

# Multi-center multi-omics integration predicts individualized prognosis in medullary thyroid carcinoma

Received: 18 March 2025

Accepted: 3 December 2025

Published online: 14 January 2026

 Check for updates

Yan Zhou  <sup>1,2,3,4,17</sup>, Yingrui Wang  <sup>2,3,4,17</sup>, Xiao Shi  <sup>5,6,17</sup>, Jiatong Wang  <sup>2,3,4,17</sup>, Zelin Zang  <sup>7,17</sup>, Likun Zhang  <sup>8,17</sup>, Zhiqiang Gui <sup>9</sup>, Xue Cai <sup>2,3,4</sup>, Pingping Hu <sup>2,3,4</sup>, Jiaxi Wang  <sup>10</sup>, Hanqing Liu <sup>11</sup>, Zhihong Wang <sup>9</sup>, Haixia Guan  <sup>12</sup>, Xiaohong Wu <sup>13</sup>, Cenkai Shen <sup>5,6</sup>, Yi He <sup>14</sup>, Mo Li <sup>15</sup>, Hao Zhang  <sup>9</sup>, Jianbiao Wang <sup>16</sup>, Yijun Wu  <sup>11</sup>, Chuang Chen  <sup>10</sup>, Yi Zhu  <sup>2,3,4</sup>, Yaoting Sun  <sup>2,3,4</sup> , Zhiyan Liu  <sup>8</sup> , Yu Wang <sup>5,6</sup>  & Tiannan Guo  <sup>1,2,3,4</sup> 

Medullary thyroid carcinoma (MTC) is a rare, aggressive neuroendocrine tumor with limited treatment options and frequent recurrence. Comprehensive recurrence risk stratification remains lacking. Here, we profile 482 MTC samples from 452 patients across ten Chinese clinical centers, identifying 10,092 proteins and mutations in 87.0% of patients. Clinically, MTC grading, concurrent papillary thyroid carcinoma, and lymph node metastasis are significant recurrence risk factors, whereas at the genetic level, *RET* M918T and *RET* S891A mutations are correlated with high recurrence risk in sporadic and hereditary MTC, respectively. Ubiquitinomics show downregulated E3 ligases CUL4B and TRIM32 are associated with structural recurrence. We define three molecular subtypes with distinct outcomes and present an integrative machine learning model combining clinical, genomic, and proteomic features, validated in an independent test dataset of 105 patients and a published dataset. This multi-center, multi-omics study enhances the understanding of MTC heterogeneity and facilitates personalized patient management.

Medullary thyroid carcinoma (MTC) is a rare neuroendocrine tumor arising from parafollicular C cells<sup>1</sup>. The incidence of MTC represents only 2% of all thyroid cancers<sup>2,3</sup>, yet it accounts for 8% of thyroid cancer-related deaths. MTC is characterized by aggressive behavior and elevated metastatic potential, leading to a median disease-specific survival of 8.6 years<sup>4</sup>. At primary diagnosis, 75% of patients with MTC had metastases in cervical lymph nodes, while 10–15% present with distant metastasis<sup>5,6</sup>. The disease's inherent resistance to radioiodine therapy significantly restricts viable treatment options. The primary treatment for MTC is surgical intervention, specifically total thyroidectomy and bilateral central neck lymph node dissection<sup>7</sup>. However, postoperative recurrence remains a significant issue, with a reported reoperation rate of 16.3% and a median time to reoperation of

6.4 months<sup>8</sup>. Disease recurrence significantly affects disease-free survival and quality of life, underscoring the importance of effective risk stratification to predict outcomes and optimize long-term follow-up strategies for improved results.

The current postoperative prognostic assessment of MTC primarily relies on the TNM staging system, which evaluates maximum primary tumor size, extrathyroidal extension (ETE), lymph node metastasis (LNM), and distant metastasis. However, this system does not incorporate other critical prognostic factors, such as age, sex, heredity, and postoperative levels of calcitonin and carcinoembryonic antigen<sup>9–11</sup>. Consequently, there is a need for a more comprehensive tool that integrates these diverse factors to accurately assess prognostic risk in MTC patients. Xu et al. recently introduced the

A full list of affiliations appears at the end of the paper.  e-mail: [sunyaojing@westlake.edu.cn](mailto:sunyaojing@westlake.edu.cn); [zhiyanliu@shsmu.edu.cn](mailto:zhiyanliu@shsmu.edu.cn); [neck130@sina.com](mailto:neck130@sina.com); [guotiannan@westlake.edu.cn](mailto:guotiannan@westlake.edu.cn)

international MTC grading system (IMTCGS), which incorporated measures of proliferative activity, including the mitotic index and/or Ki67 proliferative index, as well as tumor necrosis<sup>12</sup>. This grading system stratifies MTC into high and low grades, with high-grade tumors demonstrating worse disease-specific survival and higher recurrence rates at both local and distant sites. While the system has been validated in several cohorts from Europe, the United States and Australia, its generalizability to large Asian populations remains uncertain<sup>12–16</sup>.

Genomic and transcriptomic technologies have been extensively utilized to investigate MTC prognosis, revealing genomic features associated with heredity<sup>17,18</sup>, IMTCGS grade<sup>14</sup>, and clinical outcomes<sup>11</sup>. The *RET* mutation plays a pivotal role in both hereditary and sporadic MTC. Hereditary MTC involves germline *RET* mutations, accounting for approximately 25% of all MTC cases. For sporadic MTC, somatic *RET* M918T mutations are linked to a worse prognosis<sup>11</sup>. Notably, 10–20% of sporadic MTC cases lack known driver mutations, necessitating other technologies to elucidate the underlying pathogenesis<sup>11,17</sup>.

Proteins represent the final product of gene expression, but their abundances do not always align with DNA or RNA levels. Mass spectrometry (MS)-based proteomic analysis<sup>19</sup> remains underexplored despite its potential to enhance prognostic assessment in MTC. One study involving three MTC patients demonstrated that decreased fibronectin 1 (FNI) expression was correlated with reduced progression-free survival<sup>20</sup>. In a prior study involving 102 Chinese patients with MTC, three proteomic subtypes with distinct biological traits and prognoses were identified<sup>21</sup>. However, these studies are limited by small sample sizes and/or a lack of validation in independent test cohorts. Proteomic analyses focusing on structural recurrence in MTC are particularly scarce, yet such research could illuminate the molecular mechanisms associated with MTC aggressiveness and metastatic potential.

Ubiquitination is a versatile post-translational modification in which the 76-amino-acid ubiquitin protein is covalently attached to lysine residues of substrate proteins. Substrate proteins tagged by ubiquitin will be recognized and degraded by proteasome, through which to regulate protein stability, trafficking, and signaling transduction<sup>22</sup>. Ubiquitination is reversible and dynamically regulated by E3 ubiquitin ligases and deubiquitinases (DUBs)<sup>23</sup>. Increasing evidence indicates that perturbations of the ubiquitin–proteasome system intersect with *RET* signaling, cell-cycle control, and apoptosis, highlighting ubiquitination as a mechanistically relevant layer in MTC<sup>24–27</sup>. MS-based ubiquitinomics enables proteome-wide site mapping of ubiquitination by identifying the characteristic di-glycine (K-GG) remnant that remains on modified lysines after tryptic digestion<sup>28–30</sup>. Proteome-scale studies on ubiquitination dynamics in MTC remain limited. Such investigations may provide insights into potential therapeutic strategies, including the development of proteasome or ubiquitin pathway inhibitors<sup>31</sup>.

In this study, we systematically analyze data from 482 samples obtained from 452 Chinese MTC patients from 10 clinical centers. We integrate proteomics, ubiquitinomics, gene panel data, and clinical information and identify three molecular subtypes based on 52 proteins, each demonstrating distinct clinical, prognostic, and molecular characteristics. Furthermore, we develop a predictive model combining proteomic and clinical features to assess the risk of structural recurrence in MTC.

## Results

### Clinicopathological characteristics of MTC patients

After careful curation, we procured 377 ( $N=347$ ) and 105 ( $N=105$ ) eligible formalin-fixed, paraffin-embedded (FFPE) samples in the discovery and the independent test datasets, respectively (Fig. 1a). The study design is summarized in Fig. 1b. We collected 12 clinical indicators, sequenced a panel of 28 genes, and quantified the proteome

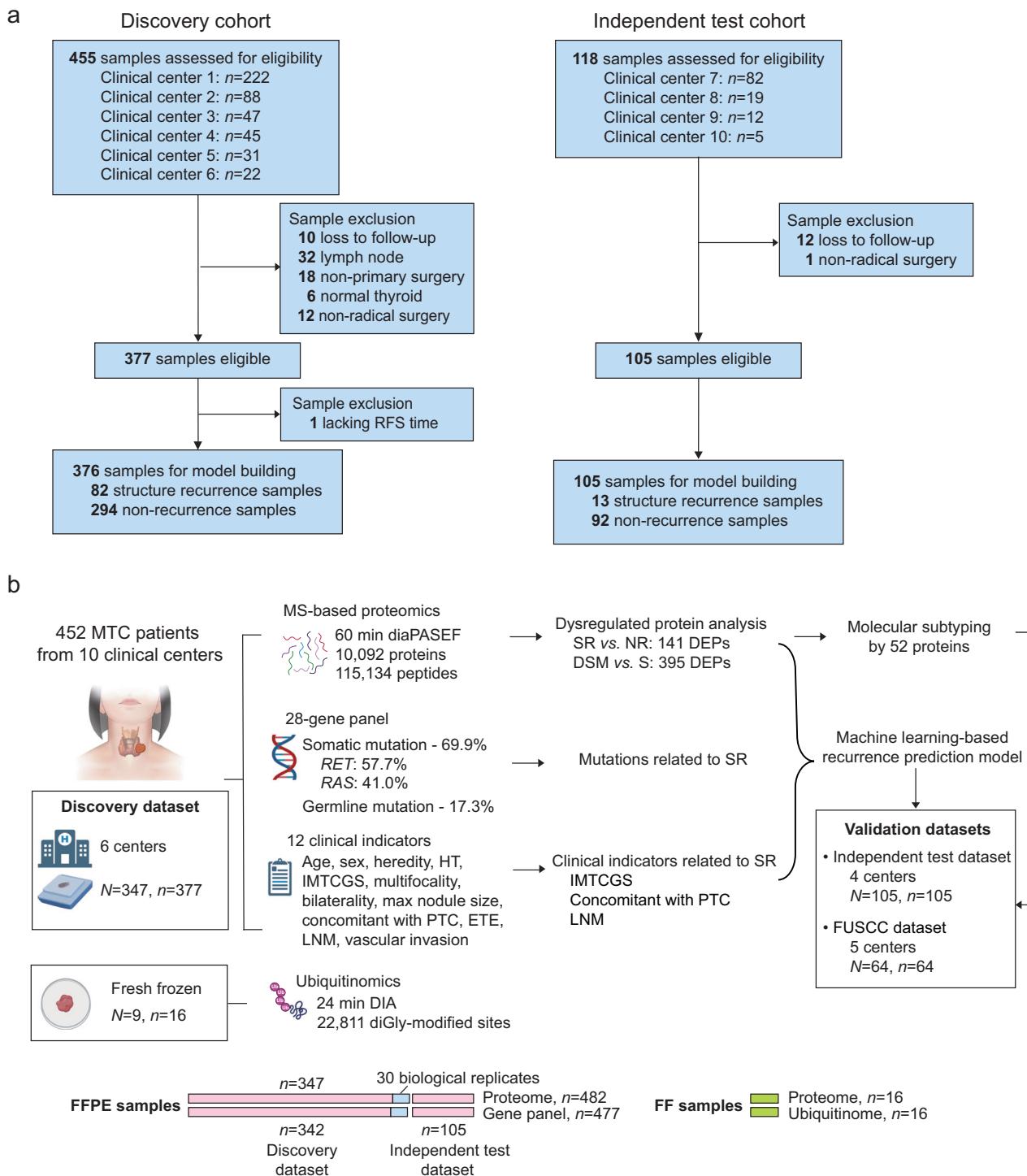
using pressure cycling technology (PCT) for sample preparation followed by data-independent acquisition (DIA) MS<sup>32,33</sup>. The baseline clinical characteristics are summarized in Table 1. The mean age at primary surgery for all patients in the study is  $49.7 \pm 12.2$  (mean  $\pm$  standard deviation) years, with a mean maximum nodule size of  $2.2 \pm 1.4$  cm. Male patients comprise 45.8% of the cohort, and hereditary MTC represents 17.0% of cases. During the follow-up period in the discovery dataset, 20.7% of MTC patients experienced structural recurrence (SR), and 3.5% died from MTC-specific causes. The mean time from primary surgery to SR or disease-specific mortality (DSM) is 52.5 and 78.6 months, respectively. In comparison, the corresponding control groups have mean follow-up times of 102.0 and 106.5 months, respectively. Compared to non-recurrence (NR) cases, patients with SR exhibit larger maximum nodule sizes, a higher proportion of males, and more frequent multifocal and bilateral lesions, extrathyroidal extension (ETE), and lymph node metastasis (LNM). Similarly, the DSM group has larger maximum nodule sizes, higher tumor grades, and more ETE and LNM than the survival (S) group.

### Association of *RET* and *RAS* mutations with clinicopathologic characteristics and disease prognosis

In addition, we explored the clinical features stratified by *RET* and *RAS* mutation status. Gene sequencing revealed germline *RET* mutations in 17.3% of MTC patients, somatic mutations in 69.9%, and no detectable mutations in 12.8% (Fig. 1b). Patients with germline *RET* mutations were younger than those with *RET/RAS* wild-type patients ( $P=0.003$ , Table 2). Five patients were not sequenced for mutation, so they were not included in the table. Both germline and somatic *RET* mutations were associated with higher incidences of ETE and LNM compared to *RET/RAS* wild-type cases (germline:  $P=0.038$  (ETE) and  $P=0.017$  (LNM); somatic:  $P=0.011$  (ETE) and  $P=0.013$  (LNM)). *RET* somatic mutations present a higher rate of structural recurrence than the *RET/RAS* wild-type group (29.0% vs. 14.9%). However, *RET* or *RAS* mutations do not significantly correlate with DSM. There are no statistically significant differences between *RAS* mutations and *RET/RAS* wild-type cases regarding clinical characteristics or prognosis. In summary, our results show that *RET/RAS* mutation is not related to SR or DSM when specific mutation sites are not considered.

To further investigate the recurrence risk associated with specific mutation sites, Kaplan-Meier (K-M) survival analyses were performed. Among sporadic MTC cases, patients harboring the *RET* M918T mutation exhibit a higher recurrence risk than those without the mutation ( $P=0.0013$ , Fig. 2b). In hereditary MTC, the *RET* S891A mutation presents a higher recurrence risk, whereas the *RET* C634 mutation is associated with a lower recurrence risk ( $P=0.0015$ , Fig. 2c).

