European Radiology
https://doi.org/10.1007/s00330-020-07467-4

NEURO

The T2-FLAIR mismatch sign as a predictor of IDH-mutant, 1p/19q-noncodeleted lower-grade gliomas: a systematic review and diagnostic meta-analysis

Sang Ik Park[1] · Chong Hyun Suh[1] · Jeffrey P. Guenette[2] · Raymond Y. Huang[2] · Ho Sung Kim[1]

Received: 16 July 2020 / Revised: 30 September 2020 / Accepted: 4 November 2020

Abstract

Objectives To evaluate the diagnostic performance of the T2-FLAIR mismatch sign for prediction of isocitrate dehydrogenase (*IDH*)-mutant, 1p/19q-noncodeleted lower-grade gliomas (LGGs) and review studies with false positive results.

Methods The MEDLINE and EMBASE databases were searched up to March 13, 2020, to identify articles reporting the diagnostic performance of the T2-FLAIR mismatch sign for prediction of *IDH*-mutant, 1p/19q-noncodeleted LGGs (*IDH*mut-Noncodel) using the search terms (T2 FLAIR mismatch). Pooled sensitivity, specificity, and correlation coefficient for interobserver agreement were calculated.

Results Twelve studies including a total of 1053 patients were included. The median age was 43 (median; range, 14–56). The pooled sensitivity and specificity were 42% (95% CI, 28–58%) and 100% (95% CI, 88–100%), respectively. According to the HSROC curve, the area under the curve was 0.77 (95% CI, 0.73–0.80). Considerable heterogeneity was possible among the studies in terms of both sensitivity and specificity. A threshold effect was suggested and was considered to explain most of the heterogeneity. Four studies reported false positive results for the T2-FLAIR mismatch sign, including dysembryoplastic neuroepithelial tumor, pediatric-type gliomas, and non-neoplastic lesions. The 2 original articles with false positive results showed the highest sensitivities among the 10 studies included in the quantitative analysis, supporting the probability of the threshold effect. The pooled correlation coefficient was 0.87 (95% CI, 0.73–0.94).

Conclusions The T2-FLAIR mismatch sign had a high specificity and interobserver agreement for the prediction of *IDH*mut-Noncodel. However, the sign demonstrated low sensitivity, and a few studies with false positive cases were also reported.

Key Points

- *The pooled sensitivity and specificity of the T2-FLAIR mismatch sign for prediction of IDH-mutant, 1p/19q-noncodeleted lower-grade gliomas were 42% and 100%, respectively.*
- *Four studies reported false positive results.*
- *The pooled correlation coefficient was 0.87, suggesting almost perfect interobserver agreement.*

Keywords Astrocytoma · Oligodendroglioma · Glioma · Brain neoplasms · Magnetic resonance imaging

Abbreviations

DCE	Dynamic contrast enhanced imaging
DNET	Dysembryoplastic neuroepithelial tumor
DSC	Dynamic susceptibility contrast imaging
FISH	Fluorescence in situ hybridization
FLAIR	Fluid-attenuated inversion recovery
HSROC	Hierarchical summary receiver operating characteristic
IDH	Isocitrate dehydrogenase
*IDH*mut-Codel	*IDH*-mutant, 1p/19q-codeleted lower-grade glioma
*IDH*mut-Noncodel	*IDH*-mutant, 1p/19q-noncodeleted lower-grade glioma
*IDH*wt	*IDH* wild-type lower-grade glioma
IHC	Immunohistochemistry

✉ Chong Hyun Suh
chonghyunsuh@amc.seoul.kr

[1] Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Olympic-ro 33, Seoul 05505, Republic of Korea

[2] Division of Neuroradiology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

Published online: 06 January 2021

LGG	Lower-grade glioma
MP2RAGE	Magnetization-prepared 2 rapid acquisition gradient echoes
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
PRISMA-DTA	Preferred Reporting Items for Systematic Reviews and Meta-analysis of Diagnostic Test Accuracy Studies
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
T2WI	T2-Weighted imaging
WHO	World Health Organization

Introduction

The revised World Health Organization (WHO) 2016 classification for central nervous system tumors added molecular and genetic features as well as microscopic findings for the classification of diffuse lower-grade gliomas (LGGs) [1]. According to the classification, based on the mutation status of the isocitrate dehydrogenase (*IDH*) 1 and 2 genes and the codeletion status of chromosomes 1p and 19q, LGGs are classified into the following: (i) *IDH*-mutant, 1p/19q-codeleted LGGs (*IDH*mut-Codel); (ii) *IDH*-mutant, 1p/19q-noncodeleted LGGs (*IDH*mut-Noncodel); and (iii) *IDH* wild-type LGGs (*IDH*wt) [1]. The outcomes of patients with LGGs are known to be stratified across the subtypes, with the worst outcomes associated with *IDH*wt, the most favorable with *IDH*mut-Codel, and intermediate with *IDH*mut-Noncodel [2].

The T2-FLAIR mismatch sign, initially described by Patel et al, is defined as a complete or near-complete homogeneous high signal intensity on T2-weighted images (T2WI) and a relative suppression of the signal intensities on the fluid-attenuated inversion recovery (FLAIR) sequence [3]. Following the initial description by Patel et al [3] and subsequent validation by Broen et al [4], the T2-FLAIR mismatch sign was reported to demonstrate a near-perfect positive predictive value and specificity for the prediction of *IDH*mut-Noncodel [3–7]. However, it also exhibited relatively lower sensitivities (from 22 to 89%) and a few studies even reported false positive results [3, 8, 9]. In order to use it as a biomarker for *IDH*mut-Noncodel, it is important to ascertain that the T2-FLAIR mismatch sign has a high positive predictive value with little or no false positive results and to identify conditions in which false positive results are observed.

Although two reviews were recently published dealing with this subject [10, 11], none of them has quantitatively evaluated the diagnostic performance and interobserver agreement of the T2-FLAIR mismatch sign. Therefore, this systematic review and meta-analysis was performed to evaluate the diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel and to review the studies that reported the false positive results.

