Supplementary materials

Material and methods

Protein extraction

The samples were deparaffinized with xylene and washed with acetonitrile (ACN). Thereafter, 25 mM of ammonium bicarbonate was added, followed by dithiothreitol at a final concentration of 5 mM, and heated at 95 °C for 30 min. Protein lysates without sodium dodecyl sulfate were added using a sonicator to promote proteolysis. Dithiothreitol was added at a final concentration of 10 mM and heated for 30 min at 37 °C. Iodoacetamide was added at a final concentration of 55 mM, and the cells were incubated in the dark for 45 min.

Proteolysis

For each sample, 5–20 μg of protein solution was taken, and 2.5 μg of trypsinase was added at a protein-to-enzyme ratio of 40:1, and the enzyme was digested at 37 ºC for 4 h. The digested peptides were desalted using a Strata X column and vacuum-dried.

High pH separation

Peptides of equal amounts were extracted from the samples and mixed, and the mixture was diluted with mobile phase A (5% ACN, pH 9.8) and injected. The Shimadzu LC-20AB high-performance liquid chromatography (HPLC) system coupled with a Gemini high pH C18 column (5 μm, 4.6 × 250 mm) was used. The

Shao M, *et al. J Clin Pathol* 2024;0:1–9. doi: 10.1136/jcp-2024-209489

sample was subjected to the column and eluted at a flow rate of 1 mL/min using a gradient of 5% mobile phase B (95% CAN, pH 9.8) for 10 min, 5–35% mobile phase B for 40 min, and 35–95% mobile phase B for 1 min. Flow phase B lasted 3 min, and 5% mobile phase B was equilibrated for 10 min. The elution peak was monitored at a wavelength of 214 nm, and the components were collected every minute. The components were combined into 10 fractions and freeze-dried.

Data-dependent acquisition (DDA) library construction and data-independent acquisition (DIA) quantitative detection (Nano-LC-MS/MS)

The dried peptides were reconstituted with mobile phase A (2% ACN, 0.1% formic acid [FA]), centrifuged at 20 000 ×g for 10 min, and the supernatant was used for injection. The peptides were separated using a Thermo UltiMate 3 000 ultra-HPLC (UHPLC). The peptides were first enriched in a trap column and desalted, and then separated on a tandem self-packed C18 column (150 μm internal diameter, 1.8 μm column size, and 35 cm column length) with a flow rate of 500 nL/min. The separations were directly injected into a mass spectrometer (MS) for analysis.

DDA library construction

For DDA library construction, each peptide fraction was separated using an HPLC gradient for 180 min before MS: 0–5 min, 5% mobile phase B (98% ACN, 0.1% FA); 5–120 min, mobile phase B linearly increased from 5% to 25%; 120–160 min, mobile phase B increased from 25% to 35%; 160–170 min, mobile phase B increased from

35% to 80%; 170–175 min, 80% mobile phase B; and 175–180 min, 5% mobile phase B. The liquid chromatography (LC)-separated peptides were ionized using nanoelectrospray ionization (nanoESI) and injected into a tandem MS (Q-Exactive HF, Thermo Fisher Scientific, San Jose, CA) in DDA detection mode. The main settings were: ion source voltage, 1.9 kV; MS scan range, 350–1 500 m/z; MS resolution, 120 000; and maximal injection time. (MIT) 100 ms. For MS/MS collision-type higher-energy C-trap dissociation (HCD), the collision energy NCE, MS/MS resolution, MIT, and dynamic exclusion duration were 28, 30 000, 100 ms, and 30 s, respectively. The starting m/z for MS/MS was fixed at 100. The precursor for the MS/MS scan satisfied the charge range of 2+–6+, with the top 20 precursors having an intensity >5E4. The automatic gain control (AGC) was MS 3E6 and MS/MS 1E5.