A comparative analysis was subsequently conducted to evaluate the frequencies of genetic mutation sites between our dataset of Chinese patients (Westlake) and those of Western populations (Pisa<sup>17</sup>, MSKCC<sup>11</sup>). Among sporadic MTC patients, the *RET* M918T, *HRAS*, and *RETC634* mutations are the three most prevalent mutation sites across the three datasets (Supplementary Fig. 1a, Supplementary Data 1). Notably, the Westlake dataset exhibits an *HRAS* mutation frequency higher than the other two datasets (30.0% vs. 17.1% and 14.8%). For hereditary MTC cases, the *RET* C634 mutation is the most common in the Westlake dataset, accounting for 44.1%, compared to 27.5% in the MSKCC dataset (Supplementary Fig. 1b). Conversely, the *RET* C609Y mutation, which is the most prevalent in the MSKCC dataset, is rare in the Westlake cohort (35.0% vs. 1.7%). Additionally, the *RET* S891A mutation is identified as the second most frequent mutation in the Westlake dataset (11.9%) but is absent in the MSKCC dataset. In conclusion, Chinese MTC patients have a unique mutation pattern compared to Western patients, especially with more *HRAS* and germline *RET* C634 mutations and fewer germline *RET* C609Y mutations.



**Fig. 1 | Study Design.** **a** Sample exclusion criteria; **b** The diagram shows the study workflow, including sample collection, MS-based proteomics, ubiquitinomics, gene-panel-based sequencing, and data analysis (Created in BioRender. lab, g, (2025) <https://BioRender.com/vsdjlxg>). IMTCGS International Medullary Thyroid

Carcinoma Grading System, HT Hashimoto's thyroiditis, PTC papillary thyroid carcinoma, ETE extrathyroidal extension, LNM lymph node metastasis, SR structural recurrence, NR non-recurrence, DSM disease-specific mortality, S survival.

### Clinical, genomic, and proteomic characterization of MTC patients with different prognoses

To identify risk factors related to the disease prognosis of MTC, univariate Cox proportional hazards (CoxPH) models were constructed for each of the 12 clinicopathological features. The analysis identifies sex, IMTCGS, max nodule size, multifocality, concurrence of papillary thyroid carcinoma (PTC), vascular invasion, ETE, and LNM as risk factors associated with SR (Supplementary Table 1). Subsequently, a

multivariate CoxPH model was employed to assess these 12 factors collectively (Fig. 2a). The results indicate that IMTCGS ( $P=0.0096$ , hazard ratio, HR = 4.5), concurrence of PTC ( $P=1.83e-04$ , HR = 5.6), and LNM ( $P=1.65e-04$ , HR = 16.2) are independent risk factors associated with SR in MTC.

Using parallel accumulation-serial fragmentation (PASEF) combined with DIA techniques<sup>34</sup>, a total of 115,134 peptides and 10,092 proteins were quantified from 377 samples in the discovery dataset.

**Table 1 | Clinical characteristics of the MTC discovery dataset**

Clinical characteristics	All cases, n = 347 (%)	Outcome 1		P value	Outcome 2		P value
		Structure recurrence, n = 72 (%)	Non-recurrence, n = 275 (%)		Death, n = 12 (%)	Survival, n = 335 (%)	
Age at surgery (years)	49.7 ± 12.2	48.4 ± 10.6*	50.1 ± 12.5	0.311	54.8 ± 10.4	49.6 ± 12.2	0.146
Max nodule size (cm)	2.2 ± 1.4	2.6 ± 1.7	2.1 ± 1.3	<b>0.017</b>	4.6 ± 2.3	2.1 ± 1.3	<b>2.2e-4</b>
Sex				<b>0.033</b>			0.767
Male	159 (45.8)	41 (56.9)	118 (42.9)		6 (50.0)	153 (45.7)	
Female	188 (54.2)	31 (43.1)	157 (57.1)		6 (50.0)	182 (54.3)	
Heredity				0.536			0.673
Hereditary	59 (17.0)	14 (19.4)	45 (16.4)		1 (8.3)	58 (17.3)	
Sporadic	288 (83.0)	58 (80.6)	230 (83.6)		11 (91.7)	277 (82.7)	
IMTCGS				0.333			<b>0.001</b>
High	14 (5.1)	5 (8.3)	9 (4.2)		4 (40.0)	10 (3.8)	
Low	262 (94.9)	55 (91.7)	207 (95.8)		6 (60.0)	256 (96.2)	
Unknown	71						
Hashimoto's thyroiditis				0.679			1.000
No	257 (84.3)	55 (85.9)	202 (83.8)		10 (83.3)	247 (84.3)	
Yes	48 (15.7)	9 (14.1)	39 (16.2)		2 (16.7)	46 (15.7)	
Unknown	42						
Multifocality				<b>3.5e-4</b>			0.781
No	263 (75.8)	43 (59.7)	220 (80.0)		10 (83.3)	253 (75.5)	
Yes	84 (24.2)	29 (40.3)	55 (20.0)		2 (16.7)	82 (24.5)	
Bilaterality				<b>0.043</b>			1.000
No	296 (85.3)	56 (77.8)	240 (87.3)		10 (83.3)	286 (85.4)	
Yes	51 (14.7)	16 (22.2)	35 (12.7)		2 (16.7)	49 (14.6)	
Vascular invasion				0.074			0.120
No	210 (84.7)	44 (77.2)	166 (86.9)		7 (63.6)	203 (85.7)	
Yes	38 (15.3)	13 (22.8)	25 (13.1)		4 (36.4)	34 (14.3)	
Unknown	99						
Concurrence with PTC				0.191			0.574
No	317 (91.4)	63 (87.5)	254 (92.4)		12 (100.0)	305 (91.0)	
Yes	30 (8.6)	9 (12.5)	21 (7.6)		0 (0.0)	30 (9.0)	
Extrathyroidal extension				<b>4.0e-4</b>			<b>0.002</b>
No	278 (80.1)	47 (65.3)	231 (84.0)		5 (41.7)	273 (81.5)	
Yes	69 (19.9)	25 (34.7)	44 (16.0)		7 (58.3)	62 (18.5)	
Lymph node metastasis				<b>7.3e-11</b>			<b>0.003</b>
No	146 (42.1)	6 (8.3)	140 (50.9)		0 (0.0)	146 (43.6)	
Yes	201 (57.9)	66 (91.7)	135 (49.1)		12 (100.0)	189 (56.4)	
Biochemical recurrence				<b>1.0e-19</b>			<b>0.013</b>
No	167 (53.5)	3 (4.5)	164 (66.9)		1 (10.0)	166 (55.0)	
Yes	145 (46.5)	64 (95.5)	81 (33.1)		9 (90.0)	136 (45.0)	
Unknown	35						
Disease specific mortality				<b>6.5e-9</b>			
No	335 (96.5)	61 (84.7)	274 (99.6)				
Yes	12 (3.5)	11 (15.3)	1 (0.4)				
Follow-up <sup>RFS</sup> (months)	91.8 ± 54.2	52.5 ± 45.4	102.0 ± 51.7	<b>1.8e-13</b>			
Follow-up <sup>DSS</sup> (months)	105.5 ± 52.6				78.6 ± 41.4	106.5 ± 52.7	0.063

IMTCGS International Medullary Thyroid Carcinoma Grading System, PTC papillary thyroid carcinoma, RFS recurrence-free survival, DSS disease-specific survival.

P values were calculated by independent double-sided Student's t-test for age, nodule size and follow-up time, and Chi-square test for the other variables. \*marks variables that did not conform to the normal distribution, which were analyzed using the nonparametric Mann-Whitney U test. Bold values mark statistical significance.

After excluding proteins with over 90% missing values, the final dataset comprises 9380 proteins. Pearson correlations for pooled samples and each pair of biological and technical replicates are above 0.92, showing the high quality and reproducibility of the proteomic data (Supplementary Fig. 2a). Given the multi-center origin of the data, batch

effects were evaluated and minimized. Visualization of the corrected data by the t-distributed stochastic neighbor embedding (t-SNE) method reveals no significant batch effects (Supplementary Fig. 2b, c).

*RET* and *RAS* are two of the most frequently mutated genes in MTC. We examined the effects of these mutations on proteins in their

**Table 2 | Clinical characteristics between *RET* and *RAS* mutations in the discovery dataset**

Clinical characteristics	All cases, n = 342 (%)	Genotype				P value <sup>a</sup>	P value <sup>b</sup>	P value <sup>c</sup>
		RET germline, n = 59 (%)	RET somatic, n = 138 (%)	RAS somatic, n = 98 (%)	RET/RAS wildtype, n = 47 (%)			
Age at surgery (years)	49.8 ± 12.1	43.5 ± 11.9*	51.1 ± 11.1*	51.4 ± 12.7*	50.5 ± 11.5	<b>0.003</b>	0.750	0.668
Max nodule size (cm)	2.2 ± 1.4	2.2 ± 1.4	2.3 ± 1.5	2.2 ± 1.3	2.0 ± 1.6	0.210	0.095	0.146
Sex						0.527	0.904	0.920
Male	157 (45.9)	24 (40.7)	66 (47.8)	45 (45.9)	22 (46.8)			
Female	185 (54.1)	35 (59.3)	72 (52.2)	53 (54.1)	25 (53.2)			
Extrathyroidal extension						<b>0.038</b>	<b>0.011</b>	0.256
No	273 (79.8)	45 (76.3)	102 (73.9)	83 (84.7)	43 (91.5)			
Yes	69 (20.2)	14 (23.7)	36 (26.1)	15 (15.3)	4 (8.5)			
Lymph node metastasis						<b>0.017</b>	<b>0.013</b>	0.549
No	142 (41.5)	19 (32.2)	48 (34.8)	49 (50.0)	26 (55.3)			
Yes	200 (58.5)	40 (67.8)	90 (65.2)	49 (50.0)	21 (44.7)			
Structural recurrence						0.257	0.055	0.531
No	270 (78.9)	45 (76.3)	98 (71.0)	87 (88.8)	40 (85.1)			
Yes	72 (21.1)	14 (23.7)	40 (29.0)	11 (11.2)	7 (14.9)			
Disease specific mortality						1.000	0.537	1.000
No	330 (96.5)	58 (98.3)	130 (94.2)	96 (98.0)	46 (97.9)			
Yes	12 (3.5)	1 (1.7)	8 (5.8)	2 (2.0)	1 (2.1)			

<sup>a</sup>The P value represents the result of the comparison between the *RET* germline and *RET*/RAS wildtype groups. <sup>b</sup>The P value represents the result of the comparison between the *RET* somatic and *RET*/RAS wildtype groups. <sup>c</sup>The P value represents the result of the comparison between the *RAS* somatic and *RET*/RAS wildtype groups.

P values were calculated by independent double-sided Student's t test for age and nodule size, and Chi-square test for the other variables. \*marks variables that did not conform to the normal distribution, which were analyzed using the nonparametric Mann-Whitney U test. Bold values mark statistical significance.

respective pathways. *RET* mutations led to an increased abundance of *RET* protein (Supplementary Fig. 3), consistent with previous findings<sup>35</sup>. In the RAS pathway, proteins like neurofibromin 1 (NF1) and protein sprouty homolog 4 (SPRY4) were regulated oppositely in *RET* and *RAS* mutation patients. NF1 and SPRY4 were downregulated in the *RET* mutation patients but upregulated in the *RAS* mutation patients. A previous study suggest that low expression of NF1 could serve as a potential biomarker for high recurrence risk in MTC<sup>21</sup>, while SPRY4 has been identified as a candidate susceptibility gene for familial non-medullary thyroid cancer<sup>36</sup>. These proteins may contribute to recurrence risk differentiation between *RET* and *RAS* mutation patients.

By exploring the expression of reported markers of MTC<sup>20,21</sup>, we find that the calcitonin gene-related peptide 1(CGRP1,  $P = 0.021$ ) and NF1 ( $P = 2.8e-05$ ) are differentially expressed between the SR and NR groups (Supplementary Fig. 4a). CGRP1 and NF1 are also differentially expressed between DSM and S groups (Supplementary Fig. 4b).

