Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

YB-1 binds the 3' splice site of EGFR exon 3in glioblastoma cells, but a YB-1 decoycannot act on EGFR alone

Kenji Ohe

oekenji@gmail.com

Fukuoka University https://orcid.org/0000-0002-1958-7572

Kyungseung Kwon

Fukuoka University

Research Article

Keywords: glioblastoma, EGFR, YB-1, YBX1, exon definition, 3' splice site, polypyrimidine tract, U2AF65, decoy RNA, antisense oligonucleotide, nonsense-mediated decay, CLIP reanalysis

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11720196/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare potential competing interests as follows: The authors declare no competing interests. K.O. is an author of all three preprints from the same department that are cited here: the companion study whose design premises are examined (doi:10.64898/2026.10.25.761620), and two studies whose observations are discussed in relation to the present findings (doi:10.64898/2026.11.02.762382; doi:10.64898/2026.10.29.762206). No data from those studies are reported here.

YB-1 binds the 3' splice site of EGFR exon 3 in glioblastoma cells, but a YB-1 decoy cannot act on EGFR alone

Running title. A YB-1 site at the 3′ splice site of EGFR exon 3

Authors. Kyungseung Kwon[1], Kenji Ohe[1]

[1] Department of Clinical Pharmacotherapy, Faculty of Pharmaceutical Sciences, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan

Keywords. glioblastoma; EGFR; YB-1 (YBX1); exon definition; decoy RNA; nonsense-mediated decay; reanalysis of deposited CLIP and RNA-seq data

Abstract

RNA oligonucleotides directed at a splice site can suppress a gene without touching its protein, and a companion preprint reports that oligonucleotides against the *3′ splice site of EGFR exon 3* reduce proliferation of A172 glioblastoma cells. That work rests on a SpliceAid [21] prediction of a Y-box binding protein 1 (YB-1, *YBX1*) site at the targeted sequence, which has never been tested. This study tests three premises of that design against deposited CLIP, RIP, knockdown RNA-seq and splice-junction data.

First, whether exon 3 is a defensible target among the 26 internal exons of *EGFR*. Applying four criteria in sequence — encoding sequence upstream of the kinase domain, frameshifting when skipped, lying within the first 200 codons, and carrying the YB-1 CA(U/C)C core at its 3′ splice site — leaves exactly one exon, exon 3. The frameshift requirement is not cosmetic: EGFRvIII, the in-frame deletion of exons 2 to 7 (801 nt, 267 codons), is constitutively active, so inducing an in-frame skip in *EGFR* risks producing a signalling-competent receptor.

Second, whether YB-1 binds there. A deposited YBX1 eCLIP peak from U251 glioblastoma cells (GSE213046) maps to chr7:55,143,305-55,143,308, the first four nucleotides of exon 3, immediately 3′ of the splice acceptor AG ($p = 1.3 \times 10^{-3}$, log2FC = 3.29). Six YBX1 CLIP datasets from non-glial cells also detect *EGFR* binding, but not at this position: across the four with mapped site coordinates, none of 103 sites overlaps exon 2, intron 2 or exon 3, and exonic binding is concentrated in exon 28, the 3′UTR. The U251 dataset differs in that its two most 5′ exonic peaks lie in exon 1 and in exon 3.

Third, whether a decoy carrying the YB-1 consensus can act specifically on *EGFR*. It cannot. In U251, YB-1 binding sites fall in 8,801 genes, 7,840 of which are expressed there — 42.8% of the expressed transcriptome — and the CA(U/C)C core occurs 48 times in the *EGFR* mRNA alone. Across four glioblastoma YBX1-knockdown RNA-seq datasets, *EGFR* mRNA changes by |log2FC| 0.23 to 0.59 and rises in three of the four, ranking 2,418th to 16,274th by effect size.

Public U2AF65 iCLIP adds a constraint on mechanism. The 30 nt upstream of the exon 3 acceptor carry 92 crosslinks in pooled HeLa data, ranking tenth among the 26 internal exons of EGFR, whereas the CD44 variable exon v5 tract, on which the YB-1/U2AF65 mechanism was established, carries 15. The EGFR tract is not short of U2AF65, and the two proteins do not share it: the U2AF65 footprint runs from −21 to −3 while the YB-1 peak sits at +1 to +4, across the acceptor and inside the exon. The route by which YB-1 is known to act at a 3′ splice site therefore does not obviously apply here. Separately, skipping of exon 3 shifts the reading frame and generates a premature stop 2,771 nt upstream of the final exon junction, making the transcript a canonical nonsense-mediated decay substrate. Across 43 public RNA-seq samples the skipping junction carried 4 reads against approximately 1,836 inclusion reads (PSI ≈ 0.998), all 4 in a glioblastoma line.

Together these results support YB-1 contact at the exon 3 3′ splice site in glioblastoma specifically, provide a reproducible rationale for that target, and set a quantitative bound on what a YB-1 decoy can be claimed to do.

Introduction

Glioblastoma remains close to incurable. *EGFR* is focally amplified in roughly 40% of IDH-wildtype cases, and some alteration of *EGFR* — amplification, rearrangement, aberrant splicing or point mutation — is present in about 57% [2,3]; amplification is itself one of the molecular criteria for the diagnosis [4]. That makes the receptor a rational target, and yet anti-EGFR therapy has not improved survival: a Cochrane review of first-line and recurrent disease found no demonstrable overall survival benefit [5,6]. One reason to look below the protein is that the lesion is often genomic amplification rather than a kinase-domain mutation, so lowering the amount of receptor made may matter more than blocking the activity of each molecule.

Splicing offers a point of attack. Human introns are long relative to exons, so the spliceosome recognises an exon across its length rather than each splice site in isolation — exon definition [7,8] — and RNA-binding proteins engaging exonic sequence near an intron boundary are central to that recognition. An oligonucleotide placed over such a sequence can block exon definition and remove the exon from the mature transcript.

That approach carries a specific hazard in *EGFR*. EGFRvIII, among the commonest *EGFR* alterations in glioblastoma, arises from an in-frame deletion of exons 2 to 7 that removes 801 nt from the coding sequence and shortens the extracellular domain by 267 residues, with a novel glycine created at the junction [9,10]; the product is constitutively active [11] and oncogenic [9-11]. Inducing skipping of an in-frame exon in this gene could therefore produce a shortened receptor that signals better, not worse. The requirement is to remove an exon whose loss shifts the reading frame, so that the transcript carries a premature termination codon and is degraded by nonsense-mediated decay rather than translated [22,23].

Y-box binding protein 1 (YB-1, gene *YBX1*) is a plausible protein to target in this setting. It acts in transcription, splicing, mRNA stability and translational control; SELEX defines CAUC and CACC as its high-affinity motifs [13], a crystal structure of the cold-shock domain bound to 5′-UCAUCU-3′ resolves the CAUC core as four pi-pi stacking pairs with His-87, Phe-85, Phe-74 and Trp-65 [12], and iCLIP in glioblastoma multiforme cells recovers a UYAUC consensus [14]. There is direct precedent for YB-1 acting at a 3′ splice site: it binds CAUC motifs in and upstream of CD44 variable exon v5 and promotes inclusion of that exon by recruiting U2AF65 to a weak polypyrimidine tract [13]. YB-1 is overexpressed in glioblastoma and associated with growth, progression and treatment resistance [15,16].

