Preoperative metabolic classification of thyroid nodules using mass spectrometry imaging of fine-needle aspiration biopsies

Rachel J. DeHoog[a,1], Jialing Zhang[a,1], Elizabeth Alore[b,2], John Q. Lin[a,2], Wendong Yu[c], Spencer Woody[d], Christopher Almendariz[a], Monica Lin[a], Anton F. Engelsman[e,f], Stan B. Sidhu[e], Robert Tibshirani[g,h,3], James Suliburk[b,3], and Livia S. Eberlin[a,3]

[a]Department of Chemistry, The University of Texas at Austin, Austin, TX 78712; [b]Department of Surgery, Baylor College of Medicine, Houston, TX 77030; [c]Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX 77030; [d]Department of Statistics and Data Sciences, The University of Texas at Austin, Austin, TX 78712; [e]Endocrine Surgery Unit, University of Sydney, Sydney, NSW2065, Australia; [f]Department of Surgery, Amsterdam UMC, University of Amsterdam, 1105 AZ Amsterdam, The Netherlands; [g]Department of Biomedical Data Science, Stanford University, Stanford, CA 94305; and [h]Department of Statistics, Stanford University, Stanford, CA 94305

Thyroid neoplasia is common and requires appropriate clinical workup with imaging and fine-needle aspiration (FNA) biopsy to evaluate for cancer. Yet, up to 20% of thyroid nodule FNA biopsies will be indeterminate in diagnosis based on cytological evaluation. Genomic approaches to characterize the malignant potential of nodules showed initial promise but have provided only modest improvement in diagnosis. Here, we describe a method using metabolic analysis by desorption electrospray ionization mass spectrometry (DESI-MS) imaging for direct analysis and diagnosis of follicular cell-derived neoplasia tissues and FNA biopsies. DESI-MS was used to analyze 178 tissue samples to determine the molecular signatures of normal, benign follicular adenoma (FTA), and malignant follicular carcinoma (FTC) and papillary carcinoma (PTC) thyroid tissues. Statistical classifiers, including benign thyroid versus PTC and benign thyroid versus FTC, were built and validated with 114,125 mass spectra, with accuracy assessed in correlation with clinical pathology. Clinical FNA smears were prospectively collected and analyzed using DESI-MS imaging, and the performance of the statistical classifiers was tested with 69 prospectively collected clinical FNA smears. High performance was achieved for both models when predicting on the FNA test set, which included 24 nodules with indeterminate preoperative cytology, with accuracies of 93% and 89%. Our results strongly suggest that DESI-MS imaging is a valuable technology for identification of malignant potential of thyroid nodules.

ambient mass spectrometry | thyroid nodule diagnosis | metabolic profiles | molecular biomarkers | cancer diagnosis

Thyroid cancer incidence is one of the most rapidly increasing malignancies in the United States (1). While some of the increase is attributable to diagnosis of small lesions and recent efforts have limited biopsy of small clinically irrelevant lesions, current data trends reflect an increase in larger-stage neoplasms that can be with associated mortality increase (1, 2). Ultrasound-guided fine-needle aspiration (FNA) biopsy is a standard of care technique for preoperative diagnosis of imaging suspicious thyroid lesions (3, 4). However, discrimination between malignant and benign thyroid nodules from FNA cytology can be challenging. Interpretation of even basic thyroid cytology requires availability of an expertly trained cytopathologist for accurate diagnosis. This is especially true for follicular neoplasms which are indistinguishable by cytopathology. Currently, diagnosis of follicular thyroid adenoma (FTA) versus follicular thyroid carcinoma (FTC) can be accomplished only after surgical excision of the nodule as FTC is defined by invasion through the thyroid capsule or vascular invasion, both of which are morphologic features not captured by FNA biopsy. Although less frequent, indeterminate FNA diagnosis can also occur for different thyroid malignancies including papillary thyroid carcinoma (PTC) due to

overlapping cytological features, limited sample size, or lack of clear histological pattern in FNA samples. Clinically, ~20% of FNAs are classified as indeterminate, and thus thyroid surgery may be required for final diagnosis. However, in the majority of these cases, the lesion is found to be benign on final histopathology, and the diagnostic surgery was, therefore, possibly unnecessary and in many cases may lead to lifelong hypothyroidism (5–7). Thus, there is a critical need for new technologies for accurate, timely, and comprehensive diagnosis of thyroid FNA biopsies.

Gene- and RNA-sequencing molecular tests have been developed to determine the likelihood that an indeterminate thyroid

Significance

Fine-needle aspiration (FNA) biopsy is a well-established technique for diagnosis of suspicious thyroid lesions. However, histologic discrimination between malignant and benign thyroid nodules from FNA can be challenging. Patients with an indeterminate FNA diagnosis often require diagnostic surgery, with the majority ultimately receiving a benign diagnosis. Here, we employ desorption electrospray ionization mass spectrometry (DESI-MS) imaging to diagnose thyroid lesions based on the molecular profiles obtained from FNA biopsy samples. Based on the molecular profiles obtained from malignant thyroid carcinomas and benign thyroid tissues, classification models were generated and used to predict on DESI-MSI data from FNA material with high performance. Our results demonstrate the potential for DESI-MSI to reduce the number of unnecessary diagnostic thyroid surgeries.

