Appendix 1 Materials and methods

Study inclusion and exclusion criteria

The inclusion criteria defined for the study cohort included the following: (I) cases clinically advised for fine-needle aspiration (FNA) with complete medical records from ultrasound examination and thyroid function serology for free thyroxine (FT4) and free tri-iodothyronine (FT3); (II) patients with nodules classified under Bethesda III to V who received surgery owing to local compressive symptoms due to large nodule size, substernal goiter, nodule growth or patient preference; and (III) patients who have not previously received partial or total thyroidectomy. Due to the considerable proportion of malignant histopathology findings observed from patients who opted for surgical intervention despite the benign diagnoses from FNA cytopathology (Bethesda II), these cases were also included as part of the study cohort.

Cases with the following characteristics were excluded from the study: (I) specimen evaluated with poor RNA quality or inadequate number of cells for quantitative chromogenic imprinted gene *in-situ* hybridization (QCIGISH) detection; (II) non-diagnostic cytopathology (Bethesda) I or determinate cytopathology assessed with 97–99% malignancy risk (Bethesda VI); (III) cases not recommended for surgery or refusal to undergo surgical treatment; and (IV) indeterminate postsurgical histopathology.

Ultrasound examination

All ultrasound examinations were performed using a 9–15 MHz linear-array probe (LOGIQ E9, GE Healthcare, Wauwatosa, WI, USA; EPIQ7, Philips Healthcare, Bothell, WA, USA; Aplio 500, Canon Medical Systems, Tokyo, Japan; iU22, Philips Healthcare, Bothell, WA, USA) by experienced radiologists in thyroid imaging and reviewed by two of the authors (Y. Zhang and H. Wu). The main ultrasound features which predicted the probability of thyroid malignancy, including echogenicity, composition, margin, shape and echogenic foci, as outlined by American College of Radiology Thyroid Imaging, Reporting and Data System (ACR TI-RADS) were recorded (7). Points were subsequently assigned for each ultrasound feature for the individual nodule.

Thyroid function serology test

The FT3 and FT4 measurements were obtained using Beckman Coulter UniCel DxI 800 Access Immunoassay System (Beckman Coulter, Boston, MA, USA). The reference normal ranges applied were 2.8–6.3 pmol/L for FT3 and 10.5–24.4 pmol/L for FT4.

Fine-needle aspiration cytopathology

FNA was performed on nodules with relatively high-grade ACR TI-RADS categories and other clinical risk indications. All FNAs were conducted by experienced radiologists under ultrasound guidance. The samples were obtained and smeared onto glass slides with 95% alcohol. The biopsy samples were immediately analyzed by pathologists and reported according to the Bethesda system (20). Patients with Bethesda II thyroid nodules were treated following the American Association of Endocrine Surgeons Guidelines stating that these cases can be safely observed, and that surgery might be considered for cases associated with significant local compressive symptoms due to large nodule size (>3 cm), or per the preference of the patient (5).

QCIGISH detection

For each patient, the same thyroid FNA specimen was divided into two parts for simultaneous cytopathology evaluation and blinded QCIGISH testing. *In-situ* hybridization (ISH) was performed following the procedure previously described (30). Briefly, samples were fixed immediately after sampling in 10% neutral buffered formalin (NBF) for 48 hours at room temperature. The dissociated cells were directly mounted onto positively charged slides. After sample pretreatment, the ISH was performed using probes targeting the non-coding intronic regions of nascent RNAs from small nuclear ribonucleoprotein polypeptide N (SNRPN) and minor histocompatibility antigen H13 (HM13) following the manual instruction of RNAscope 2.5 HD Assay kit (Advanced Cell Diagnostics, Newark, CA, USA) (58). After signal amplification, the detected gene-expressing site appeared as a distinct red or brown dot under common bright field microscope (*Figure S1A*). Data collected from

https://dx.doi.org/10.21037/cco-23-89

microscopic images were used to determine the biallelic expression (BAE), multiallelic expression (MAE) and total expression (TE) according to the equations shown in *Figure S1B*. The QCIGISH detection results were classified into five grades (Grades 0, I, II, III and IV) with the diagnostic grading model development process detailed from a previous thyroid diagnosis study (35) using various sensitivity targets representative of progressive thyroid malignancy risks. Grade 0 indicated a benign result while grade I suggested a possible but low malignancy potential, with both being classified as QCIGISH-negative. Grades II, III and IV were all considered QCIGISH-positive, indicating low, moderate and high malignancy risks, respectively. QCIGISH-negative and QCIGISH-positive classifications represent minimal and elevated aberrant allelic expressions corresponding to low and high malignancy risks, respectively.