Whether YB-1 can be drugged is less settled, and the position should be stated plainly. A cell-permeable **peptide** decoy to YB-1 suppressed growth of breast and prostate cancer cells [17], and RNA decoys directed at YB-1 slowed tumour growth and improved survival in a glioblastoma xenograft model [15]. Also reported in glioma is an oncolytic adenovirus that exploits YB-1 rather than inhibiting it [18], and the small molecule SU056, reported in ovarian cancer [19]. A YB-1 decoy therefore has in vivo precedent in this tumour type. What that precedent does not establish is specificity for any one transcript: the decoy used there was directed at YB-1 itself, and its effect was attributed to the YB-1/CCT4/mLST8/mTOR axis rather than to the splicing of a named gene.

A companion preprint (Nishi et al.) designed antisense and decoy RNA oligonucleotides against the *3′ splice site region of EGFR exon 3* in A172 glioblastoma cells and reported reduced EGFR protein and reduced proliferation, with no comparable effect in A549 lung adenocarcinoma cells. Those experiments were single runs without independent replication, and three premises of the design were untested. The target was chosen after inspecting only exons 1 to 3, so it is unknown whether any of the other 25 junctions would

have scored as well. YB-1 binding to the targeted sequence rests on a database prediction. And the word *specific*, applied to a reagent that sequesters a multifunctional RNA-binding protein, has not been examined against what that protein actually binds.

Three preprints from this department bear on the same junction, and two of them reach it from a different direction. Nishi et al. [1] report the oligonucleotide experiments whose premises this study tests. Naito et al. [35] treated A172 glioblastoma cells with the SF3b inhibitor herboxidiene and found no *EGFR* intron 2-retained product at any concentration from 0.125 to 1 µM, although *CDKN1B* introns were retained in the same cells. Fukagawa et al. [36] reanalysed deposited RNA-seq from spliceostatin-treated cells and found intron 2 among the least responsive junctions in the gene, reporting its acceptor at the 88.2nd percentile of a genome-wide panel. The 3′ splice site of *EGFR* intron 2 is the 3′ splice site of exon 3, so all three address one junction. None of them asked which proteins occupy it. That is the question taken up here.

Substantial deposited data bear on all three questions and had not been brought to them. This study uses deposited CLIP, RIP, knockdown RNA-seq and splice-junction data, together with sequence analysis of the *EGFR* locus, to test the premises rather than the experimental results. Nothing here reproduces or audits the companion preprint's own experiments; the aim is to establish what can be asserted about the target before more experimental effort is spent on it.

Methods

Gene model and coordinates

EGFR ENSG00000146648, canonical transcript ENST00000275493 (MANE Select), GRCh38, plus strand. Exon structure and sequence were retrieved from the Ensembl REST API [30]; ENST00000275493 is the MANE Select transcript [31]. Exon 2 chr7:55,142,286-55,142,437 (152 nt); exon 3 chr7:55,143,305-55,143,488 (184 nt); exon 4 chr7:55,146,606-55,146,740 (135 nt). The transcript has 28 exons, of which 26 are internal, and a coding sequence of 3,633 nt encoding 1,210 residues.

Files deposited on hg19 were handled by deriving the offset from the gene start (hg19 chr7:55,086,714 minus GRCh38 chr7:55,018,820 = +67,894) and validating it against each file's own features at both ends of the locus, rather than by applying a nominal figure. The assembly used for every dataset is stated in the results.

Frame arithmetic

For each internal exon, the coding length within that exon was computed from the intersection of the exon with the CDS. An exon whose coding length is divisible by three was classed as in-frame when skipped; otherwise frameshifting. The position of each exon was expressed as the number of codons completed before it. The procedure was validated against EGFRvIII: the summed coding length of exons 2 to 7 is 801 nt = 267 codons, in frame, matching the known 267-residue deletion.

For exon 3 specifically, the skipped cDNA was assembled and translated to locate the premature termination codon, and the distance from that codon to the final exon-exon junction was measured against the 50 to 55 nt rule for nonsense-mediated decay.

Splice site features

Polypyrimidine content was computed as the percentage of C and T over the 18 nt immediately upstream of each 3′ splice site AG, excluding the AG itself; including it would give 14 of 18 rather than 16 of 18 for EGFR exon 3. U2AF65 crosslinks were counted over the 30 nt upstream of each acceptor. Positions are numbered with −1 as the last intronic nucleotide, the G of the AG, and +1 as the first exonic nucleotide; there is no position 0. CD44 variable exon v5 was identified as ENSE00000824450 (chr11:35,201,671-35,201,787), the ninth exon of ENST00000428726, by its position among the nine variable exons and by

the tract features reported for it in the original work [13]. Coordinates in hg19 and hg18 files were converted with the Ensembl assembly mapping service and checked against the data themselves, each locus mapping as a single contiguous block with the same offset at both ends. Occurrences of the YB-1 core CA(U/C)C were counted, with overlapping matches, in a window spanning 50 nt of upstream intron plus the first 50 nt of the exon, and separately across whole exons, the mature mRNA and the full pre-mRNA span.

Public datasets

Processed files were obtained from the GEO FTP archive. No raw sequencing reads were downloaded or aligned.

GSE	Assay	Cells	Assembly	File used
GSE213046	YBX1 eCLIP	U251	hg38	significant peak annotation
GSE213046	siYBX1 RNA-seq	U251	—	authors' DESeq2 table
GSE161523	shYB-1 RNA-seq	U251, U87	—	htseq counts
GSE294052	shYB-1 RNA-seq	U87, U373	—	authors' DE table
GSE249897	YB-1/YB-3 KO RNA-seq	HEK293T	—	read counts
GSE130781	RNA-seq, Ribo-seq, RIP	HEK293T	—	gene counts
GSE150924	YBX1 KD RNA-seq	DAOY, MED08, UW228	—	FPKM
GSE264220	YBX1 RIP	PANC-1	—	peak annotations
GSE349934	ecCLIP	HEK293T	hg38	CIMS bed
GSE311747	RBProximity-CLIP	HEK293	hg38	cluster statistics
GSE133620	PAR-CLIP	T24	hg19	PARalyzer clusters
GSE63604	HITS-CLIP	MDA-shID	hg19	peak coordinates
GSE150925, GSE249896	PAR-CLIP	HEK293, HEK293T	hg38	binding-site tables
GSE83664	YB-1 RIP	multiple myeloma	—	cuffdiff output
GSE311240	RNA-seq	iPSC, NPC, neurons	hg38	STAR SJ.out.tab
GSE312830	RNA-seq	KNS-60	hg38	STAR SJ.out.tab
GSE49417	RNA-seq	7 ENCODE lines	hg19	junction bedRnaElements
GSE99700	U2AF65 iCLIP	HeLa	hg19	crosslink bigWig, 6 libraries
GSE61603	U2AF65 CLIP-seq	HeLa	hg18	coverage bedGraph
GSE91570	U2AF2 eCLIP	HepG2	hg38	peaks and signal
GSE126693	RNA-seq	HeLa	—	gene counts, for expression

Splice junction quantification

Percent spliced in was computed from STAR [27] unique-mapping junction counts as PSI = mean(inclusion) / (mean(inclusion) + skipping), where inclusion is the mean of the exon 2-exon 3 junction (intron chr7:55,142,438-55,143,304) and the exon 3-exon 4 junction (intron chr7:55,143,489-55,146,605), and skipping is the exon 2-exon 4 junction (chr7:55,142,438-55,146,605). Junction reads supporting skipping were inspected individually for splice motif, annotation status and maximum overhang. Read counts are reported alongside every PSI.