DIA sample analysis

For single-sample DIA analysis, each sample was separated using an HPLC gradient for 120 min before MS: 0–5 min, 5% mobile phase B (98% ACN, 0.1% FA); 5–90 min, mobile phase B linearly increased from 5% to 25%; 90–100 min, mobile phase B increased from 25% to 35%; 100–105 min, mobile phase B increased from 35% to 80%; 105–115 min, 80% mobile phase B; and 115–120 min, 5% mobile phase B. The LC-separated peptides were ionized using nanoESI and injected into a tandem MS (Q-Exactive HF, Thermo Fisher Scientific, San Jose, CA) in DDA detection mode. The main settings were as follows: ion source voltage, 1.9 kV; MS scan range, 400–1 250 m/z; MS resolution, 120 000; and MIT, 50 ms, and were equally divided into 45

continuous window MS/MS scans. The MS/MS collision-type HCD and MIT were in the auto mode. Fragment ions were scanned using Orbitrap; MS/MS resolution was 30 000; collision energy was distributed across three modes: 22.5, 25, and 27.5; and the AGC was 1E6.

DDA library construction (Nano-LC-MS/MS)

The dried peptide samples were reconstituted with mobile phase A (2% ACN, 0.1% FA), centrifuged at 20 000 ×*g* for 10 min, and the supernatant was collected for injection. The separation was performed using a Thermo UltiMate 3000 UHPLC. The sample was first enriched in the trap column and desalted, entered a tandem self-packed C18 column (150 μm internal diameter, 1.8 μm column size, and 35 cm column length), and separated at a flow rate of 500 nL/min by the following effective gradient: 0–5 min, 5% mobile phase B (98% ACN, 0.1% FA); 5–120 min, mobile phase B linearly increased from 5% to 25%; 120–160 min, mobile phase B was increased from 25% to 35%; 160–170 min, mobile phase B increased from 35% to 80%; 170–175 min, 80% mobile phase B; and 175.5–180 min, 5% mobile phase B. The nanoliter liquid-phase separation end was directly connected to the MS using the following settings.

DIA quantification (Nano-LC-MS/MS)

The dried peptide samples were reconstituted with mobile phase A (2% ACN, 0.1% FA), centrifuged at 20 000 ×*g* for 10 min, and the supernatant was collected for

injection. The separation was performed using a Thermo UltiMate 3000 UHPLC. The sample was first enriched in the trap column and desalted, entered a tandem self-packed C18 column (150 μm internal diameter, 1.8 μm column size, 35 cm column length), and separated at a flow rate of 500 nL/min by the following effective gradient: 0–5 min, 5% mobile phase B (98% ACN, 0.1% FA); 5–90 min, mobile phase B linearly increased from 5% to 25%; 90–100 min, mobile phase B increased from 25% to 35%; 100–108 min, mobile phase B increased from 35% to 80%; 108–113 min, 80% mobile phase B; and 113.5–120 min, 5% mobile phase B. The nanoliter liquid-phase separation end was directly connected to the MS using the following settings.

DDA library construction detection

The peptides separated via liquid phase chromatography were ionized using a nanoESI source and passed through an Oritrap Exploris 480 tandem MS (Thermo Fisher Scientific, San Jose, CA) for DDA mode detection. The main parameters were: ion source voltage, 1.9 Kv; MS1 scanning range, 350–1 650 m/z; resolution, 120 000; and MIT, 90 ms. For MS/MS collision type HCD, collision energy NCE was 30, MS/MS resolution was 30 000, MIT was in the auto mode, and dynamic exclusion duration was 120 s. The start m/z for MS/MS was fixed to the auto mode. The precursor for the MS/MS scan satisfied the charge range of 2+–6+, with the top 30 precursors having an intensity >2E4. AGC was MS 300% and MS/MS 100%.

DIA mass spectrometry detection

The peptides separated via liquid phase chromatography were ionized using a nanoESI source and passed through an Oritrap Exploris 480 tandem MS (Thermo Fisher Scientific, San Jose, CA) for DDA mode detection. The main parameters were: ion source voltage, 1.9 kV; MS1 scanning range, 400–1 250 m/z; resolution, 120 000; and MIT, 90 ms, and were equally divided into 50 continuous window MS/MS scans. For MS/MS collision-type HCD, collision energy NCE was 30 and MIT was in the auto-mode. The fragment ions were scanned using an Orbitrap with an MS/MS resolution of 30 000. AGC was MS 300% and MS/MS 1000%.