Model predictor variable pre-modeling transformation

Predictor variables used in the study to model thyroid malignancy include relevant factors, namely ultrasonography, thyroid function serology, FNA cytopathology and molecular imprinting detection through QCIGISH, which directly represent the diagnostic procedures in the order these are implemented in the clinic. To simulate the process of clinical diagnosis, these factors were transformed into binary categories, as applicable, and modeled against postsurgically confirmed benign and malignant thyroid cases, both individually and collectively, in sequential combination depending on how these diagnostic steps are clinically administered (*Figures S2*,*S3*).

For the ultrasonographic factor, the risk-stratification categories determined using ACR TI-RADS involving categories 2 (not suspicious), 3 (mildly suspicious), 4 (moderately suspicious) and 5 (highly suspicious) were applied. Since the malignancy risks for ACR TI-RADS categories 2 and 3 were relatively low (<2% and 5%, respectively) as compared to categories 4 and 5 (5–20% and >20%, respectively) (1), these categories were aggregated into two levels consisting of category 2 and 3 (assigned as the reference category) against categories 4 and 5 combined.

The serological factors identified for the study, which included the biochemical serum markers FT4 and FT3 for thyroid hormone status were similarly transformed prior to inclusion as predictor variables for model development. The ratios of the serum FT4 and FT3 measurements were determined, as this factor has been similarly reported as an effective indicator for thyroid cancer (17). The range of values for the computed FT4/FT3 ratio was dichotomized into high and low categories using a threshold equal to 3.3 based from a related study (17). As the risk for thyroid malignancy has been associated with higher FT4/FT3 ratio, a low FT4/FT3 level was used as the reference category for the model.

The FNA cytopathology examination results categorized under the Bethesda system consisting of Bethesda II (benign cytopathology), Bethesda III (atypia of undetermined significance or follicular lesions of undetermined significance), Bethesda IV (follicular neoplasm or suspicious for a follicular neoplasm) and Bethesda V (suspicious for malignancy) were used to represent the cytopathologic factor for the model. The categories were transformed from four to two levels. Combined Bethesda II, III and IV categories (relatively low malignancy risks) were used as the reference category and evaluated against the Bethesda V classification (relatively high malignancy risk).

For the imprinting factor, the QCIGISH measurements were stratified into QCIGISH-negative (Grades 0 and I) and QCIGISH-positive (Grades II, III and IV) categories as described from a previous study (35), representing low and high malignancy risks, respectively. The QCIGISH-negative category was assigned as the reference category.

References

58. Wang F, Flanagan J, Su N, *et al.* RNAscope: a novel *in situ* RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn 2012;14:22-9.

https://dx.doi.org/10.21037/cco-23-89

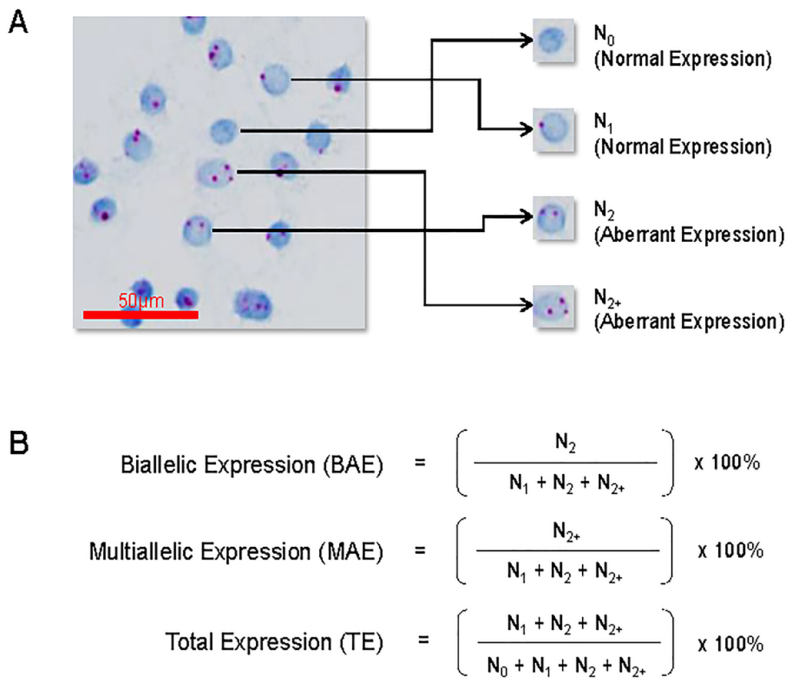

Figure S1 QCIGISH visualization and quantification of the allelic expression status for imprinted genes. (A) QCIGISH staining showing different imprinted gene expression status in cell nuclei; (B) formulas for calculating BAE, MAE and TE measurements. QCIGISH, quantitative chromogenic imprinted gene in-situ hybridization; BAE, biallelic expression; MAE, multiallelic expression; TE, total expression.

