

# **Appendix**

## **A protein-based classifier for differentiating follicular thyroid adenoma and carcinoma**

Yaoting Sun, *et al.*

### **Table of contents**

#### **Appendix Figures**

Appendix Figure S1. Discovery proteomics design and quality control. —Page 2

Appendix Figure S2. Comparison of machine learning models and feature counts performances. —Page 3

Appendix Figure S3. Protein feature characteristics. —Page 4 to Page 9

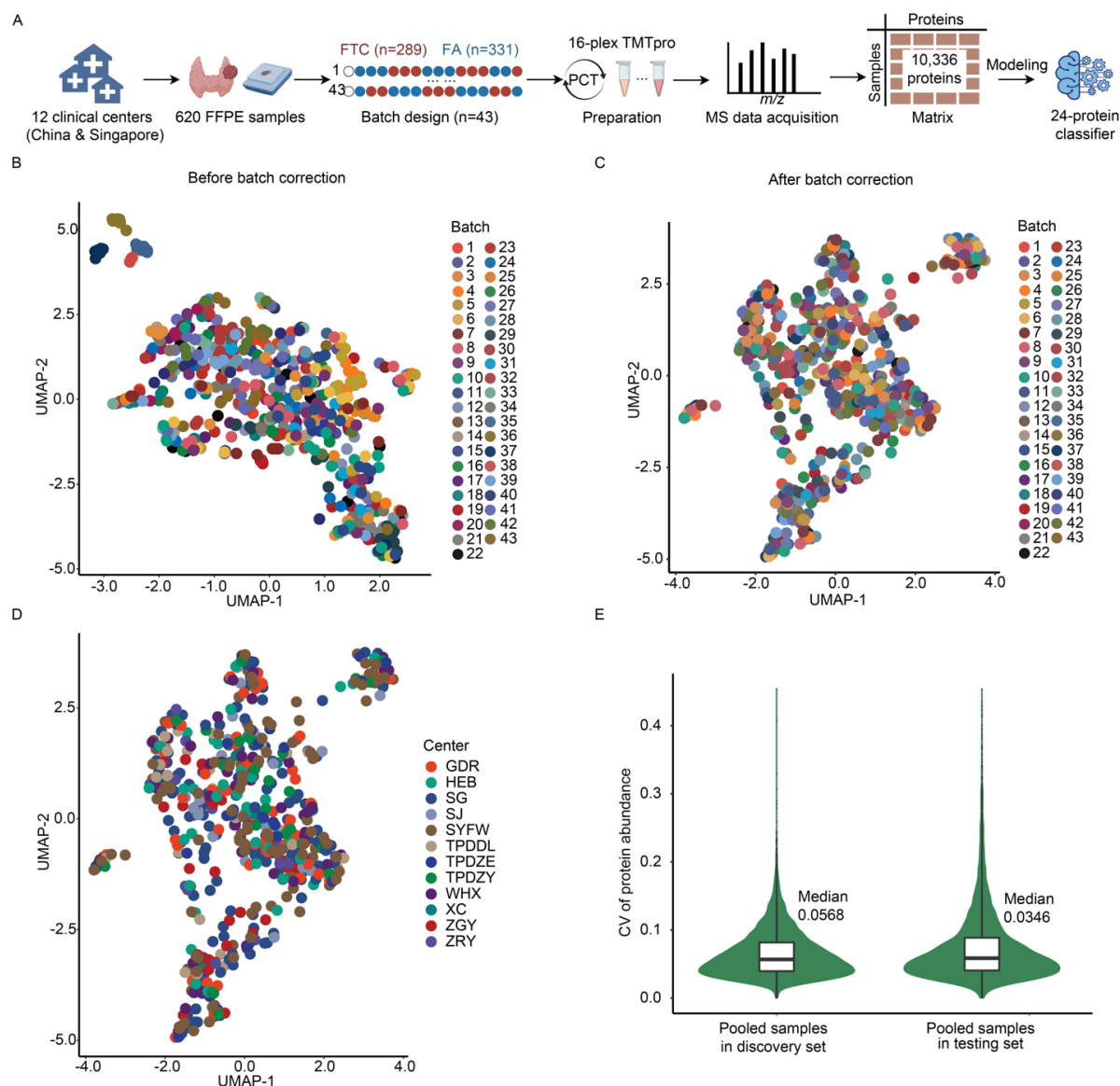
#### **Appendix Tables**

Appendix Table S1. Gene mutation frequency in FTA and FTC. —Page 10

Appendix Table S2. Twenty-four proteins established associations with thyroid physiology or pathology. —Page 11

