Digital gene expression analysis might aid in the diagnosis of thyroid cancer

H. Armanious MD PhD,*† B. Adam MD,*† D. Meunier MD,*† K. Formenti BSc,† and I. Izevbaye MBBS PhD*†

ABSTRACT

Background Thyroid cancer represents approximately 90% of endocrine cancers. Difficulties in diagnosis and low inter-observer agreement are sometimes encountered, especially in the distinction between the follicular variant of papillary thyroid carcinoma (FVPTC) and other follicular-patterned lesions, and can present significant challenges. In the present proof-of-concept study, we report a gene-expression assay using NanoString nCounter technology (NanoString Technologies, Seattle, WA, U.S.A.) that might aid in the differential diagnosis of thyroid neoplasms based on gene-expression signatures.

Methods Our cohort included 29 patients with classical papillary thyroid carcinoma (PTC), 13 patients with FVPTC, 14 patients with follicular thyroid carcinoma (FTC), 14 patients with follicular adenoma (FA), and 14 patients without any abnormality. We developed a 3-step classifier that shows good correlation with the pathologic diagnosis of various thyroid neoplasms. Step 1 differentiates normal from abnormal thyroid tissue; step 2 differentiates benign from malignant lesions; and step 3 differentiates the common malignant entities PTC, FTC, and FVPTC.

Results Using our 3-step classifier approach based on selected genes, we developed an algorithm that attempts to differentiate thyroid lesions with varying levels of sensitivity and specificity. Three genes—namely SDC4, PLCD3, and NECTIN4/PVRL4—were the most informative in distinguishing normal from abnormal tissue with a sensitivity and a specificity of 100%. One gene, SDC4, was important for differentiating benign from malignant lesions with a sensitivity of 89% and a specificity of 92%. Various combinations of genes were required to classify specific thyroid neoplasms.

Conclusions This preliminary proof-of-concept study suggests a role for nCounter technology, a digital gene expression analysis technique, as an adjunct assay for the molecular diagnosis of thyroid neoplasms.

Key Words Thyroid neoplasms, papillary thyroid carcinoma, follicular neoplasms, gene expression profiling, NanoString

INTRODUCTION

Thyroid cancer represents approximately 90% of endocrine cancers. Thyroid cancers are classified into distinct types, including papillary thyroid carcinoma (PTC), follicular variant of papillary thyroid carcinoma (FVPTC), follicular adenoma (FA), and follicular thyroid carcinoma (FTC). The most common of these lesions is PTC. Surgical pathology diagnosis of PTC, largely dependent on nuclear features, is usually straightforward; nevertheless, those cytologic criteria are subjective, with no uniform minimum diagnostic criteria.

Diagnostic dilemmas are sometimes encountered in which the distinction between FVPTC and other follicular-patterned lesions presents challenges. Occasionally, nuclear features are not sufficiently developed for an unequivocal diagnosis of PTC. Difficulties also attend the differentiation of subtypes of follicular lesions, including separating reactive lesions such as adenomatous goiter and Hashimoto thyroiditis from neoplastic conditions such as FA. Given the differences in prognosis and management for those entities, accurate diagnosis is imperative.

The diagnostic challenges encountered in the assessment of thyroid lesions are demonstrated by significant
inter-observer variability, with observer agreement being as low as 58% in some studies. Factors implicated by such studies include the interpretive significance of microfoci-
ules and capillaries within tumour capsules, variations in the assessment of nuclear features such as nuclear clearing for PTC, and lack of strict criteria for distinguishing adenomas
matous goiter and FA.

Various attempts to address these common diagnostic
challenges have been made. Past approaches have largely
been confined to the use of various immunohistochemical
stains, singly and in combination. Numerous antibodies
have been explored—including those against HMB1-1,
CK19, galectin-3, galectin-1, CITED1, CD44, Trop-2, CD56,
and TPO—with varying degrees of success. Previous
investigators have also assessed the utility of immuno-
histochemistry for confirming malignancy, identifying
vascular invasion in follicular lesions, and differentiating
FVPTC from FA. However, immunohistochemistry is also
confounded with issues of inter-observer sensitivity and
specificity and inter-laboratory variation in application.