CHEMISTRY

MEDICAL SCIENCES

lesion is malignant or benign and thus aid in the diagnosis and clinical management of patients with indeterminate thyroid FNAs (8–13). The original Afirma gene expression classifier (GEC), for example, measures the expression of 167 genes in an FNA sample to predict whether the nodule is either benign or suspicious with a sensitivity of 90% and a specificity of 52% when predicting on a set of 265 FNAs with indeterminate cytology (8). More recently, Afirma has published a genomic sequencing classifier (GSC), with an improved specificity of 68%, sensitivity of 91%, negative predictive value (NPV) of 96%, and positive predictive value (PPV) of 47% when predicting on a set of 190 cytologically indeterminate FNAs (11). Alternatively, the ThyroSeq molecular test uses next-generation sequencing to evaluate various gene mutations and expression alterations, with a reported sensitivity of 94%, specificity of 82%, NPV of 97%, and PPV of 66% on a cohort of 247 FNAs with indeterminate cytology (13). While these tests have a high sensitivity and NPV, allowing patients with a benign test result to safely undergo observation rather than surgery, they often suffer from a low specificity and PPV.

Mass spectrometry (MS) imaging technologies are powerful tools for discriminating between normal and diseased tissues based on the analysis of thousands of analytes and their spatial distributions directly from tissue sections (14–23). Matrix-assisted laser desorption ionization (MALDI)-MS imaging, for example, has been used to analyze protein profiles in thyroid tissues

and FNA samples (21–23). These studies have focused on distinguishing different types of thyroid cancer tissue samples (22) or on discriminating PTC from benign thyroid (23), while the greatest clinical challenge in discriminating FTA from FTC has not been investigated. Our group has recently applied desorption electrospray ionization (DESI)-MS imaging to investigate the lipid profiles of oncocytic and nononcocytic thyroid tumors, as well as metastatic PTC in lymph nodes (24, 25). DESI is an ambient ionization technique that requires minimal sample preparation prior to analysis, thus allowing for direct and rapid analysis of tissue samples. DESI is particularly powerful in the detection of small molecules such as lipids, fatty acids, and metabolites directly from human tissue samples. In our studies, we have shown that the metabolic and lipid information acquired by DESI-MS imaging from thyroid tissue sections is diagnostic of disease state and indicative of tissue type (24, 25). Other studies have also shown that DESI-MS analysis of tissue smears and FNA samples can be a valuable diagnostic tool (26–28). However, the use of DESI-MS to address the major clinical need to differentiate FTA from FTC in analysis of thyroid FNA samples has not yet been explored or demonstrated prior to our work.

Here, we describe the development and application of nondestructive DESI-MS imaging workflow to characterize and diagnose thyroid tissues and clinical FNAs, including indeterminate FNAs prospectively collected from patients undergoing clinical

Fig. 1. DESI-MS imaging of human thyroid tissue samples and development of statistical models. (*A*) DESI-MS tissue analysis workflow. Samples were obtained from tissue banks, sectioned, and analyzed using histologically nondestructive DESI-MS imaging, stained and evaluated by a pathologist. The diagnosis obtained of the DESI-MS slide is compared the postoperative surgical pathology of the same sample as a measure of agreement. Image courtesy of https://mecovisuals.com. (*B*) Representative DESI-MS ion images and histology of representative normal thyroid, FTA, PTC, and FTC tissues. (H&E-stained scale bars, 2 mm; histology scale bars, 100 μm.) (*C*) Representative DESI-MS metabolic profile of normal thyroid, FTA, PTC, and FTC tissues. (*D*) Using a training set of data, 2 statistical models were built: benign versus PTC (*Top*) and benign versus FTC (*Bottom*), with per-patient cross-validation results shown at *Right*.

DeHoog et al.

www.manaraa.com

care. Rich molecular profiles obtained in 114,125 mass spectra collected from 178 human PTC, FTC, FTA, and normal thyroid tissues were characterized and used to build and test statistical classification models to predict disease status of thyroid lesions. The predictive models were then used to predict on the DESI-MS imaging data obtained from clusters of thyroid cells in clinical FNA samples. In a prospective clinical study, we demonstrate the high performance of this method for the preoperative diagnosis of indeterminate thyroid nodules from FNA biopsy and the potential of this method to thereby possibly reduce the number of diagnostic thyroid surgeries.

Results

Molecular Characterization of Thyroid Tissues Using DESI-MS Imaging. DESI-MS imaging was performed in the negative and positive ion modes on 206 banked human thyroid tissue samples, including 43 normal thyroid, 81 FTA, 40 FTC, and 42 PTC (FTA and FTC classes composed of both oncocytic, or Hurthle cell subtype, and nononcocytic samples) (Fig. 1). The mass spectral profiles presented trends in ion abundances that seemed to be largely characteristic of thyroid tissue types and spatially colocalized to specific histologic features within the tissue sections. In the negative ion mode, several glycerophospholipid (GP) species, sphingolipids, fatty acids (FA), and small metabolites were detected, while phosphatidylcholine (PC), diacylglycerol (DG), and triacylglycerol (TG) were detected in the positive ion mode. Note that potential analyte degradation associated with our experimental method was explored (*SI Appendix*, Fig. S3) (29), with minimal evidence of analyte degradation observed during our analyses.

Based on pathologic evaluation, 28 samples including 6 normal thyroid, 10 FTA, 7 FTC, and 5 PTC were excluded from the sample set due to absence of typical histologic features or poor preservation. In total, 114,125 mass spectra acquired from 178 validated tissue samples were used to build statistical models, which were further employed for prediction on FNA samples.