Materials and methods

This meta-analysis was performed according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis of Diagnostic Test Accuracy Studies (PRISMA-DTA) statement [12].

Literature search

The MEDLINE and EMBASE databases were searched up to March 13, 2020, to identify articles that reported the diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel. The search term used was (T2 FLAIR mismatch), and the results were limited to English publications. The references provided in the selected articles were also screened for identification of additional eligible studies.

Study selection

Inclusion criteria

Articles were included based on the fulfillment of all the following criteria: (1) patients with pathologically confirmed LGGs; (2) MRI showing the presence or absence of the T2-FLAIR mismatch sign as an index test; (3) histopathological examination with the *IDH* mutation and 1p/19q codeletion status as a reference standard; (4) sufficient data for reconstruction of 2 × 2 tables in terms of the diagnostic performance of the T2-FLAIR mismatch sign.

Conference abstracts with sufficient data for 2 × 2 tables were included in the meta-analysis. Case reports or case series including 5 patients or fewer were included for the purpose of qualitative but not quantitative synthesis.

Exclusion criteria

Articles were discarded if they fulfilled any of the following criteria: (1) reviews, letters, guidelines, editorials, or errata; (2) insufficient data for the reconstruction of 2 × 2 tables; (3) studies with overlapping cohorts.

Two authors (S.I.P. and C.H.S. [1 and 7 years of experience in performing systematic reviews and meta-analyses, respectively]) independently evaluated the eligibility of the articles, and any disagreement was resolved via discussion with a third author (H.S.K., 22 years of experience in neuro-oncology).

Data extraction and quality assessment

A standardized form was used for extracting the following data.

1. Study characteristics—authors, year of publication, institution, country of origin, study period, study design (prospective vs. retrospective), and type of enrollment (consecutive vs. non-consecutive).
2. Patient and clinical characteristics—number of patients, males:females, mean or median age, age range, and number of patients with *IDH*-mutant gliomas, *IDH*wt, *IDH*mut-Codel, *IDH*mut-Noncodel.
3. Technical characteristics of MRI—magnetic field strength (T), vendor, scanner, head coil, and pulse sequences.
4. Interpretation of MRI—number of readers, reader experience, blinding to *IDH* mutation and 1p/19q codeletion status, definition of the T2-FLAIR mismatch sign, and interobserver agreement.
5. Reference standard—type of *IDH* genes tested (only *IDH1* or both *IDH1* and *IDH2*), *IDH* mutation testing method, and 1p/19q codeletion testing method
6. Diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel—number of true positives, false positives, false negatives, and true negatives.

The quality of the selected studies was evaluated using Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) [13]. The quality was independently evaluated by the two authors (S.I.P. and C.H.S.), and disagreements were resolved through discussion with the third author (H.S.K.).

Data synthesis and analysis

The diagnostic performance of the T2-FLAIR mismatch sign for prediction of *IDH*mut-Noncodel was regarded as the primary outcome of this meta-analysis. Two by two tables were constructed for each study for the calculation of pooled estimates.

Pooled sensitivity and specificity were calculated using the bivariate and hierarchical summary receiver operating characteristic (HSROC) models [14–16]. Coupled forest plots and HSROC curves were constructed for illustrating the results. Deeks' funnel plot with asymmetry test was used for evaluating the publication bias and its statistical significance [17]. For interobserver agreement, pooled correlation coefficient was calculated using a random-effects model with Fisher's Z transformation of correlations [18].

The presence or absence of heterogeneity was assessed using the following methods: (1) visual evaluation of the difference in area between the 95% confidence and prediction regions of the HSROC curve; (2) Cochran's Q test, with $p < 0.05$ suggestive of the possibility of heterogeneity; and (3) Higgins' I^2 index, with a value > 50% suggestive of the possibility of heterogeneity [19]. The threshold effect was assessed using the following methods: (1) visual evaluation of the coupled forest plot to observe the correlation between sensitivity and false positive rate among the studies; and (2) Spearman correlation coefficient between sensitivity and false positive rate, with a value ≥ 0.6 suggestive of a considerable threshold effect [20].

For statistical analyses, the "midas" and "metandi" modules in Stata 15.0 (StataCorp LP) and the "meta" and "mada" packages in the R software version 3.6.2. (R Foundation for Statistical Computing) were used. $p < 0.05$ was considered to denote statistical significance.

Results

Literature search

The study selection process is illustrated in Fig. 1. The initial literature search yielded 83 articles: 28 from the MEDLINE and 55 from the EMBASE databases, respectively. After removing 30 duplicate articles, the remaining 53 articles were screened on the basis of their title and abstract, and 38 articles were excluded. Full texts of the remaining 15 articles were obtained and reviewed, and 3 articles were again excluded because they were either conference abstracts with insufficient data for the reconstruction of 2 × 2 tables ($n = 2$) [21, 22] or outside the field of interest ($n = 1$) [23]. Finally, 12 studies (8 original articles, 3 conference abstracts, 1 case series, and 1 case report) that involved the reports of 1053 patients were included in this study [3–9, 24–28].

Characteristics of included studies

The characteristics of the included patients and studies are shown in Tables 1 and 2, respectively. The median number of patients for each study was 106 (range, 1–154). The mean or median age of the patients per study was 43 (median; range, 14–56). The median number of patients with *IDH*mut-Noncodel was 38 (range, 0–110). Nine studies were retrospective in design [3–9, 26, 27], whereas the other 3 did not report the study design [24, 25, 28]. Patient enrollment was consecutive in 3 studies [5, 24, 25], non-consecutive in 2 studies [26, 27], and not reported in the remaining 7 studies [3, 4, 6–9, 28].

