Diagnostic Utility of Molecular and Imaging Biomarkers in Cytological Indeterminate Thyroid Nodules

Elizabeth J. de Koster,1 Lioe-Fee de Geus-Oei,2 Olaf M. Dekkers,3,4 Ilse van Engen-van Grunsven,5 Jaap Hamming,6 Eleonora P.M. Corssmit,3 Hans Morreau,7 Abbey Schepers,6 Jan Smit,8 Wim J.G. Oyen,1,9 and Dennis Vriens2

1Department of Radiology and Nuclear Medicine, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands; 2Department of Radiology, Section of Nuclear Medicine, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; 3Department of Endocrinology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; 4Department of Epidemiology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; 5Department of Pathology, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands; 6Department of Surgery, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; 7Department of Pathology, Leiden University Medical Center, 2333 ZA Leiden, the Netherlands; 8Department of Endocrinology, Radboud University Medical Center, 6525 GA Nijmegen, the Netherlands; and 9Division of Radiotherapy and Imaging, Institute of Cancer Research, and Department of Nuclear Medicine, Royal Marsden Hospital, London SW3 6JJ, United Kingdom

ABSTRACT Indeterminate thyroid cytology (Bethesda III and IV) corresponds to follicular-patterned benign and malignant lesions, which are particularly difficult to differentiate on cytology alone. As ~25% of these nodules harbor malignancy, diagnostic hemithyroidectomy is still custom. However, advanced preoperative diagnostics are rapidly evolving.

This review provides an overview of additional molecular and imaging diagnostics for indeterminate thyroid nodules in a preoperative clinical setting, including considerations regarding cost-effectiveness, availability, and feasibility of combining techniques. Addressed diagnostics include gene mutation analysis, microRNA, immunocytochemistry, ultrasonography, elastosonography, computed tomography, sestamibi scintigraphy, [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET), and diffusion-weighted magnetic resonance imaging.

The best rule-out tests for malignancy were the Afirma® gene expression classifier and FDG-PET. The most accurate rule-in test was sole BRAF mutation analysis. No diagnostic had both near-perfect sensitivity and specificity, and estimated cost-effectiveness. Molecular techniques are rapidly advancing. However, given the currently available techniques, a multimodality stepwise approach likely offers the most accurate diagnosis, sequentially applying one sensitive rule-out test and one specific rule-in test. Geographical variations in cytology (e.g., Hürthle cell neoplasms) and tumor genetics strongly influence local test performance and clinical utility. Multidisciplinary collaboration and implementation studies can aid the local decision for one or more eligible diagnostics. (Endocrine Reviews 39: 154 – 191, 2018)
current form in 2009, recognizing six diagnostic categories with an incremental risk of malignancy and clinical management guidelines. Although the Bethesda system created a much-used handhold by standardizing the cytological diagnosis and consecutive management of thyroid nodules worldwide, the system does not provide a clear answer for the heterogeneous group of nodules with indeterminate cytology (4, 5). This includes cytology with atypia of undetermined significance or follicular lesion of undetermined significance (AUS/FLUS, Bethesda III), and cytology (suspicious for a) follicular neoplasm (SFN/FN) or (suspicious for a) Hürthle cell neoplasm (SHCN/HCN, Bethesda IV). Similar indeterminate cytological categories are found in the British Thyroid Association Thy system and classification of the Società Italiana di Anatomia Patologica e Citopatologia Diagnostica/International Academy of Pathology, Italian Division (SIAPEC-IAP): Thy3a and Thy3f, and TIR3A and TIR3B, respectively (Table 1) (6, 7).

Alongside a doubled incidence of thyroid carcinoma over the past two decades and a prevalence of thyroid nodules stretching far beyond the 5% for palpable nodules, explained by the incidental detection of nonpalpable nodules and clinically occult thyroid cancers on imaging studies, the need for a more accurate diagnostic procedure has grown (9). This urge was further emphasized when other research groups were unable to reproduce the prevalence of the cytological categories and corresponding malignancy risks proposed by Cibas et al. (4), especially those of the AUS category (4, 10, 11). Insuperable variations in the worldwide patient populations, and intra- and interobserver variation in the assessment of thyroid cytology, were named as likely underlying causes (4, 5, 11, 12). Yet, it raised questions concerning the overall approach of thyroid nodule diagnosis and whether cytology is the best starting ground. Cost-effectiveness is a major benefit of cytological examination, yet a more accurate test may eventually replace cytological examination completely (13, 14). At present, however, a supplemental diagnostic procedure is specifically warranted for cytologically indeterminate thyroid nodules. Diagnostic hemithyroidectomies are still customarily performed to obtain a definite histological diagnosis. With a benign histopathological result in approximately three in four cases, surgery was not only beneficial but also exposed the patient to unnecessary surgical risks. In the case of malignant lesions, a second-stage completion thyroidectomy is often indicated, which is associated with additional costs and higher risks of surgical complications (15–18). An additional preoperative test or combination of tests for thyroid nodules with indeterminate cytology should prevent unfavorable diagnostic hemithyroidectomies for benign nodules, limit the number of two-stage surgeries for thyroid malignancies, or both. With rapidly advancing technology, the possibilities for additional diagnostic techniques seem endless: the applications of existing diagnostics such as ultrasound (US), positron emission tomography (PET)/computed tomography (CT), and immunocytochemistry (ICC) are extended and more clearly demarcated for use in indeterminate thyroid nodules. High-tech molecular tests such as gene mutation panels, gene or microRNA expression profiles, and sequencing techniques are hot-topic (4, 19–23). Every currently known engagement point from the genotype to the phenotype of the tumor is being explored. Combined, the various research fields encompass an extensive range of investigative methods. Individually they usually focus on one or two methods only, making one-to-one comparison of these diagnostics difficult. The 2015 American Thyroid Association (ATA) guidelines suggested several additional tests, but a definitive answer or complete overview of all available tests is still lacking (8).

Alongside higher-level expert discussions and lobbying of med tech companies, clinical endocrinologists
and thyroid surgeons ponder about the best solution for their individual patients. Their choices depend on the characteristics of their patient populations, availability and costs of a certain test, and personal preference. In any case, a useful additional test should be accurate, accessible, affordable, and affect patient management.

This review aims to provide practical considerations for physicians involved in the management of patients with thyroid nodules. It gives an overview of the available literature on additional diagnostic tests for thyroid nodules with indeterminate cytology. We will work our way down from genotype to phenotype, discussing both anatomical and functional techniques, from the state-of-the-art molecular and imaging biomarkers, as well as widely available conventional imaging techniques. The ability of a test to distinguish between malignant and benign nodules in a preoperative setting is discussed, focusing on clinical validation and utility, and including the development phase, cost-effectiveness, and availability of each technique, where appropriate.

Table 1. Overview of Classification Systems for Thyroid Cytology

<table>
<thead>
<tr>
<th>Bethesda System for the Reporting of Thyroid Cytology (4)</th>
<th>British Thyroid Association (BTA) (6)</th>
<th>SIAPEC-IAP (Italy) (7)</th>
<th>Malignancy Rate (4)</th>
<th>Proposed Management [2015 ATA Guidelines (8)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Description</td>
<td>Category Description</td>
<td>Category Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I Nondiagnostic/unsatisfactory</td>
<td>Thy1 Nondiagnostic</td>
<td>TIR1 Nondiagnostic</td>
<td>1%–4%</td>
<td>Repeat FNAC with US guidance</td>
</tr>
<tr>
<td></td>
<td>Thy1c Nondiagnostic cystic lesion</td>
<td>TIR1c Nondiagnostic-cystic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II Benign</td>
<td>Thy2 Nonneoplastic</td>
<td>TIR2 Nonmalignant/benign</td>
<td>0%–3%</td>
<td>No clinical follow-up or treatment required</td>
</tr>
<tr>
<td></td>
<td>Thy2c Nonneoplastic cystic lesion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III Atypia of undetermined significance/follicular lesion of undetermined significance (AUS/FLUS)</td>
<td>Thy3a Atypical features present</td>
<td>TIR3a Low-risk indeterminate lesion</td>
<td>~5%–15%</td>
<td>Repeat FNAC if second Bethesda III result, consider additional tests and/or diagnostic hemithyroidectomy</td>
</tr>
<tr>
<td>IV Follicular neoplasm/suspicious of a follicular neoplasm, including Hurthle cell (oncocytic) type</td>
<td>Thy3f Suspicious of follicular neoplasm</td>
<td>TIR3b High-risk indeterminate lesion</td>
<td>15%–30%</td>
<td>Consider additional tests and/or diagnostic hemithyroidectomy</td>
</tr>
<tr>
<td>V Suspicious of malignancy</td>
<td>Thy4 Suspicious of malignancy</td>
<td>TIR4 Suspicious of malignancy</td>
<td>60%–75%</td>
<td>Thyroid surgery recommended. Consider preoperative additional (molecular) testing to determine extent of surgery</td>
</tr>
<tr>
<td>VI Malignant</td>
<td>Thy5 Malignant</td>
<td>TIR5 Malignant</td>
<td>97%–99%</td>
<td>Thyroid surgery recommended</td>
</tr>
</tbody>
</table>

Molecular Biomarkers

Gene mutation analysis and gene expression

In the last decades, researchers have unraveled important molecular mechanisms behind the thyroid tumorigenesis, and designated a great number of genetic alterations that are related to the various types of thyroid carcinoma. Several of these mutational markers have found their way to the preoperative diagnosis of indeterminate thyroid nodules. The most common markers are the somatic BRAF and RAS point mutations, and RET/PTC rearrangement, all of which involve the mitogen-activated protein kinase (MAPK) signaling pathway (186–188).

In the 2015 ATA guidelines, the potentially strong diagnostic impact of molecular testing is explicitly unfolded, focusing on BRAF testing and the, at that date, two main commercially available tests: the seven-gene mutation panel miRInform™ thyroid (Asuragen, Inc., Austin, TX) and the Afirma™ gene expression classifier (GEC) (Veracyte, Inc., South San Francisco, CA). The ATA recommends considerate application of one of these molecular tests for Bethesda III and IV nodules, provided that the result could change the treatment strategy (8).
In the following chapters, the diagnostic potential of mutation analysis in indeterminate thyroid nodules is discussed, including the tests mentioned in the guidelines, as well as other individual molecular biomarkers and multigene panels addressed in literature.

**BRAF mutation**

B-type RAF kinase (BRAF) is a serine–threonine kinase belonging to the rapidly accelerated fibrosarcoma (RAF) family, and the most potent MAPK pathway activator. Point mutations in the BRAF proto-oncogene occur in various human cancers. The somatic BRAFV600E mutation is the most common activating mutation in many carcinomas, including thyroid carcinoma (186). This missense mutation consists of a thymine-to-adenine substitution at nucleotide 1799 (c.1799T>A), resulting in an amino acid substitution where valine is replaced with glutamate at codon 600 (hence V600E) (189, 190). BRAF has an important function in cell proliferation, differentiation, and apoptosis. Upregulation of BRAF through the BRAFV600E activating mutation is associated with tumorigenesis (190). In differentiated thyroid cancer, the BRAFV600E mutation is exclusive to PTC, occurring in 50% to 80% of these tumors (186, 187, 24, 28–34, 69, 191, 192). The BRAFV600E mutation has been prognostically associated with poor clinicopathological outcomes, such as increased incidence of extrathyroidal invasion, recurrence of disease, and distant metastasis of the tumor (35, 36, 193).

BRAF mutation analysis has been extensively studied as a rule-in test for thyroid carcinoma. The BRAF mutation is superior to other mutations in its oftentimes 100% specificity, a positive mutation could prevent two-stage surgery for an indeterminate thyroid nodule (22, 28–30, 31–34, 35–27, 35–67). Even though the BRAF mutation was found in a majority of PTC in a number of studies, the prevalence of the BRAF mutation in indeterminate cytology ranged from 0% to 48% in individual studies (25, 27, 38, 40, 42, 61). Reported sensitivities were therefore heterogeneous and generally poor, ranging from 0% to 83% (24, 28, 40, 69). Other types of thyroid carcinoma occurring in indeterminate nodules, including FTC, FVPTC, and Hurthle cell carcinoma [oncocytic variant of follicular thyroid carcinoma (FTC-OV)], were respectively never or infrequently BRAF mutation-positive (30, 33, 34, 36, 44, 51, 67). Predominated by follicular type carcinoma, the BRAF mutation rarely occurs in Bethesda IV cytology (25, 27, 28, 30, 33, 35, 44, 46, 48, 51, 53, 55, 57, 61, 64–67, 74, 90–92).

Likely contributors to the observed heterogeneity are known global variations in the occurrence rates of PTC and BRAF mutations. In South Korea, where iodine consumption is high, 90% to 95% of thyroid cancers are PTC. More specifically, the proportion of BRAF-mutated PTC is very high: rates of 80% to more than 90% are reported (24, 40, 90). Consequently, BRAFV600E mutation analysis might have both high specificity and high sensitivity in these populations. Studies with higher sensitivities were more often of South Korean origin and frequently demonstrated sensitivity above 40%, with the prevalence of BRAF mutations reported as high as 30% to 48% (24, 40, 41, 69, 64, 93–95). Conversely, the majority of studies with sensitivity below 10% were conducted in Western countries (United States, Europe, or Canada), with some studies reporting no BRAF mutations at all (22, 25–27, 30, 33, 39, 42, 46, 50–52, 54, 60, 61, 67, 74).

Some South Korean studies based surgical decision-making on the result of the BRAF mutation analysis: surgery was relatively less often performed in BRAF mutation-negative indeterminate nodules (24, 69, 91, 95). Such a surgical management strategy is not oncologically safe for Western countries (e.g., Europa or Northern America), where 80% to 90% of thyroid carcinomas are PTC and reported rates of BRAF-mutated PTC vary from 30% to 40% (24, 40, 90). Moreover, even though the true sensitivity of BRAF mutation analysis is presumably high in South Korea for the mentioned epidemiological reasons, the conservative management of BRAF mutation-negative nodules likely magnified test sensitivity by underestimating the rate of BRAF-negative malignant nodules in these studies. Altogether we estimate that approximately one in five South Korean patients would benefit from BRAF mutation analysis, opposite mere one in 25 patients from other countries.

**BRAF mutation in papillary microcarcinoma**

Papillary microcarcinoma (mPTC) have lower BRAF mutation rates (47, 52, 55, 59, 64, 67, 96). The ATA guidelines are reserved with regard to the recommended clinical management of positive BRAF mutation in mPTC, as its relation to extrathyroidal spread and positive lymph node metastases is not as clear as in larger thyroid carcinoma. Although there are studies that associate mPTC to factors of poorer prognosis, the 2015 guidelines recommend that BRAF-mutated mPTC are treated as low-risk malignancies (8, 31).

**BRAF**<sub>600E</sub>** point mutation**

A less common activating BRAF mutation is BRAF<sup>K601E</sup> (c.1801A>G), which occurs considerably less frequently than the BRAFV600E variant and is associated with FVPTC with high specificity (97). Clinically, the characterization of a small cohort of thyroid malignancies with a BRAF<sup>K601E</sup> mutation showed better outcomes than for BRAFV600E mutated tumors: no extrathyroidal tumor extension, recurrence, lymph node, or distant metastasis were reported in indeterminate BRAF<sup>K601E</sup> positive tumors with a median follow-up of 20 months (range, 4 to 47) (98).
Table 2. Overview of Test Performance and Utility of Main Additional Diagnostics in Indeterminate Thyroid Nodules

<table>
<thead>
<tr>
<th>Molecular biomarkers</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Main Advantages</th>
<th>Main Limitations</th>
<th>Cost-Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene mutation analysis and gene expression</td>
<td>0%–83% (24–27)</td>
<td>99%–100% (22, 25–67)</td>
<td>Perfect specificity at low cost</td>
<td>Strong geographical variation in occurrence, clinical utility likely limited to gene mutation panels in countries other than South Korea</td>
<td>Presumed, though unpublished; $7.50 to $123 per test (47, 55, 63, 68)</td>
</tr>
</tbody>
</table>

- **BRAF**: 0%–83% (24–27) | 99%–100% (22, 25–67) | Perfect specificity at low cost | Strong geographical variation in occurrence, clinical utility likely limited to gene mutation panels in countries other than South Korea | Presumed, though unpublished; $7.50 to $123 per test (47, 55, 63, 68) |

- **RAS**: 0%–77% (53, 69) | 75%–100% (53, 69, 70) | High prevalence, frequently detected | Often found in follicular adenomas (false-positive); clinical utility limited to gene mutation panels | Unpublished |

- **RET/PTC**: 0%–29% (30, 43) | 73%–100% (28, 30) | Specific for PTC | Low prevalence, clinical utility limited to gene mutation panels | Unpublished |

- **PAX8/PPARγ**: 0%–29% (25, 49, 71) | 96%–100% (25, 33, 61) | No significant advantages | Low prevalence, utility limited to gene mutation panels | Unpublished |

- **7-gene mutation panel**: 18%–69% (23, 26) | 86%–99% (26, 33) | Comparatively inexpensive mutation panel | Specificity often insufficient for surgical decision-making | United States: likely (72); Europe: unlikely (15); $425 to $1700 per test (72, 73) |

- **NGS**: 71%–91% (22, 62) | 89%–93% (62, 74) | Highly accurate, rapidly advancing technology | Limited availability outside the United States; limited clinical validation studies | Unpublished; €230 to $3200 per test (75, 76) |

- **Afi rma® GEC**: 83%–100% (77–80) | 10%–52% (77, 81) | High rule-out capacity | Limited availability outside the United States; limited high-quality clinical validation studies | Unlikely (15, 73, 82–85); $3500 ($1750 to $7000) per test (73, 78, 82) |

- **MicroRNA**: 57%–100% (23, 86, 87) | 58%–100% (87, 88) | Stable expression irrespective of preservation medium (87, 89) | Limited clinical validation, research ongoing | Unpublished |

<table>
<thead>
<tr>
<th>Conventional imaging</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Main Advantages</th>
<th>Main Limitations</th>
<th>Cost-Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultrasound</td>
<td>Dependent on (combination of) feature(s)</td>
<td>Dependent on (combination of) feature(s)</td>
<td>Global availability, low cost</td>
<td>Operator dependency, limited prospective clinical validation; diagnostic accuracy of individual US features insufficient for surgical decision-making</td>
<td>Presumed, though unpublished.</td>
</tr>
</tbody>
</table>

- **Elastosonography**: 47%–97% (20, 173) | 6%–100% (174–176) | Global availability, low cost, easily performed during standard US workup | Operator dependency; limited clinical utility studies; alternative elasticity cut-off possibly more useful | Presumed, though unpublished. |

- **CT**: Unavailable | Unavailable | Unavailable | Not investigated in indeterminate thyroid nodules | NA |

(Continued)
Availability, cost-effectiveness, and limitations of BRAF mutation analysis

Altogether, the consistent specific utilization in a large number of studies supports the use of BRAF mutation analysis in obviating two-stage surgery. The technique is increasingly available in the clinical setting worldwide. A prior meta-analysis of eight studies questioned the cost-effectiveness of BRAF mutation analysis in indeterminate thyroid nodules based on a mere 4.6% mean prevalence of the mutation (99). Cost-effectiveness studies concerning sole BRAF mutation analysis in indeterminate thyroid nodules are lacking. Regardless, cost-effectiveness is generally presumed, as average costs for testing are relatively low and decreasing over time. Depending on the applied molecular technique, reported costs for BRAF mutation analysis ranged between $7.50 and $123 per tested sample (47, 55, 63, 68).

