Molecular Biomarkers in Thyroid FNA Samples

Rosa Marina Melillo and Massimo Santoro

Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università di Napoli Federico II, 80131 Naples, Italy

In this issue of the JCEM, Walsh et al. (1) report the analytical validation of a molecular diagnostic method for indeterminate fine-needle aspiration (FNA) thyroid samples. Together with previous reports describing clinical performance and cost effectiveness of this method, this study highlights the capability of such a molecular test to reduce unnecessary surgery for benign thyroid diseases (2–5).

Although rarely malignant, thyroid nodules are very common. FNA cytology is routinely used to assess whether thyroid nodules are benign or malignant (6). Four major types of thyroid malignancy are recognized, including papillary (PTC), follicular (FTC), anaplastic, and medullary carcinomas (7). PTC, particularly follicular-variant PTC, and FTC can be difficult to differentiate with respect to benign thyroid nodules (7). According to the National Cancer Institute, FNA results stratify thyroid nodules as benign, malignant, nondiagnostic (when the aspirate contains scant cellularity), and indeterminate. Indeterminate lesions (nearly one third of the cases) represent the most challenging conditions and feature an overall risk of 20–30% of being malignant. Indeterminate lesions include: 1) atypia of undetermined significance/follicular lesions of undetermined significance (with a malignancy risk of 5–15%); 2) follicular or Hurthle/suspicious for follicular or Hurthle neoplasms (with a malignancy risk of 15–30%); and 3) suspicious for malignancy (with a malignancy risk of 60–75%) (8). Commonly, patients with an indeterminate result undergo diagnostic surgical resection of the thyroid, with the risk of performing unneeded operations if a lesion is revealed to be benign (9). Therefore, methods are urgently needed to guide clinical decisions, in adjunct to cytology, in the case of indeterminate FNA samples, to reduce surgery-associated morbidity and additional costs.

Understanding molecular pathogenesis of thyroid nodules has great potential to discover novel genetic biomarkers that might guide the decision for surgery. Recently, two approaches based on nucleic acids determination have been proposed to improve FNA accuracy: 1) detection of cancer-driving oncogene mutations to identify cancer; and 2) gene expression profiling, aimed at exploiting transcriptional reprogramming of malignant vs. benign cells to identify benign lesions.

Nikiforov et al. (10–12) first proposed systematically applying oncogene mutation detection to molecularly classify FNA samples; this approach was further pursued by other investigators (13–15). Oncogene detection was proved feasible also in routine air-dried FNA smears (16). Thus, FNAB determination of BRAF V600E mutation and RET/PTC (RET/PTC1 and RET/PTC3) rearrangements (commonly associated to PTC), RAS (HRAS codon 61, KRAS codons 12/13, and NRAS codon 61) mutations (commonly associated to follicular-variant PTC and FTC), and PAX8/PPAR\gamma fusion (commonly associated to FTC) was shown to improve diagnostic accuracy and enable cancer identification. One important caveat is that RAS mutations (20–40% of the cases) and PAX8/PPAR\gamma rearrangement (2–10% of the cases) can also be associated to benign adenomas. Thus, the presence of RAS and PAX8/PPAR\gamma mutations in adenomas may generate false-positive results (7). However, adenomas carrying either RAS or PAX8/PPAR\gamma genetic alterations are considered precancerous lesions (11, 14), and precautional thyroidectomy in these patients would be a reasonable choice. In the largest prospective trial so far reported, mutation-positive lesions with indeterminate cytology (i.e. atypia of undetermined significance/follicular lesion of undetermined significance, follicular neoplasm/suspicious for a follicular neoplasm, and suspicious for malignancy) re-

Abbreviations: FNA, Fine-needle aspiration; MTC, medullary thyroid carcinoma; NPV, negative predictive value; PPV, positive predictive value; PTC, papillary thyroid carcinoma.
revealed malignancy in 88% (sensitivity, 63%), 87% (sensitivity, 57%), and 95% (sensitivity, 68%) of the cases on final pathology. Mutation-negative lesions were cancers in 6, 14, and 28% of the cases, respectively (11). Thus, a positive result at this test quite definitely indicates malignancy [positive predictive value (PPV), 87–95%]. Consequently, this test can be viewed as useful to identify patients likely harboring a malignant disease. In contrast, a negative result indicates that, although low, the probability of cancer is still significant (up to 30%) (11). One possible explanation for the false-negative results is that as many as 30–40% of thyroid carcinomas do not display known somatic oncogene mutations and may harbor novel genetic alterations. An alternative explanation is the lack of accuracy of methodological tests due either to technical limitations or to the dilution of cancer cells into analyzed samples.