Among the studies included in the meta-analysis, 9 reported 2 readers [3–5, 7–9, 24, 25, 28], whereas 1 did not submit any reader count [6]. Among the 9, the reader experience was reported in 4 studies and ranged from 2 to 19 years [3–5, 9]. The readers were blinded to the reference standard in 6 studies [3–5, 7–9], whereas no such information was reported in the remaining 4 studies [6, 24, 25, 28]. Seven studies [3–6, 8, 9, 24] used the same definition of the T2-FLAIR mismatch sign as

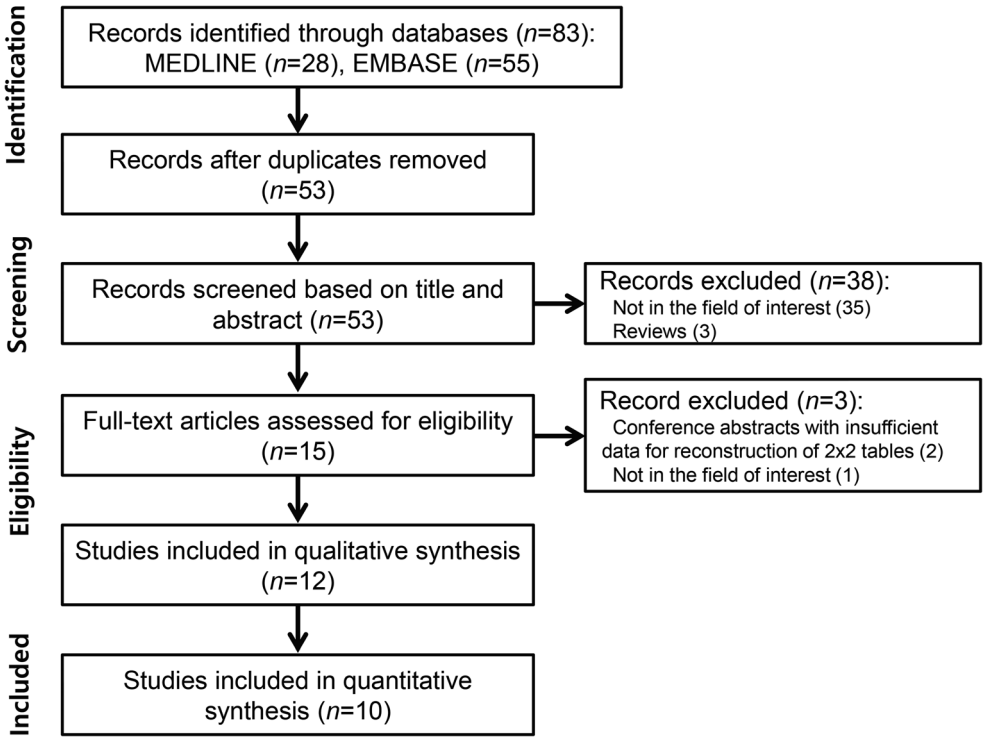

Fig. 1 Flow diagram of the study selection process

initially described by Patel et al as follows: (i) complete or near-complete homogeneous hyperintensity on T2WI and (ii) a hypointensity on FLAIR except for a hyperintense rim [3]. A subjectively determined proportion of T2-FLAIR mismatch of > 50% was considered confirmation of the T2-FLAIR mismatch sign in 1 study [7]. Two conference abstracts did not provide the definition of the T2-FLAIR mismatch sign [25, 28].

The magnetic field strength of the MRI machine was reported in 4 studies [4, 6, 7, 9]: 3 T in 2 studies [6, 9] and 1.5 T or 3 T in 2 studies [4, 7]. Details of MRI machines and pulse sequences are shown in Table 2.

Quality assessment

A quality assessment summary of the included studies using the QUADAS-2 tool is shown in Supplemental Fig. 1. The overall quality of the included studies was moderate.

With regard to patient selection, 8 studies indicated an unclear risk of bias as they failed to mention the method of patient enrollment (consecutive or not) [3, 4, 6–9, 26, 28]. One study caused high concern regarding applicability as the authors had included patients with dysembryoplastic neuroepithelial tumor (DNET) as well as with LGGs [28].

With regard to the index test, 6 studies were considered to have an unclear risk of bias, as they did not mention whether the readers had been blinded to the reference standard [6, 24–28]. One study was considered as being unclear regarding applicability, as the authors had used a different criterion for the T2-FLAIR mismatch sign (> 50% area of T2-FLAIR mismatch) [7].

With regard to the reference standard, all the studies presented an unclear risk of bias as they failed to mention whether they had performed a blinded review of the molecular classification. One study was considered to have high concerns regarding applicability, as it used a different definition of the target condition (i.e., *IDH* mutation with concomitant *TP53*/*ATRX* inactivation, or the absence of 1p/19q codeletions and/or *TERT*p mutations) [8].

With regard to the flow and timing, all the studies were considered to have an unclear risk of bias because they did not mention the imaging to surgery intervals.

Diagnostic performance of the T2-FLAIR mismatch sign

Diagnostic performance of the T2-FLAIR mismatch sign

For evaluating the diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel, 10 studies with 11 cohorts were evaluated [3–9, 24, 25, 28]. The coupled forest plot is presented in Fig. 2. The sensitivity, specificity, and diagnostic accuracy of the individual studies ranged from 10.9 to 89.5%, 69.2 to 100%, and 13.3 to 88.9%, respectively. The pooled sensitivity and specificity were 42% (95% CI, 28–58%) and 100% (95% CI, 88–100%), respectively. According to the HSROC curve, the area under the curve was 0.77 (95% CI, 0.73–0.80) (Fig. 3).