Low sensitivity remains the main limitation of BRAF mutation analysis, irrespective of the type of indeterminate cytology. Proficiency of the test in preoperative patient management depends on the regional occurrence rate of BRAF-mutated PTC; in South Korea, more patients will benefit from BRAF mutation analysis, and the probability and extent of cost-effectiveness are likely to increase (57). In other health care systems, such as in the United Kingdom, cost-effectiveness is likely more constrained. Nonetheless, BRAF testing could still save approximately half the surgical costs in BRAF mutation-positive carcinoma (27, 55). These global variations should be considered before local implementation of sole BRAF mutation analysis.

RAS point mutation

Point mutations in the gene family of retrovirus-associated DNA sequences (RAS) together constitute the second most frequently occurring genetic alteration in thyroid carcinoma. In indeterminate thyroid nodules, they are the most common genetic alteration, due to a strong association of RAS mutations with the follicular-patterned lesions that make up these cytological categories: FA, FTC, FVPTC, and NIFTP (1, 3, 25, 30, 100, 70). Originally, two of the three homologous RAS genes were identified as viral genes of the oncogenic Harvey (HRAS) and Kirsten (KRAS) murine sarcoma virus; the third, NRAS, was first identified in neuroblastoma cells (101, 102). The genes code for guanosine triphosphate-binding RAS proteins, which are involved in intracellular signaling in the MAPK/extracellular signal-regulated kinase (ERK) pathway. Mutation causes overactive RAS signaling and could ultimately induce malignant transition (188).

RAS mutation in thyroid carcinoma has been associated with favorable prognostic factors, such as encapsulation of the tumor and absence of lymph node metastases, but also with factors indicative of an adverse prognosis, such as poor cell differentiation (2). RAS mutations are not specific for carcinoma and found in both malignant and benign lesions (30, 26, 70). According to the 2015 ATA guidelines, Bethesda III or IV nodules with a RAS mutation should be treated similar to the Bethesda V category, as approximately four out of five are malignant (4, 8). HRAS, KRAS, and NRAS mutations are mutually exclusive. They are each associated with slightly different types of cytology and histology, and consequently a different clinical course. In general, point mutations in NRAS codon 61 and HRAS codon 61 are said to occur most frequently (3, 56). KRAS is associated with oncocytic lesions and a lower malignancy rate than other RAS mutations (103).

A RAS point mutation is found in 0% to 38% of the indeterminate nodules (53, 69). Moreover, approximately one-third of all reported malignancies resulting from indeterminate thyroid cytology are RAS mutation positive, frequently FVPTC or FTC (30, 33, 34, 69,
Sporadic cases of RAS mutation-positive FTC-OV and MTC are reported (33, 34). In individual studies, sensitivity and specificity of RAS mutation analysis ranged from 0% to 77% and from 75% to 100%, respectively (53, 69, 70). Test performance was similar for Bethesda III and IV categories, although the mutation occurred more frequently in Bethesda IV nodules (22, 25, 28, 30, 44, 53, 69, 67, 74, 70). Histopathologically benign nodules carrying a RAS mutation are histopathological FA in most cases, but also oncocyctic variant of FA (Hurthle cell adenoma) or hyperplastic nodules (28, 30, 34, 44, 70). There is an ongoing discussion regarding the interpretation of a false-positive RAS mutation. It is presumed that an oncogenic RAS mutation predisposes a FA for progression into follicular carcinoma. A RAS-mutated FA should be considered a premalignant preinvasive follicular neoplasm. These assumptions put false-positives in a different light, as it would justify resection of such lesions through hemithyroidectomy. Consequently, the lesions could also be considered histopathologically benign nodules carrying a RET/PTC rearrangement in a single lesion are also reported (2, 104, 106, 107). RET/PTC rearrangements, especially RET/PTC3, occur more frequently in PTC in children or patients who were exposed to ionizing radiation and are clinically associated with the presence of lymph node metastases (2). Worldwide variations in frequency of RET/PTC rearrangements exist, dependent on demographics and ethnicity. The RET/PTC rearrangement is present in 42% of PTC in Western populations with a predominance of RET/PTC1, and in 37% of PTC in Asian populations with a predominance of RET/PTC3. Without radiation exposure, in female PTC patients, RET/PTC1 is predominant (108). The rearrangements are also found in benign nodules, especially in patients who were exposed to ionizing irradiation (28, 107). Alike RAS mutations, it is assumed to be an activating genetic alteration and it is argued that a histopathologically benign nodule with an RET/PTC rearrangement should be considered a precancerous lesion.

RET/PTC rearrangements are seldom found in indeterminate nodules. In many studies, no RET/PTC translocation was found at all. Most studies investigated RET/PTC in light of a gene mutation panel and paid it no specific attention (22, 25, 26, 28, 30, 33, 34, 39, 43, 44, 49, 54, 56, 61, 67, 74, 106). Only Guerra et al. (106) solely investigated the RET/PTC rearrangement in 101 thyroid nodules of all cytological categories. In this Italian study, RET/PTC rearrangements were found in 18 of the 50 PTC (36%) using reverse transcription polymerase chain reaction (RT-PCR) and Southern-Blot. All these RET/PTC-positive carcinomas were Thy4 or Thy5 nodules on cytology. Among the 24 Thy3 nodules, two nodules with an RET/PTC3 rearrangement were histopathologically benign (106).

Noteworthy, Sapio et al. (39) detected two RET mutations during their RET/PTC assessments. In contrast to the RET/PTC translocation, RET point mutations are related to sporadic and familial MTC (39, 109). Surgery confirmed histopathological MTC in the RET-mutated nodules (39).

Even though previous histological studies undeniably associated RET/PTC1 and RET/PTC3 rearrangements to PTC, the low prevalence of the rearrangement in indeterminate cytology is a major downside. Testing exclusively for this genetic alteration
in indeterminate nodules is not advantageous, even if issues regarding the number of tested variants and sensitivity of molecular techniques are overcome. The 2015 ATA guidelines only advise RET/PTC testing in context of a gene mutation panel (8).

**PAX8/PPARγ rearrangement**

The PAX8/PPARγ rearrangement arises from a fusion of the promoter and 5’-coding portion of the thyroid-specific transcription factor PAX8 gene to the gene of the nuclear receptor peroxisome proliferator-activated receptor γ (PPARγ) (110, 111). The role of the product of this translocation, the PAX8/PPARγ fusion protein, is not yet understood, as the DNA binding sites of both original proteins are uniquely preserved in the fusion (111). In the normal thyroid, transcription factor PAX8 is involved in differentiation of thyrocytes and regulation of the expression of thyroid-specific genes encoding thyroperoxidase, thyroglobulin, and the sodium/iodide symporter (112). Nuclear receptor PPARγ has multiple presumed functions, including involvement in the regulation of lipid metabolism, adipogenesis, and insulin sensitivity (111, 113).

The chromosomal translocation PAX8/PPARγ was first discovered in, and traditionally associated with, FTC and FA (110). It is reported in 30% to 45% of FTC and in up to 33% of FA (44, 114–116). However, several studies have also uncovered varying amounts of FVPTC carrying the translocation, with published rates up to 38% (114, 117, 118). It has not been reported in benign or malignant Hurthle cell neoplasms (71, 115).

PAX8/PPARγ is often related to well-differentiated malignancies with a relatively favorable prognosis. Capsular and vascular invasion are reported to a lesser extent in FTCs with a PAX8/PPARγ rearrangement than in RAS-mutated tumors (115). Widely invasive features are not reported. PAX8/PPARγ-mutated FVPTC are mostly encapsulated, following an indolent clinical course with minimal disease recurrence despite the presence of some capsular and vascular invasion at presentation (115–117). In contrast to the BRAF, RAS and RET/PTC genetic alterations in thyroid carcinoma, the PAX8/PPARγ rearrangement does not involve the RAS-RAF-MAPK pathway. Nikiforova et al. (115) hypothesized that oncogenesis of follicular-type tumors likely takes place through two different molecular pathways: a RAS-mutation driven and PAX8/PPARγ rearrangement driven pathway (115).

Similar to the RET/PTC rearrangement, the PAX8/PPARγ rearrangement rarely occurred in indeterminate thyroid cytology. Approximately two-thirds of the indeterminate nodules carrying the rearrangement were histopathologically malignant, most often FVPTC or FTC (22, 25, 26, 30, 33, 34, 43, 49, 54, 61, 67, 71, 74). False-positive results corresponded to follicular adenomas (26, 54, 61). Similar to RAS mutations, histopathologically benign PAX8/PPARγ-mutated nodules are likely premalignant lesions, or preinvasive FTC. Eslunger et al. (26) observed a microfollicular morphological growth pattern in two of the PAX8/PPARγ-positive FA, supporting this hypothesis (26).

Still, PAX8/PPARγ rearrangement is a rare rearrangement associated with (encapsulated) follicular tumors. Similar to RET/PTC rearrangements, the PAX8/PPARγ rearrangement should only be assessed in indeterminate thyroid nodules in combination with more frequently occurring genetic alterations in a gene mutation panel.

**Other genetic alterations**

**Human telomerase reverse transcription**

The enzyme human telomerase is involved in the maintenance of the chromosomes’ telomeres, which are essential for cell life and proliferation. The catalytic subunit of telomerase is human telomerase reverse transcription (hTERT). In normal thyroid cells, it is inactive. Inappropriate reactivation is associated with malignancy and inflammatory thyroid disease (119). hTERT promoter mutations were previously observed in both PTC and FTC, sometimes together with a BRAF mutation. The mutation is strongly correlated to mortality in differentiated thyroid carcinoma (120). hTERT gene expression is potentially accurate in the preoperative differentiation of indeterminate nodules, with 57% to 88% sensitivity and 75% to 85% specificity demonstrated in two small clinical series of cytological follicular neoplasms (121, 122).

**Tyrosine receptor kinase**

The tyrosine receptor kinase (TRK) rearrangement arises from a translocation of the NTRK1 gene, which is normally expressed in the central and peripheral nervous system and involved in cell differentiation. The TRK rearrangement is associated with PTC and presumably with an adverse prognosis, although evidence is limited (123). In feasibility studies in indeterminate thyroid cytology, not a single TRK rearrangement has been detected; it is most likely not a useful marker (39, 43, 44).

**HMGA2**

Proteins high mobility group AT-hook (HMGA) 1 and 2 regulate the structure and function of chromatin. Normally only expressed during embryogenesis, the overexpression of HMGA in adult tissues is associated with malignancy (124). Laptinga et al. (125) demonstrated that HMGA2 could be a promising additional biomarker. Using receiver operating characteristic (ROC) curve analysis, a >5.9-fold HMGA2 overexpression had 76% sensitivity and 98% specificity in SFN/FN nodules (125). To date no other studies attempted to validate these results.
**Galectin-3 and CD44v6**

One Croatian study used RT-PCR to investigate the simultaneous expression of galectin-3 and CD44v6, two molecular biomarkers better known for their application in immunohistochemistry (IHC) of their expression products (126). CD44v6 normally functions as the cell-surface receptor for hyaluronic acid. Overexpression is found in various human cancers, including thyroid (126, 127). In indeterminate thyroid nodules, a positive test for either one of the two biomarkers resulted in 100% sensitivity and 60% specificity. It is presumed that similar results for these markers are achieved with the more economical IHC techniques (127, 128).

**Gene mutation panel**

Ongoing research in the past years has demonstrated that assessment of individual oncogenic mutations generally has limited clinical utility in indeterminate thyroid cytology. Combining forces of individual genetic alterations into a gene mutation panel, however, likely improves diagnostic accuracy, especially as mutations are mutually exclusive in most cases. These gene mutation panels typically assess the seven genetic alterations, gene mutations as well as gene fusions, that occur most frequently in differentiated thyroid carcinoma, including BRAFV600E, BRAFV600E, NRAS codon 61, HHRAS codon 61, and KRAS codon 12-13 point mutations and RET/PTC1, RET/PTC3, and PAX8/PPARY gene rearrangements (30, 49). The best-known panel is the commercially available miRInform® thyroid (Asuragen, Inc.), currently rebranded as the ThyGenX® Thyroid Oncogene Panel (Interpace Diagnostics, Parsippany, NJ). The miRInform® thyroid tests 17 specific genetic alterations in these seven genes (25). It is marketed as a rule-in test for thyroid malignancy.

The first large clinical utility study to investigate the miRInform® thyroid test was published in 2011. Nikiforov et al. (30) prospectively included 1056 FNAC samples, 92% of which had sufficient epithelial cells and nucleic acids to pursue molecular testing. Residual FNAC material was used for mutation analysis; no additional aspirates were required. Unfortunately, surgery was performed for only 461 of 900 (51%) indeterminate thyroid nodules, independent of the test outcome; these operated cases were included in their final analysis. It is not reported whether non-surgically managed nodules were mutation-positive or -negative. Sensitivity and specificity were 63% and 99% in the 247 Bethesda III nodules, and 57% and 97% in the 214 Bethesda IV nodules, respectively. The authors suggested that the high positive predictive value (PPV) of the miRInform® thyroid in these indeterminate thyroid nodules (88% and 87%, respectively) warrants a direct total thyroidectomy instead of two-step surgery in patients with a positive test (30).

None of the subsequent studies matched the initially reported excellent specificity. The next industry-sponsored prospective study by Beaudenon-Huibregtse et al. (25) reported 47% sensitivity and 88% specificity in 80 Bethesda III and IV nodules. Surprisingly, not a single BRAF mutation was detected (25). Valderrabano et al. (67) reported not a single mutation in 47 included Bethesda III nodules. Moreover, only 1 of 18 nodules with Hürthle cell cytology in this study tested positive, suggesting that Hurthle cell nodules may carry different mutations than the ones investigated by the miRInform® thyroid (67). Ohori et al. (33) demonstrated that genetic alterations less frequently occurred in the textbook colloid-poor Bethesda IV cytology compared with the less common colloid-rich variant. Differences in etiology are unknown, but the authors hypothesized that the two types have subtle histopathological differences. The colloid-rich thyroid carcinomas likely more often develop through the well-known mutations included in the miRInform® thyroid test, whereas mutations that elicit colloid-poor thyroid carcinoma are yet unknown (33).

Simultaneously with the American miRInform® studies, five European studies independently investigated whether a panel of the same seven genes could reliably be assessed using different methods (26, 34, 49, 54, 61). In three separate studies, Eszlinger et al. (26, 34, 49) demonstrated that testing was also feasible on routine air-dried FNAC samples from indeterminate thyroid nodules. Over the course of these studies, sensitivity of this method improved from 18% to 49% and specificity from 86% to 93%, respectively. The use of air-dried FNAC samples for mutation analysis could advance the implementation of mutation analysis in daily practice, as specific storage conditions of fresh FNAC samples for mutation analysis are no longer required (26, 34, 49). Mancini et al. (49) showed that high-resolution melting (HRM) analysis is an accurate screening method for the seven genetic alterations, with 56% sensitivity and 90% specificity. HRM is a post-PCR procedure that does not require substantial additional resources. This could reserve the costlier direct sequencing procedures solely for samples with abnormal HRM results, thereby reducing the overall costs of mutation analysis (49).