Proteome data analysis

DDA data were identified using the Andromeda search engine within MaxQuant, and identification results were used for spectral library construction. For large-scale DIA data, the mProphet algorithm was used to complete analytical quality control; hence, many reliable quantitative results were obtained. Gene ontology, clusters of orthologous groups, and pathway functional annotation analyses were also performed. Based on the quantitative results, the differentially expressed proteins between the comparison groups were identified, and functional enrichment analysis and protein–protein interaction of the differentially expressed proteins were performed.

RNA extraction of formalin-fixed paraffin-embedded samples

We added 1 mL of xylene to the sample and incubated it in a thermo mixer at 50 °C

and 350 rpm for 3 min. Centrifugation was performed at room temperature and 17 500 ×*g* for 5 min, and the supernatant was discarded to remove the xylene. Anhydrous ethanol (1 mL) was added, and the mixture was centrifuged at 17 500 ×*g* for 3 min at room temperature to remove the ethanol. We added 200 µL of digestion buffer and 4 µL of protease. We then incubated the solution in a thermomixer at 50 °C and 350 rpm for 30 min and subsequently incubated at 80 °C and 350 rpm for 15 min. Subsequently, we added 240 µL of isolation additive and 550 µL of anhydrous ethanol. We added 700 µL of the mixture to a collection tube fitted with a filter cartridge and centrifuged at 10 000 ×*g* for 30 s–2 min at room temperature; subsequently, we transferred the filter cartridge to a new 2.0 mL collection tube. We added 30–50 µL of elution solution to the filter cartridge and centrifuged at 16 000 ×*g* for 2 min at room temperature to recover the RNA solution.

Sequencing

We sent the samples to BGI Genomics (Cambridge, MA, USA) for library preparation and sequencing. Library preparation was performed using the Trio RNA-seq kit (NuGEN, San Carlos, CA, USA) following the manufacturer's instructions. Briefly, rRNA depletion was performed to enrich the mRNA. The enriched mRNA was then fragmented into small pieces using a fragmentation buffer, purified using a QiaQuick polymerase chain reaction (PCR) extraction kit, resuspended in EB buffer, and subjected to end repair and poly (A) tail addition. Subsequently, the fragments were connected to adaptors. The library was then purified using the MiniElute PCR

purification kit before PCR amplification. The libraries were amplified using PCR, and the yield was quantified. Sequencing was performed using 100 bp-paired end sequencing.

Data filtering

The sequencing data were filtered using SOAPnuke by removing reads (1) containing sequencing adapters, (2) whose low-quality base ratio (base quality ≤15) was >20%, (3) whose unknown base ('N' base) ratio was >5%, afterwards clean reads were obtained and stored in a FASTQ format. Subsequent analysis and data mining were performed using Dr. Tom's multi-omics data mining system (https://biosys.bgi.com).

Reads alignment

Bowtie 2 was used to align the clean reads to the gene set, in which known, novel coding, and non-coding transcripts were included.

Gene quantification and differential expression analysis

Gene expression levels were calculated using RSEM (v1.3.1). A heatmap was drawn using pheatmap (v1.0.8) according to the gene expression differences in different samples. Essentially, differential expression analysis was performed using the DESeq2 (v1.4.5), DEGseq, or Poisson distribution with Q value ≤ 0.05 and the absolute value of log fold change >1.

Gene annotation

Gene ontology (http://www.geneontology.org/) and Kyoto Encyclopedia of Genes and Genomes (https://www.kegg.jp/) enrichment analyses of the annotated differentially expressed genes were performed using Phyper (https://en.wikipedia.org/wiki/Hypergeometric_distribution) based on a hypergeometric test. The significant levels of terms and pathways were corrected using a Q value with a rigorous threshold (Q value $\leqslant$ 0.05).