DESI-MS to Predict Papillary Thyroid Carcinoma. We first built a predictive model to discriminate PTC from benign thyroid using 45,940 pixels or mass spectra extracted from 85 tissue samples (Fig. 1*C*). The predictive model consisted of 67 m/z features and allowed discrimination of PTC from benign thyroid with an overall per-pixel cross-validation (CV) prediction accuracy of 94.5% with an area under the receiver operating characteristic curve (AUC) of 0.98 (Fig. 1*D* and *SI Appendix*, Table S1). Among the selected features with the greatest weight for characterizing benign tissue were phosphatidylserine (PS) 36:2 (m/z 786.530), ceramide (Cer) d42:0 (m/z 682.593), and phosphatidylethanolamine (PE) 38:4 (m/z 766.541). On the other hand, phosphatidylinositol (PI) 36:1 (m/z 863.564), phosphatidic acid (PA) 36:1 (m/z 701.517), and cardiolipin (CL) 74:8 (m/z 737.494) were among the highest-weighted predictive features for PTC. Note that m/z 701.517, attributed to PA (36:1), could also be a fragmentation product of other lipids, such as PS (36:1) (m/z 788.545) (*SI Appendix*, Figs. S1 and S3); however, the exact ion assignment does not impact the performance of the lasso model. When predicting on the withheld validation set of data (48,353 pixels, from 60 samples), an AUC of 0.98 was achieved, with a 93% per-sample agreement with pathology, sensitivity of 92%, and specificity of 94% (Table 1).

DESI-MS to Predict Follicular Thyroid Carcinoma. We next tackled the most challenging clinical problem of differentiating benign thyroid, i.e., FTA subclass, from FTC, by generating a second classifier: benign thyroid versus FTC. The elastic net model built on a training set of 37,966 pixels from 70 patients using cross-validation provided an overall per-pixel CV agreement with pathology of 79.9%, with 69.1% sensitivity, 84.1% specificity, and

Table 1. Predictive performance of statistical models on the tissue validation set according to histopathological subtype

		Prediction	
Model	Sample no.	Benign	Malignant
Benign versus PTC			
Histopathological subtype			
Benign			
Total	48	45	3
Normal thyroid	12	12	0
Follicular adenoma	23	21	2
Hurthle cell adenoma	13	12	1
Malignant			
Papillary carcinoma	12	1	11
Test performance, % (95% CI*)			
Sensitivity	92 (65–99)		
Specificity	94 (83–98)		
NPV	98 (89–100)		
PPV	79 (52–92)		
Benign versus FTC			
Histopathological subtype			
Benign			
Total	55	44	11
Normal thyroid	22	22	0
Follicular adenoma	25	17	8
Hurthle cell adenoma	8	5	3
Malignant			
Total	16	3	13
Follicular carcinoma	11	2	9
Hurthle cell carcinoma	5	1	4
Test performance, % (95% CI*)			
Sensitivity	81 (57–93)		
Specificity	80 (68–88)		
NPV	94 (83–98)		
PPV	54 (35–72)		

*CI, confidence interval.

AUC of 0.83 (Fig. 1*D* and *SI Appendix*, Table S1). The model was composed of a total of 640 m/z features for discriminating benign thyroid from FTC. Among the ions selected as most important for characterizing benign thyroid tissue were FA 20:4 (m/z 303.233), several PS species including PS 36:1 (m/z 788.546) and PS 36:2 (m/z 786.530), and phosphatidylglycerol (PG) 32:1 (m/z 719.490). Among the ions selected as most important for characterizing FTC were PI 38:4 (m/z 885.550), Cer d36:1 (m/z 600.513), PE 36:1 (m/z 744.554), and several metabolites, including succinate (m/z 117.018) and malate (m/z 133.015). The model was then used to predict on the withheld validation set of data (30,369 pixels, 71 samples), yielding an overall per-pixel agreement with pathology of 83.5%, with a sensitivity of 65.2%, specificity of 91.3%, and AUC of 0.89. Remarkably, 80% accuracy, 81% sensitivity, and 80% specificity were achieved on a per-patient basis, with only 3 of 16 FTC samples misclassified as benign in the validation set (Table 1).

Use of DESI-MS to Predict an Independent Set of Clinical Thyroid FNA Biopsies. We then tested the performance of our classifiers built from tissue-imaging data on the independent test set of clinical FNA samples (Fig. 2*A*). As anticipated, the distribution of the thyroid cells was not uniform throughout the FNA smear; therefore, nondestructive DESI-MS imaging was performed on the whole sample area to ensure all of the cells potentially present in the slide were analyzed. As such, the DESI-MS imaging analysis time ranged from ~30 min to 7.5 h per sample, depending on the area of the FNA smear (~0.5 to 14 cm²). When plotting the distribution of various analytes within the FNA samples (Fig. 2*B*),

CHEMISTRY
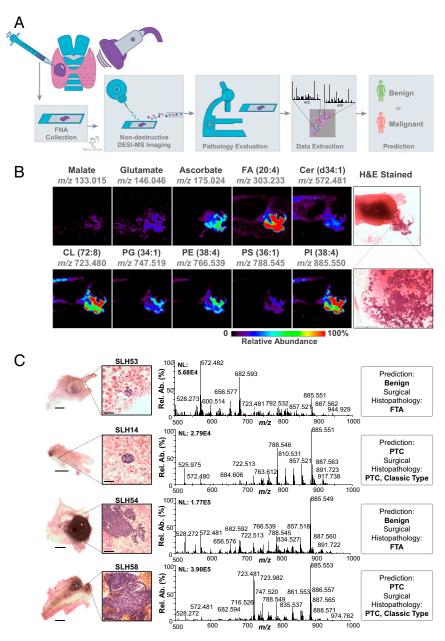
MEDICAL SCIENCES

www.manaraa.com

Fig. 2. FNA analysis and classification using the statistical models built from DESI-MS imaging data. (*A*) FNA analysis workflow. Image courtesy of https://mecovisuals.com. (*B*) DESI-MS ion images and the corresponding H&E-stained image of FNA sample SLH12. (*C*) Optical images of the FNA which were H&E stained after DESI-MS imaging. DESI mass spectra and prediction of various thyroid cell clusters obtained from FNA analysis are shown. (*Left*) Mass spectra obtained from the corresponding cell cluster shown in the H&E images. (*Right*) The prediction for the mass spectra shown using the benign versus PTC lasso model along with the final surgical pathology for the nodule. (Scale bars for the FNA smear, 2 mm; scale bars for the zoomed-in region, 50 μm.)