Overall, reported sensitivities and specificities of a seven-gene mutation panel in indeterminate thyroid nodules ranges from 18% to 69% and 86% to 99%, respectively (23, 26, 33). It is an adequate diagnostic tool with a high rule-in capacity in indeterminate nodules. Test performance was similar in Bethesda III and Bethesda IV nodules, although the latter more frequently had a positive test result based on the higher prevalence of RAS mutations (25, 30). Due to the common RAS mutations, PPV of the seven-gene mutation panel never exceeds 90% in a range of
To improve the sensitivity of the miRInform® thyroid test, the existing seven-gene mutation panel was expanded to include additional gene mutations, fusions, and translocations, and a microRNA gene expression panel. In addition, it adopted promising next generation sequencing (NGS) techniques. NGS enables the simultaneous targeted testing for multiple mutations in large gene panels and is faster, more sensitive, and more cost-effective than traditional Sanger sequencing. This detection of germline or clinically significant low-level somatic mutations in benign nodules could decrease NGS specificity (23, 74). Nikiforov et al. (74) suggested that the next improvement of the NGS-related tests should therefore be to determine accurate threshold levels for the various gene variations (74).

NGS encompasses crucial technology that is rapidly advancing. The ThyroSeq® v3 was recently announced, promoting to encompass no less than ~95% of genetic alterations occurring in PTC. Extraordinary diagnostic accuracy above 90% is anticipated, including high accuracy in Hurthle cell lesions. Results of the prospective studies validating this new version will likely be published shortly. Nonetheless, NGS techniques currently have limited global availability, with the exception of some European countries and the United States. The ThyroSeq® v3 is available for $3200 per test (75). In contrast, the thyroid nonspecific AmpliSeq™ panel can be ordered online for only realistic 15% to 40% prevalence of malignancy. As such it is debatable whether a positive test warrants immediate single-stage total thyroidectomy. It translates into an inappropriate overtreatment in a substantial number of patients with a positive test but benign final histology at higher risk of surgical complications and all requiring lifelong levothyroxine supplementation. Deliberate surgical decision-making should consider the underlying positive mutation rather than merely the positive test itself.

The limited size of the seven-gene mutation panel keeps the costs per test low compared with other, larger molecular panels. Reported prices of the seven-gene mutation panel all concern the commercial miRInform® thyroid (or ThyGenX® Thyroid Oncogene Panel) and range between $425 and $1700 (72, 73). Implementation of miRInform® testing for indeterminate nodules theoretically resulted in a 20% cost reduction in the United States: the prevented two-step surgical procedures would outweigh the added expenses for miRInform® testing and increased number of total thyroidectomies, including those for nodules with a false-positive test (72). In a European setting, treatment and hospitalization costs are generally lower and miRInform® would most likely not be cost-effective (15). However, these cost-effectiveness studies both adopted the unequalled test performance from the initial key publication, true cost-effectiveness may be less optimistic (15, 30, 72, 73).

Next generation sequencing

To improve the sensitivity of the miRInform® thyroid test, the existing seven-gene mutation panel was expanded to include additional gene mutations, fusions, and translocations, and a microRNA gene expression panel. In addition, it adopted promising next generation sequencing (NGS) techniques. NGS enables the simultaneous targeted testing for multiple mutations in large gene panels and is faster, more sensitive, and more cost-effective than traditional Sanger sequencing and other PCR-based methods (62, 74, 129). As NGS only requires a very small amount (5 to 10 ng) of nucleic acids, remainder material from regular FNAC passes sufficient and no additional aspirates are required (74, 129). The first thyroid-specific NGS-based gene panel was the ThyroSeq® v1 (CBLpath, Ocala, FL), presented in 2013. It detected gene variations in 110 of 145 investigated thyroid cancer tissue samples and 5 of 83 benign specimens. Unfortunately, indeterminate FNAC samples were not analyzed separately in this study. Nonetheless, Nikiforova et al. (129) demonstrated that NGS had a very high success rate and could be a promising molecular technique for thyroid FNAC samples.

Following the ThyroSeq® v1, the road was paved for further exploration of NGS-based diagnostics. Soon, the ThyroSeq® v2 (CBLPath, Ocala, FL) was developed, with a number of primers for TERT promoter variants added to its panel. It simultaneously tested for point mutations in 13 genes and for 42 types of gene fusion products (74). The ThyroSeq® v2 was tested on 143 Bethesda IV thyroid nodules. Forty-two genetic alterations were found, most frequently NRAS.

Diagnostic accuracy of the ThyroSeq® v2 was 92%, with astonishing 90% sensitivity and 93% specificity (74).

More recently, Nikiforov et al. (22) tested the ThyroSeq® v2.1, including point mutations in 14 genes and 42 gene fusion transcripts, in 462 Bethesda III nodules. Based on the promising results of the previous study, surgery was withheld for 362 of 431 ThyroSeq®-negative patients. In the 95 patients with available histopathology, the ThyroSeq® v2.1 demonstrated 91% sensitivity and 92% specificity. Additionally, diagnostic accuracy was estimated for malignancy rates varying between 6% and 48%. PPV would range from 42% to 91%, negative predictive value (NPV) from 92% to 99%. Within reasonable limits, the ThyroSeq® v2.1 is highly reliable to rule out malignancy (22).

Le Mercier et al. (62) retrospectively tested a different commercially available 50-gene NGS panel, the AmpliSeq™ Cancer Hotspot Panel v2 (ThermoFisher, San Diego, CA), which is a tumor-nonspecific NGS panel for detection of somatic tumor variants. This panel does not include thyroid-specific RET/PTC, PAX8/PPARY, and NTRK1 rearrangements. Albeit the study only assessed 34 FNAC samples, with 71% sensitivity and 89% specificity in indeterminate thyroid nodules, the AmpliSeq™ panel seems less accurate than the ThyroSeq® (62).

The high diagnostic accuracy is also a downside to NGS. Highly sensitive, NGS is able to identify mutant alleles at very low levels (<1%). A low percentage of mutant alleles might reflect a subclone within the nodule, which is not histopathologically identified as carcinoma. This detection of germline or clinically insignificant low-level somatic mutations in benign nodules could decrease NGS specificity (23, 74). Nikiforov et al. (74) suggested that the next improvement of the NGS-related tests should therefore be to determine accurate threshold levels for the various gene variations (74).

"Deliberate surgical decision-making should consider the underlying positive mutation rather than merely the positive test itself"
Independent prospective studies are needed to validate its performance and predicted cost-utility in different patient populations, and confirm the superior position of the ThyrSeq® and other NGS techniques.

Afirma® GEC

In molecular diagnostics, the chief competitor of the seven-gene mutation panel is the commercial Afirma® GEC (Veracyte, Inc.). The GEC uses quantification of the messenger RNA (mRNA)-expression of 167 genes and a proprietary classification algorithm to determine the probabilities of malignancy in the samples’ expression patterns. The classification algorithm to discern a “benign” (negative test) from a “suspicious” (positive test) thyroid nodule results from a successful designer study that trained the GEC in both a tissue set (positive test) thyroid nodule results from a successful prospective, blinded, industry-sponsored clinical study to validate this Afirma® GEC in patients with indeterminate thyroid nodules. From 49 hospitals 577 Bethesda III, IV, and V FNAC samples were collected, obtained by two additional needle aspirates from thyroid nodules with a diameter of at least 1 cm. After exclusion of over half (312/577, 54%) of the samples for reasons such as nodules that were not surgically resected, duplicate specimens from the same nodule, and issues with specimen shipments to Veracyte, finally 265 FNAC samples were included in the analysis. Sensitivity of the Afirma® GEC was 90% in the 129 Bethesda III as well as the 81 Bethesda IV nodules with a useful GEC-negative test result in 38% (100/265), but specificity was merely 53% and 49%, respectively (52% on average). Despite the relatively high malignancy rate in Bethesda III nodules and the high number of exclusions, this study is well conducted and recognized worldwide as the landmark study that demonstrated the strength of the Afirma® GEC (81). After the overwhelming results from this key publication, popularity of the GEC took flight. It is marketed as a highly accurate rule-out test for malignancy in thyroid nodules with indeterminate cytology. In 2014, the first multicenter study that retrospectively assessed the clinical utility of the Afirma® GEC was published. Only 6% of reported GEC-negative Bethesda III, IV, and V nodules eventually underwent surgery, of which one resulted in a 6-mm mPCT. Unfortunately, data on GEC negative nodules were only reported on an aggregate level; exact test performance rates in Bethesda III and IV nodules cannot accurately be extracted from the publication. Less than half of the GEC-negative nodules without surgery (71/163, 44%) had clinical or radiological follow-up, ranging from 1 to 24 months (median 8 months), a limited duration compared with the natural, indolent course of differentiated thyroid carcinoma. The published paper does not describe whether the remaining 92 patients with GEC-negative nodules received any follow-up at all. Despite evident limitations to the applied reference standards, Alexander et al. (131) concluded that their results confirm both the accurate test performance from their prior study, as well as the large impact that the Afirma® GEC has on clinical decision-making for cytologically indeterminate thyroid nodules.

Yet, physicians indeed seemed reassured by a negative GEC result based on the first studies alone (81, 132). In many institutions in the United States, the Afirma® GEC was immediately implemented in clinical practice. The retrospective studies that followed were mere postimplementation utility studies, and generally reported very high but moderately consistent sensitivities. GEC-negative nodules were largely managed without surgery and considered true-negative, resulting in possible overestimation of test sensitivity. Long-term follow-up is not yet available to endorse a benign diagnosis in these cases (77, 133). The high degree of missing histology was recognized by most of these studies as a major limitation (78, 79, 131, 133–135). This was confirmed by the 2015 ATA guidelines: recognizing the Afirma® GEC as a promising diagnostic tool, the guidelines stress that it is a major shortcoming that external clinical validation studies with full pathological follow-up of Afirma® GEC-negative nodules are still lacking (8).

Not all studies were able to confirm the potential of the Afirma® GEC. Some struggled with a low benign call rate (i.e., useful negative test result that could lead to management change) (77, 136). McIver et al. (77) questioned the cost-effectiveness of the Afirma® GEC in their population, as the mere 22% (16/72) negative test rate was much lower than anticipated. Moreover, one-quarter of these GEC-negative patients rejected the proposed conservative treatment of US-based follow-up and underwent surgery anyway; one of them was diagnosed with a 3.2-cm FTC with focal capsular and vascular invasion. Also, 84% of GEC-positive nodules proved histopathologically benign, overall resulting in a disappointing 83% sensitivity and 10% specificity (77).

Besides concerns regarding adequate clinical validation of test performance, the postimplementation influence of the GEC on surgical decision-making for individual patients was also questioned. In line with the results of their preliminary study, Noureldine et al. (136) demonstrated that Afirma® GEC testing had not aided surgical decision-making (136, 137). In 93% (206/222) of the included indeterminate nodules, a “benign” or “suspicious” GEC result did not affect management at all: the surgical strategy would have been identical had it been based merely on clinical, cytological, or radiological suspicion. However, if management changes were based on the GEC result, they were more often wrong than right: 11 times GEC-positive results inappropriately tempted physicians into more aggressive surgery, and total thyroidectomy...
was performed instead of the initially recommended lobectomy for nodules that proved histopathologically benign. In contrast, in just four GEC-positive cases the more aggressive surgery was appropriate and the nodule was histopathologically malignant. Also, in just one patient surgery was withheld specifically due to a negative Afirma® GEC result. In the other unresected GEC-negative nodules, surgery was not clinically indicated to begin with; the negative GEC-result merely endorsed conservative management (136). As the GEC was still a new technology when this study was conducted, it is possible that the involved physicians were unsure of the correct interpretation of the GEC results or hesitant to rely on a negative GEC result. However, clinical suspicions and physician and patient preference will always be considered when making surgical decisions.

Yang et al. (78) elegantly tried to solve the shortcoming (histological) follow-up by comparing their findings of GEC performance to a pre-GEC cohort of similar patients from their hospital in all of whom surgery was performed (12, 78). The reported malignancy rates were comparable pre- and post-GEC implementation (18% vs 17%), and obviously relatively more surgeries were performed for benign nodules in the pre-GEC period. Assuming the true malignancy rates in the successively studied populations are indeed similar, the GEC only modestly reduced the number of futile surgeries for benign thyroid nodules from 66% to 52% (78). Altogether, the contribution of the Afirma® GEC to the surgical decision-making may be more limited than expected based on its diagnostic accuracy.

**Availability, cost-effectiveness, and limitations of the GEC**

The Afirma® GEC is currently only available for routine use in the United States. There are high demands for the FNAC specimens regarding sample preservation and shipping. Cytology is revised by Veracyte cytologists and declined if not strictly Bethesda III or IV, with 14% to 17% discordancy between local assessment and central review, comparable to known interobserver rates for thyroid cytology (5, 81, 136). Reported rates of nondiagnostic GEC test results due to insufficient quantity or quality of the mRNA are substantial, varying from 1% to 17% (77, 134). Insufficient mRNA quality was often caused by problems with long duration of the sample shipment to Veracyte (77, 81). Fourth, Afirma® GEC testing is expensive and is currently marketed for $3500 (range $1750 to $7000) per test (73, 78, 82). Testing for medullary carcinoma and BRAF mutation is not included in the Afirma® GEC, but can be performed by Veracyte at additional costs (78). Yet, ancillary BRAF mutation testing may not be relevant, as Kloos et al. (51) found that it improved sensitivity nor specificity of the GEC.

Studies of cost-effectiveness yielded variable results, but most concluded that GEC testing would not be cost-effective over conventional surgical management or other diagnostic modalities in various clinical settings (15, 73, 82–85). The first of these studies proclaimed cost-effectiveness of the GEC even prior to publication of the first validation study by Alexander et al. (81), and has been criticized for several important methodological caveats. This study professedly overestimated test specificity at 75%, overestimated the rate of permanent complications from thyroid surgery, and did not consider the regularly reported GEC test failures (16–18, 77, 82, 83, 138). A recent study determined population-dependent thresholds for feasible cost-effectiveness by comparing GEC performance to conventional surgical management in a local Bethesda III/IV population. GEC-guided management was not cost-effective, adding $1197 to the $11.119 expenses for conventional treatment while hardly improving quality-adjusted life years. Sensitivity analysis showed that the GEC would only become cost-effective if its specificity exceeds 71%, if it costs less than $2640, or if the population malignancy rate decreases from the actual 24% to below 9.2%. This price threshold for cost-effectiveness decreases as the malignancy rate increases, as low as $2023 per test at 35% cancer prevalence (82).

Furthermore, existing interinstitutional differences in test performance have consequences for local applicability and effectiveness (79, 131). Marti et al. (79) compared GEC performance in distinct populations of two large hospitals. The reproducibility of the tests’ sensitivity and specificity was good, but utility strongly depended on the local prevalence of malignancy: as the population malignancy rate increased, a rarer negative GEC became less reliable to rule out malignancy. Oppositely, at low malignancy rates, a negative GEC merely confirmed that the probability of cancer was low. In neither situation, the GEC changed the management strategy. GEC testing was most useful if the malignancy rate ranged between 15% and 21%, comparable to the prevalence reported by Alexander et al. (79, 81).

Finally, the degree of missing histology is a major limitation to the performed studies. None of the studies following the key publication by Alexander et al. (81) had complete histopathologic follow-up: histopathological confirmation ranged between 35% and 82% of specimens.

Missing histology mainly comprised GEC negative nodules, likely resulting in overestimated sensitivity (i.e., missing some malignancies in the many unoperated GEC-negative nodules) and underestimated specificity [i.e., relatively more GEC-positive nodules with benign histology (false-positives) were operated on than GEC-negative nodules with benign histology (true-negatives)]. The trend that studies with higher surgical rates for GEC-negative nodules showed more
A recent meta-analysis by Sathanam et al. (140) included seven studies and reported 96% pooled sensitivity and 31% pooled specificity for the GEC in Bethesda III, IV, and V thyroid nodules with histopathological follow-up. The authors expected that more than 90% of patients with a negative test would be treated conservatively (140). However, in individual studies up to 25% of patients pursued surgery or conservative treatment despite GEC-based recommendation to do the opposite (77, 131). This observation is crucial to cost-utility analyses. In addition, expensive rule-out tests such as the Afirma® GEC should not be performed in case surgery is considered for other reasons, such as cosmetic or mechanical complaints.

**GEC in Hürthle cell cytology**

Brauner et al. (134) specifically validated the Afirma® GEC in 72 cytology samples suspicious for Hürthle cell neoplasm. They demonstrated that GEC testing could accurately have reduced the number of futile surgeries, although through a less profound reduction than in nononcocytic indeterminate thyroid nodules (134). Similar results were noticed in other studies: despite a relatively low risk of malignancy, the majority of Hürthle cell nodules were GEC-positive. Regardless of good sensitivity, this unfavorable benign call rate in Hürthle cell cytology limits diagnostic efficacy in these nodules, increasing the number needed to test and negatively affecting possible cost-effectiveness (78, 80, 81, 135, 136, 139). Diagnostic accuracy of the GEC would likely improve if Bethesda IV cytology suspicious for a Hürthle cell lesion was excluded from GEC testing. Otherwise, similar to the additional testing for medullary carcinoma, adaptations should be made to the Afirma® GEC to improve its clinical utility for Hürthle cell lesions.