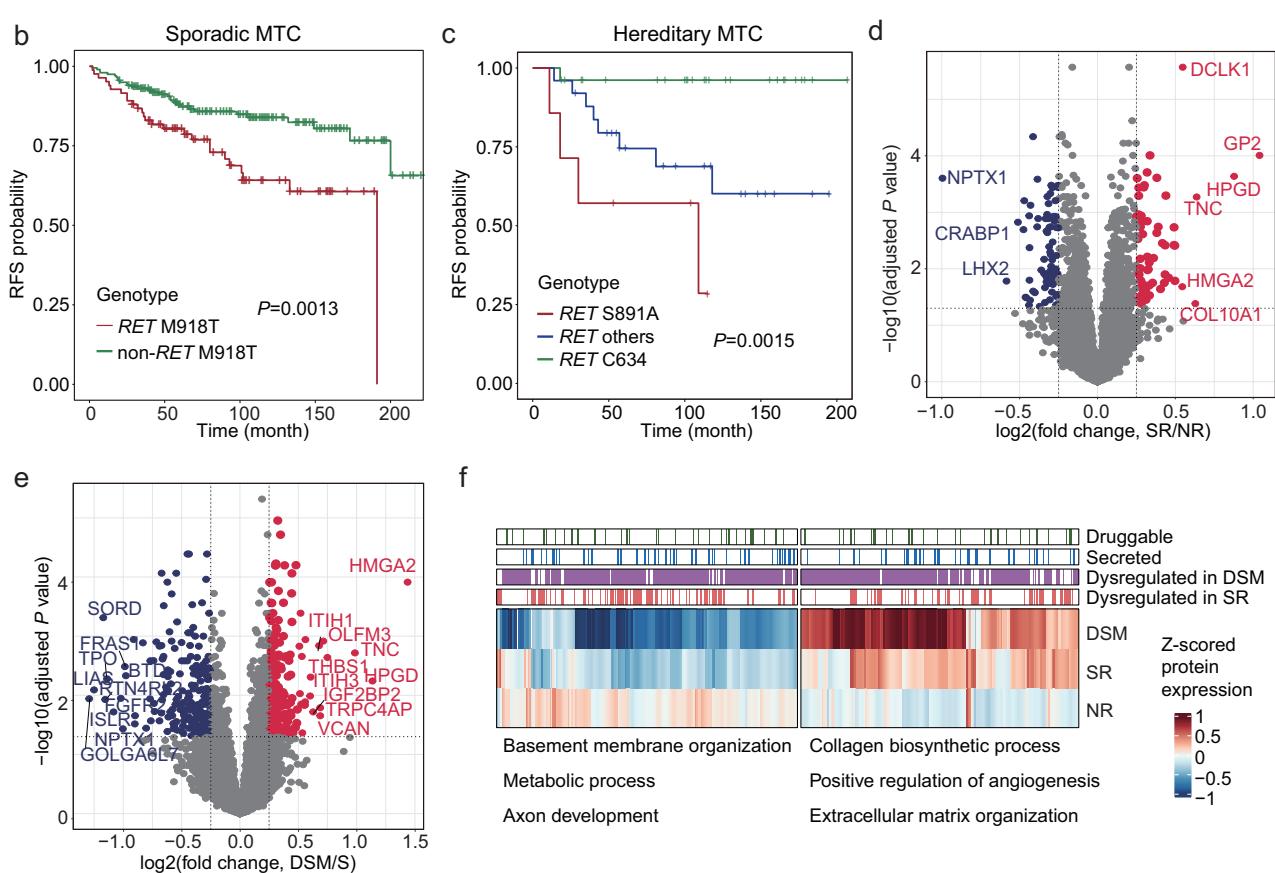
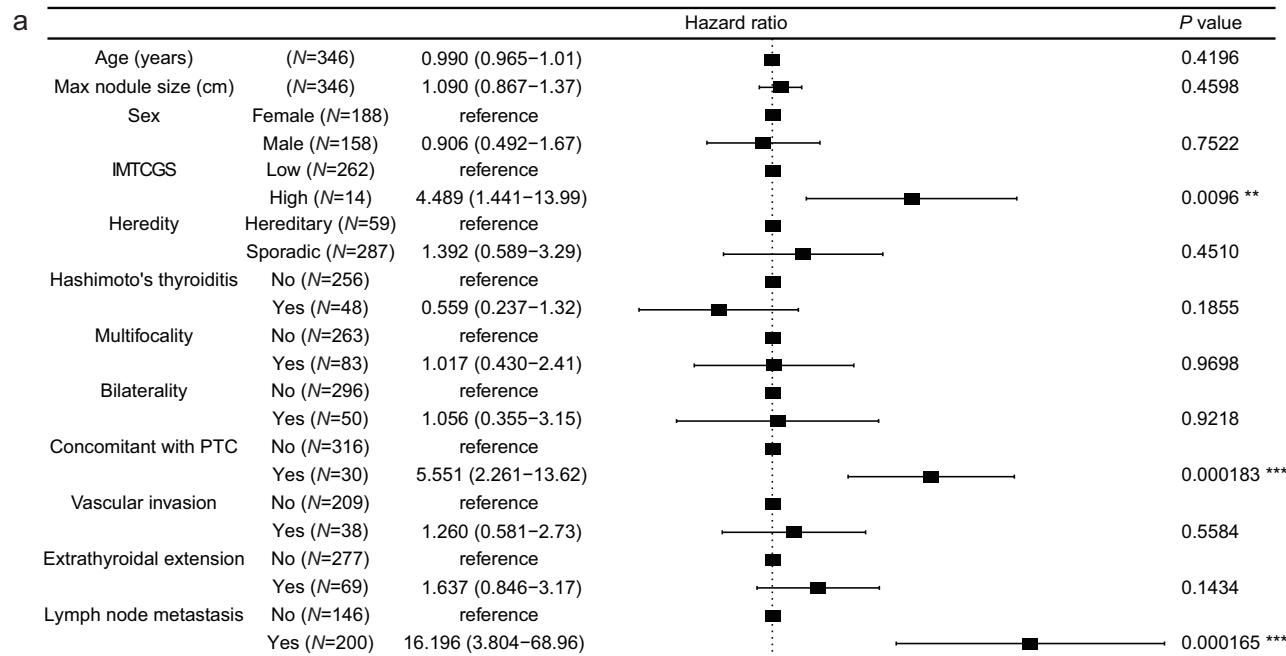
To further investigate expression differences between MTC prognostic groups in the proteomic scale, we analyzed proteins with dysregulated expression in SR and DSM. The volcano plot shows 141 differentially expressed proteins (DEPs) that are significantly increased (in red) or decreased (in blue) in patients with structural recurrence compared to those without recurrence, highlighting potential biomarkers related to recurrence risk (Fig. 2d, Supplementary Fig. 5a). Similarly, there are 395 DEPs between DSM and S groups of MTC patients (Fig. 2e, Supplementary Fig. 5b). We further examined whether the DEPs varied between male and female patients. Only four upregulated proteins and two downregulated proteins were shared between both sexes (Supplementary Fig. 6): pancreatic secretory granule membrane major glycoprotein GP2 (GP2), 15-hydroxyprostaglandin dehydrogenase (HPGD), tenascin C (TNC), and serine/threonine-protein kinase DCLK1 (DCLK1) (upregulated), as well as neuronal pentraxin-1 (NPTX1) and LIM/homeobox protein Lhx2 (LHX2) (downregulated).

Interestingly, DEPs up-regulated in SR compared to NR showed even higher expression in DSM, while down-regulated DEPs exhibited a progressive decrease, revealing a continuous expression trend across the NR-SR-DSM progression (Fig. 2f). Each column of the heatmap represents a DEP, with red indicating higher and blue indicating lower expression levels. The top color bar marks druggable or secreted proteins, suggesting potential molecular markers and therapeutic targets relevant to disease progression. Pathways enriched at the bottom right highlight that proteins positively associated with worse prognosis are related to extracellular matrix (ECM) remodeling, collagen synthesis, and angiogenesis. In contrast, those negatively associated are related to basement membrane reorganization and metabolism (Fig. 2f, Supplementary Fig. 5c). Transforming growth factor beta receptor 2 (TGFBR2), cyclin-dependent kinase inhibitor 1A (CDKN1A), erb-b2 receptor tyrosine kinase 4 (ERBB4), and ERBB2 are predicted to be upstream regulators of dysregulated proteins in SR samples (Supplementary Fig. 5d). These findings suggest that different prognosis groups have various proteomic patterns, especially the dysregulated functions happening in the ECM, which may lead to SR and DSM.

### Characterization of the MTC ubiquitinome landscape

Proteomic analysis revealed pathways associated with different prognoses in MTC, which may also be affected by post-translational modifications. Previous studies have reported that E3 ligases, such as HUWE1, and DUBs, including USP9X and UBP7, exhibit differential expressions and oncogenic functions in PTC primary and metastatic tumors<sup>37,38</sup>. Building on these findings, we further characterized the ubiquitinome of MTC to gain insights into the regulation of recurrence.

We collected fresh-frozen tissues of nine MTC samples (four NR and five SR), and seven normal adjacent thyroid (NAT) samples from clinical center 1 (Supplementary Table 2), and quantified both the



**Fig. 2 | Clinical, genomic, and proteomic profiling between patients with different prognoses.** **a** The forest plot showed the results of a multivariate Cox proportional hazards (CoxPH) regression of 12 clinical risk factors. N represents the number of patients. MTC recurrence risks were represented by hazard ratios (square) and 95% confidence intervals (whisker). P values were calculated under the Cox proportional hazard model. \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ . Kaplan–Meier survival curves for sporadic (**b**) and hereditary (**c**) MTC showed different risk levels of *RET* mutations (*RET* M918T,  $n = 84$ ; non-*RET* M918T,  $n = 199$ ; *RET* S891A,  $n = 7$ ; *RET* C634,  $n = 26$ ; *RET* others  $n = 26$ ). P values were calculated using the log-rank test. Differentially expressed proteins (DEPs) were shown in the volcano plots.

**d** structural recurrence (SR) vs. non-recurrence (NR) samples and **e** disease-specific mortality (DSM) vs. survival samples. P values were calculated by empirical Bayes moderated  $t$  test (limma). Cutoff:  $|\log_2(\text{fold change, FC})| > 0.25$ , Benjamini-Hochberg (B-H) adjusted  $P < 0.05$ . Proteins with  $|\log_2(\text{FC})| > 0.5$  were labeled with protein names. **f** The heatmap illustrated the results of functional analysis and DEPs across different prognostic groups. The protein expression was represented by the median value of the protein in each group. Color bars labeled the DEPs in SR and DSM groups, secreted proteins, and druggable proteins. Sample size in **d,e,f**: NR,  $n = 294$ ; SR,  $n = 83$ ; S,  $n = 358$ ; DSM,  $n = 19$ .

global proteome and ubiquitinome. In total, 22,811 diGly-modified sites and 6505 diGly-modified proteins were identified. After filtering for sites with < 75% missing values, 7090 diGly-modified sites (5925 in NAT, 6517 in NR, 6780 in SR) and 3161 diGly-modified proteins (2719 in NAT, 2968 in NR, 3086 in SR) remained for downstream analysis.

Although NR and SR showed extensive overlap in diGly-modified sites and proteins, a subset of sites was uniquely detected in only one group (Fig. 3a), suggesting potential roles in recurrence. The numbers of diGly-modified sites and proteins were both higher in MTC than in NAT, whereas NR and SR showed no significant difference (Fig. 3b). t-SNE analysis of all modified sites clearly separated MTC from NAT, whereas NR and SR clustered closely (Fig. 3c).

Ubiquitin can be attached to a substrate to form mono-ubiquitination, or multiple ubiquitins can be linked through one of its seven lysine residues or the N-terminal methionine to generate polyubiquitin chains. Differently branched ubiquitin chains encode distinct signals that mediate diverse cellular outcomes<sup>39</sup>. By quantifying seven diGly-modified lysine residues (K6, K11, K27, K29, K33, K48, K63) of ubiquitin, we found that K33-linked sites were detected only in MTC tissues but absent from NAT, implicating the potential role of K33 linkage in tumorigenesis (Fig. 3d). In addition, K11 ubiquitination was significantly reduced in SR compared with NR. K11-linked chains are best known for their role in cell-cycle regulation, particularly during mitosis, and act as degradation signals similar to K48 linkage<sup>40,41</sup>.

To further investigate the role of ubiquitination in MTC development and recurrence, we analyzed diGly-modified sites with significantly different intensities in both MTC *vs.* NAT and SR *vs.* NR. In total, 1067 sites were differentially regulated between MTC and NAT groups (Supplementary Data 2). By contrast, only 15 sites exhibited significant differences between SR and NR (Fig. 3e, Supplementary Data 3), with the enriched pathways shown in Fig. 3f. The label in the volcano plot shows the increasing diGly-modified sites (red) and decreasing diGly-modified sites (blue), with the number representing the position of amino acid at the protein where the ubiquitin is at. These enriched pathways were primarily related to intracellular and transmembrane transport, processes whose aberrant regulation has been reported to enhance the invasiveness and metastatic potential of cancer cells<sup>42,43</sup>.

Given that ubiquitination regulates diverse signaling networks through the coordinated action of ubiquitin-activating enzymes (E1), ubiquitin-conjugating enzymes (E2), ubiquitin ligases (E3), and deubiquitinases (DUBs), we compiled lists of these enzymes from GeneCards<sup>44</sup>, UniProt<sup>45</sup> and UbiBrowser<sup>46</sup> and mapped them with our dataset. Differential expression analysis revealed three E2 ubiquitin-conjugating enzymes, seven E3 ligases and one DUB were significantly changed between SR and NR groups, including ubiquitin-conjugating enzyme E2 I (UBE2I, E2), ubiquitin-conjugating enzyme E2 L3 (UBE2L3, E2), ubiquitin-conjugating enzyme E2 E2 (UBE2E2, E2), interferon regulatory factor 2-binding protein-like (IRF2BPL, E3), UBX domain-containing protein 7 (UBXN7, E3), tripartite motif-containing protein 32 (TRIM32, E3), F-box/LRR-repeat protein 18 (FBXL18, E3), E3 ubiquitin-protein ligase DTX3L (DTX3L, E3), cullin-4B (CUL4B, E3), F-box only protein 2 (FBXO2, E3) and ubiquitin carboxyl-terminal hydrolase 10 (USP10, DUB).

Among these, E3 ligases IRF2BPL, CUL4B, FBXO2, and TRIM32 showed consistent differences across both comparisons (SR *vs.* NR, MTC *vs.* NAT; Fig. 3g). To validate these findings, we examined the protein intensities of the four E3 ligases in external fresh frozen samples<sup>21</sup>, and found that CUL4B and TRIM32 were consistently downregulated in SR relative to NR in the external FUSCC dataset (Fig. 3h). Finally, by integrating significantly altered diGly-modified sites, E2 ubiquitin-conjugating enzymes, E3 ligases and DUB, we constructed an interaction network using STRING<sup>47</sup>, providing ubiquitin-mediated regulatory pathways potentially involved in MTC recurrence (Fig. 3i).

In summary, we characterized the ubiquitinome of MTC, revealed differentially expressed diGly-modified sites, E2 ubiquitin-conjugating enzymes, E3 ligases, and DUBs, and reconstructed their potential regulatory networks, with key dysregulated E3 ligases CUL4B and TRIM32 validated in an independent larger dataset. These findings raise important considerations regarding their potential functions in MTC pathogenesis.