Statistics

Differential expression was taken from the depositors' own tables where available (GSE213046, GSE294052) and otherwise computed with PyDESeq2 0.5.4 [29] using the DESeq2 Wald test [28] with Benjamini-Hochberg correction. GSE150924 supplies FPKM only and was analysed by Welch's t test on log2(FPKM + 0.1) with Benjamini-Hochberg correction. RIP enrichment was assessed at adjusted $p < 0.05$ with log2 fold change > 1. The comparison of skipping-read frequency between glioblastoma and neural samples used Fisher's exact test, two-sided, at read level and at sample level. All tests were two-sided at the 0.05 level; no adjustment was made across the separate questions addressed here, and the analysis is descriptive rather than confirmatory.

Sites were assigned to exons and introns by interval overlap rather than by midpoint, so that a site partially overlapping a feature is counted for that feature; a site spanning a boundary is therefore counted for both. Counts were restricted to the plus strand, the strand of EGFR, and were deduplicated; for GSE63604, the shuffled control entries deposited alongside the real sites, which carry identical coordinates and a -rand suffix, were removed before counting. Binding-region classification used 700 significant peaks sampled at random and annotated against Ensembl GRCh38 by genomic overlap.

Data not reachable

The ENCODE portal, Zenodo, UCSC Xena, the GDC API, EBI FTP, Bioconductor, recount3 and POSTAR could not be accessed from the analysis environment. ENCODE YBX1 eCLIP in K562 and HepG2 was therefore not examined, and no GTEx or TCGA junction-level data were obtained. These gaps are stated where they bear on a conclusion and none of them underpins a negative claim.

Results

Exon 3 is the only internal exon of EGFR that satisfies four independent design criteria

The companion preprint selected its target after examining exons 1 to 3 only. To ask whether that choice survives a wider view, all 26 internal exons of ENST00000275493 were evaluated against criteria drawn from the design logic itself.

An in-frame skip is the wrong outcome. EGFRvIII is an in-frame deletion of exons 2 to 7 and is constitutively active. Summing coding lengths across those exons gives 801 nt = 267 codons, exactly in frame and matching the known 267-residue deletion; this confirms the arithmetic applied below. An oligonucleotide that induced skipping of an in-frame *EGFR* exon could therefore produce a shortened but signalling-competent receptor. Within the EGFRvIII region, skipping exon 4 (135 nt) or exon 5 (69 nt) alone stays in frame, while exons 2, 3, 6 and 7 are frameshifting; exons 4 and 5 were never viable candidates.

Applying the criteria in sequence:

Step	Criterion	Exons remaining
0	Internal exons (2–27)	26
1	Encodes sequence upstream of the kinase domain	17 — exons 2–18
2	Skipping shifts the reading frame	12 — exons 2, 3, 6, 7, 9–15, 17
3	Lies within the first 200 codons	2 — exons 2 and 3
4	Carries a CA(U/C)C core in its 3′ splice site window	1 — **exon 3**

The funnel converges on a single exon and reproduces an asymmetry reported independently by the companion preprint: exon 2 survives to the final step and fails it because its 3′ splice site window contains no CA(U/C)C, consistent with SpliceAid placing

YB-1 at the exon 2 5′ splice site rather than its 3′ site. A database prediction and a motif count against the structurally defined consensus therefore agree.

Three features do not support the selection and should not be claimed. The exon 3 polypyrimidine tract is 89% pyrimidine over the 18 nt upstream of the AG, ranking 17th of 26 internal exons where rank 1 is weakest; the weakest tracts belong to exons 25, 4, 13 and 21. The argument that a weak 3′ splice site is more enhancer-dependent, and so more disruptable, does not apply. The CA(U/C)C count in the exon 3 3′ splice site window is 1, equal to the median across internal exons, while exons 8, 17 and 22 contain three or four; the motif is present, not enriched. And this funnel is a post-hoc reconstruction, not the procedure by which the target was chosen.

A combinatorial hazard. Skipping exons 2 and 3 together removes 336 nt, which is in frame (112 codons). Either exon alone is frameshifting, but a reagent disrupting both would generate an in-frame internal deletion in the extracellular domain. Any experiment in this region should report the exon 2-exon 4 junction explicitly.

A YB-1 eCLIP peak maps to the 3' splice site of EGFR exon 3 in U251 glioblastoma cells

The deposited YBX1 eCLIP peak table from U251 glioblastoma (GSE213046 [26], hg38, FLAG-YBX1, two biological replicates) contains 27 peaks across the *EGFR* locus. One maps to chr7:55,143,305-55,143,308, the first four nucleotides of exon 3, immediately 3′ of the splice acceptor AG:

```
intron 2, last 20 nt  (...55,143,304):  TTTATGTTTTCTCTTCTTAG
exon 3, first 12 nt   (55,143,305...):  ACCATCCAGGAG
peak interval         (55,143,305-308): ACCA
```

The peak gives −log10 P = 2.90 ($p = 1.3 \times 10^{-3}$) and log2 fold change 3.29, passing the depositors′ threshold. *EGFR* is robustly expressed in these cells (baseMean 10,174 against a median of 209 across tested genes), so the assay had its substrate. The position is the canonical crosslink location for a protein bound at a 3′ splice site.

Three qualifications travel with this observation.

The table is exon-annotated, so the intron was never assayed. Peak identifiers run to peak_66786 while only 40,165 rows are present; 26,621 peaks were removed before deposition. The missing identifiers are positionally traceable — peak_58626 to peak_58631 would fall within *EGFR* intron 1 and are absent — and the Location column holds exon numbers and never "intron". The SpliceAid-predicted site, if it lies in intron 2 upstream of the AG, cannot be addressed with this file.

The 4 nt width is probably an artefact of clipping. Among the 27 *EGFR* peaks, three begin exactly at an annotated exon start and four end exactly at an exon end, far above chance for 27 intervals against 28 exons. Median peak width in the file is 61 nt and 4 nt falls in the narrowest 0.41%. The most economical interpretation is that the read pileup extended across the splice site into intron 2 and was truncated at the exon edge, which would strengthen rather than weaken the 3′ splice site reading, but this cannot be established from the deposited file.