Supplementary table 1

Supplementary Table 1A: The top 100 genes upregulated in epithelioid PT compared to classical PT.

Gene	log2_EpithelioidClassical	Qvalue_EpithelioidClassical
'OR52A1'	24.31690269	2.06E-12
'TMEM196'	22.44754304	9.61E-11
'IBSP'	22.3964129	1.03E-10
'FCRL5'	22.32201492	1.18E-10
'ANKRD1'	14.21479593	2.74E-04
'ZNF256'	10.21052019	1.68E-04
'OR1C1'	9.664908736	1.13E-04
'SETDB2-PHF11'	9.397913755	0.001119662
'PPP2R2C'	8.905785249	2.42E-04
'SUPT20HL1'	7.978199533	1.59E-05
'LOC107985021'	7.853848524	2.38E-08
'ARMC7'	7.844707966	0.001394428
'CSAG2'	23.08774343	2.72E-11
'DEPDC7'	7.547416406	4.73E-06
'ABRA'	7.196103974	0.007615725
'PLEKHM1'	7.103391909	4.30E-06
'OR2AG2'	6.984670421	8.54E-05
'DGKE'	6.97600711	5.74E-08
'ZNF366'	6.921784074	0.003393989
'PIDD1'	6.887565973	2.15E-05
'ZNF846'	6.828415971	1.52E-06
'CGA'	21.15581364	1.82E-09
'OR51E2'	6.793701262	0.006600671
'CPNE8'	6.766425556	2.80E-07
'PLAU'	6.719327402	0.001001841
'XCL1'	6.717618171	0.001109309
'FAM234B'	6.683363795	1.43E-05
'VMAC'	6.668520445	0.002447549
'ANO5'	6.603869821	0.005394902
'OR2AG1'	6.547582884	5.01E-06
'OR8B8'	6.489868844	0.007717999
'LOC112267855'	23.64950278	1.06E-11
'LOC101929805'	6.384992277	1.43E-05
'OLFM3'	25.45154677	1.06E-13
'ZNF491'	6.277326415	2.96E-07
'OR52E2'	22.74340442	4.73E-11
'HLA-DQB2'	6.258359932	0.00240893
'PIGL'	6.216394702	5.52E-06
'HSPA6'	6.211828049	1.82E-04

'CCDC170'	6.17148545	0.004638637
'TRPC4'	6.155286167	5.59E-05
'C5orf63'	6.142659506	6.03E-09
'RBM15B'	6.097002673	1.82E-04
'GPR21'	6.014994463	1.33E-05
'RBM20'	5.90964845	7.45E-05
'CLVS2'	9.663584591	0.004192843
'CCL17'	5.903728108	4.61E-05
'CR2'	21.92305382	3.15E-10
'CADPS'	5.860785445	0.005036688
'PRG4'	5.850597866	3.05E-04
'F5'	5.844256666	1.93E-05
'SLC25A30'	5.839902301	9.23E-05
'SNRPB2'	5.792729197	2.04E-04
'THEMIS2'	5.782137111	1.87E-04
'MILR1'	5.760206156	3.02E-06
'CYP24A1'	5.749984419	0.004994974
'AGMO'	5.739546575	2.25E-08
'PDK3'	5.636927479	9.08E-05
'METRNL'	5.603974761	9.89E-05
'GRN'	5.589114501	6.57E-04
'OR3A2'	5.574608012	4.04E-04
'AGAP9'	5.532581892	8.95E-04
'LYSMD1'	5.530705273	3.08E-04
'SMIM5'	5.492154285	0.001879331
'SLC9A5'	5.432422716	9.58E-04
'NUBP1'	5.423087052	2.45E-04
'GNG2'	5.38875038	9.08E-05
'ICA1'	5.373438768	0.001399666
'CCDC102B'	5.360290169	3.70E-05
'GPR52'	5.357977285	0.001511453
'AKTIP'	5.333308059	1.82E-04
'CD96'	5.307707027	3.29E-06
'DOP1A'	5.299545147	4.15E-06
'GYPB'	5.291076866	0.002365232
'HEATR6'	5.290955621	8.89E-05
'PDE4C'	5.26870525	4.88E-04
'KCNIP4'	5.265569837	1.63E-04
'TCIRG1'	5.221786851	2.42E-04
'TOX'	5.217382501	0.001286549
'MKX'	5.183231524	0.001300726
'FAM186B'	5.15025192	0.003681024
'SCAI'	5.139424911	2.35E-05
'STAP1'	5.058776941	0.002692894

