Strategic Approach to Heterogeneity Analysis of Cutaneous Adnexal Carcinomas Using Computational Pathology and Genomics

Yuuki Nishimura[1,2], Eijitsu Ryo[3], Satoshi Inoue[4], Masahito Kawazu[4], Toshihide Ueno[4], Kenjiro Namikawa[5], Akira Takahashi[5], Dai Ogata[5], Akihiko Yoshida[1], Naoya Yamazaki[5], Hiroyuki Mano[4], Yasushi Yatabe[1,3,2] and Taisuke Mori[1,3]

Cutaneous adnexal tumors are neoplasms that arise from skin appendages. Their morphologic diversity and phenotypic variability with rare progression to malignancy make them difficult to diagnose and classify, and there is currently no established treatment strategy. To overcome these difficulties, this study investigated the transcription factor SOX9 expression, morphology, and genetics of skin adnexal tumors for understanding their biology, especially their histogenesis. We showed that cutaneous adnexal tumors and their nontumor counterparts of skin and appendages exhibit expression patterns similar to that of SOX9. Its expression intensity and pattern, as well as histopathologic evaluation of tumors, were analyzed using digital images of 69 normal skin adnexal 9-type organs and 185 skin adnexal 29-type tumors as references. It was possible to distinguish basal cell carcinoma from squamous cell carcinoma, sebaceous carcinoma, and pilomatrixoma with significant differences, along with porocarcinoma from squamous cell carcinoma. Furthermore, unsupervised machine learning "computational pathology" was used to derive a multiregion whole-exome sequencing fusion method termed "genocomputed pathology." The genocomputed pathology of three representable adnexal carcinomas (porocarcinoma, hidradenocarcinoma, and spiradenocarcinoma) was evaluated for total nine cases. We showed that there was more heterogeneity than expected within the tumors as well as the coexistence of components lacking driver fusion genes. The presence or absence of potential driver genes, such as *PIK3CA*, *YAP1*, and *PTEN*, in each region was identified, highlighting a therapeutic strategy for cutaneous adnexal carcinoma encompassing heterogeneous tumors.

INTRODUCTION

Skin adnexal tumors are neoplasms of the skin that show differentiation toward various skin appendages (Elder et al., 2018). They can be classified into following three main groups: hair follicle (HF) tumors, sebaceous tumors, and sweat gland tumors (eccrine and apocrine).

Adnexal tumors occasionally show multilineage differentiation (Elder et al., 2018). Although some adnexal carcinomas are clinically aggressive and exhibit poor clinical outcome, their morphological diversity and rarity of adnexal carcinomas make them difficult to diagnose and classify for precise management (Elder et al., 2018; Hile and Harms, 2021; Prieto-Granada et al., 2020); therefore, a better understanding of the molecular features of adnexal tumors is required. Recently, significant progress has been made toward understanding the molecular biology of these tumors (Hile and Harms, 2021). Notable findings include *YAP1*−*MAML2* and *YAP1*−*NUTM1* fusions in poroid neoplasms (Sekine et al., 2019) and the *ALPK1* hot spot alteration in spiradenomas and spiradenocarcinomas (Rashid et al., 2019). Nevertheless, there is a lack of exome or genome-wide studies and studies into the pathogenesis, such as the mutational signatures of these presumed nonultraviolet neoplasms (Hile and Harms, 2021; Lu and Fuchs, 2014).

Sex determining region Y−box 9 (SOX9) is a member of the SOX (Sry−type HMG box) family of transcription factors. SOX9 is essential during various organ development. In skin, studies using mouse models have revealed that SOX9 is not only pivotal for initial stem cell function and specification during HF and sebaceous organs (pilosebaceous unit) morphogenesis but also for other skin adnexal lineage including sweat glands. SOX9 is also required for maintenance of adult HF stem cells, which are thought to reside in structures called bulge and secondary hair germ which is located at the base of bulge (Jo

[1]Department of Diagnostic Pathology, National Cancer Center Hospital, Tokyo, Japan; [2]Course of Advanced Clinical Research of Cancer, Juntendo University Graduate School of Medicine, Tokyo, Japan; [3]Division of Molecular Pathology, National Cancer Center Reserch Institute, Tokyo, Japan; [4]Division of Cellular Signaling, National Cancer Center Reserch Institute, Tokyo, Japan; and [5]Dermatologic Oncology, National Cancer Center Hospital, Tokyo, Japan

Correspondence: Department of Diagnostic Pathology, National Cancer Center Hospital, 5-1-1, Chuo-ku, Tokyo, 104-0045, Japan. E-mail: tamori@ncc.go.jp

Abbreviations: BCC, basal cell carcinoma; FFPE, formalin-fixed paraffin-embedded; HF, hair follicle; IHC, immunohistochemistry; M-WES, multiregion whole-exome sequencing; ROI, regions of interest; SCC, squamous cell carcinoma; SOX9, sex determining region Y−box 9; TMB, tumour mutational burden; WES, whole-exome sequencing

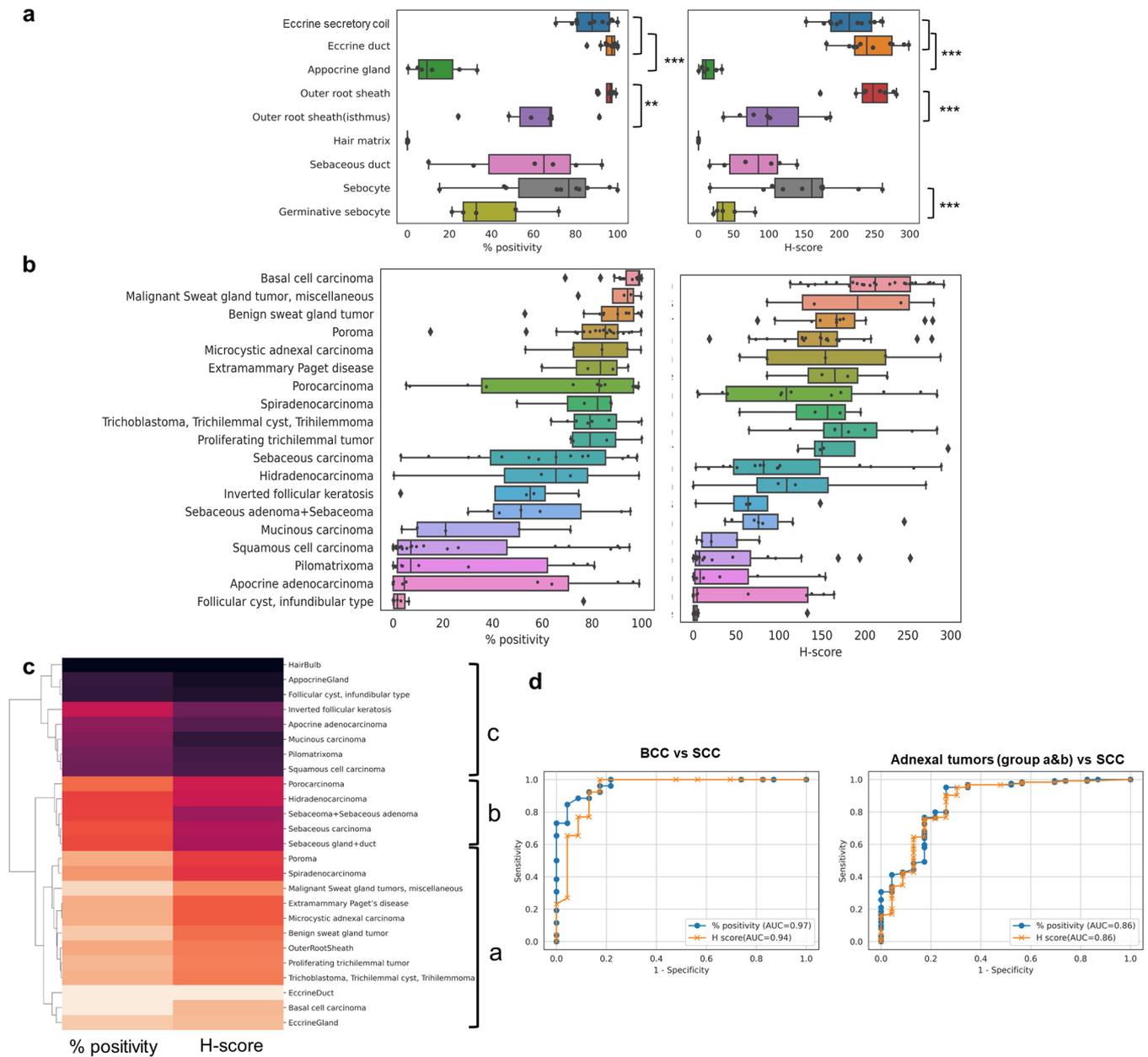Cite this article as: JID Innovations 2023;3:100229

Figure 1. Nontumor skin adnexa and adnexal tumors show characteristic SOX9 expression. Percent positivity and H-score of SOX9 IHC in normal skin adnexa **(a)** and skin adnexal tumors **(b)** All data points are shown; boxes define the interquartile range, with whiskers extending to lowest and highest data points. **$P <$ 0.01, ***$P <$ 0.001. "Benign sweat gland tumors" include tubular apocrine adenoma, hidradenoma, cylindroma, mixed tumor, spiradenoma, and syringoma. "Malignant Sweat gland tumor, miscellaneous" include adenoid cystic carcinoma, cribriform carcinoma, and digital papillary carcinoma. **(c)** Hierarchical clustering by standardized mean values of SOX9 IHC percent positivity and H-score. Skin adnexa and adnexal tumors were divided into three groups (a, b, c). **(d)** ROC curve for differential diagnosis of BCC versus SCC and adnexal tumors belonging to group a and b versus SCC. AUC, area under curve; BCC, basal cell carcinoma; SCC, squamous cell carcinoma.

et al., 2014; Lee et al., 2017; Ming et al., 2022; Morita et al., 2021; Nowak et al., 2008; Sarkar and Hochedlinger, 2013; Vidal et al., 2005). These important roles of SOX9 in skin appendages led to studies that focused on SOX9 expression in basal cell carcinoma (BCC) and other adnexal neoplasms (Kim et al., 2018; Krahl and Sellheyer, 2010; Vidal et al., 2005, 2008). More recent studies have focused on BCC, which presumably arises from hair bulges and secondary hair germs and naturally expresses SOX9 (Peterson et al., 2015).

However, SOX9 expression is not well-understood in most adnexal neoplasms. This study examined SOX9 expression in nontumor human adnexa as well as benign and malignant adnexal tumors using SOX9 immunohistochemistry (IHC) staining. We focused on three malignant sweat gland tumors (porocarcinoma, hidradenocarcinoma, and spiradenocarcinoma) and performed multiregion whole-exome sequencing (M-WES) to examine intratumoral heterogeneity, combined with digital image analysis and unsupervised machine-learning.
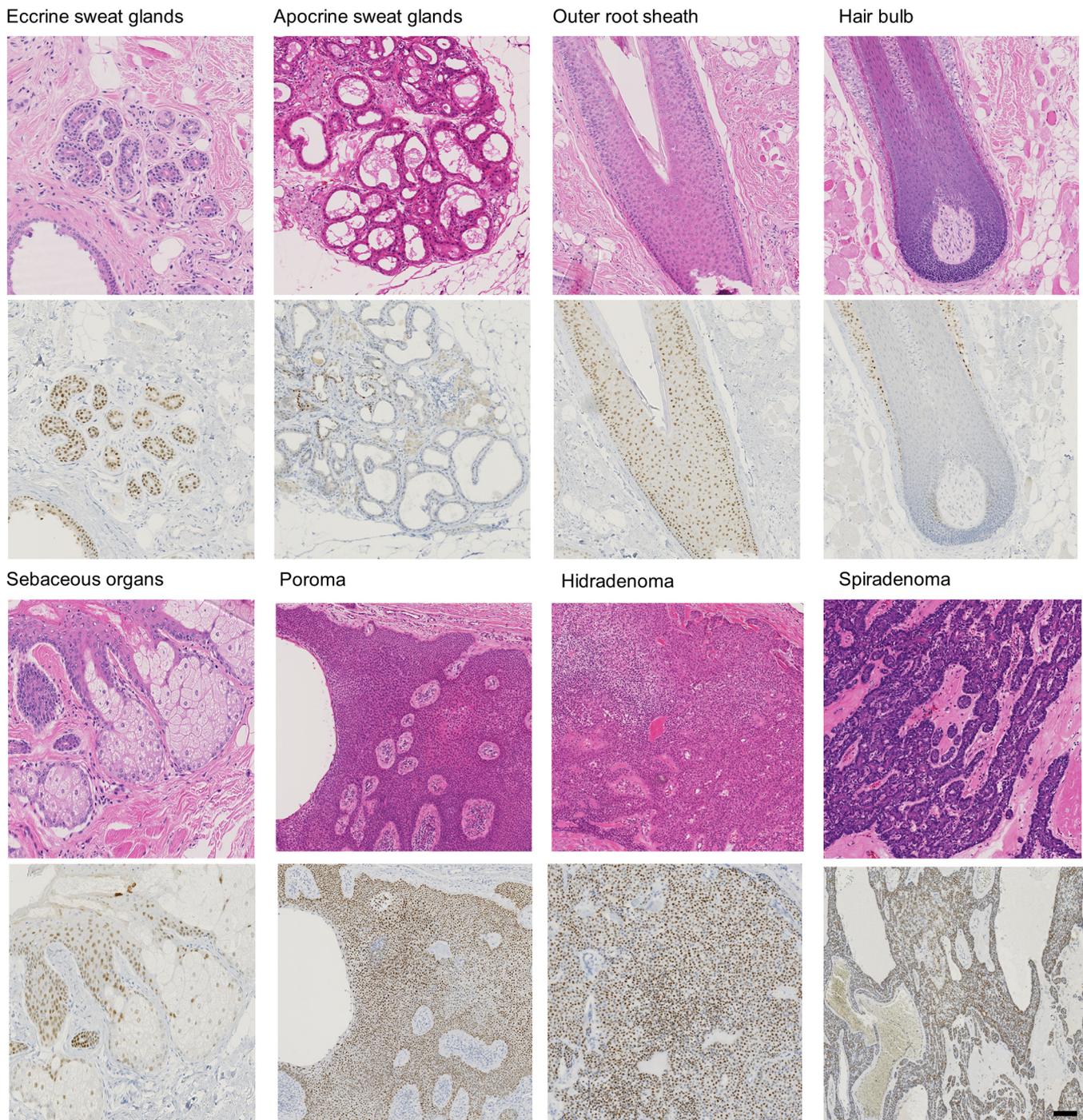
Figure 2. Representative H&E and SOX9 IHC images of normal skin adnexa and benign sweat gland tumors. Benign tumors include poroma, hidradenoma and spiradenoma. Bar = 100μm.

RESULTS

SOX9 expression is characteristic of nontumor human skin adnexa

We used digital whole-slide images and open-source software QuPath (Bankhead et al., 2017) to examine SOX9 expression using IHC (percentage positivity and H-score) of nontumor human adnexa (Figure 1a and Figure 2; Table 1). H-score is considered more useful for quantitatively evaluating protein expression than percentage positivity. SOX9 expression was high in eccrine ducts (96%, mean value) and

secretory coil (87%) but significantly lower in apocrine glands (14%, $P < 0.001$). In HFs, SOX9 expression was high in the outer root sheath (96%), lower in the isthmus of the outer root sheath (61%, $P < 0.01$), and was not expressed in the hair matrix (0%). These findings are consistent with those of a previous study that showed that β-catenin was exclusively expressed in the hair matrix (Krahl and Sellheyer, 2010). Sebaceous organs showed varied SOX9 expression, with percentage positivity and H-score ranging from low to high. Mature sebocytes generally showed higher SOX9

Table 1. SOX9 Expression Using IHC in Nontumor Human Adnexa

	% Positivity		H-score	
Skin adnexa	mean	median	mean	median
Appocrine gland (n = 6)	13.65	9.31	14.00	10.15
Eccrine duct (n = 10)	95.85	97.28	245.82	238.95
Eccrine secretory coli (n = 12)	87.50	87.88	214.83	214.09
Germinative sebocyte (n = 5)	40.81	32.76	42.86	34.48
Hair matrix (n = 5)	0.06	0.00	0.06	0.00
Outer root sheath (n = 7)	95.61	96.59	243.90	247.66
Outer root sheath (isthmus) (n = 7)	61.08	67.78	105.86	97.78
Sebaceous duct (n = 6)	57.36	64.92	79.68	85.08
Sebocyte (n = 10)	69.65	76.73	149.64	160.96

expression (69%) than germinative sebocytes (41%, $P = 0.05$), with significantly different H-scores ($P < 0.001$). Taken together, SOX9 expression was unique for each adnexa.

Skin adnexal tumors recapitulate SOX9 expression of corresponding nontumor adnexa
We next examined SOX9 expression in various skin adnexal tumors (Figure 1b and Figures 2–4; Tables 2 and 3). Most sweat gland tumors showed generally high SOX9 expression, whereas apocrine carcinoma and mucinous carcinoma showed very low expression. Porocarcinomas and hidradenocarcinomas, which are representative sweat gland carcinomas, showed highly variable SOX9 percentage positivity and H-scores. HF tumors, such as trichoblastomas and proliferating trichilemmal tumors, showed high SOX9 expression. Pilomatrixomas showed generally no or very low expression, although some exhibited high SOX9 expression, likely reflecting differentiation into organs other than hair matrix. Sebaceous tumors showed variable expression patterns similar to those of their corresponding nontumor sebaceous organs. All BCCs examined showed high percentage positivity and H-scores for SOX9 expression. SOX9 expression was also detected in some squamous cell carcinomas (SCCs) but was weakly expressed in a small proportion of cells.

We observed similar SOX9 expression in many adnexal tumors and their corresponding nontumor adnexa. Hierarchical clustering based on standardized mean SOX9 percentage positivity and H-score revealed that cutaneous nontumor adnexa and tumors were broadly clustered into three groups (Figure 1c). The group with the highest SOX9 expression (group a) included eccrine sweat glands, all benign sweat gland tumors, most malignant sweat gland tumors, and most HF tumors. The group with lowest SOX9 expression (group c) included hair matrix, pilomatrixoma, apocrine glands, and apocrine carcinoma. The group with intermediate SOX9 expression (group b) included sebaceous organs and sebaceous glands, hidradenocarcinoma, and porocarcinoma. Clustering analysis indicated that many adnexal tumors and their corresponding nontumor adnexa showed similar SOX9 expression.

We next explored the diagnostic and clinical use of SOX9. With receiver operating characteristic curves, percentage

positivity and H-score showed an outstanding (>0.9) area under the curve (Mandrekar, 2010) for differentiating BCC and SCC (0.97 and 0.94) and also showed an excellent area under the curve (0.8–0.9) for differentiating adnexal tumors belonging to groups a and b and SCC (both 0.86) (Figure 1d). For the distinction of several tumors, *P*-values, the sensitivity and specificity with cutoff of both SOX9 percent positivity and H-score are listed in Table 4. For BCC versus SCC, sensitivity 96% and specificity 82% were obtained with cutoff 80% positivity, and sensitivity 85% and specificity 87% were obtained with cutoff H-score 150. For BCC versus pilomatrixoma or sebaceous carcinoma, specificity was 90% and 73% for each with same cutoff 80% positivity. Thus, in practice, SOX9 diffuse staining would indicate BCC in these differential diagnoses. Similarly, for the distinction of several kinds of sweat gland tumors and apocrine adenocarcinoma, diffuse percentage positivity of SOX9 (80%) could be quite useful with sensitivity around 80% and specificity 75%. SOX9 percentage positivity and H-score were lower for malignant tumors with metastasis or recurrence than those without, though the differences were not statistically significant (Figure 5; Table 5).

Morphological single-cell unsupervised clustering revealed intratumoral heterogeneity
Analysis of SOX9 expression in various skin adnexal tumors revealed intertumor and intratumor divergence of expression, especially in malignancies. We hypothesized that some adnexal tumors could be biologically heterogenous, thus SOX9 expression as a differentiation marker would also be heterogenous. We focused on three representative malignant sweat gland tumors, porocarcinoma, hidradenocarcinoma, and spiradenocarcinoma, three cases for each. Three cases (porocarcinoma 1, hidradenocarcinoma 1, and spiradenocarcinoma 1) contained both pathologically benign and pathologically malignant regions. Other cases contained only pathologically malignant regions.