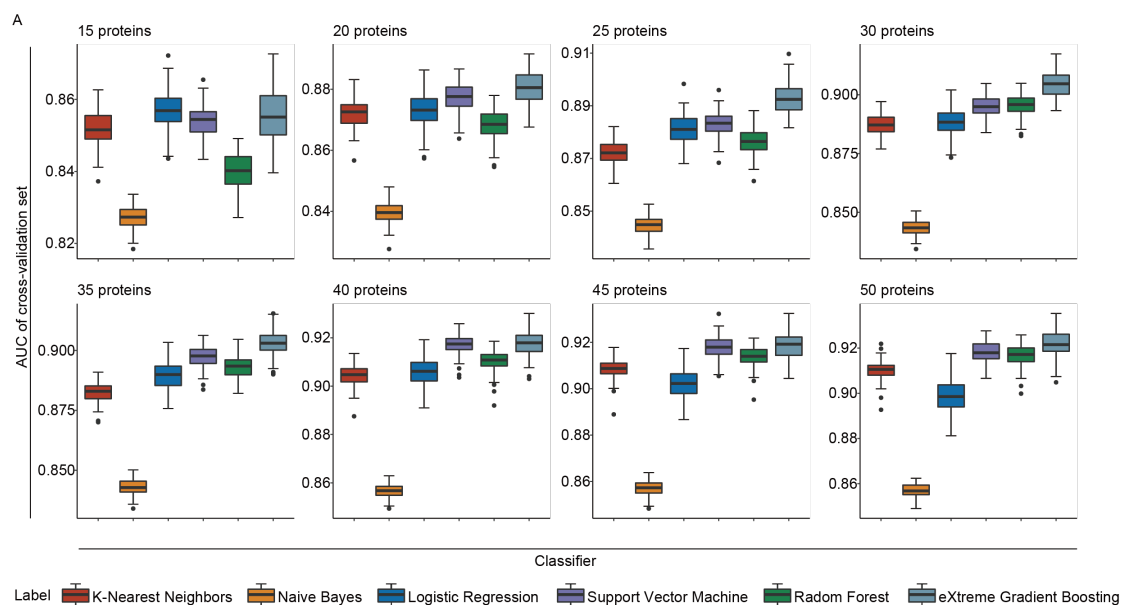
Appendix Table S3. List of 66-gene panel of the thyroid cancer. —Page 12

**Non-contributing We-TEC Investigators.** —Page 13



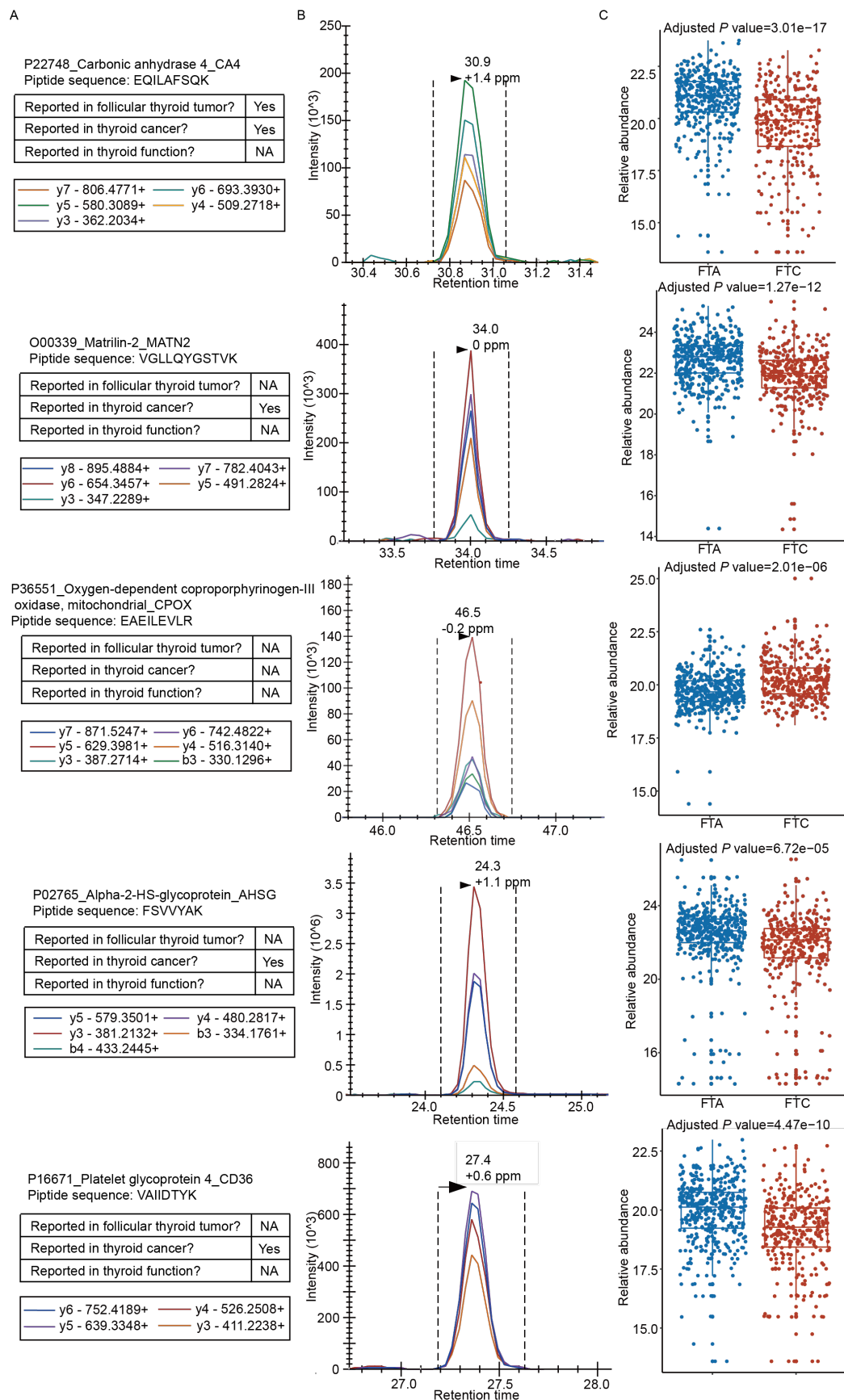
### Appendix Figure S1. Discovery proteomics design and quality control.