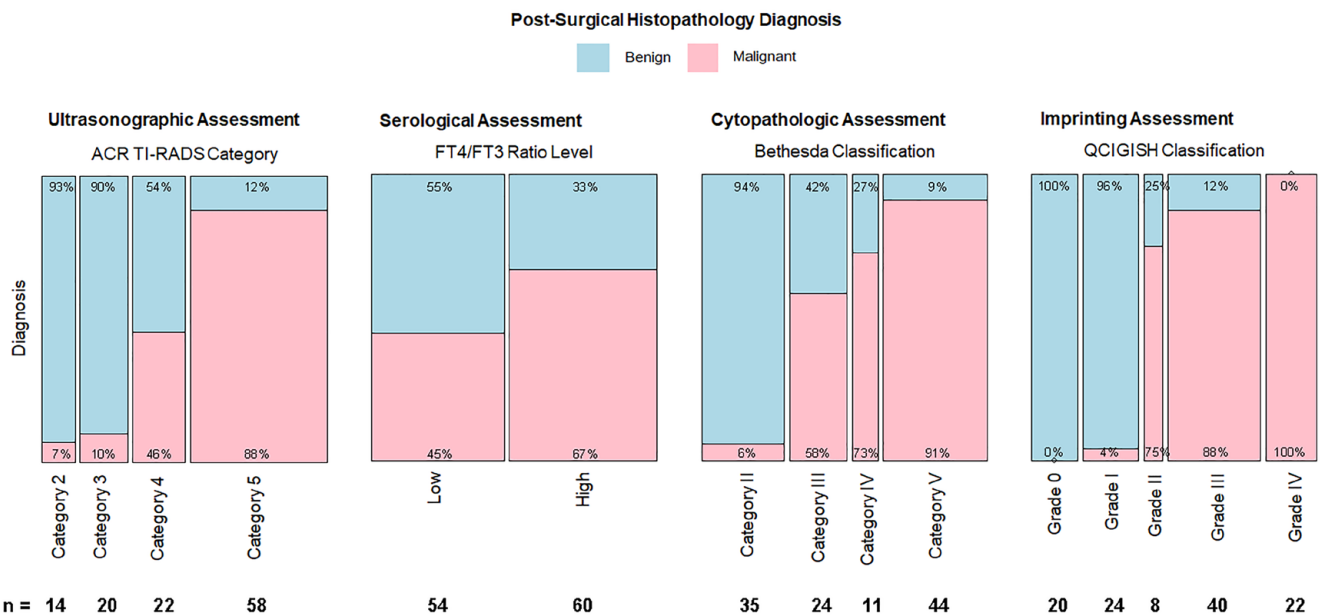

Figure S2 Raw categorical predictors for the ultrasonographic, serological, cytopathologic and imprinting factors prior to logistic regression model development. ACR TI-RADS, American College of Radiology Thyroid Imaging, Reporting and Data System; FT4, free thyroxine; FT3, free tri-iodothyronine; QCIGISH, quantitative chromogenic imprinted gene in-situ hybridization.

