diagnostics

MDPI

Systematic Review

Molecular Testing in Oncocytic (Hürthle Cell) Bethesda III/IV Thyroid Nodules: A Systematic Review and Diagnostic Test Accuracy Meta-Analysis

Xiaoyi Wang, Chenxiang Cao, Lin Shi, Jianzhong Xiao and Wenhui Zhao *

Department of Endocrinology, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua Medicine, Tsinghua University, No. 168 Litang Road, Changping District, Beijing 102218, China; wxya05531@btch.edu.cn (X.W.); ccxa00369@btch.edu.cn (C.C.); sla04348@btch.edu.cn (L.S.); xjza01150@btch.edu.cn (J.X.)
* Correspondence: zwha01637@btch.edu.cn; Tel./Fax: +86-10-5611-9057

Abstract

Background/Objectives: Oncocytic (Hürthle) cell-predominant indeterminate thyroid nodules challenge molecular testing; oncocytic-specific accuracy has not been pooled. The objective is to estimate the benign call rate (BCR), diagnostic accuracy and alteration-specific malignancy risk of molecular tests in oncocytic Bethesda III/IV nodules. **Methods:** PubMed, Scopus, Web of Science and Cochrane Library were searched from January 2010 to October 2026 for studies with extractable FNA molecular-testing data for an oncocytic Bethesda III/IV subgroup; 75 of 2774 records were included. This study involved a two-round model-assisted screening and extraction, each value verified by investigators; a stratified re-screening sample of 400 excluded records identified no falsely excluded study (0/300 at title/abstract; 0/100 at full text); one dataset per center and period was used; QUADAS-2 and GRADE were used. **Results:** Bivariate and univariate binomial–normal models estimated the accuracy among the resected nodules (scenario A) and with unoperated test-negative nodules as benign (scenario B). The pooled BCR was 74.2% (95% CI, 67.9–79.6; k = 8 datasets) for Afirma GSC, 59.6% (48.4–69.9; k = 8) for ThyroSeq v3 and 24.7% (20.0–30.1; k = 15) for Afirma GEC. In scenario A, the sensitivity/specificity were 87.5% (76.9–93.6) and 46.5% (36.2–57.1) for GSC (k = 4) and 93.9% (82.9–98.0) and 29.9% (23.4–37.5) for ThyroSeq v3 (k = 9). In scenario B, the specificity was 87.4% (k = 5) and 76.3% (k = 6); NPVs near 99% are upper bounds contingent on unoperated test-negative nodules being benign—if 4% were missed cancers, the NPV would fall to approximately 95%. The pooled malignancy rates were 9% with no alteration, 22% with CNA only and 30% with *RAS*-like alterations; for *TP53/TERT* alterations, the crude rate was 68% (17 of 25). The certainty was low to very low. **Conclusions:** In oncocytic Bethesda III/IV nodules, a negative GSC or ThyroSeq v3 result lowers the malignancy risk, if the unoperated test-negative nodules are benign, whereas a positive result is weakly informative. GSC yields more benign calls and was more specific in indirect comparisons, which may reflect the differences in populations and surgical selection as well as test performance.

Keywords: thyroid nodule; Hürthle cell; molecular testing; Afirma; ThyroSeq

Check for updates

Academic Editor: Mara Carsote

Received: 27 October 2026
Revised: 16 November 2026
Accepted: 23 November 2026
Published: 27 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A substantial minority of thyroid nodules that undergo fine-needle aspiration (FNA) receive an indeterminate cytological diagnosis, either Bethesda category III (atypia of

undetermined significance) or IV (follicular neoplasm). For these nodules, molecular testing of the aspirate has become an accepted means of refining the malignancy risk and of avoiding diagnostic surgery when the result is benign [1,2], and the 2023 Bethesda System now lists it among the usual management options for both categories [3]. Within this indeterminate group, however, nodules composed predominantly of oncocytic (formerly Hürthle) cells remain a particularly difficult subset. The 2023 Bethesda System retains the separate designation of "oncocytic follicular neoplasm" within category IV, for which a risk of malignancy of 25–50% is quoted (30% [range, 23–34%] for category IV overall), and places oncocytic atypia in the "AUS-other" subcategory [3], whereas the 2022 WHO classification recognizes oncocytic carcinoma as an entity distinct from follicular carcinoma and discourages the term "Hürthle cell" [4].

The difficulty is both morphological and biological. Cytology cannot distinguish oncocytic adenoma from oncocytic carcinoma; up to a quarter of aspirates diagnosed as oncocytic neoplasms prove to be hyperplastic nodules or lymphocytic thyroiditis, and oncocytic change also occurs in papillary and medullary carcinoma [5]. At the genomic level, oncocytic carcinoma differs from other differentiated thyroid cancers in that *RAS* and *BRAF* mutations are infrequent [6,7], whereas mitochondrial DNA mutations and widespread chromosomal losses, often amounting to near-haploidization, are the characteristic events [4,6,7]. *TERT* promoter mutations and the extensive loss of heterozygosity, in turn, are associated with invasive and metastatic disease [6,7]. These features pose a challenge for molecular tests that were designed around the mutational and expression profiles of follicular-cell neoplasms in general.

This weakness became evident with the first-generation Afirma gene expression classifier (GEC), which frequently classified oncocytic aspirates as suspicious in its validation study [8]. Later cohorts showed that most GEC-suspicious oncocytic nodules were benign, giving a positive predictive value well below that observed in other indeterminate nodules [1,9]. Two successor tests were subsequently developed: the Afirma genomic sequencing classifier (GSC), which added dedicated Hürthle cell and Hürthle neoplasm classifiers [10], and ThyroSeq v3, which expanded DNA- and RNA-based panel testing to include copy-number alterations [11]. Both, however, were validated in cohorts in which oncocytic nodules were a small minority [10,11], so how well either test performs in this subgroup remains uncertain.

Several meta-analyses have pooled the accuracy of molecular tests across indeterminate nodules [12–15], but none has reported oncocytic-specific sensitivity or specificity, largely because subgroup data are scattered across single-center series and are often reported only in tables or supplements. The only pooled oncocytic estimate to date is a comparison of the benign call rate (BCR) of GSC with that of GEC [16]. Clinicians therefore lack evidence-based estimates of how a benign or positive molecular result should alter the management of an oncocytic Bethesda III/IV nodule. To address this gap, we performed a systematic review and meta-analysis of cytologically oncocytic Bethesda III/IV nodules to estimate the BCR and diagnostic accuracy of current molecular tests, to compare these tests with one another, and to quantify the risk of malignancy associated with specific classes of molecular alteration.

2. Materials and Methods

2.1. Protocol and Reporting

This systematic review and meta-analysis was conducted according to a written protocol (version 1.0) and is reported according to the PRISMA extension for diagnostic test accuracy studies (PRISMA-DTA) [17] and the PRISMA 2020 statement [18]. The protocol was finalized after a pilot screening round and a preliminary analysis, but before the second

screening round and the final data extraction; it is provided in full in the Supplementary Material, together with the rules codified from the pilot round. The review was not registered in PROSPERO or any other public registry. No minimally acceptable accuracy was pre-specified; as a clinical benchmark for a negative (benign) molecular result, we used the 4% (range, 2–7%) risk of malignancy of benign (Bethesda II) cytology [3].

Analytical decisions made after the preliminary analysis are listed in full in Supplementary Protocol Section S10 and are summarized here:

- Fixed after the pilot round and before the second screening round:
  - The boundary-case rules of Protocol Section S4.2;
  - Acceptance of three oncocytic-present (not predominant) subgroups as a documented relaxation of criterion I1 [19–21], with a pre-specified exclusion sensitivity analysis;
  - Replacement of the planned official-call restriction by exclusion of the two studies defining any reportable alteration as positive;
  - Replacement of the NIFTP-as-malignant analysis by exclusion of NIFTP-pooling studies;
  - The one-dataset-per-center-and-period overlap adjudications.
- Decided after the final data extraction, none affecting eligibility or primary outcomes:
  - Gauss–Hermite quadrature increased from 40 to 150 nodes (50 × 50 for bivariate models) for numerical stability;
  - Hierarchical summary ROC curves replaced by summary operating points with confidence regions;
  - Platform meta-regression not performed (platform largely confounded with center and period);
  - GRADE Summary-of-Finding rows added at the writing stage as required by PRISMA-DTA.

2.2. Data Sources and Searches

We searched PubMed/MEDLINE, Scopus, Web of Science Core Collection (Science Citation Index Expanded and Conference Proceedings Citation Index–Science) and the Cochrane Library from 1 January 2010 to 8 October 2026, without language restrictions. The search combined thyroid terms with either named commercial tests (Afirma gene expression classifier [GEC] and genomic sequencing classifier [GSC], ThyroSeq, ThyGeNEXT/ThyraMIR, RosettaGX Reveal) or oncocytic/Hürthle cell terms combined with molecular-testing terms. Full strategies for each database are given in Supplementary Table S2.

2.3. Eligibility Criteria

Studies were eligible if they included thyroid nodules with Bethesda category III (atypia of undetermined significance) or IV (follicular neoplasm) cytology that were oncocytic (Hürthle) cell-predominant, or reported such nodules as a separate subgroup within a broader indeterminate cohort; subgroups defined by histology (oncocytic adenoma or carcinoma) rather than by cytology were accepted but analyzed separately. Subgroups defined by the presence of oncocytic cells rather than their predominance were also accepted where results for indeterminate cytology were reported separately [19–21], and their exclusion was examined in sensitivity analyses. Cohorts spanning Bethesda categories III–V, or enriched through selective testing for *TERT* or *TP53* mutations, were eligible only for the histology-defined, sensitivity and descriptive analyses. The index test could be any molecular test performed on FNA material, including Afirma GEC or GSC, ThyroSeq v2 or v3, ThyGeNEXT/ThyraMIR, RosettaGX Reveal, other next-generation sequencing or mutation panels, and copy-number analysis. The reference standard was surgical histopathology or clinical or sonographic follow-up, although studies reporting only the distribution of

test results were eligible for the analysis of benign call rate. Studies also had to report oncocytic-subgroup counts sufficient to derive at least one of a 2 × 2 table, the positive predictive value (PPV), the BCR or the malignancy rate by molecular alteration class, and had to be original research including at least five oncocytic nodules. Reviews, editorials, case reports, studies limited to Bethesda V/VI or to established carcinoma, studies testing only resected tissue, and conference abstracts duplicating a full report were excluded. Reasons for full-text exclusion were coded hierarchically.

2.4. Selection Process

After deduplication, titles and abstracts were screened with a deliberately inclusive rule: any original study of molecular testing on indeterminate thyroid FNA was retained for full-text review even if the abstract did not mention oncocytic cells, because oncocytic subgroup data are frequently reported only in tables. Screening was performed in two rounds, both with the assistance of a large language model (Claude-Sonnet-5-5, Anthropic, San Francisco, CA, USA) applying the same written, rule-based protocol. The first round served as a pilot, and every inclusion and every uncertain decision was checked against the source by an investigator. The second round was performed by a separately instructed model session that was blinded to the first-round decisions. Agreement between the two rounds was 95.5% ($\kappa = 0.89$) at title/abstract level and 92.3% ($\kappa = 0.76$) at full-text level; these statistics therefore describe the reproducibility of model-assisted screening rather than agreement between two human reviewers. All disagreements were resolved by an investigator re-reading the full text, with adjudication by a senior investigator.

Verification of screening: To bound the false-negative rate of the model-assisted screening, a stratified random sample of 400 excluded records (300 of the 928 title/abstract exclusions and 100 of the 273 full-text exclusions; proportional allocation within original exclusion-code strata; random seed 20261114) was re-adjudicated against the written eligibility criteria, without access to the original decisions. Full texts were re-procured where possible; five conference-abstract records with no retrievable abstract text were adjudicated at record level (Supplementary Table S10a,b).

Re-verification was performed by one author (X.W.) who was not involved in the model-assisted screening rounds and remained blinded to the original decisions: all sampled records with oncocytic or molecular signals were re-adjudicated individually (50 of the 300 title/abstract records and 65 of the 95 full texts, including all seven containing a recognizable oncocytic subgroup or subgroup characteristic); random subsets of the signal-free records were independently re-checked (13 full texts, re-running the full-text keyword search, and 30 title/abstract records; seed 20261115); and the five conference-abstract records without retrievable text were closed at record level. The completed verification worksheet is provided as an audit file.

2.5. Data Extraction

From each study, we extracted the study characteristics (country, center, study period, design, cytological or histological definition of the oncocytic subgroup and Bethesda category), the index test, its version and the threshold used, and the numbers of nodules tested, nodules with a valid result, test-negative (benign) results and nodules resected, together with the histological outcome by test result. Where reported, the molecular alteration class and the corresponding histology were also recorded. Data were first extracted with large-language-model assistance using a standardized form applied to the full text and all Supplementary Files, with each value traced to the specific table or sentence. Every extracted value was then verified against the source document by an investigator, and

ambiguous or low-confidence values were re-checked by a second investigator. Authors were not contacted for missing data.

2.6. Definitions

A positive index test was defined as reported by each study: a "suspicious" Afirma result, and, for ThyroSeq, the official positive call where available, or a result reported as positive by the study (risk of malignancy above 10%, or intermediate/high probability of cancer). Two studies defined any reportable alteration as positive [22,23] and were excluded in a sensitivity analysis. Malignancy comprised all carcinomas, including oncocytic carcinoma, papillary thyroid carcinoma and its variants, follicular carcinoma, and medullary carcinoma. Non-invasive follicular thyroid neoplasm with papillary-like nuclear features (NIFTP) was classified as benign in the primary analysis where separable [24]; studies published before the 2016 reclassification generally pooled NIFTP with malignant outcomes, as did three later studies [20,25,26]; the published classifications were retained, and sensitivity analyses excluded the affected studies from the primary analyses in which they appeared.

Because most test-negative nodules are not resected, diagnostic accuracy was estimated under two scenarios. Scenario A used resected nodules only and therefore underestimates specificity, because nodules with a benign call are under-represented, whereas scenario B additionally counted unoperated test-negative nodules as true negatives and therefore overestimates specificity and the negative predictive value (NPV). The true specificity is expected to lie between the two scenarios, whereas sensitivity and NPV are upper bounds in both, because cancers missed among unoperated test-negative nodules are not counted in either. BCR was defined as the number of test-negative results divided by the number of nodules with a valid result. Molecular alterations were grouped into five classes: negative (test-negative or no alteration detected, including low-level variants reported as negative); copy-number alteration (CNA) only; *RAS*-like alterations (*RAS*, *EIF1AX*, *PAX8::PPARG*, *BRAF* K601E or *THADA* fusion, with or without CNA); high-risk alterations (*TP53* or *TERT* promoter mutations, alone or combined); and *BRAF* V600E or kinase fusions. When a nodule carried alterations from more than one class, classes were assigned hierarchically: *TP53/TERT* first, then *BRAF* V600E or kinase fusions, then *RAS*-like, then CNA only. In total, 27 resected nodules with alterations outside these classes (for example GEA alone, *TSHR*, *EZH1*, *DICER1*, *PAX8::GLIS3* or *BRAF* G466R; 5 of them malignant) were left unclassified and excluded from this analysis.