In conclusion, it is generally assumed that the Afirma® GEC accurately reclassifies approximately two out of five indeterminate thyroid nodules as benign with published sensitivities ranging between 83% and 100% and similar test performance in Bethesda III and IV nodules. Withholding diagnostic surgery from these patients seems safe (77–80). However, the diagnostic strength and potential cost-utility of Afirma® GEC strongly rely on its NPV, thus on the prevalence of malignancy and benign call rate in the targeted population. There are important concerns regarding the currently insufficient number of clinical validation studies with adequate rates of histopathological confirmation or long-term clinical follow-up. Physicians are strongly advised to locally validate Afirma® GEC test performance before considering test implementation in daily practice. Nonetheless, further large validation studies on the Afirma® GEC may soon become obsolete, as an updated version of the test, the Gene Sequencing Classifier (Veracyte, Inc.), is currently being put into operation. Improved diagnostic accuracy is anticipated, with specific attention to the differentiation of Hürthle cell nodules.

**MicroRNA**

First described in thyroid cytology in 2006, evaluation of the expression levels of microRNA (also called miRNA) is among the newer and more promising approaches to differentiate between benign and malignant thyroid neoplasms (141, 142). MicroRNAs are small endogenous noncoding RNAs of ~22 nucleotides in length. As negative regulators (i.e., silencers) of protein synthesis at a posttranscriptional level, they are involved in many intracellular processes, including cell growth, differentiation, and proliferation. Dysregulation of microRNA expression is found in almost all types of human cancers (143). It reflects the deregulated expression of oncogenes and tumor suppressor genes (142, 88, 89, 144). MicroRNA overexpression is present before morphological tissue changes are seen and therefore considered to be a part of premalignant changes in carcinogenesis (141). MicroRNA expression profiles are tissue-specific and cannot only identify the tissue of origin, but also the histopathological subtype of the cancer and whether it concerns the primary tumor or a metastasis (89, 145).

MicroRNA expression profiles are similar among the various types of thyroid carcinoma, even though expression levels are often distinctively different (89). In histopathological studies, PTC was associated with an up to 11- to 19-fold upregulation of miR-146b, miR-221, miR-222, miR-181b, miR-187, and a downregulation of miR-1 and miR-138 compared with healthy thyroid tissue and benign nodules. Upregulation of miR-221, miR-222, and miR-187 was also found in FTC, FTC-OV, poorly differentiated and anaplastic carcinoma (89, 141, 142, 146, 147). Overexpression of miR-146b-3p, miR-146b-5p, and miR-375 was seen in both PTC and FVPTC (146, 148). Furthermore, expression levels of miR-221 and miR-222 were reported approximately twice as high in FVPTC as compared with PTC or FTC (146). Only a few microRNAs were differently expressed between follicular neoplasm and FTC (149). FA was associated with the expression of miR-200a, whereas high expression of miR-31 was found in Hürthle cell adenoma (89). FTC is related to the differential expression of miR-146b, miR-7-5p, miR-346, miR-197, and miR-21, but results among studies are more heterogeneous (89, 149, 150). FTC-OV showed an expression pattern slightly similar to FTC, but also distinct overexpression of other microRNAs, such as miR-339, miR-183, miR-197, and miR-885-5p (89, 147).

Accordingly, a diagnostic panel of a carefully selected combination of microRNAs and appropriate
expression levels could aid in the preoperative distinction of indeterminate thyroid cytology (151). Recent meta-analyses struggled to reconcile the studies on microRNA in FNAC, as the investigated set of microRNAs was never identical and individual microRNA performance was infrequently described. In unselected cytology, estimated sensitivity of microRNA expression analysis ranged from 75% to 78% regardless of the investigated set; estimated specificity from 73% to 81% (150–152).

In indeterminate thyroid cytology, different sets of microRNAs were evaluated; only several individual microRNAs were analyzed in more than one study. The selected microRNAs were first assessed in a test set of cytological and/or histopathological specimens and a cut-off for their expression level was determined. Subsequently, the significantly up- or downregulated microRNAs were validated in an independent set of (indeterminate) thyroid FNAC samples. Some studies developed a decision model for the validation step (86, 88, 148).

The most promising results were presented by Keutgen et al. (86). Of the six microRNAs investigated in their test set, miR-21, miR-146b, miR-181a, and miR-222 were differentially expressed in malignant nodules with prior indeterminate cytology. The subsequently developed support vector machine model incorporated miR-21, miR-222, and the insignificantly expressed miR-197 and miR-328. Prospective validation in an independent set of 72 indeterminate FNAC samples resulted in 100% sensitivity and 86% specificity. Five of the seven false-positives had Hurthle cell cytology; excluding these, raised specificity to 95% (86). Notably, even though overexpression of miR-146b is often related to thyroid carcinoma, it proved not useful to Keutgen et al. (86) to include in their prediction model. In contrast, Agretti et al. (88) and Shen et al. (148) included miR-146b as the key differentiators in their models. Agretti et al. (88) assessed a frequently quoted set of microRNAs consisting of miR-146b, miR-155, miR-187, miR-197, miR-221, miR-222, and miR-224 (88, 89). Published in 2008, Nikiforova et al. (89) had demonstrated that this seven-microRNA set in FNAC samples had 100% sensitivity and 94% specificity if one of the included microRNAs showed an at least twofold overexpression. Analytic validation of this model by Agretti et al. (88) showed differential upregulation in PTC of all of these microRNAs except miR-197. In particular, miR-146b showed a >30-fold higher expression in PTC. A decision tree including miR-146b, miR-155, and miR-221 was 98% accurate in the test set, but validation in an independent set of indeterminate FNAC samples was unsuccessful, yielding mere 60% sensitivity and 58% specificity (88).

Vriens et al. (153) used a microRNA array to detect 10 genes that were up- or downregulated by at least fivefold in thyroid malignancies. Four microRNAs (miR-100, miR-125b, miR-138, and miR-768-3p) were significantly downregulated and accurately differentiated between benign and malignant follicular and Hurthle cell neoplasms in the test set. In their validation set of 125 indeterminate FNAC samples, only miR-138 was moderately distinctive with 81% NPV. For Hurthle cell carcinoma, miR-138 and miR-768-3p were both 98% accurate (153).

Finally, in a recent Italian study, only miR-375 accurately differentiated between benign and malignant neoplasms. Subsequently, in TIR3 cytology excluding Hurthle cell lesions, a 12-fold or higher overexpression of miR-375 perfectly distinguished benign from malignant lesions with 100% accuracy. It was also significantly differently expressed between TIR3A and TIR3B categories and correlated with a different malignancy risk (87).

Availability and limitations of microRNA expression analysis

MicroRNA expression analysis has advantages over other techniques. MicroRNAs are more stable than mRNA at maintaining their expression in formalin-fixed paraffin-embedded tissue samples as well as FNAC specimens, irrespective of the preservation method (e.g., archived FNAC slides or nucleic acid preservation solutions) (87, 89). Recently microRNA expression was even successfully measured in serum (154). Moreover, microRNA expression levels measured with generic methods (e.g., quantitative RTPCR) correspond well to their biological effect, as microRNAs affect biological processes without the additional step of protein synthesis (89).

However, general limitations of FNAC also translate to concerns with microRNA analysis: scant cellularity or low levels of malignant cells in FNAC specimens could cause a false-negative microRNA test result (88). Another limitation is the plurality of microRNAs associated with differentiated thyroid carcinoma in histopathological studies, causing vast heterogeneity between the limited number of studies in indeterminate cytology. Validation studies of the same microRNA set are lacking. Simultaneously, new microRNAs are still correlated to thyroid carcinoma. Ongoing research has yet to compose the optimal set of microRNAs. Recently, the first commercial test was marketed as the ThyraMIR™ (Interpace Diagnostics). It evaluates the expression levels of miR-29b-1-5p, miR-31-5p, miR-138-1-3p, miR-139-5p, miR-146b-5p, miR-155, miR-204-5p, miR-222-3p, miR-375, and miR-551b-3p. The ThyraMIR™ demonstrated 57% sensitivity and 92% specificity in 109 Bethesda III and IV FNAC specimens (23). Prospective clinical validation of the ThyraMIR™ could affirm the diagnostic value of microRNA expression profiling in indeterminate thyroid nodules in the preoperative setting.
**Immunocytochemistry**

Tissue characterization through selective staining of expressed proteins, i.e. IHC, is a technique that combines histopathology and biochemistry. Exploiting basic antigen–antibody interactions, IHC is able to visualize the distribution and localization of specific cellular components within the cell and in the proper tissue context. This includes tissue biomarkers specific for infection or malignancy. IHC has been fully incorporated in the histopathological routine and is crucial to morphological and molecular tissue characterization. When ICC, the application of this immunology-based technique in cytology, became available, the possibilities were extended to the preoperative setting, too. Specific immunomarkers have been developed to differentiate between benign and malignant thyroid nodules. The 2015 ATA guidelines acknowledge ICC as a technique under development with limited prospective validation studies in indeterminate cytology (8). In unselected thyroid cytology, the much-used immunomarkers galectin-3, Hector Battifora mesothelial-1 (HBME-1), and cytokeratin 19 (CK-19) demonstrated 85%, 83%, and 80% sensitivity, and 90%, 79%, and 79% specificity, respectively (155).

**Galectin-3**

Galectin-3 is a β-galactosyl-binding protein from the lectin group. It is involved in cell-cycle regulation, including cell migration and adhesion. Its exact function is still to be unraveled, but a role in the pathogenesis and progression of PTC is presumed (38, 155–157). It is related to inhibition of apoptosis, induced by abnormal p53 expression (157). Galectin-3 can be present both in the intracellular as well as the extracellular matrix (158). Normal thyrocytes do not express galectin-3, but the physiological expression of galectin-3 in macrophages, neutrophils, mast cells, and Langerhans cells provides an internal positive control of the investigated FNAC samples (159, 160). Positive cytoplasmic staining, as opposed to nuclear staining, for galectin-3 is suspicious for malignancy and mainly associated with PTC (126, 161, 162). Galectin-3 expression has also been associated with the malignant transformation of follicular neoplasms, as it was present in FA as well as FTC (158, 163, 164). Encapsulated FVPTC and minimally invasive FTC showed less frequent and weaker staining (164, 165).

In 2001, Bartolazzi et al. (126) argued that galectin-3 staining could accurately diagnose thyroid carcinoma in unselected thyroid cytology. Subsequent studies in indeterminate thyroid cytology mostly could not reproduce these promising results. With a positive stain in approximately one-third of all nodules, sensitivity and specificity of galectin-3 ranged from 0% to 92% and from 68% to 100%, respectively (38, 166–169). Merely Saggiorato et al. (166) demonstrated that galectin-3 accurately differentiated follicular adenomas from FTC with 92% sensitivity and 94% specificity if a cytoplasmic stain in ≥10% of the cells was considered positive (166). The prospective multicenter clinical validation study by Bartolazzi et al. (156) demonstrated 78% sensitivity and 93% specificity in Thy nodules if a cytoplasmic galectin-3 stain in >5% of the cells was considered positive. Nineteen of the 22 false-positive nodules were FA. However, a group of 33 difficult-to-diagnose (follicular) tumor of unknown malignant potential lesions was disregarded, 22 of which were galectin-3 negative. If these neoplasms were considered malignant, sensitivity dropped to 69% (156).

**HBME-1**

HBME-1 is a monoclonal antibody targeting an unknown antigen on the microvilli of mesothelial cells. It is usually negative in normal thyroid follicular cells. Abnormal expression of HBME-1 shows cytoplasmic location with membrane accentuation. It is associated with, but does not necessarily indicate PTC (157, 161, 194, 195). Its low detection limit enables assessment in liquid based cytology (170). Reported sensitivity and specificity of HBME-1 in indeterminate nodules ranged from 61% to 100% and from 73% to 96%, respectively (166, 168, 170, 171). Approximately two out of five nodules showed positive staining. If only nononcocytic follicular neoplasms were selected, Saggiorato et al. (166) demonstrated that HBME-1 had excellent 93% sensitivity and 98% specificity in indeterminate thyroid nodules.

**Cytokeratin 19**

CK-19 is a type I keratin. It belongs to the group of intermediate filament proteins, which arrange the cell cytoskeleton and structural integrity. CK-19 is widely present in epithelial cells, but also found in basal cells layers of stratified epithelium (166, 172). Strong and diffuse abnormal expression of CK-19 indicates PTC, including FVPTC. Expression in FTC is less intense and more variable, warranting nuanced interpretation of CK-19 staining intensity. CK-19 usually shows no or only focal expression in follicular neoplasms, hyperplastic nodules, and adenomatous goiter (166, 194, 172, 196). The reported sensitivities and specificities for CK-19 staining in indeterminate cytology ranged from 76% to 88% and 80% to 100%, respectively (166, 170, 172). Lacoste-Collin et al. (170) demonstrated the importance of an accurate threshold. CK-19 staining in 31 Bethesda IV nodules accurately diagnosed five out of six malignancies, including a PTC, two FTC, and two out of three FVPTC. At a threshold of ≥30% stained cells, 5 of 25 benign lesions tested false-positive; at a more sensitive threshold of ≥10% stained cells, 12 of 25 tested false-positive (170).
Other ICC markers

IHC studies identified more potential ICC markers. Some, like CD44v6, have not yet been investigated in indeterminate cytology (126). Other markers were sporadically investigated in preclinical studies, including Ki-67, TROP-2, emerin, keratan sulfate, thyroperoxidase, CD57, and GLUT-1.

Nuclear protein Ki-67 is expressed in nearly all cell cycle phases in proliferating cells. It is associated with poor prognosis in PTC (197). The percentage of cells with Ki-67 expression is considered the tissues’ proliferative index. At a cutoff of ≥1% Ki-67 was 85% sensitive and 71% specific for thyroid carcinoma in Bethesda IV nodules. A combination of HBME-1, CK-19, and Ki-67 immunomarkers was 91% accurate to diagnose malignancy (170). Ki-67 expression is likely only distinctive for follicular type carcinoma; expression in PTC is generally low (170, 197, 198).

Glycoprotein human trophoblast cell surface marker (TROP-2) is overexpressed on the cell surface of different epithelial carcinoma (e.g., breast, colon) and associated with tumor aggressiveness and poor prognosis. In indeterminate thyroid cytology, it was only assessed in one small subspecies of Bethesda III samples, correctly diagnosing the three included carcinoma and all but one of the nine benign nodules (199).

Emerin staining emphasizes features of the nuclear membrane often seen in PTC, such as irregularities and invaginations. Consequently, the staining could facilitate the morphological diagnosis of PTC and especially the more difficult-to-diagnose FVPTC (168, 200). In 53 Thy3 nodules assessed by Asioli et al. (168), positive emerin staining was highly specific for PTC (including FVPTC), but misdiagnosed all FTCs.

Another immunomarker associated with PTC is keratan sulfate, an abnormal glycosaminoglycan complex. It was 98% specific in indeterminate cytology, but correctly predicted PTC only; its sensitivity was poor at 48% (166).

The expression of thyroid peroxidase (TPO) is related to benign follicular neoplasms. A negative TPO stain was 80% sensitive and 86% specific for thyroid malignancy (166).

Finally, CD47 (Leu7) expression is associated with epithelial and nonepithelial malignancies, including thyroid carcinoma. Cytological staining was only investigated in a small series of indeterminate cytology, but seemed specific for PTC. In the same series, GLUT-1 was not a useful ICC marker, there were no positive stains (201).

Combined use of ICC markers

Some research groups have suggested that evident single-marker galectin-3 positivity is sufficient to refer a patient for total thyroidectomy (156, 162, 169). The ATA guidelines did not adopt these suggestions, and many other researchers advocate that a panel of ICC markers should be applied to strengthen the suspicion of malignancy (8, 157, 166, 195). Several panels were investigated in literature. Zhang et al. (157) assessed a triple stain of galectin-3, HBME-1 and p27. P27 is a cyclin-dependent kinase inhibitor related to cell life span in normal thyroid cells. Downregulated in malignancy, positive P27 stain is related to benign histopathology. In a set of Bethesda III cytology samples, positive p27 staining with negative galectin-3 and HBME-1 staining was 100% predictive of a benign nodule and occurred in 38% of samples. Loss of p27 staining in combination with positive galectin-3 and/or HBME-1 staining was 100% sensitive and 86% specific (157). Another study investigated galectin-3 and HBME-1 in combination with a RET proto-oncogene stain, which reflects abnormal intracellular RET proto-oncogene activity and presence of the RET/PTC rearrangement. Unfortunately, RET staining was inaccurate in indeterminate thyroid nodules (171).

To find the most accurate combination of immunostains, Saggiorato et al. (166) explored the expression of galectin-3, HBME-1, thyroperoxidase, CK-19, and keratan-sulfate in 125 cytological follicular neoplasms, 24 of which were Hürthle cell lesions. Galectin-3 was not only the most accurate marker individually, but also in combination with other stains. Sequential HBME-1 staining of galectin-3-negative cases reached 98% sensitivity and 98% specificity in nononcocytic lesions. In oncocyic lesions, sequential CK-19 staining was more preferred with 100% sensitivity and 100% specificity (166).