metabolites, fatty acids, and lipids were strongly localized to regions of cells observed in the hematoxylin and eosin (H&E)-stained smears. Note, however, that there is a possibility for delocalization of analytes to occur due to the freeze/thaw cycle, particularly on a scale that is unresolved with the DESI-MS spatial resolution of 250 μm, although this effect should not have a major impact on the performance of this method. Visualization of the molecular information allowed us to precisely correlate the DESI-MS ion image to the H&E optical image obtained from the sample stained after analysis, allowing mass spectra corresponding to thyroid cell clusters only to be identified and annotated by cytology and carefully extracted, even when surrounded by blood in both in vivo preoperative FNA and ex vivo FNA samples (*SI Appendix*, Fig. S2).

A total of 69 FNA smears analyzed by DESI-MS imaging from 57 patients (11 patients had both in vivo and ex vivo FNA smears, and 1 patient had 2 nodules) had regions of thyroid cells from which data could be extracted, while the remaining 11 FNA samples either did not contain any thyroid cells upon postanalysis histologic evaluation or did not have sufficient DESI-MS signal and were thus excluded from data analysis (see *SI Appendix*, Table S2 for patient demographics information). Depending on the cellularity of the FNA sample, there were anywhere from a single pixel to over 1,000 pixels extracted per FNA smear, with anywhere from ~20 to hundreds of cells within cell clusters identified in each pixel area (250 × 250 μm). Within the 69 FNA test set samples, there were 21 benign, 19 FTA, 28 PTC, and 1 FTC FNA smears from 34 benign and 24 malignant nodules. Among the 34 benign nodules,

21404 | www.pnas.org/cgi/doi/10.1073/pnas.1911333116

DeHoog et al.

Downloaded at Palestinian Territory, occupied on November 25, 2021

www.manaraa.com

Table 2. Predictive performance of statistical models on the FNA test set according to histopathological subtype, per nodule

Histopathological subtype*	N	Benign vs. PTC model prediction		Benign vs. FTC model prediction	
		Benign	Malignant	Benign	Malignant
Benign					
Total	34				
Benign (FNA cytology only)	11	11	0	9	2
Benign thyroid tissue	2	2	0	2	0
Hyperplastic nodule	5	5	0	4	1
Multinodular goiter	2	2	0	2	0
Follicular adenoma	14	11	3	13	1
Malignant					
Total	24				
Follicular carcinoma, Hurthle cell variant	1	NA		0	1
Papillary carcinoma, classic type	21	0	21	NA	
Papillary carcinoma, follicular variant	1	0	1	NA	
Papillary carcinoma, cystic changes	1	1	0	NA	
Test performance, % (95% CI†)					
Sensitivity		96 (79–99)		100 (21–100)	
Specificity		91 (77–97)		88 (73–95)	
NPV		97 (84–99)		100 (88–100)	
PPV		88 (70–96)		20 (4–62)	

*Among the benign nodules, 11 patients did not undergo surgery so FNA cytology was used for the diagnosis rather than histopathology.

†CI, confidence interval.

there were 11 with benign cytology, 2 benign thyroid tissues, 5 hyperplastic, 2 multinodular goiter, and 14 FTA nodules, and among the 24 malignant samples, there were 1 FTC, 21 PTC classic type, 1 PTC follicular variant, and 1 PTC with cystic changes (Table 2).

The benign versus PTC model was first used to predict on the smears from the benign, FTA, and PTC FNA nodules, yielding an AUC of 0.89 and per-nodule specificity of 91% and sensitivity of 96%, with only 4 of the 57 FNA nodules misclassified (Table 2). For patients that had both in vivo and ex vivo FNA smears, the data from both FNA smears were combined to provide an overall per-nodule prediction (per-FNA sample predictions can be found in *SI Appendix*, Table S3). Note that an FNA-based benign versus PTC classifier was also built, yielding comparable prediction performance in CV (*SI Appendix*). Next, the FTC versus benign thyroid model was used to predict on the 20 benign, 14 FTA, and 1 FTC FNA biopsies. Note that the FNA samples were obtained from 35 patients, including 24 patients who had an indeterminate preoperative diagnosis from FNA cytology. Of the 34 benign FNA samples, 30 were correctly classified as benign. A sensitivity of 100%, specificity of 88%, and overall accuracy of 89% were achieved for the FNA samples with an AUC of 0.96.

Discussion

In this study, we describe a method to diagnose thyroid lesions from FNA samples using metabolic data collected via DESI-MS imaging. Two classification models were developed from a large dataset of DESI mass spectra obtained from tissue sections to discriminate PTC and FTC from benign thyroid in a validation set of thyroid tissue sections and a completely independent test set of prospectively collected FNA samples. This strategy was employed as tissue sections are vastly cellular, allowing extraction of metabolic data from multiple pixels within tissue regions composed of regions with dominant histologic features. On the other hand, FNA samples can be highly heterogeneous, presenting sparse and often overlapping cellular content. Tens of thousands of pixels of mass spectra for each tissue type were obtained from tissue sections for building the models, while hundreds of FNA samples would be needed to generate the same volume of data. Using this approach, the PTC versus benign and FTC versus benign models built provided overall CV accuracies of 96% and 74% and validation accuracies of 93% and 80%, respectively, in our tissue dataset. In the FNA test set, overall accuracies of 93% and 89% were achieved, which are extremely promising results especially considering FTA and FTC are cytologically indistinguishable.