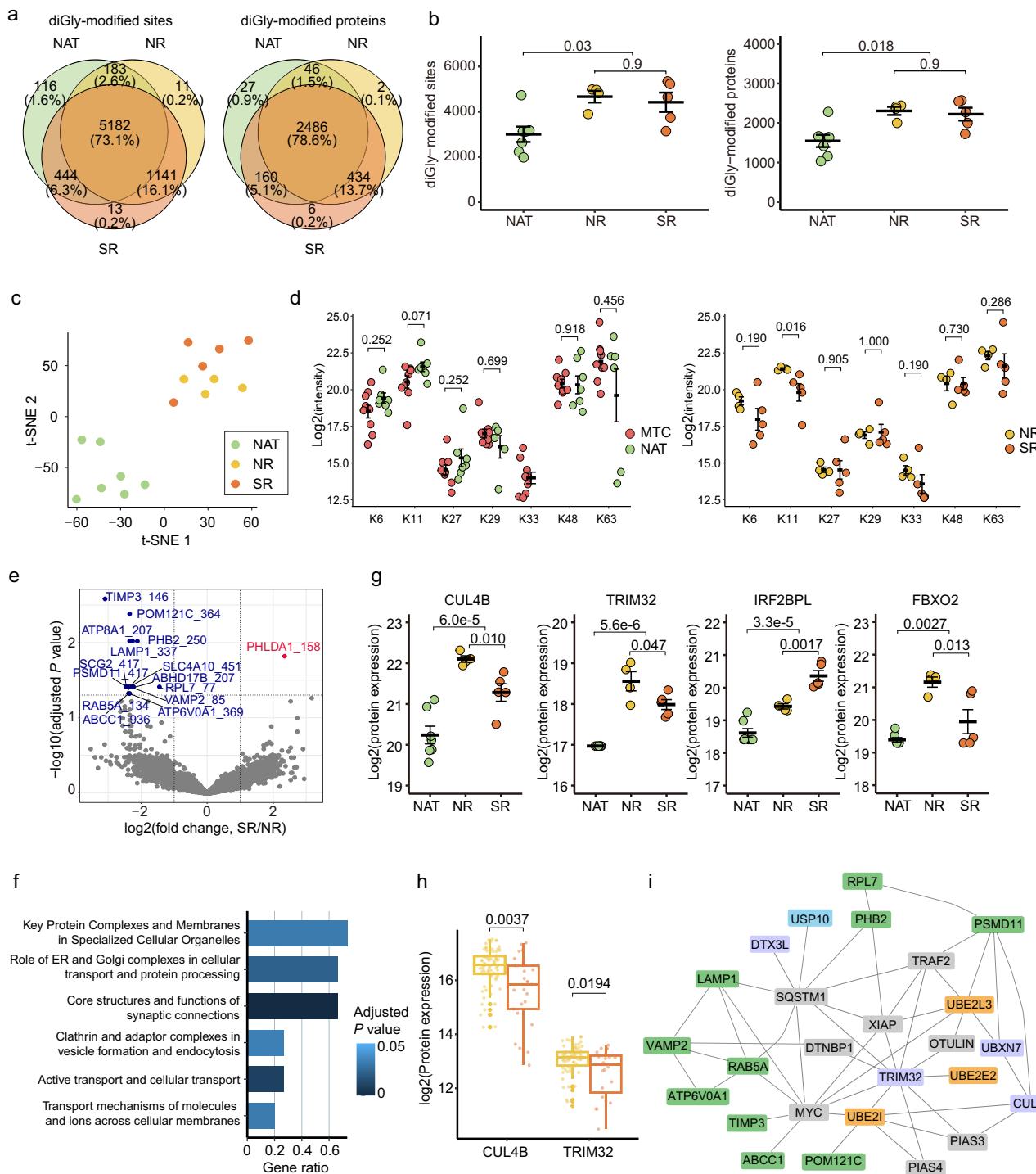
### Molecular subtyping of MTC based on proteomic profiles

To investigate the prognostic relevance of proteomic heterogeneity in MTC, we applied the non-negative matrix factorization (NMF) algorithm to perform unsupervised clustering of differentially expressed proteins (Supplementary Fig. 7). The algorithm prioritized 52 proteins, stratifying the cohort into three distinct molecular subtypes: M1, M2, and M3. These subtypes exhibit divergent clinicopathological features, mutational landscapes, and proteomic profiles.

Comparative analysis reveals significant demographic and genetic disparities among subtypes. The three subtypes significantly differ in terms of structural recurrence rate. The M2 subtype has a poorer prognosis, with a 5-year recurrence-free survival (RFS) rate of 68.8% and a DSM rate of 6.3%, while samples of the M3 subtype have a better prognosis, with a 5-year RFS of 94.2% and no DSM (Table 3). In the heatmap, each column represents a patient, and each row represents either a clinical feature (Fig. 4a), a gene mutation (Fig. 4b), or the expression level of a protein (Fig. 4c). The M3 has a lower proportion of males and a higher proportion of hereditary MTC patients than the other two subtypes (Fig. 4a, Supplementary Fig. 8a). The M2 subtype has larger max nodule sizes and more cases of ETE, LNM, and high-grade MTC (Supplementary Data 4). M2 also contains a higher proportion of patients with the *RET* M918T mutation (Fig. 4b). The 52 subtyping proteins cluster into three groups, each highly expresses in one of the subtypes (Fig. 4c). Notably, MTC-specific marker (calcitonin), pan-neuroendocrine markers have higher protein abundances in M3 (Supplementary Fig. 8b). TNC, a reported prognosis biomarker of MTC<sup>21</sup>, is enriched in M2 and decreased in M3.

To further explore the association between protein subtypes and prognosis, we used the same molecular typing method based on these 52 proteins to validate in our independent test dataset consisting of four hospitals and a published dataset (FUSCC<sup>21</sup>). Quality control details for the independent test datasets are shown in Supplementary Fig. 2d–g. Survival analysis shows that the RFS rates of M2 and M3 are significantly different in the discovery dataset ( $P = 4.57e-06$ ), independent test set ( $P = 0.034$ ), and the 64 MTC samples from the published literature<sup>21</sup> ( $P = 0.017$ ) (Fig. 5a–c), demonstrating the robustness of our 52-protein classifier in classifying patients into proteomic subtypes with different prognostic outcomes.

Functional analysis by Metascape reveals that proteins highly expressed in M1, M2, and M3 are related to metabolic process, cell adhesion, and ECM regulation, respectively (Fig. 4c). The interaction network of subtyping proteins highly expressed in M2 reveals their involvement in multicellular organismal processes, cell adhesion, protein digestion and absorption, and non-integrin membrane-ECM interactions (Fig. 5d). Pathway enrichment analysis using GSVA highlights distinct functional characteristics that may play important roles across the three subtypes (M1, M2, M3), suggesting potential mechanisms underlying disease heterogeneity. A larger absolute  $t$  value indicates greater significance, and a positive value denotes higher pathway activity. The M1 subtype is linked to integrin activation, metabolism, and monocyte aggregation (Fig. 5e), while M2 is enriched for apoptotic process, response to prostaglandin, modification of amino acids, and cell signaling. M3 is associated with neurotransmission, synaptic transmission, and intracellular environmental regulation. Given the association between epithelial-mesenchymal transition (EMT) and tumor aggressiveness, we quantified EMT pathway activity. M2 tumors exhibit elevated expression of EMT drivers, such as Zinc



**Fig. 3 | Ubiquitinome landscape of MTC.** **a** Overlap of quantified diGly-modified sites and proteins across groups. **b** Numbers of identified diGly-modified sites and proteins in each group. The significance is determined by Wilcoxon rank sum test. **c** t-SNE plot of the MTC ubiquitinome. **d** Intensities of ubiquitination at different lysine residues of ubiquitin. The significance is determined by Wilcoxon rank sum test. **e** Differentially regulated diGly-modified sites between SR and NR. Numbers after protein names indicate the position of ubiquitinated lysine residues. *P* values were calculated by empirical Bayes moderated *t* test (limma). Cutoff:  $|\log_2(\text{FC})| > 1$ , B-H adjusted  $P < 0.05$ . **f** Pathways enriched by diGly-modified sites with significant changes between SR and NR groups (one-sided Fisher's Exact Test). **g** E3 ligases differentially expressed in both SR vs. NR and MTC vs. NAT. *P* values were

calculated by empirical Bayes moderated *t* test (limma). **h** Validation of E3 ligases CUL4B and TRIM32 expression in the FUSCC dataset<sup>21</sup> (yellow, NR,  $n = 77$ ; orange, SR,  $n = 25$ ). Box plots show the first quartile, median, and third quartile, and the whiskers represent  $\pm 1.5$  interquartile range. *P* values were calculated by two-sided Welch's *t* test. **i** Interaction network of differentially expressed E3 ligases, DUBs and ubiquitinated proteins between SR and NR. Green, downregulated ubiquitinated proteins; orange, dysregulated E2 ubiquitin-conjugating enzymes; purple, dysregulated E3 ligases; blue, dysregulated DUB. Abbreviations: NAT, normal adjacent thyroid tissue; NR, non-recurrence; SR, structural recurrence. Data in panel b, d and g are presented as mean values  $\pm$  SEM. Sample size in panel b, c, d and g: NAT,  $n = 7$ ; NR,  $n = 4$ ; SR,  $n = 5$ .

**Table 3 | Summary of the clinical features of the MTC proteomic subtypes**

Proportion of patients (%)	Prognosis	5-year RFS	Disease-specific mortality	Hereditary MTC	RET mutation	High-grade MTC	Other clinical information
<b>M1</b> 51.6%	middle	87.2%	4.5%	13.4%	57.3%	4.9%	more micro-MTC (size < 1 cm), less multifocality, bilaterality
<b>M2</b> 18.4%	poor	68.8%	6.3%	10.9%	66.7%	9.8%	more macro-MTC (size ≥ 1 cm), LNM, multifocality, vascular invasion, ETE, males
<b>M3</b> 30.0%	good	94.2%	0.0%	26.9%	52.5%	2.5%	more macro-MTC (size ≥ 1 cm), and bilaterality, less LNM, vascular invasion, ETE, males

RFS recurrence-free survival, LNM lymph node metastasis, ETE extrathyroidal extension.

finger E-box binding homeobox 2 (ZEB2), alongside downregulation of epithelial markers cadherin 1 (CDH1), indicating a more mesenchymal phenotype (Supplementary Fig. 9).

Since the three subtypes of MTC showed differences in clinical indicators, genetic mutations, and protein expression, we further explored potential therapeutic targets related to MTC subtypes. We first matched all the proteins that differed among the three subtypes (at least one group  $P < 0.05$ ,  $n = 6756$ ) with the drug targets in the Ingenuity Pathway Analysis (IPA) database and the Human Protein Atlas<sup>48</sup> (HPA) database. Then, we used multivariate Cox to remove confounding factors to find 12 significant targets. The left heatmap shows the expression patterns of the 12 drug-target proteins across the three subtypes, particularly highlighting their differential expression between the SR and NR groups. The right panel presents the corresponding risk scores, reflecting the prognostic impact of each drug-target protein. Among these, macrophage migration inhibitory factor (MIF) has the highest risk factor (Fig. 5f).

We computed the immune infiltration scores by xCell<sup>49</sup> and found that M1 had the highest immune scores. Further exploration of cell-type specific immune cell infiltration reveals different levels of CD8+ T cells, Tregs, dendritic cells, NK cells, macrophages, and granulocytes in the three subtypes (Supplementary Fig. 10a). M2 and M3 have more CD8+ T cells, and M3 has the lowest Tregs. Higher levels of infiltration of dendritic cells, NK cells, macrophages, and granulocytes are found in M1. The expression of immune checkpoints in each subtype is also examined. Among the 24 immune checkpoints quantified in our dataset, 7 are significantly higher in M1 (Supplementary Fig. 10b).

To validate the immune infiltration results estimated by xCell, we performed immunohistochemistry (IHC) staining for immune cells using a CD45 antibody on the three molecular subtypes. Quantification of CD45+ cells showed higher immune cell infiltration in the M1 subtype compared with M2 and M3, while no significant difference was observed between M2 and M3 (Supplementary Fig. 11).

In conclusion, we identified three proteomic subtypes of MTC based on 52 proteins. M1 with an intermediate prognosis is featured in the upregulated metabolic process and immune infiltration score. M2 with the worst prognosis is featured in the upregulated EMT pathway and the lowest immune score. The M3 subtype exhibits favorable prognosis and distinct neuroendocrine differentiation, characterized by overexpression of pan-neuroendocrine markers and enhanced synaptic signaling pathways.

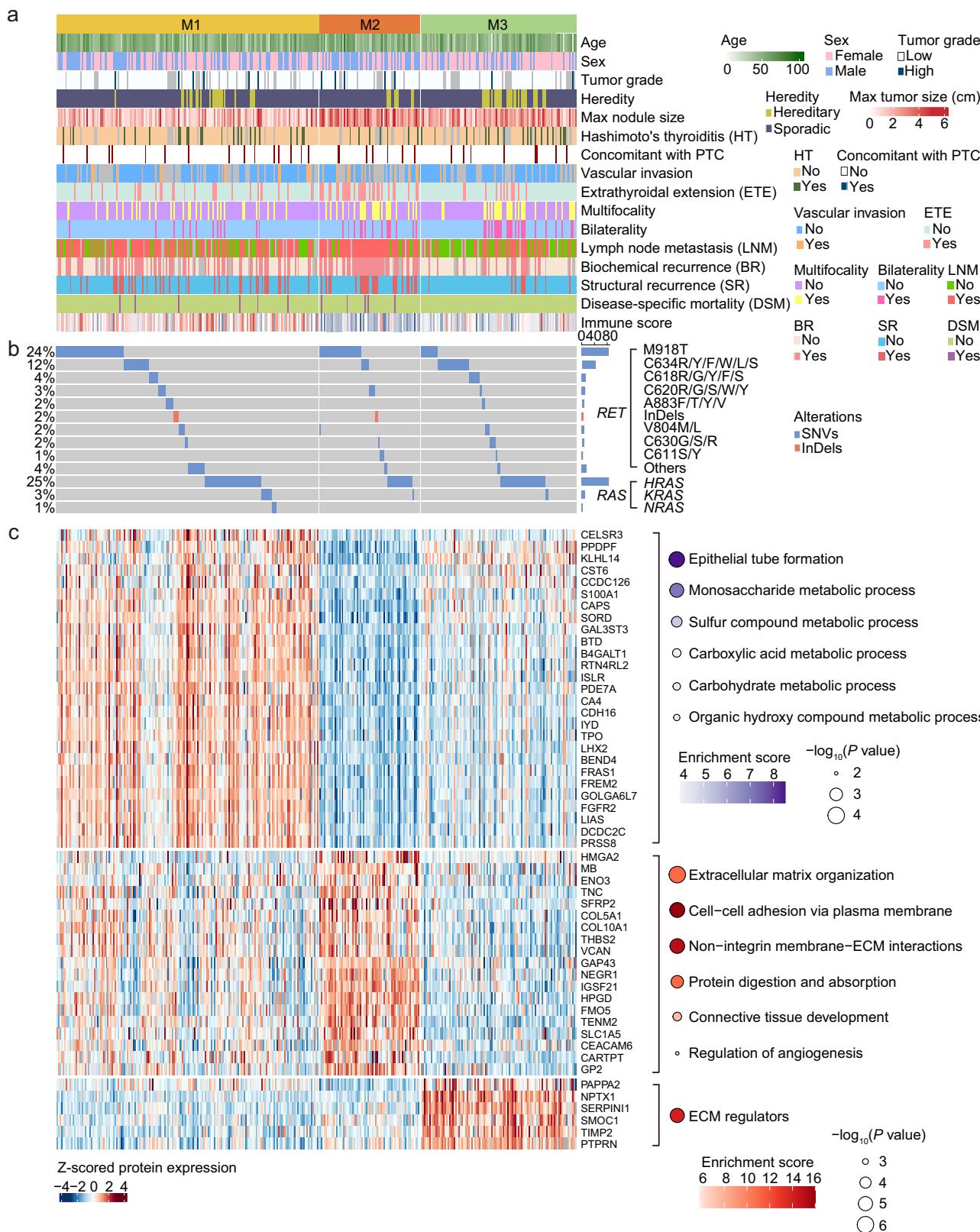
### Machine learning model for recurrence prediction

To predict prognostic risk and develop individualized treatment and follow-up plans for patients, four machine learning models were developed to predict the probability of structural recurrence after initial surgery, each derived from clinical indicators, gene mutations, and proteomic data, respectively, and one that integrated three kinds of data. The construction of the model consisted of three phases: feature selection, model training and cross-validation, and model prediction (Fig. 6a). The detailed process of model construction is described in the Methods section.