The common denominator between all these studies is the combined use of galectin-3 and HBME-1.

Performance of ICC in Hürthle cell cytology

Expression of ICC markers in Hürthle cell nodules differs from nononcocytic indeterminate cytology. Hürthle cell carcinomas were distinguished in the cytological samples by typical overexpression of markers associated with a high degree of cell proliferation, disorganized tissue structure, and intermediate differentiation, such as Ki-67, laminin, cyclin D1, and cyclin D3. Overexpression reflects the known more erratic behavior of Hürthle cell carcinoma (198, 202). Moreover, markers that were highly diagnostic in indeterminate nodules in general, also seem differently expressed in Hürthle cell lesions. Saggiorato et al. (166) demonstrated that two combinations of ICC markers were extraordinarily accurate: galectin-3 and CK-19 staining was 100% sensitive and 100% specific; galectin-3 and thyroperoxidase staining was 100% sensitive and 85% specific (166).
In a previous meta-analysis, inclusion of Hürthle cell lesions was related to between-study heterogeneity (155). Hürthle cell lesions require a biotin-free ICC method, as Hürthle cells themselves are rich in biotin. Thus, much-used biotin-based methods may consequently cause false-positive and highly intensive staining in Hürthle cell neoplasms (158, 164, 171).

**Availability, cost-effectiveness, and limitations of ICC**
Current application of ICC is limited. Clinical validation studies for all of the described immunomarkers are scarce, and no cost-effectiveness studies are available to date. Yet, the technique is widely available, relatively inexpensive and fast in comparison with other (molecular) techniques. Costs per immunostain vary up to €20, partly depending on simultaneous local application of the technique and similar stains for IHC.

ICC is preferably performed on cell block FNAC specimens, but can be performed in all types of cytology, from direct smears to liquid-based cytology (171, 195). ICC is impossible when the FNAC specimen has poor cellularity or too much obscuring blood (203). Also, immunostaining of cytology is technically more difficult than histological staining, especially in (destained) cytology smears. Technical inconsistency and interobserver variation likely lead to false-negative results (156, 172). Stain intensity thresholds or percentage of stained cells necessary to raise suspicion of malignancy vary in the available literature. Consistent methodology and assessment thresholds should be determined to improve reproducibility of ICC results.

Clinical validation studies of existing ICC markers are ongoing. Meanwhile, new markers are also playing the field, searching for the interfaces between mutation analysis of highly specific oncogenic driver mutations and accessible ICC techniques. For example, Leslie et al. (92) investigated ICC of the BRAFV600E mutation using the mutation specific antibody VE1 in a small series of thyroid FNAC samples. Concordance between ICC and conventional BRAFV600E mutation analysis was 85%. All samples that were BRAFV600E positive by either method were confirmed as BRAFV600E positive PTC on histopathology. Of the eight included indeterminate thyroid nodules, seven were histopathologically malignant and BRAFV600E mutation was detected in two nodules: one by both methods, one only by molecular analysis. The BRAFV600E specific antibody (VE1) stain was much weaker in cytology than in histology. Moreover, costs of the VE1 antibody are currently high and optimization of methodology is warranted. Yet, Leslie et al. (92) demonstrated that BRAFV600E mutation analysis using ICC is a promising alternative to mutation analysis. If future studies could validate these results in larger cohorts of indeterminate thyroid nodules and detect reliable immunomarkers for other oncogenic driver mutations, this technique unites the strengths of gene mutation analysis and ICC in one technique, though likely at lower costs.

In general, ICC is a widely available and relatively inexpensive technique with a reasonable diagnostic accuracy. Many immunomarkers seem to have a pronounced association with PTC. Galectin-3 and HBME-1 were most frequently investigated, but their specificities and sensitivities seem to fall short of justifying ICC-based surgical decision making. Diagnostic accuracy of their combined use seems promising, yet current evidence is limited. Prospective validation trials are warranted to confirm the diagnostic potential of ICC, including validation of thresholds for stain positivity, panels of multiple immunostains, and other methodology.

**Conventional Imaging**

**Ultrasound**
US is one of the principal steps in the initial workup of thyroid nodules. It is cheap, fast, noninvasive, and globally available, but accurate assessment strongly depends on operator experience (204). Multiple meta-analyses showed that well-known US features such as nodule hypoechogenicity, microcalcifications, irregular margins (including microlobulated or ill-defined margins), and a taller-than-wide shape raise the suspicion for thyroid malignancy and are mostly associated with PTC (204, 205). Nonetheless, no single US feature is sufficiently sensitive or specific to accurately identify a malignant nodule in an unselected population (204). Certain combinations of US features, however, may offer accurate closure. The current ATA guidelines now include a flowchart recommending FNAC dependent on nodule size and various combinations of US characteristics with an incremental risk of malignancy (8). Despite the obvious importance of both US and cytology, the ATA guidelines do not provide recommendations regarding (re)interpretation of US characteristics after FNAC has resulted in indeterminate cytology. Follicular-type malignancies typically have a different US appearance. More often FTC may be iso- to hyperechoic, with a spherical shape, smooth regular margins, and no calcifications (206, 207). FVPTC may also show FTC-like or benign features rather than the classic suspicious features, although microcalcifications may be distinctive (207–209). In the past years, Brito et al. (204) and Remonti et al. (205) performed meta-analyses on US assessment of unselected thyroid nodules. Both also briefly discussed its diagnostic value in indeterminate nodules, including a mere limited number of studies and also including cytology suspicious for malignancy. Increased central vascularization was most predictive of malignancy with reported 96% specificity (205). Yet, in general US seemed less accurate in
indeterminate nodules than in unselected thyroid nodules (204, 205).

In the dozens of available original ultrasound studies, individual US features generally demonstrated limited sensitivity in indeterminate thyroid nodules. Only the appearance of a solid thyroid nodule, as opposed to varying degrees of cystic content, had high sensitivity. Ranging between 46% and 100%, multiple studies demonstrated sensitivity above 90% (19, 20, 210–214).

A number of classic suspicious US characteristics, such as a taller-than-wide shape, presence of irregular margins, and presence of microcalcifications, demonstrated valid specificity in indeterminate thyroid nodules. Specificities for each of these characteristics ranged from 72% to 99% (214–217), 65% to 100% (215, 218, 219), and 36% to 100%, respectively (220, 221). Despite the wide range, presence of microcalcifications was more than 90% specific in many studies (210, 211, 213, 215–217, 219, 220, 222–224). Large nodule size (defined as a diameter larger than 4 cm) was only investigated in a limited number of studies. Reported specificities ranged between 69% and 94% (220, 225).

Other features, such as a solitary nodule, hypoechoicity and absence of a hypoechoic halo were associated with thyroid malignancy, but less accurately differentiated between benign and malignant indeterminate thyroid nodules (174, 211, 214, 219, 225–229). Additionally, opposing the results from one of the mentioned meta-analyses, central vascularization also does not seem very accurate in indeterminate thyroid nodules. Specificity ranged from 0% to 100%, although multiple studies demonstrated extremely poor specificity (19, 215, 228–232).

Results regarding two US features are remarkably contradicting. First, the absence of a hypoechoic halo is typically considered suspicious for malignancy, but showed overall poor and very heterogeneous diagnostic potential in indeterminate thyroid nodules (204). Sensitivity and specificity ranged from 17% to 99% and 0% to 93%, respectively (173, 175, 213, 214, 218). Presence of a hypoechoic halo is typically considered a benign feature, but has also been associated with follicular types of thyroid carcinoma (233). Dogan et al. (234) reported 88% specificity for presence of a halo in AUS/FLUS nodules and 78% in FN/SFN nodules. Second, the ultrasonographic nodule shape seems ambiguous. Similar to the unselected population, a typically suspicious taller-than-wide shape was generally specific for carcinoma, with reported specificities up to 99% (214, 215, 217). A spherical shape is generally considered benign, but has also been associated with FTC (204, 206, 235). In two studies in cytological follicular neoplasms, a spherical shape had an increased risk of malignancy, with 86% to 97% sensitivity and 19% to 26% specificity (236, 237). Chin et al. (237) even suggested that follicular neoplasms with a taller-than-wide shape could be treated conservatively. The uniquely balanced rates of PTC, FVPTC, and FTC resulting from indeterminate cytology may explain why these and various other US characteristics have different diagnostic accuracy than in the unselected population. Dependent on the local case mix, accurate differentiation of indeterminate nodules using the classical suspicious US features may or may not be feasible.

**Combination of US characteristics**

A combination of US characteristics likely provides more accurate differentiation than individual features. Different combinations were investigated in multiple studies (93, 210, 211, 213, 217–219, 228, 238–243). Yoo et al. (214) reported 100% specificity for the combination of marked hypoechogenicity and taller-than-wide shape, a pattern that occurred in 9.6% (24/249) of the included Bethesda III nodules. In the elastosonography study by Rago et al. (173), absence of a hypoechoic halo in combination with presence of microcalcifications was 95% specific for thyroid malignancy, but only 64% sensitive. Maia et al. (240) found 62% sensitivity and 89% specificity in Bethesda III and IV nodules if hypoechogenicity, microcalcifications, an irregular margin, and increased intranodular vascularity were considered suspicious. Gulcelik et al. (244) demonstrated that the US pattern of a solid, hypoechogenic nodule with microcalcifications had 95% sensitivity and 99% specificity. The pattern was seen in 21% of cytological follicular neoplasms.

In multiple studies, it was argued that cytological follicular neoplasms with a typically benign US pattern (a regular shape, isoechoic, homogeneous, with well-defined margins, cystic components or peripheral vascularity only, and not a single malignant feature) could be safely followed up clinically instead of undergoing diagnostic surgery (213, 235, 238). Consideration of more features generally increased the sensitivity of the US assessment at the cost of its specificity (211, 212). The terms of their interpretation were crucial: Norlen et al. (241) demonstrated that US was 95% sensitive and 48% specific if a Bethesda III nodule had either hypoechoic appearance, irregular margins, or microcalcifications. If solely the simultaneous presence of all three features was considered suspicious for malignancy, sensitivity dropped to 37% but specificity increased to 96%.

 Altogether, diagnostic US scores or step-by-step algorithms could aid the classification of US patterns and consequent risk of malignancy.”
irregular margins, microcalcifications, and a taller-than-wide shape. Nodules without any of these features are likely benign and categorized as TIRADS 3. Their risk of malignancy is ~1.7% in a cytologically unselected population. TIRADS 4 includes suspicious nodules, which are further classified according to an increasing malignancy risk into 4a (one suspicious US feature), 4b (two suspicious features), and 4c (three or four suspicious features). Nodules with all five suspicious US features are classified as TIRADS 5 and associated with a high 88% risk of cancer in an unselected population (247). Studies that validated the TIRADS specifically in indeterminate thyroid nodules showed that diagnostic accuracy depended on the chosen cutoff score and type of cytology (215, 245, 248–251). Although TIRADS 5 scores were infrequently assigned in indeterminate nodules, a higher TIRADS score (4b/4c/5) was an accurate predictor of malignancy, especially in Bethesda IV cytology (215, 248, 250). In Bethesda III nodules, lower TIRADS scores (3/4a) could also rule out malignancy (248, 251). Prospective validation studies applying the TIRADS in indeterminate cytology are warranted to assess its possible clinical utility in indeterminate nodules.

**US performance in Hürthle cell nodules**

Cytological Hürthle cell nodules expressed a large variation of US characteristics (216, 221, 225, 233). Many malignant and most benign Hürthle cell nodules had a benign US appearance (221, 233). Only three US features possibly predictive of malignancy were reported in individual studies: both hypoechoogenicity and hyperechogenicity (as opposed to isoechogenicity) (221), large nodule size (233), and microcalcifications (216). Despite limited evidence, US evaluation does not seem reliable to differentiate Hürthle cell lesions.

**Availability and limitations of ultrasonography**

The major advantages of US over other additional diagnostics are its already permanent position in the workup of thyroid nodules, global availability, and low costs. No additional resources or hospital visits are needed to include US interpretations in preoperative management decisions and the investigation is non-invasive. Nonetheless, besides known limitations concerning interobserver variability and less reliable interpretation of small nodules, US feasibility in indeterminate thyroid nodules is limited by the presumed differences in US appearance of papillary and follicular thyroid malignancies, illustrated by the conflicting results for nodule shape and hypoechoic halo in indeterminate nodules. Consequently, local diagnostic accuracy likely follows variations in the local histopathological case mix.

In addition, many of the available ultrasound studies are retrospective, limiting the power of the evidence. As the decision to perform FNAC is customarily based on the results of the prior US, the prevalence of suspicious US features in indeterminate cytology in these studies is presumably overestimated.

Nonetheless, several individual US characteristics seem to have reasonable specificity in indeterminate nodules, although insufficient for accurate diagnosis. A combination of US features is likely more accurate, although current evidence does not support US-based surgical decision-making. We propose that a future meta-analysis should use the individual patient data from the large number of available original studies to develop a US algorithm specifically for indeterminate thyroid nodules. The existing TIRADS needs prospective validation.

Even though more advanced and less operator-dependent techniques might be preferred, US features should always be assessed in current clinical practice. The presence of one or more suspicious US features in a Bethesda III or IV nodule increases the suspicion of malignancy and underpins the need for a definite diagnosis. Moreover, to centers or regions with limited access to other (molecular) diagnostics, US may definitively have clinical utility, pending local validation in the indeterminate population.

**Elastosonography**

Firm consistency of a thyroid nodule upon palpation is considered suspicious for malignancy, an established principle during physical examination (252). US elastosonography (USE) is a dynamic US technique that is sometimes referred to as "electronic palpation." Tissue elasticity is evaluated by measuring tissue distortion while applying a standardized crossed external force by the US transducer. It was first applied to the thyroid gland by Lyschik et al. (253) in 2005. Classic real-time qualitative USE is performed by free-hand compression and a sine-wave or numerical scale showing how much pressure the operator applies with the probe. A color-coded elastosonography image is superimposed on the grayscale US images: red and orange visualizes high tissue elasticity (soft tissue), green represents intermediate elasticity, and blue low elasticity (firm tissue). Several score systems are available. The original score was developed by Itoh et al. (254) in 2006 for the evaluation of breast tumors and considers scores 1 to 3 benign on a scale of 1 (highest elasticity) to 5 (no elasticity). Rago et al. (173, 255) first applied it to thyroid tumors and modified it to a three-point score. Astera et al. (256) derived a modified four-point score.

The earliest studies in thyroid nodules reported opportune results of USE as an additional modality to B-mode US, but were heterogeneous in USE technique and study population (257). A recent meta-analysis by Nell et al. (258) included 20 studies on qualitative USE prior to FNAC and concluded that qualitative USE is fit to diagnose benign nodules and safely dismiss FNAC, provided that the usual elasticity score cutoff is
abandoned and only completely soft nodules (score 1 of all systems) are classified as benign. Pooled 99% sensitivity and 99% negative predictive value demonstrated the ability of USE to reliably rule-out malignancy in entirely soft thyroid nodules, composing 14% of their pooled study population (258).

In individual studies on USE in indeterminate nodules, sensitivity and specificity of qualitative USE ranged from 47% to 97% and from 6% to 100%, respectively (20, 173–176). Results of several qualitative USE studies stand out. Lippolis et al. (174) showed an aberrant 6.1% specificity, because they reported only eight nodules with high elasticity, 62 of 66 benign nodules were not elastic. The authors themselves suggest that a rather homogenous study population with predominantly small nodules with a solid US pattern, absence of cystic areas, and follicular histology with minimal colloid could be explanatory for the poor specificity rather than operator-dependent causes (174). Such possible relations remain undescribed in other studies. A meta-analysis on the value of USE in indeterminate thyroid nodules demonstrated meager pooled 69% sensitivity and 75% specificity (259).

As manually applied pressure is difficult to standardize, qualitative USE is strongly operator dependent (260). Different USE techniques have been developed to improve objectivity, such as semi-quantitative tissue-to-nodule strain ratio indices (also based on manual compression). Studies investigating semi-quantitative USE in indeterminate thyroid nodules reported sensitivity and specificity ranging from 82% to 100% and from 88% to 100%, respectively (176, 227, 229, 231). Furthermore, quantitative shear wave USE measures the propagation velocity of focused acoustic pulses, shear waves, from the probe, which correlate to tissue stiffness (Young’s modulus) (19, 261). It had 82% sensitivity and 88% specificity in a recent prospective pilot study by Samir et al. (19). Performance of (semi)quantitative USE seems better than qualitative USE, but results are subject to overfitting from the ROC analysis performed to determine the strain ratio cutoff value with the highest sensitivity and specificity. None of the studies applied a predefined cutoff or validated their own cutoff externally. Consequently, the resulting thresholds were hardly comparable (176, 227, 229, 231).