We first assessed intratumoral morphological variation. For each tumor, a sufficient number of regions of interest (ROIs) were set in one slide (Figure 6a–c and Figure 7a–f). Each ROI was selected rather randomly to decipher morphological heterogeneity in objective manner as possible. For each ROI, morphological data for every cell were measured with QuPath. StarDist, which is a deep learning–based method of nuclei segmentation, was used within QuPath for precise detection (Schmidt et al., 2018) (Figure 8a). Dimensional reduction of measurement data was performed using uniform manifold approximation and projection (McInnes et al., 2018). After selecting an appropriate number of clusters based on Akaike information criterion and Bayesian information criterion plots, the Gaussian-mixture model of unsupervised clustering of every cell was performed (Figure 8b–d, Figure 9a–c, and Figure 10a–c). The proportion of clustered cells for each ROI was calculated and the similarities with each other were compared with hierarchical clustering. In porocarcinoma 1, the ROIs of histologically benign regions were clustered into distinct group from ROIs in malignant regions, which were also clustered into several large clusters (Figure 8e). In hidradenocarcinoma 1, benign and malignant regions were also
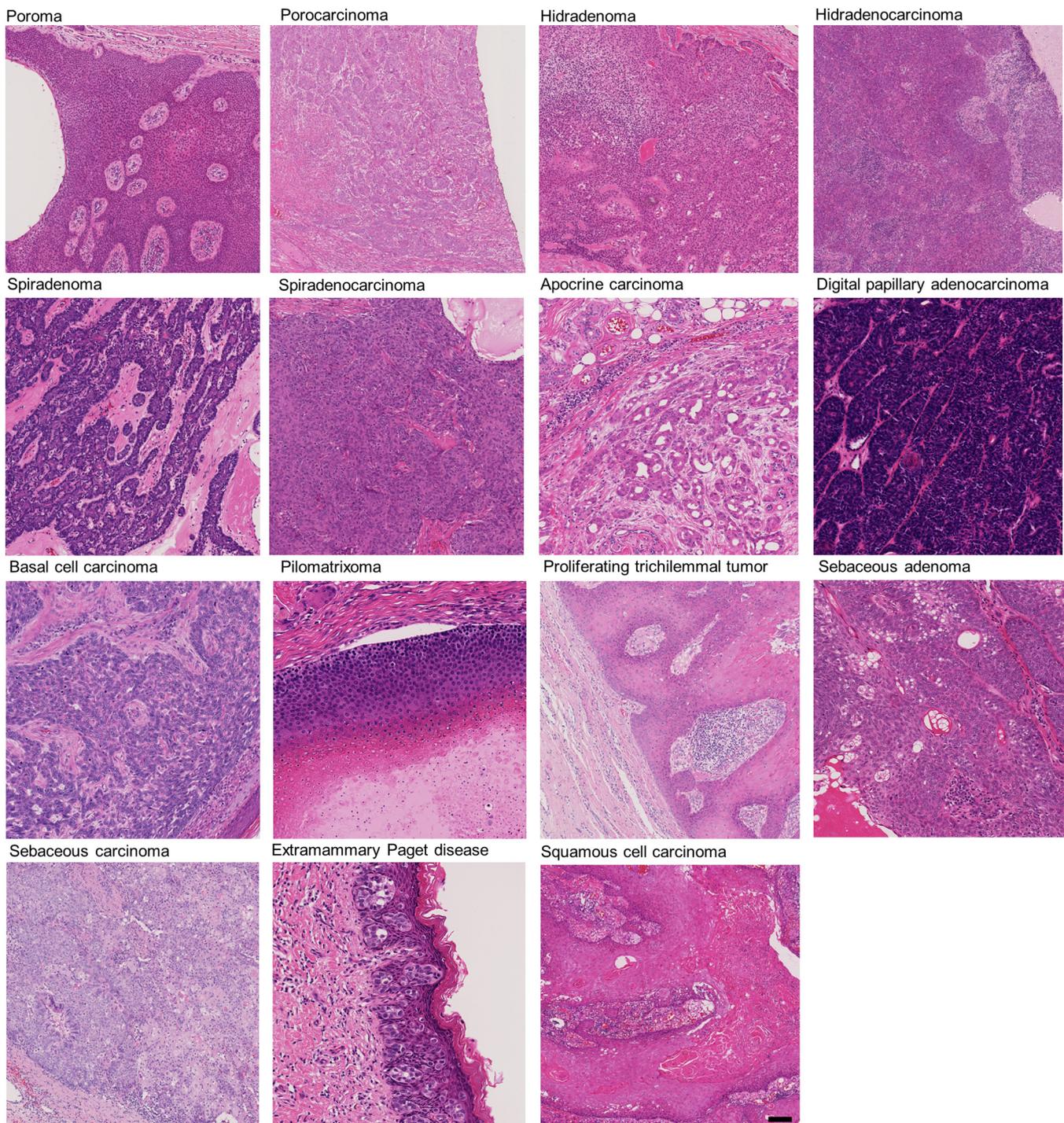
Figure 3. Representative H&E images of various adnexal tumors and SCC. Bar = 100μm.

clustered into separate groups and further clustered into several groups (Figure 9d). However, in spiradenocarcinoma 1, benign regions K and L were split into two. Region L was clearly distinct from other malignant ROIs, but K was in the same cluster of malignant P region. The ROIs of the malignant regions were clustered into several clusters (Figure 10d). ROIs in other tumors were clustered similarly into several distinct groups (Figure 11a−f)

M-WES revealed distinct genomic profiles of sweat gland carcinomas

We then performed sampling from multiple sites within one slide for M-WES to investigate genomic heterogeneity (Figure 6d−f, Figure 12a−f, and Figure 13). Sampling sites were basically chosen for each group (clusters) based on the above clustering results. We additionally measured the areas of each group and calculated their proportions in
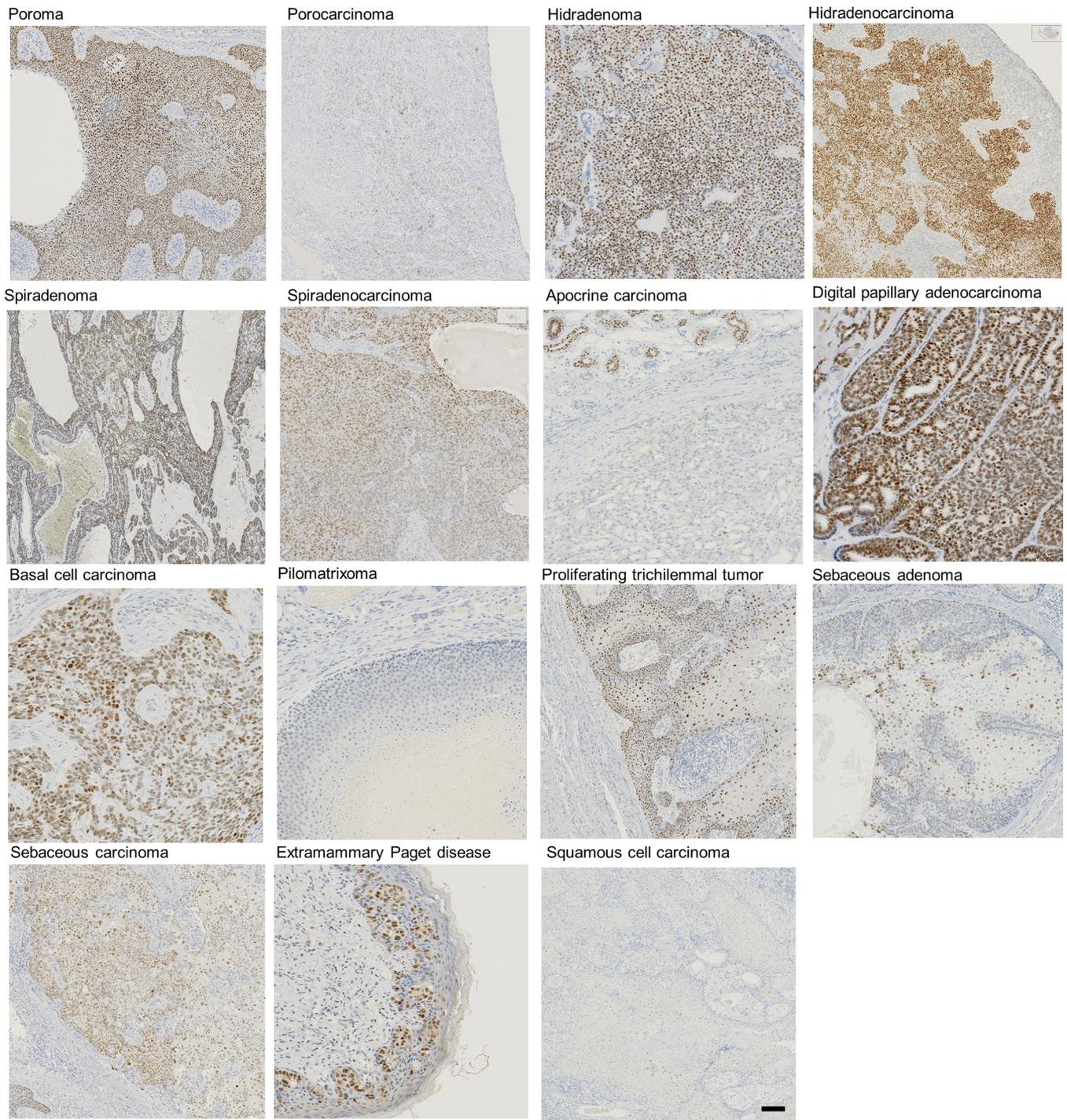
Figure 4. Representative SOX9 IHC images of adnexal tumors. IHC image of SCC is also shown. Bar = 100μm.

porocarcinoma 1, hidradenocarcinoma 1, and spiradenocarcinoma 1. SOX9 expression was clearly different among benign and malignant areas in these three cases. In porocarcinoma 1, tumor cells in the benign regions showed high SOX9 expression, whereas those in the malignant regions showed almost no SOX9 expression. Similar patterns of SOX9 expression were shown in hidradenocarcinoma 1, whereas in spiradenocarcinoma 1, tumor cells in malignant areas showed higher expression than benign areas. On the other hand, for other tumors, percentage positivity of SOX9 did not change significantly in each sampling sites. Somatic alterations and copy number aberrations in each sampling sites were compared for each tumor and phylogenetic trees were constructed (Figures 14—29).

Three cases of porocarcinoma were stained with NUT and YAP1 (both anti-N- and C-terminal) immunostaining, and FISH was also used to confirm the presence of the fusion gene (Macagno et al., 2021; Sekine et al., 2019).

Table 2. SOX9 Expression in Various Skin Adnexal Tumors

Pathological diagnosis	Total cases	Percentage positivity		H-score		Gender		Body site					Diameter, mm (mean)	Metastasis		Multiple		Recurrence	
		mean	median	mean	median	M	F	Head and neck	Trunk	axilla	limbs	Genitalia/groin		+	-	+	-	+	-
Adenoid cystic carcinoma	2	83.7	83.7	113.5	113.5	2		1				1	45		2		2		2
Apocrine adenocarcinoma	12	34.7	4.5	54.9	4.5	10	2	4	2	6			39	7	5		12		12
Basal cell carcinoma	26	95.6	98.8	212.3	212.0	14	12	21	3		1		11		26	3	23		26
Cribriform carcinoma	1	95.6	95.6	280.0	280.0		1		1				2		1		1		1
Cylindroma	1	99.9	99.9	201.0	201.0		1		1				40		1		1		1
Digital papillary adenocarcinoma	1	99.7	99.7	242.0	242.0	1	2				1		10		1		1		1
Extramammary Paget disease	4	80.3	83.4	160.5	165.0	2	2					4	91	1	3	1	3		4
Follicular cyst, infundibular type	7	12.6	1.7	20.6	2.0	6	1	3	3		1		13		7		7		7
Hidradenocarcinoma	4	57.5	65.5	122.3	109.0	2	2	1	1			2	30	2	2		4		4
Hidradenoma	3	86.5	83.1	195.0	167.0	0	3	1	2				13		2		2		2
Inverted follicular keratosis	4	47.0	55.1	69.8	64.0	2	2	4					7		4		4		4
Microcystic adnexal carcinoma	5	80.7	84.0	161.2	154.0	2	3	4	1				77	1	4		5		5
Mixed tumor	2	75.8	75.8	177.0	177.0	2		1			1		21		2		2		2
Mucinous carcinoma	5	31.3	21.1	32.6	21.0	5		2	1		1		19	1	4		5		5
Pilomatrixoma	10	28.1	7.1	42.9	8.0	6	4	6			4		12		10		10		10
Porocarcinoma	12	66.1	82.9	125.8	108.5	3	9	5	3		3	1	54	3	9		12		10
Poroma	20	80.0	85.3	145.6	148.5	13	7	6	7		7		10		20		20		20
Proliferating trichilemmal tumor	4	82.4	79.2	179.8	150.0		4	3	1				26		4		4		4
Sebaceoma	4	67.2	71.6	118.8	96.0	4		2	1		1		12		4		4		4
Sebaceous adenoma	3	46.6	42.7	65.7	71.0		3	3					7		3	1	2		3
Sebaceous carcinoma	15	60.9	65.5	108.5	82.0	5	10	11	4				17	2	13		15	2	11
Spiradenocarcinoma	4	75.5	82.2	140.5	156.5		4	3			1		32	3	1	1	3	2	2
Spiradenoma	2	94.3	94.3	131.0	131.0	2		1	1				19		2		2		2
Squamous cell carcinoma	23	26.3	7.1	46.2	7.0	14	9	5	2		14	2	34	3	20	6	17	3	20
Syringoma	2	84.6	84.6	156.5	156.5	1	1	1				1	5		2		2		2
Trichilemmal cyst	1	73.7	73.7	165.0	165.0		1	1					20		1		1		1
Trichilenmoma	1	78.3	78.3	181.0	181.0		1	1					2		1		1		1
Trichoblastoma	6	83.2	83.4	180.3	182.5	3	3	5	1				9		6		6		6
Tubular apocrine adenoma	1	90.3	90.3	171.0	171.0		1	1					10		1		1		1
Total	185					99	88	96	35	6	35	11		23	161	12	172	7	173

Abbreviations: F, female; M, male.

Table 3. Correlation between Skin Appendage Differentiation Tendency and Clinical Information

Clinicopathological factors	Tumors with follicular differentiation (n = 33)[1]	Tumours with sebaceous differentiation (n = 22)	Tumours with apocrine and eccrine differentiation (n = 77)[2]	P-value	Total
Age, y (median, mean ± SD)	15−90 (55, 54 ± 16)	44−93 (70, 71 ± 13)	22−89 (65, 62 ± 14)	<0.001	15−93 (64, 62 ± 15)
		Number(%)			
Sex					
Female	16 (48)	13 (59)	34 (44)		63 (48)
Male	17 (52)	9 (41)	43 (56)	0.56	69 (52)
Body site					
Head and neck	23 (70)	16 (73)	32 (42)		70 (54)
scalp	9 (27)	8 (36)	15 (19)		32 (25)
face	13 (39)	6 (27)	14 (18)		32 (25)
ear	0 (0)	2 (9)	2 (3)		4 (3)
neck	1 (3)	0 (0.0)	1 (1)		2 (2)
Trunk	5 (15)	5 (23)	19 (25)		29 (22)
Chest/Abdomen	1 (3)	3 (14)	13 (17)		17 (13)
back	4 (12)	2 (9)	6 (8)		12 (9)
axilla	0 (0)	0 (0.0)	6 (8)		4 (3)
Limbs	5 (15)	1 (5)	14 (18)		20 (16)
Upper limbs	5 (15)	1 (4)	3 (4)		11 (9)
Lower limbs	0 (0)	0 (0.0)	11 (14)		9 (7)
Genitalia/groin	0 (0)	0 (0.0)	6 (8)		6 (5)
Diameter, mm (median, mean ± SD)	2−45 (8, 12 ± 9) (n = 32)	3−50 (10, 15 ± 13) (n = 22)	2−240 (18, 30 ± 37) (n = 75)	0.006	2−240 (14, 23 ± 30)

Abbreviation: BCC, basal cell carcinoma.
[1]not include BCC.
[2]not include extramammary Paget disease.

YAP1-NUTM1 fusion was confirmed in porocarcinoma 1 (Figure 15a) whereas *YAP1* with non-*NUTM1* partner (presumably *MAML2*) fusion was confirmed in porocarcinoma 2 and 3. Porocarcinoma 1 showed few shared alterations between benign and malignant regions, showing early branching (Figure 14; Supplementary data 1). Benign regions P1 and P2 were also relatively distant, with multiple "private" alterations that were unique to specific regions, such as *ARID1A* in P2. On the other hand,

malignant sites shared a number of alterations showing a linear pattern. All malignant sites had a *PIK3CA* driver alteration in common. In porocarcinoma 2 and 3, more alterations in genes involved in epigenetic mechanisms, such as *ARID1A*, *EP300*, *PBRM1*, and *KMT2B* were observed. In addition, several copy number aberrations were observed, including amplification of genes related to cell proliferation, such as *CCND1*, *MYC*, and *FGF3*, 4, and 19, and loss of *CDKN2A* (Figure 16; Supplementary data

Table 4. Sensitivity and Specificity of Tumor Distinction by SOX9 Expression

Pathological diagnosis	Percentage Positivity			H-Score			Cutoff values (% Positivity, H-Score)
	P value	sensitivity, %	specificity, %	P value	sensitivity, %	specificity, %	
BCC (n = 26) vs SCC (n = 23)	<0.001	96	82	<0.001	85	87	80%, 150
BCC (n = 26) vs pilomatrixoma (n = 10)	<0.001	96	90	<0.001	85	90	80%, 150
BCC (n = 26) vs sebaceous carcinoma (n = 15)	0.017	96	73	0.002	84	73	80%, 150
Poroma (n = 20) vs SCC (n = 23)	<0.001	95	74	0.003	80	83	50%, 100
Porocarcinoma (n = 12) vs SCC (n = 23)	0.002	67	74	0.065	67	83	50%, 100
Benign sweat gland tumors (n = 11) vs apocrine adenocarcinoma (n = 12)	0.007	82	75	0.058	82	67	80%, 100
Malignant sweat gland tumors, miscellaneous (n = 4) vs apocrine adenocarcinoma (n = 12)	0.016	75	75	0.194	75	67	80%, 100

Abbreviations: BCC, basal cell carcinoma; SCC, squamous cell carcinoma.
*"Benign sweat gland tumors" include tubular apocrine adenoma, hidradenoma, cylindroma, mixed tumor, spiradenoma and syringoma.
**"Malignant sweat gland tumors, miscellaneous" include adenoid cystic carcinoma, cribriform carcinoma, and digital papillary carcinoma.
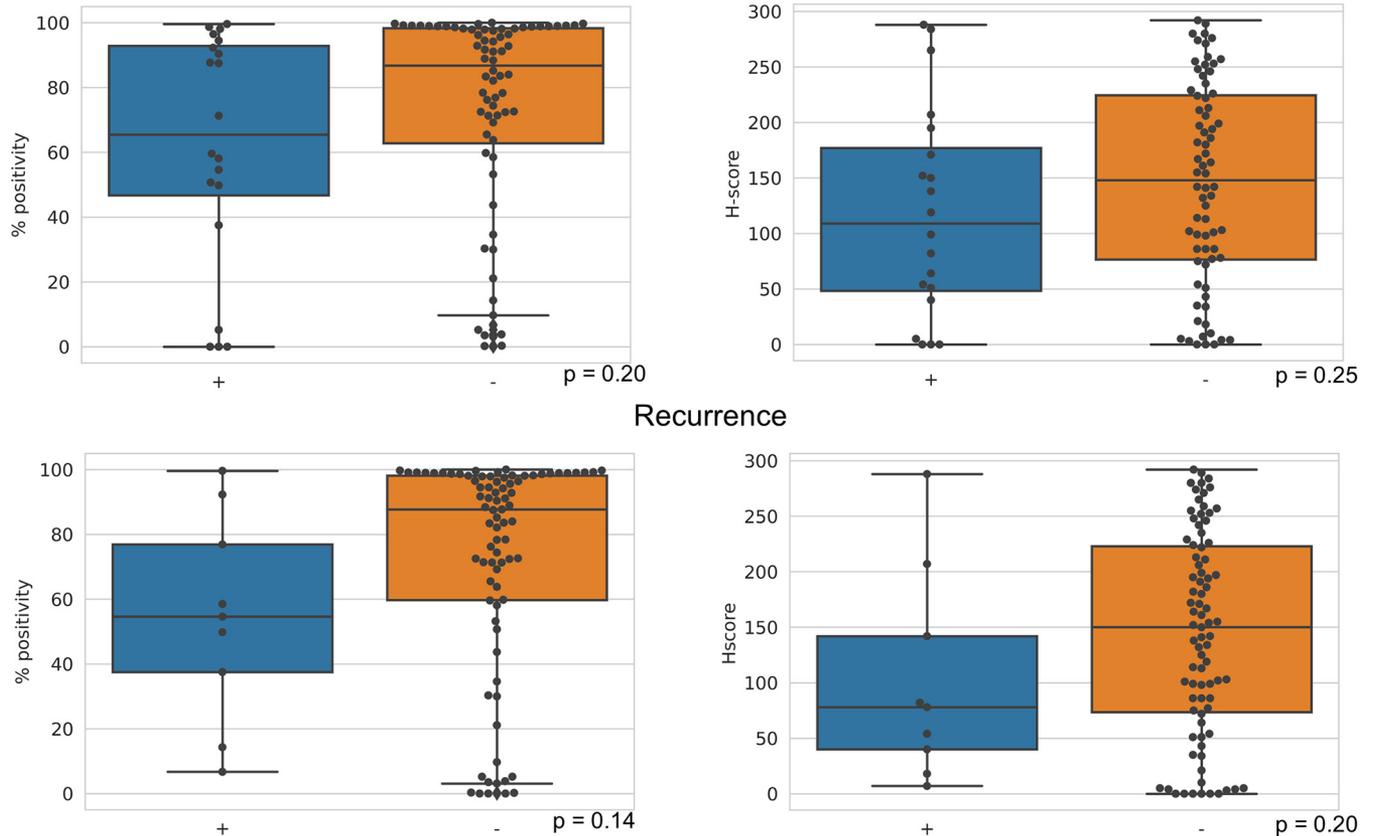
Metastasis

Recurrence

Figure 5. Malignant adnexal tumors with ("+") metastasis or recurrence tend to show lower SOX9 expression than those without ("−").

2−4). This may be due to differences in the original fusion gene status. Tumor mutational burden (TMB) was yet generally the same in all three tumors with an overall low TMB of 1∼1.5/Mb, except for P1 and P2 which were pathologically benign and even lower TMB (Figure 19a).

In only hidradenocarcinoma 1, fusion analysis using FISH showed *MAML2* rearrangement, which was previously found in both hidradenoma and hidradenocarcinoma (Kazakov et al., 2009; Kuma et al., 2017; Winnes et al., 2007) (Figure 15b). Hidradenocarcinoma 1 showed a branching pattern among benign (H1−3) and malignant (H4−8) sites

(Figure 14; Supplementary data 5). *FLCN* alteration was found in H1. *TP53* and *CDKN2A* alterations were shared in all the malignant regions, but H4 and H8 were genetically distant and showed a different branching evolution. In particular, H4 had a number of private alterations as well as private copy number aberrations, such as *ERBB2*, and *MAP3K3* amplifications (Supplementary data 6). *CCND1*, *FGF3*, *FGF4*, *FGF19*, and *YAP1* amplification were shared among regions H5−H8, whereas *AURKA*, *JUN*, and *ZNF217* amplifications were only detected in region H8. FISH analysis exhibited *MAML2* split signals with occasional

Table 5. Comparison of Clinical Characteristics between Patients with and without metastasis or recurrence.

