

A benign genomic classifier for thyroid nodules with indeterminate cytology: a critical appraisal

Sam M. Wiseman

Department of Surgery, St. Paul's Hospital & University of British Columbia, Canada

Corresponding to: Sam M. Wiseman, MD, FRCSC, FACS. Department of Surgery, St. Paul's Hospital, University of British Columbia C303 - 1081 Burrard Street, Vancouver, British Columbia, Canada, V6Z 1Y6. Email: smwiseman@providencehealth.bc.ca.



Submitted Jul 20, 2012. Accepted for publication Aug 21, 2012.

DOI: 10.3978/j.issn.2227-684X.2012.08.04

Scan to your mobile device or view this article at: <http://www.glandsurgery.org/article/view/1002/1193>

Alexander and colleagues recently published, in the *New England Journal of Medicine*, a 19 month prospective multicenter genomic classifier validation study for benign thyroid nodules with indeterminate cytology, involving 49 clinical sites, 3,789 patients, and 4,812 fine needle aspiration biopsy (FNAB) specimens. This study was sponsored by Veracyte Incorporated (Veracyte Inc. South San Francisco California U.S.A.) (1). The development of the FNAB-trained genomic classifier utilized was described in a prior Veracyte sponsored study published 2 years earlier, and was composed of 167 genes (142 genes in the main classifier and 25 genes to filter out rare neoplasms) (2). FNAB were sent to Veracyte for genomic analysis that was carried out on custom-built arrays, and performance on these custom arrays was validated with data from the Affymetrix Human Exon ST 1.0 arrays (the genomic profiling platform upon which the classifier was originally developed) (Affymetrix, Santa Clara California U.S.A.). A total of 367 FNABs (47 benign, 55 cancer, and 265 indeterminate (129 atypia or follicular lesion of undetermined significance, 81 follicular or Hurthle cell neoplasm, 55 suspicious for malignancy) were evaluated by the genomic classifier. For the 312 cases that had genomic data and reference standard data available their genomic classifier had a sensitivity of 87% (79-93%) and a specificity of 53% (46-60%), PPV 47% (39-54%), and NPV 90% (83-94%) for diagnosing thyroid lesions as “suspicious” for cancer. When the 265 indeterminate FNABs are considered alone, the classifier correctly identified 78 of 85 cancers as being “suspicious” for cancer and 93 of 180 benign lesions as being benign. Thus, in the indeterminate FNAB group the classifier had a sensitivity of 92% (84-97%), specificity of

52% (44-59%), a PPV of 47% (40-55%), and a NPV of 93% (86-97%). The reasons for the observed reduction in their genomic classifier specificity, when compared to their prior study (specificity 83.9%), was not elucidated upon by the authors. These investigators concluded that a more conservative, or a less surgically aggressive, management approach should be considered for most individuals who have indeterminate FNAB results and a benign genomic classifier diagnosis.

Based upon a sensitivity of 92%, there will be 8% of FNABs diagnosed as benign by the genomic classifier that will actually be cancer (false negatives). In their report the authors showed that 6/7 false negatives were due to inadequate sampling of thyroid nodules. It is not clear whether this is an inherent limitation of the FNAB based genomic test or if changes in their operating procedures could reduce this number. Based upon a test specificity of 52%, there will be 48% of patients that will be incorrectly diagnosed as “suspicious” for cancer (false positive) when they actually have benign disease. Thus, the high sensitivity (92%) and low specificity (52%) could lead to close to half of individuals with an indeterminate FNAB diagnosis, being given a false positive “suspicious” for cancer diagnosis. Therefore, a key question is whether a “suspicious” for cancer genomic classifier diagnosis lead to an increase in thyroid operations? Another major drawback of the genomic classifier is that its low positive predictive value (47%) does not allow surgeons to have confidence in tailoring their operative approach (i.e. total thyroidectomy and central neck dissection) for cancer. As well, will individuals classified as being “suspicious” for cancer by a genomic classifier, with a 47% PPV, undergo inappropriately aggressive thyroid

surgery and central neck lymph node dissection?

In another recent study, also sponsored by Veracyte Inc., Duick *et al.* reported on how a benign diagnosis from their genomic classifier (now named Afirma) influenced the decision of the endocrinologists and patients to proceed with a thyroid operation (3). Their genomic classifier is now termed the 'Afirma Gene Expression Classifier' (AGEC) and is described by these authors as a proprietary diagnostic test developed by Veracyte that is offered through a sole source, Clinical Laboratory Improvement Amendments (CLIA)-certified reference laboratory. As mentioned above the genomic classifier test classifies thyroid nodules diagnosed as being indeterminate by cytology as either benign (NPV 93%) or "suspicious" for cancer (PPV 47%). This study was carried out through survey of 51 endocrinologists at 21 practice sites that had requested ≥ 3 molecular classifier tests from Veracyte. They found that the historical operative rate of 74% for cytologically indeterminate nodules fell to 7.6% after the molecular classifier was adopted into their practice ($P < 0.001$). The rate of surgery on cytologically indeterminate nodules that were diagnosed as benign by the genomic classifier did not differ from the historically reported rate of operation on benign thyroid nodules ($P = 0.41$). In this report the influence of a molecular classifier test with a "suspicious" for cancer diagnosis was not specifically evaluated, though the authors commented that the cost savings from not operating on the genomic classifier 'benign' patients may partially offset the costs of any possible increase in the rate of operation for genomic classifier diagnosed "suspicious" for cancer patients. This comment is worrisome, and the impact of their molecular classifier "suspicious" for malignancy diagnosis on rates of thyroid surgery and costs, given the genomic classifier specificity being 53%, is concerning and warrants further study.

Overall, the report by Alexander *et al.* that describes a multicenter clinical study evaluating a benign thyroid tumor genomic classifier for lesions with an indeterminate FNAB diagnosis, despite its limitations, is exciting not only because it serves to validate a diagnostic test with a high NPV, but because it is one of the few published studies that demonstrates the feasibility of conducting a multicenter thyroid cancer diagnostic molecular marker trial. While ultimately the clinical acceptance, implementation, and economic impact of this or some other thyroid cancer molecular diagnostic test are yet to be determined, it seems likely that such tests for thyroid cancer are here to stay, and over time will likely become important adjuncts to traditional thyroid cytomorphology.

Acknowledgements

Disclosure: The author declares no conflict of interest.

References

1. Alexander EK, Kennedy GC, Baloch ZW, et al. Preoperative Diagnosis of Benign Thyroid Nodules with Indeterminate Cytology. *N Engl J Med* 2012;367:705-15.
2. Chudova D, Wilde JI, Wang ET, et al. Molecular classification of thyroid nodules using high-dimensionality genomic data. *J Clin Endocrinol Metab* 2010;95:5296-304.
3. Duick DS, Klopper JP, Diggans JC, et al. The Impact of Benign Gene Expression Classifier Test Results on the Endocrinologist-Patient Decision to Operate on Patients with Thyroid Nodules with Indeterminate Fine-Needle Aspiration Cytopathology. *Thyroid* 2012. [Epub ahead of print].

Cite this article as: Sam M. Wiseman. A benign genomic classifier for thyroid nodules with indeterminate cytology: a critical appraisal. *Gland Surg* 2012;1(2):87-88. DOI: 10.3978/j.issn.2227-684X.2012.08.04