Altogether, the results from currently available studies cannot support surgical decision-making in thyroid nodules with indeterminate cytology using elastosonography in any of its forms. Whereas color-coded qualitative USE has insufficient sensitivity and specificity, the semiquantitative method lacks validation. The power of the available evidence is additionally limited by both methodological heterogeneity and the use of different USE techniques, image processing, and elasticity scoring methods across studies. Nevertheless, the suggested promising rule-out capacity of qualitative USE when applying an alternative cutoff score of 1 in unselected nodules deserves clinical validation in indeterminate thyroid nodules. Major advantages of the technique are the minor extra costs of USE, as it can be performed during regular thyroid US with the same equipment, and only adds −5 minutes to the procedure time per patient. Cost-effectiveness will largely depend on performance of USE, but no cost-effectiveness studies in indeterminate thyroid nodules are available to date.

**Computed tomography**

There are no studies that investigated CT scanning in thyroid nodules with indeterminate cytology. Prior studies indicated that CT cannot accurately differentiate thyroid carcinoma (262, 263).

**Functional and Molecular Imaging**

**99mTc-MIBI**

Hexakis(2-methoxy-2-methylpropionitrile)technetium$[^{99mTc}]$ (99mTc-MIBI) is a Technetium-99m-labeled radiopharmaceutical, primarily known for its use in myocardial perfusion imaging since the 1980s and more recently the evaluation of hyperparathyroidism. Uptake of 99mTc-MIBI, a lipophilic cation, reflects both perfusion and the number of active mitochondria in the cells of the thyroid nodule and thus its oxidative burden (61, 264). 99mTc-MIBI scintigraphy is more suitable for the differentiation between benign and malignant thyroid nodules than scintigraphy with 99mTc-pertechnetate ($[^{99mTc}]$O$_4^-$) or radioisotopes of iodide (often $^{131I}$–, $^{123I}$– or $^{124I}$–). These latter tracers interrogate the sodium-iodide symporter of the thyrocyte and are frequently used to assess thyroid nodule functioning to distinguish autonomous (“hot”) from hypofunctioning (“cold”) nodules. They are neither specific nor effective to detect malignancy: benign nodules can be anything from hyper- to hypofunctioning, and far outnumber the carcinomas. Still, thyroid malignancies are almost always hypofunctioning: decrease of the sodium-iodide symporter or TPO are hallmarks of cell dedifferentiation and lead to loss of iodide-trapping function and thus $[^{99mTc}]$-pertechnetate or radiiodine uptake (8, 264–266). 99mTc-MIBI uptake is independent of iodide trapping and organization in the thyrocytes. Nodules with increased uptake and late retention of 99mTc-MIBI are suspicious for malignancy (61, 264). A 2013 meta-analysis by Treglia et al. (264) demonstrated 82% sensitivity and 63% specificity for 99mTc-MIBI scintigraphy in clinically suspicious, hypofunctioning, cytologically unselected thyroid nodules. Hyperfunctioning benign adenomas can show false-positive increased uptake of 99mTc-MIBI due to their increased metabolic needs, thereby decreasing test specificity.
Only three studies investigated the role of $^{99m}$Tc-MIBI in indeterminate thyroid nodules. In all studies, evaluation of thyroid nodules was performed by dual-time planar imaging; an early image was made ranging from 10 to 20 minutes after injection of the radiopharmaceutical and a delayed image 60 to 120 minutes post injection. The intensity of the $^{99m}$Tc-MIBI uptake within the nodule, and possible increased uptake or denoting retention on delayed imaging were assessed within the nodule, and possible increased uptake or denoting retention on delayed imaging were assessed. A visual pattern of increased $^{99m}$Tc-MIBI uptake on early images that persisted or further increased on the delayed images was generally considered suspicious for malignancy. The individual study sensitivity and specificity for this interpretation ranged from 56% to 79% and from 52% to 96%, respectively (20, 61, 177). Despite the limited number of available studies, the performance of $^{99m}$Tc-MIBI in indeterminate thyroid nodules seems insufficient and less accurate than in cytologically unselected nodules (264).

Nonetheless, Giovanna et al. (61) demonstrated that NPV for this method could increase from 88% to 100% if only the pattern of $^{99m}$Tc MIBI uptake lower than or equal to the pertechnetate uptake within the nodule was considered benign. As few benign lesions expressed this uptake pattern, this would decrease the yield of this diagnostic.

Piccardo et al. (20) did not preselect hypofunctioning lesions, but included all indeterminate thyroid nodules. As expected given the explanation above, the specificity of $^{99m}$Tc-MIBI was poor: 52%.

Assessment of a retention index of the tracer based on semiquantitative measurements of the lesion to nonlesion uptake ratios for early and delayed $^{99m}$Tc-MIBI images yielded better accuracy. Optimal thresholds for the retention index were determined using ROC analysis and unfortunately not externally validated (61, 177). As such, it is unclear whether semiquantitative $^{99m}$Tc-MIBI retention indices are truly more accurate than conventional visual assessment. Moreover, semiquantitative analysis is still operator dependent, as it depends on the manual definition of ranges of interest (ROI) (20).

$^{99m}$Tc-MIBI in thyroid nodules with Hürthle cell cytology

Oncocytic cells are rich in mitochondria. Therefore, Hürthle cell lesions, malignant as well as benign, frequently show a more intense and persistent $^{99m}$Tc-MIBI uptake (177, 267, 268). Boi et al. (268) investigated $^{99m}$Tc-MIBI in cold thyroid nodules with varying proportions of Hürthle cells in the cytology samples. A relation between $^{99m}$Tc-MIBI uptake and increased tissue density of oncocyes was suggested (268). Subsequent studies also concluded that $^{99m}$Tc-MIBI is not specific enough to differentiate indeterminate lesions with Hürthle cell cytology (61, 177, 264). Excluding Hürthle cell nodules from $^{99m}$Tc-MIBI assessment likely excludes many false-positive tests while improving benign call rate, specificity, and overall diagnostic accuracy in indeterminate thyroid nodules.

Availability, cost-effectiveness, and limitations of $^{99m}$Tc-MIBI

Imaging of $^{99m}$Tc-MIBI requires conventional gamma cameras [with or without single-photon emission computed tomography (SPECT) and CT], which are more widely available than PET, especially in non-Western countries. Furthermore, the tracer itself is more widely available due to relatively simple complexity using $^{99m}$Tc-MIBI-kits together with the favorable half-life of $^{99m}$Tc (~6 hours) obtained from on-site generators. The radiation burden of the recommended whole-body adult dose is 5 to 6 milisievert, but can be lowered by a factor 2 to 3 by partial-body imaging (269). However, the system resolution of state-of-art gamma cameras is a factor 3 lower than of PET/CT cameras. This decreases the measured signal of lesions smaller than 30 mm, increasingly limiting test sensitivity in smaller nodules. Average costs of $^{99m}$Tc-MIBI scanning range from €119 to €500 in Europe and from $69 to $1156 in the United States (84, 178, 179). From a German perspective, $^{99m}$Tc-MIBI-based management was cost-effective over Afirma® GEC-testing and conventional management. However, this study inappropriately extrapolated auspicious performance parameters of $^{99m}$Tc-MIBI in unselected thyroid nodules (96% sensitivity and 46% specificity) to the indeterminate population, and likely underestimated modeled costs for $^{99m}$Tc-MIBI scanning and thyroid surgery (15, 72, 83, 84, 178, 179, 270). Therefore, these assumptions regarding cost-effectiveness in indeterminate thyroid nodules are decidedly questionable and require careful re-evaluation.

Altogether, there is an increased risk of malignancy in thyroid nodules that show increased $^{99m}$Tc-MIBI uptake, provided that hypofunctioning nodules are preselected. Nonetheless, test performance in indeterminate thyroid nodules seems insufficient. Excluding Hürthle cell lesions suggests high specificity, but does not resolve the reported poor sensitivity. However, the number of studies currently available for indeterminate thyroid nodules is limited. We believe prospective validation studies in nononcocytic indeterminate thyroid nodules should be performed. Future studies should also focus on external threshold validation for retention indices to reduce operator dependency and increase accuracy and objectivity of $^{99m}$Tc-MIBI. Based on the current evidence, we recommend that $^{99m}$Tc-MIBI scanning is not used in surgical management decisions in indeterminate thyroid nodules without another adjunctive test.
FDG-PET is an imaging modality that exploits the basic principle that (malignant) tumors and inflammatory tissues are much more metabolically active than normal tissues. Whereas normal tissues predominantly produce energy by low rates of aerobic glycolysis followed by the citric acid cycle in mitochondria, glycolytic rates of rapidly growing cancers can be up to 200 times higher. Subsequent lactic acid fermentation takes place even if oxygen is plentiful (the Warburg effect). Similarly to regular glucose, the glucose analog $^{18}$F-FDG is internalized by transmembranous GLUT transporters and converted by hexokinase to $^{18}$F-FDG-6-phosphate. However, unlike the 6-phosphorylation product of regular glucose, $^{18}$F-FDG-6-phosphate cannot be metabolized further. It is trapped intracellularly and thus accumulates in the tissue. Subsequently, PET scanning can visualize the increased glucose metabolism of the (abnormal) tissue.

FDG-PET is generally performed in combination with CT (FDG-PET/CT), mainly to correlate metabolically active regions to their anatomic substrates and to correct for tissue-attenuation of the radioactive signal. It is increasingly applied in the diagnostic workup, staging, and therapeutic response monitoring of various malignancies. For thyroid cancer, FDG-PET is frequently used to characterize recurrent disease, especially if dedifferentiation is expected in thyroid carcinomas that lost the capacity to concentrate radioiodide, yet still have measurable serum values of the tumor marker thyroglobulin. It may also be considered in the initial staging of poorly differentiated or invasive Hurthle cell carcinoma. Moreover, FDG-avid thyroid incidentalomas require additional workup by FNAC when $>1$ cm (8, 21). In the current ATA guidelines FDG-PET is not routinely recommended for the diagnostic workup of indeterminate thyroid nodules due to limited clinical validation, despite a 2011 meta-analysis by Vriens et al. (272) that demonstrated $95\%$ sensitivity and $96\%$ NPV in indeterminate thyroid nodules larger than $15$ mm (8, 272).

Results of available individual studies were mutually consistent despite limited sample sizes. The first studies showed extremely promising results; each reporting $100\%$ sensitivity (180–183). De Geus-Oei et al. (182) argued that implementation of FDG-PET could reduce the number of futile hemithyroidectomies for benign nodules by $66\%$, likely outweighing the costs of the extra scans and suggesting cost-effectiveness of this technique in the preoperative setting. A subsequent study suggested a less optimistic $39\%$ reduction in futile surgeries, following a lower benign call rate (183). More recent studies demonstrated more modest performance of FDG-PET/CT (20, 21, 184, 185, 273). Overall, reported sensitivity and specificity of FDG-PET/CT to detect thyroid carcinoma in indeterminate thyroid nodules ranged from $77\%$ to $100\%$ and from $33\%$ to $64\%$, respectively. A negative index test was reported in $40\%$ of patients (20, 21, 180–185, 273, 274).

Several reasons for false-negativity were proposed, foremost small nodule size. It is how Traugott et al. (273) explained their $20\%$ false-negative FDG-PET scans: eight lesions were histopathologically smaller than $1$ cm. Excluding these, sensitivity and NPV increased to $100\%$. FDG-avidity in very small nodules may be missed on FDG-PET due to the low volume of malignant cells and due to the partial volume effect: the detected FDG-concentration is underestimated dependent on nodule size in relation to the (limited) spatial resolution of the scanner. In larger nodules, this effect is negligible (21, 182). Although the improving resolution of state-of-the-art PET scanners pushes the detection limit toward $10$ mm, PET is less sensitive in lesions smaller than $15$ mm on US. It is less reliable to rule-out microcarcinomas (272). Theoretically, the improving spatial resolution could also become a limitation of the technique: not only will there be less false-negatives, but likely also more false-positive results, leading to a decrease in the already limited specificity over time. In the currently available literature, no such downward trend is noted, but future studies should monitor this possibility.

Semiquantitative FDG-PET

Semiquantitative analysis of FDG-PET is performed using the maximum standardized uptake value (SUV$_{\text{max}}$); the ratio between the maximum radioactivity concentration measured within a region of interest on the PET image (the "hottest" voxel) and the decay-corrected amount injected radiotracer per unit of body mass. It reflects the FDG-concentration factor compared with a homogenous distribution of the radiotracer (275). The SUV$_{\text{max}}$ is generally significantly higher in malignant than in benign lesions (21, 182–185, 275, 276). There is a possible correlation between higher SUV$_{\text{max}}$ values and increasing size in nodules, insufficiently explained by the above-mentioned partial volume effect (21, 276). Also, in FTC a higher SUV was associated with capsular or vascular invasion (274). Nonetheless, even though Kresnik et al. (181) demonstrated that all carcinoma and Hurthle cell adenoma had $aSUV_{\text{max}} \geq 2$ and all other benign lesions a $SUV_{\text{max}} < 2$, in multiple other studies the SUV$_{\text{max}}$ of benign and malignant indeterminate thyroid nodules overlapped. No threshold could accurately tell them apart (21, 182–185, 275, 276). Moreover, as SUV$_{\text{max}}$ calculations strongly depend on image acquisition and reconstruction methods, type of PET-scanner, and other variable methodology, reported absolute SUV$_{\text{max}}$ thresholds are not simply valid for other institutions (21). Standardized optimized FDG-PET protocols are...
required for interinstitution comparison of study results and advancement of PET research (277, 278).

**FDG-PET in thyroid nodules with Hurthle cell cytology**

Multiple studies observed aberrant FDG-PET characteristics in indeterminate nodules with Hurthle cell cytology: both benign and malignant lesions are mostly FDG-positive. Twenty-nine Hurthle cell lesions were reported by Deandreis et al. (184), consisting 52% of their study population and providing an explanation for their limited sensitivity. Moreover, Hurthle cell adenoma generally demonstrated a significantly higher SUVmax than other benign lesions (21, 180–182, 184, 279). The proportion of Hurthle cell cytology in individual studies is relatively small, but overall FDG-PET seems inadequate in these neoplasms.

**Availability, cost-effectiveness, and limitations of FDG-PET**

PET systems are less widely available than conventional gamma cameras. Moreover, 18F-FDG synthesis is produced in cyclotrons, and transport distances are limited due to the short half-life (~4.0 h) of this isotope (~24 h). Transport distances are limited due to the short half-life (~59 min) of 99mTc-MIBI. The radiation exposure of FDG-PET/CT is largely accounted for by the FDG dosage at ~19 µSv/MBq, i.e., ~3 to 4 mSv for a typical activity of 185 MBq administered to an average adult (280). Insights regarding common practice total-body FDG-PET/CT imaging are changing (281, 282). The CT radiation dose greatly varies, and can be less than 0.5 mSv for a low-dose CT of the neck region only. When scanning the thyroid region only, a longer imaging time can compensate for a reduction in FDG dose, which would lower the radiation burden as well as the costs. Such solutions may counter prevailing reservations regarding ionizing radiation exposure. Additionally, partial-body acquisition could limit the number of coincidental PET-positive findings. Much of the criticism on FDG-PET focuses on these potential incidental findings, which require additional diagnostics, are not always clinically relevant and may negatively impact potential cost-effectiveness (281, 282). Malignant ipsi- or contralateral thyroid incidentalomas are reported while the nodule under investigation was histopathologically benign (185, 273). PET-positive incidentalomas are histopathologically malignant in ~20% of patients (282). Cost-effectiveness of FDG-PET/CT was modeled by Vriens et al. (15). From a Dutch health care perspective, FDG-PET/CT driven treatment would decrease the rate of unbeneficial diagnostic hemithyroidectomies for benign thyroid nodules by 35% and reduce the costs per patient by €822 compared with the €5,804 expenses for conventional surgical treatment. Also, FDG-PET/CT was favored over the miRInform® and Afirma® GEC (15).

Contrasting the generally strong sensitivity, specificity of FDG-PET is consistently poor. The underlying mechanism is not yet fully elucidated. The negative influence of Hurthle cell cytology may be partly responsible. It could also be explained by cellular atypia, which was significantly and independently related to FDG uptake, and found in both benign and malignant lesions. Atypia was also related to the presence of Hurthle cells (184). Sebastianes et al. (183) hypothesized that FDG uptake is related to variations in gene expression patterns. They suggested that genetic variations between populations may also explain the varying diagnostic accuracy of FDG-PET between studies.

In conclusion, FDG-PET/(CT) has the potential to accurately rule-out malignancy in all indeterminate nodules except Hurthle cell lesions. It could prevent unnecessary diagnostic surgery for a substantial number of benign thyroid nodules. Sample sizes of existing studies are small, but larger prospective trials are currently ongoing to settle the diagnostic value of this technique and its utility in clinical practice. We recommend that these studies also focus on identifying (genetic) causes for the occasional false-negativity and generally low specificity of this technique.