The models were further validated in the independent test dataset ( $n = 105$ ) and FUSCC dataset ( $n = 64$ ) to evaluate the generalization ability (Fig. 6b). In the independent test set, the integrated model and the protein model obtain similar areas under the curve (AUCs) (0.87 and 0.85), which are higher than the clinical (0.76), genomic (0.47) models, and IMTCGS (0.51). In the FUSCC dataset, the proteomic model obtains the highest AUC (0.78), followed by the integrated model (0.77), the clinical model (0.76), and the genomic model (0.53). It is worth noting that the samples in the FUSCC dataset are fresh frozen, which is different from the FFPE samples in the discovery dataset used to train our model. The similar performance of the model further demonstrates the robustness of our model. In addition, the number of features in the integrated model is 31% fewer than that in the protein model, so with a similar AUC, the integrated model is considered better.

Based on the expression of proteins in the integrated model and clinical indicators, we divided the patients in the two test sets into high-risk and low-risk groups, and their RFS probabilities are shown in Fig. 6c ( $P = 1e-4$  and  $P = 0.046$ ). Figure 6d and Supplementary Fig. 12 illustrate the ranking of feature importance, where features with higher scores contribute more significantly to the model. Of the 18 proteins in the integrated model, five proteins are reported to be associated with thyroid: melanotransferrin (MELTF)<sup>50</sup>, VGF nerve growth factor inducible (VGF)<sup>51</sup>, inter-alpha-trypsin inhibitor heavy chain 1 (ITIH1)<sup>52-54</sup>, platelet-activating factor acetylhydrolase 1b catalytic subunit 3 (PAFAH1B3)<sup>55</sup>, and selenoprotein 1 (SELENO1)<sup>56</sup>. Of the remaining 13 proteins, all have been reported to be associated with the progression of multiple cancers, except for lysophospholipase D GPD1 (GPD1), which is not reported to be related to cancer. The relative abundance of the 18 model proteins and clinical characteristics in the different risk groups is illustrated in the heatmap (Fig. 6e). MELTF, sorcin (SRI), protein tyrosine phosphatase receptor type M (PTPRM), laminin subunit alpha 5 (LAMA5), protein associated with Lin7 1 (PALS1), and transmembrane channel like 4 (TM4) have lower abundance in the high-risk group, while the other 12 proteins have higher abundance. Pathway enrichment by STRING database<sup>47</sup> revealed that GPD1, SELENO1, and PAFAH1B3 were significantly enriched in the ether lipid metabolism pathway (false discovery rate = 0.0041). Ether lipid metabolism is elevated in several cancers and supports tumor proliferation and survival, suggesting its potential role in MTC progression and recurrence<sup>57,58</sup>. The network established by IPA shows that there are 13 proteins out of the 18 proteins connected with each other (Fig. 6f), which are regulated directly or indirectly by CDH1, CDKN1A and cyclin-dependent kinase inhibitor 2A (CDKN2A).

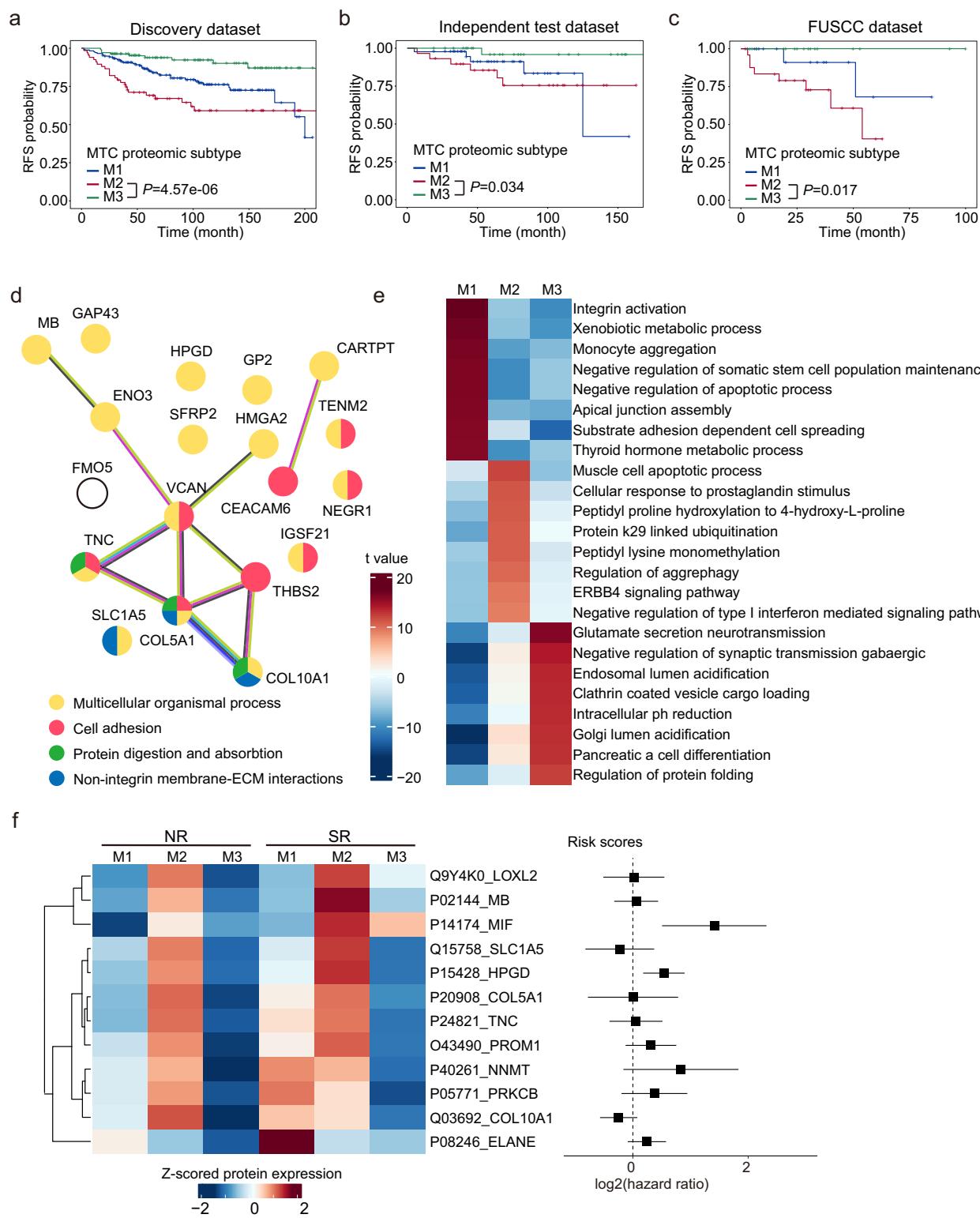
To investigate the expression levels of the three key regulators, we explored the expression levels in a dataset comprises various thyroid nodule pathologies from multiple clinical centers (<https://www.ebi.ac.uk/pride/archive/projects/PXD061183>). Due to its very low abundance, CDKN1A was not detected in this dataset. Additionally, we examined cyclin-dependent kinase 4 inhibitor C (CDKN2C), which, together with CDKN2A, inhibits cyclin binding to CDK4 and CDK6<sup>59</sup>. CDH1 was significantly lower in MTC compared to normal,



**Fig. 4 | Molecular subtyping based on 52 proteins by non-negative matrix factorization.** **a** Clinical characteristics, **b** genomic alterations, and **c** proteomic expressions and pathway enrichment of the three MTC subtypes.  $P$  values were calculated by one-sided Fisher's Exact Test for pathway enrichment.

MNG, and FTC samples, but higher than in PTC, PDT, and ATC (Supplementary Fig. 13a), indicating a negative correlation with metastasis tendency. CDKN2A was higher in PTC, while CDKN2C was elevated in MNG, PTC, and FTC than in MTC. We further examined

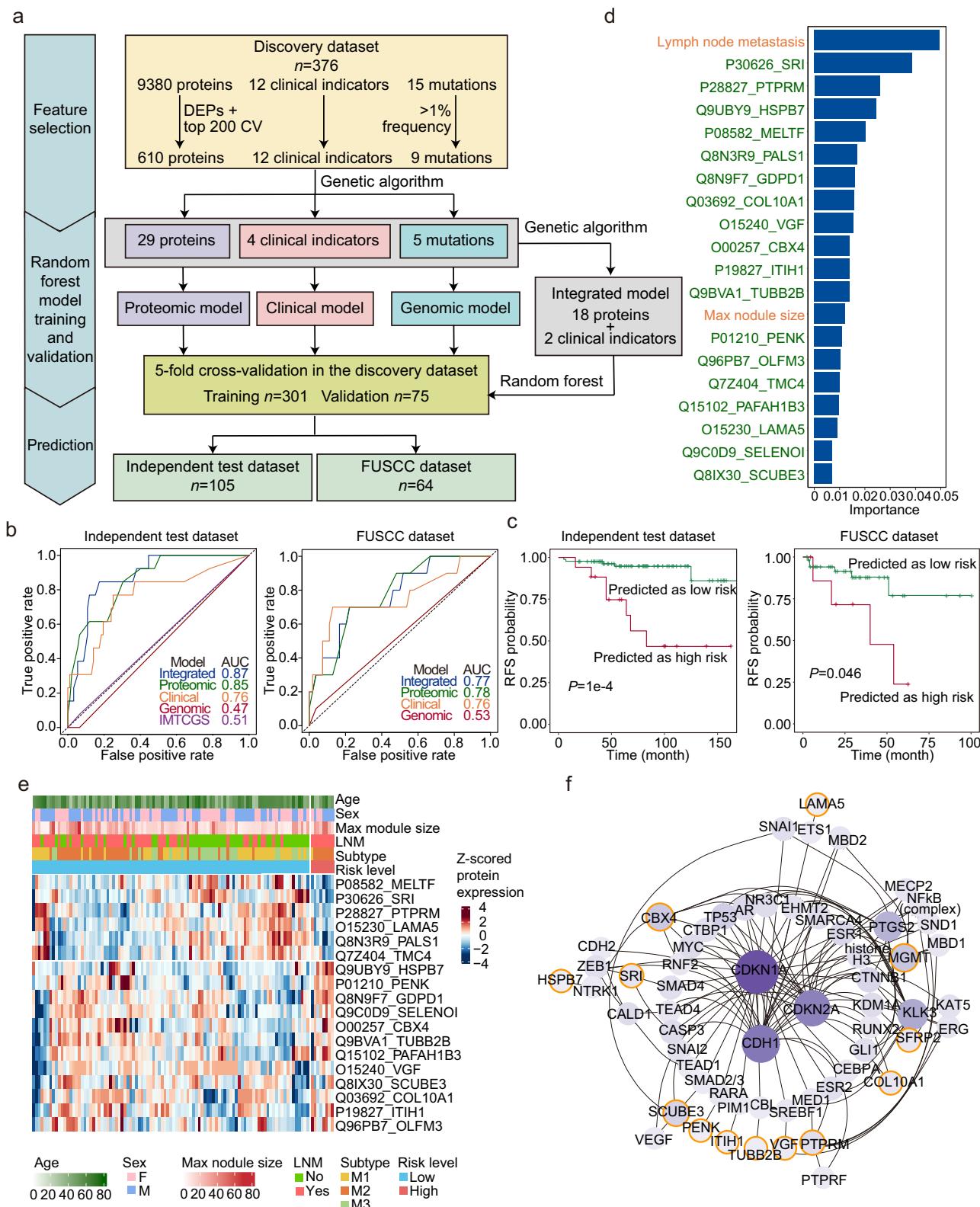
these proteins across the identified molecular subtypes and found that M2 was characterized by low CDH1 and CDKN2C but high CDKN2A expression, a pattern similar to that of ATC (Supplementary Fig. 13b).



**Fig. 5 | Functional profiling and potential drug target identification.**

Kaplan-Meier survival curves for **a** the discovery dataset (M1,  $n = 178$ ; M2,  $n = 64$ ; M3,  $n = 104$ ), **b** the independent test dataset (M1,  $n = 43$ ; M2,  $n = 29$ ; M3,  $n = 33$ ), and **c** the test dataset from published data (M1,  $n = 17$ ; M2,  $n = 27$ ; M3,  $n = 20$ ) showing different prognostic risks of MTC subtypes.  $P$  values were calculated by Peto-Peto test. **d** The protein interaction network of subtyping proteins that were highly expressed in the M2 subtype based on the STRING database. **e** Pathways enriched in

each proteomic subtype by gene set variation analysis (GSVA) using gene ontology biological process gene set (M1,  $n = 178$ ; M2,  $n = 64$ ; M3,  $n = 104$ ). The enrichment score was represented by  $t$ -values. **f** Potential drug targets for MTC. The Proteins that can be targeted by FDA-approved drugs or candidate drugs were shown on the left, with respective multivariate CoxPH risk scores on the right. Risk scores were represented by hazard ratios (square) and 95% confidence intervals (whisker). Sample size from left to right: 139, 41, 95, 39, 23, 9.



**Fig. 6 | Development and validation of prognostic models.** **a** Schematic workflow of the model development. **b** The ROC curves of random forest models in the independent test ( $n = 105$ ) and the FUSCC ( $n = 64$ ) datasets. **c** Kaplan-Meier plot for recurrence-free survival of the 20-feature integrated model in the independent test ( $n = 105$ ) and the FUSCC ( $n = 64$ ) datasets.  $P$  values were calculated by Peto-Peto test. **d** Permutation importance of the 20 features from the integrated model

ranked by the Shapley additive explanations (SHAP values). **e** Expression levels of the 18 proteins from the integrated model in the two risk groups ( $n=105$ ). Color bars above the heatmap indicate clinical characteristics and molecular subtypes of patients in the independent test dataset. **f** Network showing the 18 proteomic features of the integrated model with the connected proteins enriched by the Ingenuity Pathway Analysis software.

The differential expressions of CDH1, CDKN2A, and CDKN2C across various histological types suggest associations with disease aggressiveness and metastatic potential. CDH1, a regulator of epithelial-mesenchymal transition, has been linked to thyroid cancer progression and metastasis<sup>60,61</sup>. Our findings are consistent with previous studies but extend them by incorporating a broader range of histological types.

We also investigated how the 18 proteins in the integrated model were affected by mutation patterns and sex. COL10A1 expression was higher in males and in patients harboring *RET* mutations, including *RET* M918T (Supplementary Fig. 14). OLFM3 and ITIH1 exhibited opposite expression trends based on *RET* and *RAS* mutations. VGF expression was significantly elevated in *RAS* mutation patients. In addition, age influenced protein expressions differently among the three subtypes. For instance, COL10A1 increased with age in M1 patients, while OLFM3 and SCUBE3 increased, and MELTF decreased with age in M3 patients (Supplementary Fig. 15).