'TTC26'	5.05688589	7.54E-04
'UPK1B'	5.030705897	0.001001841
'EPHA7'	5.029612881	3.46E-04
'OR7D2'	4.925553391	0.00141971
'HHLA2'	4.924279256	9.93E-04
'EVC2'	4.915583893	2.33E-07
'L3MBTL1'	4.910767781	8.05E-04
'ERG'	4.879780122	3.05E-04
'ARL17B'	4.86436331	1.12E-05
'RGPD4'	4.856425137	7.66E-04
'ERMARD'	4.834505091	1.97E-04
'SEMA3C'	4.828269594	0.002580216
'C1QL3'	4.820412438	0.006769759
'S100A8'	4.755521339	0.007395818
'OR52N1'	4.680380461	4.35E-08
'FZD8'	4.620423231	0.002447549
'FAM153A'	4.59959155	0.002184959

Supplementary Table 1B: The top 100 genes downregulated in epithelioid PT compared to classical PT.

Gene	log2_EpithelioidClassical	Qvalue_EpithelioidClassical
'CLEC3A'	-17.81486992	6.13E-07
'ALPL'	-7.335891722	1.31E-06
'KLF15'	-6.880161589	0.001160622
'PRR15L'	-6.80654665	1.23E-04
'KCNB2'	-6.496252421	0.001171513
'HOXC12'	-6.293058183	0.002306164
'HSPB6'	-6.250534445	1.10E-05
'HOXA7'	-6.247455241	5.43E-05
'LYPD3'	-6.188590435	0.003750054
'SCGB2A1'	-6.167886733	0.003261714
'DDT'	-6.109968927	7.71E-05
'MRPS18A'	-6.036279061	0.009049638
'UGT2B11'	-6.023186937	0.005729287
'TMEM256'	-5.998298018	3.69E-04
'ROPN1'	-5.96211544	0.002260201
'MASP1'	-5.944130829	3.71E-04
'LRAT'	-5.821907927	4.88E-04
'SLC39A4'	-5.798366915	0.001479167
'PTX3'	-5.75694776	2.89E-04
'ACOT2'	-5.656407501	4.56E-04
'ATP6V1G2'	-5.61279233	0.005678555
'IL1RN'	-5.497143897	1.82E-04
'LAMC3'	-5.449209118	0.006782667
'TNXB'	-5.392970389	0.002430576
'RBP7'	-5.338951168	0.00844243
'CYP4F22'	-5.32250198	0.00397346
'SLC16A12'	-5.302791586	2.30E-04
'PENK'	-5.285055168	8.08E-04
'MUCL1'	-5.2753939	0.002318416
'MMP7'	-5.163972144	2.33E-05
'SLC24A3'	-5.159838365	4.23E-04
'DHRS4'	-5.149591876	0.008214256
'SCGB1D2'	-5.141643068	0.005188583
'JUND'	-5.110760589	5.60E-08
'ATAD3C'	-5.061101472	0.003696229
'POMGNT2'	-5.041250733	0.007097011
'ISLR'	-5.032938167	5.65E-07
'ECRG4'	-4.995282322	0.00395278
'ATP5MC1'	-4.991960148	0.00113162
'FBP1'	-4.982122779	0.00844243
'SERPINA5'	-4.973881721	0.001634297