A genomics study from The Cancer Genome Atlas
(TCGA) has proposed the classification of thyroid cancers
based on molecular subtypes that better reflect the bio-
logic pathways involved with each cancer subtype. Using
gene-expression profiling (GEP), that project was able to
identify unique gene signatures for PTC. Molecular meth-
ods using next-generation sequencing (ThyroSeq; UPMMC,
Pittsburgh, PA, U.S.A.) and GEP (Affirma: Thermo Electron
Scientific Instruments, Madison, WI, U.S.A) are established
adjunct methods in indeterminate thyroid cytology. Those
genomics technologies have thus far been limited to thyroid
cytology. Their implementation in routine surgical pathol-
ogy has been hampered because of prohibitive operational
and infrastructural costs. Some investigators have also
used gene-expression techniques such as DNA microarrays
to identify biomarkers of PTC, FTC, and FA. However, those
studies have been largely investigational and have not as
yet been translated into a clinical assay for routine use in
surgical pathology.

In the present proof-of-concept study, we used a rap-
id, cost-effective digital counting technology, nCounter
(NanoString Technologies, Seattle, WA, U.S.A.) for gene-
extpression analysis. We identified diagnostic biomarkers
and developed a preliminary diagnostic algorithm that
might aid in resolving diagnostically challenging cases in
the surgical pathology assessment of thyroid neoplasms.

METHODS

Thyroid Samples
This study was approved by the Health Research Ethics
Board of Universities of Alberta and Calgary.

The laboratory information system at the University of
Alberta and Royal Alexandra hospitals in Edmonton, Alber-
ta, were searched for all patients with a diagnosis of thyroid
neoplasm between January 2010 and December 2016. That
search identified 81 patients with thyroid neoplasms and
14 patients with normal results. Predetermined sample
inclusion criteria were a tumour area of 5 mm or more, tumour
cellularity of 80% or more, RNA quantity 100 ng or more,
and satisfactory intrinsic quality control metrics in

RNA Isolation
The RNeasy FFPE kit (Qiagen, Hilden, Germany) was used
to isolate RNA from five 10-μm sections from each sample.
The RNA concentration and purity were measured using a
NanoDrop 2000 spectrophotometer (Thermo Fisher Scien-
tific, Waltham, MA, U.S.A.).

Code Set Design
A literature search was performed to identify genes
involved in thyroid cancer biology, and TCGA was also
reviewed. Key words for our literature search included
“thyroid carcinoma,” “gene expression,” “FTC,” “gene
profiling,” “PTC,” “follicular neoplasms (FN),” “FVPTC,” and
“FA,” singly and in combination. The 83 genes selected as a
result included 5 housekeeping genes: ACT2, ACTB, ATF4,
RPL27, and RPS13. The gene set included all 71 genes in the
thyroid differentiation signature identified by TCGA. Custom
nCounter probes for the gene set were designed by NanoString
Technologies and manufactured by Integrated DNA
Technologies [Coralville, IA, U.S.A. (supplemental Table 1)].

Expression Quantification
Details of the nCounter technology have been reported
previously. Briefly, a custom nCounter code set consist-
ing of multiplexed probes targeting the 83-gene set, were
used for gene-expression analysis. Each code set includes
selected housekeeping genes to control for variations in
RNA input and quality.

Probe pairs with sequences specific to a 100-base
region of each target messenger RNA were designed using
a 3’ biotinylated capture probe and a 5’ reporter probe
tagged with a specific fluorescent barcode, thus creating
2 sequence-specific probes for each target transcript.
Probes were hybridized with RNA at 65°C, and then applied
to the nCounter Preparation Station for automated removal
of excess probe and immobilization of probe-transcript complexes on a streptavidin-coated cartridge. Data were collected using the nCounter Digital Analyzer by counting the individual barcodes. Raw gene expression counts were quality-controlled and normalized using the nSolver Analysis Software (version 2.0: NanoString Technologies) and manufacturer-included positive and negative controls. Quality control (qc) parameters include imaging qc, binding density qc, overall assay efficiency, assay linearity, and limit of detection (supplemental text: qc parameters). Data were normalized to the 5 housekeeping genes to correct for variation in rna input quantity.