The peak is modest. It sits at the 37th percentile of all peaks in the file (median −log10 P = 3.79), well below the *EGFR* 3′UTR peaks, which reach −log10 P = 25.1. Per-replicate reproducibility is not recoverable.

The position is not seen outside glioblastoma

Dataset	Method	Cells	Sites in *EGFR*	exon 1	exon 2	intron 2	**exon 3**	exon 28 (3′UTR)
GSE213046	eCLIP	U251 (GBM)	27	**1**	0	0	**1**	14

Dataset	Method	Cells	Sites in *EGFR*	exon 1	exon 2	intron 2	**exon 3**	exon 28 (3′UTR)
GSE311747	RBProximity-CLIP	HEK293	40	0	0	0	**0**	7
GSE349934	ecCLIP	HEK293T	13	0	0	0	**0**	11
GSE133620	PAR-CLIP	T24	31	2	0	0	**0**	7
GSE63604	HITS-CLIP	MDA-shID	19	0	0	0	**0**	4
GSE150925	PAR-CLIP	HEK293	0	—	—	—	—	—
GSE249896	PAR-CLIP	HEK293T	0	—	—	—	—	—

Counts are of sites on the plus strand, the strand of *EGFR*, after removing duplicates and, for GSE63604, the shuffled control entries deposited alongside the real sites. Across the four non-glial datasets with mapped coordinates, 103 sites fall in the *EGFR* locus and **none overlaps exon 2, intron 2 or exon 3**; two overlap exon 1, both in T24. Exonic binding in these cells is concentrated in exon 28, the 3′UTR, and the 5′ region of the gene is represented almost entirely by clusters within intron 1, which spans 123 kb.

The HEK293 and HEK293T negatives carry limited weight: *EGFR* there is expressed at roughly 2% of the U251 level, and GSE150925 and GSE249896 report fewer than 2,100 sites genome-wide, finding none in *EGFR* at all. The T24 and MDA-MB-231-derived datasets are more informative, since each recovered 19 to 31 *EGFR* sites, so the transcript was present and bound but not at exon 3. GSE311747 includes a nuclear fraction, the compartment relevant to a co-transcriptional splicing interaction, and it too has no site in exon 2, intron 2 or exon 3.

Cell-type-restricted binding at this position would be consistent with the companion preprint's observation that its oligonucleotides acted in A172 and not in A549, although that comparison involves two differences at once and cannot be attributed to binding alone.

Cell-type-restricted binding at this position would be consistent with the companion preprint's observation that its oligonucleotides acted in A172 and not in A549, although that comparison involves two differences at once and cannot be attributed to binding alone.

YB-1 binds 43% of the expressed transcriptome, and EGFR is not a leading responder to its loss

A decoy carrying the YB-1 consensus sequesters the protein, so its reach is bounded by what that protein engages rather than by the gene the experimenter has in mind.

Dataset	Cells	Peaks	Distinct target genes
GSE213046 eCLIP	U251 (GBM)	40,165	8,801
GSE213046, P ≥ 3 (−log10)	U251 (GBM)	23,656	5,913
GSE264220 RIP	PANC-1	60,489 / 55,417	11,461 / 10,864
GSE133620 PAR-CLIP	T24	48,047 / 47,576	not annotatable
GSE130781 RIP	HEK293T	—	3,388 enriched

In U251, 18,304 genes meet a mean control count above 5. Of the 8,801 eCLIP target genes, 7,840 are expressed there, so **42.8% of the expressed transcriptome carries a YB-1 site**; at the stricter threshold it remains 29.6%. Of the eCLIP targets also testable in the matching knockdown, 6,071 of 7,899 (77%) changed at adjusted $p < 0.05$, so binding and functional response overlap broadly.

EGFR is a well-bound target, with 27 peaks ranking 87th of 8,801 genes, and 48th to 63rd of about 11,000 in PANC-1. It is not an exceptional one: 86 genes carry more peaks. Binding region, from 700 randomly sampled significant peaks, was CDS 64.4%, 3′UTR 30.7%, 5′UTR 2.9%; intron could not be assessed for the reason given above.

EGFR mRNA changes little, and the direction is not reproducible

Dataset	Cells	*EGFR* log2FC	adj. p	Rank by effect size
GSE213046 (siRNA)	U251	**−0.228**	2.3e−7	16,274 / 22,854
GSE161523 (shRNA)	U251	**+0.589**	4.9e−17	2,735 / 15,665
GSE161523 (shRNA)	U87	**+0.472**	1.5e−11	2,418 / 14,440
GSE294052 (shRNA)	U87, U373	**+0.544**	1.4e−127	5,176 / 18,191

Three observations follow. The effect is small: every significant value lies between 0.23 and 0.59 in absolute log2 fold change, or 1.17- to 1.50-fold in mRNA. The direction disagrees: three of four glioblastoma datasets show *EGFR* mRNA rising when YB-1 is depleted, and the two U251 datasets disagree with each other. The rank by effect size is middling to low, with 16,273 genes moving more than *EGFR* in GSE213046; the better p-value ranks (170th to 662nd) reflect high expression and the statistical power that brings rather than a large change.

Knockdown efficiency does not explain this. *YBX1* itself gives log2 fold change −2.54 to −4.75 and ranks first or second by p value in seven of nine datasets. Depleting YB-1 moved 13.6% to 56% of tested genes in glioblastoma lines.

Datasets in which *EGFR* is poorly expressed — HEK293T (baseMean 29 to 101), three medulloblastoma lines (FPKM 0 to 1.4) and multiple myeloma (not detected) — showed no significant change and are discounted rather than counted as evidence.

Skipping of exon 3 is real, rare, and coupled to nonsense-mediated decay

EGFR has 23 Ensembl transcripts; 15 span exons 2 to 4 and **all 15 contain exon 3**. No annotated transcript carries an exon 2 to exon 4 junction. The only boundary variant is ENST00000898199, which uses an acceptor 9 nt downstream (exon 3 = 175 nt), an alternative 3′ splice site rather than skipping. Among RefSeq entries, the only mRNA lacking exon 3 is EGFRvIII, which removes exons 2 to 7 as a single in-frame block.

This has a methodological consequence: annotation-driven tools enumerate candidate events from a supplied transcript annotation, so rMATS [24] and MISO [25] can never report this one. Two public splicing datasets were inspected and neither contains an exon-2-anchored event at this locus; their silence is not evidence against the event.

Source	Sample type	n	e2–e3	e3–e4	skip e2–e4	PSI
GSE311240	iPSC, NPC, cortical neurons	15	417	280	0	1.0000
GSE312830	KNS-60 glioblastoma, IDH-wt	21	476	460	**4**	0.9915
GSE49417	7 ENCODE human cell lines	7	1,753	1,392	0	1.0000

The four skipping reads come from two KNS-60 samples (1 and 3 reads). Both carry canonical GT/AG dinucleotides at the authentic exon 2 donor and exon 4 acceptor, are flagged unannotated, and have maximum overhangs of 23 and 28 nt, so they are genuine spliced reads rather than short-overhang misalignments. Pooled across all 43 samples the figure is 4 skipping reads against approximately 1,836 mean-inclusion reads, **PSI ≈ 0.998**.