Considerable heterogeneity was possible among the studies in terms of both sensitivity and specificity according to the

Table 1 Patient characteristics

Author (year of publication)	Institution	Period	Study design	Consecutive enrollment	No. of patients (*n*)	Mean age (years)	Age range	Male:female
Kinoshita M, et al (2020) [6]	Osaka International Cancer Institute, Japan	NA	Retrospective	NA	9	46	20–81	4:5
Batchala PP, et al (2019) [5]	University of Virginia Health System, USA	2010–2017	Retrospective	Yes	106	38.5[a]	17–70	53:53
Juratli TA, et al (2019) [8]	Massachusetts General Hospital, USA; University Hospital Dresden, Germany	NA	Retrospective	NA	133	NA	NA	NA
Lee MK, et al (2019) [9]	Asan Medical Center, South Korea	2015 May–2017 May	Retrospective	NA	110	56	19–82	56:54
Broen MPG, et al (2018) [4]	Erasmus MC, University Medical Center Rotterdam and Maastricht University Medical Center, Netherlands	Three different cohorts (2003–, 2013–, 2015–)	Retrospective	NA	154	43	20–82	86:68
Lasocki A, et al (2018) [7]	Peter MacCallum Cancer Centre, Australia	2010 August–2016 August	Retrospective	NA	59	NA	NA	NA
Patel SH, et al (2017) (1) [3]	NYU Langone Medical Center, USA	2011–2014	Retrospective	NA	60	NA	NA	NA
Patel SH, et al (2017) (2) [3]	TCGA/TCIA database	NA	Retrospective	NA	125	45.5[a]	20–75	62:63
Conference abstracts								
Foltyn M, et al (2019) [24]	University of Heidelberg Medical Center, Germany	NA	NA	Yes	113	NA	NA	NA
Galldiks N, et al (2019) [25]	University of Cologne, Research Center Juelich, Germany	NA	NA	Yes	134	NA	NA	NA
Onishi S, et al (2019) [28]	National Hospital Organization Kure Medical Center and Chugoku Cancer Center, Hiroshima University Hospital, Japan	NA	NA	NA	44	NA	NA	NA
Studies not included in meta-analysis								
Johnson DR, et al (2019) [26]	Multi-institutional case series	NA	Retrospective	No	5	14[a]	2–44	5:0
Niemeyer B, et al (2018) [27]	Instituto Estadual do Cérebro Paulo Niemeyer, Brazil	NA	Retrospective	No	1	33	NA	0:1

IDH-Mutant glioma (*n*)	*IDH*mut-Codel (*n*)	*IDH*mut-Noncodel (*n*)	*IDH* wt (*n*)	*IDH*	Reference test (*IDH* mutation)	Reference test (1p/19q-codeletion)
4	2	2	5	*IDH1/2*	NA	NA
106	56	50	0	*IDH1*	IHC, DNA pyrosequencing	FISH
124	42	82[c]	9	*IDH1/2*	NGS, Sanger sequencing, or IHC	FISH
65	46	19	45	*IDH1*	Standard genomic sequencing	FISH
142	67	75	12	*IDH1/2*	NGS or Sanger sequencing	FISH
59	21	38	0	*IDH1*	IHC	FISH
53	31	22	7	*IDH1*	IHC, NGS	PCR loss of heterozygosity
102	34	68	23	*IDH1/2*	Whole-exome sequencing	Affymetrix SNP6.0 arrays
Conference abstracts						
NA	NA	110	NA	NA	NA	NA
65	0	65	69	NA	NA	NA
NA	NA	18	NA	NA	NA	NA
Studies not included in meta-analysis						
1[b]	1	0	1[d]	NA	NA	NA
1	0	1	0	NA	NA	NA

NA, not available; *IDH*, isocitrate dehydrogenase; *IDHmut-Codel*, *IDH*-mutant, 1p/19q-codeleted lower-grade glioma; *IDHmut-Noncodel*, *IDH*-mutant, 1p/19q-noncodeleted glioma; *IDHwt*, *IDH* wild-type lower-grade glioma; *IHC*, immunohistochemistry; *NGS*, next-generation sequencing; *FISH*, fluorescence in situ hybridization; *PCR*, polymerase chain reaction

[a] Median age

[b] Tested in 2 patients

[c] *IDH* mutation with concomitant *TP53/ATRX* inactivation, or the absence of 1p/19q codeletions and/or *TERT*p mutations

[d] Low-grade isomorphic astrocytoma

Table 2 Study characteristics

Author (year of publication)	No. of readers (*n*)	Experience of readers (years)	Blinding	Definition of the T2-FLAIR mismatch sign	Magnetic field strength (T)	Vendor	Scanner	Head coil	Pulse sequences
Kinoshita M, et al (2020) [6]	NA	NA	NA	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	3	Siemens	Prisma	NA	T1 and T2 mapping, MP2RAGE, T2WI, FLAIR
Batchala PP, et al (2019) [5]	2	3, 19	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	NA	NA	NA	NA	NA
Juratli TA, et al (2019) [8]	2	NA	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	NA	NA	NA	NA	FLAIR, T2WI, T1WI, CE T1WI
Lee MK, et al (2019) [9]	2	2, 5	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	3	Philips	Achieva	8-Channel SENSE	T2WI, T1WI, FLAIR, DWI, DSC perfusion, CE T1WI
Broen MPG, et al (2018) [4]	2	3, 10	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	1.5 or 3	NA	NA	NA	FLAIR, T2WI, T1WI, CE T1WI
Lasocki A, et al (2018) [7]	2	NA	Yes	Subjectively determined as proportion of tumor demonstrating high signal on T2WI and substantial suppression on FLAIR; > 50% suggesting T2-FLAIR mismatch	1.5 or 3	NA	NA	NA	NA
Patel SH, et al (2017) (1) [3]	2	3, 17	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	NA	NA	NA	NA	NA
Patel SH, et al (2017) (2) [3]	2	3, 17	Yes	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	NA	NA	NA	NA	NA
Conference abstracts									
Foltyn M, et al (2019) [24]	2	NA	NA	Presence or absence of complete/near-complete hyperintense signal on T2WI and relatively hypointense signal on FLAIR except for a hyperintense peripheral rim	NA	NA	NA	NA	NA
Galldiks N, et al (2019) [25]	2	NA	NA	NA	NA	NA	NA	NA	NA
Onishi S, et al (2019) [28]	2	NA	NA	NA	NA	NA	NA	NA	NA