https://dx.doi.org/10.21037/cco-23-89

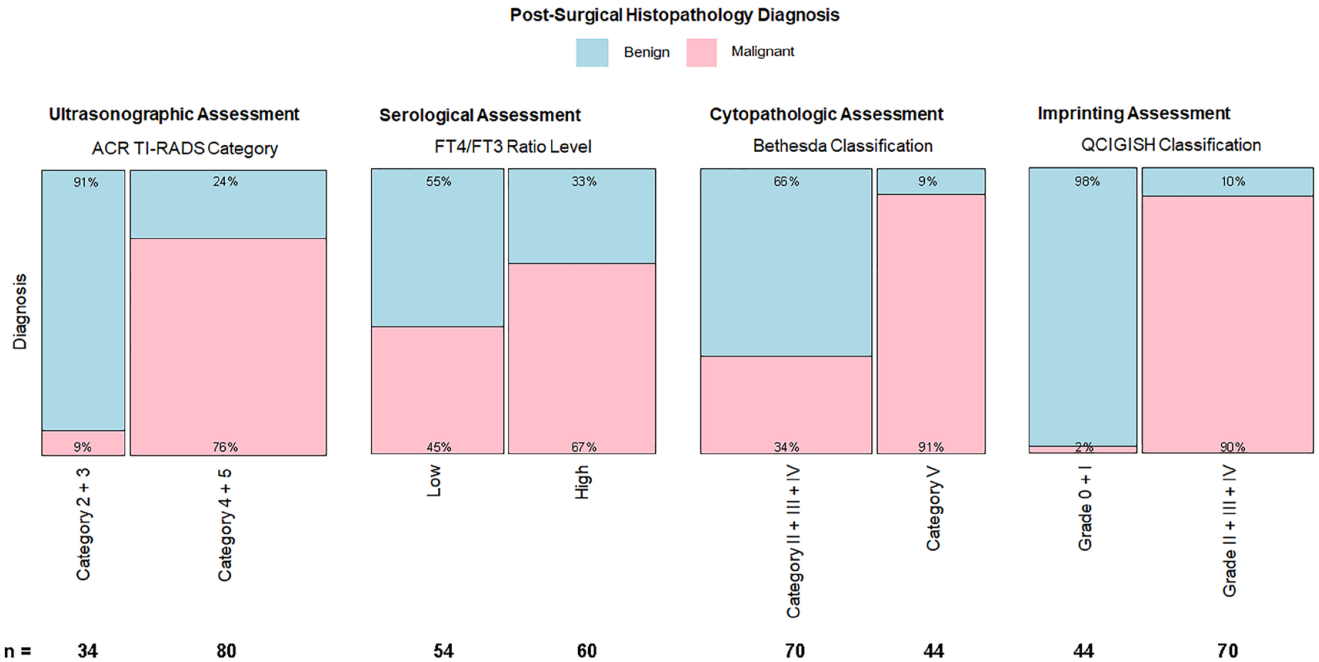

Figure S3 Transformed binary categorical predictors for the ultrasonographic, serological, cytopathologic and imprinting factors prior to logistic regression model development. ACR TI-RADS, American College of Radiology Thyroid Imaging, Reporting and Data System; FT4, free thyroxine; FT3, free tri-iodothyronine; QCIGISH, quantitative chromogenic imprinted gene in-situ hybridization.

https://dx.doi.org/10.21037/cco-23-89

Table S1 AUROC comparison of the stepwise and individual diagnostic models

Diagnostic model assessment	AUROC (95% CI)
Individual diagnostic model	
Utrasonographic[†] (A)	0.787 (0.714 to 0.859)
Serological[‡] (B)	0.613 (0.522 to 0.703)
Cytopathologic[§] (C)	0.773 (0.702 to 0.843)
Imprinting[¶] (D)	0.922 (0.871 to 0.973)
P values[††]	
A *vs.* B	0.001
A *vs.* C	0.75
A *vs.* D	<0.001
B *vs.* C	0.003
B *vs.* D	<0.001
C *vs.* D	<0.001
Stepwise diagnostic model	
Ultrasonographic[†] (E)	0.787 (0.714 to 0.859)
Ultrasonographic[†] + serological[‡] (F)	0.816 (0.737 to 0.896)
Ultrasonographic[†] + serological[‡] + cytopathologic[§] (G)	0.875 (0.810 to 0.939)
Ultrasonographic[†] + serological[‡] + cytopathologic[§] + imprinting[¶] (H)	0.954 (0.909 to 0.999)
P values[††]	
E *vs.* F	0.23
F *vs.* G	0.02
G *vs.* H	0.007

AUROC of the different diagnostic models were compared using the [††]DeLong's test for paired ROC curves. [†], ACR TI-RADS categories used as ultrasonographic diagnostic factors. [‡], FT4/FT3 ratio categories used as serological diagnostic factors. [§], Bethesda classification categories used as cytopathologic diagnostic factors. [¶], QCIGISH classification categories used as imprinting diagnostic factors. AUROC, area under the receiver operating characteristics curve; CI, confidence interval; ACR TI-RADS, American College of Radiology Thyroid Imaging, Reporting and Data System; FT4, free thyroxine; FT3, free tri-iodothyronine; QCIGISH, quantitative chromogenic imprinted gene in-situ hybridization.

https://dx.doi.org/10.21037/cco-23-89