Handling of clinical heterogeneity: Oncocytic subgroups entered the primary analyses only when defined by cytological oncocytic predominance; three datasets defined by cytological oncocytic presence were accepted as a documented relaxation of criterion I1 [19–21] and tested in a pre-specified exclusion analysis; histology-defined oncocytic cohorts were analyzed only in the pre-specified histology-defined sensitivity analysis. Bethesda III and IV strata were retained separately wherever a study reported them, and platform positivity rules (official classifier call versus any reportable alteration) were recorded and tested in sensitivity analyses.

2.7. Overlapping Populations

Several institutions contributed multiple publications with overlapping periods, for example, UCLA, UPMC, Cleveland Clinic and the University of Miami. For each analysis, we used one dataset per center and period, choosing the dataset with the most complete oncocytic-subgroup data, then the largest, then the most recent. A study could therefore contribute to one analysis but not another. When a study reported Bethesda III and IV strata

separately, each stratum entered the analysis as a separate dataset. The study-level data used in each analysis are listed in Supplementary Tables S3–S5.

2.8. Assessment of Risk of Bias and Applicability

Risk of bias and applicability concerns were assessed with QUADAS-2 [27], with signaling questions tailored to this review. These addressed whether enrolment of oncocytic nodules was consecutive or complete, whether test thresholds were pre-specified, whether pathologists were blinded to molecular results, whether verification was partial or differential, and whether non-diagnostic or unresected nodules were excluded. Judgments for all 75 studies were drafted with large-language-model assistance, each with a written justification, and were reviewed and finalized by two investigators (Supplementary Table S6).

2.9. Statistical Analysis

Sensitivity and specificity were pooled with the bivariate binomial–normal model of Reitsma et al., fitted as the exact binomial–normal generalized linear mixed model (GLMM) of Chu and Cole [28,29], which models within-study binomial variability exactly and accounts for the correlation between sensitivity and specificity across studies. Integration over the random effects used Gauss–Hermite quadrature (150 nodes for univariate and 50 × 50 nodes for bivariate models), and standard errors were obtained from the full observed information matrix, falling back to profile likelihood with the variance parameters fixed when that matrix was singular, as occurs when a variance component is estimated at its boundary. Where the bivariate model could not be fitted, for example, with fewer than four datasets or when all studies had zero false negatives (Afirma GEC), sensitivity and specificity were pooled separately with univariate binomial–normal GLMMs. When all studies had 0% or 100% events, the pooled count is reported with an exact (Clopper–Pearson) 95% confidence interval (CI). BCR, PPV, NPV and malignancy rate by alteration class were pooled with univariate random-effects logistic GLMMs. Positive and negative likelihood ratios and the diagnostic odds ratio were derived from the pooled bivariate parameters, or from the univariate parameters for GEC, in which no study reported a false negative. Summary operating points are shown in summary ROC space with 95% confidence regions; hierarchical SROC curves were not drawn because between-study variance was poorly estimated with few studies.

Platforms were compared with Wald tests of the difference between pooled logit estimates, and the difference in malignancy rate between the CNA-only and *RAS*-like classes was tested descriptively with Fisher's exact test on pooled counts, ignoring between-study clustering. Sensitivity analyses (Supplementary Table S7) added the conference abstract from a high-volume center, excluded studies that defined a positive ThyroSeq v3 result as any reportable alteration, excluded studies that enrolled resected nodules only, analyzed histology-defined cohorts separately, excluded a study in which NIFTP could not be separated from malignant outcomes, excluded studies that pooled NIFTP with malignancy from each primary analysis in which they appeared, stratified the analysis by Bethesda category and excluded the study that defined any reportable alteration as a positive result from the ThyroSeq v3 benign call rate, and omitted each study in turn (leave-one-out analysis).

Small-study effects were assessed with Deeks' funnel plot asymmetry test [30] across all accuracy datasets, with test platform as a covariate. Verification bias was considered qualitatively [31].

Certainty of evidence for sensitivity and specificity was rated with the GRADE approach for diagnostic tests [32–34]. Results were expressed as true and false positives and negatives per 1000 oncocytic Bethesda III/IV nodules at an assumed malignancy prevalence

of 20%, slightly below the risks quoted for Bethesda category III (22%) and category IV (30%) nodules [3] and close to the pooled prevalence among resected GEC-tested oncocytic nodules (20%). All analyses were performed in Python 3.12.5 (NumPy 2.0.1, SciPy 1.17.0, statsmodels 0.14.6) with custom code that was written with large-language-model assistance and checked by an investigator against the source data; the code is provided as Supplementary Data. Two-sided $p < 0.05$ was considered statistically significant.

3. Results

3.1. Study Selection

The searches identified 2774 records. After the removal of 1451 duplicates, 1323 titles and abstracts were screened and 928 were excluded. Of the 395 reports sought, 47 could not be retrieved; most of these (39) were conference abstracts without accessible text. Of the 348 full-text reports assessed, 273 were excluded (Figure 1; Supplementary Table S8). The most common reason was the absence of a separable oncocytic subgroup ($n = 187$). In a further 31 reports, an oncocytic subgroup was mentioned but no extractable counts were given.

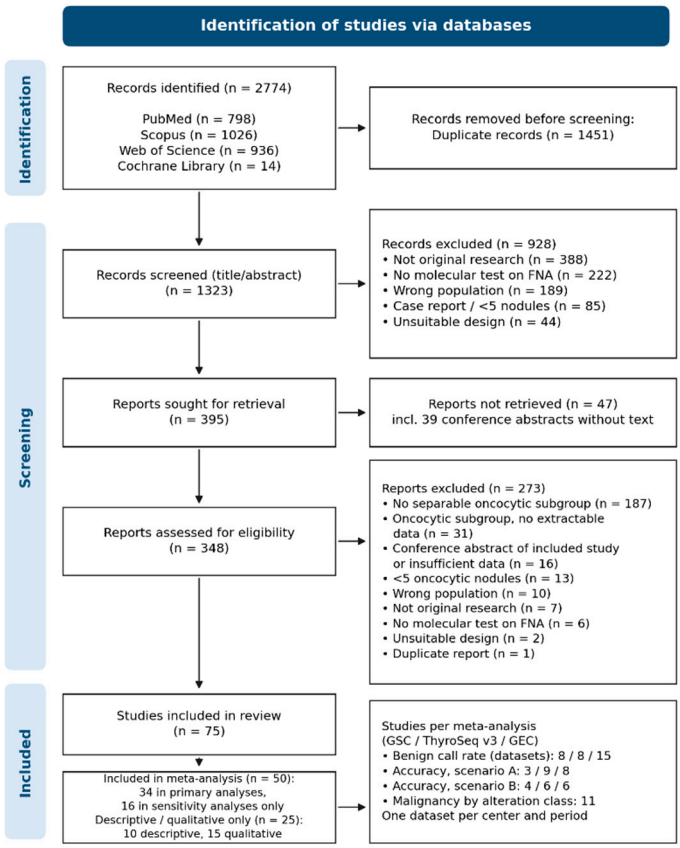

Figure 1. PRISMA 2020 flow diagram of study selection. Reasons for full-text exclusion are listed in Supplementary Table S8.

Seventy-five studies were included [8–11,19–23,25,26,35–98]. Fifty contributed to at least one meta-analysis: 34 to the primary analyses and 16 to sensitivity analyses only. Ten reported only tests other than Afirma or ThyroSeq v3, or provided data suitable only for descriptive synthesis. The remaining 15 were summarized qualitatively.

Verification of screening: No sampled exclusion was falsely excluded: 0 of 300 records at title/abstract (95% CI 0.0–1.2%) and 0 of 95 re-verifiable full texts (0.0–3.8%), with five

conference-abstract records closed at record level (0 of 100 overall; 50 abstracts required individual adjudication; Supplementary Table S10a,b). Seven re-verified full texts contained an oncocytic subgroup or subgroup characteristic, and each was confirmed ineligible: in one the subgroup had fewer than five nodules, in two oncocytic-dominant cytology was reported only as a baseline characteristic, and in four binary outcome data were not extractable. Two further studies had excluded oncocytic cytology by design. The reports and the reasons for exclusion are given in Supplementary Tables S8 and S10b. An audit of the 149 records originally excluded as 'full text not available' confirmed that the 15 provisionally carried from the pilot round all appear in the final included set. All signal-positive records and the random subsets were re-verified by one author (X.W.) blinded to the original decisions; no disagreement arose in any tier.

3.2. Study Characteristics

The included studies were published between 2010 and 2026, and 57 (76%) appeared from 2017 onward (Table 1). Sixty-five were conducted exclusively in the United States and one in the United States and Singapore, most of them at a small number of high-volume academic centers. Fifty-six were retrospective cohorts; the remainder comprised six test validation studies, five prospective cohorts, three secondary analyses of randomized trials, three case series and two conference abstracts. The oncocytic subgroup was defined cytologically in 57 studies and histologically in 18. Thirty-two studies included both Bethesda III and IV nodules, 20 included Bethesda IV only and 14 Bethesda III only, and 9 included Bethesda V or other categories. Afirma GEC was the most frequently studied test (31 studies), followed by ThyroSeq v3 (25), Afirma GSC (19) and ThyroSeq v2 (7). Together, the studies reported molecular testing in approximately 3780 oncocytic nodules, of which about 1850 were resected; these totals are approximate because some cohorts overlapped. Sources of funding and declared conflicts of interest for the 75 included studies are summarized in Supplementary Table S9; 21 studies (28%) disclosed ties to a manufacturer of an evaluated test (Veracyte, $n = 14$; UPMC/ThyroSeq, $n = 8$; Interpace, $n = 1$; two studies disclosed ties to more than one manufacturer), including several of the pivotal platform-validation cohorts. Where reported, invalid or uninterpretable results were uncommon: 7 of 86 Afirma GSC results (8%) and 2 of 63 ThyroSeq v3 results (3%) were insufficient for analysis in the trial cohort reported by Raghunathan et al. [50], and 1 of 9 Afirma GEC results in the Hürthle-cell subgroup was non-diagnostic [90]; other studies did not report invalid-result rates.

3.3. Risk of Bias and Applicability

The risk of bias was most often high in the flow and timing domain (61 of 75 studies, 81%), reflecting partial verification: test-negative nodules were rarely resected (Supplementary Figures S1 and S2; Supplementary Table S6). In the reference standard domain, the risk was unclear in 54 studies (72%) because the blinding of the pathologists to the molecular results was seldom reported, and high in 12. Patient selection was at a high risk of bias in 35 studies (47%), mainly because of the resected-only enrolment or non-consecutive sampling. The index test domain was at low risk in 64 studies (85%). Applicability concerns for the index test were high in 43 studies (57%), chiefly because earlier test versions (GEC, ThyroSeq v2) are no longer in clinical use.

Table 1. Overview of the 75 included studies and of the datasets contributing 2 × 2 tables to the primary diagnostic accuracy analyses. The complete study-level characteristics table is provided as Supplementary Table S1.

Panel A. Summary Characteristics of all 75 Included Studies.		
Characteristic	**75 Included Studies**	**Note**
Inclusion periods (start years)	2001–2019	Complete list: Supplementary Table S1
Countries/regions	USA (65), South Korea (2), other or multinational (8)	
Design	Retrospective cohort (56), prospective (8), RCT (2), other (9)	
Oncocytic subgroup definition	Cytology-predominant (43), cytology-present (3; I1 relaxation), histology-defined (18), mixed/other (11)	Histology-defined cohorts analyzed only in the pre-specified sensitivity analysis
Bethesda category	III/IV (32), IV (20), III (15), V or other (8)	
Tests evaluated	ThyroSeq v3 (19), Afirma GEC (15), Afirma GSC (6), other panels (35)	A study may evaluate more than one platform

Panel B. Datasets Contributing 2 × 2 Tables to the Primary Diagnostic Accuracy Analyses.					
Study	**Platform**	**Bethesda**	**Oncocytic Definition**	**Analyses Contributed**	**Nodules, Scenario A/B**
Ahmadi 2026 [35] (III stratum)	GSC	III	oncocytic-present cytology	BCR, A, B	50/198
Ahmadi 2026 [35] (IV stratum)	GSC	IV	oncocytic-present cytology	BCR, A, B	56/149
Elsheikh 2026 [25]	GSC; TSv3	III/IV	oncocytic cytology (resected)	A	49 (GSC); 11 (TSv3)
Jin 2022 [56]	GSC	III	Hürthle-predominant AUS	BCR, A, B	6/17
Endo 2019 [20]	GSC; GEC	III/IV	oncocytic-present cytology	BCR (GSC); B	17 (GSC); 58 (GEC)
San Martin 2020 [69]	GSC; GEC	III/IV	oncocytic cytology	BCR; B	17 (GSC); 24 (GEC)
Raghunathan 2024 [50]	GSC; TSv3	III/IV	Hürthle cytology (RCT subset)	BCR; A and B (TSv3)	24/53 (TSv3)
Brauner 2015 [9]	GEC	III/IV	Hürthle-cell neoplasm	BCR, A, B	46/69
Parajuli 2019 [74]	GEC; panel	III/IV	Hürthle-predominant	A, B	20/31
Lastra 2014 [93]	GEC	IV	suspicious for Hürthle-cell neoplasm	A, B	15/21
McIver 2014 [94]	GEC	IV	Hürthle-cell neoplasm	A	13/-
Zhang 2021 [64]	GEC	III	AUS with Hürthle cells	A, B	8/10
Chaudhary 2016 [83]	GEC	IV	Hürthle-cell neoplasm	BCR	12/-
Glass 2021 [62]	GEC; panel	III	AUS with oncocytic features	A; B (panel)	4 (GEC); 5/14 (panel)

Table 1. *Cont.*

Panel B. Datasets Contributing 2 × 2 Tables to the Primary Diagnostic Accuracy Analyses.					
Study	**Platform**	**Bethesda**	**Oncocytic Definition**	**Analyses Contributed**	**Nodules, Scenario A/B**
Landau 2021 [23]	TSv3	IV	follicular neoplasm, Hürthle-cell type	BCR, A, B, alterations	79/140
Tjendra 2023 [54]	TSv3	IV	oncocytic follicular neoplasm	A, B, alterations	27/44
Gajzer 2022 [55]	TSv3	III	AUS, Hürthle-cell predominance	A, B, alterations	19/82
Hsiao 2025 [44]	TSv2/v3	III	AUS, oncocytic	A, B	21/62
Abi-Raad 2022 [22]	TSv3	III/IV	Hürthle cytology	A	31/-
Spagnuolo 2022 [57]	TSv3	III/IV	Hürthle cytology	A, alterations	29/-

Nodules are scenario A/scenario B totals of each 2 × 2 table (TP + FP + FN + TN). Benign-call-rate datasets without a 2 × 2 table are listed in Supplementary Table S4. Datasets feeding only the alteration-class analysis (Table 2): Radkay 2014 [96] (41 resected nodules), Schatz-Siemers 2019 [75] (33), Gilani 2021 [61] (42), Geetha 2025 [43] (18). A, scenario A; B, scenario B; Studies are ordered by publication year (newest first). "Use in this review" lists every analysis to which the study contributes; overlapping cohorts were handled by including one dataset per center and period in each analysis (most complete subgroup data, then largest, then most recent). Fifty studies contributed to at least one meta-analysis (34 to primary analyses, 16 to sensitivity analyses only); studies using other or older tests were summarized descriptively; the remainder were synthesized qualitatively. Endo 2019 [20] did not report overall tested or resected counts; its GSC ($n = 18$) and GEC ($n = 58$) oncocytic subsets were used. Stack 2026 [38] tested 19 of 47 resected FN/SFN-O nodules. Abbreviations: BCR, benign call rate; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier.