'AAMP'	-4.95837116	6.44E-04
'UBL5'	-4.958284561	4.49E-06
'SCGB2A2'	-4.943049672	8.67E-04
'MFAP4'	-4.921542024	7.18E-05
'GIPC2'	-4.914443362	0.002000789
'ATP5ME'	-4.875260167	5.84E-04
'GRPR'	-4.867786952	0.0076586
'ELOB'	-4.84300387	3.32E-06
'FAM222A'	-4.827167321	0.006203139
'AZGP1'	-4.813221927	0.002556316
'TUSC1'	-4.797045392	0.001001841
'CYTL1'	-4.794555383	0.003097359
'COX4I1'	-4.717105304	1.32E-04
'LMOD1'	-4.711462271	2.03E-05
'TOMM22'	-4.693724134	1.16E-05
'ADH1B'	-4.646399633	0.004472055
'RPS21'	-4.637726933	1.43E-05
'UBE2M'	-4.635152901	0.005489742
'TMC4'	-4.601800319	0.002704382
'TMEM54'	-4.591560408	0.006739324
'SOX10'	-4.544769234	0.009184728
'TRIR'	-4.52249724	8.88E-08
'TUSC2'	-4.511154589	0.004339255
'SLPI'	-4.510748246	0.00177724
'EFHD1'	-4.479884766	8.86E-04
'MUC6'	-4.457991354	0.007027363
'MYCL'	-4.441310012	0.007636001
'SLC24A2'	-4.430517241	0.004091301
'DLK1'	-4.422150218	0.009820677
'VPS25'	-4.42159786	0.004192843
'ASB13'	-4.403158872	8.08E-04
'ZNF358'	-4.327247714	0.004352061
'SDF2'	-4.325973525	1.50E-04
'SERPINA3'	-4.290257549	0.002056034
'CSTA'	-4.264959855	0.002199819
'SMDT1'	-4.262911749	0.002788887
'CLDN8'	-4.247633677	0.002941499
'ALOX15B'	-4.239375206	0.006290105
'SFRP1'	-4.236244577	2.34E-04
'SAA1'	-4.234099553	0.005764225
'EFNA1'	-4.214946588	0.002867936
'MOSPD3'	-4.189371608	0.009507033
'RAB25'	-4.188147635	0.002720736
'SCN4B'	-4.178328001	0.00209872

'CNN1'	-4.144000808	0.001578109
'CRABP1'	-4.134478876	0.002290821
'PLEKHB1'	-4.10707348	0.00395278
'C15orf61'	-4.097709163	0.00151939
'FBL'	-4.08476993	1.10E-05
'NDUFB11'	-4.084091819	0.001069242
'CDC42EP4'	-4.066483885	1.30E-05
'RPL41'	-4.065523149	3.78E-04
'NDUFA7'	-4.042251697	0.002447549
'PHYHD1'	-4.039302318	9.18E-04
'OLFM2'	-4.012799848	0.001861799
'MSX2'	-4.00625068	0.0014483
'TYROBP'	-4.003614647	0.006138352
CDK2AP2'	-4.000405792	0.009282342
'RBX1'	-3.997984224	1.66E-04

Supplementary table 2

Expression of p16 and Rb in different grades of PT and epithelioid PT.

	P16+/Rb+	P16+/Rb-	P16-/Rb+	P16-/Rb-
benign PT (N=7)	1	0	0	6
borderline PT (N=8)	0	1	0	7
malignant PT (N=5)	3	1	1	0
epitheloid PT (N=5)	0	3	1	1

Supplementary table 3

Pairwise comparison of p16 and Rb expression patterns among four groups.

	benign PT	borderline PT	malignant PT	epitheloid PT
benign PT				
borderline PT	1			
malignant PT	0.015*	0.005*		
epitheloid PT	0.072	0.032*	1	

*Statistical differences.