Clinicopathological factors	Metastasis			Recurrence		
	+(n = 20)	−(n = 72)		+(n = 9)	−(n = 83)	
Age, y (median, mean ± SD)	69 (66 ± 16)	70 (65 ± 16)	$P = 0.85$	66 (66 ± 12)	70 (65 ± 16)	$P = 0.87$
Gender						
Female	7	38		6	39	
Male	13	34	$P = 0.25$	3	44	$P = 0.44$
Diameter, mm	40 (55 ± 53)	15 (25 ± 30)	$P = 0.001$	25 (54 ± 75)	18 (28 ± 31)	$P = 0.05$
SOX9 % positivity, mm	65 (62 ± 36)	87 (73 ± 33)		55 (54 ± 32)	87 (72 ± 34)	
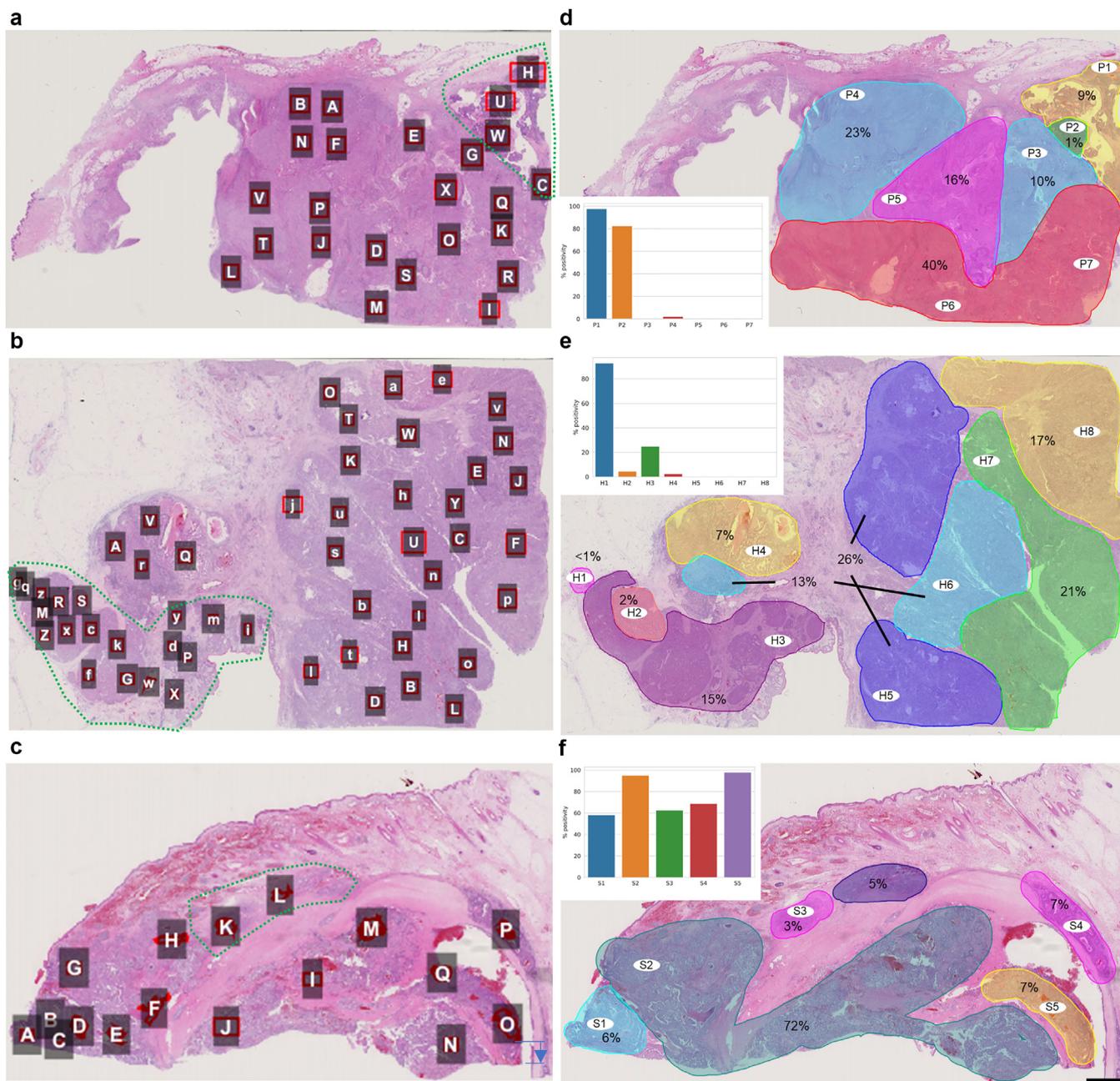
Figure 6. Region of interests (ROIs) for morphological single-cell clustering and sample sites for M-WES in sweat gland carcinomas with pathologically benign and malignant regions. **(a)** Porocarcinoma 1, **(b)** hidradenocarcinoma 1, **(c)** spiradenocarcinoma 1. Green dotted line shows pathologically benign areas in each tumor. **(d–f)** Sample sites for M-WES in each tumor. Colors indicate morphologically similar groups based on machine-learning described in the text and % shows the area percentage of each group. Sampling areas were approximately from 0.5 mm² to 1 mm². Inset: Sox9 % positivity in each sampled site. Bar = 2 mm.

duplication in H4, but split signals were not confirmed in H6 (Figure 15b). *TP53* alterations were found in both hidradenocarcinoma 2 and hidradenocarcinoma 3, whereas frameshift alterations in *BAP1* were found only in hidradenocarcinoma 2. *RAS* alterations were found in these two tumors, namely *HRAS* somatic alteration and *KRAS* somatic alteration and amplification, as were *CDH1* alterations (Figure 17; Supplementary data 2, 7, and 8). TMB of histologically benign area of hidradenocarcinoma 1 was as low as

0.5/Mb whereas in other malignant regions the TMB range was 2.5∼4/Mb depending on the sample site (Figure 19b).

In spiradenocarcinoma 1, sampling sites showed few shared alterations, except regions S2 and S5, thus showing the most prominent branching pattern (Figure 14; Supplementary data 9). Interestingly, the recently reported *ALPK1* hotspot alteration (p. Val1092Ala) was only detected in benign region S3. Another recently reported *BCOR* alteration in spiradenoma and cylindroma (Davies et al., 2019)
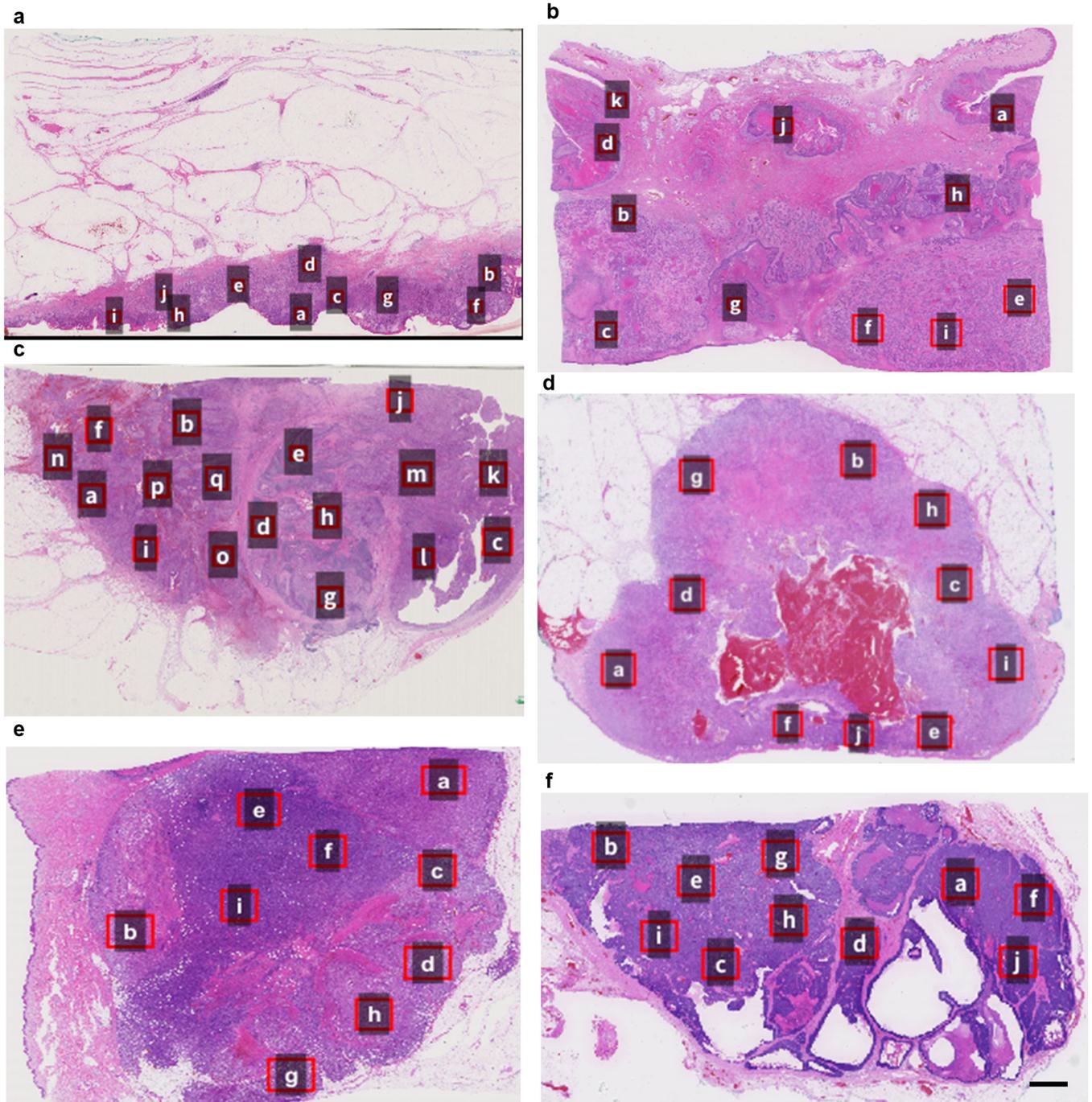
Figure 7. Region of interests (ROIs) for morphological single-cell clustering in sweat gland carcinomas with pathologically malignant only regions. (a) porocarcinoma 2, **(b)** porocarcinoma 3, **(c)** hidradenocarcinoma 2, **(d)** hidradenocarcinoma 3, **(e)** spiradenocarcinoma 2, **(f)** spiradenocarcinoma 3. Bar = 2 mm.

was found in region S1 and histologically tumor cells showed relatively spindle shape. Region S1, S2, and S5 shared *PTEN* loss/deletion. Region S4 harbored *FLT1*, *FLT3*, and *CDK8* amplification, whereas regions S2 and S5 harbored *AURKA*, *MYC*, and *DDR1* amplification (Supplementary data 10). In Spiradenocarcinoma2, *ALPK1* alterations were common in all three regions, *TERT* was amplified in two regions, and *MYC* was amplified in only region S2−2. *APOBEC3* copy number loss was observed in S2−2 and S2−3 (Figure 18; Supplementary data 2, 11, and 12). This is of interest

considering that APOBEC signature (signature 2) was observed in spiradenocarcinoma 1 as detailed below. No presumable driver alterations were found in spiradenocarcinoma 3. The TMB was highest (4/Mb) in the region S1 and lowest in the benign region S3 spiradenocarcinoma 1 (0.5/Mb) among three tumors. Generally, there was more regional variation of TMB in three spiradenocarcinomas (Figure 19c).

Mutational signature analysis (COSMIC v2) was performed for selected sites with >70 somatic alterations and revealed
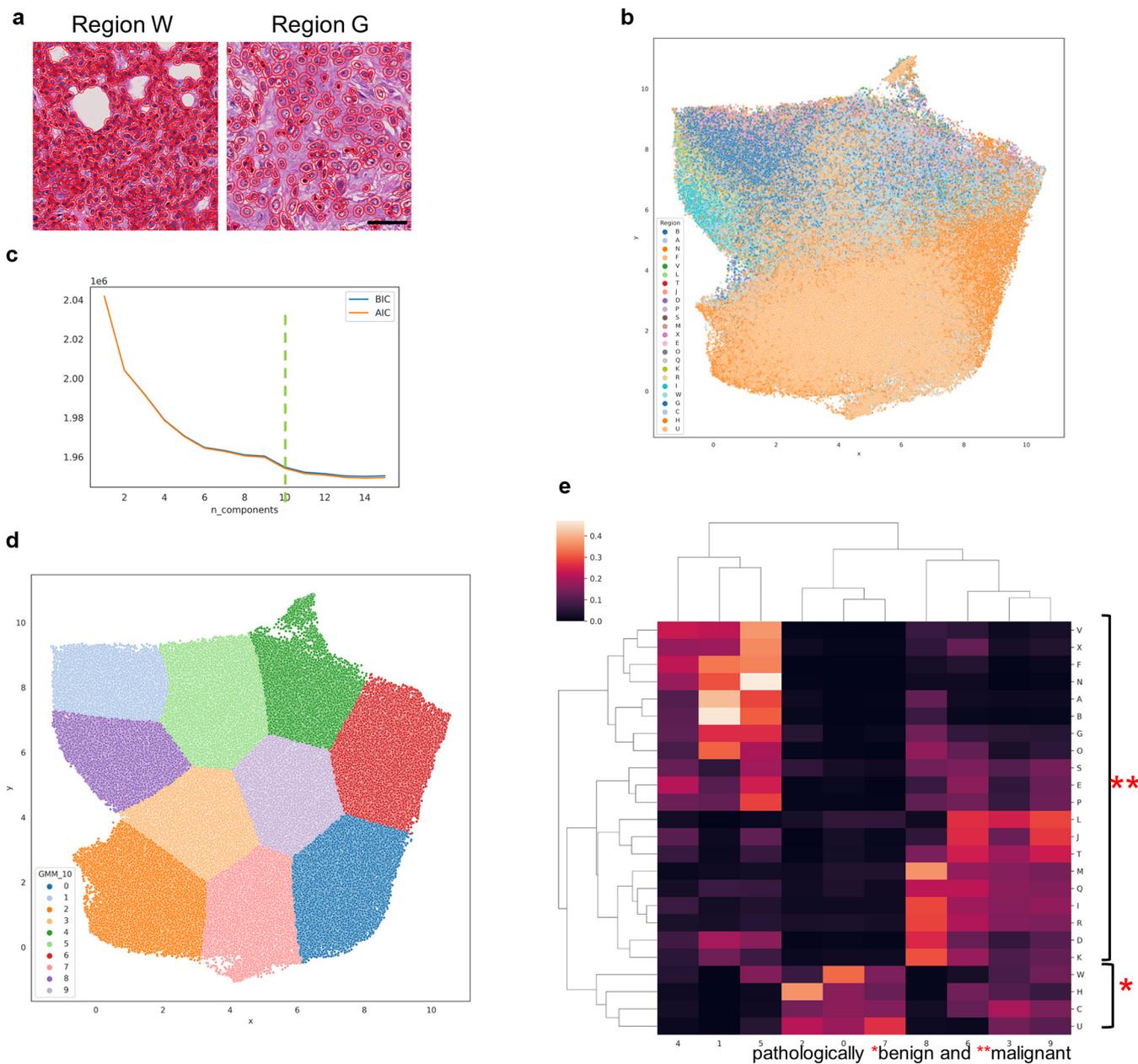
Figure 8. Morphological single-cell unsupervised clustering helps revealing intratumoral heterogeneity. (a) Example images of nuclear segmentation in region W and G of porocarcinoma 1. **(b)** Uniform Manifold Approximation and Projection (UMAP) embedding of all cells in ROIs of the porocarcinoma 1. **(c)** Akaike information criterion (AIC) and Bayesian information criterion (BIC) plots to determine appropriate number of clusters for gaussian mixture model (GMM) clustering. Green dotted line indicates the optimal cluster number chosen. **(d)** UMAP clustering by GMM. **(e)** Hierarchical clustering of all ROIs. As shown in heatmap, each ROI has different proportion of cell clusters.

an APOBEC signature (signature 2 and 13) (Alexandrov et al., 2020) in spiradenocarcinoma 1 (Figure 20a and b). Intriguingly, region H4 of hidradenocarcinoma 1 also showed the APOBEC signature and clustered into the same group as spiradenocarcinoma 1. All three hidradenocarcinomas harbored age-related signature 1.

DISCUSSION

This study highlighted SOX9 as an important marker for understanding the histogenesis and diagnosis of skin adnexal tumors. SOX family transcription factors regulate cell fate

during development (Kamachi and Kondoh, 2013; Sarkar and Hochedlinger, 2013) and several SOX proteins are currently used in cancer diagnosis to specify the cell type of origin. For example, SOX10 is crucial for the maintenance of Schwann cells and melanocytes derived from the neural crest, hence it is widely used as a marker for melanocytic and schwannian tumor diagnosis (Nonaka et al., 2008). SOX10 is also expressed in acini and the intercalated duct of the major salivary gland from the developmental stage and is used to diagnose salivary gland tumors (Ohtomo et al., 2013). This study showed that SOX9 could greatly improve the diagnosis
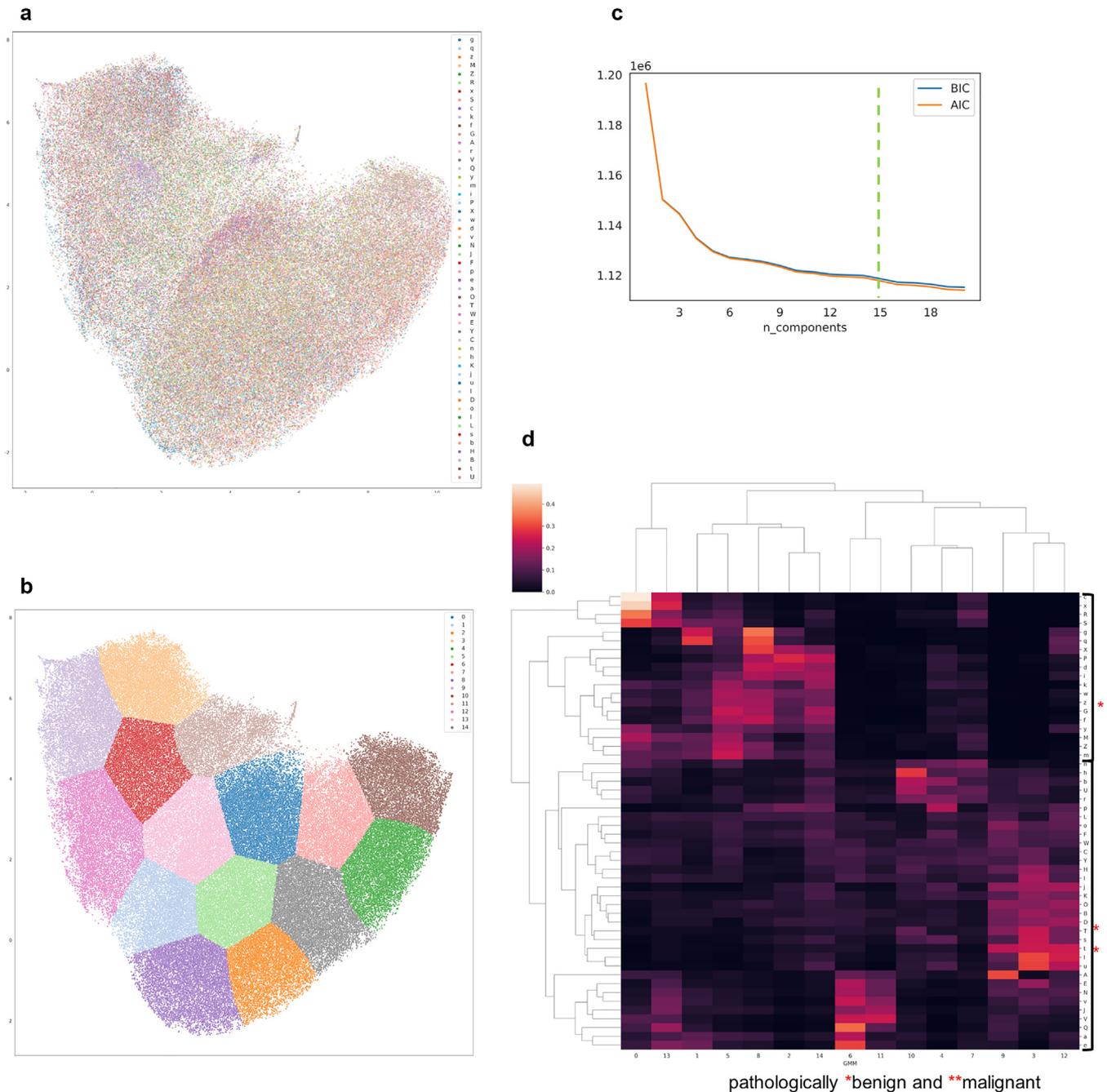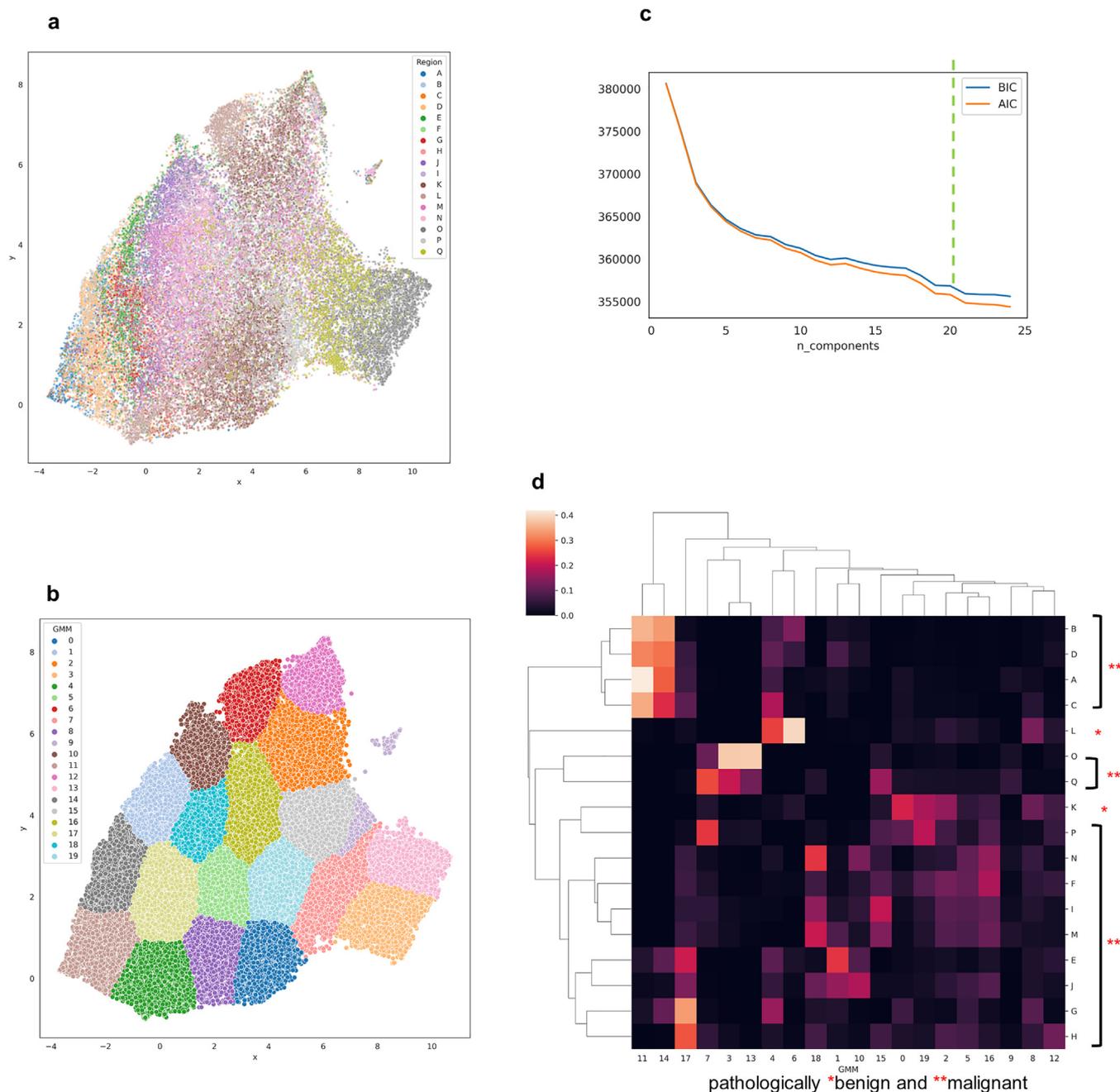
Figure 9. Morphological clustering of the hidradenocarcinoma 1. (a) UMAP embedding of all cells in ROIs. **(b)** UMAP clustering by GMM (Upper right and bottom right). **(c)** AIC and BIC plots. Green dotted line indicates the optimal cluster number chosen. **(d)** Hierarchical clustering of all ROIs. AIC, Akaike information criterion; BIC, Bayesian information criterion; GMM, gaussian mixture model.