Statistical Analysis
Post-normalization statistical analysis and visualization were performed with the R software application (version 3.3.2: R Foundation for Statistical Computing, Vienna, Austria). Normalized transcript counts were used for individual gene analysis (supplemental Table 2). Geometric means of normalized transcript counts were used for aggregate gene-set analysis. Exploratory analysis was performed using heat maps with unsupervised hierarchical clustering by Euclidean distance (heatmap.2 function in gplots package for R) and unsupervised principal component analysis (prcomp function in stats package for R). Receiver operating characteristic (roc) curve analysis (roc function in pROC package for R) was used to assess individual gene and gene-set performance in each of the diagnostic scenarios. Individual genes were first ranked by area under the curve values and then compiled into respective upregulated and downregulated gene sets. That is, 1-gene set is the top gene; 2-gene set is the geometric mean of the top 2 genes; 3-gene set is the geometric mean of the top 3 genes; and so on. The roc curve analysis was then repeated for each gene set and the one with the highest area under the curve was identified for each diagnostic scenario (in the event of a tie, the one with the most transcripts was selected). Benjamini–Hochberg (false discovery rate) correction for multiple comparisons was used for all differential expression analyses. Only genes with corrected p values or false discovery rates less than 0.05 were included in the gene sets. The Youden J-statistic (point on the roc curve furthest from the diagonal index line) defined the diagnostic thresholds. Leave-one-out cross-validation analysis (train function in caret package for R) was used to validate individual gene and gene-set roc results. Mann–Whitney U-tests (wilcox.test function in stats package for R) were used for diagnostic class comparisons. A p value less than 0.05 was considered statistically significant.

RESULTS
RNA and Quality Control
The mean rna yield obtained from five 10 μm sections per ffpe block was 7480.6 ng (range: 161.4 ng–31204.4 ng) with a mean concentration of 374.03 ng/μL (8.07 ng/μL–1560.22 ng/μL) and a mean A260/A280 rna purity ratio of 1.93.

Analytic Cohort of 84 Thyroid Cases
Of 95 identified samples with available paraffin-embedded material, 11 were excluded from downstream analysis because of quality control (n = 5) or normalization (n = 6) flags encountered during nSolver processing. Gene expression data were adequate for 84 of 95 samples (88%) using ffpe material, and those samples constituted the analytic cohort for the study.

Gene Expression Cluster Analysis Shows Clustering of PTC Compared with Other Lesions Using 78-Gene Set
Using all 78 genes and all 84 samples, hierarchical clustering identified molecular signatures for 3 diagnostic groups: ptc, follicular neoplasms, and normal thyroid tissue. Papillary thyroid cancer appears to be molecularly distinct and clustered with 100% specificity within the ptc signature, which contained both classical ptc and some fvpTC samples. The follicular neoplasms cluster included fvpTC, ftc, and fa [Figure 1(A)]. Normal thyroid tissue samples were predominantly clustered together with some overlap with ptc, fa, and ftc. Using all 78 genes and all 84 samples, principal component analysis demonstrated findings similar to those for the heat-map analysis—that is, the genes mainly separated ptc from "not ptc" with fvpTC cases scattered between both groups [Figure 1(B)].

A 3-Step Gene Expression Algorithm Classifies Thyroid Lesions
We developed a 3-step algorithm for diagnostic classification: Step 1 distinguishes normal thyroid samples (n = 14) from abnormal (n = 70), including benign and malignant lesions (ptc, ftc). Step 2 differentiates benign tissue from malignant, and step 3 subclassifies malignant lesions.