While the most established and commercially available molecular tests for thyroid FNA are based on gene expression and mutation profiles, here we report the use of metabolomic and lipidomic information for thyroid cancer diagnosis. Our statistical models were based on hundreds of metabolites and lipids detected by DESI-MS in benign and malignant tissue that are predictive of disease state and biologically relevant. Many of the differentially expressed genes reported in the Afirma GEC are involved in metabolic pathways, such as glycolysis, gluconeogenesis, and the tricarboxylic acid cycle, processes which are highly relevant in cancer metabolism (30–32). In our benign versus FTC model, various metabolite intermediates involved in critical metabolic pathways were selected for FTC characterization, including malate, glutamate, and succinate. A previous GC-MS study by Wojakowska et al. (33) reported increased levels of succinate in thyroid cancers compared to FTA, which corroborates our findings. Furthermore, many of the differentially expressed genes in benign and malignant thyroid tumors described in the Afirma GEC are involved in cell adhesion, extracellular matrix, and plasma membrane pathways (8, 34). In the benign versus PTC model, for example, several lipid species including PI were selected for discriminating PTC from benign thyroid. When examining the saturation levels of their fatty acid chains, the PI species selected for characterizing benign thyroid such as PI 36:4, PI 38:5, and PI 40:5 contained higher degrees of unsaturation compared to those selected for characterizing PTC, such as PI 34:1, PI 34:2, and PI 36:1. Previous studies have indicated that tumors that undergo de novo lipogenesis favor the production of lipids with higher degrees of saturation compared to normal tissues, possibly protecting cancer cells from peroxidation (35). A similar trend was observed for other GP classes

besides PI selected in the PTC model. For example, mono-unsaturated PA 36:1, PE 36:1, and PS 36:1 lipids were selected for characterizing PTC, while many polyunsaturated lipids such as PA 36:2, PE 38:3, and PS 36:2 were indicative of benign thyroid. These results suggest that by leveraging both genetic alterations and metabolomic/lipidomic information obtained by molecular tests and DESI-MS, respectively, insights into the biochemical processes involved in thyroid tumorigenesis could be achieved.

Importantly, the results obtained in the independent test set of thyroid nodule FNA samples collected from patients under clinical care provide compelling evidence of the potential impact DESI-MS imaging could have in better defining treatment options for patients and in reducing unnecessary diagnostic surgeries. From the FNA test set of 57 patients, 24 patients had an indeterminate preoperative diagnosis from FNA cytology. Remarkably, the FTC versus benign model correctly classified 18 of the 20 indeterminate nodules subjected to prediction, while the PTC versus benign model correctly classified 21 of 23 preoperatively indeterminate nodules. Afirma GEC results were available for 6 of these patients with indeterminate cytology (SLH4, SLH13b, SLH15, SLH19, SLH53, and SLH54), and all patients had a "suspicious" result except for patient SLH19, who had a "benign" result but elected to undergo surgery to remove the thyroid nodule. When predicting on these patients, both of our models classified all 6 of these samples as benign, which agreed with the final surgical histopathology diagnosis. Patient SLH54, for example, had a preoperative diagnosis of follicular neoplasm with a suspicious Afirma GEC result and therefore had to undergo surgery for final diagnosis. From this patient, 2 FNAs were collected for DESI-MS analysis: a preoperative in vivo FNA with 17 pixels of extracted mass spectra and an ex vivo surgical FNA that also had 17 pixels extracted, with as few as 30 cells in a single pixel. Using the PTC lasso model, all pixels predicted as benign, which agreed with the final surgical diagnosis of FTA (Table 2 and Fig. 2B). Using our FTC versus benign classifier, a benign classification was also obtained for this nodule, which was in agreement with the final histopathology of FTA. Another patient, SLH4, had 2 nodules: one with an indeterminate preoperative diagnosis with an Afirma GEC result of suspicious and one with a benign cytology. When predicting on the data extracted from the FNA obtained for the indeterminate nodule, a benign prediction was obtained, which agreed with the final surgical histopathology of benign nodular hyperplasia. When predicting on the data extracted from the FNA obtained for the benign nodule from this same patient, agreement was also achieved. Patient SLH43 also had an indeterminate preoperative diagnosis from FNA cytology and had an additional preoperative microRNA testing performed instead of Afirma GEC, resulting in a suspicious diagnosis. Using our PTC lasso classifier, all 548 pixels extracted were predicted as PTC, which agreed with the final histopathology of PTC, classic type. Similarly, patient SLH53 had a suspicious preoperative diagnosis from FNA cytology (no molecular test) and was recommended for surgery. Using our method, an ex vivo FNA sample was analyzed and a single pixel of data was extracted, corresponding to a cluster of ~20 follicular cells (Fig. 2B). Remarkably, this sample was predicted as benign for both the PTC and FTC models, which agreed with the final surgical histopathology diagnosis of FTA. Altogether, our method could have improved diagnosis in 17 of the 19 patients with indeterminate FNA diagnosis that were further confirmed to have benign nodules by surgical pathology and potentially obviate the need for diagnostic surgery by correctly classifying the preoperative FNA samples as benign thyroid, even for cases where commercial clinical genomic classifier Afirma results were suspicious (Fig. 3). It is important to note, however, that patients with a preoperative benign diagnosis may still need to undergo surgery as a result of mass effect, growth of the lesion, or patient concern. In addition to improved benign diagnosis, our method could have correctly confirmed a malignant
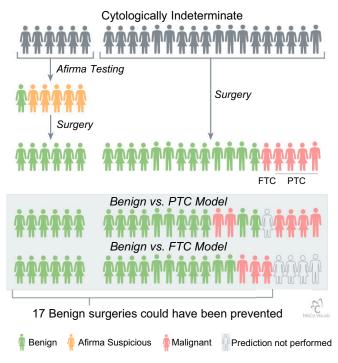

Fig. 3. Statistical predictions on the cytologically indeterminate FNA samples. Of these 24 FNA samples, 6 had Afirma testing done prior to surgery. The benign versus FTC model was used to predict on the benign and FTC samples, with 18 of 20 predicted diagnoses agreeing with the final surgical histopathology. The benign versus PTC model was used to predict on the benign and PTC samples, with 21 of 23 predicted diagnoses agreeing with the final surgical histopathology. Using the DESI method, 17 of 19 benign surgeries could have been prevented. Image courtesy of https://mecovisuals.com.

diagnosis from the other 5 preoperative indeterminate FNA samples of 4 patients with PTC and 1 patient with FTC, further strengthening their need for surgery.