of adnexal tumors to differentiate between SCC and BCC or adnexal tumors. The result that SOX9 expression was significantly weaker in apocrine glands were controversial given that they are believed to share the same stem cells with the pilosebaceous units, which generally showed stronger SOX9 expression. The precise origin of apocrine glands is still poorly understood, however, perhaps because mouse basically lack apocrine glands so mouse models could not be obtained (Lu and Fuchs, 2014). Single-cell RNA sequencing

may help to reveal presumed unique expression profile of human apocrine glands. Although we did not focus on the biology of cutaneous stem cell itself in this study, it could be predicted that SOX9 expression is strong around bulge, which belong to isthmus of the outer root sheath from previous studies. However, these studies are predominantly based on mouse developmental models, so the distribution of SOX9 expression in human adult skin could be more or less different. We specified bulge and hair germ area with Ber-EP4

Figure 10. Morphological clustering of the spiradenocarcinoma 1. (a) UMAP embedding of all cells in ROIs. (b) UMAP clustering by GMM (Upper right and bottom right). (c) AIC and BIC plots. Green dotted line indicates the optimal cluster number chosen. (d) Hierarchical clustering of all ROIs. AIC, Akaike information criterion; BIC, Bayesian information criterion; GMM, gaussian mixture model.

IHC (Figure 30) and confirmed that SOX9 expression was strong from bulge to hair germ and weaker in the upper part. Previous study by Purba et al. reported that SOX9 expression was most prominent in the outer-root sheath below the bulge (Purba et al., 2015). Given that they seemed to define bulge area rather widely, their result would be consistent with our result, but further study would be necessary.

SOX proteins are important regulators of development and differentiation and their dysregulation has been implicated in cancer development (Grimm et al., 2020; Sarkar and Hochedlinger, 2013). Aberrant SOX9 expression was observed at very high or low levels in some adnexal

carcinomas in our cohort. This presumed SOX9 dysregulation was observed in three of the sweat gland carcinomas in which M-WES was performed in this study. No genetic changes were seen for *SOX9* in all 9 carcinomas, however, thus loss of SOX9 expression in malignant regions in porocarcinoma 1 and hidradenocarcinoma 1 may stem from epigenetic changes, such as hypermethylation (Aleman et al., 2008; Cheng et al., 2015).

Multiregion sequencing revealed intratumoral heterogeneity and evolutionary history of both primary and metastatic malignancies including lung (Jamal-Hanjani et al., 2017), renal (Gerlinger et al., 2012, 2014), esophageal (Hao et al.,
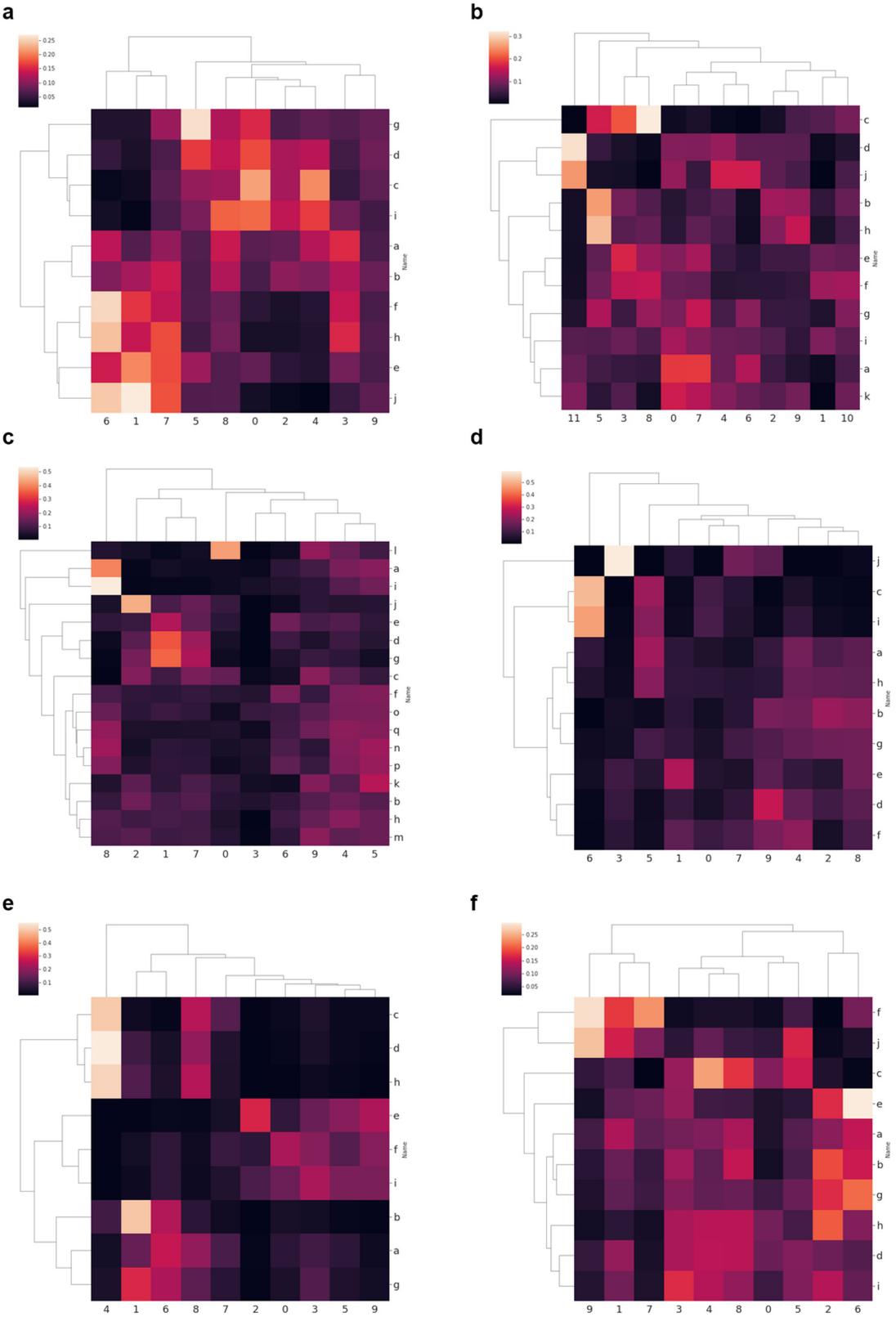
Figure 11. Hierarchical clustering of all ROIs in procarcinoma 2−3, hidradenocarcinoma 2−3, spiradenocarcinoma 2−3. (a) Porocarcinoma 2, **(b)** porocarcinoma 3, **(c)** hidradenocarcinoma 2, **(d)** hidradenocarcinoma 3, **(e)** spiradenocarcinoma 2, **(f)** spiradenocarcinoma 3.
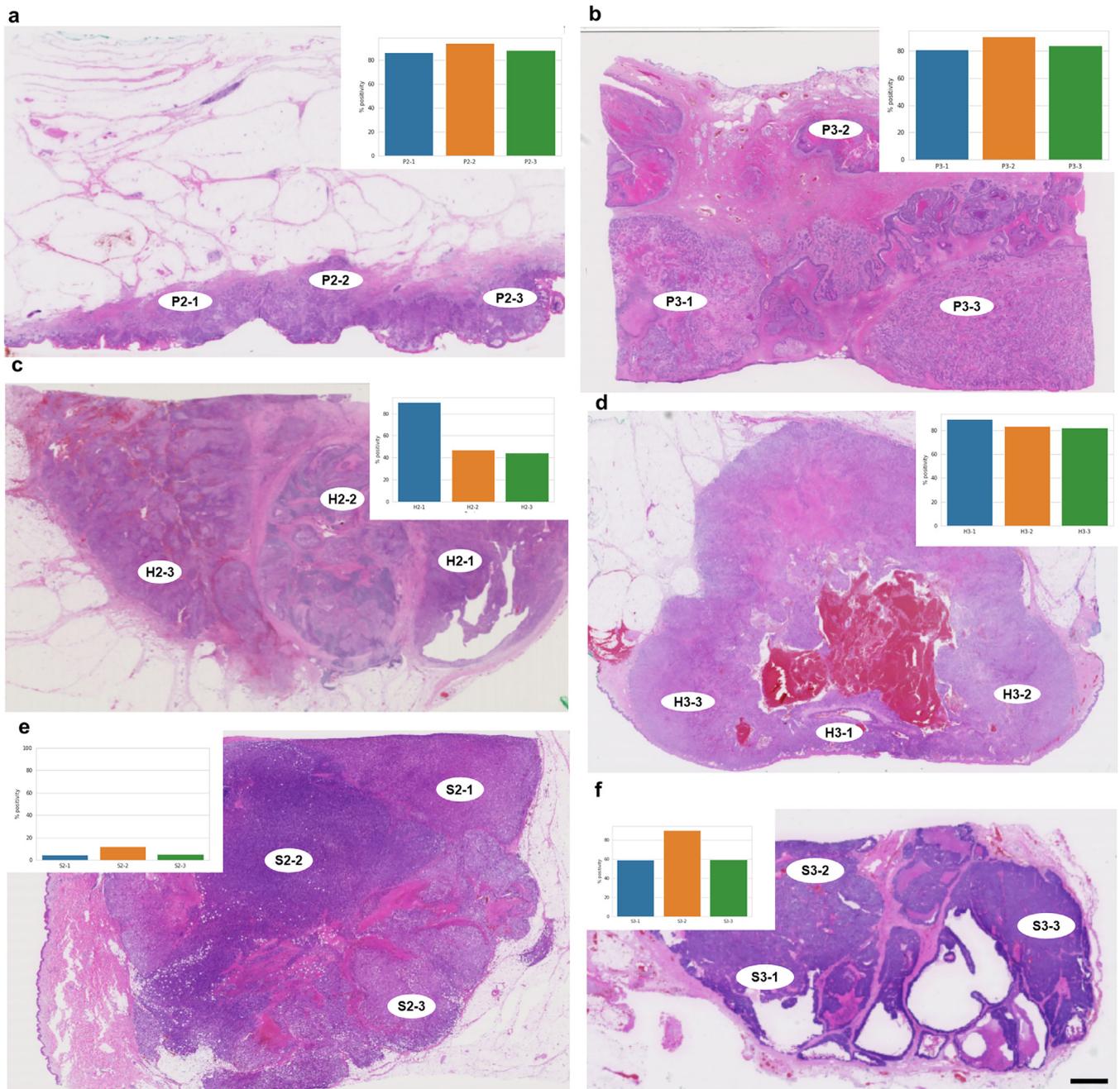
Figure 12. Sampled regions for M-WES in porocarcinoma 2–3, hidradenocarcinoma 2–3, and spiradenocarcinoma 2–3. Inset: Sox9 % positivity in each sampled region. **(a)** Porocarcinoma 2, **(b)** porocarcinoma 3, **(c)** hidradenocarcinoma 2, **(d)** hidradenocarcinoma 3, **(e)** spiradenocarcinoma 2, **(f)** spiradenocarcinoma 3. Bar = 2 mm.

2016), and breast cancers (Ullah et al., 2018). The findings from these studies provide various implications for clinical management. This study revealed significant genetic heterogeneity that may have been of multiclonal origin. One important finding from our M-WES study is that some of the driver alterations were often limited to one or two regions. Several previous histological studies have reported discordant driver mutational status among spatially adjacent but morphologically distinct regions in colon and bone cancer (Sekine et al., 2014; Yoshida et al., 2019). In fusion-driven

tumors, a heterogenous fusion state was reported in a case of *ETV6–NTRK2* acute myeloid leukemia, in which the *TRK* fusion-negative AML clone was present at baseline (Taylor et al., 2018). Similarly, the observations from our phylogenetic study indicate that some of adnexal carcinomas may share an origin but branched during the very early stages and hence share few genetic alterations. The clinical history of porocarcinoma 1, hidradenocarcinoma 1, and spiradenocarcinoma 1 showed that it was present for several years to decades and grew rapidly in recent years; therefore,
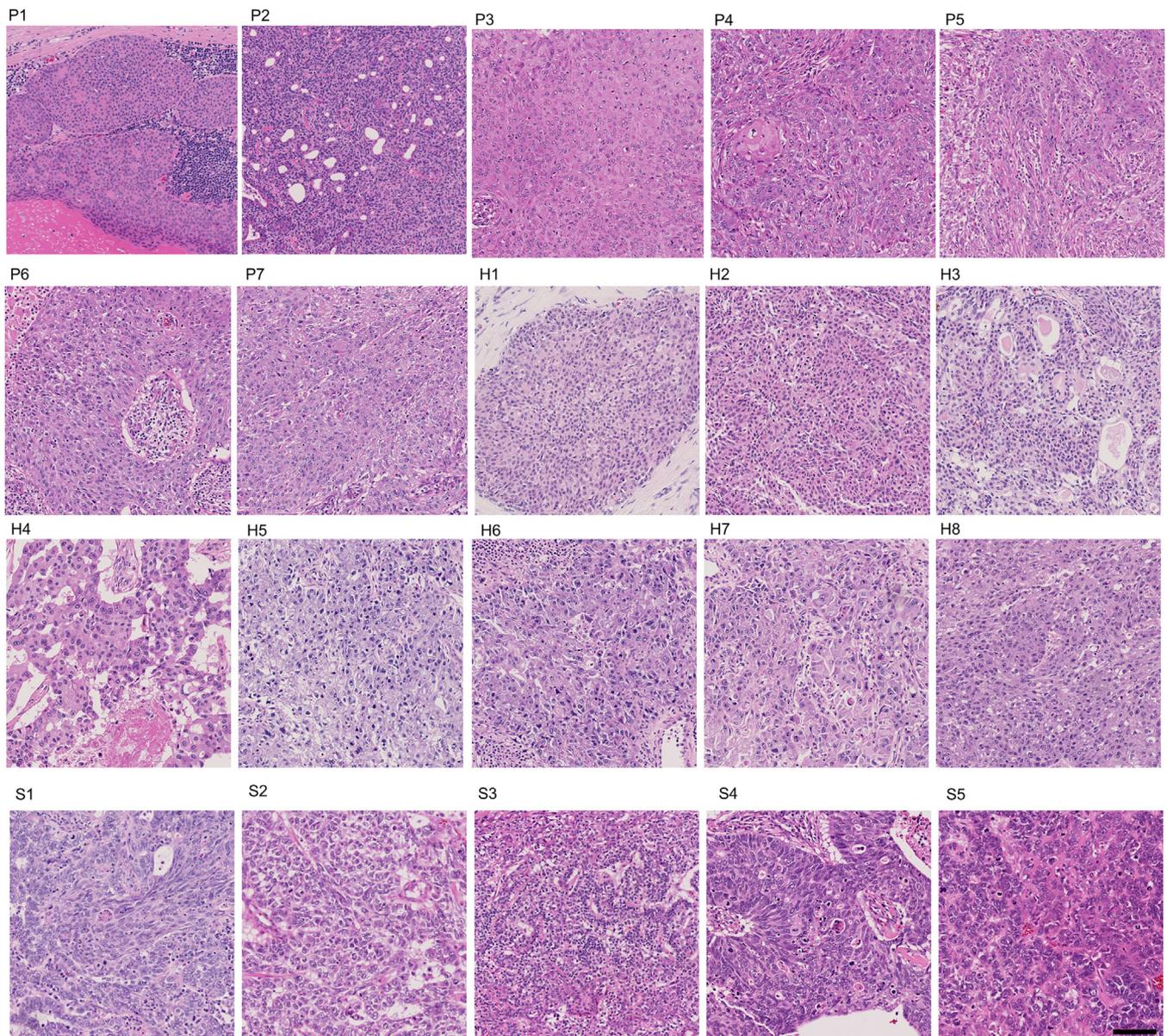
Figure 13. Representative H&E images of sampling regions for M-WES. P: porocarcinoma 1, H: hidradenocarcinoma 1, S: spiradenocarcinoma 1. Bar = 100μm.

some tumor cells may have been dormant in their early stage (Naxerova and Jain, 2015).

NUTM1-rearranged tumors usually lack additional driver alterations (Lee et al., 2017; McEvoy et al., 2020), and the rearrangement status may explain only two additional presumable driver alterations (*PIK3CA* and *ARID1A*) in porocarcinoma 1. FISH splits were indistinct in malignant regions of porocarcinoma 1 and hidradenocarcinoma 1 compared with benign regions, and IHC in some regions of porocarcinoma 1 showed weakened NUT and altered staining patterns of YAP1 (Figure 15a and b). Therefore, heterogeneity in the status of the fusion gene was possible, but this is only speculation at this time and may contradict with the mutational status of porocarcinoma 1. Another important implication is that almost all malignant regions of each tumor shared several clinically targetable driver alterations, such as

alpelisib for the canonical *PIK3CA* alteration in porocarcinoma (André et al., 2019, 2021; Jin et al., 2021). YAP/TAZ−TEAD inhibitors or G-protein coupled receptors inhibitor targeting upstream regulators of the Hippo pathway could be effective for *YAP1−NUTM1* fusion in porocarcinoma and *YAP1* amplification in hidradenocarcinoma (Cunningham and Hansen, 2022).

To use whole-exome sequencing (WES) for extracting mutational signatures clearly has limitations; compared with whole-genome sequencing, WES reveals far less alterations and thus the power to reliably discern signatures is reduced (Koh et al., 2021). However, Alexandrov et al. previously reported that WES can be useful for extracting at least some of the signatures, based on the simulated data and 100 breast cancer exomes containing ∼7,000 somatic alterations (Alexandrov et al., 2013; Stephens et al., 2012). We thus
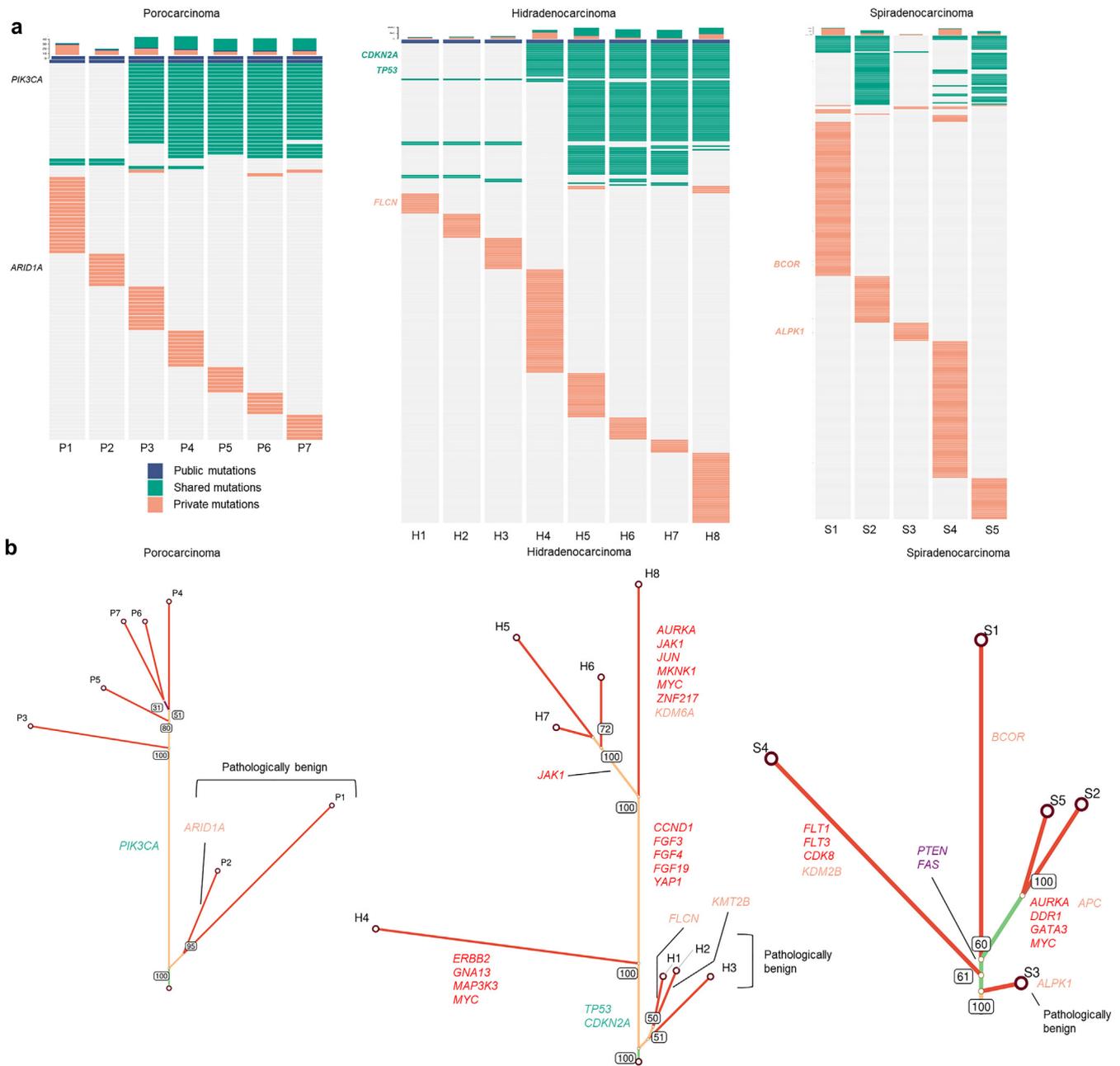
Figure 14. M-WES reveals distinct genomic profiles of sweat gland carcinomas. (a) Somatic alterations in each ROIs of porocarcinoma 1, hidradenocarcinoma 1, and spiradenocarcinoma 1. Each gene is arranged in a row, and cancer-related genes with putative driver alterations are indicated. "Public" alteration is shared by all ROIs, "Shared" is shared by not all but some ROIs, and "Private" is unique to each ROI. **(b)** Phylogenetic trees are constructed from all somatic alterations by the neighbor-joining. CNAs are also indicated (red: gain, purple: loss). CNAs, copy number abberations.

arbitrarily set the threshold of somatic substitutions to 70. We believe that extracted mutational signatures, such as APO-BEC signatures would be robust, whereas signatures with less proportions would be uncertain. In addition, number of substitutions in three porocarcinoma sites were all under threshold 70. Whole-genome study would be necessary for studying poroid tumors.