3.4. Benign Call Rate

The pooled BCR differed substantially between platforms (Figure 2; Table 3). It was 74.2% (95% CI, 67.9–79.6; 8 datasets from 7 studies; 379 of 503 nodules) for Afirma GSC, 59.6% (48.4–69.9; 8 datasets; 318 of 550) for ThyroSeq v3 and 24.7% (20.0–30.1; 15 studies; 108 of 446) for Afirma GEC. GSC had a higher BCR than ThyroSeq v3 ($p = 0.017$), and both had higher BCRs than GEC ($p < 0.001$ for each). In smaller datasets, the BCR was 74.5% (39.3–92.9; 2 studies) for ThyGeNEXT/ThyraMIR and 74.0% (52.3–88.1; 3 datasets) for ThyroSeq v2 or earlier. For ThyroSeq v3, the BCR was higher in Bethesda III (74.6%; 4 studies) than in Bethesda IV (50.9%; 2 studies) oncocytic nodules, and adding the high-volume abstract from the University of Pennsylvania [21] changed the pooled estimate only slightly, to 59.6% (49.9–68.6). Excluding the study that counted any alteration as a positive result [22] raised the ThyroSeq v3 BCR to 63.4% (53.7–72.1), and its difference from GSC became borderline ($p = 0.047$).

The platform pools differ in case mix: the malignancy prevalence among resected nodules was 40% in GSC, 30% in ThyroSeq v3 and 20% in GEC datasets, and Bethesda composition, enrolment periods and positivity rules also differed (Supplementary Table S13). The only within-population comparisons support the direction of the pooled differences: in the randomized trial of Raghunathan et al., the benign call rate was 84% (66/79) for GSC versus 56% (34/61) for ThyroSeq v3 in the same patients, and Elsheikh 2026 [25] reported both platforms in overlapping Cleveland Clinic cohorts. Between-platform differences in the pooled analyses therefore cannot be attributed to intrinsic test performance.

Table 2. Risk of malignancy by molecular alteration class among resected oncocytic Bethesda III/IV nodules.

Alteration Class	Definition	Studies (k)	Malignant/ Resected	Pooled ROM, % (95% CI)	τ^2
Negative (no alteration)	Test-negative; no alteration detected (includes low-level variants reported as negative)	8	5/54	9.3 (3.9–20.4)	0.00
CNA only	Copy-number alteration(s) without a co-existing mutation or fusion	10	15/64	21.6 (9.7–41.5)	0.09
RAS-like	*N/H/KRAS*, *EIF1AX* alone, *PAX8::PPARG*, *BRAF K601E*, *THADA* fusion (±CNA)	11	53/144	30.2 (17.9–46.3)	0.68
High-risk (*TP53/TERT*)	*TP53* and/or *TERT* promoter mutation (±other alterations)	9	17/25	88.5 (18.8–99.6)	7.53
BRAF V600E/fusion	*BRAF* V600E or *RET/NTRK/ALK* fusion	4	6/6	100.0 (54.1–100.0)	–

Data from ThyroSeq (v2/v3) and other mutation panels (including Afirma Xpression Atlas) in cytology-defined oncocytic subgroups; one dataset per center. Classes were assigned hierarchically when several alterations co-existed (*TP53/TERT* > *BRAF* V600E or fusion > *RAS*-like > CNA only); 27 resected nodules with other alterations (e.g., GEA alone, *TSHR*, *EZH1*, *DICER1*, *PAX8::GLIS3*, *BRAF* G466R; 5 malignant) were unclassifiable and excluded. CNA only vs. *RAS*-like, pooled counts, Fisher exact $p = 0.078$. The pooled estimate for *TP53/TERT* alterations is the typical-study value from a random-effects model with large heterogeneity; the crude rate was 17 of 25 (68%). Excluding the pre-reclassification study of Radkay et al. lowered the pooled *RAS*-like risk to 27.2% (15.0–44.1; 41 of 122; CNA-only vs. *RAS*-like, $p = 0.18$). The wide CI for the high-risk class reflects between-study heterogeneity (two small series with mostly benign *TP53/TERT*-positive nodules). Alteration classes partly derive from selectively tested samples (clinician-ordered Afirma Xpression Atlas and ThyroSeq v2 in Elsheikh 2026 [25] and a matched 33-positive/33-negative sample in Geetha 2025 [43]); excluding Elsheikh 2026 [25] lowered the pooled *RAS*-like risk to 27.9% (17.4–41.6; 40 of 126). Abbreviations: CI, confidence interval; CNA, copy-number alteration, ROM, risk of malignancy.

Table 3. Pooled benign call rate and diagnostic accuracy in oncocytic Bethesda III/IV nodules.

A. Benign Call Rate

Test	Studies (Datasets)	Nodules Tested	Benign Calls	Pooled BCR, % (95% CI)	τ^2
Afirma GSC	7 (8)	503	379	74.2 (67.9–79.6)	0.06
ThyroSeq v3	8 (8)	550	318	59.6 (48.4–69.9)	0.34
Afirma GEC	15 (15)	446	108	24.7 (20.0–30.1)	0.07
Thy-GeNEXT/Thyra-MIR	2 (2)	41	29	74.5 (39.3–92.9)	0.84
ThyroSeq v2 or earlier	3 (3)	224	145	74.0 (52.3–88.1)	0.43

B. Diagnostic Accuracy

Scenario/Test	Studies (Datasets)	Nodules	Sensitivity	Specificity	PPV	NPV	LR+	LR−	DOR
Scenario A—Afirma GSC	3 (4)	161	87.5 (76.9–93.6)	46.5 (36.2–57.1)	49.5 (24.6–74.7)	89.6 (58.9–98.1)	1.64	0.27	6.1
Scenario A—ThyroSeq v3	9 (9)	246	93.9 (82.9–98.0)	29.9 (23.4–37.5)	35.3 (27.6–43.8)	91.4 (78.1–96.9)	1.34	0.20	6.5
Scenario A—Afirma GEC *	8 (8)	136	100.0 (87.2–100.0)	10.1 (5.7–17.3)	20.2 (10.4–35.5)	100.0 (71.5–100.0)	1.11	0.00	–

Table 3. *Cont.*

Scenario B—Afirma GSC	4 (5)	398	91.2 (76.0–97.1)	87.4 (83.5–90.4)	44.7 (24.6–66.6)	99.1 (97.1–99.7)	7.21	0.10	71.4
Scenario B—ThyroSeq v3	6 (6)	395	94.8 (69.8–99.3)	76.3 (65.3–84.6)	33.3 (23.1–45.3)	99.1 (94.4–99.9)	4.00	0.07	58.2
Scenario B—Afirma GEC *	6 (6)	212	100.0 (90.0–100.0)	40.7 (33.7–48.1)	21.5 (11.9–35.5)	100.0 (95.0–100.0)	1.69	0.00	–

Scenario A uses nodules with surgical histology only (verification bias lowers specificity and NPV). Scenario B additionally counts unoperated test-negative nodules as true negatives (bounds specificity and NPV from above). Studies (datasets): number of contributing studies (number of contributing 2 × 2 tables). Pooled sensitivity and specificity from the bivariate binomial–normal model (Chu & Cole); PPV, NPV and BCR from univariate binomial–normal models. * Univariate models because no study reported a false negative. NIFTP was counted as benign where separable. Between-test comparisons (Wald test on the logit scale): BCR GSC vs. ThyroSeq v3 $p = 0.017$; specificity GSC vs. ThyroSeq v3, scenario A $p = 0.010$, scenario B $p = 0.016$; sensitivity GSC vs. ThyroSeq v3, scenario A $p = 0.26$, scenario B $p = 0.64$. Abbreviations: BCR, benign call; DOR, diagnostic odds ratio; GEC, Gene Expression Classifier; GSC, Genomic Sequencing Classifier; PPV, positive predictive value; NPV, negative predictive value; LR, likelihood ratio.

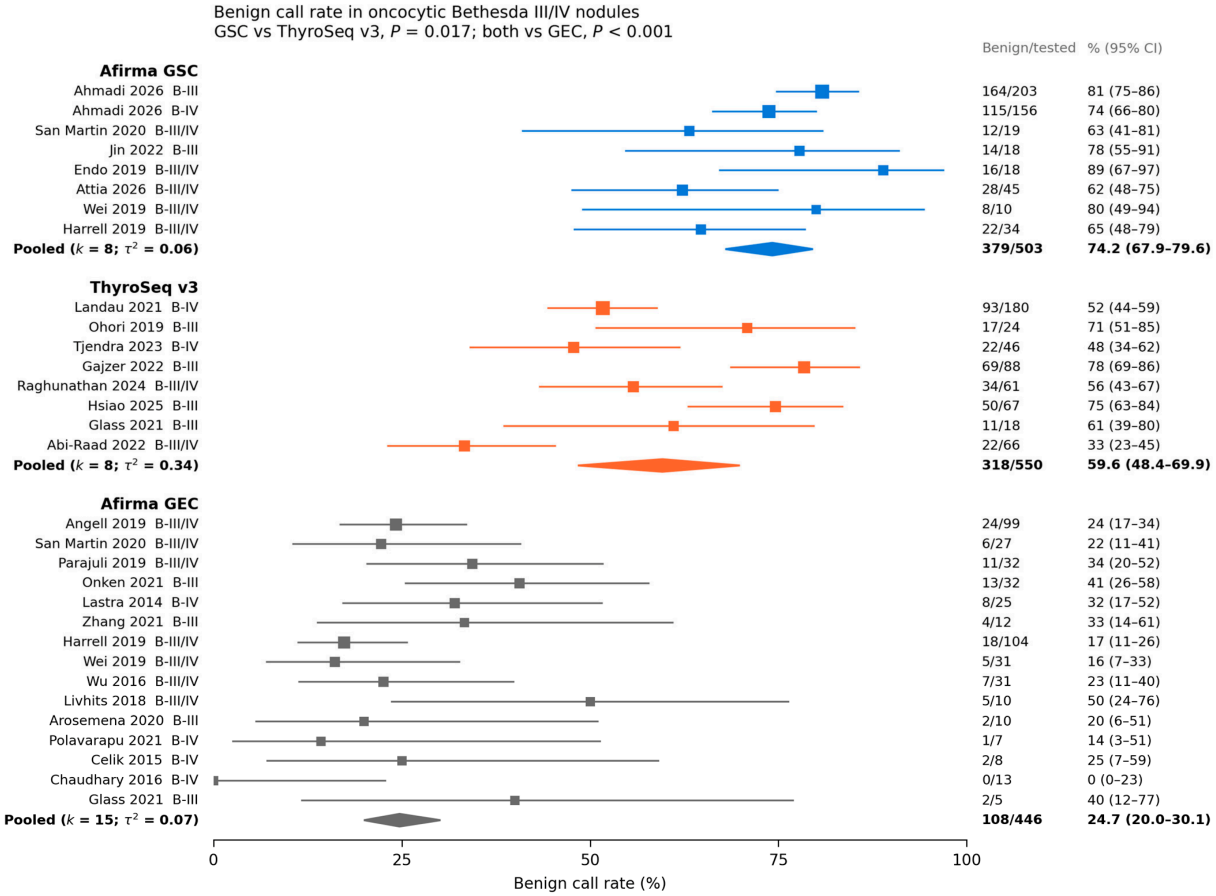

Figure 2. Forest plot of the benign call rate (BCR) of Afirma GSC, ThyroSeq v3 and Afirma GEC in cytologically oncocytic Bethesda III/IV nodules. Squares, study estimates (size proportional to the number tested) with Wilson 95% CIs; diamonds, pooled estimates from univariate binomial–normal random-effects models. One dataset per center and period was included; studies reporting Bethesda III and IV separately contribute two rows. GSC vs. ThyroSeq v3, $p = 0.017$; GSC and ThyroSeq v3 vs. GEC, $p < 0.001$. Datasets (top to bottom): Afirma GSC: Ahmadi 2026 [35], Ahmadi 2026 [35], San Martin 2020 [69], Jin 2022 [56], Endo 2019 [20], Attia 2026 [19], Wei 2019 [76], Harrell 2019 [72]; ThyroSeq v3: Landau 2021 [23], Ohori 2019 [73], Tjendra 2023 [54], Gajzer 2022 [55], Raghunathan

2024 [50], Hsiao 2025 [44], Glass 2021 [62], Abi-Raad 2022 [22]; Afirma GEC: Angell 2019 [70], San Martin 2020 [69], Parajuli 2019 [74], Onken 2021 [26], Lastra 2014 [93], Zhang 2021 [64], Harrell 2019 [72], Wei 2019 [76], Wu 2016 [88], Livhits 2018 [78], Arosemena 2020 [65], Polavarapu 2021 [63], Celik 2015 [90], Chaudhary 2016 [83], Glass 2021 [62].

3.5. Diagnostic Accuracy

Among the resected nodules (scenario A), the prevalence of malignancy was 40% in GSC datasets, 30% in ThyroSeq v3 datasets and 20% in GEC datasets (Figure 3; Table 3). GSC (4 datasets; 161 nodules) had a pooled sensitivity of 87.5% (76.9–93.6) and a specificity of 46.5% (36.2–57.1), with a PPV of 49.5% (24.6–74.7) and an NPV of 89.6% (58.9–98.1). ThyroSeq v3 (9 datasets; 246 nodules) had a sensitivity of 93.9% (82.9–98.0) and a specificity of 29.9% (23.4–37.5), with a PPV of 35.3% (27.6–43.8) and an NPV of 91.4% (78.1–96.9). No GEC study reported a false-negative result, so the sensitivity of GEC was 100% (87.2–100), but its specificity was only 10.1% (5.7–17.3) and its PPV 20.2% (10.4–35.5) (8 datasets; 136 nodules). Specificity was higher for GSC than for ThyroSeq v3 ($p = 0.010$) and higher for both than for GEC ($p < 0.001$), whereas sensitivity did not differ significantly between GSC and ThyroSeq v3 ($p = 0.26$). Accordingly, the positive likelihood ratio was low for both current tests, at 1.6 for GSC and 1.3 for ThyroSeq v3.