To sum up, the integrated model comprising two clinical features and 18 proteins demonstrated the best performance among the models and performed well in predicting structural recurrence in MTC patients.

## Discussion

This study represents a multi-center data resource of MTC, integrating proteomic, genomic, and clinical data from 452 MTC patients from 10 clinical centers and quantified 10,092 proteins through MS-based proteomics. Our work identified clinical, genomic, and proteomic risk factors of structural recurrence and produced an MTC subtyping classifier that divided patients into three subtypes with various characteristics and prognoses. We built a machine learning model for predicting the risk of SR in MTC and validated it in two independent test datasets. The model performed better compared to reported prognostic indicators such as IMTCGS and *RET* mutations, which may facilitate tailored risk stratification and informing individualized follow-up strategy.

The univariate CoxPH analysis identified male<sup>62</sup>, high-grade tumor<sup>13–16,63</sup>, multifocality<sup>64</sup>, concurrence with PTC, vascular invasion<sup>65,66</sup>, ETE<sup>10</sup>, and LNM<sup>10</sup> as risk factors of SR in MTC, consistent with previous studies. Our study applies the IMTCGS to a large Chinese dataset, validating its utility for identifying SR and DSM risk. The co-occurrence of PTC and MTC has rarely been studied. We report the risk of this co-occurrence in MTC, hinting at potential interactions. In contrast to preoperative<sup>67</sup> and postoperative<sup>10</sup> unstimulated serum calcitonin levels, which are reported to be associated with prognostic risk, we found no significant association between calcitonin levels in tumor tissues and recurrence. Instead, CGRP1, from the same coding gene of calcitonin and alternative RNA splicing, exhibited a positive correlation with structural recurrence, indicating a potential MTC prognosis marker. Future work could further examine CGRP1 levels in serum and validate its effect.

This study reported an overall mutation rate of 60.0% for *RET* and 26.4% for *RAS* in MTC patients. Although the *RAS* mutations were relatively frequent, neither our study nor the literature found a significant association between *RAS* mutations and the prognosis of MTC<sup>11</sup>. Additionally, our data revealed distinct germline *RET* mutation patterns in Chinese hereditary MTC patients compared to Western populations. Similar frequencies were reported in other published Asian datasets. Separate studies from Japan<sup>68</sup> (# families = 75), South Korea<sup>69,70</sup> (# families = 48), and China<sup>71–74</sup> (# families = 62) consistently identified *RET* C634 as the most common mutation site, with frequencies of 60%, 47.9%, and 66.1%, respectively. Our previous study<sup>21</sup> also reported a 60% prevalence of *RET* C634 locus in patients with germline *RET* mutations. According to the 2015 American Thyroid Association (ATA) guidelines<sup>1</sup>, *RET* C634 is classified as a high-risk mutation, suggesting a higher risk for Asian hereditary MTC patients.

In our study, we found differential expression of the E3 ubiquitin ligases CUL4B and TRIM32 across NAT, NR, and SR groups, and the network analysis further predicted regulatory pathways involving these ligases and downstream dysregulated diGly-modified sites. CUL4B participates in DNA damage repair, chromatin remodeling, and cell cycle regulation<sup>75</sup>. It is overexpressed in various solid tumors, including thyroid carcinoma, compared to normal tissues, and correlates with tumor malignancy and poor prognosis<sup>76</sup>. In thyroid cancer specifically, emerging data demonstrate that CUL4B promotes dedifferentiation and poor prognosis by ubiquitinating ARID1A and repressing PAX8 expression<sup>77</sup>. This process contributes to the development of anaplastic thyroid carcinoma and modulates sensitivity to MAPK inhibitors. TRIM32 enhances proliferation, migration, and chemoresistance by degrading tumor suppressors such as p53 and activating oncogenic pathways including Wnt/β-catenin and TGF-β<sup>78,79</sup>. High expression of TRIM32 is associated with metastasis and poor prognosis in various cancers, including prostate cancer and lung cancer<sup>79–81</sup>. Although direct evidence in MTC remains limited, their known roles suggest that CUL4B and TRIM32 may promote recurrence and metastasis. Our findings highlight them as potential regulators of MTC pathogenesis and therapeutic targets, providing a valuable direction for future investigation.

Compared with the previously reported molecular subtypes of MTC<sup>21</sup>, the subtypes in this study not only reproduced certain clinical and molecular associations observed in earlier studies but also provided several notable advances. First, the subtypes in this study are based on fewer proteins (52 vs. hundreds) and validated in more samples and multiple independent testing datasets (Supplementary Table 3). Second, it demonstrated stronger associations with patient outcomes. For example, the M3 subtype was characterized by favorable prognosis and very low disease-specific mortality, while M2 showed poor outcomes, highlighting improved prognostic stratification. Finally, the proteomic signatures also corresponded to immune contexture, with M1 exhibiting higher immune scores, suggesting potential therapeutic implications. We further observed significantly lower expressions of CGRP1 and CGRP2 in M1 than in M2/M3. This finding aligns with prior evidence<sup>82</sup> linking elevated CGRP levels to abnormal dendritic cell development and impaired tumor-infiltrating T-cell activity, which collectively foster an immune-suppressive microenvironment. Importantly, subtype malignancy (prognostic risk) positively correlates with CD276 abundance. This immune checkpoint protein enabling tumor immune evasion via suppression of cytotoxic T cells and NK cells<sup>83</sup>, and has recently emerged as a potential therapeutic target for MTC<sup>84</sup>. Together, these data indicate distinct immune microenvironments across subtypes and suggest that molecular heterogeneity drives divergent prognostic outcomes in MTC.

This study has several limitations. First, its retrospective design and reliance on postoperative FFPE samples may limit the generalizability of the findings. Future studies should validate the identified MTC prognostic biomarkers using preoperative plasma or fine-needle aspiration (FNA) samples to improve their clinical utility. Second, targeted proteomic approaches could be applied to quantify selected model proteins in independent cohorts, offering a cheaper, more accurate technology to validate the model's performance and robustness. Additionally, expanding the number of centers for molecular subtyping validation and experimentally validating potential drug targets in cell lines or animal models are important directions for future research. Despite these limitations, our study contributes significantly to MTC recurrence risk stratification.

In conclusion, this study provides a valuable data resource on proteomics and gene mutations in Chinese MTC patients while identifying potential biomarkers and therapeutic targets. The proteomic subtypes associated with varying prognoses offer a foundation for investigating disease heterogeneity and developing personalized therapies. Furthermore, the machine learning model, which utilizes

multidimensional features to stratify recurrence risk in MTC patients, may facilitate the development of individualized treatment, potentially benefiting patient outcomes.

## Methods

### Patients and samples

This study was conducted in accordance with the ethical principles in the Declaration of Helsinki. The Ethics Committee of Westlake University reviewed and approved the study protocol with the study number 20240527GTN001. Informed consent was waived because this retrospective study analyzed previously archived clinical data and specimens, and all patient identifiers were removed to ensure minimal risk and protection of privacy.

This retrospective study included 543 patients with MTC who underwent surgical treatment between 2002 and 2022. The exclusion criteria were as follows: (a) non-primary surgery, (b) loss to follow-up, or (c) a prior history of MTC-related medication. Disease remission and recurrence were determined according to the ATA management guidelines<sup>1</sup>. Only structural recurrences were considered in this study, which was defined as recurring histological or radiologic evidence of MTC after radical surgery. RFS was defined as the time from initial surgery to the first SR. For patients without SR, RFS was the time from initial surgery to the last follow-up.

FFPE slides were collected for each sample. All samples were reviewed by at least two experienced pathologists. Ninety-one patients were excluded for the following reasons: (1) loss to follow-up ( $n=22$ ), (2) not-thyroid tissue ( $n=32$ ), (3) non-primary surgery ( $n=18$ ), (4) normal thyroid tissue ( $n=6$ ) and (5) non-radical surgery ( $n=13$ ). After careful review, 377 and 105 samples were included in the two datasets respectively (Fig. 1a, Supplementary Table 4). The discovery dataset includes 30 biological replicate samples from the same patient. The MTC tumor grades were determined according to the IMTCGS<sup>12</sup>.

FF samples were collected from the tissue bank of clinical center 1 under the same criteria for FFPE slides, including four NR samples, five SR samples, and seven normal adjacent thyroid (NAT) samples.

### Gene-panel based sequencing

We obtained both MTC and adjacent cancer-free normal tissue samples for gene sequencing during sample collection. Sequencing libraries were created using a next-generation sequencing panel developed by RigenBio to detect variations in 28 genes associated with thyroid cancer (see Supplementary Table 5 for the gene list). DNA was extracted from 477 FFPE thyroid samples, excluding five samples from the discovery dataset, using the DNA Extraction Kit (Rigen Biotech, China). The extraction and sequencing protocol was previously described<sup>50</sup>. In brief, the DNA underwent multiplex amplification of target regions after extraction, followed by PCR amplification to incorporate unique dual-index and Illumina sequencing adapters. After purification using beads, the indexed libraries were quantified using a Thermo Fisher Qubit fluorometer and sequenced on the Illumina NovaSeq 6000 System, generating 150 bp paired-end reads.

The quality of the raw sequencing data was assessed using FastQC (v0.11.9). The raw reads were preprocessed to remove adapters and low-quality bases through Cutadapt (v1.18). The processed reads were then aligned to the hg19 human reference genome using Burrows-Wheeler Aligner software (v0.7.17). Single nucleotide variants (SNVs) and insertions/deletions (InDels) were identified using VarScan2 (v2.4.4), and Ensembl Variant Effect Predictor was used for variant annotation to assess potential impacts.

### Proteomic sample preparation

FFPE tissues were prepared as previously described<sup>33,85</sup>. Briefly, the FFPE slides were dewaxed, rehydrated, and de-crosslinked using heptane, three different concentrations of ethanol (100%, 90%, and 75%),

100% of water and 100 mM Tris-HCl solution (pH=10.0), sequentially. Samples were then lysed, assisted by PCT, in a buffer containing 6 M urea, 2 M thiourea, 10 mM Tris(2-carboxyethyl)phosphine, and 40 mM iodoacetamide. Trypsin and lysC were mixed and used for digestion using PCT. Finally, the digested peptides were quenched with trifluoroacetic acid and desalting using C18 columns (Thermo Fisher Scientific, USA).

### Ubiquitinomic sample preparation

Fresh frozen samples were lysed in 8 M urea, supplemented with protease inhibitor cocktail (Roche, Germany). The lysis process was assisted by a cryogenic grinder under 65 Hz and -20 °C. The protein yield was determined by BCA protein assays (Thermo Fisher Scientific, USA). The lysates were reduced (Tris(2-carboxyethyl)phosphine, 10 mM), alkylated (iodoacetamide, 40 mM), and digested (trypsin, 1:50 w/w). The first digestion took 4 hours and the second digestion took 12 hours. The digested peptides were desalting using 100 mg C18 SepPak cartridges (Waters, USA). The desalting peptides were dried in a vacuum concentrator. Ubiquitin remnant were enriched from 1.5 mg cleaned peptides by immunoaffinity purification (Cell Signaling Technology, USA) according to the manufacturer's protocol. The enriched peptides were desalting using C18 columns (Thermo Fisher Scientific, USA).

### DIA-MS data analysis

Peptide samples were injected into a custom-packed C18 separation column (15 cm × 75 μm × 1.9 μm, 120 Å) equipped with a nanoElute® system (Bruker Daltonics, Germany). Then, samples were separated by a 60 min liquid chromatography (LC) gradient, from 5% to 27% buffer B in 50 min, then to 40% buffer B in 10 min. Buffer A contains 0.1% formic acid in water and buffer B contains 0.1% formic acid in 100% acetonitrile.

Peptides eluted from the LC were analyzed in a hybrid trapped ion mobility spectrometry quadrupole time-of-flight mass spectrometer (timsTOF Pro, Bruker Daltonics, Germany) through a CaptiveSpray nanoflow electrospray ion source. PASEF was performed in DIA mode. The dual TIMS analyzer had an accumulation and ramp time of 100 ms and a total cycle time of 1.17 s, consisting of 14 PASEF scans with four ion mobility-  $m/z$  two-dimensional isolation windows per scan. Ion mobility scans ranged from 0.6 to 1.6 Vs/cm<sup>2</sup>. MS1 and MS2 acquisition was performed within the  $m/z$  range from 100 to 1700 Th. Precursor ions with single charge were excluded.

DIA raw files were analyzed by DIA-NN<sup>86,87</sup> (v1.8.1) against a thyroid-specific spectral library<sup>88</sup>, containing 12,000 proteins and 215,000 precursors. Variable modifications were set to include methionine oxidation and N-terminal acetylation. The fixed modification included cysteine carbamidomethylation. Peptide length range, precursor  $m/z$  range, and fragment ion  $m/z$  range were set as 6–30, 300–1800, and 200–1800, respectively. The false discovery rate for both precursor and protein was set to 1%. “unrelated run” and “use isotopologues” options were selected. Protein inference was set to “off”. Other parameters were kept as default.

For ubiquitinome, the enriched peptides were separated using a custom-packed C18 separation column (15 cm × 75 μm × 1.9 μm, 120 Å) in a Vanquish™ Neo UHPLC system (Thermo Fisher Scientific, USA). The separation process consisted of a 19.5-min gradient running step and a 4.5-min column washing step. During running step, buffer B increased from 3% to 7% in 1 min, then to 30% in 18.5 min. The mobile phase comprised buffer A (98% MS grade H<sub>2</sub>O, 2% MS grade ACN, and 0.1% FA) and buffer B (98% MS grade ACN, 2% MS grade H<sub>2</sub>O, and 0.1% FA).

Following LC separation, the eluted peptides were analyzed in an Orbitrap Astral mass spectrometer equipped with a Nanospray Flex ion source (Thermo Fisher Scientific, USA). The total carrier gas flow was set to 4 L/min. The Orbitrap MS1 full scan settings included a

resolution of 240k, a scan range of 380-980  $m/z$ , a FAIMS CV of  $-48$  V, a radio frequency (RF) lens of 40%, a normalized automatic gain control (AGC) target of 500%, and a maximum injection time (IT) of 3 ms. The Astral DIA-MS2 parameters included a scan range of 150 to 2000  $m/z$ , an isolation window of 2  $m/z$ , a higher-energy collisional dissociation (HCD) collision energy of 25%, and a loop control time of 0.6 s. Other parameters were kept identical to those used for MSI.