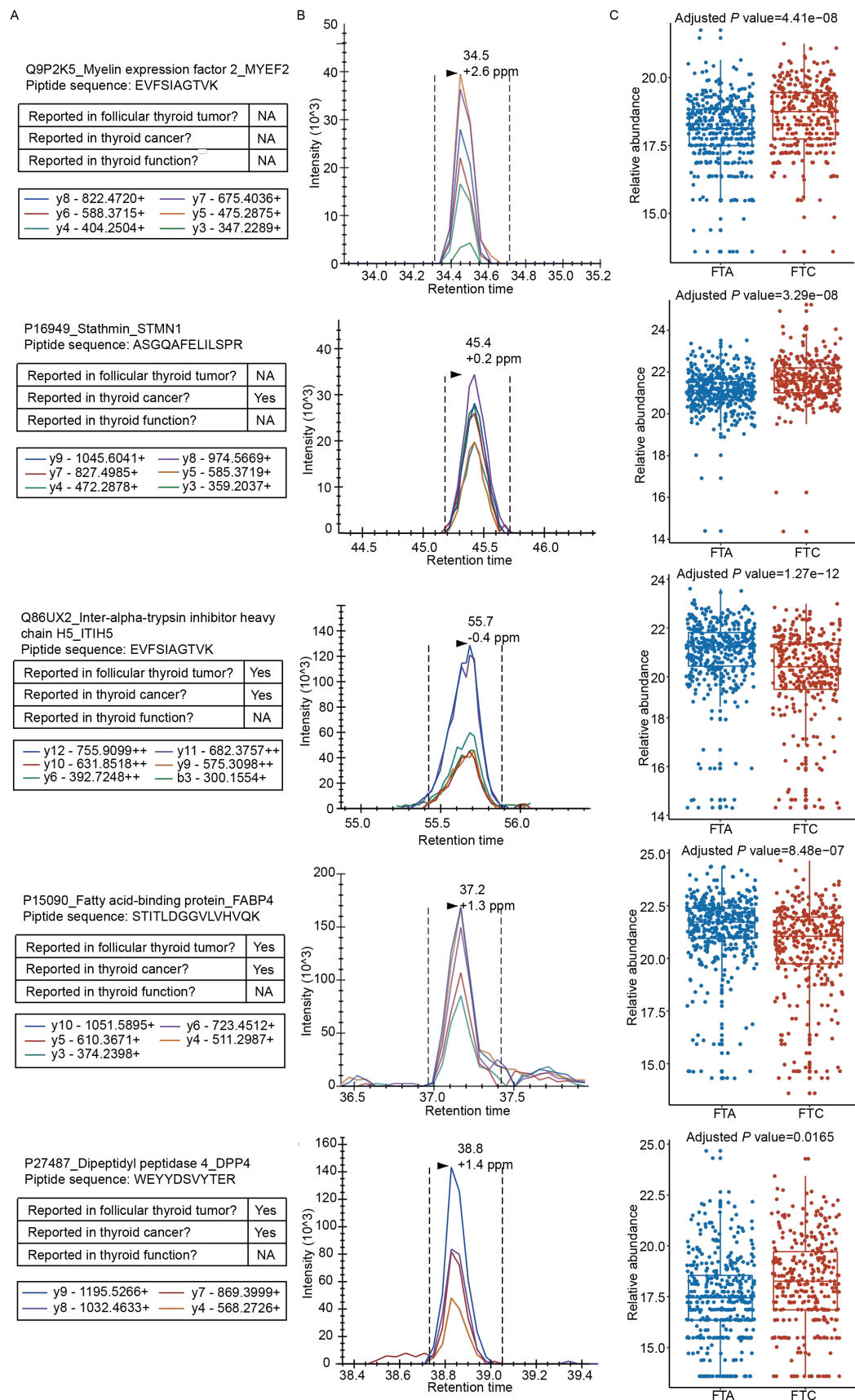
Study flowchart. Samples (n=620) were retrospectively collected from 12 clinical centers. The samples were randomly allocated into 43 batches and prepared by pressure cycling technology (PCT) assisted sample preparation. Peptides were further labeled by 16-plex tandem mass tag (TMT) and data were acquired through the mass spectrometer with data-dependent acquisition (DDA) mode. There were 10,336 proteins quantified, and a 24-protein classifier was built based on them. (B-D) Uniform manifold approximation and projection (UMAP) plot showing the distribution of batches of samples using 7876 proteins (B) before and (C) after batch correction. Dots are colored according to the TMT batches. (D) UMAP analysis of clinical centers using 7876 proteins. (E) Violin plots show the CV of protein abundance across pooled samples in the discovery and testing sets. Medians are labeled upon the plot. The boxes represent the first and third quartiles, with the center line indicating the median. The whiskers extend to  $\pm 1.5$  times the interquartile range, while the violin shapes depict the density of the data points.



### Appendix Figure S2. Comparison of machine learning models and feature counts performances.

(A) The compared six models (x-axis) are k-nearest neighbors (KNN), naive bayes (NB), linear regression (LR), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost). Each panel indicates the different protein feature counts applied. Each boxplot was derived from 100 times cross-validation on the training set. Y-axis showing the values of area under the curve (AUC) from the 100 iterations. The boxes represent the first and third quartiles, the center line indicates the median, the whiskers extend to  $\pm 1.5$  times the interquartile range, and the dots represent individual data points.