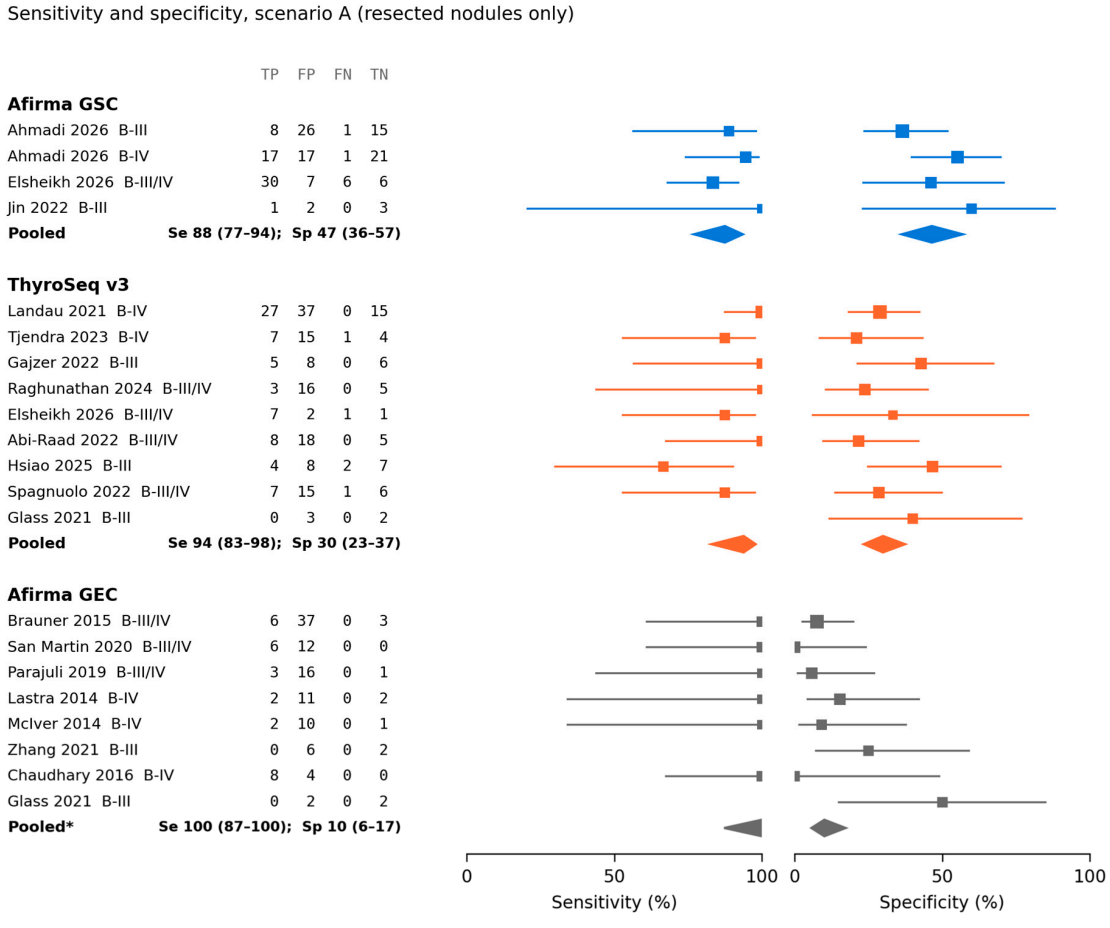

Figure 3. Paired forest plot of sensitivity and specificity in scenario A (nodules with surgical histology only). Pooled estimates for GSC and ThyroSeq v3 from the bivariate binomial–normal model; for GEC, univariate models, because no study reported a false negative (*). No sensitivity estimate is shown for

Glass 2021 [62] (ThyroSeq v3), Zhang 2021 [64] (GEC) or Glass 2021 [62] (GEC), which reported no malignant nodules. Datasets (top to bottom): Afirma GSC: Ahmadi 2026 [35], Ahmadi 2026 [35], Elsheikh 2026 [25], Jin 2022 [56]; ThyroSeq v3: Landau 2021 [23], Tjendra 2023 [54], Gajzer 2022 [55], Raghunathan 2024 [50], Elsheikh 2026 [25], Abi-Raad 2022 [22], Hsiao 2025 [44], Spagnuolo 2022 [57], Glass 2021 [62]; Afirma GEC: Brauner 2015 [9], San Martin 2020 [69], Parajuli 2019 [74], Lastra 2014 [93], McIver 2014 [94], Zhang 2021 [64], Chaudhary 2016 [83], Glass 2021 [62].

When unoperated test-negative nodules were counted as benign (scenario B), specificity rose to 87.4% (83.5–90.4) for GSC (5 datasets from 4 studies; 398 nodules), 76.3% (65.3–84.6) for ThyroSeq v3 (6 datasets; 395 nodules) and 40.7% (33.7–48.1) for GEC (6 datasets; 212 nodules) (Supplementary Figure S3; Figure 4). Sensitivity was 91.2% (76.0–97.1), 94.8% (69.8–99.3) and 100% (90.0–100), respectively, and the NPV reached 99.1% for both GSC and ThyroSeq v3 and 100% for GEC. The PPV, however, remained modest at 44.7% (24.6–66.6) for GSC, 33.3% (23.1–45.3) for ThyroSeq v3 and 21.5% (11.9–35.5) for GEC. The specificity advantage of GSC over ThyroSeq v3 persisted ($p = 0.016$), while the difference in sensitivity again did not reach significance ($p = 0.64$).

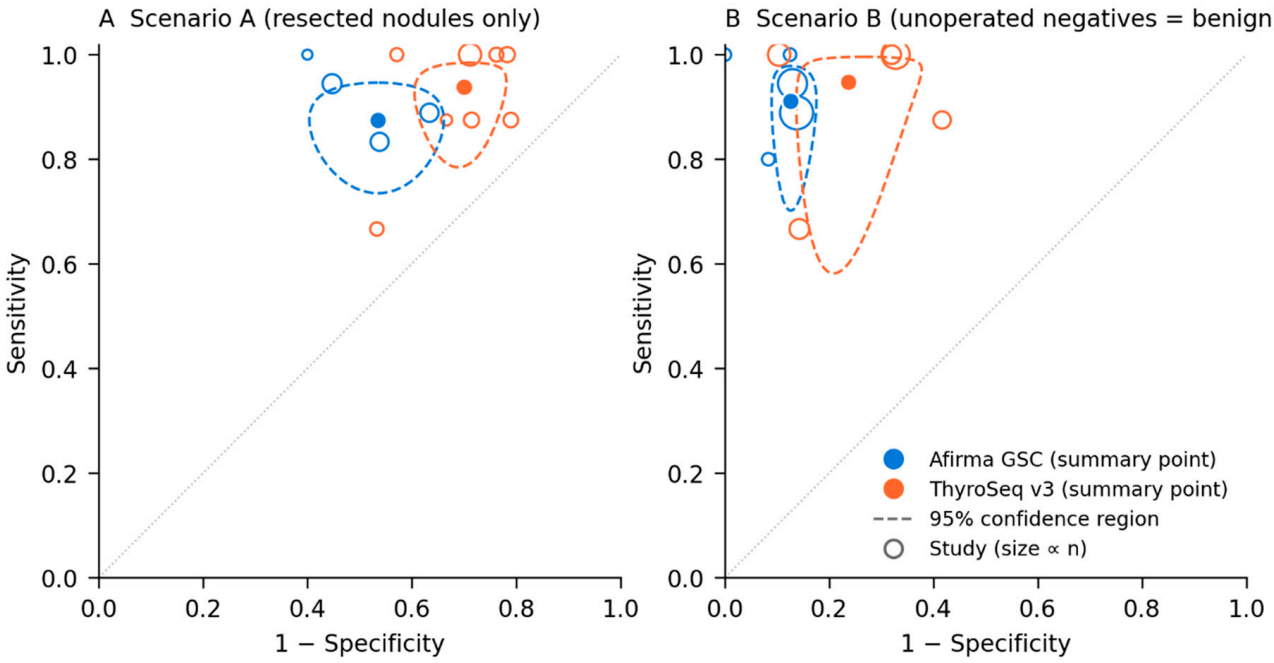

Figure 4. Summary receiver operating characteristic space for Afirma GSC and ThyroSeq v3: (**A**) scenario A, resected nodules only; and (**B**) scenario B, unoperated test-negative nodules counted as benign. Open circles, individual datasets (size proportional to sample size); filled circles, bivariate summary points; dashed lines, 95% confidence regions. SROC curves were not drawn because of the small number of datasets.

Scenario B rests on 563 unoperated test-negative nodules assumed benign (GSC 279, ThyroSeq v3 220, GEC 64; Supplementary Table S11). Only 4 of the 17 contributing datasets specified a minimum follow-up duration (≥6 months: Zhang 2021 [64], Jin 2022 [56]; annual ultrasound in a randomized trial: Raghunathan 2024 [50]; ≥1 year for unoperated benign calls: Ahmadi 2026 [35], median 24 months), and no dataset reported a malignancy diagnosed during observation of a test-negative oncocytic nodule. The scenario B negative predictive values are therefore upper bounds.

3.6. Malignancy by Molecular Alteration Class

Among the resected nodules tested with ThyroSeq or comparable panels, the pooled risk of malignancy increased across alteration classes (Figure 5; Table 2). It was 9.3% (3.9–20.4; 5 of 54) in nodules with a negative result (no alteration detected), 21.6% (9.7–41.5; 15 of 64) in those with CNA only, and 30.2% (17.9–46.3; 53 of 144) in those with *RAS*-like alterations, and, among nodules with *TP53* or *TERT* alterations, 17 of 25 were malignant (crude rate 68%; random-effects estimate 88.5%, 95% CI 18.8–99.6); all six nodules with *BRAF* V600E or a kinase fusion were malignant (100%; 54.1–100). The difference between CNA-only and *RAS*-like nodules was not statistically significant ($p = 0.078$). *BRAF* V600E and kinase fusions define papillary carcinoma, including variants with oncocytic features that can mimic an oncocytic neoplasm on cytology [23].

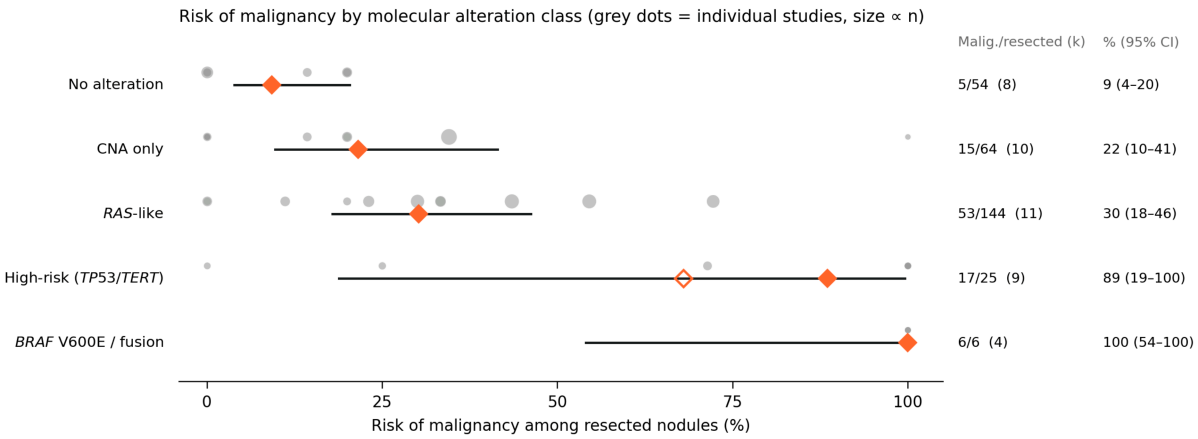

Figure 5. Risk of malignancy among resected oncocytic Bethesda III/IV nodules according to molecular alteration class. Diamonds, pooled estimates (univariate binomial–normal model; for the *BRAF* V600E/fusion class, an exact Clopper–Pearson interval on the pooled count) with 95% CIs; grey dots, individual studies (size proportional to n). For *TP53/TERT* alterations, the filled diamond is the model (typical-study) estimate under large heterogeneity and the open diamond the crude rate, 17 of 25 (68%).

Alteration-level detail is provided in Supplementary Table S12. *KRAS* predominated among *RAS*-like alterations in oncocytic aspirates in several cohorts (6 of 9 *RAS*-positive aspirates in Gajzer 2022 [55]; *KRAS*-positive nodules had oncocytic features in 75% in Radkay 2014 [96]) with low malignancy rates; *EIF1AX* splice-site and non-splice variants were separable in Landau 2021 [23] and Schatz-Siemers 2019 [75]; the divergence between the crude high-risk rate (68%) and the random-effects estimate (88.5%) is explained by two small series whose *TP53/TERT*-positive nodules were mostly benign (Geetha 2025 [43]: three *TERT*-positive nodules, all non-malignant; Raghunathan 2024 [50]: one of four high-risk nodules malignant, the three *TERT*-bearing nodules benign, one reclassified as a follicular lesion of uncertain malignant potential after multidisciplinary review).

3.7. Sensitivity Analyses

The ThyroSeq v3 estimates were robust in sensitivity analyses (Supplementary Table S7). Adding the University of Pennsylvania abstract [21] gave a sensitivity of 92.8% and a specificity of 32.2%, excluding the two studies that defined a positive result as any reportable alteration [22,23] gave 86.8% and 31.6%, and excluding studies that enrolled resected nodules only gave 96.8% and 30.1%. By Bethesda category, ThyroSeq v3 specificity was 44.1% (28.6–60.8) in Bethesda III and 26.8% (17.8–38.2) in Bethesda IV, with sensitivities of 81.9% and 97.4%, respectively, although these subgroup estimates had wide CIs.

Excluding the three oncocytic-present datasets accepted under the I1 relaxation changed the pooled GSC benign call rate only from 74.2% to 75.4% (Supplementary Table S7).

For GSC, excluding the study in which NIFTP could not be separated from malignant outcomes [25] gave a sensitivity of 92.9% (75.5–98.2) and a specificity of 46.7% (34.7–59.1), and adding the study of Black and Hispanic patients [19] to scenario B gave a sensitivity of 90.2% and a specificity of 85.7%.

Restricting the *RAS*-like class to studies published after the NIFTP reclassification (excluding Radkay et al. [96]) lowered the pooled risk to 27.2% (15.0–44.1; 41 of 122) and the difference from CNA-only remained non-significant ($p = 0.18$).