The second approach, i.e. gene expression profiling, has been conceived to prospectively identify FNA lesions that are truly benign and, therefore, to reduce the number of unneeded thyroidectomies. Thus, this approach should be applied to increase the negative predictive value (NPV) of FNA. To this aim, Chudova et al. (2) developed a molecular classifier based upon commercial (Affymetrix, Santa Clara, CA) gene chips that achieved 96% NPV and 84% specificity. The molecular classifier was developed by training a large set of prospectively collected FNA samples (including prevalent and rare malignant and benign thyroid lesions) (2). FNAs can contain variable fractions of malignant cells mixed with normal cell types, mostly lymphocytes. By estimating the proportion of thyrocyte markers (including DIO1, DIO2, EGFR, KRT19, KRT7, MUC1, TG, and TPO) and lymphocyte markers (including CD3, FOXP3, and others) in FNAs, the presence of contaminating lymphocytes was found highly variable. Nevertheless, by performing in silico simulations, the Gene Expression Classifier (GEC) robustly identified PTCs even in the presence of an 80% dilution with benign cells (2). Based upon these results, a GEC called Afirma was developed by Veracyte, Inc. (South San Francisco, CA) and used in subsequent studies (1, 3, 4). The GEC featured 167 different genes (142 genes in the main classifier and 25 genes used to filter out rare neoplasms of thyroid, medullary and Hurthle, and metastatic tumors of nonthyroid origin). As suggested by previous functional studies, some of these genes, besides representing diagnostic markers, may contribute to cancer cell malignant features (17–19).

In a prospective, multicenter, blinded study, the Afirma test NPV resulted 95% for thyroid nodules classified as atypia of undetermined significance/follicular lesion of undetermined significance, 94% for those classified as follicular neoplasms, and 85% for a lesion suspicious for cancer (3). These values met requirements of a test designed to identify benign samples. However, up to 15% of samples classified as benign, particularly in the case of suspicious for malignancy lesions, were false negative. This erroneous interpretation was probably due to a paucity of thyroid follicular cells in the sample (3, 20). Perhaps the oncogene mutation tests may help with correctly classifying these false-negative samples by identifying those at high risk of being malignant.

These results were encouraging, but a clinical implementation will require a test that is robust and reproducible in multiple laboratories. Thus, in this issue of the JCEM, Walsh et al. (1) have analyzed reliability of the Afirma test with respect to specific caveats such as RNA stability, benign/malignant mixture, effects of blood contamination, and results reproducibility. Their data demonstrate that RNA is stable for up to 6 d; that although 15 ng RNA are needed, the RNA input can be scaled down to as little as 5 ng; and that malignant cells can be diluted down to 20% without a loss of test reliability (1). Dilution still remains a concern, as in the previous clinical study (3), because false-negative calls were often due to the paucity of thyrocytes in the sample. As also shown in the case of oncogene mutation testing, expression of epithelial (like cytokeratin-7, KRT7) with respect to house-keeping (like glyceraldehyde 3-phosphate dehydrogenase) markers may be worth implementing to determine sample adequacy in terms of thyrocyte density before a test (11). One further concern is that blood contamination of the sample may determine false-positive suspicious calls (1); this might be corrected by predetermining blood content (as an example, by measuring hemoglobin concentration) before subjecting FNA to the RNA processing. Finally, another concern is that the assay is not performed on the same aspirate that is assessed cytologically, but rather on a sequential needle pass; this may create inconsistencies between molecular and cytology results. Thus, it might be worth checking whether the test may be performed on routine air-dried FNA material that is also used for morphological evaluation (16).

Despite these concerns, the article published by Walsh et al. (1) represents another step toward the molecular diagnosis of thyroid nodules. The goal of effectively differentiating benign from malignant thyroid nodules will probably at the end be achieved by exploiting some crucial research tracks, such as: 1) combination of tests based on cancer-specific mutations and on expression classifiers, particularly (see also point 4) if the complement of mutations associated to thyroid cancer will be rendered more exhaustive by next generation sequencing (indeed, with a PPV of 87–95%), the mutation assessment test may serve
best as a diagnostic algorithm to identify suspected malignancy, whereas, with a negative predictive value of up to 95%, the Afirma test may serve to exclude malignancy (20); 2) identification of additional expression classifiers including regulatory RNAs (in particular microRNAs), because several studies have identified microRNAs differentially expressed between cancerous and benign thyroid lesions (21–24); 3) design of gene-specific hypermethylation classifiers, as proved feasible in the differentiation of melanomas from nevi (25); and 4) perform systematic next-generation sequencing analysis of malignant and benign thyroid lesions will enable implementing the already available molecular markers.

### Acknowledgments

Address all correspondence and requests for reprints to: Massimo Santoro, M.D., Ph.D., Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università di Napoli Federico II, via S. Pansini 5, 80131 Naples, Italy. E-mail: masantor@unina.it.

Disclosure Summary: The authors have no conflict of interest to disclose.

### References