**Diffusion-weighted magnetic resonance imaging**

Diffusion-weighted magnetic resonance imaging (DW-MRI) is a functional nuclear magnetic resonance imaging technique that evaluates the rate of random (Brownian) motion of water in tissue, also called diffusivity. By applying diffusion-sensitizing magnetic gradients (the strength and duration of which are expressed as b-values) different levels of diffusion-weighting are obtained: from nondiffusion images (b-value = 0 s/mm²) to highly diffusion weighted images (i.e., b-value > 800 s/mm²) (283). Lesions that show high signal intensity on DW-MRI images with a high b-value thus show restricted diffusion. The apparent diffusion coefficient (ADC, in mm²/s) is calculated based on the exponential relationship between signal intensity and the corresponding b-value according to $S(b) = S(0)e^{-b/ADC}$. A high ADC represents a high degree of diffusion; a low ADC represents diffusion restriction (283, 284). DW-MRI thus allows noninvasive quantification of tissue properties without ionizing radiation exposure for the patient. Differentiation between benign and malignant tissues by DW-MRI is based on the assumption that increased cell proliferation, cellular-density, and disorganized structures in malignant tissue restrict random motion and thus diffusion of water: a lower ADC value, together with high signal intensity at high b-values, is more suspicious for malignancy (283, 284). Oppositely, increased ADC-values suggest free movement of water molecules in the tissue. It is found in for example...
edema, colloid follicles, fibrous tissue, hemorrhage, and calcification, all of which associated with benign tissues (285). Prior application of DW-MRI in i.e., neuroradiology, breast, and lymph nodes showed high diagnostic accuracy (286, 287).

Recent exploratory studies in small cohorts of thyroid nodules found distinctively higher ADC values for benign than malignant nodules (283–285, 288–292). A recent meta-analysis in 765 cytopathologically unselected thyroid nodules estimated that DW-MRI had 90% sensitivity and 95% specificity to distinguish thyroid carcinoma (293). Among the individual studies, however, presented optimum ADC thresholds varied and were not externally validated (283–285, 288–292).

Only one small study had assessed DW-MRI in indeterminate thyroid nodules to date. Nakahira et al. (283) reported a mean ADC value of 1.27 ± 0.29 × 10⁻³ mm²/s in malignancies opposite 1.95 ± 0.24 × 10⁻³ mm²/s in benign nodules with indeterminate cytology. These results were similar to those of their entire study population (n = 42), in which a cutoff ADC value of 1.95 ± 0.24 × 10⁻³ mm²/s was 95% sensitive and 83% specific.

**Availability and limitations of DW-MRI**

DW-MRI is infrequently and only experimentally used in the workup of thyroid nodules. Nonetheless, the worldwide availability and application of MRI is growing. As it uses no ionizing but only radiofrequency radiation, the associated risk to the patients is limited, provided that specific measures are taken for patients with MRI-incompatible implanted devices or metal. No MRI-contrast is necessary for DW-MRI, thus avoiding gadolinium-associated toxicity. As the spatial resolution of MRI-scanners is still improving, technical limitations of DW-MRI with regard to minimal lesion size are becoming less relevant compared with SPECT and probably also PET. Still, spatial resolution of DW-MRI sequences is less than that of conventional anatomical MRI-sequences.

There are several major limitations to DW-MRI. MRI is still a rather costly technique; additional sequences such as DW-MRI adds scanner time (~5 to 10 minutes) per patient and thus further increases costs. DW-MRI methodology is not standardized yet and its optimal settings still unsettled, leading to varying ADC and b-values (283, 289, 292). Suboptimal methodology or artifacts cause poor image quality, impede accurate interpretation, and caused undesirable exclusions from already small-sized studies, with reported exclusion rates up to 28% (283, 284, 292). Image artifacts are often caused by inhomogeneity in pathologic tissues or by their vicinity to interfaces between soft-tissues and bone or air, a source of MRI-arts specific in the thyroid region. Besides viable tumor tissue, malignant tumors partly exist of components with high diffusivity, such as necrosis, cystic components, or intratumoral hemorrhage (283, 285). For accurate ADC measurement, such macroscopic areas should be manually avoided when drawing a region-of-interest.

However, avoiding microscopic areas of similar origin, invisible to the human eye, is an impossible task (285). Furthermore, it is hypothesized that the substantial amounts of follicular or Hürthle cells limit the diagnostic accuracy of DW-MRI, specifically in indeterminate thyroid neoplasms. Follicular and Hürthle cell neoplasms are known for their varying colloid tissue involvement. Histologically they contain more fluid. Thus, DW-MRI would inaccurately provide a more benign image (283, 292). These hypotheses are currently based on very limited evidence. Further prospective validation studies are desired to determine the possible diagnostic value of DW-MRI in indeterminate thyroid nodules. Future prospects also include improvements of the technique, including consensus on methodology and standardization of acquisition techniques.

**Combined and Multistep Diagnostics**

The previous sections of this review addressed the large number of available diagnostic tools to assess indeterminate thyroid nodules. Most studies focused on a single diagnostic technique only. The elimination of between-study population-level differences is a major advantage when comparing the performance of multiple diagnostics independently in one study, optimally in a prospective, independent, and blinded fashion. Moreover, assessment of multiple techniques in one study allows investigation of the complementary value of multiple techniques as a diagnostic tool by means of simultaneous or sequential testing while at the same time aiding to further unravel tumor biology as a research tool, especially in the current multidisciplinary in-hospital working environment. For example, the question how the presence of a certain oncogenic mutation relates to the (positive) result of an FDG-PET scan could be addressed.

Piccardo et al. (20) compared ⁹⁹mTc-MIBI, FDG-PET/CT, and US plus USE in 87 indeterminate TIR³ nodules with a 21% malignancy rate. FDG-PET/CT was the superior technique with 94% sensitivity and 58% specificity. Following a nonspecific positive FDG-PET result, review of US characteristics offered slight further differentiation; it improved specificity to 77%. However, an additional negative ⁹⁹mTc-MIBI scan increased specificity to 94%; this combination was found in 13% of patients.

Giovanella et al. (61) performed both ⁹⁹mTc-MIBI and a seven-gene mutation panel in cold indeterminate thyroid nodules. Combined testing did not improve diagnostic accuracy. Performance of the gene mutation panel was inferior to ⁹⁹mTc-MIBI imaging. Of the seven (11%) mutation-positive
nODULES (FOUR RAS MUTATIONS AND THREE PAX8/PPARγ RARREARRANGEMENTS), ONLY FOUR WERE MALIGNANT. IT IS UNCLEAR WHETHER THE LOW SENSITIVITY OF THE GENE MUTATION PANEL IN THIS STUDY CAN BE EXPLAINED BY THE SELECTED POPULATION OF HYPOFUNCTIONING NODULES.

ELASTOSONOGRAPHY AND ULTRASONOGRAPHY
USE IS SUPERIOR TO US IN INDETERMINATE THYROID NODULES, BOTH INDIVIDUAL US CHARACTERISTICS AS WELL AS COMBINED US PATTERNS DESCRIBED IN VARIOUS ARTICLES (173–175, 223, 227, 229, 231). TWO RECENT PROSPECTIVE STUDIES DEMONSTRATED THAT ADDITIONAL USE EVIDENTLY IMPROVED THE DIAGNOSTIC ACCURACY OF US. GARINO ET AL. (223) INCLUDED NODULE STIFFNESS AS ADDITIONAL CHARACTERISTIC INTO A PANEL OF US CHARACTERISTICS AND DEMONSTRATED THAT USE IDENTIFIED EIGHT ADDITIONAL MALIGNANCIES THAT WOULD HAVE BEEN MISSED BY US ASSESSMENT ALONE. PRESENCE OF ONE OR MORE SUSPICIOUS US/USE FEATURES WAS 100% SENSITIVE; TWO OR MORE 88% SENSITIVE AND 77% SPECIFIC. BENIGN TEST RESULTS WERE FOUND IN 57% OF PATIENTS. THE AUTHORS SUGGESTED THAT THE 6.4% REMAINING RISK OF MALIGNANCY, SIMILAR TO THE BENIGN CYTOLOGY CATEGORY, WOULD JUSTIFY FOLLOW-UP INSTEAD OF DIAGNOSTIC HEMITHYROIDECTOMY IN THIS GROUP (223). IN ANOTHER STUDY OF 315 THY3 NODULES, SEMIQUANTITATIVE USE CORRECTLY DIAGNOSED 75% OF THE HISTOPATHOLOGICALLY BENIGN LESIONS THAT WERE CONSIDERED SUSPICIOUS FOR MALIGNANCY ON US, AND 83% OF THE MALIGNANCIES THAT WERE MISDIAGNOSED AS BENIGN ON US (229). THESE RESULTS SUGGEST THAT THE EXISTING TIRADS CLASSIFICATION COULD BE EXTENDED WITH TISSUE ELASTICITY FEATURES. IN UNSELECTED THYROID NODULES, THIS IMPROVED TIRADS SENSITIVITY, BUT NOT SPECIFICITY (250, 294). THE COMBINATION IS A SUITABLE TOPIC FOR FUTURE RESEARCH IN INDETERMINATE THYROID NODULES. MAJOR BENEFIT IS THAT THE TWO TECHNIQUES ARE INDIVIDUALLY INEXPENSIVE AND OBVIOUSLY EASILY COMBINED DURING ONE DIAGNOSTIC PROCEDURE. COST-EFFECTIVENESS CAN BE ANTICIPATED.

US AND MUTATION ANALYSIS


ICC AND MUTATION ANALYSIS
IN HISTOPATHOLOGY SAMPLES, CERTAIN GENETIC ALTERATIONS WERE CORRELATED TO POSITIVE STAINING FOR SPECIFIC IMMUNOMARKERS: PAX8/PPARγ rearrangement was associated with galectin-3 reactivity, and RAS Point mutation with HBME-1 (115). ONLY ONE STUDY INVESTIGATED THIS COMBINATION OF TECHNIQUES IN INDETERMINATE THYROID CYTOLOGY. ALTHOUGH NO SIGNIFICANT CORRELATION WAS DEMONSTRATED BETWEEN POSITIVE BRAFV600E MUTATION AND GALECTIN-3 OVEREXPRESSION, BENEFITING POSSIBLE COMPLEMENTARY USE, NO ADDITIONAL DIAGNOSTIC VALUE WAS DEMONSTRATED EITHER (38).

MICRORNA AND MUTATION ANALYSIS
COMBINED MICRORNA EXPRESSION PROFILING AND MUTATION ANALYSIS COULD ACCURATELY AID DIAGNOSIS AND PROGNOSIS OF THYROID MALIGNANCY. DISTINCT MICRORNAs HAVE BEEN RELATED TO ONCOGENIC MUTATIONS. FOR EXAMPLE, miR-221, miR-222, AND miR-146b WERE MORE OVEREXPRESSED IN BRAF- AND RAS-MUTATED PTC. HIGH EXPRESSION OF miR-187 WAS ASSOCIATED WITH RET/PTC Rearrangement (89, 295). THE FIRST STEP TOWARD DIAGNOSTIC INTEGRATION OF THE TWO TECHNIQUES WAS TAKEN BY LABORIER ET AL. (23), WHO TESTED THE COMMERCIAL DEVELOPED 10-MICRORNA THYROID CLASSIFIER

178 de Koster et al Diagnostics in Indeterminate Thyroid Nodules Endocrine Reviews, April 2018, 39(2):154–191

Downloaded from https://academic.oup.com/edrv/article-abstract/39/2/154/4782710 by Radboud University user on 12 June 2018
ThyraMIR™ simultaneously with the miRInform® thyroid. The ThyraMIR™ was designed to increase the sensitivity of the miRInform® without affecting its specificity. Combined use demonstrated 89% sensitivity and 85% specificity (23). A recent decision analytics model for Bethesda III and IV nodules estimated that combined miRInform® and ThyraMIR™ testing was cost-effective, reducing the rate of unnecessary surgery (diagnostic hemithyroidectomy as well as two-step thyroidectomies) from 88% to 20% and saving $1384 per patient in the first year of treatment or $3170 per avoided surgery. However, it is not described how the economic consequences of the 15% missed malignancies are accounted for in this model (85). The economic as well as medical-ethical consequences of such a high number of missed malignancies question the current clinical utility of this combination of expensive techniques.

In brief, the combined or sequential use of multiple diagnostics in indeterminate thyroid nodules was infrequently studied. Regrettably, the available studies also mostly remained within their own field of expertise: comparing tests either within the domain of pathological (molecular) techniques or within the domain of imaging. Although a sequential combination of a sensitive and an uncorrelated specific test might bring the solution that this clinical issue has been waiting for, the most accurate combination of tests cannot reliably be determined yet.

**Recent Developments and Future Prospects**

**The cancer genome atlas**

Papillary thyroid cancer was one of the cancers targeted by the cancer genome atlas (TCGA) research network, a large collaborative project by the National Cancer Institute and National Human Genome Research Institute. The incentive of the project is to map genomic alterations occurring in 33 types of cancer in 11,000 patients and improve the understanding, classification, and extending possibilities for targeted therapy of these cancers (296). Genetic alterations of all kinds were detected in nearly 500 clinically non-aggressive PTCs (classical, follicular, and tall cell variants) using one proteomic and six genomic platforms. PTC harbored fewer somatic mutations than other human cancer types, but if they were present, driver mutations were detected in the majority of the cancer cells. As expected, the known driver mutations in the MAPK/ERK pathway were dominant, confirming the mutually exclusive relation for BRAF and RAS point mutations and RET/PTC rearrangements. Other detected genetic alterations included genetic variations of the TERT promoter, PI3K and PPARγ pathways, as well as new alterations of known and new drivers, such as EIF1AX, PPM1D, and CHEK2.

Moreover, molecular subtypes of, for example, BRAF-mutated PTC were identified and linked to different clinical subtypes. The role of microRNA in determining cancer phenotype was elaborated, allowing better understanding of clinical behavior of various genetic variants of PTC. Somatic copy number alterations were mostly linked to FVPTC. Ultimately, the TCGA Research Network envisions a reclassification of thyroid carcinoma, abandoning the discrimination between PTC and FTC, and classifying according to molecular subtypes instead of by histopathological subtype first (297). The identified markers may not just have an application in the diagnosis of thyroid carcinoma, but also in better risk-stratification of the different cancers and in targeted therapies. The plurality of applications is best known for the BRAFV600E mutation, which has an association with clinically more aggressive tumor behavior on several fronts. Also, nonthyroid malignancies carrying a BRAF mutation are now (experimentally) treated with RAF inhibitors (298, 299).

There is little doubt that molecular classification systems are the future of oncology diagnostics in all types of human cancers. The position of histopathological assessment is changing, but cannot be renounced. With the current knowledge of thyroid genomics, the need to distinguish the mutated malignant from the mutated benign, premalignant, neoplasms remains, with all due consequences for the surgical and postoperative treatment strategy.

Cytological application of the TCGA set was also already investigated in a recent study. Pagan et al. (300) validated a panel containing the genomic alterations identified by the TCGA in 88 FNAC samples selected from a previous cohort study, including 22 indeterminate thyroid nodules (81, 300). In the latter, 33% sensitivity and 84% specificity were demonstrated. In the same set of patients, Pagan et al. (300) also performed the Afirma® GEC. The GEC yielded less false-negatives and a much higher sensitivity. Even though technical limitations of the applied sequencing techniques could leave RNA transcriptions with low expression levels undetected and thus negatively influence sensitivity of the TCGA set, the scopes of the TCGA and GEC most likely explain their difference in performance. The TCGA was developed using PTC only. It did not include follicular lesions and their distinctive genetic alterations. Moreover, in contrast to the GEC, the TCGA set was not optimized for preoperative diagnostic application in indeterminate thyroid nodules (300). Consequently, the comparison performed by this Veracyte-sponsored study seems unjust: it is obvious that the Afirma® GEC yielded better diagnostic performance in this specific clinical setting. Yet, the results of this study did prove that a large panel of genetic alterations such as the TCGA was not useful in clinical practice without further expansion of the scope of the panel toward follicular tissue testing.

"There is little doubt that molecular classification systems are the future of oncology diagnostics in all types of human cancers."
thyroid neoplasms. Still, the genetic alterations and their relations detected by TCGA are groundbreaking for the progression of research. From these comprehensive sets of biomarkers, we may select new combinations of genetic alterations for future clinical research to develop an accurate rule-in or rule-out molecular test for indeterminate thyroid nodules.

**Proteomics**

Other molecular advances include protein expression diagnostics, or proteomic profiling. These techniques allow for more detailed insight in the molecular biology and protein expression of thyroid neoplasms. For example, matrix-assisted laser desorption ionization/mass spectrometry imaging is able to simultaneously visualize the spatial distribution of proteins and profile up- and downregulated protein expression in relation to the morphological features of the thyroid specimen. These and related proteomic techniques could identify new biomarkers for preoperative cytological diagnosis, but require high levels of expertise. Application to thyroid cytology has so far been investigated by few studies (301, 302). Ex-vivo cytology studies show accurate and reproducible differentiation between various lesions, including the currently difficult to diagnose Hurthle cell neoplasms (302). No studies investigated the diagnostic value of proteomics in in-vivo indeterminate thyroid cytology yet.