Deep learning has great potential in the field of histopathology for the optimal treatment of patients (Huss and Coupland, 2020; van der Laak et al., 2021). In contrast,

unsupervised clustering algorithms, such as hierarchical clustering, and recent dimensionality reduction methods, such as uniform manifold approximation and projection and t-SNE, have been extensively used mainly in medical and biology research and might be expected not to be as versatile as deep learning. However, we made the most of unsupervised algorithms and attempted to decipher biological characteristics of cancer. This approach using cellular and nuclear information extracted from digital images for machine learning was expected to be rather straightforward as
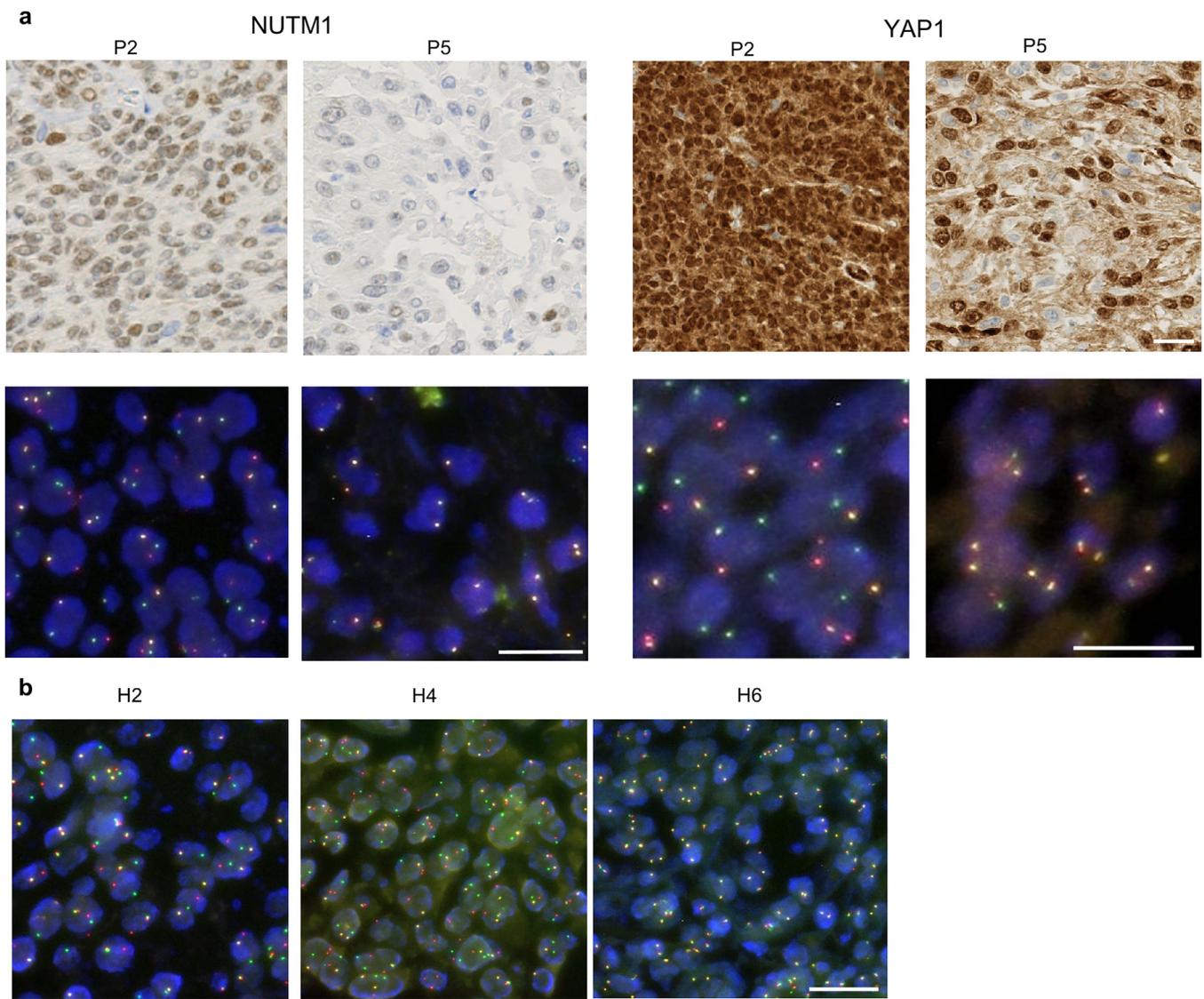
Figure 15. Gene rearrangements of porocarcinoma 1 and hidradenocarcinoma 1. (a) NUTM1 and YAP1 IHC (upper row) and FISH (bottom row) analysis in porocarcinoma 1 region P2 and P5. (b) *MAML2* FISH analysis in hidradenocarcinoma 1 in H2, H4 and H6. Bar = 20 μm.

previously described (Grys et al., 2017; Uhler and Shivashankar, 2018). In fact, in our workflow, we were able to clearly distinguish benign and malignant areas solely from morphological data in tumors with benign and malignant areas, such as porocarcinoma 1 and hidradenocarcinoma 1. We were also able to objectively classify malignant ROIs in the nine tumors. This is of great advantage for objectively examining morphological differences which tend to be necessarily subjective. It should be noted, however, that unsupervised learning is not completely objective. Although Akaike information criterion or Bayesian information criterion was used as the criterion for determining the number of clusters in morphological clustering by Gaussian-mixture model, the final number of clusters was determined somewhat subjectively. For example, the optimal number of clusters for porocarcinoma 1 was determined to be 10; decrease in Akaike information criterion or Bayesian

information criterion was limited after the number of components 6, thus some may consider 6 to be more optimal. However, the final hierarchical clustering results were almost identical when the number of clusters is 6 (Figure 31). We believe this suggests the robustness of the present method.

Assuming that the cancer cell morphology could reflect biological variation, such as genomic and/or epigenetic variations, formalin-fixed paraffin-embedded (FFPE) whole-slide image-based analysis of intratumoral heterogeneity described in this study could be a powerful tool for the diagnosis and treatment of tumors. Subclones with common and private alterations could be estimated quantitatively by measuring the areas or cell numbers; therefore, targeting common alteration and subpopulations with the largest proportions may be effective. In our case of spiradenocarcinoma 1, *PTEN* loss was a common alteration in most subpopulations. The estimated subpopulation with *BCOR* alteration was small
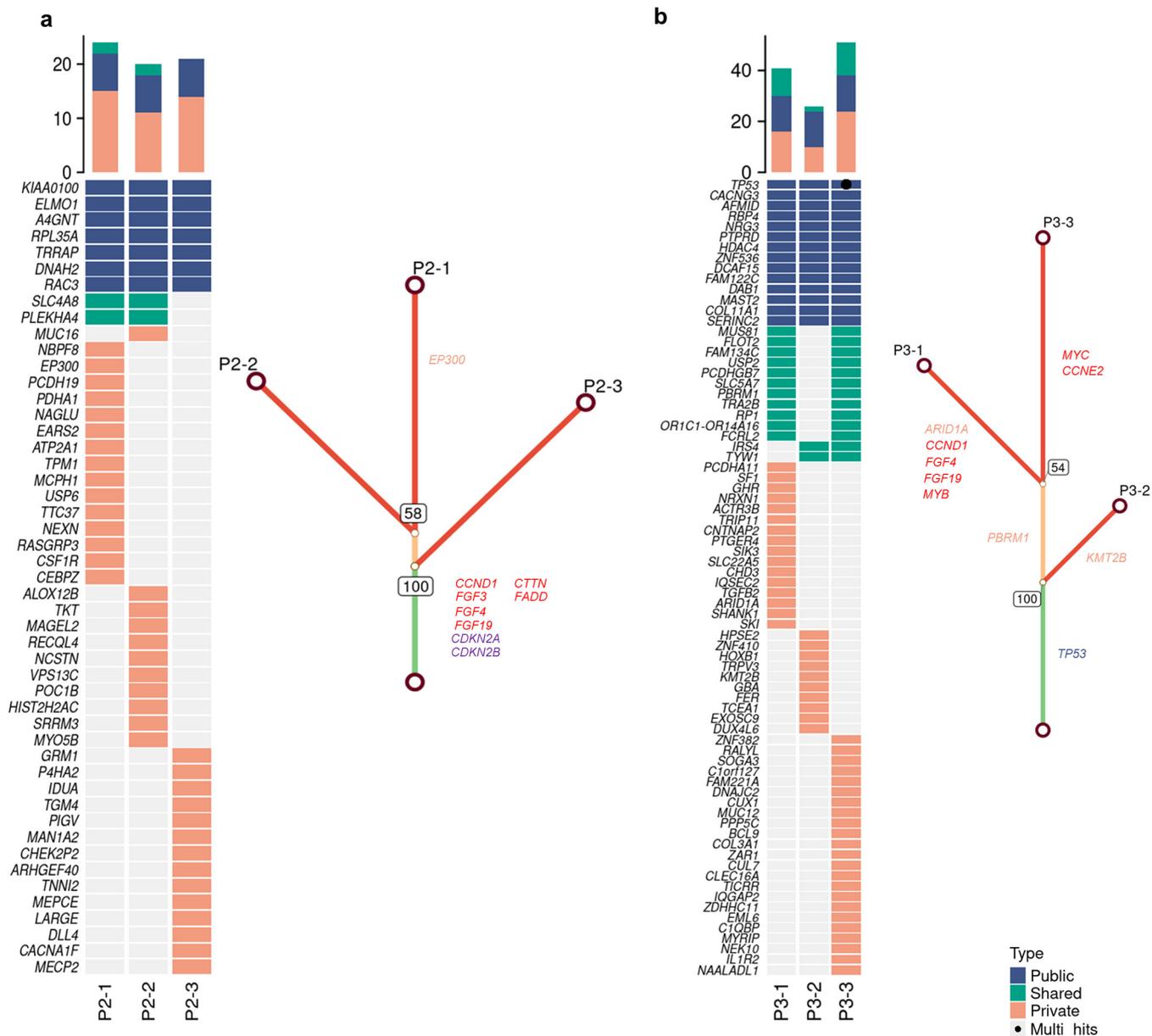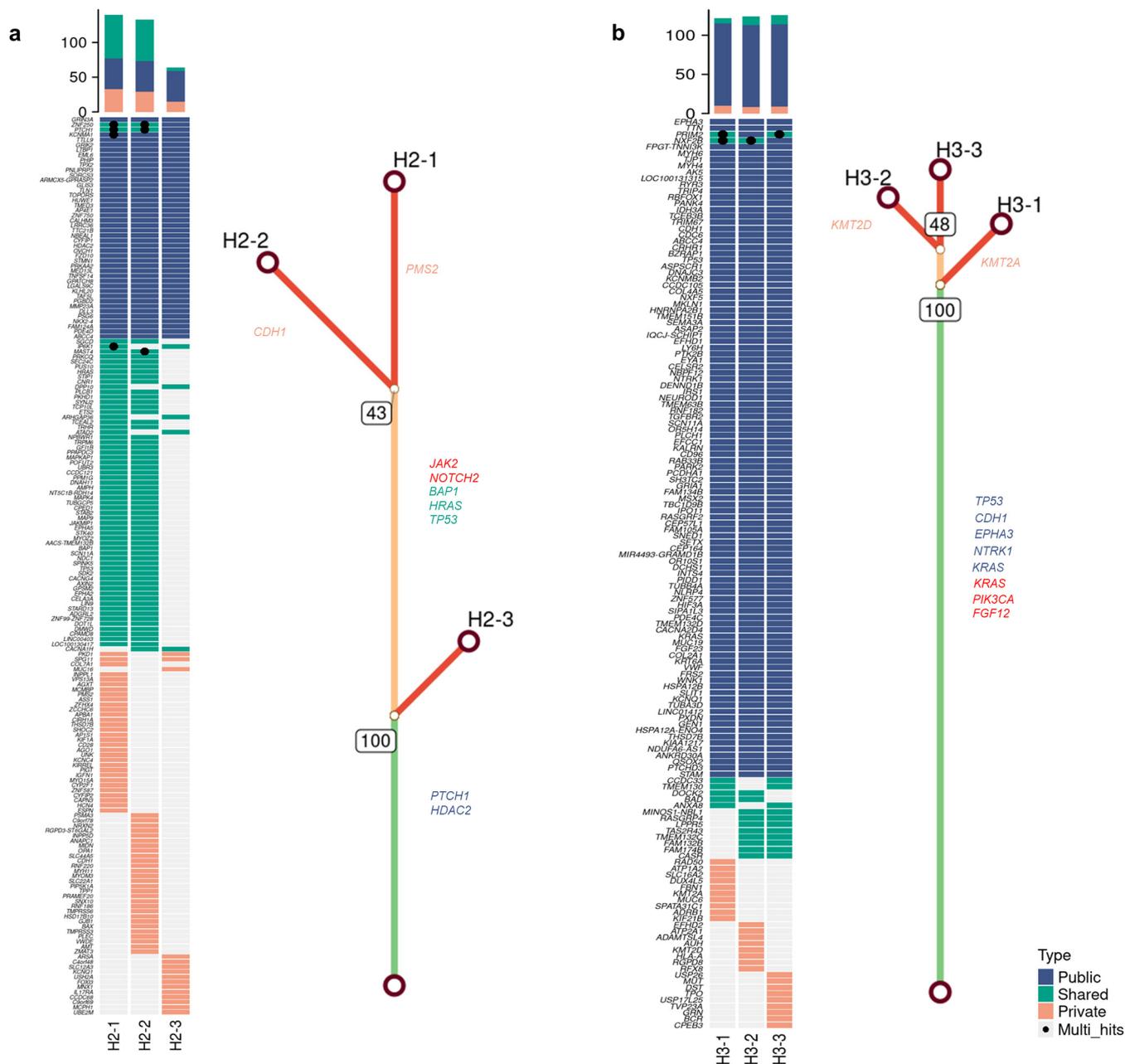
Figure 16. Somatic alterations in each ROIs and phylogenetic trees of porocarcinoma 2 and 3. Phylogenetic trees were constructed from all somatic alterations by the neighbor-joining of (a) porocarcinoma 2 and (b) porocarcinoma 3. CNAs are also indicated (red: gain, purple: loss). CNAs, copy number abberations.

(6%), whereas the estimated subpopulation with gene amplifications, such as *AURKA, DDR1*, and *MYC* was largest (72%) (Figure 3f). A straightforward approach would be to target *PTEN* and the largest subpopulations, though smaller subpopulations could be potentially more malignant. Analysis combined with metastatic lesions could bring further insight to evaluate potential malignancy and proportions. Some types of tumors, such as melanoma or lung adenocarcinoma, harbor genetic homogeneity and show a linear progression pattern (McGranahan and Swanton, 2017). The present case of porocarcinoma 1 and hidradenocarcinoma 3 showed strong genetic homogeneity. The morphological approach described in this study may not be effective for these tumors, but epigenetic changes may affect morphological variations.

In conclusion, this study revealed heterogeneity of skin adnexal tumors using "genocomputed pathology" analysis combining SOX9 expression, machine-learning, and multi-region exome analysis. There may be more heterogeneity than previously assumed within the tumors as well as the coexistence of other tumors lacking the driver fusion gene. The presence or absence of potential driver genes was also identified in each region, highlighting therapeutic strategies for the clinical management of cutaneous adnexal carcinoma encompassing heterogeneous tumors. At present, the rarity of adnexal carcinomas and their convoluted evolution is challenging for clinical management. However, comprehensive studies could provide further insight into skin adnexal tumors for rational therapeutic strategies and ultimately improve outcomes in patient.
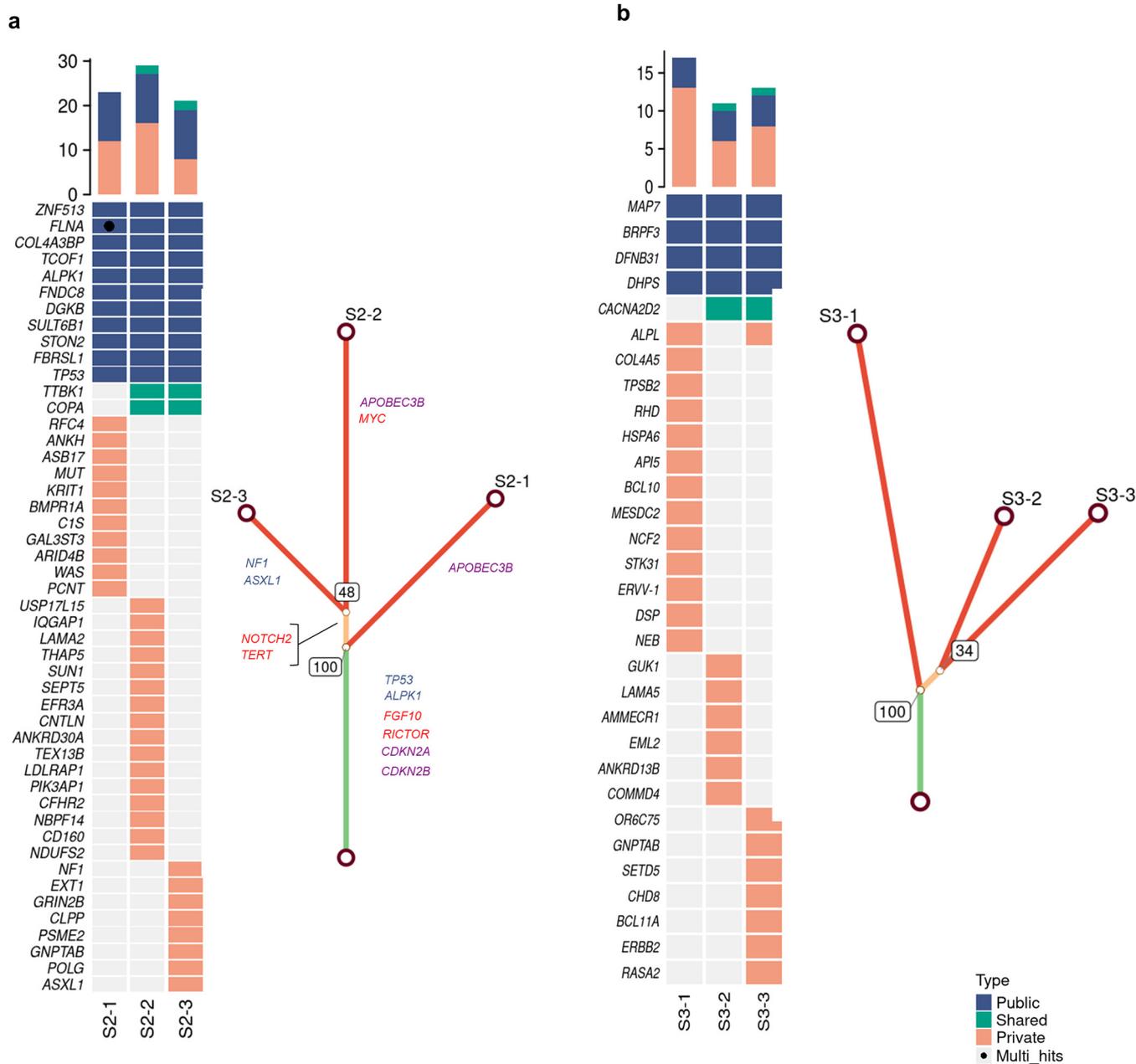
Figure 17. Somatic alterations in each ROIs and phylogenetic trees of hidradenocarcinoma 2 and 3. Phylogenetic trees were constructed from all somatic alterations by the neighbor-joining of **(a)** hidradenocarcinoma 2 and **(b)** hidradenocarcinoma 3. CNAs are also indicated (red: gain, purple: loss). CNAs, copy number abberations.

MATERIALS AND METHODS

Study approval

All of the patients provided written informed consent, and the study was approved by the Institutional Review Boards of the National Cancer Center Hospital, Japan (approval no. 2017-078)

Tissue samples

All cancers were diagnosed at the National Cancer Center Hospital, Tokyo, Japan between June, 2007 and September, 2020.

This study analyzed 185 FFPE specimens of skin tumors, including 33 HF tumors, 77 sweat gland tumors (including three cases of lymph node metastasis), 22 sebaceous tumors, 26 BCCs, 23 SCCs, and 4 extramammary Paget disease. Tumors were reviewed by two

pathologists (YN and TM) and classified according to the World Health Organization classification (Elder et al., 2018).