MP2RAGE, magnetization-prepared 2 rapid acquisition gradient echoes; *FLAIR*, fluid-attenuated inversion recovery; *NA*, not available; *WI*, weighted image; *CE*, contrast enhanced; *DWI*, diffusion-weighted imaging; *DSC*, dynamic susceptibility contrast imaging

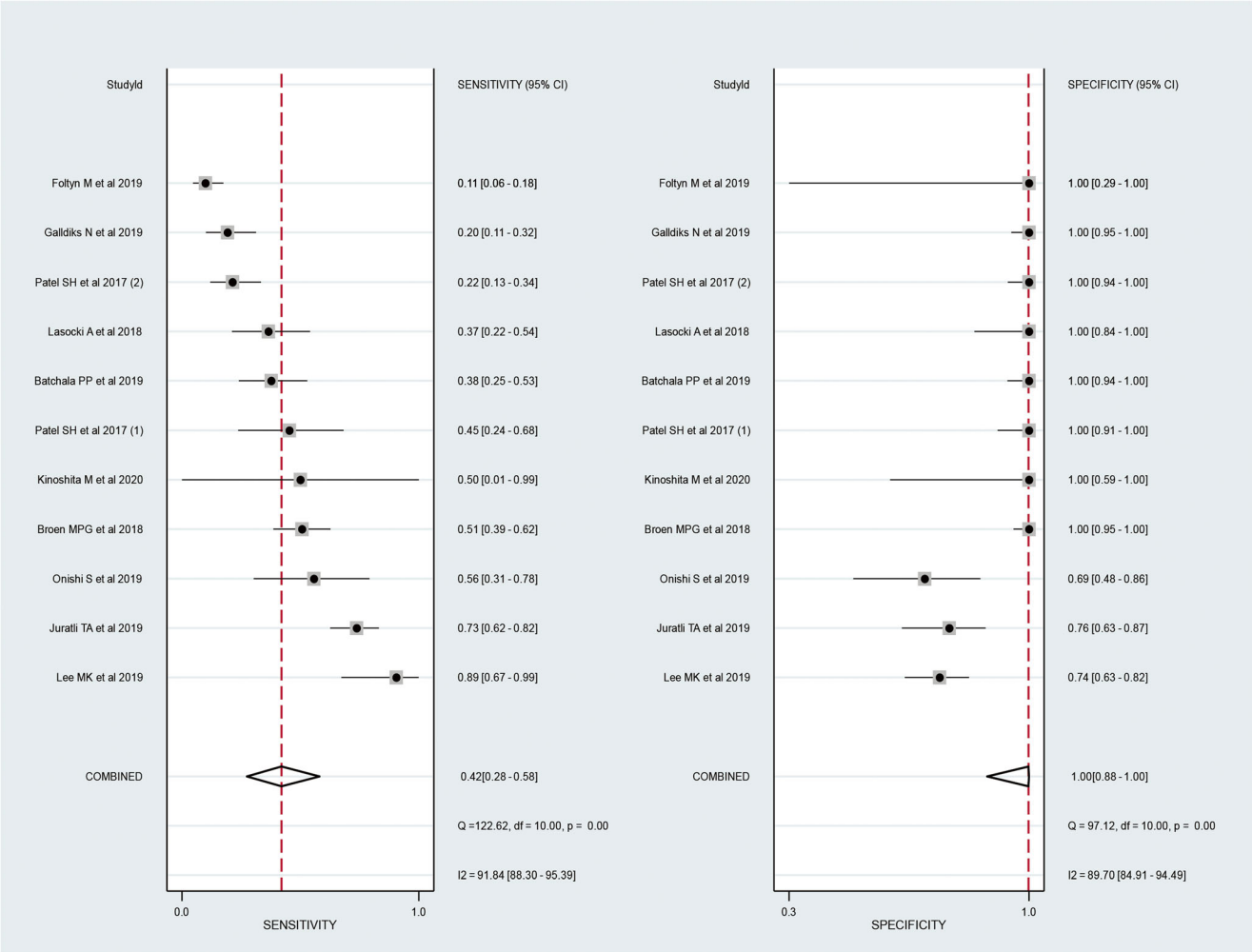

Fig. 2 Coupled forest plot of pooled sensitivity and specificity for evaluating the diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel

Cochran's Q test and Higgins' I^2 statistics ($p < 0.01$ for Q test, $I^2 = 91.8\%$ for sensitivity; $p < 0.01$ for Q test, $I^2 = 89.7\%$ for specificity). A large difference in the area between the 95% confidence and prediction regions on the HSROC curve also indicated possible heterogeneity among the studies (Fig. 3).

The near V-shape of the coupled forest plot on visual evaluation and the Spearman correlation coefficient (between sensitivity and false positive rate) of 0.63 (95% CI, 0.05–0.89) suggested the possibility of a threshold effect, which accounted for most of the heterogeneity among the studies. The Deeks funnel plot with an asymmetry test revealed a low probability of publication bias ($p = 0.11$) (Supplemental Fig. 2).

Studies with false positive results

The T2-FLAIR mismatch sign provided false positive results for the prediction of *IDH*mut-Noncodel in 4 studies (2 original articles [8, 9], 1 conference abstract [28], and 1 case series [26]); 3 of them were included in the meta-analysis [8, 9, 28].

Among the 10 studies included in the quantitative analysis, 2 original articles (Lee et al [9] and Juratli et al [8]) demonstrated the highest sensitivities of 89.5% and 73.2%, respectively (Fig. 2). Lee et al reported that 39.1% (18/46) of *IDH*mut-Codel and 15.6% (7/45) of *IDH*wt cases exhibited the T2-FLAIR mismatch sign [9]. In the study by Juratli et al, 28.6% (12/42) of *IDH*mut-Codel cases were positive for the T2-FLAIR mismatch sign, whereas no *IDH*wt case was positive for the T2-FLAIR mismatch sign [8].