Our study has some limitations that might affect generalization of the results. In this study, the tissues obtained for development and validation of statistical classifiers were from tissue banks and thus access to detailed patient and clinical information including potential therapies that could impact molecular information was limited. Another limitation is the number of clinical FNA samples, particularly regarding FTC, which contained only 1 sample due to the low prevalence of this condition and timeline of our clinical study. To further validate and optimize the performance of our method and confidently establish FNA quality metrics, a larger sample cohort is currently being prospectively collected. Another limitation relates to the foreseen clinical implementation of DESI-MS imaging, which requires FNA samples to be frozen to ensure stability and quality of metabolic markers and thus specific logistic considerations in the clinical storage and transport of samples. A separate FNA pass than what is used for clinical cytological assessment was used for DESI-MS in our study and might be needed in clinical workflows to ensure adequate sample quality for accurate pathologic readings, although this process is already in place in centers that have implemented gene expression or gene sequencing classifier testing for thyroid neoplasia. Finally, the current time per imaging experiment is arguably long (a few hours) for FNA samples that have been smeared over large areas (a few square centimeters) of the histologic glass slide and further extended by the time requirement related to data extraction and processing steps. Advanced optimization and standardization of sample smearing approaches, time per scan for DESI-MS analyses, and development of software

tools for data analysis are being pursued to expedite and automate the analysis.

In conclusion, our results provide compelling evidence that the metabolic information obtained by DESI-MS imaging coupled to statistical analysis is a powerful approach to aid in the diagnosis of thyroid nodules. DESI-MS imaging allows direct analysis of FNA smears without any sample modification or preparation requirements. Further, the DESI-MS imaging workflow we implemented is nondestructive and histologically compatible (36), allowing the cytology of the FNA smears to be preserved for further evaluation and thus unambiguous spatial colocalization and correlation between molecular data and cell clusters for precise data extraction. Thus, we envision our method to be used in conjunction with current pathology/cytology workflows. With further clinical validation, the addition of the DESI-MS imaging into clinical practice has tremendous promise to improve the diagnosis of thyroid neoplasia, especially in cases of indeterminate cytology, potentially reducing the number of diagnostic surgeries while improving clinical management and outcomes for patients.

Methods

Study Design and Participants. A total of 206 frozen human thyroid tissue samples, including 43 normal thyroid, 62 nononcocytic FTA, 19 oncocytic FTA (or Hurthle cell adenoma), 30 nononcocytic FTC, 10 oncocytic FTC (or Hurthle cell carcinoma), and 42 PTC, were obtained from the Cooperative Human Tissue Network, Baylor College of Medicine Tissue Bank, Asterand Bioscience, and the Kolling Institute of Medical Research Tumor Bank. The samples were sectioned, imaged using DESI-MS, stained, and evaluated by an expert endocrine cytopathologist after analysis. A total of 28 thyroid tissues that did not present with regions of clear diagnosis as determined by the endocrine cytopathologist were excluded from the dataset. Classification models were generated using the mass spectra corresponding to the regions of clear histology determined by the pathologist.

A total of 80 deidentified FNA smears from 68 patients that provided informed consent were prospectively collected at Baylor College of Medicine from March 2017 to March 2018 with institutional review board approval. Aspirates were collected from patients undergoing standard outpatient FNA biopsy, preoperatively from patients undergoing thyroid nodule surgery, and from ex vivo thyroid nodules upon resection from the surgical field. Only 1 patient, SLH4, had 2 nodules from which FNA biopsies were obtained. Samples were then analyzed using DESI-MS imaging, followed by staining and pathologic evaluation. Eleven FNA samples that did not contain thyroid cells or did not have sufficient DESI-MS signal were excluded from statistical analysis. The classification models built from the tissue imaging data were then used to predict on the data obtained from the remaining 69 FNA samples.

Procedures. Frozen tissues were sectioned at a thickness of 16 μm and the glass slides stored at −80 °C until analysis. Tissue sections were thawed at room temperature for ~10 min and then analyzed using a Q Exactive (Thermo Scientific) mass spectrometer fitted with a 2D Omni spray stage (Prosolia Inc.) and a laboratory-built DESI sprayer. DESI-MS imaging was performed at a spatial resolution of 200 μm, mass range of m/z 100 to 1,500, and a resolving power of 70,000 (at m/z 200). To ensure nondestructive DESI imaging, thus allowing sequential analysis of the same tissue section by histopathology, histologically compatible solvent systems were used (36). In the negative ion mode, dimethylformamide:acetonitrile (DMF:ACN) 1:1 (vol/vol) was used at a flow rate of 1.2 μL/min, while pure ACN was used in the positive ion mode at a flow rate of 3 μL/min. The N₂ pressure was set to 180 psi. DESI ion images were assembled using Firefly and BioMap software. High mass accuracy measurements and tandem mass spectrometry using higher-energy collisional dissociation (HCD) were used for lipid and metabolite identification (*SI Appendix*, Fig. S1). After DESI-MS imaging, all tissue sections were stained using standard H&E protocol. Pathologic evaluation was performed by our study team (W.Y.) per standard of care microscopy. Regions of clear diagnosis were annotated on the glass slides and then correlated to the metabolic signatures obtained by DESI-MS from the same tissue regions. As FTA and FTC are difficult to diagnose from frozen sections, we relied on the previously verified tissue diagnosis provided by the tissue bank for these samples.