Comparing glioblastoma with normal neural cells at read level (348 / 0 against 468 / 4) gives **Fisher's exact test, two-sided, p = 0.14**; by sample presence (0 of 15 against 2 of 21), **p = 0.50**. Skipping reads appeared only in the glioblastoma line and in none of 22 other samples, but four reads cannot establish a difference.

The frame shift explains the scarcity

Exon 2 ends on a codon boundary (240 coding nt consumed). Exon 3 is 184 nt and 184 mod 3 = 1, so skipping shifts the frame by one position. Translating the skipped cDNA gives residues 1 to 80 identical to wild-type EGFR — the signal peptide and the start of domain L1 — then a frameshift at codon 81 and a stop at codon 106, yielding a 105-residue truncation that is neither a soluble ectodomain nor a signalling receptor.

The premature stop lies in exon 4, 2,771 nt upstream of the final exon-exon junction with 24 junctions downstream, far beyond the 50 to 55 nt rule [22,23]. The transcript is an unambiguous substrate for nonsense-mediated decay, which is consistent with 4 reads in roughly 1,840 and means that detecting the product requires NMD inhibition. The companion preprint's use of an NMD inhibitor was therefore necessary rather than incidental, and the public data constrain the untreated steady-state level to well under 1% without contradicting that observation.

U2AF65 occupies the EGFR exon 3 tract, unlike the CD44 v5 tract that depends on YB-1

The route by which YB-1 is known to act on a 3′ splice site is to recruit U2AF65 to a **weak** polypyrimidine tract: in *CD44*, YB-1 binds CAUC motifs in and upstream of variable exon v5 and promotes its inclusion by that mechanism [13]. The tract at *EGFR* exon 3 is not weak, so whether the same route is available can be tested directly by asking how much U2AF65 each tract carries in public CLIP data.

In pooled U2AF65 iCLIP from HeLa cells (GSE99700, six libraries, 7.5×10^7 crosslinks [32]), the 30 nt upstream of the *EGFR* exon 3 acceptor carry **92 crosslinks**, peaking at position −15 and forming a continuous footprint from −21 to −3 — the canonical position for U2AF65 on a polypyrimidine tract. Among the 26 internal exons of *EGFR* this ranks 10th, close to the neighbouring exons 2 (76 crosslinks) and 4 (87) and above the gene's median of about 71. An independent HeLa CLIP-seq dataset (GSE61603, hg18 [33]) gives 143 reads of coverage over the same window, and HepG2 eCLIP gives a low but non-zero signal.

The contrast with *CD44* v5 is the point. Using the same pooled HeLa data, and identifying v5 as ENSE00000824450 (chr11:35,201,671-35,201,787):

3′ splice site	Crosslinks, −30 to −1	Pyrimidine, 18 nt	Longest pyrimidine run
EGFR exon 3	**92**	89%	12
EGFR exon 2	76	89%	12
EGFR exon 4	87	72%	7
CD44 v5	**15**	61%	3

CD44 v5 carries about one sixth of the *EGFR* exon 3 signal, and roughly a third after normalising to total crosslinks at each locus (1.59×10^{-3} against 5.25×10^{-4}). It is the lowest of the nine *CD44* variable exons and lower than all eight constant exons. In GSE61603 the 50 nt upstream of v5 carry zero coverage, the only *CD44* variable exon for which that is true. Both genes are expressed in HeLa (*EGFR* 8,139-9,203 and *CD44* 13,233-15,946 raw counts in matched RNA-seq), so the difference is not one of expression.

At single-nucleotide resolution the difference is one of shape, not only of amount. Positions are numbered with −1 as the last intronic nucleotide, the G of the AG, and +1 as the first exonic nucleotide.

Position	−21	−20	−19	−18	−17	−16	−15	−14	−13	−12	−11	−10	−9	−8	−7	−6	−5
EGFR exon 3	4	6	1	1	1	2	**18**	13	9	3	7	1	1	1	4	7	4
CD44 v5	1	3	1	0	0	0	0	0	1	0	1	1	0	2	0	0	0

EGFR exon 3 carries a single-peaked footprint: 19 consecutive positions from −21 to −3 are occupied, the maximum at −15 accounts for about a fifth of the window's total, and the peak sits in the middle of a 12 nt uninterrupted pyrimidine run (TTTTCTCTTCTT, −20 to −9). *CD44* v5 reaches a maximum of 3 counts, forms no peak, and its longest run of consecutive occupied positions is 3 nt. A structure recognisable as a U2AF65 footprint is present at one site and absent at the other.

The motifs sit where the occupancy does not. *EGFR* exon 3 contains no CAUC or CACC anywhere in the 40 nt upstream of the acceptor; its only CAUC lies at +3 to +6, inside the exon. Conversely, *CD44* v5 carries two overlapping CAUC motifs at −14 to −8, and those positions hold 5 crosslinks in total. What holds U2AF65 at *EGFR* exon 3 is the pyrimidine run itself, not a YB-1 motif.

YB-1 and U2AF65 occupy adjacent but non-overlapping positions

Placing the YB-1 eCLIP peak from Results section 2 on the same axis:

Factor	Dataset	Cells	Occupied positions	Value
U2AF65	GSE99700 iCLIP	HeLa	−21 to −3, peak −15	92 crosslinks, peak 18
U2AF65, secondary	GSE99700 iCLIP	HeLa	+3 to +5	10 crosslinks
YB-1	GSE213046 eCLIP	U251 (GBM)	+1 to +4	$p = 1.3 \times 10^{-3}$, log2FC 3.29

The two are adjacent and do not overlap. U2AF65 occupies the intronic polypyrimidine tract; the YB-1 peak lies wholly across the AG, on the first four nucleotides of the exon. The *CD44* mechanism requires both proteins on the same tract, since YB-1 recruits U2AF65 to it [13]. Here YB-1 has no signal on the tract at all.

Three caveats bound this comparison. The two factors were measured in different cells (HeLa and U251); no reachable dataset profiles both in one cell type. The methods differ in resolution — single-nucleotide iCLIP crosslinks against eCLIP peak intervals — so the 1 nt precision of the U2AF65 maximum and the 4 nt extent of the YB-1 peak are not comparable quantities. And the secondary U2AF65 signal at +3 to +5 coincides with both the exonic CAUC and the YB-1 peak; that coincidence is an observation, and nothing in these data establishes a relation between them.

The same comparison cannot be made at *CD44* v5. In U251, YB-1 peaks are present on eight constant *CD44* exons, reaching $p = 1.6 \times 10^{-40}$ in the 3′UTR, but none on any of the nine variable exons. The most economical reading is that U251 expresses the CD44s isoform, from which the variable exons are spliced out, so v5 is absent from the mature transcript. Absence of a YB-1 peak there therefore cannot be distinguished from absence of substrate.

Sequence and occupancy agree: the tract that depends on YB-1 is short of pyrimidines and sparsely occupied, while the *EGFR* exon 3 tract is pyrimidine-rich and constitutively occupied. Across the 26 internal exons of *EGFR*, however, pyrimidine content alone explains occupancy only loosely (Spearman rho = 0.289, p = 0.15), so tract strength is not the whole account of where U2AF65 sits.