Raw files for ubiquitinome were analyzed by DIA-NN (v2.2.0)<sup>89</sup> using a predicted spectral library. The predicted library was generated by DIA-NN using the FASTA file downloaded from UniProt (v 29/06/2025). Variable modifications were set to include methionine oxidation and diGly-modification. The fixed modification included N-term methionine excision and cysteine carbamidomethylation. Peptide length range was set to 6–30. “Unrelated run” option was selected, and “MBR” was disabled. Mass accuracy was set to 10, and MS1 accuracy was set to 4. Scoring was set to “Proteoforms” mode. Other parameters were kept as default.

### Proteomic data quality control and preprocessing

To minimize possible bias during sample preparation and mass spectrometry acquisition, we randomly assigned recurrent and non-recurrent samples. Each batch had 15 tissue samples and one thyroid pooled peptide sample (quality control). Samples from the same patient were used as biological replicates. One sample per batch was randomly selected as a technical replicate, which was injected twice to evaluate the quantification stability of mass spectrometry.

Missing values in the protein matrix were imputed by the ridge regression method<sup>90</sup> through *N4guideR*<sup>91</sup>. The resulting protein matrix was corrected for batch effects using the empirical Bayesian framework Combat in the R package *sva*<sup>92</sup> (v3.48.0). Batch effects were corrected for different clinical centers and sample batches. Each pair of technical replicate samples was combined into one sample by calculating the average protein abundance.

### Proteomic-based unsupervised clustering

The coefficient of variation (CV) was calculated for all proteins, and the 200 proteins with the highest CV values were selected and combined with DEPs. These highly variable proteins were subjected to unsupervised clustering using the non-negative matrix factorization method (*NMF*<sup>93</sup> package, v0.26) in R (v4.3.1). The optimal number of subtypes was obtained using the rank survey function. A total of 150 iterations were conducted to ensure the robustness of the clustering results.

### In silico immune infiltration analysis

The immune score was calculated by xCell<sup>49</sup> (<https://comphealth.ucsf.edu/app/xcell>) using the xCell gene signatures which contained 64 cell types. For in silico immune infiltration, we used CIBERSORTx<sup>94</sup> (<https://cibersortx.stanford.edu/>) to deconvolute the proportions of nine immune cells. An MS-based proteomic data of immunocytes<sup>95</sup> was used as a protein expression profile for generating a custom signature matrix.

### Immunohistochemistry

FFPE sections (5  $\mu$ m) were deparaffinization, antigen retrieval, and nonspecific binding blockage (10% normal goat serum). Subsequently, the slides were incubated overnight with anti-CD45 (1:500, DAKO) and then subsequently probed with secondary antibody (DAKO DAB kit). Stained slides were scanned by KF-SCAN-ST (Kfbio, China) at 10  $\times$  40. Qupath (v0.5.1) was used for image processing and positive cell counting.

### Cox proportional hazards analysis

To identify risk factors associated with structural recurrence, univariate and multivariate CoxPH analyses were performed using 12 clinicopathological features: age, sex, heredity, tumor grade, presence

of Hashimoto’s thyroiditis (HT), multifocality, bilaterality, max nodule size, concurrence of PTC, ETE, LNM, and vascular invasion.

Univariate CoxPH models were initially constructed for each of the 12 factors to identify those with a significance level of  $P < 0.05$ . Subsequently, a multivariate CoxPH model was developed, incorporating all 12 factors to identify independent risk factors associated with structural recurrence.

### Druggable protein screening

Proteins differentially expressed among the three subtypes were identified using the Kruskal-Wallis test and Dunn test. Proteins with a B-H adjusted  $P < 0.05$  were considered significant and matched to drug targets in the IPA and the HPA databases. Subsequent filtering retained only DEPs from this protein pool. Univariate CoxPH analysis was employed to identify significant drug targets, followed by multivariate CoxPH analysis to control for potential confounders.

### Dataset partitioning and cross-validation in machine learning

The dataset consisted of a discovery dataset and two test datasets. The two test datasets included one independent test dataset collected from four independent medical centers ( $n = 105$ ) and the other from published literature<sup>21</sup> (FUSCC dataset,  $n = 93$ ). Of the 93 patients in the FUSCC dataset, 29 were also included in the discovery set. After removing the overlapping patients, 64 patients remained in the FUSCC dataset. Due to variations in the instruments, laboratories, and sample types, batch effects were observed between the proteomic data of the three datasets. These effects were corrected using the Combat<sup>92</sup> method. The preprocessing steps for the test datasets were consistent with those applied to the discovery dataset. To optimize model performance, the discovery dataset was evenly divided into five folds for cross-validation. During each iteration, four subsets were used to train the model, and one subset was used to validate the model.

### Feature selection and model generation

Initially, proteins with missing values greater than 90% were excluded, with 9380 proteins remaining. Then, the proteins were screened in the discovery dataset to identify those significantly associated with prognosis by differential protein analysis and the CV values. DEPs were defined as  $|\log_2(\text{FC})| > 0.25$  and B-H adjusted  $P < 0.05$  between SR vs. NR and DSM vs. S groups. The DEPs and top 200 CV proteins were combined, yielding 610 prognosis-related and most dynamic proteins. Additionally, gene mutation sites with mutation rates  $> 1\%$  were retained. Subsequently, a genetic algorithm (GA) was employed for feature selection, with a detailed methodology previously described<sup>96,97</sup>. Briefly, the genetic algorithm was implemented using the eaSimple function (Python package *DEAP*, v1.4.1), with the following parameters: a crossover probability (cxpb) of 0.5, a mutation probability (mutpb) of 0.2, and a total of 400 generations (ngen). The population was iteratively optimized to retain subsets of features with better performance.

Based on the optimal feature subset selected by GA, three random forest (RF) models were constructed, each utilizing one of the feature types: clinical, genomic, and proteomic. In addition, the genetic algorithm was employed for all 38 features of three types, leading to a further reduction in the number of features. This process culminated in an integrated model encompassing 18 proteins and 2 clinical features.

The RF models predict the final classification result by aggregating the predictions from all decision trees through majority voting. During each iteration, the fitness of feature subsets was evaluated using a RandomForestClassifier (with random\_state=42) as the primary performance metric. The classifier assessed the contribution of each feature subset to the classification task, according to which the samples were divided into high-risk and low-risk groups. The model was optimized in the discovery set using 5-fold cross-validation (training set to validation set ratio 4:1).

## Bioinformatics and statistical analysis

The tables of baseline clinical characteristics were calculated by SPSS Statistics (v23.0, IBM Corporation). The Kolmogorov-Smirnov test was used for normality testing of continuous variables, and the independent double-sided Student's *t* test was used for variables that conform to the normal distribution. *P* values were calculated using the non-parametric Mann-Whitney *U* test for variables that did not meet normality. The chi-square test was used for calculating the *P* values of categorical variables.

The differentially expressed analyses were performed using the *limma* package<sup>98</sup> (v3.56.2). The *P* values were adjusted using the B-H method. Samples lacking RFS information or exhibiting persistent disease were excluded when conducting the differentially expressed analyses. The secreted protein data were obtained from the Human Protein Atlas portal ([www.proteinatlas.org](http://www.proteinatlas.org)). Druggable proteins were annotated from the IPA software database. IPA was also used for pathway enrichment of the differentially expressed proteins. Functional analysis of the 52 subtyping proteins was performed by Metascape<sup>99</sup>. The network was visualized using the STRING database<sup>47</sup>. Functional enrichment of proteins in subtypes was performed using the R package *gsva* with the C5 gene ontology biological process gene sets (v2024.1) from the Molecular Signatures Database (MSigDB). Chi-square test was used to calculate the statistical differences in the characteristics among the three proteomic subtypes. The *P* values in the Kaplan-Meier curve were derived from the Peto-Peto test. The RAS and RET pathway protein list was curated from the MSigDB (v2025.05)<sup>100</sup>. The differences in protein expression between the gene mutation and wildtype groups were calculated by unpaired two-sided Wilcoxon rank sum test. The *P* values were adjusted using the B-H method.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The mass spectrometry proteomic raw data generated in this study have been deposited in the ProteomeXchange Consortium<sup>101</sup> (<https://proteomecentral.proteomexchange.org>) via the PRIDE database<sup>102</sup> under accession code **PXD063677**. The FUSCC<sup>21</sup> dataset used in this study is available in the iProX database<sup>103,104</sup> under accession code **IPX0004234000**. The thyroid dataset is publicly available in the PRIDE database under accession code **PXD061183**. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

## Code availability

Code relevant to data analysis in this study is available at <https://github.com/guomics-lab/MTC>.

## References

- Wells, S. A. Jr. et al. Revised American Thyroid Association guidelines for the management of medullary thyroid carcinoma. *Thyroid* **25**, 567–610 (2015).
- Lim, H., Devesa, S. S., Sosa, J. A., Check, D. & Kitahara, C. M. Trends in thyroid cancer incidence and mortality in the United States, 1974–2013. *JAMA* **317**, 1338–1348 (2017).
- Pereira, M., Williams, V. L., Hallanger Johnson, J. & Valderrabano, P. Thyroid cancer incidence trends in the United States: association with changes in professional guideline recommendations. *Thyroid* **30**, 1132–1140 (2020).
- Roman, S., Lin, R. & Sosa, J. A. Prognosis of medullary thyroid carcinoma: demographic, clinical, and pathologic predictors of survival in 1252 cases. *Cancer* **107**, 2134–2142 (2006).
- Moley, J. F. Medullary thyroid carcinoma: management of lymph node metastases. *J. Natl. Compr. Cancer Netw.* **8**, 549–556 (2010).
- Moley, J. F. & DeBenedetti, M. K. Patterns of nodal metastases in palpable medullary thyroid carcinoma: recommendations for extent of node dissection. *Ann. Surg.* **229**, 880–887 (1999).
- Bartz-Kurycki, M. A., Oluwo, O. E. & Morris-Wiseman, L. F. Medullary thyroid carcinoma: recent advances in identification, treatment, and prognosis. *Ther. Adv. Endocrinol. Metab.* **12**, 20420188211049611 (2021).
- Kuo, E. J. et al. Risk factors associated with reoperation and disease-specific mortality in patients with medullary thyroid carcinoma. *JAMA Surg.* **153**, 52–59 (2018).
- Jung, C. K. et al. Update on C-cell neuroendocrine neoplasm: prognostic and predictive histopathologic and molecular features of medullary thyroid carcinoma. *Endocr. Pathol.* **34**, 1–22 (2023).
- Kotwal, A., Erickson, D., Geske, J. R., Hay, I. D. & Castro, M. R. Predicting outcomes in sporadic and hereditary medullary thyroid carcinoma over two decades. *Thyroid* **31**, 616–626 (2021).
- Xu, B. et al. Association of the genomic profile of medullary thyroid carcinoma with tumor characteristics and clinical outcomes in an international multicenter study. *Thyroid* **34**, 167–176 (2024).
- Xu, B. et al. International medullary thyroid carcinoma grading system: a validated grading system for medullary thyroid carcinoma. *J. Clin. Oncol.* **40**, 96–104 (2022).
- Censi, S. et al. Tumor grade and molecular characteristics associated with survival in sporadic medullary thyroid carcinoma. *Thyroid* **34**, 177–185 (2024).
- Ruz-Caracuel, I. et al. Transcriptomic differences in medullary thyroid carcinoma according to grade. *Endocr. Pathol.* **35**, 207–218 (2024).
- Vissio, E. et al. External validation of three available grading systems for medullary thyroid carcinoma in a single institution cohort. *Endocr. Pathol.* **33**, 359–370 (2022).
- Nigam, A. et al. Tumor grade predicts for calcitonin doubling times and disease-specific outcomes after resection of medullary thyroid carcinoma. *Thyroid* **32**, 1193–1200 (2022).
- Ciampi, R. et al. Genetic landscape of somatic mutations in a large cohort of sporadic medullary thyroid carcinomas studied by next-generation targeted sequencing. *iScience* **20**, 324–336 (2019).
- Abraham, D. et al. MicroRNA profiling of sporadic and hereditary medullary thyroid cancer identifies predictors of nodal metastasis, prognosis, and potential therapeutic targets. *Clin. Cancer Res.* **17**, 4772–4781 (2011).
- Guo, T. J. A. S. a. M. M. Mass-spectrometry-based proteomics: from single cells to clinical applications. *nature* **638**, 901–911 (2025).
- Zhan, S., Li, J., Wang, T. & Ge, W. Quantitative proteomics analysis of sporadic medullary thyroid cancer reveals FN1 as a potential novel candidate prognostic biomarker. *Oncologist* **23**, 1415–1425 (2018).
- Shi, X. et al. Integrated proteogenomic characterization of medullary thyroid carcinoma. *Cell Discov.* **8**, 120 (2022).
- Damgaard, R. B. The ubiquitin system: from cell signalling to disease biology and new therapeutic opportunities. *Cell Death Differ.* **28**, 423–426 (2021).
- Glickman, M. H. & Ciechanover, A. The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiol. Rev.* **82**, 373–428 (2002).
- Xiong, X. et al. Ubiquitin-modifying enzymes in thyroid cancer: mechanisms and functions. *Heliyon* **10**, e34032 (2024).
- Hyndman, B. D. et al. Differential recruitment of E3 ubiquitin ligase complexes regulates RET isoform internalization. *J. Cell Sci.* **130**, 3282–3296 (2017).

26. Dilla, T., Velasco, J. A., Medina, D. L., Gonzalez-Palacios, J. F. & Santisteban, P. The MDM2 oncoprotein promotes apoptosis in p53-deficient human medullary thyroid carcinoma cells. *Endocrinology* **141**, 420–429 (2000).

27. Takano, T. et al. PGP9.5 mRNA could contribute to the molecular-based diagnosis of medullary thyroid carcinoma. *Eur. J. Cancer* **40**, 614–618 (2004).

28. Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat. Protoc.* **8**, 1950–1960 (2013).

29. Kim, W. et al. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340 (2011).

30. Wagner, S. A. et al. A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell Proteom.* **10**, M111 013284 (2011).

31. Fanciulli, G. et al. Proteasome inhibitors in medullary thyroid carcinoma: time to restart with clinical trials?. *Front. Endocrinol.* **14**, 1145926 (2023).

32. Guo, T. et al. Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* **21**, 407–413 (2015).

33. Cai, X. et al. High-throughput proteomic sample preparation using pressure cycling technology. *Nat. Protoc.* **17**, 2307–2325 (2022).

34. Meier, F. et al. diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).

35. Mule, C. et al. Higher RET gene expression levels do not represent an alternative RET activation mechanism in medullary thyroid carcinoma. *Biomolecules* **11**, 1542 (2021).

36. Marques, I. J. et al. Identification of SPRY4 as a novel candidate susceptibility gene for familial nonmedullary thyroid cancer. *Thyroid* **31**, 1366–1375 (2021).