Excluding Chaudhary et al. [83] from the GEC scenario A analysis gave a specificity of 10.5% (5.9–17.9). Excluding Endo et al. [20], which counted NIFTP as malignant by design, from the scenario B analyses gave a GSC sensitivity of 90.9% (75.3–97.0) and specificity of 86.8% (82.8–90.0) and a GEC specificity of 40.1% (32.3–48.6), and excluding all four pre-reclassification studies from the GEC scenario A analysis gave a specificity of 11.9% (3.0–37.4).

In the 14 histology-defined datasets in which each patient contributed once (348 nodules), the pooled sensitivity was 85.8% (62.5–95.7) and specificity 49.1% (27.1–71.4) (Supplementary Figure S4); restricted to GSC and ThyroSeq, the corresponding values were 90.2% and 65.6%, and, for GEC, they were 92.9% and 5.2%. Leave-one-out analyses did not materially change any pooled estimate: ThyroSeq v3 sensitivity ranged from 89.4% to 95.5% and specificity from 27.8% to 31.3%, and GSC sensitivity ranged from 87.3% to 92.9% and specificity from 45.7% to 46.7% (omitting Ahmadi 2026 [35] left too few datasets for pooling). In three of the four primary bivariate models, a variance component was estimated at or near zero or the correlation at its boundary; univariate models gave almost identical summary estimates (Supplementary Table S7).

Tipping-point analysis: Under the assumption that 2% and, separately, 4% of the unoperated test-negative nodules counted as benign were in fact missed malignancies, the pooled scenario B sensitivity for Afirma GSC would fall from 91.2% to 77.5% and 66.0%, respectively, and the NPV from 99.1% to 97.2% and 95.0%; for ThyroSeq v3, the sensitivity would fall from 93.9% to 85.2% and 78.0% and the NPV to 96.9% and 95.0%, respectively. Specificity changes little, because assumed-benign nodules dominate the true-negative cell. The 4% benchmark approximates the malignancy risk of benign cytology, against which negative molecular results are benchmarked in this review.

3.8. Small-Study Effects and Certainty of Evidence

By Bethesda category, the benign call rate was 81% (164/203) for Bethesda III and 74% (115/156) for Bethesda IV in Ahmadi 2026 [35], and the pooled ThyroSeq v3 specificity was higher in Bethesda III (44%) than Bethesda IV (27%) datasets, with benign call rates of 75% versus 51% (Supplementary Table S7). The pooled estimates are weighted averages across Bethesda III oncocytic atypia and Bethesda IV oncocytic follicular neoplasms and should be read alongside the stratum-specific values.

Deeks' test indicated funnel plot asymmetry across 18 scenario A datasets (three of the 21 datasets, all without malignant nodules, were omitted; $p = 0.023$ after adjustment for test platform; unadjusted $p = 0.001$; Supplementary Figure S5); the adjusted test is the more appropriate, because pooling platforms with very different specificities produces spurious asymmetry. The asymmetry ran opposite to the direction expected from publication bias: smaller studies, mostly of GEC, showed lower diagnostic odds ratios. At an assumed prevalence of 20% (Table 4), testing 1000 oncocytic Bethesda III/IV nodules with GSC would be expected to detect 175–182 of the 200 malignancies, miss 18–25 and yield 101–428 false-positive results, with each range spanning scenarios A and B. The corresponding figures

were 188–190, 10–12 and 189–560 for ThyroSeq v3, and 200, 0 and 475–719 for GEC. The certainty of evidence was low for sensitivity, downgraded for risk of bias and for imprecision or indirectness, and very low for specificity, downgraded further for risk of bias and for imprecision or indirectness.

Table 4. Summary of findings (GRADE) for molecular testing in oncocytic Bethesda III/IV nodules.

Test	Studies (Nodules), Scenario A/B	Sensitivity, %	Specificity, %	Per 1000 Tested (Malignancy 20%): TP/FN	TN/FP	Certainty: Sensitivity	Certainty: Specificity
Afirma GSC	3 (161)/4 (398)	88–91	47–87	175–182/18–25	372–699/ 101–428	Low ⊕⊕◯◯ (risk of bias −1; imprecision −1: 3 studies, 8 false negatives)	Very low ⊕◯◯◯ (risk of bias −2: partial verification, estimate 47–87% depending on handling of unoperated negatives; imprecision −1)
Thy-roSeq v3	9 (246)/6 (395)	94–95	30–76	188–190/10–12	240–611/ 189–560	Low ⊕⊕◯◯ (risk of bias −1; indirectness −1: 'any alteration' positivity in 2 studies, mixed v2/v3)	Very low ⊕◯◯◯ (risk of bias −2: partial verification, 30–76%; indirectness −1)
Afirma GEC	8 (136)/6 (212)	100	10–41	200/0	81–325/ 475–719	Low ⊕⊕◯◯ (risk of bias −1; imprecision −1: zero false negatives across small studies)	Very low ⊕◯◯◯ (risk of bias −2; imprecision −1)

Ranges span scenario A (resected nodules only) and scenario B (unoperated test-negatives counted as benign); for TN and FP, the true value is expected to lie between them, whereas TP and FN may be overestimated in both scenarios because cancers missed among unoperated test-negative nodules are not counted. A 20% risk of malignancy approximates the pooled prevalence among resected GEC-tested oncocytic nodules. GRADE for diagnostic accuracy: evidence from cross-sectional accuracy studies starts at high certainty and was rated down for risk of bias (partial verification, unclear blinding of pathologists) and for imprecision or indirectness. Deeks' funnel plot asymmetry test (scenario A, k = 18, adjusted for test type): $p = 0.023$. At a 10% pre-test probability, numbers per 1000 would be as follows: GSC TP 88–91, FN 9–12, TN 419–786 and FP 114–481; ThyroSeq v3 TP 94–95, FN 5–6, TN 270–687 and FP 213–630; and GEC TP 100, FN 0, TN 91–366 and FP 534–809; publication bias was not used to rate down because the test has low power with few small studies and is confounded by test type. FN, false negative; FP, false positive; TN, true negative; TP, true positive.

4. Discussion

To our knowledge, this is the first meta-analysis to pool oncocytic-specific diagnostic accuracy for current molecular tests in Bethesda III/IV thyroid nodules. Across the 75 included studies, which contributed to the analysis of resected nodules, four GSC datasets from three studies and nine ThyroSeq v3 datasets, both current platforms retained a high sensitivity in oncocytic nodules (GSC 88–91%, ThyroSeq v3 94–95%), but the specificity among resected nodules was low, at 47% for GSC and 30% for ThyroSeq v3. GSC had a higher BCR than ThyroSeq v3 (74% vs. 60%, although the difference narrowed after excluding a study that defined any reportable alteration as a positive result) and a higher specificity under both verification scenarios in these indirect comparisons, without a significant difference in sensitivity. Within ThyroSeq-positive nodules, the risk of malignancy depended strongly on the alteration class, rising from about one in five for CNA-only results and one in three for *RAS*-like results (pooled model estimates) to 68% (crude) for *TP53/TERT* alterations (random-effects estimate 88.5%). These conclusions rest on weak evidence, however: partial verification and unblinded reference standards pervaded the literature, and the certainty of evidence was low for sensitivity and very low for specificity.

Previous meta-analyses of molecular testing in indeterminate nodules reported pooled specificities of 36–51% for GSC and 42–64% for ThyroSeq v3 [12–14]. None provided an

oncocytic estimate: oncocytic nodules were either pooled with other indeterminate nodules or mentioned only as a case-mix descriptor [12–15]. The only previous pooled oncocytic result is the BCR comparison by Vuong et al. [16], which reported 73.7% for GSC vs. 21.4% for GEC. Our estimates of 74.2% and 24.7% replicate this with a larger evidence base, and we extend it in two ways: we add ThyroSeq v3, and we give accuracy estimates that Vuong et al. could not pool for lack of data. In a 2025 meta-analysis that directly compared Afirma GSC with ThyroSeq v3 by pooling primary studies, no significant difference between platforms was detected in all-comer indeterminate nodules, and no oncocytic-specific head-to-head comparison exists [99].

Our results also suggest that the loss of specificity in oncocytic nodules depends on the platform. For GSC, the specificity in oncocytic nodules (47%) lay within the range reported for indeterminate nodules overall, consistent with the dedicated Hürthle cell and neoplasm classifiers that were introduced to correct the poor oncocytic performance of GEC [10]. For ThyroSeq v3, by contrast, the specificity (30%) was clearly lower than its all-indeterminate estimates. A plausible explanation is biological. ThyroSeq v3 reports copy-number alterations as positive findings, and chromosomal gains and losses are a hallmark of oncocytic neoplasia, benign as well as malignant: more than one-third of oncocytic adenomas carry copy-number changes [4], and near-haploidization and other widespread chromosomal losses are the dominant genomic events in oncocytic carcinoma [6,7]. A test that flags CNA will therefore label many benign oncocytic adenomas as positive, which is consistent with the low malignancy rate we observed among CNA-only nodules (22% pooled, 23% crude). Because the two platforms were evaluated predominantly in different populations, with different baseline malignancy risks, cytological classifications and surgical selection criteria, the between-platform differences in benign call rate and specificity cannot be attributed to intrinsic test performance alone.

The pooled likelihood ratios help to translate these findings to individual patients. Assuming a pre-test risk of malignancy of 20%, slightly below the risks quoted for Bethesda category III and IV nodules and close to the prevalence among resected oncocytic nodules in GEC-tested cohorts [3,5], a negative GSC result would lower the risk to about 6% (scenario A) or 2.5% (scenario B), and a negative ThyroSeq v3 result to about 5% or 2%. These values approach the 4% (range, 2–7%) risk of a benign Bethesda II aspirate [3], so a negative result in an oncocytic nodule can reasonably support surveillance, as current guidelines allow [1,2]. A positive result is much less informative. Among resected nodules, a positive GSC result raises the risk only to about 29%, and a positive ThyroSeq v3 result to about 25%, values similar to the risks quoted on cytology alone by the 2023 Bethesda System for category IV nodules (30% [range, 23–34%] overall; 25–50% for the oncocytic subtype) [3].

The alteration-class analysis refines this picture for ThyroSeq. CNA-only and *RAS*-like results carried pooled malignancy risks of 22% and 30% (crude rates 23% and 37%), so that, for most positive oncocytic nodules, a diagnostic lobectomy, rather than total thyroidectomy, remains the proportionate operation. *TP53* or *TERT* alterations were uncommon but carried a crude malignancy risk of 68% (17 of 25; random-effects estimate 88.5%), in agreement with genomic studies linking *TERT* promoter mutations and widespread chromosomal loss to widely invasive and metastatic oncocytic carcinoma [6,7]. *BRAF* V600E and kinase fusions were rare, and their presence should prompt the consideration of a papillary carcinoma with oncocytic features rather than a true oncocytic neoplasm [5,23], in keeping with the recognition of oncocytic carcinoma as a distinct entity and a third form of differentiated thyroid carcinoma [4,100]. Reporting the alteration class, rather than a binary positive or negative call, therefore matters for surgical planning in this subgroup, although these class-based implications rest on low or very-low certainty evidence. The alteration class

is hypothesis-generating for counselling about risk, but the current evidence—25 resected nodules with two divergent small series—is not sufficient to determine the extent of surgery.

The higher BCR of GSC is clinically relevant: about three of four oncocytic nodules received a benign result, compared with three of five with ThyroSeq v3 and one of four with GEC. BCR should not, however, be equated with avoided surgery. The 2023 European Thyroid Association guideline notes that this argument wrongly assumes every indeterminate nodule would otherwise undergo diagnostic surgery [2]. Moreover, the long-term safety of observing molecularly benign oncocytic nodules rests on the short follow-up in a minority of patients. Surgical-avoidance analyses of molecular testing remain contested [15,101], including in correspondence from manufacturer-affiliated authors [101]. The prospective follow-up of unoperated test-negative oncocytic nodules is needed. Quantified in Supplementary Table S11, only 4 of 17 scenario B datasets specified a minimum follow-up duration, and the 563 unoperated test-negative nodules carried 0 reported interval cancers; the prospective follow-up of test-negative oncocytic nodules remains necessary. The reproducibility statistics reported for model-assisted screening are now complemented by a sampled false-negative estimate of 0 of 400 re-adjudicated excluded records (upper bound 1.2% at title/abstract, 3.8% at full text).

Most included studies resected only a small fraction of test-negative nodules. This classic partial verification inflates sensitivity and deflates specificity when only resected nodules are analyzed [31]. Bethesda 2023 likewise warns that risk estimates from resected series are skewed by selection bias [3]. Instead of choosing one biased estimate, we bracketed specificity with two scenarios (GSC 47% vs. 87%; ThyroSeq v3 30% vs. 76%), between which its true value is expected to lie. Sensitivity differed little between the scenarios by construction, because unoperated test-negative nodules add true negatives but no false negatives, and it is an upper bound in both, as cancers missed among unoperated test-negative nodules are counted in neither. Both scenarios gave the same ranking of the platforms, which strengthens the comparative conclusion even though the absolute specificity remains uncertain. The scenario A and scenario B NPVs (for GSC, 89.6% and 99.1%) are therefore best read together, as lower and upper bounds under differential and partial verification, respectively.

This review has several strengths. The search was extensive and without language limits, and an inclusive screening rule was designed to capture subgroup data that are often reported only in tables. Screening was repeated in a blinded second round with high agreement ($\kappa = 0.89$ at title/abstract and 0.76 at full-text level), overlapping institutional cohorts were not double-counted, and exact binomial–normal models suited to sparse data were used. The risk of bias and certainty of evidence were assessed for every study with QUADAS-2 and GRADE, and the results were reported under explicit assumptions about verification.

Several limitations should also be considered. The evidence comes overwhelmingly from retrospective studies at a few US academic centers, data from Europe and Asia are scarce [52,102], and the commercial tests studied are mainly available in the United States [2]. Each platform-specific accuracy analysis included only four to nine datasets, so the confidence intervals are wide. Embase was not searched, and the platform comparisons were indirect and not fully independent; the planned meta-regression was not performed because too few datasets were available. The definition of an oncocytic nodule varied across studies and Bethesda editions [3], and interobserver variation in this designation is substantial [25]. Pathologists were rarely blinded to molecular results, and NIFTP could not be separated from malignancy in some studies [20,25,26]; excluding them did not alter our conclusions. Funnel plot asymmetry persisted after adjustment for platform and may reflect small-study effects, publication bias or residual heterogeneity. Screening, data extraction, risk-of-bias

assessment and coding were performed with large-language-model assistance; every output was verified by the investigators, but the two screening rounds were not independent human reviews, and readers should weigh this methodological novelty. Finally, the review was not registered in PROSPERO, its protocol was finalized after a pilot screening round (the full protocol is provided in the Supplementary Material), and the authors were not contacted for the missing subgroup data.