Functional work supports reading the tract as a determinant rather than a correlate. Dominski and Kole showed that beta-globin-derived internal exons of 23 to 33 nt are skipped both in vitro and in transfected cells, and that replacing purines in the upstream polypyrimidine tract with pyrimidines reverses the skipping — completely for a run of 28 pyrimidines, partially for 15 [34]. Tract composition therefore sets whether an exon is

recognised, on a scale where runs of this length are what matters. The longest uninterrupted pyrimidine run is 12 at *EGFR* exon 3 and 3 at *CD44* v5, placing the two tracts at opposite ends of that range and matching the occupancy difference measured above.

The comparison should not be pushed further than it goes. Those experiments rescued exons of 23 to 33 nt, which are too short to be defined on their own and therefore depend on an unusually strong tract; *EGFR* exon 3 is 184 nt and carries no such handicap. What transfers is the principle that pyrimidine content governs recognition, not the specific thresholds.

Two limits apply. HeLa is the only reachable dataset in which both genes are expressed, so the direct comparison rests on one cell line, albeit six pooled libraries; HepG2, K562, HEK293, HEK293T and HEL each fail on expression or depth. And high U2AF65 occupancy does not logically exclude a YB-1 contribution — it shows that this tract is not short of U2AF65, not that YB-1 is dispensable. Settling that needs U2AF65 occupancy measured under YB-1 depletion, which no public dataset provides.

The decoy sequences carry the YB-1 core, and so does much of EGFR

SELEX defines CAUC and CACC as the high-affinity motifs of YB-1, giving the consensus CA(U/C)C [13]. A 1.7 A crystal structure of the YB-1 cold-shock domain bound to the hexamer 5′-U0C1A2U3C4U5-3′ resolves the CAUC core as four pi-pi stacking pairs: C1 with His-87, A2 with Phe-85, U3 with Phe-74 and C4 with Trp-65. Substituting any one of those four residues with alanine abolishes binding to the RNA oligomer [12].

Oligonucleotide	Sequence (5′→3′)	CA(U/C)C	UCAUCU
dYB-1	GUCAUCUU	1	**yes**
dRNA2	ACCAUCAGGAGGAG	1	no
asRNA1	ACCUCCUGGAUGGUCUAAG	0	no
asRNA2	AGCCACCUCCUGGAUGGAG	0	no
targeted exonic sequence	ACCAUCAGGAGGUGGCU	1	no

Two points follow. The dYB-1 sequence contains UCAUCU exactly, the hexamer resolved in the co-crystal structures, so its design rests on structural and selection evidence rather than on a database call alone. And dRNA2, presented in the companion preprint as a decoy complementary to the asRNA2 target, also carries the CA(U/C)C core, inherited from the sequence it was built to mimic. If dRNA2 and dYB-1 sequester the same protein, their comparable effects are an expected result rather than a coincidence. The same core in the targeted sequence is also why SpliceAid returned YB-1 at that position.

The core is too common to permit gene-level selectivity

Region	CA(U/C)C	UCAUCU
EGFR mature mRNA (9,905 nt)	101	7
EGFR gene, pre-mRNA span (192,809 nt)	1,639	94
exon 3 ± 200 nt (584 nt)	3	0

A four-nucleotide core recurs roughly every 200 nt. The site at the 3′ splice site of *EGFR* exon 3 is one of 1,639 in this gene alone, and transcriptome-wide the count runs to millions. A decoy carrying that core competes for YB-1 wherever it binds; it cannot select *EGFR*, and within *EGFR* it cannot select exon 3. This argument is independent of the empirical breadth measured above and reaches the same conclusion from sequence alone.

Discussion

Three conclusions follow, and they do not all point the same way.

The target choice is defensible and the binding premise has support. Applying the design logic to every internal exon of *EGFR* leaves exon 3 alone, and the step that eliminates exon 2 — absence of the YB-1 core at its 3′ splice site — agrees with a database prediction arrived at independently. A deposited eCLIP peak from glioblastoma cells places YB-1 at the first four nucleotides of exon 3, immediately 3′ of the acceptor AG. The position of that peak is where a protein bound at a 3′ splice site would crosslink.

The in-frame hazard deserves to be stated explicitly in this gene. EGFRvIII is the standing demonstration that removing *EGFR* exons in frame can produce a more active receptor, not a less active one. Any splice-switching strategy in *EGFR* should state the frame consequence of the exon it targets, and the companion preprint's instinct to work at the 5′ end happens to coincide with a frameshifting exon. Reporting that as an explicit criterion rather than a happy accident would make such work easier to evaluate.

The specificity premise does not hold. YB-1 sites fall in 43% of genes expressed in U251 cells, the recognised core occurs 1,639 times in *EGFR* alone, and depleting YB-1 moves between 13.6% and 56% of tested genes. A decoy built on that core is a sponge for the protein, not a reagent directed at one transcript. Where a YB-1 decoy lowers EGFR, that is most parsimoniously read as one consequence among many of reducing available YB-1.

A further discrepancy should be faced rather than smoothed over. In three of four glioblastoma knockdown datasets, *EGFR* mRNA rises when YB-1 is depleted, and in the fourth it falls by 17%. If a YB-1 decoy lowers EGFR protein, a transcript-abundance route is not the obvious explanation. YB-1 is a translational regulator, and a translational mechanism would reconcile the observations, but it has not been demonstrated and should not be assumed. Polysome or ribosome profiling under decoy treatment is the experiment that would settle it.

The established splicing mechanism of YB-1 does not transfer cleanly to this exon. In *CD44*, YB-1 promotes inclusion of variable exon v5 by recruiting U2AF65 to a weak polypyrimidine tract [13]; the mechanism is defined by that weakness, which YB-1 compensates for. The public U2AF65 data show the two tracts behaving as that account predicts for *CD44* and against it for *EGFR*: v5 is pyrimidine-poor and sparsely occupied, the lowest of the nine *CD44* variable exons, while the *EGFR* exon 3 tract is pyrimidine-rich and carries about six times the crosslink density, ranking tenth of 26 internal exons in its own gene. A reagent that displaced YB-1 from the *EGFR* site would therefore not be removing a compensatory factor from a tract that needs one. The positions make the same point more directly: at EGFR exon 3 the U2AF65 footprint runs from −21 to −3 while the YB-1 peak lies at +1 to +4, wholly across the splice acceptor and off the tract. The CD44 mechanism requires the two proteins on the same tract, and here they do not share one.

Two readings remain. YB-1 may act at this exon by some route other than U2AF65 recruitment, or its contribution to the definition of this exon may be smaller than the *CD44* precedent suggests. The eCLIP peak reported above establishes contact, not function, and occupancy measured in the absence of YB-1 — which no public dataset provides — is what would separate the two. The experiment is specific and modest: U2AF65 CLIP or RIP at this tract in glioblastoma cells with and without YB-1 depletion.