IHC

IHC was performed on whole-tissue sections after antigen retrieval with Target Retrieval solution (pH 6.1; Dako, Carpinteria, CA) in a pressure cooker. IHC staining was performed on 4-μm−thick FFPE specimens following antigen retrieval with Target Retrieval Solution pH 9 (Agilent Technologies, Palo Alto, CA). Anti-SOX9 antibody (1:100 dilution; clone H-90; Santa Cruz Biotechnology Inc., Santa Cruz, CA), anti-NUT antibody (1:100 dilution; clone C52B1; Cell Signaling Technology, Beverly, MA), anti-N−terminal region of YAP1 antibody (dilution 1:200; clone 2F12; Abnova, Taipei, Taiwan) and anti-C−terminal region of YAP1 antibody (dilution 1:200; clone

Figure 18. Somatic alterations in each ROIs and phylogenetic trees of spiradenocarcinoma 2 and 3. Phylogenetic trees were constructed from all somatic alterations by the neighbor-joining of **(a)** spiradenocarcinoma 2 and **(b)** spiradenocarcinoma 3. CNAs are also indicated (red: gain, purple: loss). CNAs, copy number abberations.

D8H1X-XP; Cell Signaling Technology, Beverly, MA) were used. As to anti-N−terminal YAP1 antibody, diffuse nuclear expression of the N-terminal portion was considered significant whereas the absence of C-terminal portion expression was considered significant for anti-C−terminal antibody (Hartsough et al., 2023; Russell-Goldman et al., 2021; Sekine et al., 2019). Epithelial antigen (clone Ber-EP4; Dako Omnis) were used to specify hair germ.

Digital image analysis and IHC evaluation
All H&E stained and IHC slides were digitized using the NanoZoomer 2.0-HT (Hamamatsu Photonics K.K., Hamamatsu, Japan) platform at ×20 magnification. QuPath open-source software platform (version 0.2.3) was used for analysis (Bankhead et al., 2017). First, a "project" was created for H&E and IHC slides of nontumor cutaneous

appendages and adnexal tumors. After setting image types ("H&E" or "H-DAB" for IHC), stain vectors were estimated on representative image for better nucleus segmentation and 3,3′-diaminobenzidine; DAB intensity evaluation. The same stain vectors were used across all images in each project with the script recorded in the "command history." H-scores and ROIs were annotated using annotation tools to calculate SOX9 percentage positivity. For tumors, ROIs were selected to include sufficient tumor cells (>500) and exclude nontumoral cells, such as immune cells. Multiple ROIs were selected when apparent heterogeneity of percentage positivity was present within one tumor. For nontumor cutaneous adnexa, ROIs were selected for each separate structure in nonlesional skin from at least five patients.

Precise detection of nuclei in H&E and IHC images was achieved using the deep learning−based method of nucleus detection,

a

b

c

Figure 19. Tumor mutational burden of each region derived from M-WES. **(a)** Porocarcinoma 1−3, **(b)** hidradenocarcinoma 1−3, and **(c)** spiradenocarcinoma 1−3.
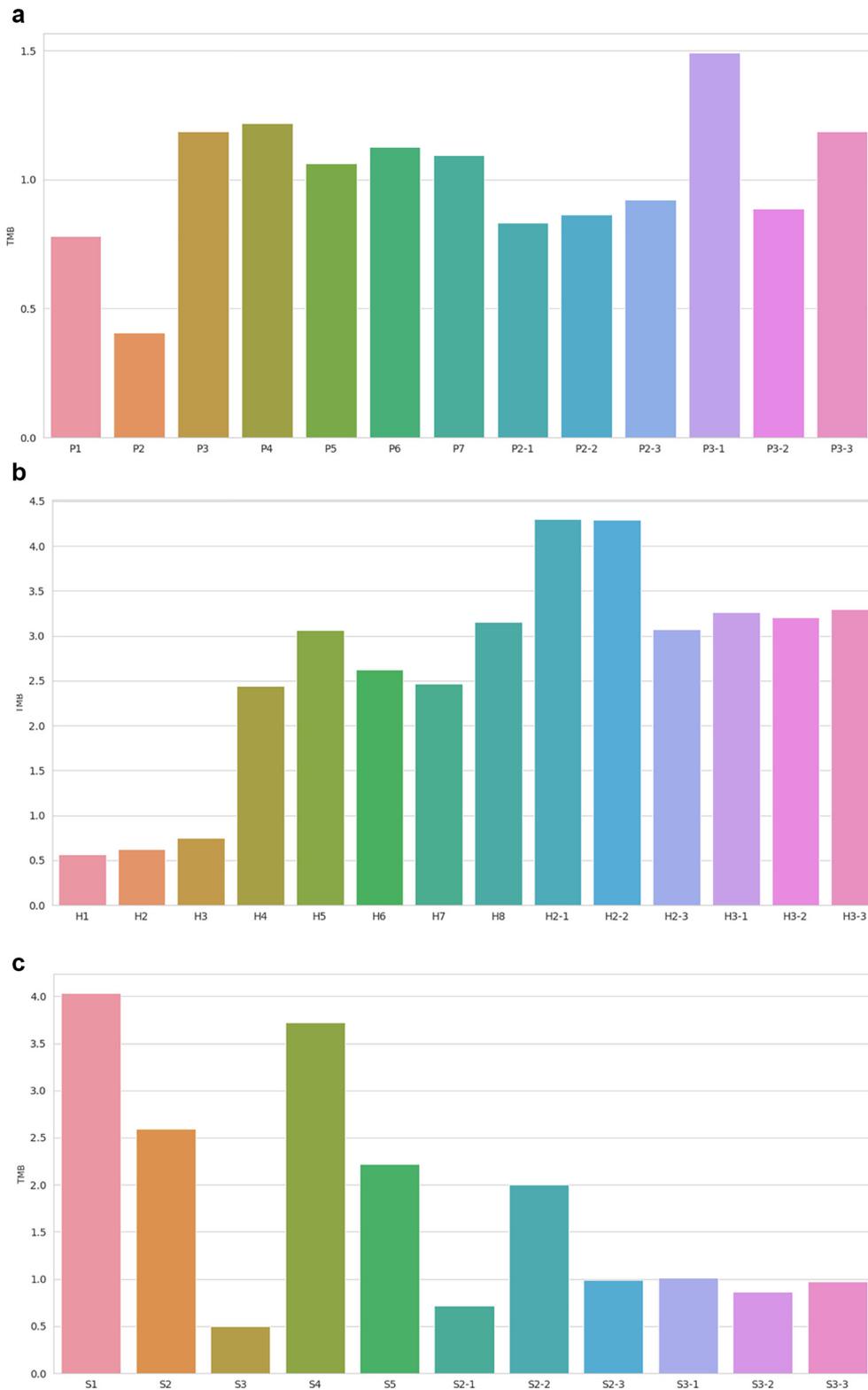
StarDist, within QuPath (Schmidt et al., 2018). Scripts were modified from those written in a QuPath tutorial for running StarDist (https://qupath.readthedocs.io/en/0.2/docs/advanced/stardist.html?highlight=stardist) and were applied to each project as batch process. Cell detection and measurement, including percentage of positivity and H-score of SOX9, were performed and measurement tables were exported via script. H-score is based on the percentage of cells stained with intensities of 0, 1+, 2+, and 3+ and calculated as $\sum$ [intensity (0, 1, 2, 3) × extent of each staining intensity (%)]. H-score ranged from 0 to 300. Three thresholds (0.1, 0.3, and 0.5) were set for intensity detection as in QuPath script. The tables were used for comparison and statistical analysis of SOX9 expression among
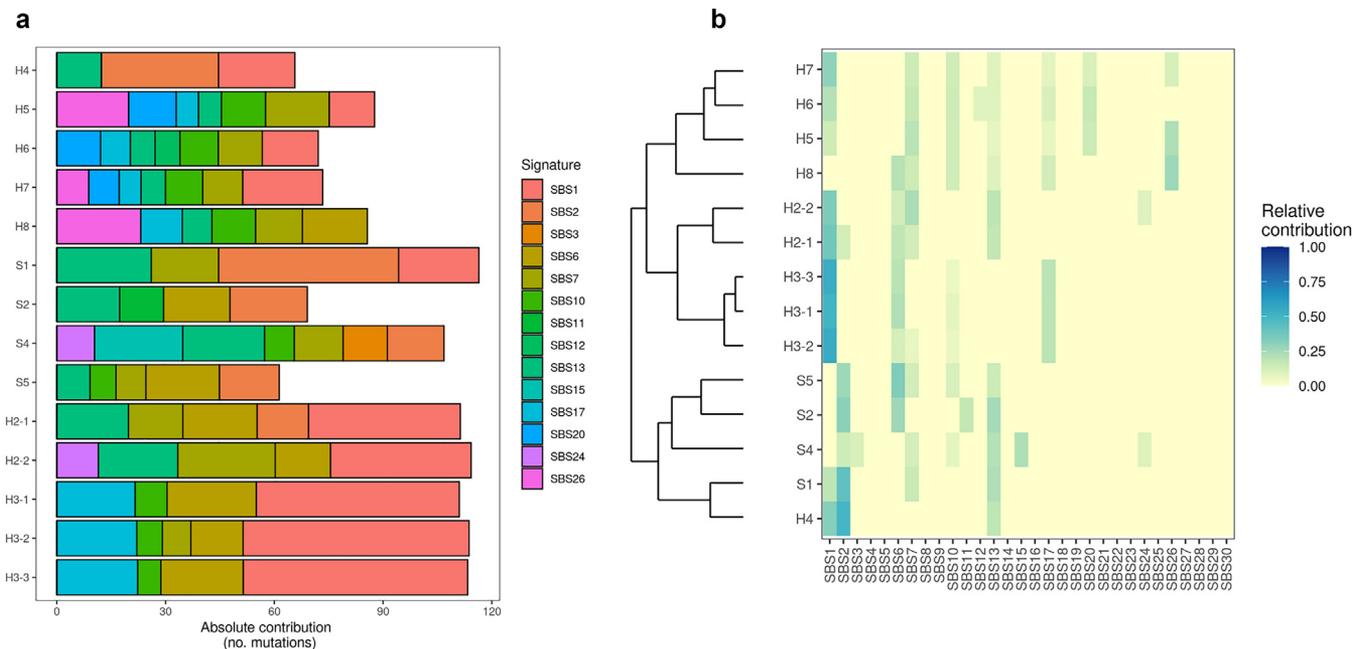
Figure 20. Mutational signatures (COSMIC v2) of selected regions with more than >70 somatic alterations in sweat gland carcinomas reveals both unique and intermingling characters. Absolute contributions of signatures **(a)**, heatmap and hierarchical clustering by signatures **(b)** of sampling sites.

tumors or nontumor adnexa using Python (3.7.0) library Pandas (1.1.3), Matplotlib (3.3.2), Seaborn (0.11.0), and SciPy (1.5.2).

Single-cell cytological clustering with QuPath measurement

Intratumoral heterogeneity was explored in three adnexal carcinomas (porocarcinoma, hidradenocarcinoma, and spiradenocarcinoma), three cases with histologically both apparent benign and malignant regions and six cases with only malignant regions. Multiple ROIs were selected in each case and every single cell was detected in all ROIs as described in the IHC section. Cell measurement tables containing 61 characteristics, such as morphological and staining features, were read into pandas dataframe, and 35 features that seemed meaningful for the further analysis were selected. The following features were selected: "nucleus: area μm^2"; "nucleus: length μm"; "nucleus: circularity"; "nucleus: max diameter μm"; "nucleus: min diameter μm"; "hematoxylin: nucleus: mean"; "hematoxylin: nucleus: median"; "hematoxylin: nucleus: min"; "hematoxylin: nucleus: max"; "hematoxylin: nucleus: SD"; "hematoxylin: cytoplasm: mean"; "hematoxylin: cytoplasm: median"; "hematoxylin: cytoplasm: min"; "hematoxylin: cytoplasm: max"; "hematoxylin: cytoplasm: SD"; "hematoxylin: cell: mean"; "hematoxylin: cell: median"; "hematoxylin: cell: min"; "hematoxylin: cell: max"; "hematoxylin: cell: SD"; "eosin: nucleus: mean"; "eosin: nucleus: median"; "eosin: nucleus: min"; "eosin: nucleus: max"; "eosin: nucleus: SD"; "eosin: cytoplasm: mean"; "eosin: cytoplasm: median"; "eosin: cytoplasm: min"; "eosin: cytoplasm: max"; "eosin: cytoplasm: SD"; "eosin: cell: mean"; "eosin: cell: median"; "eosin: cell: min"; "eosin: cell: max"; and "eosin: cell: SD." The values in each of the columns were preprocessed by standardization, that is, removing the mean value and dividing by SD using Scikit-learn ("0.23.2"). Dimensional reduction of the standardized data by uniform manifold approximation and projection was performed using Python umap-learn Library ("0.5.0") with default parameters. The Gaussian mixture model was used for unsupervised clustering of the data with sklearn.mixture.GaussianMixture with 'covariance_type': 'full.' Cluster number was determined using

analytic criterion Akaike information criterion and Bayesian information criterion. The proportion of clustered cells for each ROI was calculated and hierarchical clustering of ROIs was performed to compare similarities using seaborn.clustermap with default linkage method ("average"). ROIs were grouped based on clustering into several large clusters. The area percentage of each group were measured using QuPath on whole-slide image.

Laser capture microdissection

FFPE tissue samples were sectioned at a 10-μm thickness, mounted on polyethylene naphthalate membrane—coated slides (Leica Microsystems, Wetzlar, Germany), and stained with hematoxylin (Sakura Finetek, St. Torrance, CA) for 1 minute and eosin (Sakura Finetek) for 30 seconds. Sampling sites were separately microdissected using an LMD 7000 instrument (Leica Microsystems).

Genomic analysis

WES was performed using genomic DNA isolated from FFPE specimens using a GeneRead DNA FFPE kit (Qiagen, Hilden, Germany), and exonic fragments were enriched using a Human Core Exome kit with RefSeq spike-in (both from Twist Bioscience, South San Francisco, CA). Massively parallel sequencing of prepared libraries was performed using a NovaSeq 6000 system (Illumina, San Diego, CA) (porocarcinoma 1, hidradenocarcioma 1, and spiradenocarcinoma 1) and NextSeq 1000 system (Illumina) (porocarcinoma 2 and 3, hidradenocarcinoma 2 and 3, and spiradenocarcinoma 2 and 3). WES included masked nucleotides with a quality value <20 that were independently aligned to a human reference genome (hg38) using BWA (http://bio-bwa.sourceforge.net/) and Bowtie 2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Both somatic synonymous and nonsynonymous alterations were called using our in-house caller and two publicly available alteration callers, Genome Analysis Toolkit (https://gatk.broadinstitute.org/hc/en-us) MuTect2 and VarScan 2 (http://varscan.sourceforge.net/). The exclusion criteria for alterations were variant present in normal human genomes in either the 1000
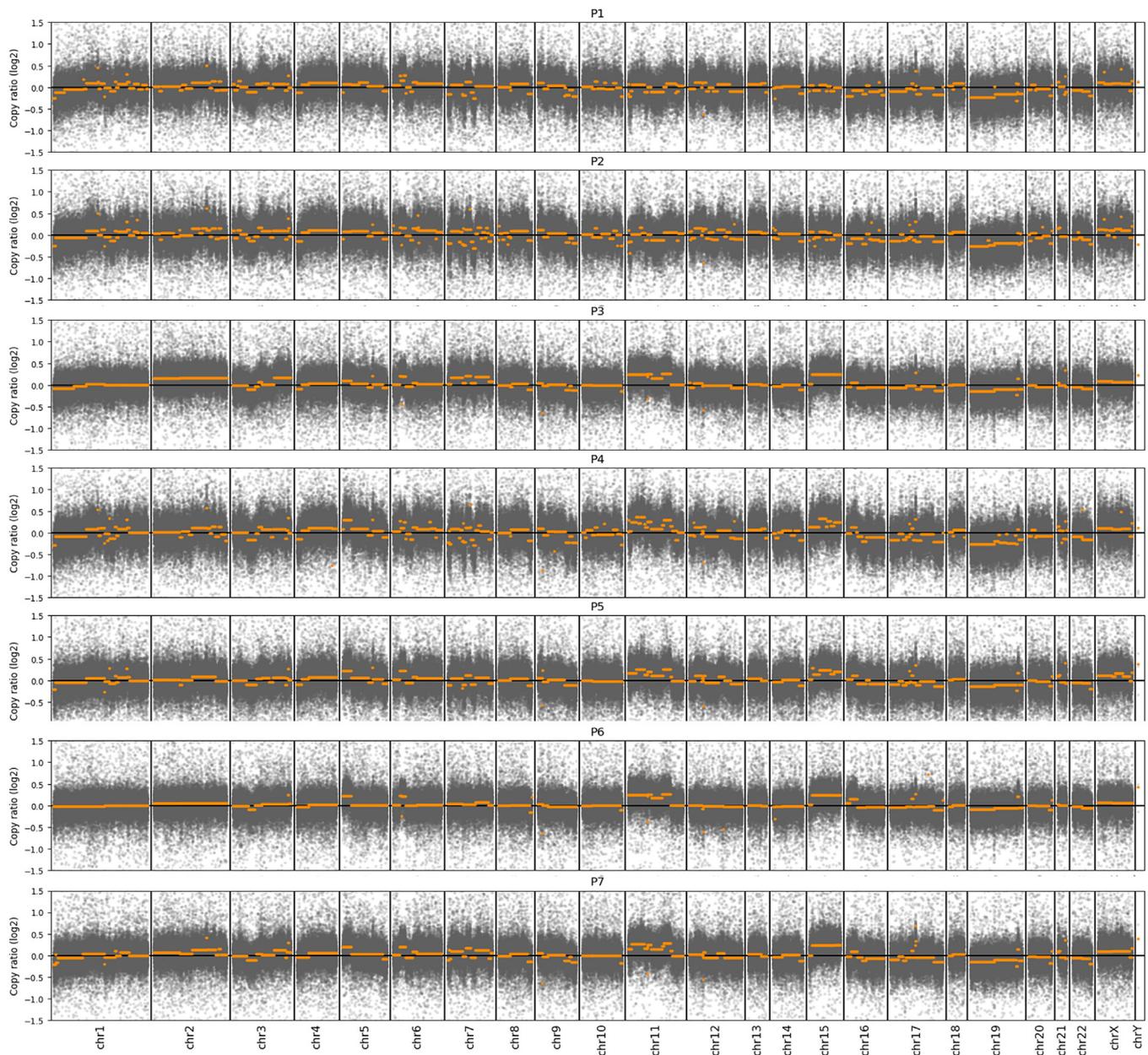
Figure 21. Copy number plot of each region of porocarcinoma 1 derived from M-WES. Y axis represents log2 copy ratio and x axis shows chromosomal numbers.

Genomes Project dataset (https://www.internationalgenome.org/) or our in-house database; alteration was called by only one of the callers; the read depth was <20 or the variant allele frequency was <0.05.

In addition, alterations were selected using the filtering pipeline MicroSEC, which removes errors caused by single-stranded DNA originating from other regions with microhomology in FFPE samples (Ikegami et al., 2021). Gene alterations were annotated by SnpEff (https://pcingola.github.io/SnpEff/). Genomic characterization was analyzed by focusing on the common oncogenic signaling pathways (15). TMB was defined as the total number of nonsynonymous alterations, including indels, per megabase (muts/Mb) in WES.

Mutational signature analysis

MutationalPatterns (3.2.0) (Blokzijl et al., 2018) was used to estimate the absolute and relative contribution of COSMIC (v.2, https://cancer.sanger.ac.uk/signatures/signatures_v2/) single-base substitutions. Only sampling sites with >70 somatic substitutions were analyzed. Known mutational signature trinucleotide frequencies were fitted to the observed alteration using the function "fit_to_signatures."

Copy number analysis

Copy number variants were assessed using the CNVkit (0.9.9). Genes with log 2 copy number ratio estimate of <−1 or >0.8 were considered to have loss (deletion) or gain (amplification), respectively.
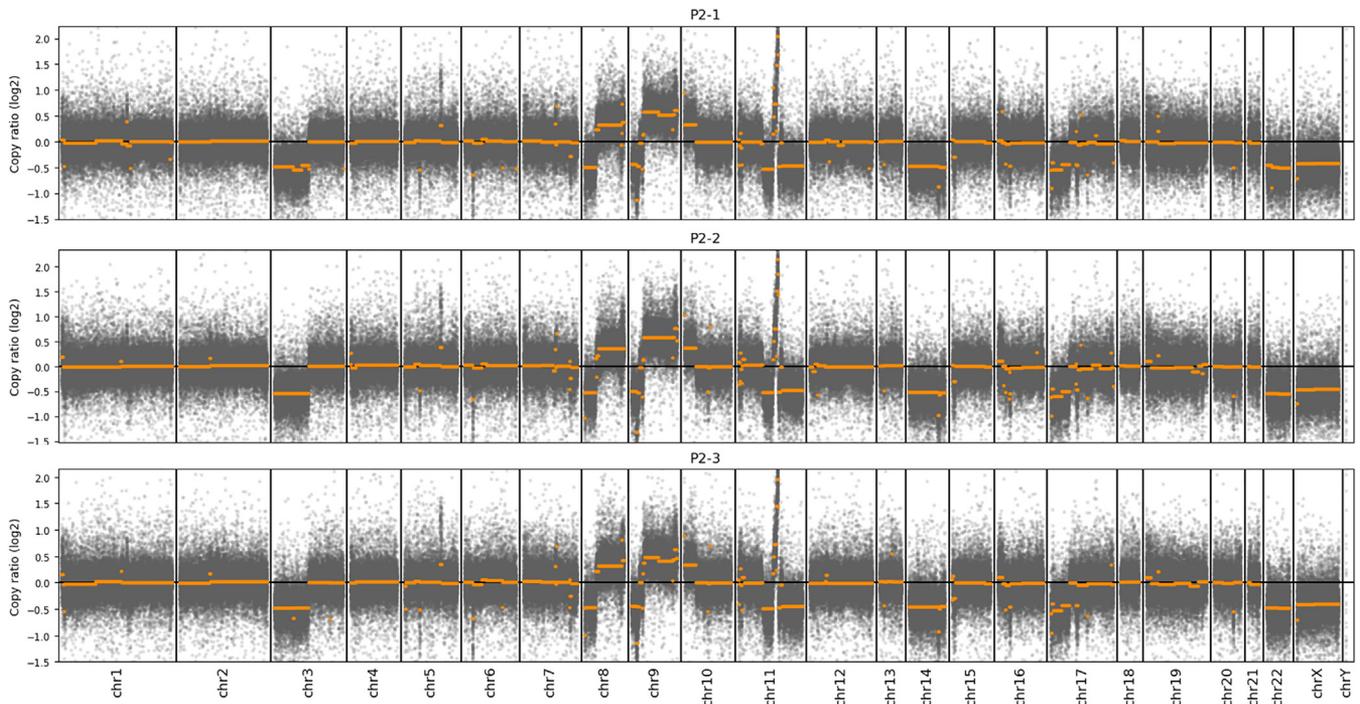
Figure 22. Copy number plot of each region of porocarcinoma 2 derived from M-WES.