37. Krishnan, A. et al. Proteogenomics analysis unveils a TGF-RET gene fusion and druggable targets in papillary thyroid carcinomas. *Nat. Commun.* **11**, 2056 (2020).

38. Renaud, E. et al. Multiomic analysis of papillary thyroid cancers identifies BAIAP2L1-BRAF fusion and requirement of TRIM25, PDE5A and PKCdelta for tumorigenesis. *Mol. Cancer* **21**, 195 (2022).

39. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).

40. Meyer, H. J. & Rape, M. Enhanced protein degradation by branched ubiquitin chains. *Cell* **157**, 910–921 (2014).

41. Haakonsen, D. L. & Rape, M. Branching out: improved signaling by heterotypic ubiquitin chains. *Trends Cell Biol.* **29**, 704–716 (2019).

42. Wenzel, E. M. et al. Intercellular transfer of cancer cell invasiveness via endosome-mediated protease shedding. *Nat. Commun.* **15**, 1277 (2024).

43. Gerstberger, S., Jiang, Q. & Ganesh, K. Metastasis. *Cell* **186**, 1564–1579 (2023).

44. Stelzer, G. et al. The GeneCards Suite: from gene data mining to disease genome sequence analyses. *Curr. Protoc. Bioinforma.* **54**, 1 30 31–31 30 33 (2016).

45. UniProt, C. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).

46. Wang, X. et al. UbiBrowser 2.0: a comprehensive resource for proteome-wide known and predicted ubiquitin ligase/deubiquitinase-substrate interactions in eukaryotic species. *Nucleic Acids Res.* **50**, D719–D728 (2022).

47. Szklarczyk, D. et al. The STRING database in 2025: protein networks with directionality of regulation. *Nucleic Acids Res.* **53**, D730–D737 (2024).

48. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

49. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

50. Barresi, G. & Tuccari, G. Immunocytochemical demonstration of melanotransferrin (p97) in thyroid tumors of follicular cell origin. *Pathology* **26**, 127–129 (1994).

51. Yamaguchi, H. et al. Peptidomic identification and biological validation of neuroendocrine regulatory peptide-1 and -2. *J. Biol. Chem.* **282**, 26354–26360 (2007).

52. Hamm, A. et al. Frequent expression loss of Inter-alpha-trypsin inhibitor heavy chain (ITIH) genes in multiple human solid tumors: a systematic expression analysis. *BMC Cancer* **8**, 25 (2008).

53. Chang, Q. H. et al. Pan-cancer analysis identifies ITIH1 as a novel prognostic indicator for hepatocellular carcinoma. *Aging* **13**, 11096–11119 (2021).

54. Kopylov, A. T. et al. Revelation of proteomic indicators for colorectal cancer in initial stages of development. *Molecules* **25**, 619 (2020).

55. Jiang, W. et al. PAFAH1B3 regulates papillary thyroid carcinoma cell proliferation and metastasis by affecting the EMT. *Curr. Med. Chem.* **31**, 1152–1164 (2024).

56. Rayman, M. P. Selenium and human health. *Lancet* **379**, 1256–1268 (2012).

57. Benjamin, D. I. et al. Ether lipid generating enzyme AGPS alters the balance of structural and signaling lipids to fuel cancer pathogenicity. *Proc. Natl. Acad. Sci. USA* **110**, 14912–14917 (2013).

58. Chen, Z. et al. Ether phospholipids are required for mitochondrial reactive oxygen species homeostasis. *Nat. Commun.* **14**, 2194 (2023).

59. Bockstaele, L. et al. Regulated activating Thr172 phosphorylation of cyclin-dependent kinase 4(CDK4): its relationship with cyclins and CDK “inhibitors. *Mol. Cell Biol.* **26**, 5070–5085 (2006).

60. Weinberger, P. et al. Cell cycle M-phase genes are highly upregulated in anaplastic thyroid carcinoma. *Thyroid* **27**, 236–252 (2017).

61. Wu, J. et al. Expression of epithelial-mesenchymal transition regulators TWIST, SLUG and SNAIL in follicular thyroid tumours may relate to widely invasive, poorly differentiated and distant metastasis. *Histopathology* **74**, 780–791 (2019).

62. Roncalli, M. et al. Prognostic value of N-myc immunoreactivity in medullary thyroid carcinoma. *Cancer* **74**, 134–141 (1994).

63. Torricelli, F. et al. Medullary thyroid carcinomas classified according to the international medullary carcinoma grading system and a Surveillance, Epidemiology, and End Results-based metastatic risk score: a correlation with genetic profile and angioinvasion. *Mod. Pathol.* **36**, 100244 (2023).

64. Liu, B., Peng, Y., Su, Y., Diao, C. & Cheng, R. Treatment and management of medullary thyroid microcarcinoma: a 10-year retrospective study from a single center. *Endocrine* **86**, 1081–1089 (2024).

65. Erovic, B. M. et al. Prognostic and predictive markers in medullary thyroid carcinoma. *Endocr. Pathol.* **23**, 232–242 (2012).

66. Rios, A. et al. Prognostic value of histological and immunohistochemical characteristics for predicting the recurrence of medullary thyroid carcinoma. *Ann. Surg. Oncol.* **17**, 2444–2451 (2010).

67. Pellegriti, G. et al. Long-term outcome of medullary thyroid carcinoma in patients with normal postoperative medical imaging. *Br. J. Cancer* **88**, 1537–1542 (2003).

68. Kameyama, K., Okinaga, H. & Takami, H. RET oncogene mutations in 75 cases of familial medullary thyroid carcinoma in Japan. *Biomed. Pharmacother.* **58**, 345–347 (2004).

69. Jung, K. Y. et al. Genotypic characteristics and their association with phenotypic characteristics of hereditary medullary thyroid carcinoma in Korea. *Surgery* **164**, 312–318 (2018).

70. Chung, Y. J. et al. RET proto-oncogene mutations are restricted to codon 634 and 618 in Korean families with multiple endocrine neoplasia 2A. *Thyroid* **14**, 813–818 (2004).

71. Wang, J. et al. Screening of RET gene mutations in Chinese patients with medullary thyroid carcinoma and their relatives. *Fam. Cancer* **15**, 99–104 (2016).

72. Qi, X. P. et al. RET proto-oncogene genetic screening of families with multiple endocrine neoplasia type 2 optimizes diagnostic and clinical management in China. *Thyroid* **22**, 1257–1265 (2012).

73. Zhou, Y. et al. RET proto-oncogene mutations are restricted to codons 634 and 918 in mainland Chinese families with MEN2A and MEN2B. *Clin. Endocrinol.* **67**, 570–576 (2007).

74. Zhang, X. et al. Is new American Thyroid Association risk classification for hereditary medullary thyroid carcinoma applicable to Chinese patients? A single-center study. *Chin. J. Cancer Res* **29**, 223–230 (2017).

75. Hu, H. et al. CRL4B catalyzes H2AK119 monoubiquitination and coordinates with PRC2 to promote tumorigenesis. *Cancer Cell* **22**, 781–795 (2012).

76. Li, Y. & Wang, X. The role of cullin4B in human cancers. *Exp. Hematol. Oncol.* **6**, 17 (2017).

77. Gu, H. et al. CUL4B regulates thyroid cancer differentiation and treatment sensitivity by ubiquitinating ARID1A. *Transl. Oncol.* **56**, 102389 (2025).

78. Liu, J. et al. E3 ubiquitin ligase TRIM32 negatively regulates tumor suppressor p53 to promote tumorigenesis. *Cell Death Differ.* **21**, 1792–1804 (2014).

79. Huang, N. et al. TRIM family contribute to tumorigenesis, cancer development, and drug resistance. *Exp. Hematol. Oncol.* **11**, 75 (2022).

80. Zhou, Q. et al. TSPAN18 facilitates bone metastasis of prostate cancer by protecting STIM1 from TRIM32-mediated ubiquitination. *J. Exp. Clin. Cancer Res.* **42**, 195 (2023).

81. Xu, J. et al. TRIM32 promotes anoikis resistance and metastasis in NSCLC by degrading CHEK2 to enhance IL-6 secretion. *Cell Death Differ.* (2025).

82. Hou, Y. et al. The neurotransmitter calcitonin gene-related peptide shapes an immunosuppressive microenvironment in medullary thyroid cancer. *Nat. Commun.* **15**, 5555 (2024).

83. Flem-Karlsen, K., Fodstad, O., Tan, M. & Nunes-Xavier, C. E. B7-H3 in cancer - beyond immune regulation. *Trends Cancer* **4**, 401–404 (2018).

84. Hincza-Nowak, K. et al. CD276 as a candidate target for immunotherapy in medullary thyroid cancer. *Int. J. Mol. Sci.* **24** (2023).

85. Zhu, Y. et al. High-throughput proteomic analysis of FFPE tissue samples facilitates tumor stratification. *Mol. Oncol.* **13**, 2305–2328 (2019).

86. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).

87. Demichev, V. et al. dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nat. Commun.* **13**, 3944 (2022).

88. Li, L. et al. Comprehensive mass spectral libraries of human thyroid tissues and cells. *Sci. Data* **11**, 1448 (2024).

89. Steger, M. et al. Time-resolved *in vivo* ubiquitinome profiling by DIA-MS reveals USP7 targets on a proteome-wide scale. *Nat. Commun.* **12**, 5399 (2021).

90. Wei, R. et al. GSimp: A Gibbs sampler based left-censored missing value imputation approach for metabolomics studies. *PLoS Comput. Biol.* **14**, e1005973 (2018).

91. Wang, S. et al. NAguideR: performing and prioritizing missing value imputations for consistent bottom-up proteomic analyses. *Nucleic Acids Res* **48**, e83 (2020).

92. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

93. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinforma.* **11**, 367 (2010).

94. Newman, A. M. et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

95. Rieckmann, J. C. et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat. Immunol.* **18**, 583–593 (2017).

96. Sun, Y. et al. Artificial intelligence defines protein-based classification of thyroid nodules. *Cell Discov.* **8**, 85 (2022).

97. Zang, Z. et al. DMT-EV: an explainable deep network for dimension reduction. *IEEE Trans. Vis. Comput. Graph* **30**, 1710–1727 (2024).

98. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

99. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

100. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

101. Deutsch, E. W. et al. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **51**, D1539–D1548 (2023).

102. Perez-Riverol, Y. et al. The PRIDE database at 20 years: 2025 update. *Nucleic Acids Res.* **53**, D543–D553 (2025).

103. Ma, J. et al. iProX: an integrated proteome resource. *Nucleic Acids Res.* **47**, D1211–D1217 (2019).

104. Chen, T. et al. iProX in 2021: connecting proteomics data sharing with big data. *Nucleic Acids Res.* **50**, D1522–D1527 (2022).

## Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFF0608403 to Yi Zhu, 2021YFA1301600 to Tiannan Guo), the Pioneer and Leading Goose R&D Program of Zhejiang (2024SSYS0035 to Tiannan Guo), the National Natural Science Foundation of China (U24A20476 to Tiannan Guo, 82473361 and 82072951 to Yu Wang, 82373008, 82573036 and 82002830 to Xiao Shi, 82472629 to Zhiyan Liu), the Natural Science Fund of Science and Technology Commission of Shanghai Municipality (23ZR1448200 to Zhiyan Liu) and the Natural Science Foundation of Hangzhou (2025SZRJJ0471 to Yingrui Wang). We thank the Westlake University Supercomputer Center and the Flow Cytometry Platform for Biomedical Research Core Facilities of Westlake University for data storage, computation, and histology staining. We thank the Westlake Thyroid Proteome Consortium (We-TEC) collaborators for their support during the project.

## Author contributions

T.G., Yu Wang, Z.L., and Y.S. supervised the project. T.G., Y.Zhou, and Yingrui Wang conceived and designed the study. Yu Wang, Z.L., C.C., Y.Wu, Jianbiao Wang, H.Z., M.L., Y.H., X.S., L.Z., Z.G., Jiaxi Wang, H.L., Z.W., and C.S. collected samples and clinical information. Z.L. and L.Z. performed the pathological evaluation. Yingrui Wang, Jiatong Wang, P.H., and Z.G. conducted proteomic data acquisition. J.W., Z.G., Y.Zhou, and X.S. conducted genomic data acquisition. Y.Zhou and Z.Z. conducted data analysis. Y.Zhou, Yingrui Wang, and T.G. drafted the manuscript. All authors provided critical revisions of the manuscript and approved the final version.

## Competing interests

T.G. and Y.Z. are shareholders of Westlake Omics (Hangzhou) Biotechnology Co., Ltd. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-67533-7>.

**Correspondence** and requests for materials should be addressed to Yaoting Sun, Zhiyan Liu, Yu Wang or Tiannan Guo.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025

<sup>1</sup>College of Life Sciences, Zhejiang University, Hangzhou, China. <sup>2</sup>Affiliated Hangzhou First People's Hospital, State Key Laboratory of Medical Proteomics, Hangzhou, China. <sup>3</sup>Westlake Center for Intelligent Proteomics, Westlake Laboratory of Life Sciences and Biomedicine, Hangzhou, China. <sup>4</sup>Research Center for Industries of the Future, School of Life Sciences, Westlake University, Hangzhou, China. <sup>5</sup>Department of Head and Neck Surgery, Fudan University Shanghai Cancer Center, Shanghai, China. <sup>6</sup>Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China. <sup>7</sup>AI Lab, Research Center for Industries of the Future, Westlake University, Hangzhou, China. <sup>8</sup>Department of Pathology, Shanghai Sixth People's Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China. <sup>9</sup>Department of Thyroid Surgery, the First Hospital of China Medical University, Shenyang, China. <sup>10</sup>Department of Breast and Thyroid Surgery, Renmin Hospital of Wuhan University, Wuhan, China. <sup>11</sup>Department of Thyroid Surgery, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>12</sup>Department of Endocrinology, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical University, Guangzhou, China. <sup>13</sup>Department of Endocrinology, Zhejiang Provincial People's Hospital, Affiliated People's Hospital, Hangzhou, China. <sup>14</sup>Department of Urology, The Second Hospital of Dalian Medical University, Dalian, China. <sup>15</sup>Department of Pathology, Liaoning Cancer Hospital & Institute, Shenyang, China. <sup>16</sup>Department of Head and Neck Surgery, Institute of Micro-Invasive Surgery of Zhejiang University, The Affiliated Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China. <sup>17</sup>These authors contributed equally: Yan Zhou, Yingrui Wang, Xiao Shi, Jiatong Wang, Zelin Zang, Likun Zhang.  e-mail: [suyaoting@westlake.edu.cn](mailto:sunyaoting@westlake.edu.cn); [zhiyanliu@shsmu.edu.cn](mailto:zhiyanliu@shsmu.edu.cn); [neck130@sina.com](mailto:neck130@sina.com); [guotiannan@westlake.edu.cn](mailto:guotiannan@westlake.edu.cn)