The study by Onishi et al included patients with both DNET and LGGs [28], and 8 out of 11 patients with DNET were positive for the T2-FLAIR mismatch sign. No LGG except *IDH*mut-Noncodel exhibited the T2-FLAIR mismatch sign. In the case series by Johnson et al, there was 1 adult patient with *IDH*mut-Codel, while the other 4 patients were children or young adults with pediatric-type gliomas or non-neoplastic lesions [26].

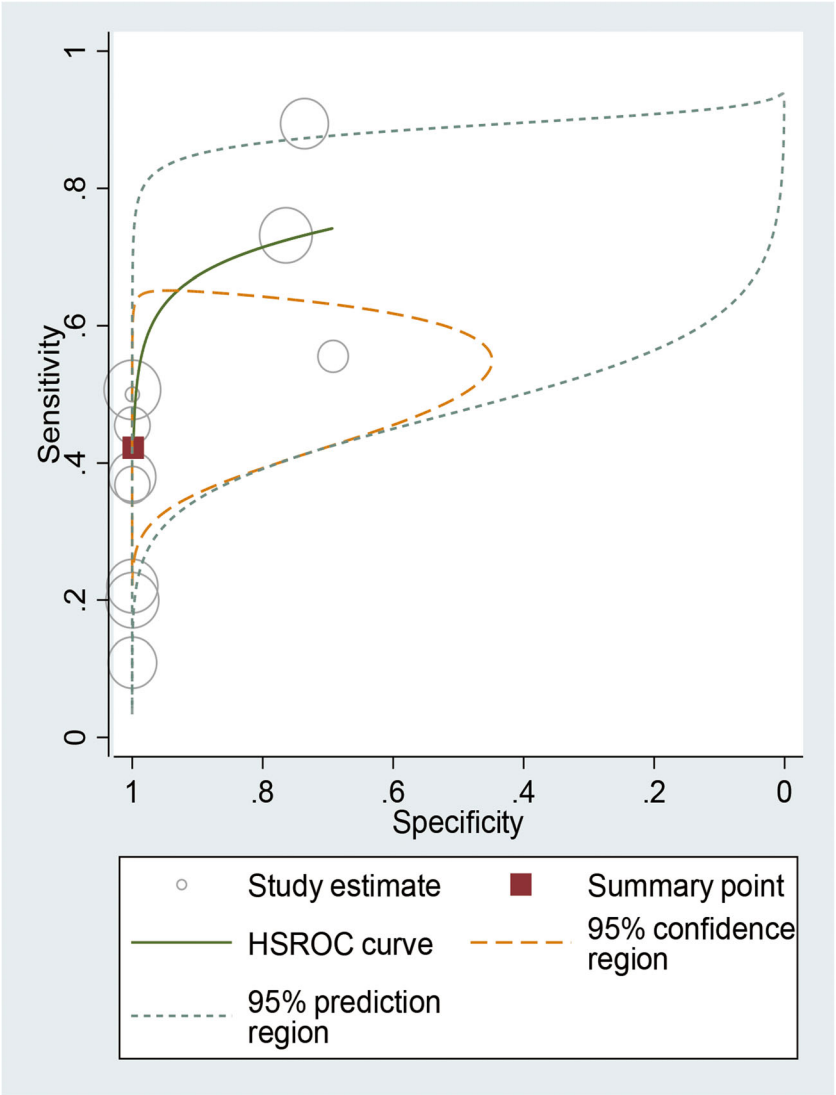

Fig. 3 Hierarchical summary receiver operating characteristic (HSROC) curve of diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel

Interobserver agreement

For testing the interobserver agreement, 9 studies with 10 cohorts were evaluated using a random-effects model for the calculation of pooled correlation coefficient [3–5, 7–9, 24, 25, 28]. The interobserver agreement among the individual studies ranged from 0.56 to 1, and the pooled correlation coefficient was 0.87 (95% CI, 0.73–0.94), suggesting almost perfect agreement (Fig. 4). Cochran's Q test ($p < 0.01$) and the Higgins I^2 statistic indicated the possibility of heterogeneity.

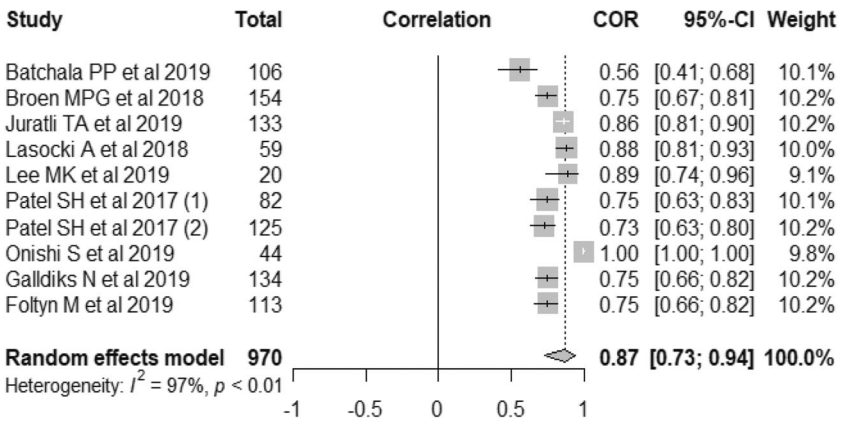

Fig. 4 Forest plot for interobserver agreement

There was no publication bias ($p = 0.11$) (Supplemental Fig. 3).

Discussion

This systematic review and meta-analysis evaluated the diagnostic performance of the T2-FLAIR mismatch sign for prediction of *IDH*mut-Noncodel. The pooled sensitivity and specificity were 42% (95% CI, 28–58%) and 100% (95% CI, 88–100%), respectively. False positive cases of the T2-FLAIR mismatch sign were observed in only 4 studies [8, 9, 26, 28]. The pooled correlation coefficient was 0.87 (95% CI, 0.73–0.94). Therefore, the T2-FLAIR mismatch sign for prediction of *IDH*mut-Noncodel exhibited high specificity and interobserver agreement. However, it also demonstrated low sensitivity, and a few studies showed false positive cases.