Phylogenetic tree construction

Phylogenetic tree was constructed through R package meskit (1.2.0) (Liu et al., 2021) using the neighbor-joining algorithm.

FISH assays

FISH assays were performed for *YAP1*, *NUM1*, and *MAML2* with *YAP1* break-apart probe set (Agilent Technologies) and

NUTM1 break-apart probe set (Agilent Technologies), and *MAML2* ZytoLight Dual Color Break Apart Probes (ZytoVision, Bremerhaven, Germany). FISH images were captured using the Metafer Slide.

Scanning Platform (MetaSystems, Altlußheim, Germany) and 200 nonoverlapping tumor nuclei with at least one signal were
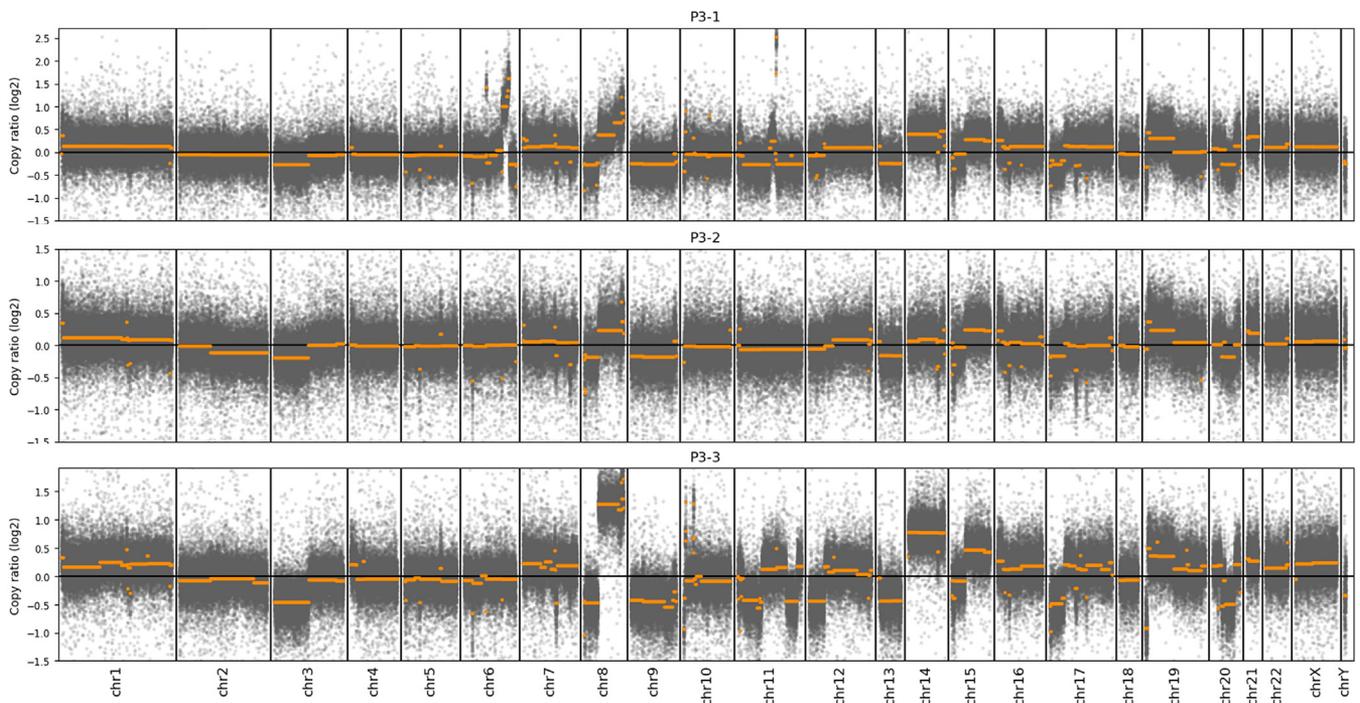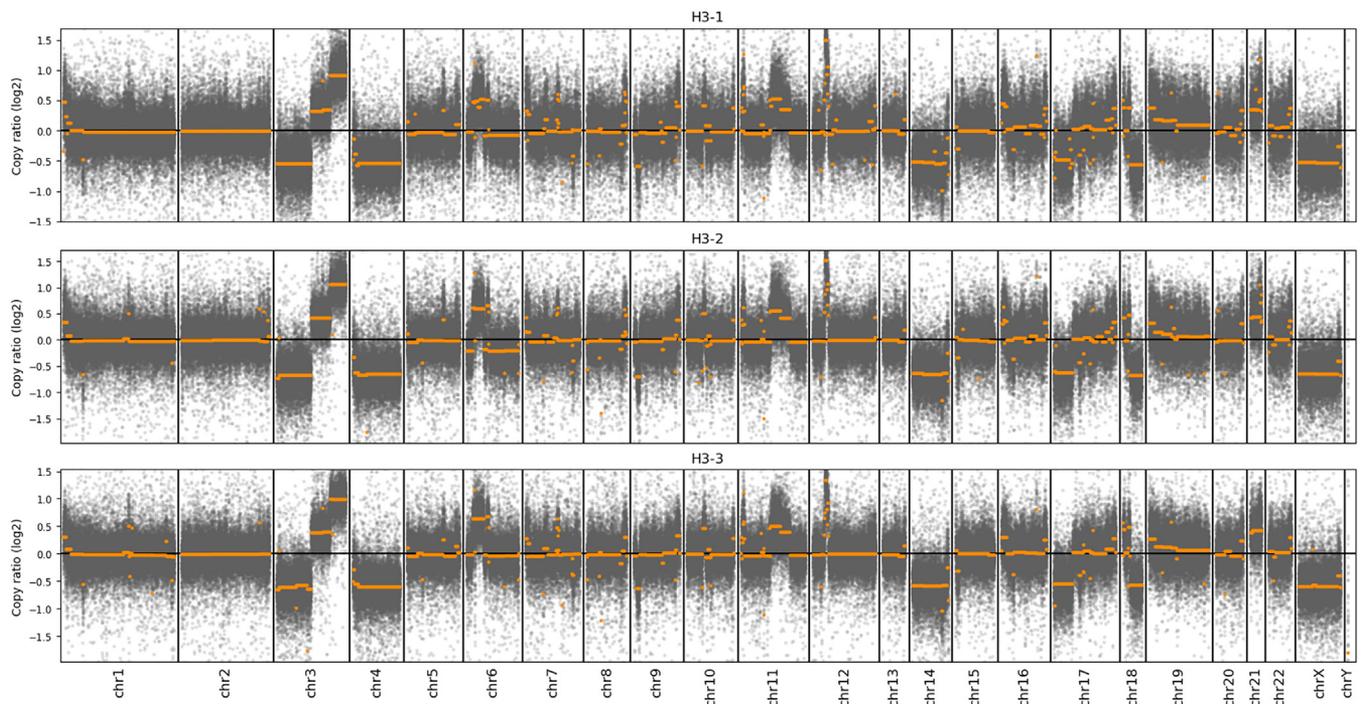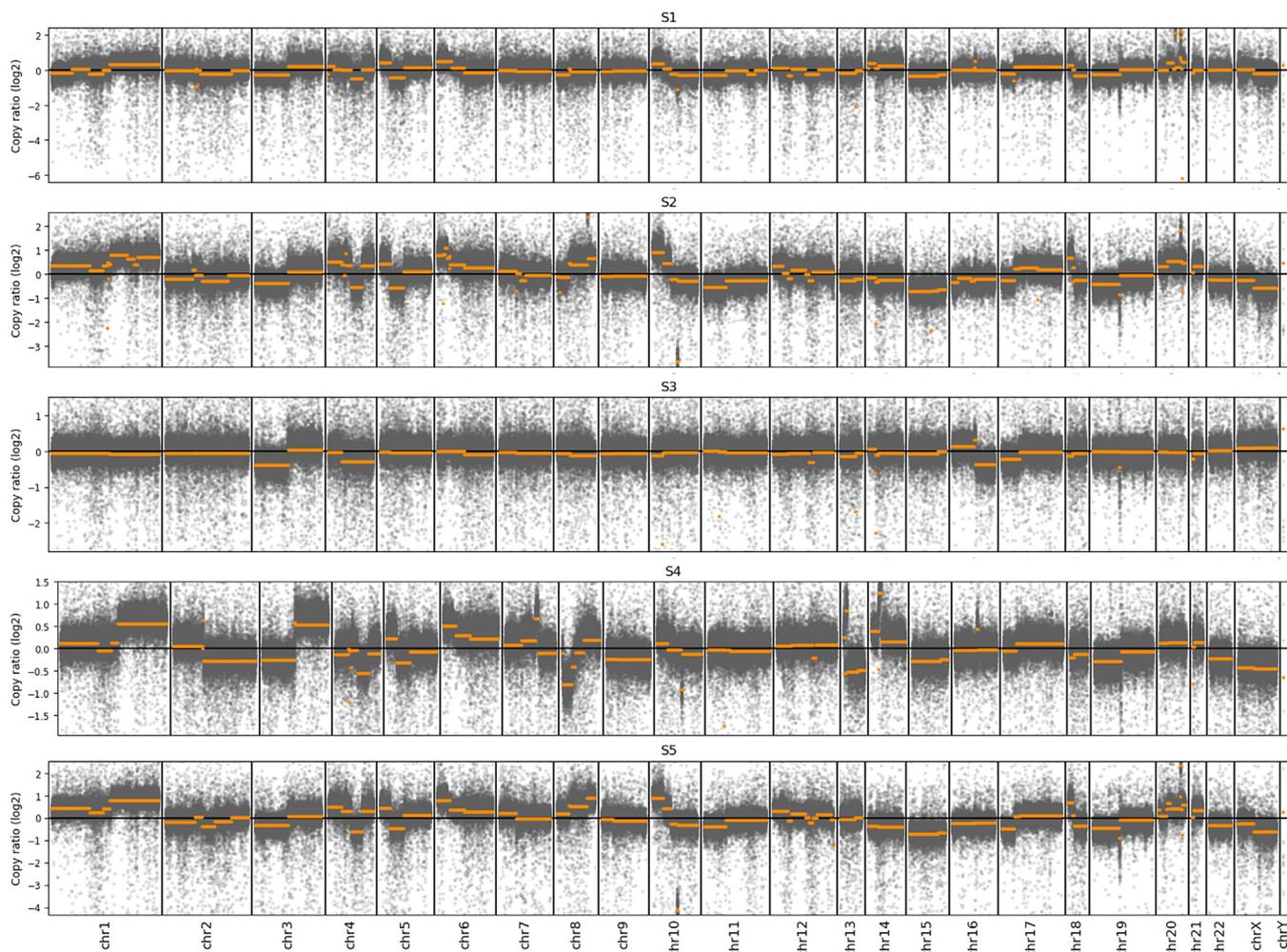

Figure 23. Copy number plot of each region of porocarcinoma 3 derived from M-WES.
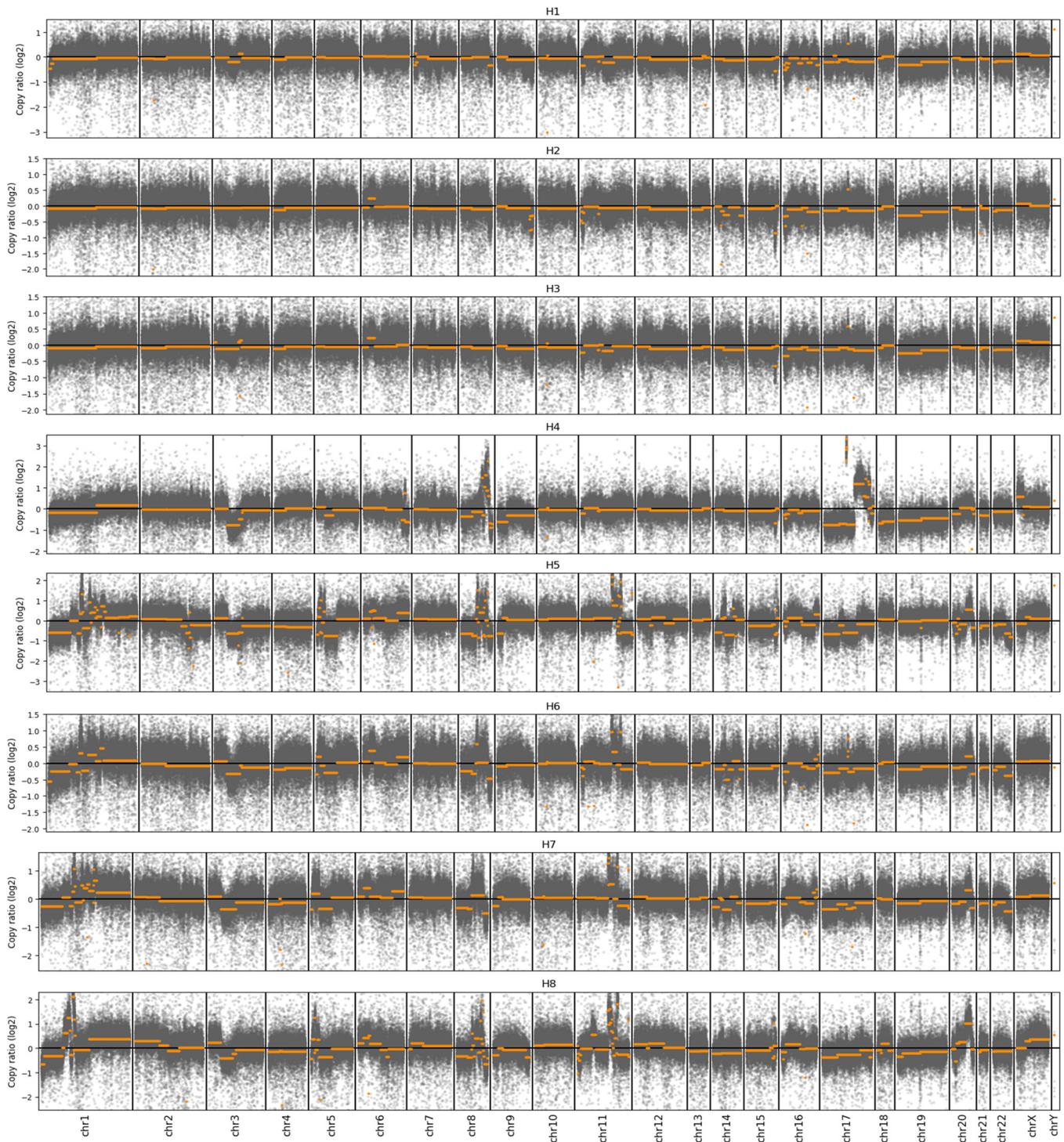
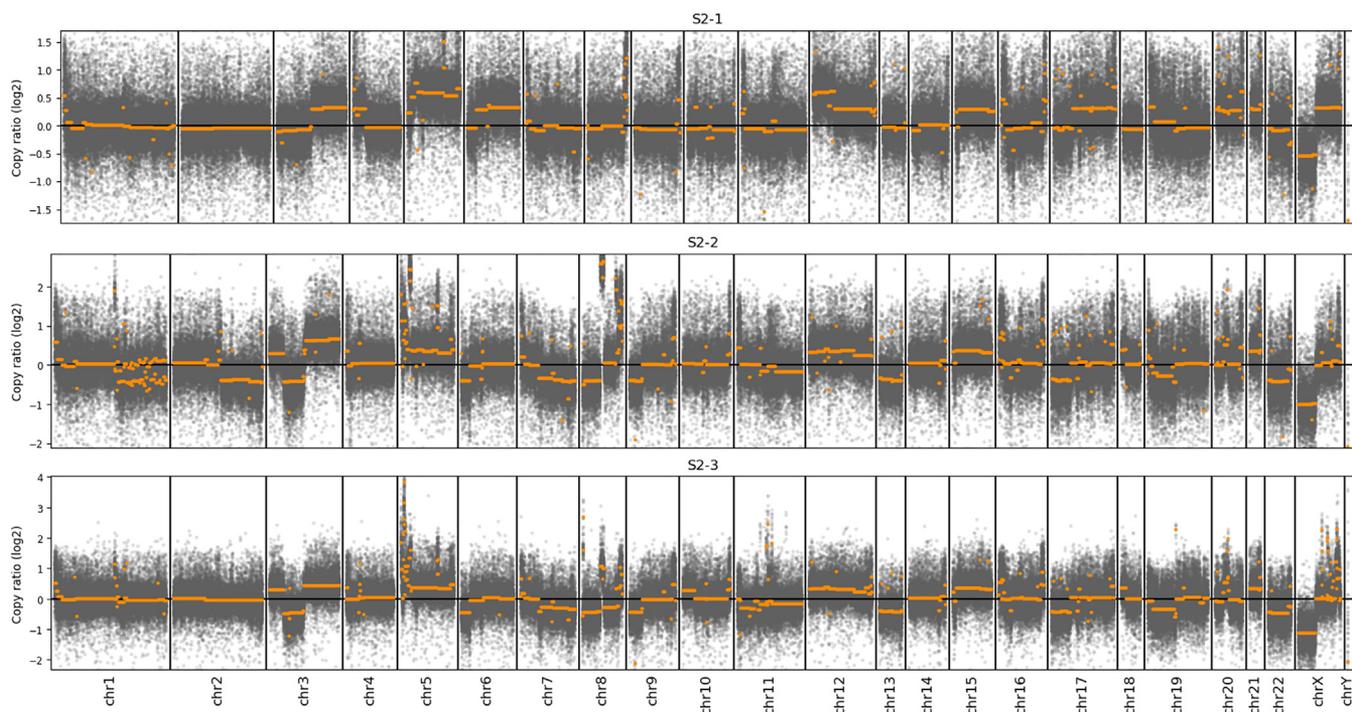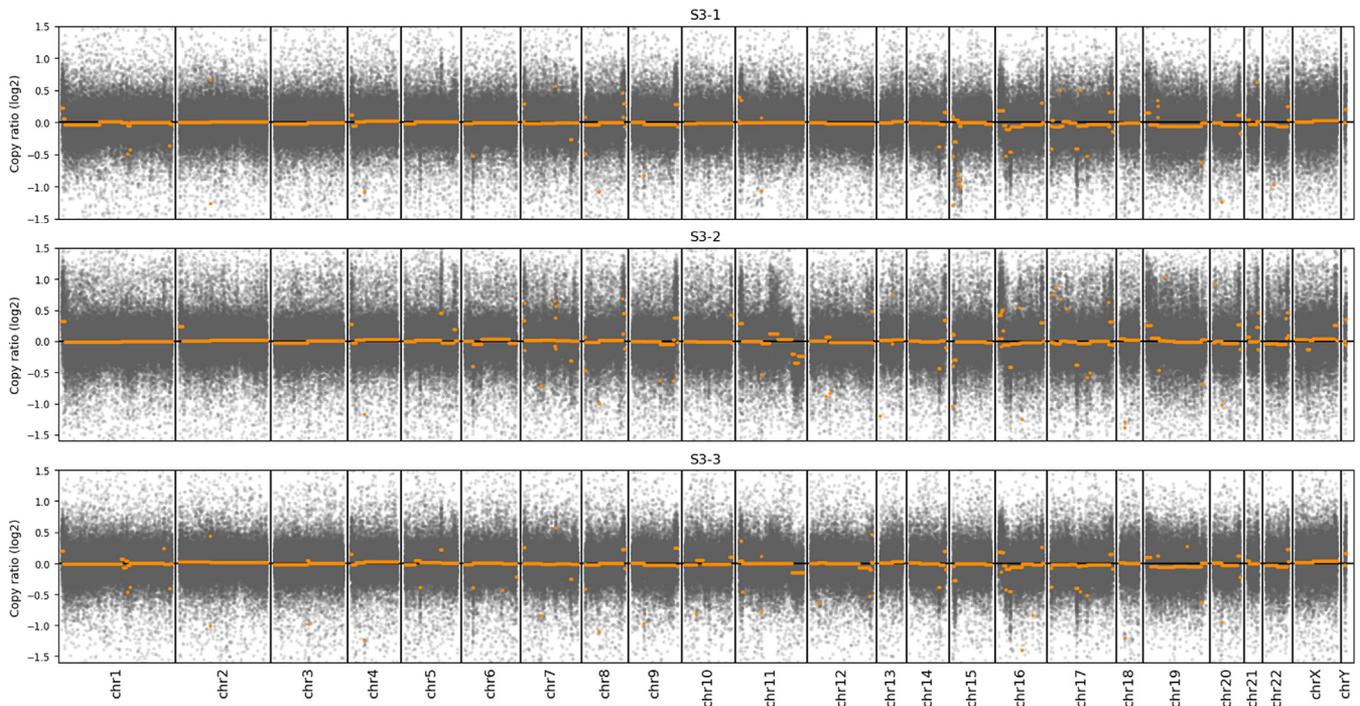**Figure 24. Copy number plot of each region of hidradenocarcinoma 1 derived from M-WES.**

examined. More than 20% of nuclei with split signals were deemed to represent rearrangement-positive cells.