A

P61916\_NPC intracellular cholesterol transporter 2\_NPC2  
Peptide sequence: LVVEWQLQDDK

|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | NA  |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

|                 |                 |
|-----------------|-----------------|
| y9 - 1160.5582+ | y8 - 1061.4898+ |
| y7 - 932.4472+  | y6 - 746.3679+  |
| y5 - 618.3093+  | y4 - 505.2253+  |

Q9Y6M1\_Insulin-like growth factor 2 mRNA  
-binding protein 2\_IGF2BP2  
Peptide sequence: LYIGNLSPAVTADDLR

|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | NA  |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

|                  |                  |
|------------------|------------------|
| y13 - 1328.6805+ | y11 - 1157.6161+ |
| y10 - 1044.5320+ | y9 - 957.5000+   |
| y6 - 690.3417+   | y3 - 403.2300+   |

Q687X5\_Metalloreductase STEAP4\_STEAP4  
Peptide sequence: TTLPSGAELVSYSEAAK

|                                       |    |
|---------------------------------------|----|
| Reported in follicular thyroid tumor? | NA |
| Reported in thyroid cancer?           | NA |
| Reported in thyroid function?         | NA |

|                  |                  |
|------------------|------------------|
| y14 - 1408.6955+ | y8 - 868.4411+   |
| y7 - 755.3570+   | y14 - 704.8514++ |
| b3 - 316.1867+   | b4 - 429.2708+   |

O15460\_Proyl 4-hydroxylase subunit alpha-2  
P4HA2  
Peptide sequence: SSWLEEDDDPVVAR

|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | NA  |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

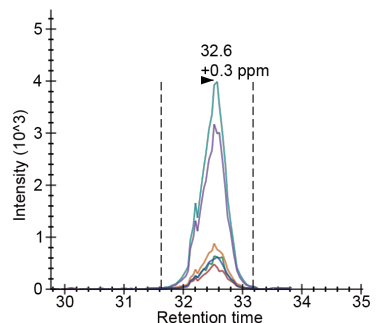
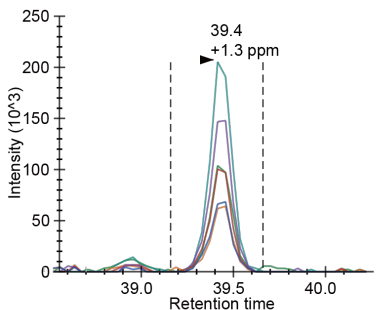
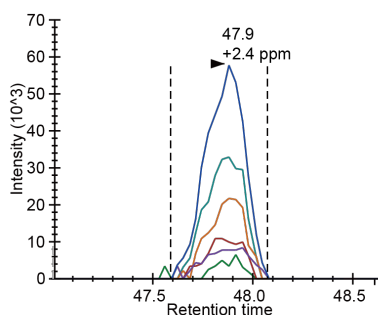
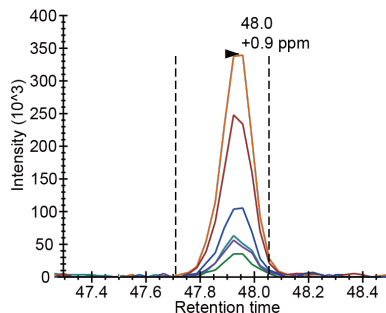
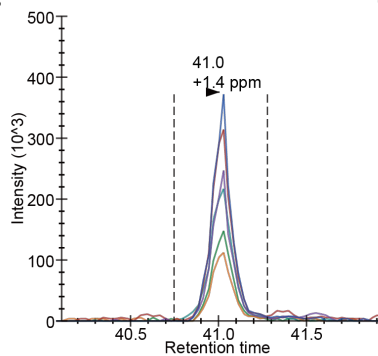
|                  |                  |
|------------------|------------------|
| y11 - 1257.5957+ | y10 - 1144.5117+ |
| y9 - 1015.4691+  | y8 - 886.4265+   |
| y5 - 541.3457+   | b3 - 361.1506+   |

P16401\_Histone H1.5\_H1-5  
Peptide sequence: ATGPPVSELITK

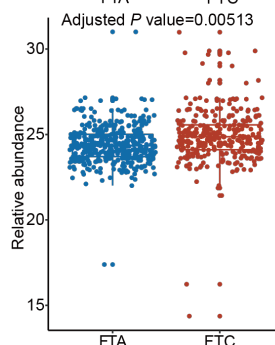
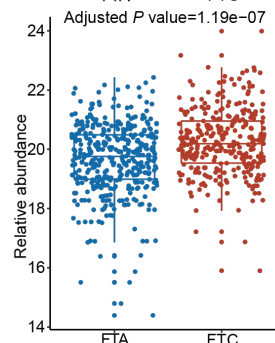
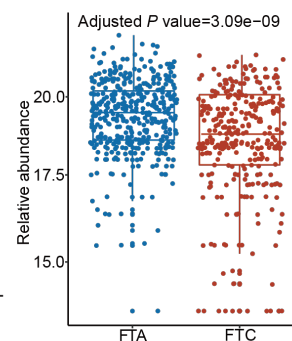
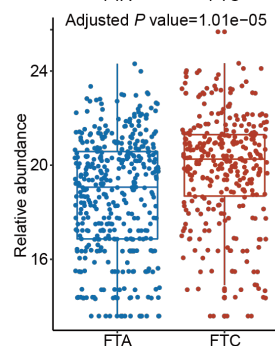
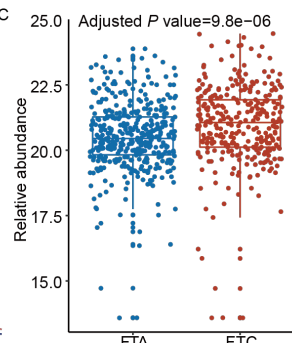
|                                       |    |
|---------------------------------------|----|
| Reported in follicular thyroid tumor? | NA |
| Reported in thyroid cancer?           | NA |
| Reported in thyroid function?         | NA |

|                  |                 |
|------------------|-----------------|
| y10 - 1040.5986+ | y8 - 886.5244+  |
| y7 - 789.4716+   | y6 - 690.4032+  |
| y9 - 492.2922++  | y8 - 443.7658++ |