What the tract implies for SF3b inhibition at the same junction

The U2AF65 occupancy reported here was measured to ask whether the *CD44*-type mechanism could operate at *EGFR* exon 3. It bears on a separate observation reported elsewhere, and the connection is worth stating because the two lines of evidence are independent.

Fukagawa et al. [36], reanalysing sudemycin D1-treated Rh18 cells, found ΔPIR +0.00 at *EGFR* intron 2, ranking 22nd of 26 measurable introns, while introns 1, 11, 13, 14, 3 and 19 each rose by more than 15 percentage points; they also placed the intron 2 acceptor at the 88.2nd percentile of a 6,000-intron panel. Naito et al. [35] found no intron 2-retained product in A172 cells at any herboxidiene concentration tested, while *CDKN1B* introns 1 and 2 were retained in the same cells, which establishes that the spliceosome was engaged and that the absence of retention at *EGFR* was not a failure of the drug to act.

The tract described here offers an account of both. A polypyrimidine tract that is 89% pyrimidine over the 18 nt upstream of the AG, carrying a continuous U2AF65 footprint from −21 to −3 with 92 crosslinks, is a tract that holds U2AF65 without assistance. SF3b inhibitors act by destabilising U2 snRNP engagement at the branch point, and reported sensitivity tracks with weak 3′ splice sites; a tract of this kind leaves correspondingly less to destabilise. The splice site score and the CLIP occupancy are separate measurements, computed from different data by different methods, and they converge on the same property of this junction.

The same property does not predict resistance to a steric oligonucleotide, which does not compete with U2 snRNP for the branch point but occludes the exonic sequence across which the exon is defined. That the oligonucleotides of Nishi et al. [1] reduced EGFR protein at a junction where an SF3b inhibitor produced no retention [35] is consistent with that distinction.

This connection was not tested here and is offered as an interpretation, not a result. Testing it would mean comparing, in one cell line, junctions matched for expression but differing in acceptor strength and U2AF65 occupancy, under both an SF3b inhibitor and a steric oligonucleotide.

The cell-type restriction of the exon 3 peak is the most interesting unresolved point, and the contrast is positional rather than quantitative. In U251, YB-1 engages the 5′ exons of *EGFR*: the two most 5′ exonic peaks are in exon 1 (−log10 P = 6.57) and in exon 3 (−log10 P = 2.90), the latter covering the four nucleotides immediately 3′ of the splice acceptor AG. In HEK293, HEK293T, T24 and an MDA-MB-231 derivative, by contrast, 103 mapped sites include none overlapping exon 2, intron 2 or exon 3, and only two overlapping exon 1, both in T24; exonic binding there is concentrated in exon 28, the 3′UTR, while the 5′ region of the gene is represented by clusters within intron 1, which spans 123 kb. Two of those datasets — in T24 and the MDA-MB-231 derivative — recovered 19 to 31 *EGFR* sites each, so the transcript was sampled. YB-1 therefore reaches the 5′ region of the *EGFR* pre-mRNA in non-glial cells without binding its 5′ exons, which argues against a simple 3′-end preference and for a cell-type-dependent interaction. If YB-1 engages the 3′ splice site of *EGFR* exon 3 only in glioblastoma, the reason is worth knowing: it could reflect the *EGFR* amplification and rearrangement characteristic of these tumours, a difference in YB-1 abundance or localisation, or the presence of a partner protein. None of these is tested here.

Finally, a point about tools. Exon 3 skipping is absent from every annotated *EGFR* transcript, so annotation-driven splicing pipelines cannot report it. Investigators looking for novel skipping events at this locus will find nothing in rMATS or MISO output and should work from junction counts directly. The event is genuine — canonical dinucleotides, long overhangs, unannotated flag — but at a level that only becomes visible when nonsense-mediated decay is inhibited.

Limitations

The decisive dataset is a single deposited peak table, and its limitations are inherited. GSE213046 is exon-annotated, with 26,621 of 66,786 peaks removed before deposition, so intronic binding cannot be assessed at all; the SpliceAid-predicted site, if intronic, was

never assayed. Peak intervals appear to be clipped at exon boundaries, so the 4 nt width of the exon 3 peak is probably not the true footprint. Per-replicate reproducibility is not recoverable from the file, and the unclipped peak calls are not available.

A172, the cell line used in the companion preprint, appears in none of the datasets analysed. It expresses *EGFR* without amplification and carries a tandem duplication of the tyrosine kinase and calcium internalization domains specific to that line [20], for which no public CLIP or junction data exist. U251 and U87 are the nearest comparators.

The exon selection funnel is a post-hoc reconstruction. It was not the procedure by which the target was chosen, and it was built with knowledge of which exon had been selected. It should be read as a check that the choice is defensible, not as evidence that it was derived systematically. Its criteria also embed judgements — the kinase-domain boundary, the 200-codon cutoff — that could reasonably be set elsewhere, though exons 2 and 3 survive any plausible variation of the third criterion.

The junction analysis rests on 4 reads. The glioblastoma-versus-neural comparison does not reach significance and should not be cited as showing a difference. No glioma tumour tissue and no adult brain tissue were available: GTEx and TCGA junction-level data could not be reached from the analysis environment, and only one glioblastoma cell line contributed junction counts. Every sample analysed was NMD-competent, which systematically depresses a PTC-bearing isoform.

ENCODE YBX1 eCLIP in K562 and HepG2 could not be examined. Its absence does not affect the positive finding but leaves the survey of non-glial cells less complete than intended.

The breadth measurements depend on peak-calling thresholds set by the original depositors, which differ between datasets and are not harmonised here. The 42.8% figure should be read as an order of magnitude rather than a precise fraction.

Finally, this study tests premises, not results. It does not reproduce, audit or validate the experiments reported in the companion preprint, which were single runs without independent replication. Nothing here bears on whether those Western blot, RT-PCR and proliferation results are correct.

Declarations

Data availability. This study analysed only publicly deposited data. Accessions are listed in Methods: GSE213046, GSE161523, GSE294052, GSE249897, GSE130781, GSE150924, GSE264220, GSE349934, GSE311747, GSE133620, GSE63604, GSE150925, GSE249896, GSE83664, GSE311240, GSE312830, GSE49417, GSE99700, GSE61603, GSE91570 and GSE126693. Gene models and sequence were obtained from Ensembl release for GRCh38 via the REST API.

Code availability. Analysis scripts, the derived tables and the per-exon output underlying the selection funnel will be deposited in a public repository on submission, together with a record of the exact processed files used from each accession.

Funding. `[to be completed]`

Competing interests. The authors declare no competing interests.

Relationship to preprints from the same department. Three preprints cited here come from the authors' own department, and K.O. is an author of all three: Nishi et al. [1], whose design premises this work examines; Naito et al. [35] and Fukagawa et al. [36], whose observations are discussed in relation to the tract described here. K.-S.K. is a member of the same department. The present work reports no experimental data from any of those studies and does not audit the experiments described in them.

Author contributions. Kwon and Ohe conceived the study. Kwon retrieved the deposited datasets and performed the analyses. Kwon and Ohe wrote the manuscript. Both authors approved the submitted version.