For clinicians, a negative GSC or ThyroSeq v3 result in an oncocytic Bethesda III/IV nodule provides useful reassurance and can support surveillance, whereas a positive result should not be taken as evidence of malignancy, because most positive oncocytic nodules are benign. In ThyroSeq-positive nodules, the specific alteration may help guide the extent of surgery, with *TP53/TERT* alterations suggesting a more aggressive approach, although this evidence is of low to very low certainty; GEC, which has been superseded, should no longer be used for this indication. For researchers, the priorities are prospective multicenter cohorts that report oncocytic nodules separately with complete 2 × 2 data, including the follow-up of unoperated test-negative nodules, a blinded histological review using the 2022 WHO criteria, a head-to-head comparison of platforms on the same aspirates, and the validation of classifiers or CNA thresholds specific to oncocytic nodules to improve specificity without loss of sensitivity.

5. Conclusions

In cytologically oncocytic Bethesda III/IV thyroid nodules, Afirma GSC and ThyroSeq v3 retain high sensitivity, and, assuming unoperated test-negative nodules are benign, a negative result lowers the risk of malignancy to a level close to that of benign cytology. The specificity is low, especially for ThyroSeq v3, so most positive oncocytic nodules are benign. GSC classifies more oncocytic nodules as benign and appeared more specific in indirect comparisons. For ThyroSeq, the alteration class, rather than the binary call, may separate intermediate-risk *RAS*-like and CNA-only results from high-risk *TP53/TERT* alterations. The low certainty of this evidence reflects pervasive partial verification. Prospective studies with a complete follow-up of test-negative oncocytic nodules are needed.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/diagnostics16234012/s1, Supplementary Protocol (version 1.0)—full written protocol, including Section S10 (protocol history and deviations). Table S1. Complete characteristics of the 75 included studies (country, design, period, oncocytic subgroup definition, Bethesda category, tests, tested and resected numbers, and use in this review). Table S2. Search strategies for all databases, with dates and hits per database. Table S3. Study-level 2 × 2 data used in each diagnostic accuracy analysis. Table S4. Study-level benign call rate data. Table S5. Study-level malignancy by molecular alteration class (resected nodules). Table S6. QUADAS-2 assessment of the 75 included studies. Table S7. Sensitivity and subgroup analyses (including the tipping-point analysis and Bethesda III/IV stratification). Table S8. Reports excluded at full-text review ($n = 273$) or not retrieved ($n = 47$), with reasons. Table S9. Funding sources and declared conflicts of interest for the 75 included studies. Table S10a. Verification sample of excluded records: design. Table S10b. Verification sample of excluded records: results. Table S11. Scenario B datasets: composition and follow-up of unoperated test-negative nodules. Table S12. Alteration-level detail underlying the alteration-class analysis. Table S13. Case mix of the platform datasets underlying the indirect between-platform comparisons. Figure S1. QUADAS-2 risk-of-bias and applicability-concerns summary. Figure S2. QUADAS-2 traffic-light plot. Figure S3. Paired forest plots of sensitivity and specificity, scenario B. Figure S4. Paired forest plots, histology-defined analysis. Figure S5. Deeks' funnel plot, scenario A. Data Workbook (all study-level data and analyses) and analysis code (Python 3.12.5).

Author Contributions: X.W., W.Z., and J.X. designed the study and developed the search strategy. X.W. and L.S. conducted the literature search and screened the retrieved records. X.W. and C.C. performed data extraction and assessed the risk of bias of included studies. X.W. conducted the

statistical analyses. X.W. wrote the initial draft of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this work, the authors used Claude-Sonnet-5-5 (Anthropic) and GLM-5.3 (Z.ai) to assist with data extraction, statistical programming, figure preparation, and the language editing of the manuscript; the model-assisted screening and risk-of-bias drafting are described in the Methods. All study decisions and all verification of model-assisted output rested with the authors, who take full responsibility for the content of this article.

Conflicts of Interest: The authors declare no conflict of interest..

References

1. Haugen, B.R.; Alexander, E.K.; Bible, K.C.; Doherty, G.M.; Mandel, S.J.; Nikiforov, Y.E.; Pacini, F.; Randolph, G.W.; Sawka, A.M.; Schlumberger, M.; et al. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* **2016**, *26*, 1–133. [CrossRef] [PubMed]
2. Durante, C.; Hegedüs, L.; Czarniecka, A.; Paschke, R.; Russ, G.; Schmitt, F.; Soares, P.; Solymosi, T.; Papini, E. 2023 European Thyroid Association Clinical Practice Guidelines for Thyroid Nodule Management. *Eur. Thyroid J.* **2023**, *12*, e230067. [CrossRef] [PubMed]
3. Ali, S.Z.; Baloch, Z.W.; Cochand-Priollet, B.; Schmitt, F.C.; Vielh, P.; VanderLaan, P.A. The 2023 Bethesda System for Reporting Thyroid Cytopathology. *Thyroid* **2023**, *33*, 1039–1044. [CrossRef] [PubMed]
4. Baloch, Z.W.; Asa, S.L.; Barletta, J.A.; Ghossein, R.A.; Juhlin, C.C.; Jung, C.K.; LiVolsi, V.A.; Papotti, M.G.; Sobrinho-Simões, M.; Tallini, G.; et al. Overview of the 2022 WHO Classification of Thyroid Neoplasms. *Endocr. Pathol.* **2022**, *33*, 27–63. [CrossRef] [PubMed]
5. Auger, M. Hürthle Cells in Fine-Needle Aspirates of the Thyroid: A Review of Their Diagnostic Criteria and Significance. *Cancer Cytopathol.* **2014**, *122*, 241–249. [CrossRef] [PubMed]
6. Ganly, I.; Makarov, V.; Deraje, S.; Dong, Y.; Reznik, E.; Seshan, V.; Nanjangud, G.; Eng, S.; Bose, P.; Kuo, F.; et al. Integrated Genomic Analysis of Hürthle Cell Cancer Reveals Oncogenic Drivers, Recurrent Mitochondrial Mutations, and Unique Chromosomal Landscapes. *Cancer Cell* **2018**, *34*, 256–270.e5. [CrossRef] [PubMed]
7. Gopal, R.K.; Kübler, K.; Calvo, S.E.; Polak, P.; Livitz, D.; Rosebrock, D.; Sadow, P.M.; Campbell, B.; Donovan, S.E.; Amin, S.; et al. Widespread Chromosomal Losses and Mitochondrial DNA Alterations as Genetic Drivers in Hürthle Cell Carcinoma. *Cancer Cell* **2018**, *34*, 242–255.e5. [CrossRef] [PubMed]
8. Alexander, E.K.; Kennedy, G.C.; Baloch, Z.W.; Cibas, E.S.; Chudova, D.; Diggans, J.; Friedman, L.; Kloos, R.T.; LiVolsi, V.A.; Mandel, S.J.; et al. Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology. *N. Engl. J. Med.* **2012**, *367*, 705–715. [CrossRef] [PubMed]
9. Brauner, E.; Holmes, B.J.; Krane, J.F.; Nishino, M.; Zurakowski, D.; Hennessey, J.V.; Faquin, W.C.; Parangi, S. Performance of the Afirma Gene Expression Classifier in Hürthle Cell Thyroid Nodules Differs from Other Indeterminate Thyroid Nodules. *Thyroid* **2015**, *25*, 789–796. [CrossRef] [PubMed]
10. Patel, K.N.; Angell, T.E.; Babiarz, J.; Barth, N.M.; Blevins, T.; Duh, Q.-Y.; Ghossein, R.A.; Harrell, R.M.; Huang, J.; Kennedy, G.C.; et al. Performance of a Genomic Sequencing Classifier for the Preoperative Diagnosis of Cytologically Indeterminate Thyroid Nodules. *JAMA Surg.* **2018**, *153*, 817–824. [CrossRef] [PubMed]
11. Steward, D.L.; Carty, S.E.; Sippel, R.S.; Yang, S.P.; Sosa, J.A.; Sipos, J.A.; Figge, J.J.; Mandel, S.; Haugen, B.R.; Burman, K.D.; et al. Performance of a Multigene Genomic Classifier in Thyroid Nodules with Indeterminate Cytology: A Prospective Blinded Multicenter Study. *JAMA Oncol.* **2019**, *5*, 204–212. [CrossRef] [PubMed]
12. Silaghi, C.A.; Lozovanu, V.; Georgescu, C.E.; Georgescu, R.D.; Susman, S.; Năsui, B.A.; Dobrean, A.; Silaghi, H. Thyroseq v3, Afirma GSC, and microRNA Panels Versus Previous Molecular Tests in the Preoperative Diagnosis of Indeterminate Thyroid Nodules: A Systematic Review and Meta-Analysis. *Front. Endocrinol.* **2021**, *12*, 649522. [CrossRef] [PubMed]

13. Vardarli, I.; Tan, S.; Görges, R.; Krämer, B.K.; Herrmann, K.; Brochhausen, C. Diagnostic Accuracy of Afirma Gene Expression Classifier, Afirma Gene Sequencing Classifier, ThyroSeq v2 and ThyroSeq v3 for Indeterminate (Bethesda III and IV) Thyroid Nodules: A Meta-Analysis. *Endocr. Connect.* **2024**, *13*, e240170. [CrossRef] [PubMed]
14. Liao, X.; Van Den Heede, K.; Lapauw, B.; Huvenne, W.; Ysebaert, D.; Meert, V.; Fang, Y.; Li, J.; Van Slycke, S.; Brusselaers, N. Comparing the Diagnostic Accuracy of Pre-Operative Genetic Testing for Thyroid Cancer on Fine Needle Aspiration Cytology Specimens: A Systematic Review and Meta-Analysis of Diagnostic Accuracy. *Discov. Oncol.* **2025**, *16*, 1797. [CrossRef] [PubMed]
15. Chowdhury, R.; Hier, J.; Payne, K.E.; Abdulhaleem, M.; Dimitstein, O.; Eisenbach, N.; Forest, V.-I.; Payne, R.J. Impact of Molecular Testing on Surgical Decision-Making in Indeterminate Thyroid Nodules: A Systematic Review and Meta-Analysis of Recent Advancements. *Cancers* **2025**, *17*, 1156. [CrossRef] [PubMed]
16. Vuong, H.G.; Nguyen, T.P.X.; Hassell, L.A.; Jung, C.K. Diagnostic Performances of the Afirma Gene Sequencing Classifier in Comparison with the Gene Expression Classifier: A Meta-Analysis. *Cancer Cytopathol.* **2021**, *129*, 182–189. [CrossRef] [PubMed]
17. McInnes, M.D.F.; Moher, D.; Thombs, B.D.; McGrath, T.A.; Bossuyt, P.M.; The PRISMA-DTA Group. Preferred Reporting Items for a Systematic Review and Meta-Analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA* **2018**, *319*, 388–396. [CrossRef] [PubMed]
18. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *Int. J. Surg.* **2021**, *88*, 105906. [CrossRef] [PubMed]
19. Attia, P.J.; Chinta, S.; Celidonio, J.; Armas, J.S.D.; Kalyoussef, E.; Roden, D.F. The Diagnostic Performance of Afirma Genomic Sequencing Classifier in Black and Hispanic Patients with Indeterminate Thyroid Nodules. *Otolaryngol.—Head Neck Surg.* **2026**, *175*, 849–858. [CrossRef] [PubMed]
20. Endo, M.; Nabhan, F.; Porter, K.; Roll, K.; Shirley, L.A.; Azaryan, I.; Tonkovich, D.; Perlick, J.; Ryan, L.E.; Khawaja, R.; et al. Afirma Gene Sequencing Classifier Compared with Gene Expression Classifier in Indeterminate Thyroid Nodules. *Thyroid Off. J. Am. Thyroid Assoc.* **2019**, *29*, 1115–1124. [CrossRef] [PubMed]
21. Khan, N.; Baloch, Z. Thyroid Nodules with Oncocytic Cells: Outcomes and Molecular Insights. *Lab. Investig.* **2026**, *106*, 104847. [CrossRef]
22. Abi-Raad, R.; Prasad, M.L.; Adeniran, A.J.; Cai, G. Copy Number Variations Identified in Thyroid FNA Specimens Are Associated with Hürthle Cell Cytomorphology. *Cancer Cytopathol.* **2022**, *130*, 415–422. [CrossRef] [PubMed]
23. Landau, M.S.; Nikiforov, Y.E.; Ohori, N.P.; Chiosea, S.I. Impact of Molecular Testing on Detecting Mimics of Oncocytic Neoplasms in Thyroid Fine-Needle Aspirates Diagnosed as Follicular Neoplasm of Hürthle Cell (Oncocytic) Type. *Cancer Cytopathol.* **2021**, *129*, 788–797. [CrossRef] [PubMed]
24. Nikiforov, Y.; Seethala, R.; Tallini, G.; Baloch, Z.; Basolo, F.; Thompson, L. Nomenclature Revision for Encapsulated Follicular Variant of Papillary Thyroid Carcinoma: A Paradigm Shift to Reduce Overtreatment of Indolent Tumors. *JAMA Oncol.* **2016**, *2*, 1023–1029. [CrossRef] [PubMed]
25. Elsheikh, T.M.; Doxtader, E.E.; Golusin, M.; Xing, J.; Bena, J.F.; Balikani, L. Interobserver Variation among Cytopathologists in the Evaluation of Oncocytic Cell-Predominant Thyroid FNA Cytology: A Validation Study of 173 Cases with Surgical Follow-up and Correlation with Molecular Testing. *Cancer Cytopathol.* **2026**, *134*, e70143. [CrossRef] [PubMed]
26. Onken, A.M.; VanderLaan, P.A.; Hennessey, J.V.; Hartzband, P.; Nishino, M. Combined Molecular and Histologic End Points Inform Cancer Risk Estimates for Thyroid Nodules Classified as Atypia of Undetermined Significance. *Cancer Cytopathol.* **2021**, *129*, 947–955. [CrossRef] [PubMed]
27. Whiting, P.F.; Rutjes, A.W.S.; Westwood, M.E.; Mallett, S.; Deeks, J.J.; Reitsma, J.B.; Leeflang, M.M.G.; Sterne, J.A.C.; Bossuyt, P.M.M.; QUADAS-2 Group. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann. Intern. Med.* **2011**, *155*, 529–536. [CrossRef] [PubMed]
28. Reitsma, J.; Glas, A.; Rutjes, A.; Scholten, R.; Bossuyt, P.; Zwinderman, A. Bivariate Analysis of Sensitivity and Specificity Produces Informative Summary Measures in Diagnostic Reviews. *J. Clin. Epidemiol.* **2005**, *58*, 982–990. [CrossRef] [PubMed]
29. Chu, H.; Cole, S. Bivariate Meta-Analysis of Sensitivity and Specificity with Sparse Data: A Generalized Linear Mixed Model Approach. *J. Clin. Epidemiol.* **2006**, *59*, 1331–1332. [CrossRef] [PubMed]
30. Deeks, J.; Macaskill, P.; Irwig, L. The Performance of Tests of Publication Bias and Other Sample Size Effects in Systematic Reviews of Diagnostic Test Accuracy Was Assessed. *J. Clin. Epidemiol.* **2005**, *58*, 882–893. [CrossRef] [PubMed]
31. Begg, C.; Greenes, R. Assessment of Diagnostic Tests When Disease Verification Is Subject to Selection Bias. *Biometrics* **1983**, *39*, 207–215. [CrossRef]
32. Schünemann, H.; Oxman, A.; Brozek, J.; Glasziou, P.; Jaeschke, R.; Vist, G. Grading Quality of Evidence and Strength of Recommendations for Diagnostic Tests and Strategies. *BMJ* **2008**, *336*, 1106–1110. [CrossRef] [PubMed]
33. Schünemann, H.; Mustafa, R.; Brozek, J.; Steingart, K.; Leeflang, M.; Murad, M. GRADE Guidelines: 21 Part 1. Study Design, Risk of Bias, and Indirectness in Rating the Certainty across a Body of Evidence for Test Accuracy. *J. Clin. Epidemiol.* **2020**, *122*, 129–141. [CrossRef] [PubMed]