B



C



A

P01033\_Metalloproteinase inhibitor 1\_TIMP1  
Peptide sequence: GFQALGDAADIR

|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | Yes |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

|                |                |
|----------------|----------------|
| y9 - 901.4738+ | y8 - 830.4367+ |
| y7 - 717.3526+ | y5 - 545.3042+ |
| b3 - 333.1557+ | b4 - 404.1928+ |

Q9HCD6\_Protein TANC2\_TANC2  
Peptide sequence: LGFLLGEK

|                                       |    |
|---------------------------------------|----|
| Reported in follicular thyroid tumor? | NA |
| Reported in thyroid cancer?           | NA |
| Reported in thyroid function?         | NA |

|                |                |
|----------------|----------------|
| y6 - 706.4134+ | y5 - 559.3450+ |
| y4 - 446.2609+ | y3 - 333.1769+ |
| b3 - 318.1812+ |                |

Q16610\_Extracellular matrix protein 1\_ECM1  
Peptide sequence: EVGPPLPQEAVPLQK

|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | Yes |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

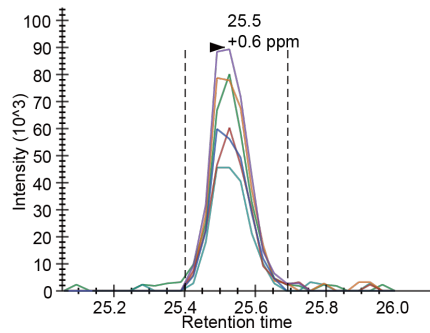
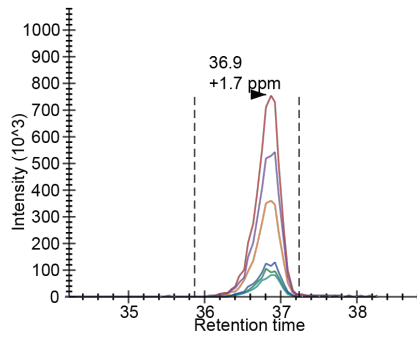
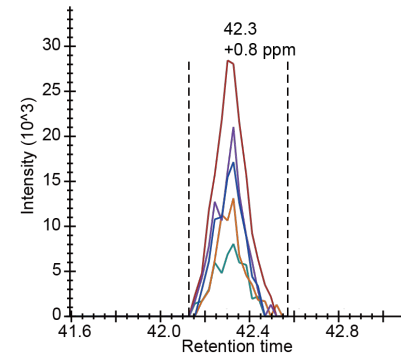
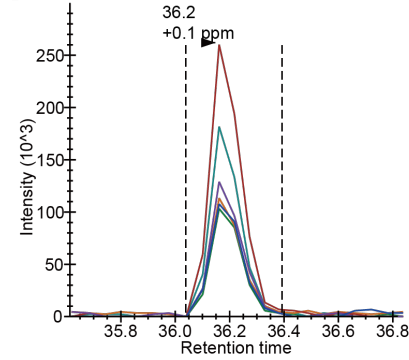
|                  |                  |
|------------------|------------------|
| y11 - 1219.7045+ | y9 - 1009.5677+  |
| y4 - 485.3082+   | y12 - 658.8823++ |
| y9 - 505.2875++  | b6 - 593.3293+   |

P47736\_Rap1 GTPase-activating protein 1  
RAP1GAP  
Peptide sequence: LPYTEGDAQQLQR

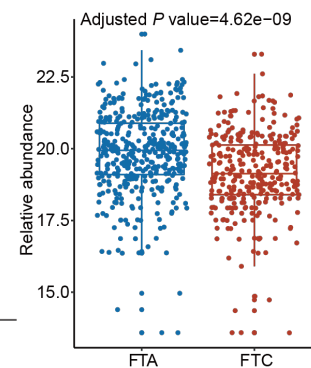
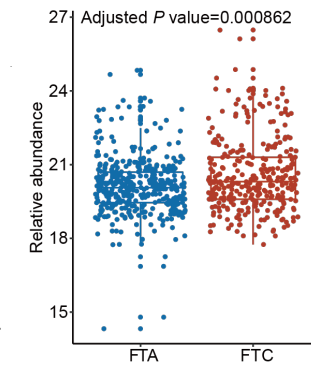
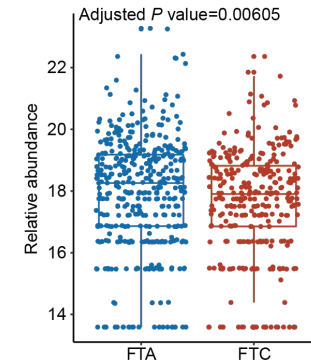
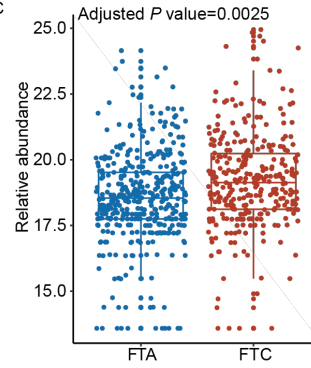
|                                       |     |
|---------------------------------------|-----|
| Reported in follicular thyroid tumor? | NA  |
| Reported in thyroid cancer?           | Yes |
| Reported in thyroid function?         | NA  |