Step 1 Analysis
We divided the cohort of 84 patients into 2 groups: normal (n = 14) and abnormal (n = 70). The 78-gene panel was then used to evaluate gene expression. The roc analysis of the results identified 3 genes—SDC4, PLCD3, and PVRL4—to be the combination of genes with the best discriminatory value between those diagnostic groups, with test performance sensitivity of 89%, specificity of 92%, and positive predictive value of 98% (Figure 2).

Step 2 Analysis
The cohort of 84 patients was stratified into 2 groups: benign, including normal tissue (n = 14) and fa (n = 14), and malignancy (ptc, ftc; n = 56). The 78-gene panel was then used to evaluate gene expression. The roc analysis identified the gene SDC4 (alone and not in combination with any other gene) to be the most discriminatory between the cohorts, with test performance sensitivity of 89%, specificity of 92%, and positive predictive value of 98% (Figure 3).

Step 3 Analysis
Finally, the 56 malignant thyroid lesion samples were separated into 3 groups: classic ptc (n = 29), ftc (n = 14) and fvpTC (n = 13). The 78-gene panel was then used to evaluate gene expression. The roc analysis identified 3 gene sets that were differentially expressed in the groups (supplemental Table 3). A 13-gene set (SPOCK2, KCNN4, AHR, RUNX1, FNI, CTSC, ITGB8, PTPRE, ANXA2P2, PDLIM4, SLC34A2, CFH, STK17B) was most discriminatory between classical ptc compared with ftc and fvpTC. A 3-gene set (CA12, CYB561,
KCNAB1) was most discriminatory between classical FvPTC compared with PTC and FTC. A single gene, SLC4A4, was most discriminatory between classical FTC compared with FvPTC and FTC (Figure 4).

The gene sets that are most discriminatory between each of the malignant groups (Figure 5) are a 2-gene set differentiating PTC and FvPTC (SPOCK2, FN1), a 7-gene set differentiating PTC and FTC (SPOCK2, KCNN4, ITGB8, LY6E, PTPRE, CFH, PDE5A), and a 9-gene set differentiating FTC and FvPTC (KCNN4, PDE5A, SDC4, ANXA1, SEL1L3, GABRB2, TBC1D2, DUSP5, CREBS).

FIGURE 1 (A) With inclusion of the entire 83-gene panel, heat-map analysis demonstrates unsupervised hierarchical clustering of papillary thyroid carcinoma (PTC), follicular thyroid carcinoma (FTC), normal thyroid, follicular variant of papillary thyroid carcinoma (FvPTC), and follicular adenoma (FA) samples. Columns show genes; rows show patient samples. Blue shows downregulation of gene expression; red shows upregulation of gene expression. (B) Principal component (PC) analysis including all 78 genes and all 84 samples. No significant separation between FTC, FA, and normal is observed, with PTC appearing molecularly distinct from those three categories, and FvPTC overlapping PTC and “not PTC.”

FIGURE 2 Three genes differentiate “normal” from “not normal.” For each diagnostic group comparison, a receiver operating characteristic analysis was performed for each individual gene. The top upregulated gene set (highest area under the curve (AUC)) is shown. Acc = accuracy; Sens = sensitivity; Spec = specificity.

FIGURE 3 SDC4 differentiates benign from malignant lesions. The top upregulated single-gene set is shown. AUC = area under the curve; Acc = accuracy; Sens = sensitivity; Spec = specificity; PPV = positive predictive value; NPV = negative predictive value.

DISCUSSION
In the present study, we report a proof-of-principle gene expression approach using NanoString nCounter technology as an adjuvant test for classifying thyroid neoplasms. The genes included in the study were selected from publicly available gene expression data sources, including TCGA, and a literature review of PubMed. Unsupervised hierarchical clustering of our gene expression results delineated 3 molecular signatures, 1 highly specific for PTC, as previously demonstrated by TCGA, and 2 relatively less-specific signatures for follicular lesions and normal tissue.
A 3-step classifier was able to improve the diagnosis of the other classes of thyroid neoplasms. Using the 3-step classifier approach based on selected genes, we were able to differentiate various tissue categories, including normal compared with abnormal; benign compared with malignant; and distinct malignant entities, including ptc, fvptc, and ftc. Three genes, SDC4, PLCD3, and PVRL4, were most informative in distinguishing normal from abnormal tissue with a sensitivity and specificity of 100%; SDC4 was most important for differentiating benign from malignant, with a sensitivity of 89% and a specificity of 92%. Various combinations of genes were required to classify specific thyroid neoplasms.

