lines reveals the major FL-BAI1 ~175 kDa isoform and its cleavage products of ~70 kDa and ~75 kDa.

**E)** *Adgrb1* mRNA expression in 13 brain regions of mouse. Source: The Human Protein Atlas(www.proteinatlas.org).[34]

**F, G)** Expression of BAI1 isoforms in different organs and brain regions of WT and exon 2 deleted C57BL/6J mice.

**FIGURE 2: Transcriptional diversity of the human *ADGRB1* gene.**

**A)** Top: Genome browser tracks of H3K4me3 and H3K27ac ChIP-seq data (dorsolateral prefrontal cortex), ATAC-seq data (caudate region), Iso-seq data and TSS-seq data as well as a predicted promoter region from Ensembl genome browser (red rectangle) at the human *ADGRB1* locus (hg38).

Middle: *ADGRB1* transcript variants identified in human cerebral cortex by long-read isoform sequencing (Iso-seq). Isoform number is indicated on the left. a, adult; f, fetal. Colors indicate the classification of transcript categories (blue = FSM; cyan = ISM; red = NNC; orange = NIC).[39]

Bottom: Transcription start sites predicted by TSS-seq in a ~1200bp. region at the intron17/exon 18 boundary in adult and fetal brains and HEK293 cell line.

**B)** Heatmap showing median read counts per base for each exon of *ADGRB1* in different brain and other tissues from GTEx database. Gene-level expression is calculated based on a collapsed gene model combining all *ADGRB1* mRNA variants into a single transcript.[43]

**FIGURE 3: Transcriptional diversity of the mouse *Adgrb1* gene.**

**A)** Top: Genome browser tracks of H3K4me3 and H3K27ac ChIP-seq data, ATAC-seq data, Iso-seq data and TSS-seq data as well as predicted promoter regions near exons 1 and 18 from Ensembl genome browser (red rectangles) at the mouse *Adgrb1* locus (mm10) in forebrain tissue of the C57BL/6J mouse strain.

Middle: *Adgrb1* transcript variants identified in mouse cerebral cortex by Iso-seq. Isoform number is indicated on the left. Transcript categories: blue = FSM; cyan = ISM; red = NNC; orange = NIC. [39]

Bottom: Transcription start sites predicted by TSS-seq in a ~1200bp region at the intron17/exon 18 boundary in mouse embryo brains at days E11, E15, E17.

**B)** Reverse transcription-polymerase chain reaction (RT-PCR) analysis of exon usage in *Adgrb1* transcripts in brain of wild type (WT), heterozygous (HET) and exon 2 deleted (KO) mice using exon-specific forward (F) and reverse (R) primers.

**C)** RT-PCR analysis showing the expression of novel *Adgrb1* transcript variant (1.44/2) from the alternate promoter in intron 17 in mouse brain. Left: Exon structure and primer design of transcript *Adgrb1*.44/2 are shown. Yellow, start of new exon 1 from intron 17 start site. Green, exons present in full length transcript. Right: agarose gel showing RT-PCR products in WT, HET and KO mice.

**FIGURE 4: Functional evaluation of intron 17 alternative promoter activity in Human.**

**A)** New exon 1 of transcript variants of *ADGRB1* starting in 3′ end of intron 17 identified in human cerebral cortex (orange rectangles). Length of the new exon 1, splicing to exon 18 and start codons (ATG) for same open reading frame (ORF) as the full length *ADGRB1* are indicated.

**B)** Two regions of interest (1190bp and 2210bp long) upstream of the new exons shown in (A) were tested for putative promoter activity in firefly luciferase reporter constructs and transiently transfected in LN229 and HEK293T cells with a EF1-renilla luciferase construct as an internal control. Fold change of the normalized firefly over luciferase activity is shown. Data are presented as mean ± SEM. $^{****}P < 0.0001$ (One-way ANOVA with Tukey's multiple comparisons correction).

**FIGURE 5: Alternative promoter leads to the formation of BAI1 isoforms.**

Top panel: Schematic showing the three main transcriptional start site regions (TSS1,2,3) in the *ADGRB1* gene (exons are shown with blue rectangles). Transcription from the 5′ core promoter near exon 1 generates two mRNAs that start in exons 1 and 2, respectively. They differ in size of their 5′ UTR regions (red rectangles) but use the same start codon in exon 2. Several transcripts are generated from the alternative promoter at the 3′ end of intron 17 with variable 5′ UTR regions and translation start codons either in the new alternate exons 1 in intron 17 or an in-frame ATG in exon 19 (see detail in figure 4 and figure S2).

Middle panel: Left, translation from full size mRNAs give rise to FL-BAI1 proteins (1,584 amino acids; predicted size of 173.5 kDa) in human. Location of functional domains described in introduction are indicated. Right, translation from shorter transcripts originating from alternative promoter in intron 17 generate hBAI1 isoforms lacking most of the N-terminal region. Those using a start codon in the new exon 1 have a predicted size of 76.4 kDa, while those using the ATG in exon 19 are 70.5 kDa. These predicted isoform sizes are calculated for transcripts that share the same exons 19-31 as the full-size transcript.

Bottom panel: Left, GAIN domain induced autoproteolysis at the GPS of FL-hBAI1 in the endoplasmic reticulum cleaves between leucine (L-926) and serine(S-927) and leaves a membrane-associated N-terminal truncated BAI1 with a remaining stalk of 19 amino acids (AA)[6] that can serve as an agonist (also called “stinger” or “Stachel” in German) to activate the receptor. Right, the hBAI1 isoforms generated from the alternative promoter transcripts likely do not undergo autoproteolysis as they lack the GAIN domain. The transcripts translated from the ATG in exon 19 generate a hBAI1 isoform with only 5 AA outside of the cell membrane. On the contrary, those translated from an ATG in the new exon 1 in intron 17, have 60 AA N-termini. AA compositions of N-termini are shown.