Supplementary figure 1

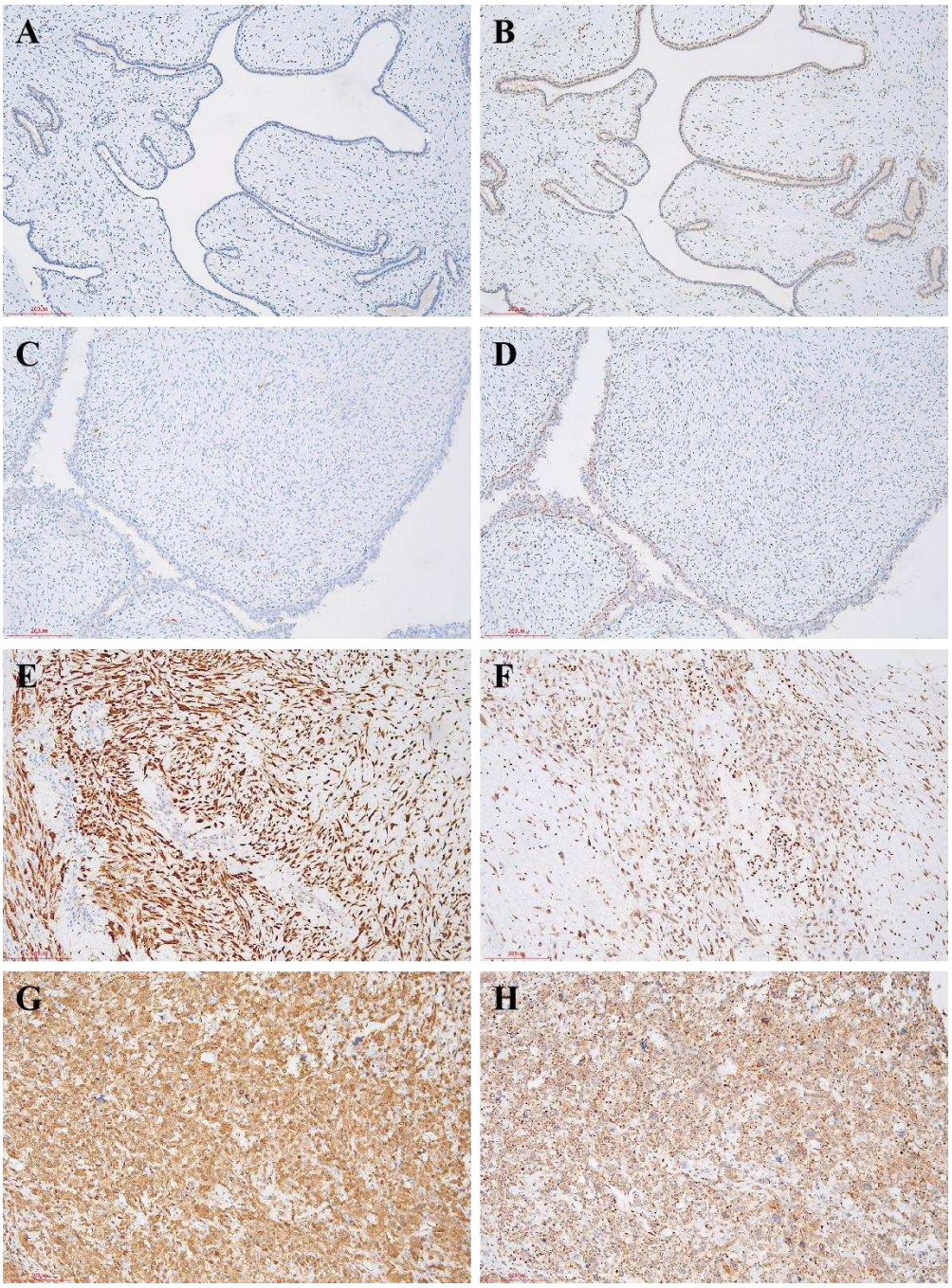

Figure legend

Supplementary figure 1

A-B. Tumor stromal cells in benign phyllodes tumors were negative for P16 and Rb. **C-D**. Tumor stromal cells in borderline phyllodes tumors were negative for P16, and locally weak positive for Rb. **E-F.** Stromal cells in malignant phyllodes tumors show diffuse strong positivity for P16 and Rb. **G-H.** Stromal cells in Epithelioid phyllodes tumors show diffuse strong positivity for P16 or Rb.

Shao M, *et al. J Clin Pathol* 2024;0:1–9. doi: 10.1136/jcp-2024-209489