Many of the genes that we identified as being most discriminatory in our study have previously been shown by other authors to be of significance in this context. SDC4 has been shown in various studies to be highly expressed in thyroid carcinoma compared with normal samples. SLC4A4, which in our study showed significant differential expression between FTC and other types of cancers, has been used in a 15-gene set for differentiating benign from malignant samples in indeterminate cytology specimens. That result is especially promising in the otherwise challenging histologic distinction between cases of FTC and fvptc. SLC4A4 could potentially serve as a distinguishing biomarker if validated as an immunohistochemical stain.

Our study demonstrates the potential utility of nCounter technology as an adjunct assay for clinical molecular diagnosis of thyroid neoplasms. NanoString is an innovative technique for the rapid digital assessment of gene expression that is amenable to in-house clinical implementation because of simplicity of use, efficient workflow, cost effectiveness, and user-friendly bioinformatics analysis. The successful analysis of 84 from among 95 of our archival FFPE samples (88%) demonstrates the robustness and efficacy of the technique for surgical pathology specimens. To date, the complexity of other methods for GEP, such as DNA microarrays and RNA sequencing, present significant challenges for the widespread clinical laboratory adoption of such techniques as an ancillary in-house assay. Gene-expression profiling using NanoString is amenable to in-house implementation by smaller laboratories with limited resources. In a number of disease entities, GEP has been adopted clinically for diagnostic, predictive,
and prognostic indications, including determining the cell of origin in diffuse large B-cell lymphoma and identifying the primary site of tumours of unknown origin.\textsuperscript{14,35} And \textit{gene expression profiling} has also been demonstrated to be a useful prognostic and predictive tool in breast carcinoma,\textsuperscript{36,37} brain,\textsuperscript{38} and gastric cancers,\textsuperscript{39} and melanomas.\textsuperscript{40}

NanoString is particularly suitable for the highly degraded nucleic acid material derived from \textit{ffpe} samples, which is the standard processing method used in the surgical pathology assessment of thyroid neoplasms. Our results demonstrate good sensitivity and specificity in thyroid diagnosis, but limitations of this preliminary study include the use of only a limited training-sample cohort. Thus, the performance characteristics of sensitivity and specificity have not been definitively established. More extensive validation in an independent validation cohort is required to establish the performance characteristics (including positive and negative predictive value) and to investigate use of the technique as both a “rule in” and a “rule out” test. To assess the clinical utility of our classifier assay in diagnostic pathology, the validation cohort should include diagnostically challenging cases and other reactive thyroid lesions such as Hashimoto thyroiditis.

CONCLUSIONS

Using the NanoString nCounter platform, we developed a 3-step gene-expression classifier assay for the diagnosis of thyroid neoplasms. This classifier has potential utility in diagnostically challenging cases of thyroid lesions.

CONFLICT OF INTEREST DISCLOSURES

We have read and understood \textit{Current Oncology}’s policy on disclosing conflicts of interest, and we declare the following interests: II has received honoraria for serving on advisory boards for AstraZeneca, Pfizer, Bayer, Novartis, and PrecisionRx-Dx. All other authors have no conflicts to disclose.

AUTHOR AFFILIATIONS

*Department of Laboratory Medicine and Pathology, University of Alberta, and †Alberta Public Laboratories, University of Alberta, Edmonton, AB.

REFERENCES

A ROLE FOR GENE EXPRESSION ANALYSIS IN THYROID CANCER DIAGNOSIS? Armanious et al.