As reported in multiple studies and this meta-analysis, the T2-FLAIR mismatch sign has the advantage of near-perfect specificity for predicting *IDH*mut-Noncodel. Thus, it can serve as a radiogenomic biomarker in the diagnosis of LGGs. As the *IDH*mut-Noncodel group has intermediate outcomes among LGGs [3], pretreatment identification of the T2-FLAIR mismatch sign may aid in deciding management options [4, 29]. Furthermore, as *IDH*-mutant astrocytomas with even small tumor remnants after surgery were shown to have worse outcomes, patients displaying the T2-FLAIR mismatch sign in preoperative MRI may need to be treated more radically (i.e., towards gross total resection) [4, 30–32]. However, the low sensitivity of the T2-FLAIR mismatch sign should be considered when using it for imaging classification of a brain tumor—its absence does not exclude the possibility of *IDH*mut-Noncodel. Despite its low sensitivity, the high specificity (i.e., positive predictive value) of the T2-FLAIR mismatch sign makes it a useful predictor for *IDH*mut-Noncodel.

Apart from its high specificity, the T2-FLAIR mismatch sign presents another advantage: that it does not require any advanced imaging techniques such as perfusion-weighted MRI, including DCE (dynamic contrast-enhanced imaging) or DSC (dynamic susceptibility contrast imaging), or MR spectroscopy. The wide availability and high interobserver agreement of the T2-FLAIR mismatch sign make it a useful biomarker for the preoperative diagnosis and classification of LGGs.

Regarding the false positive results, Jain et al argued that they may have been caused by the non-strict application of the T2-FLAIR mismatch sign [11]. It is supported by the results of this meta-analysis that there was the threshold effect accounting for the heterogeneity among the studies and that the 2 original articles with false positive results demonstrated the highest sensitivities among the included studies [8, 9]. As Patel et al initially described, the T2-FLAIR mismatch sign requires (i) complete or near-complete homogeneous hyperintensity on T2WI and (ii) a hypointensity on FLAIR except for a hyperintense rim [3]. Thus, the homogeneity of signal intensity on T2WI and a hyperintense rim on FLAIR are also required for the presence of the T2-FLAIR mismatch sign. The presence of only a discrepancy in the signal intensity on T2WI and FLAIR (such as in a cyst) in itself is insufficient to confirm the T2-FLAIR mismatch sign.

The exact mechanism underlying the presence of the T2-FLAIR mismatch sign in *IDH*mut-Noncodel cases is still to be established. Patel et al reported that higher prevalence of abundant microcysts was observed in tumors with a positive T2-FLAIR mismatch sign, but it failed to reach statistical significance ($p = 0.128$) [3]. One possible explanation is that the T2-FLAIR mismatch sign may reflect the cellularity of the tumor. Further studies with MRI-pathology correlation may provide a clue on the exact mechanism.

The results of this meta-analysis should be applied with caution to routine image interpretation. In the majority of included studies, the target subjects were limited to those with proven LGGs, and not all space-occupying lesions undergoing tissue confirmation. Thus, the observed high specificity was not proven against tumors except LGGs or non-tumorous lesions. For example, patients with dysembryoplastic neuroepithelial tumors (DNETs) were also reported to show the T2-FLAIR mismatch sign [33]. In that study, most patients with a positive T2-FLAIR mismatch sign were aged < 20 years, similarly to the study by Johnson et al [26, 33]. Therefore, the T2-FLAIR mismatch sign should be applied in patients only after considering their age and the possibility of LGGs.

Two questions remained unclear in this systematic review and meta-analysis. First, regarding the outcomes of patients depending on the presence or absence of the T2-FLAIR mismatch sign, only 2 studies reported no significant differences between the mismatch positive and mismatch negative groups [3, 8]. Second, regarding the diagnostic performance of the T2-FLAIR mismatch sign for the prediction of *IDH*mut-Noncodel according to the histologic grade of LGGs, 2 studies provided sensitivities of the T2-FLAIR mismatch sign separately for grade 2 and grade 3 groups: Broen et al (48.6% (34/70) for grade 2, 80% (4/5) for grade 3) [4] and Juratli et al (68.7% (22/32) for grade 2, 76% (38/50) for grade 3) [8]. Future studies with more patients might give clues to these questions.

This study has the following limitations. First, the number of analyzed studies was small, with 3 of them being conference abstracts [24, 25, 28]; thus, additional analyses could not be performed. However, the high specificity of the T2-FLAIR mismatch sign was consistently demonstrated among these studies except for a few exceptions. Second, heterogeneity was observed among the studies and the diagnostic performance profiles of individual studies were highly variable. The heterogeneity was partly attributed to the threshold effect

(inverse relationship between sensitivity and specificity among studies) but could not be analyzed further. Third, the selection criteria among the studies had slight differences (e.g., inclusion of *IDH*wt, inclusion of contrast-enhanced tumors).

Conclusions

The T2-FLAIR mismatch sign had high specificity and inter-observer agreement for the prediction of *IDH*mut-Noncodel. However, low sensitivity was observed, and a few studies had false positive cases.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00330-020-07467-4.

Funding The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Ho Sung Kim.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors (Chong Hyun Suh) has significant statistical expertise.

Informed consent Written informed consent was not required for this study because it was a systematic review and meta-analysis using data from published studies.

Ethical approval Institutional Review Board approval was not required because this study was a systematic review and meta-analysis using data from published studies.