34. Schünemann, H.; Mustafa, R.; Brozek, J.; Steingart, K.; Leeflang, M.; Murad, M. GRADE Guidelines: 21 Part 2. Test Accuracy: Inconsistency, Imprecision, Publication Bias, and Other Domains for Rating the Certainty of Evidence and Presenting It in Evidence Profiles and Summary of Findings Tables. *J. Clin. Epidemiol.* **2020**, *122*, 142–152. [CrossRef] [PubMed]
35. Ahmadi, S.; Kotwal, A.; Bikas, A.; Altshuler, B.F.; Patel, A.; Arora, G.; Hughes, E.G.; Sahadeo, J.; Yeh, M.W.; Livhits, M.J. Utility of Genomic Sequencing Classifier in Managing Cytologically Indeterminate Oncocytic Thyroid Nodules. *J. Clin. Endocrinol. Metab.* **2026**, *111*, e1832–e1838. [CrossRef] [PubMed]
36. Geetha, S.D.; Seth, N.; Chau, K.; Das, K. Cytologic and Surgical Correlation of *TERT*-Mutated Indeterminate Thyroid Nodules: A Case Series. *Int. J. Surg. Pathol.* **2026**, *34*, 1800–1805. [CrossRef] [PubMed]
37. Lazim, A.; Alzumaili, B.; Xu, B.; Instrum, R.; Hintze, J.; Ganly, C.; Sadow, P.M.; Morris, L.G.; Viswanathan, K.; Ghossein, R. Pathologic Diagnosis of Thyroid Nodules with Preoperatively Detected Tumor Protein 53 Mutations: A Tricenter Series of 32 Cases. *Surgery* **2026**, *199*, 110536. [CrossRef] [PubMed]
38. Stack, J.; Siddiqui, M.T.; Lam, H. Cytologic Sub-Categorization of Bethesda Category IV Nodules Has Significant Clinicopathologic Outcomes. *Cytopathol. Off. J. Br. Soc. Clin. Cytol.* **2026**. [CrossRef] [PubMed]
39. Alzumaili, B.A.; Instrum, R.; Alabkaa, A.; Sadow, P.M.; Tuttle, M.R.; Xu, B.; Morris, L.G.T.; Ghossein, R.A. Pathological Diagnosis of Thyroid Nodules with Preoperatively Detected *TERT* Promoter Mutations in the Absence of *BRAF*(V600E): A Bi-Center Series of 52 Cases. *Thyroid Off. J. Am. Thyroid Assoc.* **2025**, *35*, 1145–1152. [CrossRef] [PubMed]
40. Bang, H.; Cho, C.; Kim, M.Y.; Hyeon, J.; Lee, S.E. Subclassifying "Atypia of Undetermined Significance (AUS)" Category in the 2023 Bethesda System for Thyroid Cytopathology: Analyzing K-TIRADS, *BRAF* V600E Mutation, and Risk of Malignancy. *Endocr. Pathol.* **2025**, *36*, 12. [CrossRef] [PubMed]
41. Cocelli, C.J.; Reid, L.; Ren, S. Various Genetic Alterations Observed in Both Benign and Malignant Oncocytic (Hürthle Cell) Neoplasms of the Thyroid. *CytoJournal* **2025**, *22*, 100. [CrossRef] [PubMed]
42. Connelly, C.F.; Smithgall, M.C.; Desai, N.; Cimic, A.; Baskota, S.U. Performance Characteristics of ThyroSeq, ThyGeNEXT/ThyraMIR, and Afirma Molecular Platforms in Evaluation of 1252 Cytologically-Indeterminate Thyroid Nodules. *J. Am. Soc. Cytopathol.* **2025**, *14*, 243–254. [CrossRef] [PubMed]
43. Geetha, S.D.; Karam, P.; Ziemba, Y.; Chau, K.; Savant, D.; Rosca, O.; Khutti, S.; Gimenez, C.; Das, K. Correlation of Cytologic Features with Molecular Testing of Indeterminate Oncocytic (Hürthle Cell) Thyroid Lesions. *Diagn. Cytopathol.* **2025**, *53*, 102–110. [CrossRef] [PubMed]
44. Hsiao, W.-Y.; Saba, N.F.; Lubin, D.; Chen, A.; Shi, Q. Risk Stratification of ThyroSeq Results in Indeterminate Thyroid Lesions: A Single-Institution Experience of Clinicopathologic Correlation with Cytologic Findings. *Cancer Cytopathol.* **2025**, *133*, e22905. [CrossRef] [PubMed]
45. Potonnier, W.; Guillerm, E.; Bigorgne, C.; Ghander, C.; Roy, M.; Coulet, F.; Ansart, F.; Menegaux, F.; Leenhardt, L.; Brocheriou, I.; et al. Performance of the AmpliSeq NGS Panel in Thyroid Nodules with Indeterminate Cytology. *Eur. Thyroid J.* **2025**, *14*, e240160. [CrossRef] [PubMed]
46. Tjendra, Y.; Piñero, P.; Miltos, Y.; Zuo, Y.; Gonzalez, M.; Jorda, M.; Brachtel, E.; Garcia-Buitrago, M.; Gomez-Fernandez, C.; Ruiz-Cordero, R.; et al. AUS e Predominance of Oncocytes: Exploring the Cytologic, Histologic, and Molecular Landscape of These Thyroid Nodules. *Lab. Investig.* **2025**, *105*, 102619. [CrossRef]
47. Ahmadi, S.; Kotwal, A.; Bikas, A.; Xiang, P.; Goldner, W.; Patel, A.; Hughes, E.G.; Longstaff, X.; Yeh, M.W.; Livhits, M.J. Outcomes of Cytologically Indeterminate Thyroid Nodules Managed with Genomic Sequencing Classifier. *J. Clin. Endocrinol. Metab.* **2024**, *109*, e2231–e2239. [CrossRef] [PubMed]
48. Guzmán-Arocho, Y.D.; VanderLaan, P.A.; Nishino, M. Binary Subclassification Scheme (AUS-Nuclear versus AUS-Other) Adequately Risk-Stratifies Thyroid Fine Needle Aspiration Specimens Classified as Atypia of Undetermined Significance. *J. Am. Soc. Cytopathol.* **2024**, *13*, 23–32. [CrossRef] [PubMed]
49. Ohori, N.P.; Cuda, J.M.; Bastacky, S.I.; Yip, L.; Karslioglu-French, E.; Morariu, E.M.; Ullal, J.; Ramonell, K.M.; Carty, S.E.; Nikiforov, Y.E.; et al. Molecular-Derived Risk of Malignancy and the Related Positive Call Rate of Indeterminate Thyroid Cytology Diagnoses as Quality Metrics for Individual Cytopathologists. *Cancer Cytopathol.* **2024**, *132*, 109–118. [CrossRef] [PubMed]
50. Raghunathan, R.; Longstaff, X.R.; Hughes, E.G.; Li, S.J.; Sant, V.R.; Tseng, C.-H.; Rao, J.; Wu, J.X.; Yeh, M.W.; Livhits, M.J. Diagnostic Performance of Molecular Testing in Indeterminate (Bethesda III and IV) Thyroid Nodules with Hürthle Cell Cytology. *Surgery* **2024**, *175*, 221–227. [CrossRef] [PubMed]
51. Yang, S.P.; Nga, M.E.; Bundele, M.M.; Chiosea, S.I.; Tan, S.H.; Lum, J.H.Y.; Parameswaran, R.; Lim, M.Y.; Li, H.; Cheah, W.K.; et al. Performance of a Multigene Genomic Classifier and Clinical Parameters in Predicting Malignancy in a Southeast Asian Cohort of Patients with Cytologically Indeterminate Thyroid Nodules. *Cancer Cytopathol.* **2024**, *132*, 309–319. [CrossRef] [PubMed]
52. de Koster, E.J.; Morreau, H.; Bleumink, G.S.; van Engen-van Grunsven, A.C.H.; de Geus-Oei, L.-F.; Links, T.P.; Wakelkamp, I.M.M.J.; Oyen, W.J.G.; Vriens, D. Molecular Diagnostics and [(18)F]FDG-PET/CT in Indeterminate Thyroid Nodules: Complementing Techniques or Waste of Valuable Resources? *Thyroid Off. J. Am. Thyroid Assoc.* **2024**, *34*, 41–53. [CrossRef] [PubMed]

53. Kim, T.H.; Rodriguez, E.F.; Lim, D.; Moatamed, N.A. "Copy Number Alteration" as the Sole Molecular Finding of a Thyroseq Test Is More Commonly Seen in Hurthle Cell Neoplasms. *Diagn. Cytopathol.* **2023**, *51*, 705–711. [CrossRef] [PubMed]
54. Tjendra, Y.; Kerr, D.A.; Zuo, Y.; Menendez, S.G.; Jorda, M.; Gomez-Fernandez, C.; Velez Torres, J.M. Probability of Malignancy and Molecular Alterations as Determined by ThyroSeq v3 Genomic Classifier in Bethesda Category IV. *Cancer Cytopathol.* **2023**, *131*, 586–595. [CrossRef] [PubMed]
55. Gajzer, D.C.; Tjendra, Y.; Kerr, D.A.; Algashaamy, K.; Zuo, Y.; Menendez, S.G.; Jorda, M.; Garcia-Buitrago, M.; Gomez-Fernandez, C.; Velez Torres, J.M. Probability of Malignancy as Determined by ThyroSeq v3 Genomic Classifier Varies According to the Subtype of Atypia. *Cancer Cytopathol.* **2022**, *130*, 881–890. [CrossRef] [PubMed]
56. Jin, X.; Lew, M.; Pantanowitz, L.; Smola, B.; Jing, X. Performance of Afirma Genomic Sequencing Classifier and Histopathological Outcome Are Associated with Patterns of Atypia in Bethesda Category III Thyroid Nodules. *Cancer Cytopathol.* **2022**, *130*, 891–898. [CrossRef] [PubMed]
57. Spagnuolo, G.M.; Tierney, H.T.; Laver, N.M.V.; Eldeiry, L.S. A Retrospective Study of Clinicopathologic Outcomes of Nodules with Hürthle Cell Cytology and the Thyroid Nodule App (TNAPP) Ultrasound Recommendations. *Endocr. Pract. Off. J. Am. Coll. Endocrinol. Am. Assoc. Clin. Endocrinol.* **2022**, *28*, 593–598. [CrossRef] [PubMed]
58. Yang, Z.; Zhang, T.; Layfield, L.; Esebua, M. Performance of Afirma Gene Sequencing Classifier versus Gene Expression Classifier in Thyroid Nodules with Indeterminate Cytology. *J. Am. Soc. Cytopathol.* **2022**, *11*, 74–78. [CrossRef] [PubMed]
59. Abdelhakam, D.A.; Mojica, R.E.; Huenerberg, K.A.; Nassar, A. Impact of a Genomic Classifier on Indeterminate Thyroid Nodules: An Institutional Experience. *J. Am. Soc. Cytopathol.* **2021**, *10*, 155–163. [CrossRef] [PubMed]
60. Doerfler, W.R.; Nikitski, A.V.; Morariu, E.M.; Ohori, N.P.; Chiosea, S.I.; Landau, M.S.; Nikiforova, M.N.; Nikiforov, Y.E.; Yip, L.; Manroa, P. Molecular Alterations in Hürthle Cell Nodules and Preoperative Cancer Risk. *Endocr.-Relat. Cancer* **2021**, *28*, 301–309. [CrossRef] [PubMed]
61. Gilani, S.M.; Ross, J.A.; Prasad, M.L.; Hammers, L.; Cai, G.; Adeniran, A.J. Molecular Alterations in Hürthle Cell Neoplasms of Thyroid: A Fine Needle Aspiration Cytology Study with Cytology-Histology Correlation. *Cancer Cytopathol.* **2021**, *129*, 363–373. [CrossRef] [PubMed]
62. Glass, R.E.; Levy, J.J.; Motanagh, S.A.; Vaickus, L.J.; Liu, X. Atypia of Undetermined Significance in Thyroid Cytology: Nuclear Atypia and Architectural Atypia Are Associated with Different Molecular Alterations and Risks of Malignancy. *Cancer Cytopathol.* **2021**, *129*, 966–972. [CrossRef] [PubMed]
63. Polavarapu, P.; Fingeret, A.; Yuil-Valdes, A.; Olson, D.; Patel, A.; Shivaswamy, V.; Matthias, T.D.; Goldner, W. Comparison of Afirma GEC and GSC to Nodules Without Molecular Testing in Cytologically Indeterminate Thyroid Nodules. *J. Endocr. Soc.* **2021**, *5*, bvab148. [CrossRef] [PubMed]
64. Zhang, L.; Smola, B.; Lew, M.; Pang, J.; Cantley, R.; Pantanowitz, L.; Heider, A.; Jing, X. Performance of Afirma Genomic Sequencing Classifier vs Gene Expression Classifier in Bethesda Category III Thyroid Nodules: An Institutional Experience. *Diagn. Cytopathol.* **2021**, *49*, 921–927. [CrossRef] [PubMed]
65. Arosemena, M.; Thekkumkattil, A.; Valderrama, M.L.; Kuker, R.; Castillo, R.P.; Sidani, C.; Gonzalez, M.L.; Casula, S.; Kargi, A.Y. American Thyroid Association Sonographic Risk and Afirma Gene Expression Classifier Alone and in Combination for the Diagnosis of Thyroid Nodules with Bethesda Category III Cytology. *Thyroid Off. J. Am. Thyroid Assoc.* **2020**, *30*, 1613–1619. [CrossRef] [PubMed]
66. Lupo, M.A.; Walts, A.E.; Sistrunk, J.W.; Giordano, T.J.; Sadow, P.M.; Massoll, N.; Campbell, R.; Jackson, S.A.; Toney, N.; Narick, C.M.; et al. Multiplatform Molecular Test Performance in Indeterminate Thyroid Nodules. *Diagn. Cytopathol.* **2020**, *48*, 1254–1264. [CrossRef] [PubMed]
67. Ohori, N.P.; Landau, M.S.; Manroa, P.; Schoedel, K.E.; Seethala, R.R. Molecular-Derived Estimation of Risk of Malignancy for Indeterminate Thyroid Cytology Diagnoses. *J. Am. Soc. Cytopathol.* **2020**, *9*, 213–220. [CrossRef] [PubMed]
68. Papoian, V.; Rosen, J.E.; Lee, W.; Wartofsky, L.; Felger, E.A. Differentiated Thyroid Cancer and Hashimoto Thyroiditis: Utility of the Afirma Gene Expression Classifier. *J. Surg. Oncol.* **2020**, *121*, 1053–1057. [CrossRef] [PubMed]
69. San Martin, V.T.; Lawrence, L.; Bena, J.; Madhun, N.Z.; Berber, E.; Elsheikh, T.M.; Nasr, C.E. Real-World Comparison of Afirma GEC and GSC for the Assessment of Cytologically Indeterminate Thyroid Nodules. *J. Clin. Endocrinol. Metab.* **2020**, *105*, E428–E435. [CrossRef] [PubMed]
70. Angell, T.E.; Heller, H.T.; Cibas, E.S.; Barletta, J.A.; Kim, M.I.; Krane, J.F.; Marqusee, E. Independent Comparison of the Afirma Genomic Sequencing Classifier and Gene Expression Classifier for Cytologically Indeterminate Thyroid Nodules. *Thyroid Off. J. Am. Thyroid Assoc.* **2019**, *29*, 650–656. [CrossRef] [PubMed]
71. Hao, Y.; Duh, Q.-Y.; Kloos, R.T.; Babiarz, J.; Harrell, R.M.; Traweek, S.T.; Kim, S.Y.; Fedorowicz, G.; Walsh, P.S.; Sadow, P.M.; et al. Identification of Hürthle Cell Cancers: Solving a Clinical Challenge with Genomic Sequencing and a Trio of Machine Learning Algorithms. *BMC Syst. Biol.* **2019**, *13*, 27. [CrossRef] [PubMed]