**Discussion**

This review provides a comprehensive overview of the available literature on molecular and imaging biomarkers as additional diagnostics for thyroid nodules with indeterminate cytology (Bethesda III and IV) and their application in a clinical preoperative setting. Clinical utility requires more from a diagnostic than mere well-validated test performance and high rule-in or rule-out capacity. The 2015 ATA guidelines suggested that the ideal rule-out diagnostic for thyroid carcinoma should have an NPV similar to a benign cytological diagnosis (~96.3%) and the ideal rule-in test a PPV that is at least similar to a malignant cytological diagnosis (~98.6%) (8, 11). The balance between test sensitivity and specificity, and their prevalence-dependent derivatives PPV and NPV, directly reflects on feasibility and cost-effectiveness estimates. A diagnostic with (near) perfect sensitivity but limited specificity is inefficient and unlikely cost-effective: the NPV will be close to 100%, but the majority of nodules will test positive. Therefore, instead of focusing on the reproducible highest sensitivity or specificity, a diagnostic is better appreciated by end points such as desired minimal rates of accurately prevented beneficial surgeries or accurately diagnosed carcinomas. More importantly, clinical utility demands that implementation of the ancillary test leads to changes in patient management and overall health benefits (303). All these requirements directly depend on a plurality of epidemiological and economic factors within the tested population, such as the local test availability, professional expertise, and case mix—prevalence of malignancy as well as the balance of various subtypes of indeterminate cytology including especially Hurthle-cell neoplasms and BRAF-mutation. Additionally, clinical utility considerations should include less tangible factors such as physician and patient preference, multidisciplinary decision-making, and compatibility with everyday clinical routine and logistics in endocrine practice. All things considered, global perspectives regarding the preferred diagnostic for indeterminate thyroid nodules likely greatly differ.

**Recommendation for clinical use of rule-out tests**

The most accurate currently available rule-out tests are the Afirma® GEC and FDG-PET/CT imaging. The Afirma® GEC had strikingly high sensitivity in nearly all studies (78, 79, 131, 133–135). However, there are concerns regarding the lack of strong validation studies. With a high degree of missing histology, especially in GEC negative nodules, there is a potentially strong diminution of the tests’ sensitivity if unrecorded GEC-negative lesions were less often benign than presumed. In the United States, physicians should locally validate the tests’ utility prior to implementation. However, with its limited global availability, high costs and low probability of cost-effectiveness, clinical implementation of the Afirma® GEC outside the United States is currently not favored (15, 73, 82–85).

FDG-PET/CT may be the preferred rule-out test for indeterminate thyroid nodules in a European setting. With sufficient validation studies with complete histopathological follow-up, it demonstrated consistent high sensitivity and a benign test result in 40% of the patients, although the number of currently published patients is moderate. Cost-effectiveness of FDG-PET over other diagnostics is presumed (15). Its popularity in the United States is more limited, although the efficacy of this molecular imaging technique could likely compete with molecular biomarkers panels, even if the costs per scan are somewhat higher than in Europe. The main drawback of FDG-PET/CT is its, admitted minor, risk to the patient by using a limited dose of ionizing radiation.

The recently announced version 3 of the ThyroSeq® may become a prime contender. Dependent on the case mix, the ThyroSeq® v2.1 anticipated high negative predictive value (22). However, the number of studies to confirm test performance and clinical utility in different patient populations is limited. Clinical results for the ThyroSeq® v3 are eagerly awaited.

Semiquantitative elastosonography could be a suitable alternative, in particular in case a more economic test is required. However, overfitting and
lack of external cut-off validation likely overestimated the performance of this technique in the limited number of available studies. If future prospective studies can confirm its performance and thresholds of this operator-dependent but globally accessible method, USE could become a more important diagnostic in this field.

None of the diagnostic techniques under investigation in this review has a perfect NPV or fulfills the threshold proposed by the ATA. A number of malignant nodules will be misdiagnosed as benign on first assessment. Considering the typical indolent clinical course of differentiated thyroid cancer, follow-up of these initially false-negative nodules will most likely still result in timely diagnosis without relevant treatment delay and dismal prognostic consequences.

Recommendation for clinical use of rule-in tests
The best rule-in performance was unmistakably demonstrated by \(BRAF\) mutation analysis, which showed perfect 100% specificity in an abundance of studies. Yet, strong regional differences in prevalence of \(BRAF\) mutations have a major impact on its clinical utility, especially when comparing South Korea to other countries. Moreover, the analysis most likely has very low yield in Bethesda IV nodules, in which the mostly follicular type malignancies are more frequently \(RAS\)-mutated (30, 44, 61, 67). Testing for individual genetic alterations other than the \(BRAFV_{600E}\) point mutation is not useful. In American and European settings, a gene mutation panel is likely preferred over any individual mutation analysis.

Promising rule-in capacity was also demonstrated for Galectin-3 ICC, an infrequently applied technique with limited validation studies. Further prospective studies are warranted to validate its performance in indeterminate thyroid nodules and endorse its possible clinical use.

Besides \(BRAF\) mutation analysis, none of diagnostics meet the 2015 ATA requirements of an ideal rule-in test. Compared with ruling-out tests, ruling-in tests face an additional challenge. With a generally low frequency of thyroid carcinoma in indeterminate thyroid nodules, achieving a reliable PPV (higher than 95%) can be a major challenge despite adequate test specificity. Such high demands to a ruling-in test advocate the use of a ruling-out test in populations with a limited pretest probability of malignancy.

Clinical recommendation for a step-wise approach
Most of the diagnostic modalities are optimized for either ruling in or ruling out malignancy.

No single diagnostic addressed in the current review currently has it all: both a near-perfect sensitivity and a near-perfect specificity, and (proven) cost-effectiveness. It is extremely challenging to develop such test performance parameters in a single diagnostic. Even promising new diagnostics, such as the ThyroSeq\textsuperscript{Tm} and ThyraMir\textsuperscript{Tm}, require substantial further optimization to get near this diagnostic utopia.

With the diagnostics currently available in the clinical setting, a multimodality stepwise approach could offer a conclusive diagnosis for indeterminate thyroid nodules, sequentially combining one sensitive rule-out and one specific rule-in test. Unfortunately, thus far few studies investigated this approach (20, 61). Combinations of (molecular) imaging and somatic genetics were especially scarce. There is currently insufficient evidence to accommodate reliable interpretation of sequentially used tests, as performance of the second test is unknown in a population preselected by the first. Besides choosing two accurate and uncorrelated tests to achieve maximum diagnostic accuracy, the sequence of testing, local availability, and costs of the selected diagnostics are crucial. Costs of two or more additional tests may compromise cost-utility estimates. Available cost-effectiveness studies for individual diagnostic modalities were additionally greatly susceptible to global variations in population-dependent factors such as pretest probability of thyroid carcinoma and local test performance, and varying health care costs including the surgical reimbursement rates (15, 73, 83, 84). Reported surgical and hospitalization costs range from $4,628 to $65,49 for hemithyroidectomy, $5,472 to $7068 for completion thyroidectomy, and $5,680 to $11,265 for initial total thyroidectomy. Secondary expenses following surgery should be considered as well, including postoperative observation, thyroid hormone replacement (approximately $150 per patient per year), treatment of hypoparathyroidism (approximately $860 per patient per year), and resolution of rare but potentially serious surgical complications (15, 72, 83, 270). Secondary end points such as quality of life and survival are of minor importance to cost-effectiveness, due to the generally indolent course of differentiated thyroid cancer, adequate treatment options, and overall low disease-related mortality (15, 83, 84).

Recent discussions in thyroid histopathology
Histopathology is classically based on microscopic assessment of tumor phenotype, aided by IHC.

Histopathology is classically based on microscopic assessment of tumor phenotype, aided by IHC. However, this "gold standard test" is also subject to advancing insights regarding tumor phenotype, increasingly aided by knowledge regarding tumor genotype. Mutation-negative malignancies resulting from indeterminate cytology were frequently identified as encapsulated follicular variants of papillary thyroid carcinoma without histologic features of aggressive behavior (22, 23, 25, 30, 74). Also, several studies defined a separate intermediate histopathological category called "(follicular) tumor of uncertain malignant potential" for encapsulated, well-differentiated follicular tumors with questionable PTC-type nuclear changes (62, 156, 169, 184). These examples illustrate one of the important ongoing discussions in thyroid histopathology. In 2016, Nikiforov et al. (304) proposed an official downsampling of the classification of
proven noninvasive encapsulated FVPTCs, renaming them "noninvasive follicular neoplasm with papillary-like nuclear features" (NIFTP). The behavior of these neoplasms is benign unlike other thyroid carcinoma subtypes, showing no evidence of recurrent disease after a median 13-year follow-up. Approximately one in four of the neoplasms in the retrospective cohort were mutated, most frequently carrying RAS (NRAS) or PAX8/PPARγ alterations. Presence of a mutation likely predisposes the NIFTP to progress into an invasive encapsulated FVPTC, justifying surgical resection. Treatment of NIFTP should most likely be limited to hemithyroidectomy, waiving totalizing thyroidectomy, and radioiodine ablation (304). Although revolutionizing, this new nomenclature complicates mutation-based preoperative decision-making (22, 30, 74). The justification to skip two-stage surgery and perform a total thyroidectomy at once for mutation-positive nodules is the driving force of the seven-gene mutation panel and similar tests, but would be overkill for the subgroup of NIFTP (30). Nonetheless, most of the undesirable possible overtreatment for NIFTP is likely resolved if RAS-mutated indeterminate nodules are treated with hemi- instead of total thyroidectomy, as previously suggested. No comprehensive diagnostic test is currently available to diagnose mutation-positive NIFTP preoperatively, as follicular tumor invasiveness and encapsulation cannot be distinguished on cytology.

**Hürthle cell cytology**

The Achilles heel of many diagnostics investigated in this review is cytology suspicious for a Hürthle cell neoplasm (Bethesda IV SHCN/HCN). Hürthle cells are oxyphilic cells with abundant cytoplasm and an enlarged nucleus with a prominent nucleolus. They are found in benign thyroid diseases such as Hashimoto’s thyroiditis, but also occur in the notorious Hürthle cell adenoma and carcinoma, the oncocytic variant of FA and carcinoma (4, 198). Although Hürthle cell carcinomas (FTC-OV) are rare, their aberrant clinical course and association with invasive features justifies the special attention given to Hürthle cell cytology by the Bethesda and other classification systems. An accurate additional diagnostic is desired. Disappointingly, several studies concluded that the investigated test was accurate in all except Hürthle cell lesions (61, 134, 177, 184). ICC handed some solutions, although promising results of combined galectin-3 and CK-19 staining have not yet been validated (166). Besides that, BRAF, RAS, RET/PTC, or PAX8/PPARγ alterations are only occasionally found (61, 67). These findings support previous presumptions that oncocytic thyroid nodules are a completely separate entity with a unique molecular and phenotypic profile (305–308). Malignant transition in Hürthle cell nodules most likely involves the PIK3CA-Akt-mTOR and Wnt/beta-catenin pathways rather than the MAPK/ERK pathway (305, 308). Rare TP53 mutations, usually associated with poorly differentiated and anaplastic carcinoma, were recently also identified in well-differentiated Hürthle cell nodules (306). Also, recurrent FTC-OV have shown genome haploïdisation, a rare phenomenon in other types of differentiated thyroid carcinoma (309). Specific markers for the preoperative molecular differentiation of Hürthle cell nodules should be developed. Adaptation of existing tests to additionally suit Hürthle cell nodules (e.g., the Afirmac® GEC) is a strategy being explored, for example, by the ThyroSeq® v3 and the Afirmac® Gene Sequence Classifier. Caution should be taken that these adaptations do not decrease the diagnostic accuracy for nononcocytic lesions. MicroRNA expression profiling of these lesions is currently also under investigation (89, 146).

**Strengths and limitations of the current review**

There are several important strengths and limitations to this comprehensive review. This review provides a complete overview of the available additional diagnostics for indeterminate thyroid nodules, resulting from a careful and systematic literature selection and quality appraisal. Different types of clinical data of various levels of evidence were considerably presented. Nonetheless, this review is generally prone to inaccuracies from low study quality, study heterogeneity, and different types of bias. For some of the assessed diagnostics, the limited number of available publications and small study cohorts contribute to heterogeneity of data and loss of applicability. This mainly concerns studies on nonroutine imaging techniques. By nature, these clinical studies need to prospectively include subjects to voluntarily undergo an extra investigation with, at least in the clinical validation phase, no implications for individual patient management. These types of studies require more resources than "further use" tissue biobank studies. Consequently, the number of studies is more limited and published series often are small. In contrast, cytological biomarker research gratefully profited from available large tissue biobanks for initial validation studies. We believe consistent results from properly designed imaging studies should not be disregarded due to mere their sample size, but be appreciated by the quality of their study design and statistics.

Population-level study differences were often observed, not only related to test performance but also strongly varying malignancy rates that were oftentimes much lower or higher than expected from indeterminate thyroid nodules. Besides insuperable epidemiological variations, the selection of indeterminate cytology, and the retrospective nature of many studies may have contributed to these discrepancies.

The type of indeterminate cytology included by individual studies varied, likely leading to between-study heterogeneity. Besides global variations and
known intra- and interobserver discordance, diverse definitions of indeterminate cytology were adhered (5). Nowadays, the Bethesda system differentiates indeterminate from benign and suspicious cytology in a more standardized manner in both literature and clinic. Bethesda III and/or IV and similar categories from other classification systems were frequently applied. Unfortunately, some studies also included small numbers of Bethesda V nodules without presenting results for individual categories separately (131). Many other studies adhered to their own definition of indeterminate cytology. This especially, but not exclusively, concerns studies published before the introduction of the Bethesda system in 2009.

Retrospective study designs and subsequent selection bias—only including indeterminate thyroid nodules that had undergone both thyroid surgery and (routine) preoperative testing—likely also caused overestimation of the true efficacy of certain techniques (e.g., BRAF mutation analysis or US).

Conclusion and Recommendations

In current-day practice, there are numerous additional diagnostics available to further assess thyroid nodules with indeterminate cytology, all with advantages and disadvantages. This review provided a comprehensive overview of the available literature on these techniques, addressing both molecular and imaging biomarkers, aiming to provide an objective and nuanced comparison of their performance and cost-effectiveness with regard to rightful surgical decision-making. Many of these diagnostics have either an adequate rule-in or rule-out capacity, but no single currently available test seems to serve both purposes well. Diagnostics from the different research fields likely complement each other in a multimodality stepwise diagnostic approach toward. Notwithstanding, test performance is always population-dependent. To correctly interpret the results, the prevalence of malignancy and the performance, costs, and feasibility of the desired diagnostic in the local patient population should be known beforehand. Local implementation studies are strongly recommended to confirm clinical utility. Most importantly, the local decision favoring or opposing a certain diagnostic should be a deliberate and multidisciplinary one. Cooperation between clinical endocrinologists, endocrine surgeons, pathologists, radiologists, and nuclear medicine physicians is crucial.

References


https://academic.oop.com/edrv
REVIEW

24. Kim SK, Hwang TS, Yoo YB, Han HS, Kim DL, Song

21. Munoz Perez N, Villar del Moral JM, Muros Fuentes

184


228.

68.


51.


25. Seethala RR, Tublin ME, Yip L, Nikiforova MN.


664.


117.


160. Mohey N, Hassan TA, Abdel-Baki S. Role of combined grey scale US and US tissue elastography in...


178. Wale A, Milea KA, Young B, Zammit C, Williams A, Quin J, Didacre S. Combined (99mTc)-methoxysybenzylisonitrile scintigraphy and fine-needle aspiration cytology o...
nODULES WITH INDETERMINATE CYTOLoGY. \textit{Ann Surg Oncol.} 2010;17(8):2147–2155.


Acknowledgments

Funding Support: EJ de Koster receives a research grant from the Dutch Cancer Society (DCS).

Correspondence and Reprint Requests: EJ de Koster, MD, Radboudumc, Department of Radiology and Nuclear Medicine, PO Box 9101, 6500 HB Nijmegen, the Netherlands. E-mail: lisanne.dekoster@radboudumc.nl.

Disclosure Summary: The authors have nothing to disclose.

Abbreviations

ADC, apparent diffusion coefficient; ATA, American Thyroid Association; AUS/FLUS, atypia of undetermined significance or follicular lesion of undetermined significance; CK-19, cytokeratin 19; CT, computed tomography; DW-MRI, diffusion-weighted magnetic resonance imaging; ERK, extracellular signal-regulated kinase; FA, follicular adenoma; FDG-PET, [18F]-2-fluoro-2-deoxy-D-glucose positron emission tomography; FNAC, fine needle aspiration cytology; FTC, follicular thyroid carcinoma; FTC-OV, oncocyotic variant of follicular thyroid carcinoma; FVPTC, follicular variant of papillary thyroid carcinoma; GEC, gene expression classifier; HBME-1, Hector Battifora mesothelial-1; HRM, high-resolution melting; hTERT, human telomerase reverse transcriptase; ICC, immunocytochemistry; IHC, immunohistochemistry; MAPK, mitogen-activated protein kinase; mPTC, papillary microcarcinoma; mRNA, messenger RNA; NGS, next generation sequencing; NIFTP, noninvasive follicular thyroid neoplasm with papillary-like nuclear features; PET, positron emission tomography; RAS, retrovirus-associated DNA sequences; ROC, receiver operating characteristic; RT-PCR, reverse transcription polymerase chain reaction; SFN/FN, (suspicious for a) follicular neoplasm; SHCN/HCN, (suspicious for a) Hürthle cell neoplasm; TCGA, the cancer genome atlas; TIRADS, Thyroid Imaging Reporting and Data System; TPO, thyroid peroxidase; TRK, tyrosine receptor kinase; US, ultrasound; USE, ultrasound elastography.