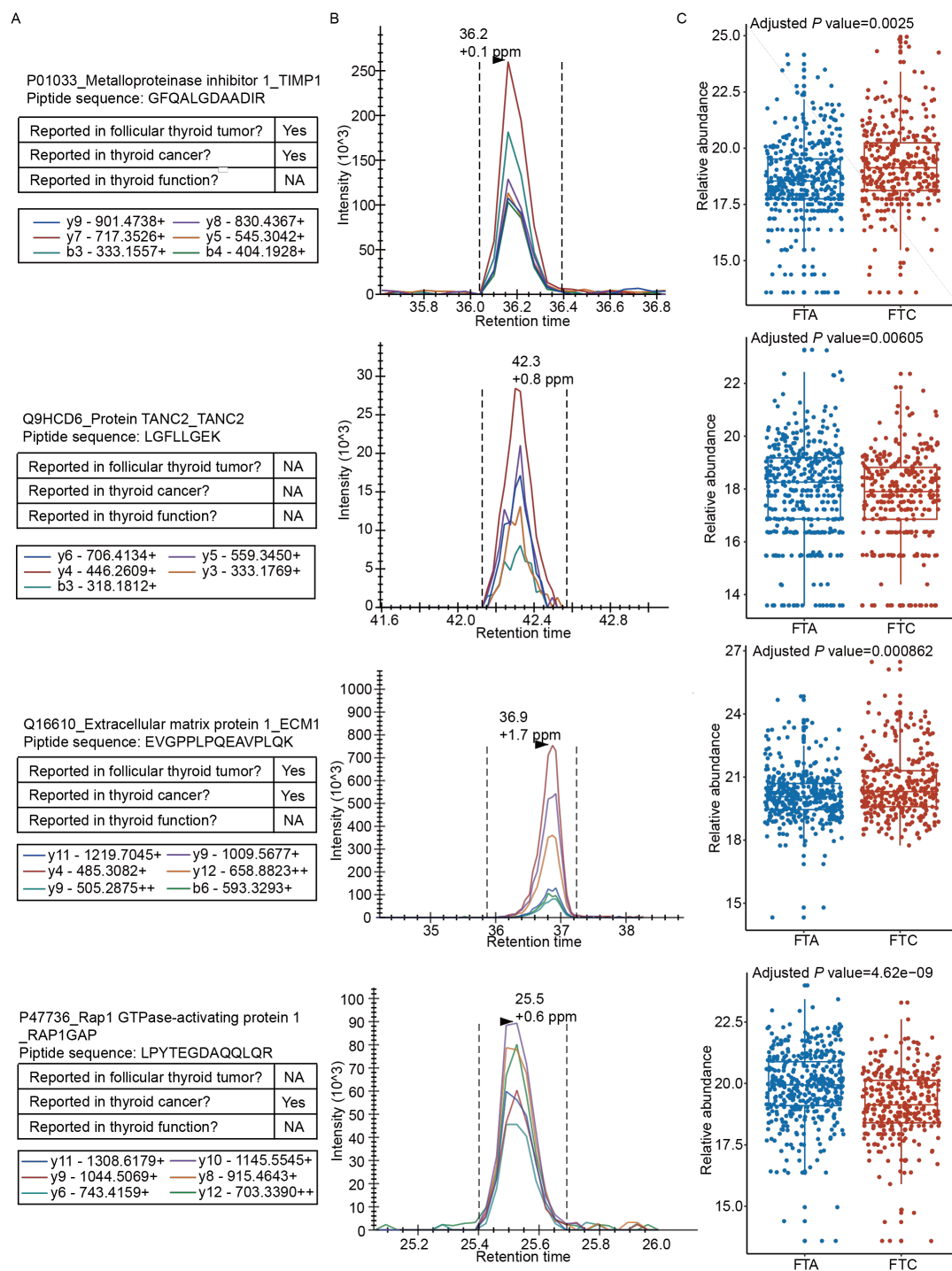
|                  |                  |
|------------------|------------------|
| y11 - 1308.6179+ | y10 - 1145.5545+ |
| y9 - 1044.5069+  | y8 - 915.4643+   |
| y6 - 743.4159+   | y12 - 703.3390++ |

B



C





**Appendix Figure S3. Protein feature characteristics.**

(A) The protein identified by Uniprot ID\_protein name\_gene name with corresponding sequences which detected by targeted proteomics. (B) A chromatographic profile of a representative peptide precursor peak group. (C) Boxplots showing the protein abundance of each sample measured by parallel-reaction monitoring. FTA (n=415) and FTC (n=314). The boxes represent the first and third quartiles, the center

line indicates the median, the whiskers extend to  $\pm 1.5$  times the interquartile range, and the dots represent individual data points.  $P$  values are calculated by two-sided Welch's  $t$ -test.