Acknowledgements. `[to be completed]`

Correspondence. Kenji Ohe, Department of Clinical Pharmacotherapy, Faculty of Pharmaceutical Sciences, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, Japan. ohekenji@fukuoka-u.ac.jp (ORCID 0000-0002-1958-7572).

References

1. Nishi Y, Horita Y, Harada H, Mori M, Murata Y, Ohe K. Decoy and antisense RNA oligonucleotides targeting the EGFR exon 3 splice site suppress proliferation of A172 glioblastoma cells. bioRxiv 2026.10.25.761620. https://doi.org/10.64898/2026.10.25.761620
2. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455(7216):1061-8. PMID: 18772890.
3. Brennan CW, Verhaak RG, McKenna A, et al. The somatic genomic landscape of glioblastoma. Cell. 2013;155(2):462-77. PMID: 24120142.
4. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol. 2021;23(8):1231-51. PMID: 34185076.
5. Lee A, Arasaratnam M, Chan DLH, Khasraw M, Howell VM, Wheeler H. Anti-epidermal growth factor receptor therapy for glioblastoma in adults. Cochrane Database Syst Rev. 2020;5(5):CD013238. PMID: 32395825.
6. Eskilsson E, Røsland GV, Solecki G, et al. EGFR heterogeneity and implications for therapeutic intervention in glioblastoma. Neuro Oncol. 2018;20(6):743-52. PMID: 29040782.
7. Robberson BL, Cote GJ, Berget SM. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol Cell Biol. 1990;10(1):84-94. PMID: 2136768.
8. De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. Wiley Interdiscip Rev RNA. 2013;4(1):49-60. PMID: 23044818.
9. Sugawa N, Ekstrand AJ, James CD, Collins VP. Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas. Proc Natl Acad Sci U S A. 1990;87(21):8602-6. PMID: 2236070.
10. Humphrey PA, Wong AJ, Vogelstein B, et al. Anti-synthetic peptide antibody reacting at the fusion junction of deletion-mutant epidermal growth factor receptors in human glioblastoma. Proc Natl Acad Sci U S A. 1990;87(11):4207-11. PMID: 1693434.
11. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA. Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. Oncogene. 2018;37(12):1561-75. PMID: 29321659.
12. Yang XJ, Zhu H, Mu SR, et al. Crystal structure of a Y-box binding protein 1 (YB-1)-RNA complex reveals key features and residues interacting with RNA. J Biol Chem. 2019;294(28):10998-11010. PMID: 31160337.
13. Wei WJ, Mu SR, Heiner M, et al. YB-1 binds to CAUC motifs and stimulates exon inclusion by enhancing the recruitment of U2AF to weak polypyrimidine tracts. Nucleic Acids Res. 2012;40(17):8622-36. PMID: 22730292.
14. Wu SL, Fu X, Huang J, et al. Genome-wide analysis of YB-1-RNA interactions reveals a novel role of YB-1 in miRNA processing in glioblastoma multiforme. Nucleic Acids Res. 2015;43(17):8516-28. PMID: 26240386.
15. Wang JZ, Zhu H, You P, et al. Upregulated YB-1 protein promotes glioblastoma growth through a YB-1/CCT4/mLST8/mTOR pathway. J Clin Invest. 2022;132(8):e146536. PMID: 35239512.
16. Fotovati A, Abu-Ali S, Wang PS, et al. YB-1 bridges neural stem cells and brain tumor-initiating cells via its roles in differentiation and cell growth. Cancer Res. 2011;71(16):5569-78. PMID: 21730024.

17. Law JH, Li Y, To K, et al. Molecular decoy to the Y-box binding protein-1 suppresses the growth of breast and prostate cancer cells whilst sparing normal cell viability. PLoS One. 2010;5(9):e12661. PMID: 20844753.
18. Mantwill K, Naumann U, Seznec J, et al. YB-1 dependent oncolytic adenovirus efficiently inhibits tumor growth of glioma cancer stem like cells. J Transl Med. 2013;11:216. PMID: 24044901.
19. Tailor D, Resendez A, Garcia-Marques FJ, et al. Y box binding protein 1 inhibition as a targeted therapy for ovarian cancer. Cell Chem Biol. 2021;28(8):1206-1220.e6. PMID: 33713600.
20. Fenstermaker RA, Ciesielski MJ, Castiglia GJ. Tandem duplication of the epidermal growth factor receptor tyrosine kinase and calcium internalization domains in A-172 glioma cells. Oncogene. 1998;16(26):3435-43. PMID: 9692551.
21. Piva F, Giulietti M, Nocchi L, Principato G. SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. Bioinformatics. 2009;25(9):1211-3. PMID: 19261717.
22. Nagy E, Maquat LE. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci. 1998;23(6):198-9. PMID: 9644970.
23. Kurosaki T, Popp MW, Maquat LE. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. Nat Rev Mol Cell Biol. 2019;20(7):406-20. PMID: 30992545.
24. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proc Natl Acad Sci U S A. 2014;111(51):E5593-601. PMID: 25480548.
25. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods. 2010;7(12):1009-15. PMID: 21057496.
26. Zheng C, Wei Y, Zhang Q, et al. Multiomics analyses reveal DARS1-AS1/YBX1-controlled posttranscriptional circuits promoting glioblastoma tumorigenesis/radioresistance. Sci Adv. 2023;9(31):eadf3984. PMID: 37540752.
27. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. PMID: 23104886.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. PMID: 25516281.
29. Muzellec B, Teleńczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. Bioinformatics. 2023;39(9):btad547. PMID: 37669147.
30. Dyer SC, Austine-Orimoloye O, Azov AG, et al. Ensembl 2025. Nucleic Acids Res. 2025;53(D1):D948-57. PMID: 39656687.
31. Morales J, Pujar S, Loveland JE, et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature. 2022;604(7905):310-5. PMID: 35388217.
32. Sutandy FXR, Ebersberger S, Huang L, et al. In vitro iCLIP-based modeling uncovers how the splicing factor U2AF2 relies on regulation by cofactors. Genome Res. 2018;28(5):699-713. PMID: 29643205.
33. Shao C, Yang B, Wu T, et al. Mechanisms for U2AF to define 3′ splice sites and regulate alternative splicing in the human genome. Nat Struct Mol Biol. 2014;21(11):997-1005. PMID: 25326705.
34. Dominski Z, Kole R. Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol. 1991;11(12):6075-83. PMID: 1944277.
35. Naito S, Horita Y, Harada H, Mori M, Murata Y, Ohe K. Herboxidiene suppresses EGFR expression in A172 glioblastoma cells without retention of EGFR intron 2. bioRxiv 2026.11.02.762382. https://doi.org/10.64898/2026.11.02.762382

36. Fukagawa W, Horita Y, Harada H, Mori M, Murata Y, Ohe K. SF3b inhibitors suppress proliferation of non-small cell lung cancer cells irrespective of EGFR mutation status. bioRxiv 2026.10.29.762206. https://doi.org/10.64898/2026.10.29.762206