For FNA samples, aspirates were deposited on a glass slide, frozen, and stored at −80 °C until analysis. Prior to analysis, samples were air dried for ~20 min. As there was only a single slide collected for DESI-MS analysis for many of the patients, each FNA sample was imaged in first the positive and subsequently the negative ion mode at a spatial resolution of 250 μm. The same imaging parameters were used for DESI-MS analysis of FNA biopsies in Q Exactive and Q Exactive HF mass spectrometers (Thermo Scientific), with an adjusted solvent flow rate of 1.5 μL/min for DMF:ACN 1:1 and 4.5 μL/min for ACN. These parameters resulted in a scan rate of 0.52 s per scan for the Q Exactive and 0.54 s per scan for the Q Exactive HF. After analysis, samples were stained and evaluated by our study team (W.Y.). Clusters of follicular thyroid cells were identified and annotated and mass spectra corresponding to these clusters were extracted to be used for statistical prediction, with the time spent on data extraction ranging from 5 to 60 min per sample. For patients with indeterminate preoperative FNA cytology, the final diagnosis was determined postoperatively based on the final clinical pathology of the surgical specimen.

Statistical Analysis. Mass spectra corresponding to regions of clear diagnosis were extracted using MSiReader software. Following data extraction, the lasso (37) and elastic net (38) statistical methods were used to build predictive models using the mass spectra obtained. The greatest clinical relevance lies in discriminating thyroid cancer from benign thyroid tissue; thus, the normal thyroid and adenoma classes were combined to create a benign thyroid class for statistical analysis. Two classification models were generated: benign thyroid vs. PTC (using lasso) and benign thyroid vs. FTC (using elastic net). Data preprocessing steps were applied prior to running lasso or elastic net (see *SI Appendix* for details). The data were then randomly split into a training and a validation set of data on a per sample basis. Lasso/ elastic net was used along with 5-fold CV to generate a model using the training set of data. The model was then used to predict on the withheld validation set of data. Sensitivity, specificity, PPV, NPV, and 95% Wilson confidence intervals (39) were calculated based on the agreement of the model prediction with pathology. The classification models generated from the tissue-imaging data were then used to predict on each mass spectrum extracted from the FNA samples, which can be performed within minutes of data extraction. The FTC versus benign model was applied to all FNA samples except those that contained cells with papillary nuclear features, while the PTC versus benign model was applied to all FNA samples but the FNA sample that had been confirmed as FTC by postoperative diagnosis. As some patients had multiple FNA smears, the predictive results from each of the FNA smears from the same patient were combined to give an overall per-nodule diagnosis. A new lasso PTC vs. benign thyroid model was also generated using the FNA data.

Data and Materials Availability. The data for this study are available for researchers and have been deposited in Dataverse (https://doi.org/10.7910/DVN/3B6NSQ).

1. R. L. Siegel, K. D. Miller, A. Jemal, Cancer statistics, 2019. *CA Cancer J. Clin.* **69**, 7–34 (2019).
2. E. G. Grant *et al.*, Thyroid ultrasound reporting lexicon: White paper of the ACR thyroid imaging, reporting and data system (TIRADS) committee. *J. Am. Coll. Radiol.* **12**, 1272–1279 (2015).
3. J. Y. Kwak *et al.*, Thyroid imaging reporting and data system for US features of nodules: A step in establishing better stratification of cancer risk. *Radiology* **260**, 892–899 (2011).
4. H. J. Moon, E. K. Kim, J. H. Yoon, J. Y. Kwak, Malignancy risk stratification in thyroid nodules with nondiagnostic results at cytologic examination: Combination of thyroid imaging reporting and data system and the Bethesda system. *Radiology* **274**, 287–295 (2015).
5. M. Bongiovanni, A. Spitale, W. C. Faquin, L. Mazzucchelli, Z. W. Baloch, The Bethesda system for reporting thyroid cytopathology: A meta-analysis. *Acta Cytol.* **56**, 333–339 (2012).
6. M. Nishino, Molecular cytopathology for thyroid nodules: A review of methodology and test performance. *Cancer Cytopathol.* **124**, 14–27 (2016).
7. C. J. Balentine, R. P. Domingo, R. Patel, J. W. Suliburk, Thyroid lobectomy for indeterminate FNA: Not without consequences. *J. Surg. Res.* **184**, 189–192 (2013).
8. E. K. Alexander *et al.*, Preoperative diagnosis of benign thyroid nodules with indeterminate cytology. *N. Engl. J. Med.* **367**, 705–715 (2012).
9. E. Labourier *et al.*, Molecular testing for miRNA, mRNA, and DNA on fine-needle aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. *J. Clin. Endocrinol. Metab.* **100**, 2743–2750 (2015).
10. G. Lithwick-Yanai *et al.*, Multicentre validation of a microRNA-based assay for diagnosing indeterminate thyroid nodules utilising fine needle aspirate smears. *J. Clin. Pathol.* **70**, 500–507 (2017).

DeHoog et al.

CHEMISTRY

MEDICAL SCIENCES

11. K. N. Patel *et al.*, Performance of a genomic sequencing classifier for the preoperative diagnosis of cytologically indeterminate thyroid nodules. *JAMA Surg.* **153**, 817–824 (2018).