Methodology
• systematic review and meta-analysis

References

1. Louis DN, Perry A, Reifenberger G et al (2016) The 2016 World Health Organization classification of tumors of the central nervous system: a summary. Acta Neuropathol 131:803–820
2. Cancer Genome Atlas Research Network, Brat DJ, Verhaak RG et al (2015) Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. N Engl J Med 372:2481–2498
3. Patel SH, Poisson LM, Brat DJ et al (2017) T2-FLAIR mismatch, an imaging biomarker for IDH and 1p/19q status in lower-grade gliomas: a TCGA/TCIA project. Clin Cancer Res 23:6078–6085
4. Broen MPG, Smits M, Wijnenga MMJ et al (2018) The T2-FLAIR mismatch sign as an imaging marker for non-enhancing IDH-mutant, 1p/19q-intact lower-grade glioma: a validation study. Neuro Oncol 20:1393–1399
5. Batchala PP, Muttikkal TJE, Donahue JH et al (2019) Neuroimaging-based classification algorithm for predicting 1p/19q-codeletion status in IDH-mutant lower grade gliomas. AJNR Am J Neuroradiol 40:426–432
6. Kinoshita M, Uchikoshi M, Sakai M, Kanemura Y, Kishima H, Nakanishi K (2020) T2-FLAIR mismatch sign is caused by long T1 and T2 of IDH-mutant, 1p19q non-codeleted astrocytoma. Magn Reson Med Sci. https://doi.org/10.2463/mrms.bc.2019-0196
7. Lasocki A, Gaillard F, Gorelik A, Gonzales M (2018) MRI features can predict 1p/19q status in intracranial gliomas. AJNR Am J Neuroradiol 39:687–692
8. Juratli TA, Tummala SS, Riedl A et al (2019) Radiographic assessment of contrast enhancement and T2/FLAIR mismatch sign in lower grade gliomas: correlation with molecular groups. J Neurooncol 141:327–335
9. Lee MK, Park JE, Jo Y, Park SY, Kim SJ, Kim HS (2020) Advanced imaging parameters improve the prediction of diffuse lower-grade gliomas subtype, IDH mutant with no 1p19q codeletion: added value to the T2/FLAIR mismatch sign. Eur Radiol 30:844–854
10. Goyal A, Yolcu YU, Goyal A et al (2019) The T2-FLAIR-mismatch sign as an imaging biomarker for IDH and 1p/19q status in diffuse low-grade gliomas: a systematic review with a Bayesian approach to evaluation of diagnostic test performance. Neurosurg Focus 47:E13
11. Jain R, Johnson DR, Patel SH et al (2020) "Real world" use of a highly reliable imaging sign:"T2-FLAIR mismatch" for identification of IDH mutant astrocytomas. Neuro Oncol 22:936–943
12. McInnes MDF, Moher D, Thombs BD et al (2018) Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. JAMA 319:388–396
13. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25
14. Kim KW, Lee J, Choi SH, Huh J, Park SH (2015) Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part I. General guidance and tips. Korean J Radiol 16:1175–1187
15. Lee J, Kim KW, Choi SH, Huh J, Park SH (2015) Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. Korean J Radiol 16:1188–1196
16. Suh CH, Park SH (2016) Successful publication of systematic review and meta-analysis of studies evaluating diagnostic test accuracy. Korean J Radiol 17:5–6
17. Deeks JJ, Macaskill P, Irwig L (2005) The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol 58:882–893
18. Hedges LV, Olkin I (2014) Statistical methods for meta-analysis. Academic Press, Orlando, Florida
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560
20. Deville WL, Buntinx F, Bouter LM et al (2002) Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol 2:9
21. Dao Trong P, Jesser J, Von Deimling A et al (2018) Preoperative predictors of malignancy in non-enhancing glioma in the ERA of molecular classification. Neuro Oncol 20:iii262
22. Graber J, Throckmorton P (2020) Exploring T2-flair mismatch among IDH-mutant astrocytomas: patterns of evolution and further characterization. J Invest Med 68:A137
23. Fujita Y, Sasayama T, Nagashima H, Tanaka K, Hashigutchi M, Kohmura E (2019) The relation between T2-FLAIR mismatch sign and ADC values reflecting pathological microstructure in lower-grade gliomas. Neuro Oncol 21:vi164

24. Foltyn M, Taborda KNN, Neuberger U et al (2019) Non-invasive detection of IDH mutant 1P19Q non-codeleted gliomas using the T2-FLAIR mismatch sign. Neuro Oncol 21:vi161
25. Galldiks N, Werner JM, Stoffels G et al (2019) The T2-FLAIR mismatch sign in IDH-mutant astrocytomas-is there an association with FET PET uptake? Neuro Oncol 21:vi162
26. Johnson DR, Kaufmann TJ, Patel SH, Chi AS, Snuderl M, Jain R (2019) There is an exception to every rule-T2-FLAIR mismatch sign in gliomas. Neuroradiology 61:225–227
27. Niemeyer B, Muniz B, Marchiori E (2018) T2-FLAIR mismatch sign as an imaging biomarker in lower-grade gliomas. Eur Neurol 79:317–318
28. Onishi S, Yamasaki F, Takano M et al (2019) T2WI-flair mismatch sign in lower grade glioma and dysembryoplastic neuroepithelial tumor. Neuro Oncol 21:vi161
29. Soffietti R, Baumert B, Bello L et al (2010) Guidelines on management of low-grade gliomas: report of an EFNS–EANO* Task Force. Eur J Neurol 17:1124–1133
30. Diehn M, Nardini C, Wang DS et al (2008) Identification of non-invasive imaging surrogates for brain tumor gene-expression modules. Proc Natl Acad Sci U S A 105:5213–5218
31. Kawaguchi T, Sonoda Y, Shibahara I et al (2016) Impact of gross total resection in patients with WHO grade III glioma harboring the IDH 1/2 mutation without the 1p/19q co-deletion. J Neurooncol 129:505–514
32. Wahl M, Phillips JJ, Molinaro AM et al (2017) Chemotherapy for adult low-grade gliomas: clinical outcomes by molecular subtype in a phase II study of adjuvant temozolomide. Neuro Oncol 19:242–251
33. Onishi S, Amatya VJ, Kolakshyapati M et al (2020) T2-FLAIR mismatch sign in dysembryoplasticneuroepithelial tumor. Eur J Radiol:108924