**Appendix Table S1. Gene mutation frequency in FTA and FTC.**

| Gene name      | FTA   | FTC   |
|----------------|-------|-------|
| <i>NRAS</i>    | 12.4% | 21.6% |
| <i>HRAS</i>    | 7.3%  | 14.1% |
| <i>TERT</i>    | 2.3%  | 18.8% |
| <i>DICER1</i>  | 4.0%  | 10.3% |
| <i>EIF1AX</i>  | 5.8%  | 6.6%  |
| <i>KRAS</i>    | 2.8%  | 3.8%  |
| <i>TSHR</i>    | 3.0%  | 2.8%  |
| <i>TP53</i>    | 0.8%  | 5.6%  |
| <i>BRAF</i>    | 2.5%  | 2.8%  |
| <i>SPOP</i>    | 2.5%  | 1.9%  |
| <i>EZH1</i>    | 2.0%  | 0.0%  |
| <i>CTNNB1</i>  | 0.5%  | 1.9%  |
| <i>GLIS3</i>   | 1.0%  | 1.9%  |
| <i>RET</i>     | 0.5%  | 1.9%  |
| <i>MEN1</i>    | 0.5%  | 1.9%  |
| <i>GNAS</i>    | 0.5%  | 0.5%  |
| <i>KMT2C</i>   | 0.3%  | 1.4%  |
| <i>NOTCH1</i>  | 0.3%  | 1.4%  |
| <i>PIK3CA</i>  | 0.8%  | 0.9%  |
| <i>POR</i>     | 1.0%  | 0.5%  |
| <i>KMT2D</i>   | 0.8%  | 0.5%  |
| <i>PTEN</i>    | 0.3%  | 0.9%  |
| <i>ALK</i>     | 0.8%  | 0.0%  |
| <i>NTRK1</i>   | 0.8%  | 0.0%  |
| <i>RB1</i>     | 0.3%  | 0.9%  |
| <i>STK11</i>   | 0.3%  | 0.9%  |
| <i>APC</i>     | 0.3%  | 0.5%  |
| <i>CDKN2A</i>  | 0.3%  | 0.0%  |
| <i>LRP1B</i>   | 0.3%  | 0.5%  |
| <i>VHL</i>     | 0.0%  | 0.9%  |
| <i>AKT1</i>    | 0.0%  | 0.5%  |
| <i>CDK12</i>   | 0.0%  | 0.5%  |
| <i>EGFR</i>    | 0.3%  | 0.0%  |
| <i>FAM193A</i> | 0.3%  | 0.0%  |
| <i>FGFR4</i>   | 0.0%  | 0.5%  |
| <i>IDH1</i>    | 0.3%  | 0.0%  |
| <i>KIT</i>     | 0.3%  | 0.0%  |
| <i>MET</i>     | 0.3%  | 0.0%  |
| <i>NCOR2</i>   | 0.0%  | 0.5%  |
| <i>ROS1</i>    | 0.3%  | 0.0%  |
| <i>TSC2</i>    | 0.3%  | 0.0%  |

**Appendix Table S2. Twenty-four proteins established associations with thyroid physiology or pathology.**

| Rank | Uniprot ID | Protein name                                        | Gene name      | Follicular tumour related | Thyroid cancer related | Thyroid function related |
|------|------------|-----------------------------------------------------|----------------|---------------------------|------------------------|--------------------------|
| 1    | P22748     | Carbonic anhydrase 4                                | <i>CA4</i>     | Yes                       | -                      | -                        |
| 2    | O00339     | Matrilin-2                                          | <i>MATN2</i>   | -                         | Yes                    | -                        |
| 3    | P36551     | Oxygen-dependent coproporphyrinogen-III oxidase     | <i>CPOX</i>    | -                         | -                      | -                        |
| 4    | P02765     | Alpha-2-HS-glycoprotein                             | <i>AHSG</i>    | -                         | Yes                    | -                        |
| 5    | P16671     | Platelet glycoprotein 4                             | <i>CD36</i>    | -                         | Yes                    | -                        |
| 6    | Q9P2K5     | Myelin expression factor 2                          | <i>MYEF2</i>   | -                         | -                      | -                        |
| 7    | P16949     | Stathmin                                            | <i>STMN1</i>   | -                         | Yes                    | -                        |
| 8    | Q86UX2     | Inter-alpha-trypsin inhibitor heavy chain H5        | <i>ITIH5</i>   | Yes                       | Yes                    | -                        |
| 9    | P15090     | Fatty acid-binding protein                          | <i>FABP4</i>   | Yes                       | Yes                    | Yes                      |
| 10   | P27487     | Dipeptidyl peptidase 4                              | <i>DPP4</i>    | Yes                       | Yes                    | -                        |
| 11   | P61916     | NPC intracellular cholesterol transporter 2         | <i>NPC2</i>    | -                         | Yes                    | -                        |
| 12   | Q9Y6M1     | Insulin-like growth factor 2 mRNA-binding protein 2 | <i>IGF2BP2</i> | -                         | Yes                    | -                        |
| 13   | Q687X5     | Metalloreductase STEAP4                             | <i>STEAP4</i>  | -                         | -                      | -                        |
| 14   | O15460     | Prolyl 4-hydroxylase subunit alpha-2                | <i>P4HA2</i>   | -                         | Yes                    | -                        |
| 15   | P16401     | Histone H1.5                                        | <i>H1-5</i>    | -                         | -                      | -                        |
| 16   | Q86XX4     | Extracellular matrix organizing protein FRAS1       | <i>FRAS1</i>   | -                         | -                      | -                        |
| 17   | P98164     | Low-density lipoprotein receptor-related protein 2  | <i>LRP2</i>    | -                         | Yes                    | Yes                      |
| 18   | Q8N6C5     | Immunoglobulin superfamily member 1                 | <i>IGSF1</i>   | -                         | Yes                    | Yes                      |
| 19   | P29762     | Cellular retinoic acid-binding protein 1            | <i>CRABP1</i>  | Yes                       | Yes                    | -                        |
| 20   | P52926     | High mobility group protein HMGI-C                  | <i>HMGA2</i>   | Yes                       | Yes                    | -                        |
| 21   | P01033     | Metalloproteinase inhibitor 1                       | <i>TIMP1</i>   | Yes                       | Yes                    | -                        |
| 22   | Q9HCD6     | Protein TANC2                                       | <i>TANC2</i>   | -                         | -                      | -                        |
| 23   | Q16610     | Extracellular matrix protein 1                      | <i>ECM1</i>    | Yes                       | Yes                    | -                        |
| 24   | P47736     | Rap1 GTPase-activating protein 1                    | <i>RAP1GAP</i> | -                         | Yes                    | -                        |