12. M. T. D. Santos *et al.*, Molecular classification of thyroid nodules with indeterminate cytology: Development and validation of a highly sensitive and specific new miRNA-based classifier test using fine-needle aspiration smear slides. *Thyroid*, 10.1089/thy.2018.0254 (2018).

13. D. L. Steward *et al.*, Performance of a multigene genomic classifier in thyroid nodules with indeterminate cytology: A prospective blinded multicenter study. *JAMA Oncol.* **5**, 204–212 (2019).

14. H. Tian *et al.*, Gas cluster ion beam time-of-flight secondary ion mass spectrometry high-resolution imaging of cardiolipin speciation in the brain: Identification of molecular losses after traumatic injury. *Anal. Chem.* **89**, 4611–4619 (2017).

15. M. Sans *et al.*, Metabolic markers and statistical prediction of serous ovarian cancer aggressiveness by ambient ionization mass spectrometry imaging. *Cancer Res.* **77**, 2903–2913 (2017).

16. D. Calligaris *et al.*, Application of desorption electrospray ionization mass spectrometry imaging in breast cancer margin analysis. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15184–15189 (2014).

17. D. R. Ifa, L. S. Eberlin, Ambient ionization mass spectrometry for cancer diagnosis and surgical margin evaluation. *Clin. Chem.* **62**, 111–123 (2016).

18. Y. Ucal *et al.*, Clinical applications of MALDI imaging technologies in cancer and neurodegenerative diseases. *Biochim. Biophys. Acta. Proteins Proteomics* **1865**, 795–816 (2017).

19. J. M. Spraggins *et al.*, Next-generation technologies for spatial proteomics: Integrating ultra-high speed MALDI-TOF and high mass resolution MALDI FTICR imaging mass spectrometry for protein analysis. *Proteomics* **16**, 1678–1689 (2016).

20. K. Schwamborn, R. C. Krieg, S. Uhlig, H. Ikenberg, A. Wellmann, MALDI imaging as a specific diagnostic tool for routine cervical cytology specimens. *Int. J. Mol. Med.* **27**, 417–421 (2011).

21. M. Galli *et al.*, Proteomic profiles of thyroid tumors by mass spectrometry-imaging on tissue microarrays. *Biochim. Biophys. Acta. Proteins Proteomics* **1865**, 817–827 (2017).

22. M. Pietrowska *et al.*, Molecular profiles of thyroid cancer subtypes: Classification based on features of tissue revealed by mass spectrometry imaging. *Biochim. Biophys. Acta. Proteins Proteomics* **1865**, 837–845 (2017).

23. F. Pagni *et al*., Proteomics in thyroid cytopathology: Relevance of MALDI-imaging in distinguishing malignant from benign lesions. *Proteomics* **16**, 1775–1784 (2016).

24. J. Zhang *et al.*, Cardiolipins are biomarkers of mitochondria-rich thyroid oncocytic tumors. *Cancer Res.* **76**, 6588–6597 (2016).

25. J. Zhang *et al.*, Detection of metastatic breast and thyroid cancer in lymph nodes by desorption electrospray ionization mass spectrometry imaging. *J. Am. Soc. Mass Spectrom.* **28**, 1166–1174 (2017).

26. M. Woolman *et al.*, An assessment of the utility of tissue smears in rapid cancer profiling with desorption electrospray ionization mass spectrometry (DESI-MS). *J. Am. Soc. Mass Spectrom.* **28**, 145–153 (2017).

27. A. K. Jarmusch *et al.*, Characteristic lipid profiles of canine non-Hodgkin's lymphoma from surgical biopsy tissue sections and fine needle aspirate smears by desorption electrospray ionization–mass spectrometry. *Analyst* **140**, 6321–6329 (2015).

28. V. Pirro *et al.*, Intraoperative assessment of tumor margins during glioma resection by desorption electrospray ionization-mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6700–6705 (2017).

29. A. L. Dill, L. S. Eberlin, A. B. Costa, D. R. Ifa, R. G. Cooks, Data quality in tissue analysis using desorption electrospray ionization. *Anal. Bioanal. Chem.* **401**, 1949–1961 (2011).

30. R. G. Coelho, R. S. Fortunato, D. P. Carvalho, Metabolic reprogramming in thyroid carcinoma. *Front. Oncol.* **8**, 82 (2018).

31. S. Beloribi-Djefaflia, S. Vasseur, F. Guillaumond, Lipid metabolic reprogramming in cancer cells. *Oncogenesis* **5**, e189 (2016).

32. R. A. Kulkarni *et al.*, A chemoproteomic portrait of the oncometabolite fumarate. *Nat. Chem. Biol.* **15**, 391–400 (2019).

33. A. Wojakowska *et al.*, Detection of metabolites discriminating subtypes of thyroid cancer: Molecular profiling of FFPE samples using the GC/MS approach. *Mol. Cell. Endocrinol.* **417**, 149–157 (2015).

34. D. Chudova *et al.*, Molecular classification of thyroid nodules using high-dimensionality genomic data. *J. Clin. Endocrinol. Metab.* **95**, 5296–5304 (2010).

35. E. Rysman *et al.*, De novo lipogenesis protects cancer cells from free radicals and chemotherapeutics by promoting membrane lipid saturation. *Cancer Res.* **70**, 8117–8126 (2010).

36. L. S. Eberlin *et al.*, Nondestructive, histologically compatible tissue imaging by desorption electrospray ionization mass spectrometry. *ChemBioChem* **12**, 2129–2132 (2011).

37. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).

38. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320 (2005).

39. A. Agresti, B. A. Coull, Approximate is better than "exact" for interval estimation of binomial proportions. *Am. Stat.* **52**, 119–126 (1998).

DeHoog et al.

www.manaraa.com