Statistics

Statistical analysis was performed using RStudio (R version 4.1.2) and SciPy (1.5.2). Percentage positivity and H-score of nontumor skin adnexa and adnexal tumors were compared using one-way ANOVA with Tukey multiple-comparison test. Patients' age, and diameters of

three groups of adnexal tumors (HF tumors, sweat gland tumors, and sebaceous tumors) were compared using one-way ANOVA, and gender was analyzed using χ^2 test. Similarly, Sox9 percentage positivity, H-score, patients' age, and diameters of adnexal carcinomas with and without metastasis or recurrence were compared using two-tailed Student t test, and gender was analyzed using χ^2 test. Corrected P-values <0.05 were considered significant. A receiver operating characteristic curve is a plot of the sensitivity (true positive rate) versus
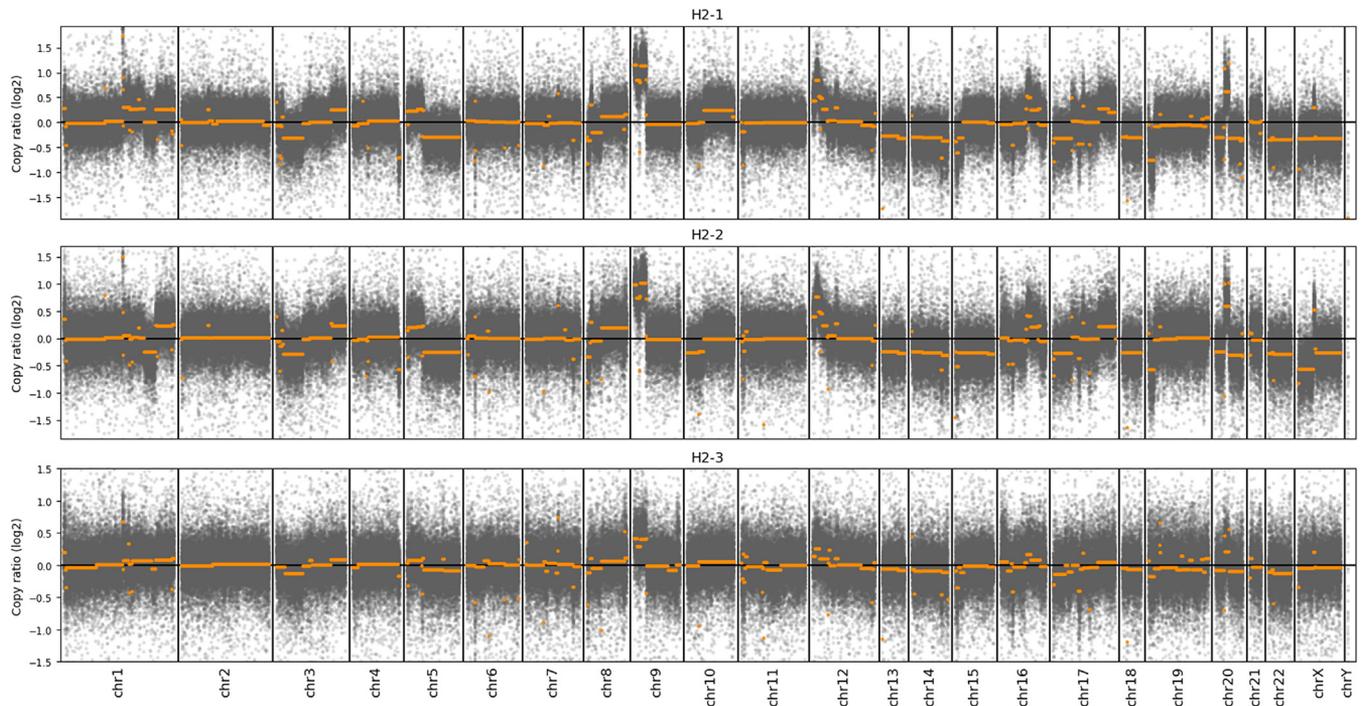
Figure 25. Copy number plot of each region of hidradenocarcinoma 2 derived from M-WES.

1-specificity (false positive rate) of some diagnostic test. True positive rate and false positive rate of classification between BCC/adnexal tumors and SCC by SOX9 percentage positivity and H-score at various threshold settings were calculated using sklearn.metrics.roc_curve. Thereafter, receiver operating characteristic curve was plotted with matplotlib.pyplot. Area under the curve was computed using sklearn.metrics.roc_auc_score.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Bam files of all tumor exome data can be viewed at https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA993176. Part of the computational codes used in this study are available in GitHub repository (https://github.com/Nash-RYU/Single-cell-cytological-clustering).

Figure 26. Copy number plot of each region of hidradenocarcinoma 3 derived from M-WES.

Figure 27. Copy number plot of each region of spiradenocarcinoma 1 derived from M-WES.

Figure 28. Copy number plot of each region of spiradenocarcinoma 2 derived from M-WES.

**Figure 29. Copy number plot of each region of spiradenocarcinoma 3 derived from M-WES.**

ORCIDs
Yuuki Nishimura: https://orcid.org/0009-0000-5489-3076
Eijitsu Ryo: https://orcid.org/0009-0003-1362-1332
Satoshi Inoue: https://orcid.org/0000-0003-1484-2228
Masahito Kawazu: https://orcid.org/0000-0003-4146-3629
Toshihide Ueno: https://orcid.org/0000-0002-7408-7298
Kenjiro Namikawa: https://orcid.org/0000-0003-2763-7346

Akira Takahashi: https://orcid.org/0000-0002-9304-7590
Dai Ogata: https://orcid.org/0000-0002-1335-7100
Akihiko Yoshida: https://orcid.org/0000-0002-3373-0099
Naoya Yamazaki: https://orcid.org/0000-0002-9638-0428
Hiroyuki Mano: https://orcid.org/0000-0003-4645-0181
Yasushi Yatabe: https://orcid.org/0000-0003-1788-559X
Taisuke Mori: https://orcid.org/0000-0003-1838-7883

Figure 30. SOX9 expression in bulge and hair germ. (a) SOX9 and **(b)** Ber-EP4 immunohistochemistry. SOX9 is strongly expressed from hair germ(sub-bulge) labeled by Ber-EP4 to bulge but weakly expressed above bulge. Bar = 50 μm.
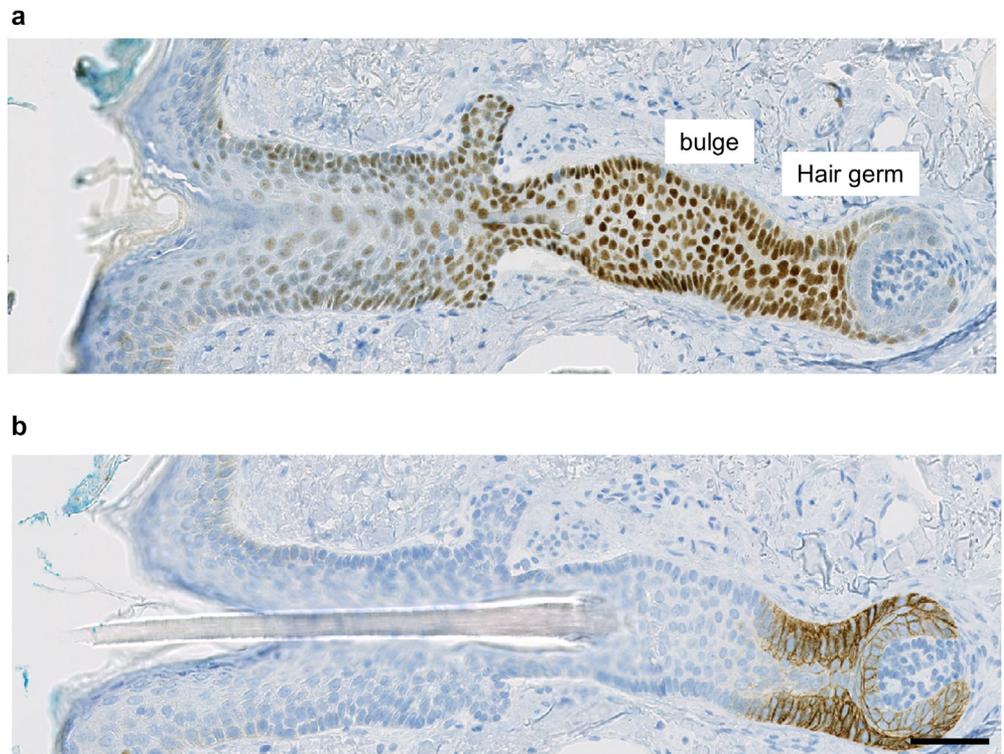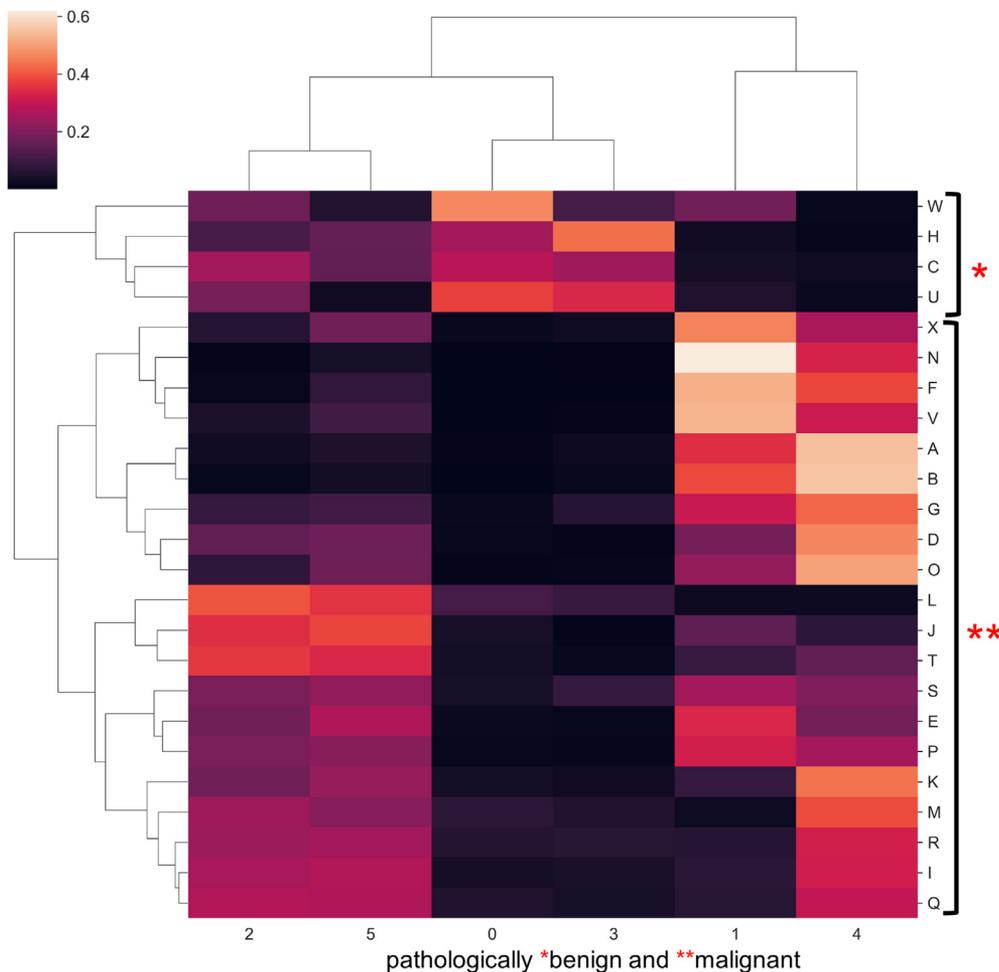
Figure 31. Morphological clustering of the porocarcinoma 1 with cluster number 6. Hierarchical clustering of ROIs is quite analogous to the original result with cluster number 10 in Figure 8e.

CONFLICT OF INTEREST

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

Conceptualization: YN and TM; Data Curation: YN, TM, and KN; Formal Analysis: YN and TU; Funding Acquisition: YN and TM; Investigation: YN, ER, AY, TM, SI, and MK; Methodology: YN, TM, SI, MK, TU, and HM; Project Administration: TM and NY; Resources: KN, AT, DO, NY, TU, and HM; Software: TU, HM, and YN; Supervision: TM and YY; Validation: YN and TM; Visualization: YN, TU, and TM; Writing-Original Draft: YN and TM; Writing-Review and Editing: all authors

REFERENCES

Aleman A, Adrien L, Lopez-Serra L, Cordon-Cardo C, Esteller M, Belbin TJ, et al. Identification of DNA hypermethylation of SOX9 in association with bladder cancer progression using CpG microarrays. Br J Cancer 2008;98: 466–73.

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578: 94–101.

Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep 2013;3:246–59.

André F, Ciruelos E, Rubovszky G, Campone M, Loibl S, Rugo HS, et al. Alpelisib for PIK3CA-Mutated, hormone receptor–positive advanced breast cancer. N Engl J Med 2019;380:1929–40.

André F, Ciruelos EM, Juric D, Loibl S, Campone M, Mayer IA, et al. Alpelisib plus fulvestrant for PIK3CA-mutated, hormone receptor-positive, human epidermal growth factor receptor-2–negative advanced breast cancer: final overall survival results from SOLAR-1. Ann Oncol 2021;32:208–17.

Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.

Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. Genome Med 2018;10:33.

Cheng PF, Shakhova O, Widmer DS, Eichhoff OM, Zingg D, Frommel SC, et al. Methylation-dependent SOX9 expression mediates invasion in human melanoma cells and is a negative prognostic factor in advanced melanoma. Genome Biol 2015;16:42.

Cunningham R, Hansen CG. The Hippo pathway in cancer: YAP/TAZ and TEAD as therapeutic targets in cancer. Clin Sci (Lond) 2022;136:197−222.

Davies HR, Hodgson K, Schwalbe E, Coxhead J, Sinclair N, Zou X, et al. Epigenetic modifiers DNMT3A and BCOR are recurrently mutated in CYLD cutaneous syndrome. Nat Commun 2019;10:4717.

Elder D, Massi D, Scolyer R, Willemze R. WHO classification of skin tumours. 4th ed. Lyon: IARC; 2018.

Gerlinger M, Horswell S, Larkin J, Rowan AJ, Salm MP, Varela I, et al. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Nat Genet 2014;46:225−33.

Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 2012;366:883−92.

Grimm D, Bauer J, Wise P, Krüger M, Simonsen U, Wehland M, et al. The role of SOX family members in solid tumours and metastasis. Semin Cancer Biol 2020;67:122−53.

Grys BT, Lo DS, Sahin N, Kraus OZ, Morris Q, Boone C, et al. Machine learning and computer vision approaches for phenotypic profiling. J Cell Biol 2017;216:65−71.

Hao JJ, Lin DC, Dinh HQ, Mayakonda A, Jiang YY, Chang C, et al. Spatial intratumoral heterogeneity and temporal clonal evolution in esophageal squamous cell carcinoma. Nat Genet 2016;48:1500−7.

Hartsough EM, Moran J, Devins K, Wszolek M, Cornejo KM. An unusual case of a scrotal porocarcinoma and review of the literature. Am J Dermatopathol 2023;45:51−5.

Hile G, Harms PW. Update on molecular genetic alterations of cutaneous adnexal neoplasms. Surg Pathol Clin 2021;14:251−72.

Huss R, Coupland SE. Software-assisted decision support in digital histopathology. J Pathol 2020;250:685−92.

Ikegami M, Kohsaka S, Hirose T, Ueno T, Inoue S, Kanomata N, et al. MicroSEC filters sequence errors for formalin-fixed and paraffin-embedded samples. Commun Biol 2021;4:1396.

Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non−small-cell lung cancer. N Engl J Med 2017;376:2109−21.

Jin N, Keam B, Cho J, Lee MJ, Kim HR, Torosyan H, et al. Therapeutic implications of activating noncanonical PIK3CA mutations in head and neck squamous cell carcinoma. J Clin Invest 2021;131:e150335.

Jo A, Denduluri S, Zhang B, Wang Z, Yin L, Yan Z, et al. The versatile functions of Sox9 in development, stem cells, and human diseases. Genes Dis 2014;1:149−61.

Kamachi Y, Kondoh H. Sox proteins: regulators of cell fate specification and differentiation. Development 2013;140:4129−44.

Kazakov DV, Ivan D, Kutzner H, Spagnolo DV, Grossmann P, Vanecek T, et al. Cutaneous hidradenocarcinoma: a clinicopathological, immunohistochemical, and molecular biologic study of 14 cases, including HER2/neu gene expression/amplification, TP53 gene mutation analysis, and t(11;19) translocation. Am J Dermatopathol 2009;31:236−47.

Kim AL, Back JH, Chaudhary SC, Zhu Y, Athar M, Bickers DR. SOX9 transcriptionally regulates mTOR-induced proliferation of basal cell carcinomas. J Invest Dermatol 2018;138:1716−25.

Koh G, Degasperi A, Zou X, Momen S, Nik-Zainal S. Mutational signatures: emerging concepts, caveats and clinical applications. Nat Rev Cancer 2021;21:619−37.

Krahl D, Sellheyer K. Basal cell carcinoma and pilomatrixoma mirror human follicular embryogenesis as reflected by their differential expression patterns of SOX9 and β-catenin. Br J Dermatol 2010;162:1294−301.

Kuma Y, Yamada Y, Yamamoto H, Kohashi K, Ito T, Furue M, et al. A novel fusion gene CRTC3-MAML2 in hidradenoma: histopathological significance. Hum Pathol 2017;70:55−61.

Lee JK, Louzada S, An Y, Kim SY, Kim S, Youk J, et al. Complex chromosomal rearrangements by single catastrophic pathogenesis in NUT midline carcinoma. Ann Oncol 2017;28:890−7.

Liu M, Chen J, Wang X, Wang C, Zhang X, Xie Y, et al. MesKit: a tool kit for dissecting cancer evolution of multi-region tumor biopsies through somatic alterations. GigaScience 2021;10:giab036.

Lu C, Fuchs E. Sweat gland progenitors in development, homeostasis, and wound repair. Cold Spring Harb Perspect Med 2014;4:a015222.

Macagno N, Kervarrec T, Sohier P, Poirot B, Haffner A, Carlotti A, et al. NUT is a specific immunohistochemical marker for the diagnosis of YAP1-NUTM1-rearranged cutaneous poroid neoplasms. Am J Surg Pathol 2021;45:1221−7.

Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol 2010;5:1315−6.

McEvoy CR, Fox SB, Prall OWJ. Emerging entities in NUTM1-rearranged neoplasms. Genes Chromosomes Cancer 2020;59:375−85.

McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell 2017;168:613−28.

McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw 2018;3:861.

Ming Z, Vining B, Bagheri-Fam S, Harley V. SOX9 in organogenesis: shared and unique transcriptional functions. Cell Mol Life Sci 2022;79:522.

Morita R, Sanzen N, Sasaki H, Hayashi T, Umeda M, Yoshimura M, et al. Tracing the origin of hair follicle stem cells. Nature 2021;594:547−52.

Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. Nat Rev Clin Oncol 2015;12:258−72.

Nonaka D, Chiriboga L, Rubin BP. Sox10: a pan-schwannian and melanocytic marker. Am J Surg Pathol 2008;32:1291−8.

Nowak JA, Polak L, Pasolli HA, Fuchs E. Hair follicle stem cells are specified and function in early skin morphogenesis. Cell Stem Cell 2008;3:33−43.

Ohtomo R, Mori T, Shibata S, Tsuta K, Maeshima AM, Akazawa C, et al. SOX10 is a novel marker of acinus and intercalated duct differentiation in salivary gland tumors: A clue to the histogenesis for tumor diagnosis. Mod Pathol 2013;26:1041−50.

Peterson SC, Eberl M, Vagnozzi AN, Belkadi A, Veniaminova NA, Verhaegen ME, et al. Basal cell carcinoma preferentially arises from stem cells within hair follicle and mechanosensory niches. Cell Stem Cell 2015;16:400−12.

Prieto-Granada C, Castner N, Chen A, Li J, Yue B, Wong JY, et al. Behavior of cutaneous adnexal malignancies: a single institution experience. Pathol Oncol Res 2020;26:239−44.

Purba TS, Haslam IS, Shahmalak A, Bhogal RK, Paus R. Mapping the expression of epithelial hair follicle stem cell-related transcription factors LHX2 and SOX9 in the human hair follicle. Exp Dermatol 2015;24:462−7.

Rashid M, van der Horst M, Mentzel T, Butera F, Ferreira I, Pance A, et al. ALPK1 hotspot mutation as a driver of human spiradenoma and spiradenocarcinoma. Nat Commun 2019;10:2213.

Russell-Goldman E, Hornick JL, Hanna J. Utility of YAP1 and NUT immunohistochemistry in the diagnosis of porocarcinoma. J Cutan Pathol 2021;48:403−10.

Sarkar A, Hochedlinger K. The Sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 2013;12:15−30.

Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2018.

Sekine S, Kiyono T, Ryo E, Ogawa R, Wakai S, Ichikawa H, et al. Recurrent YAP1-MAML2 and YAP1-NUTM1 fusions in poroma and porocarcinoma. J Clin Invest 2019;129:3827−32.

Sekine S, Ogawa R, Oshiro T, Kanemitsu Y, Taniguchi H, Kushima R, et al. Frequent lack of GNAS mutations in colorectal adenocarcinoma associated with GNAS -mutated villous adenoma. Genes Chromosomes Cancer 2014;53:366−72.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. Nature 2012;486:400−4.

Taylor J, Pavlick D, Yoshimi A, Marcelus C, Chung SS, Hechtman JF, et al. Oncogenic TRK fusions are amenable to inhibition in hematologic malignancies. J Clin Invest 2018;128:3819−25.

Uhler C, Shivashankar GV. Nuclear mechanopathology and cancer diagnosis. Trends Cancer 2018;4:320−31.

Ullah I, Karthik GM, Alkodsi A, Kjällquist U, Stålhammar G, Lövrot J, et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. J Clin Invest 2018;128:1355−70.

van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. Nat Med 2021;27:775−84.

Vidal VPI, Chaboissier MC, Lützkendorf S, Cotsarelis G, Mill P, Hui CC, et al. Sox9 is essential for outer root sheath differentiation and the formation of the hair stem cell compartment. Curr Biol 2005;15: 1340−51.

Vidal VPI, Ortonne N, Schedl A. SOX9 expression is a general marker of basal cell carcinoma and adnexal-related neoplasms. J Cutan Pathol 2008;35: 373−9.

Winnes M, Mölne L, Suurküla M, Andrén Y, Persson F, Enlund F, et al. Frequent fusion of theCRTC1 andMAML2 genes in clear cell variants of cutaneous hidradenomas. Genes Chromosomes Cancer 2007;46: 559−63.

Yoshida KI, Nakano Y, Honda-Kitahara M, Wakai S, Motoi T, Ogura K, et al. Absence of H3F3A mutation in a subset of malignant giant cell tumor of bone. Mod Pathol 2019;32:1751−61.