**Appendix Table S3. List of 66-gene panel of the thyroid cancer.**

| Gene List     |                |              |               |                |               |
|---------------|----------------|--------------|---------------|----------------|---------------|
| <i>AKT1</i>   | <i>EGFR</i>    | <i>FLT3</i>  | <i>KRAS</i>   | <i>NTRK3</i>   | <i>SMAD4</i>  |
| <i>ALK</i>    | <i>EIF1AX</i>  | <i>GLIS3</i> | <i>LRP1B</i>  | <i>PIK3CA</i>  | <i>SPOP</i>   |
| <i>APC</i>    | <i>EP300</i>   | <i>GNAQ</i>  | <i>MEN1</i>   | <i>PLEKHS1</i> | <i>STK11</i>  |
| <i>ATM</i>    | <i>ERBB4</i>   | <i>GNAS</i>  | <i>MET</i>    | <i>POR</i>     | <i>SUGCT</i>  |
| <i>BANP</i>   | <i>EZH1</i>    | <i>HRAS</i>  | <i>MTOR</i>   | <i>PPARG</i>   | <i>TERT</i>   |
| <i>BRAF</i>   | <i>FAM193A</i> | <i>IDH1</i>  | <i>NCOR2</i>  | <i>PTEN</i>    | <i>TP53</i>   |
| <i>CDK12</i>  | <i>FARSB</i>   | <i>IDH2</i>  | <i>NF1</i>    | <i>PTH</i>     | <i>TRIM61</i> |
| <i>CDKN2A</i> | <i>FGFR1</i>   | <i>KIT</i>   | <i>NF2</i>    | <i>RB1</i>     | <i>TSC2</i>   |
| <i>CHEK2</i>  | <i>FGFR2</i>   | <i>KLK1</i>  | <i>NOTCH1</i> | <i>RBM10</i>   | <i>TSHR</i>   |
| <i>CTNNB1</i> | <i>FGFR3</i>   | <i>KMT2C</i> | <i>NRAS</i>   | <i>RET</i>     | <i>VHL</i>    |
| <i>DICER1</i> | <i>FGFR4</i>   | <i>KMT2D</i> | <i>NTRK1</i>  | <i>ROS1</i>    | <i>ZNF148</i> |

### **Non-contributing We-TEC Investigators**

Bo Huang<sup>1</sup>, Gaosong Wu<sup>2</sup>, Guan Ruan<sup>3</sup>, Guang Chen<sup>4</sup>, Guoyang Wu<sup>5</sup>, Huixiong Xu<sup>6</sup>, James A. Fagin<sup>7</sup>, Jianhua Wang<sup>8</sup>, Jing Liu<sup>9</sup>, Lei Liang<sup>10</sup>, Meiping Shen<sup>11</sup>, Qiushi Zhang<sup>3</sup>, Wen Tian<sup>12</sup>, Xianghui He<sup>13</sup>, Xiao Shi<sup>14</sup>, Yan Li<sup>15</sup>

<sup>1</sup>Liaoning Cancer Hospital & Institute, Cancer Hospital of Dalian University of Technology, Shenyang, China;

<sup>2</sup>Department of Thyroid and Breast Surgery, Zhongnan Hospital of Wuhan University, Wuhan, China;

<sup>3</sup>Westlake Omics (Hangzhou) Biotechnology Co., Ltd., Hangzhou, China.

<sup>4</sup>Department of Thyroid Surgery, General Surgery Center, The First Hospital of Jilin University, Changchun, China;

<sup>5</sup>Department of General Surgery, Zhongshan Hospital Affiliated with Xiamen University, Xiamen, China;

<sup>6</sup>Department of Ultrasound, Zhongshan Hospital, Institute of Ultrasound in Medicine and Engineering, Fudan University, Shanghai, China;

<sup>7</sup>Memorial Sloan Kettering Cancer Center, New York, USA;

<sup>8</sup>Department of Thyroid and Breast Surgery, Affiliated Hospital of Integrated Chinese and Western, Nanjing University of Chinese Medicine, Nanjing, China;

<sup>9</sup>First Hospital of Shanxi Medical University, Thyroid Surgery Department, Taiyuan, China;

<sup>10</sup>Department of Ultrasound, Aerospace Center Hospital, Beijing, China;

<sup>11</sup>Department of General Surgery, the First Affiliated of Nanjing Medical University, Nanjing, China;

<sup>12</sup>Department of Thyroid & Hernia Surgery, Medical Department of General Surgery, The First Medical Center of Chinese People's Liberation Army General Hospital, Beijing, China;

<sup>13</sup>Department of General Surgery, Tianjin Medical University General Hospital, Tianjin, China;

<sup>14</sup>Department of Head and Neck Surgery, Fudan University Shanghai Cancer Centre; Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China;

<sup>15</sup>College of Basic Medical Sciences, Shanghai Jiao Tong University, Shanghai, China;