72. Harrell, R.M.; Eyerly-Webb, S.A.; Golding, A.C.; Edwards, C.M.; Bimston, D.N. Statistical Comparison of Afirma GSC and Afirma GEC Outcomes in a Community Endocrine Surgical Practice: Early Findings. *Endocr. Pract. Off. J. Am. Coll. Endocrinol. Am. Assoc. Clin. Endocrinol.* **2019**, *25*, 161–164. [CrossRef] [PubMed]
73. Ohori, N.P.; Landau, M.S.; Carty, S.E.; Yip, L.; LeBeau, S.O.; Manroa, P.; Seethala, R.R.; Schoedel, K.E.; Nikiforova, M.N.; Nikiforov, Y.E. Benign Call Rate and Molecular Test Result Distribution of ThyroSeq V3. *Cancer Cytopathol.* **2019**, *127*, 161–168. [CrossRef] [PubMed]
74. Parajuli, S.; Jug, R.; Ahmadi, S.; Jiang, X.S. Hurthle Cell Predominance Impacts Results of Afirma Gene Expression Classifier and ThyroSeq Molecular Panel Performance in Indeterminate Thyroid Nodules. *Diagn. Cytopathol.* **2019**, *47*, 1177–1183. [CrossRef] [PubMed]
75. Schatz-Siemers, N.; Brandler, T.C.; Oweity, T.; Sun, W.; Hernandez, A.; Levine, P. Hürthle Cell Lesions on Thyroid Fine Needle Aspiration Cytology: Molecular and Histologic Correlation. *Diagn. Cytopathol.* **2019**, *47*, 977–985. [CrossRef] [PubMed]
76. Wei, S.; Veloski, C.; Sharda, P.; Ehya, H. Performance of the Afirma Genomic Sequencing Classifier versus Gene Expression Classifier: An Institutional Experience. *Cancer Cytopathol.* **2019**, *127*, 720–724. [CrossRef] [PubMed]
77. Bellevicine, C.; Sgariglia, R.; Migliatico, I.; Vigliar, E.; D'Anna, M.; Nacchio, M.A.; Serra, N.; Malapelle, U.; Bongiovanni, M.; Troncone, G. Different Qualifiers of AUS/FLUS Thyroid FNA Have Distinct *BRAF*, *RAS*, *RET/PTC*, and *PAX8/PPARg* Alterations. *Cancer Cytopathol.* **2018**, *126*, 317–325. [CrossRef] [PubMed]
78. Livhits, M.J.; Kuo, E.J.; Leung, A.M.; Rao, J.; Levin, M.; Douek, M.L.; Beckett, K.R.; Zanocco, K.A.; Cheung, D.S.; Gofnung, Y.A.; et al. Gene Expression Classifier vs Targeted Next-Generation Sequencing in the Management of Indeterminate Thyroid Nodules. *J. Clin. Endocrinol. Metab.* **2018**, *103*, 2261–2268. [CrossRef] [PubMed]
79. Nikiforova, M.N.; Mercurio, S.; Wald, A.I.; Barbi de Moura, M.; Callenberg, K.; Santana-Santos, L.; Gooding, W.E.; Yip, L.; Ferris, R.L.; Nikiforov, Y.E. Analytical Performance of the ThyroSeq v3 Genomic Classifier for Cancer Diagnosis in Thyroid Nodules. *Cancer* **2018**, *124*, 1682–1690. [CrossRef] [PubMed]
80. Walts, A.E.; Sacks, W.L.; Wu, H.H.; Randolph, M.L.; Bose, S. A Retrospective Analysis of the Performance of the RosettaGX(®) Reveal™ Thyroid miRNA and the Afirma Gene Expression Classifiers in a Cohort of Cytologically Indeterminate Thyroid Nodules. *Diagn. Cytopathol.* **2018**, *46*, 901–907. [CrossRef] [PubMed]
81. Hang, J.-F.; Westra, W.H.; Cooper, D.S.; Ali, S.Z. The Impact of Noninvasive Follicular Thyroid Neoplasm with Papillary-like Nuclear Features on the Performance of the Afirma Gene Expression Classifier. *Cancer Cytopathol.* **2017**, *125*, 683–691. [CrossRef] [PubMed]
82. Roychoudhury, S.; Klein, M.; Souza, F.; Gimenez, C.; Laser, A.; Shaheen Cocker, R.; Chau, K.; Das, K. How "Suspicious" Is That Nodule? Review of "Suspicious" Afirma Gene Expression Classifier in High Risk Thyroid Nodules. *Diagn. Cytopathol.* **2017**, *45*, 308–311. [CrossRef] [PubMed]
83. Chaudhary, S.; Hou, Y.; Shen, R.; Hooda, S.; Li, Z. Impact of the Afirma Gene Expression Classifier Result on the Surgical Management of Thyroid Nodules with Category III/IV Cytology and Its Correlation with Surgical Outcome. *Acta Cytol.* **2016**, *60*, 205–210. [CrossRef] [PubMed]
84. Collet, J.-F.; Lacave, R.; Hugonin, S.; Poulot, V.; Tassart, M.; Fajac, A. *BRAF* V600E Detection in Cytological Thyroid Samples: A Key Component of the Decision Tree for Surgical Treatment of Papillary Thyroid Carcinoma. *Head Neck* **2016**, *38*, 1017–1021. [CrossRef] [PubMed]
85. Giovanella, L.; Campenni, A.; Treglia, G.; Verburg, F.A.; Trimboli, P.; Ceriani, L.; Bongiovanni, M. Molecular Imaging with 99mTc-MIBI and Molecular Testing for Mutations in Differentiating Benign from Malignant Follicular Neoplasm: A Prospective Comparison. *Eur. J. Nucl. Med. Mol. Imaging* **2016**, *43*, 1018–1026. [CrossRef] [PubMed]
86. Samulski, T.D.; LiVolsi, V.A.; Wong, L.Q.; Baloch, Z. Usage Trends and Performance Characteristics of a "Gene Expression Classifier" in the Management of Thyroid Nodules: An Institutional Experience. *Diagn. Cytopathol.* **2016**, *44*, 867–873. [CrossRef] [PubMed]
87. Valderrabano, P.; Leon, M.E.; Centeno, B.A.; Otto, K.J.; Khazai, L.; McCaffrey, J.C.; Russell, J.S.; McIver, B. Institutional Prevalence of Malignancy of Indeterminate Thyroid Cytology Is Necessary but Insufficient to Accurately Interpret Molecular Marker Tests. *Eur. J. Endocrinol.* **2016**, *174*, 621–629. [CrossRef] [PubMed]
88. Wu, J.X.; Young, S.; Hung, M.L.; Li, N.; Yang, S.E.; Cheung, D.S.; Yeh, M.W.; Livhits, M.J. Clinical Factors Influencing the Performance of Gene Expression Classifier Testing in Indeterminate Thyroid Nodules. *Thyroid Off. J. Am. Thyroid Assoc.* **2016**, *26*, 916–922. [CrossRef] [PubMed]
89. Yang, S.-E.; Sullivan, P.S.; Zhang, J.; Govind, R.; Levin, M.R.; Rao, J.-Y.; Moatamed, N.A. Has Afirma Gene Expression Classifier Testing Refined the Indeterminate Thyroid Category in Cytology? *Cancer Cytopathol.* **2016**, *124*, 100–109. [CrossRef] [PubMed]
90. Celik, B.; Whetsell, C.R.; Nassar, A. Afirma GEC and Thyroid Lesions: An Institutional Experience. *Diagn. Cytopathol.* **2015**, *43*, 966–970. [CrossRef] [PubMed]

91. Eszlinger, M.; Piana, S.; Moll, A.; Bösenberg, E.; Bisagni, A.; Ciarrocchi, A.; Ragazzi, M.; Paschke, R. Molecular Testing of Thyroid Fine-Needle Aspirations Improves Presurgical Diagnosis and Supports the Histologic Identification of Minimally Invasive Follicular Thyroid Carcinomas. *Thyroid* **2015**, *25*, 401–409. [CrossRef] [PubMed]
92. Harrell, R.M.; Bimston, D.N. Surgical Utility of Afirma: Effects of High Cancer Prevalence and Oncocytic Cell Types in Patients with Indeterminate Thyroid Cytology. *Endocr. Pract. Off. J. Am. Coll. Endocrinol. Am. Assoc. Clin. Endocrinol.* **2014**, *20*, 364–369. [CrossRef] [PubMed]
93. Lastra, R.R.; Pramick, M.R.; Crammer, C.J.; LiVolsi, V.A.; Baloch, Z.W. Implications of a Suspicious Afirma Test Result in Thyroid Fine-Needle Aspiration Cytology: An Institutional Experience. *Cancer Cytopathol.* **2014**, *122*, 737–744. [CrossRef] [PubMed]
94. McIver, B.; Castro, M.R.; Morris, J.C.; Bernet, V.; Smallridge, R.; Henry, M.; Kosok, L.; Reddi, H. An Independent Study of a Gene Expression Classifier (Afirma) in the Evaluation of Cytologically Indeterminate Thyroid Nodules. *J. Clin. Endocrinol. Metab.* **2014**, *99*, 4069–4077. [CrossRef] [PubMed]
95. Park, H.J.; Moon, J.H.; Yom, C.K.; Kim, K.H.; Choi, J.Y.; Choi, S.I.; Ahn, S.-H.; Jeong, W.-J.; Lee, W.W.; Park, S.Y. Thyroid "Atypia of Undetermined Significance" with Nuclear Atypia Has High Rates of Malignancy and *BRAF* Mutation. *Cancer Cytopathol.* **2014**, *122*, 512–520. [CrossRef] [PubMed]
96. Radkay, L.A.; Chiosea, S.I.; Seethala, R.R.; Hodak, S.P.; LeBeau, S.O.; Yip, L.; McCoy, K.L.; Carty, S.E.; Schoedel, K.E.; Nikiforova, M.N.; et al. Thyroid Nodules with *KRAS* Mutations Are Different from Nodules with *NRAS* and *HRAS* Mutations with Regard to Cytopathologic and Histopathologic Outcome Characteristics. *Cancer Cytopathol.* **2014**, *122*, 873–882. [CrossRef] [PubMed]
97. Keutgen, X.; Filicori, F.; Crowley, M.; Wang, Y.; Scognamiglio, T.; Hoda, R.; Buitrago, D.; Cooper, D.; Zeiger, M.; Zarnegar, R.; et al. A Panel of Four miRNAs Accurately Differentiates Malignant from Benign Indeterminate Thyroid Lesions on Fine Needle Aspiration. *Clin. Cancer Res.* **2012**, *18*, 2032–2038. [CrossRef] [PubMed]
98. Lappinga, P.J.; Kip, N.S.; Jin, L.; Lloyd, R.V.; Henry, M.R.; Zhang, J.; Nassar, A. HMGA2 Gene Expression Analysis Performed on Cytologic Smears to Distinguish Benign from Malignant Thyroid Nodules. *Cancer Cytopathol.* **2010**, *118*, 287–297. [CrossRef] [PubMed]
99. Dowell, N.; Begum, S.; Muzaffar, J.; Boelaert, K.; Nieto, H. Comparing the Diagnostic Accuracy of Afirma GSC to ThyroSeq V3 in Cytologically Indeterminate Thyroid Nodules. *Eur. Thyroid J.* **2025**, *14*, e250296. [CrossRef] [PubMed]
100. Ringel, M.D.; Sosa, J.A.; Baloch, Z.; Bischoff, L.; Bloom, G.; Brent, G.A.; Brock, P.L.; Chou, R.; Flavell, R.R.; Goldner, W.; et al. 2025 American Thyroid Association Management Guidelines for Adult Patients with Differentiated Thyroid Cancer. *Thyroid* **2025**, *35*, 841–985. [CrossRef] [PubMed]
101. Kaykov, A.; Geurink, C.; Klopper, J. Comment on Chowdhury et al. Impact of Molecular Testing on Surgical Decision-Making in Indeterminate Thyroid Nodules: A Systematic Review and Meta-Analysis of Recent Advancements. *Cancers* **2026**, *18*, 2295. [CrossRef] [PubMed]
102. Lončar, I.; van Velsen, E.F.S.; Massolt, E.T.; van Kemenade, F.J.; van Engen-van Grunsven, A.C.H.; van Hemel, B.M.; van Nederveen, F.H.; Netea-Maier, R.; Links, T.P.; Peeters, R.P.; et al. European Experience with the Afirma Gene Expression Classifier for Indeterminate Thyroid Nodules: A Clinical Utility Study in the Netherlands. *Head Neck* **2023**, *45*, 2227–2236. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.