Supplementary Materials of Pathomic Features Reveal Immune and Molecular Evolution from Lung Preneoplasia to Invasive Adenocarcinoma

Supplementary Tables

Supp Table 1. Statistics of patients, slides, and regions of interest (ROI) within the Japan, China, and USA cohorts belonging to distinct pathological stages (Normal, AAH, AIS, MIA, and ADC).

Entity	**Patient (n=98)**			**Slide (n=162)**			**ROI (n=669)**		
Cohort	Japan	China	USA	Japan	China	USA	Japan	China	USA
Stages									
Normal	0	0	0	0	0	9	66	16	61
AAH	6	2	1	22	3	17	47	12	70
AIS	12	3	1	22	5	1	69	15	10
MIA	21	2	3	22	4	3	67	10	21
ADC	20	14	13	20	18	16	86	63	56
Total	59	21	18	86	30	46	335	116	218

Note: Normal ROIs within Japan and China cohorts were annotated from slides diagnosed with severe pathological stages, such as AAH and AIS.

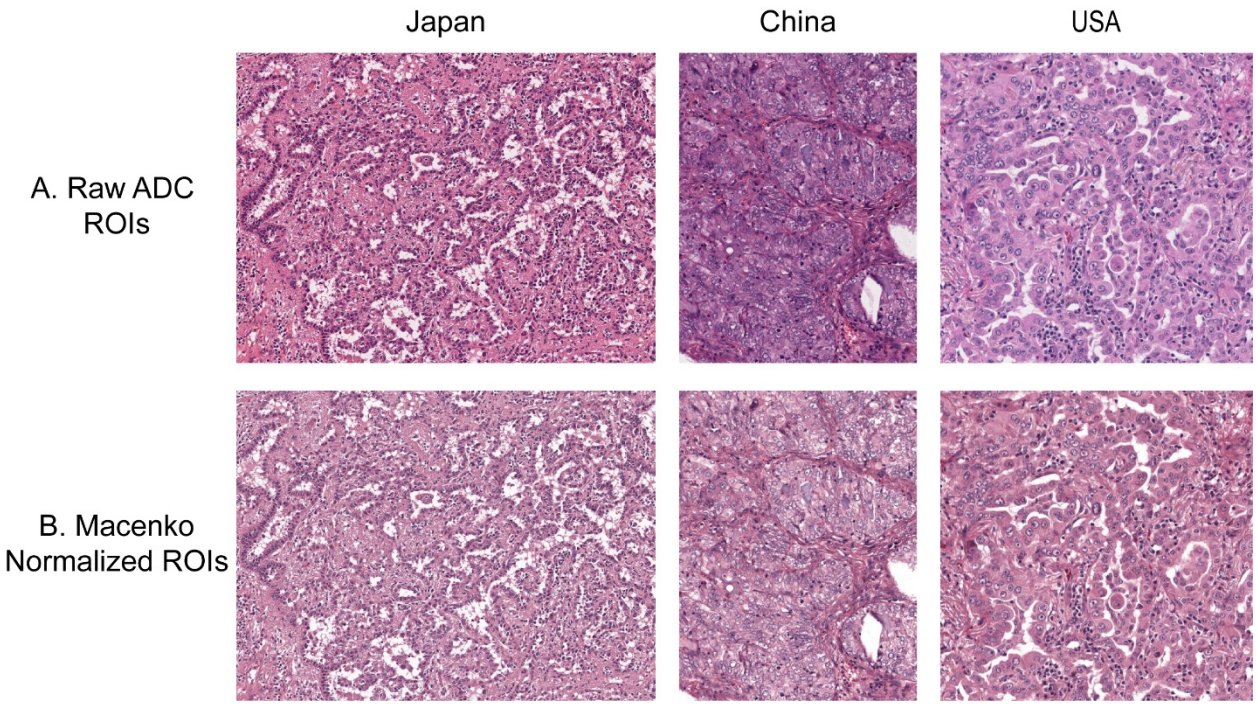
Supp Table 2. The numbers of slides and annotated regions of interest (ROI) with different diagnosis in each pathological stage in three study cohorts.

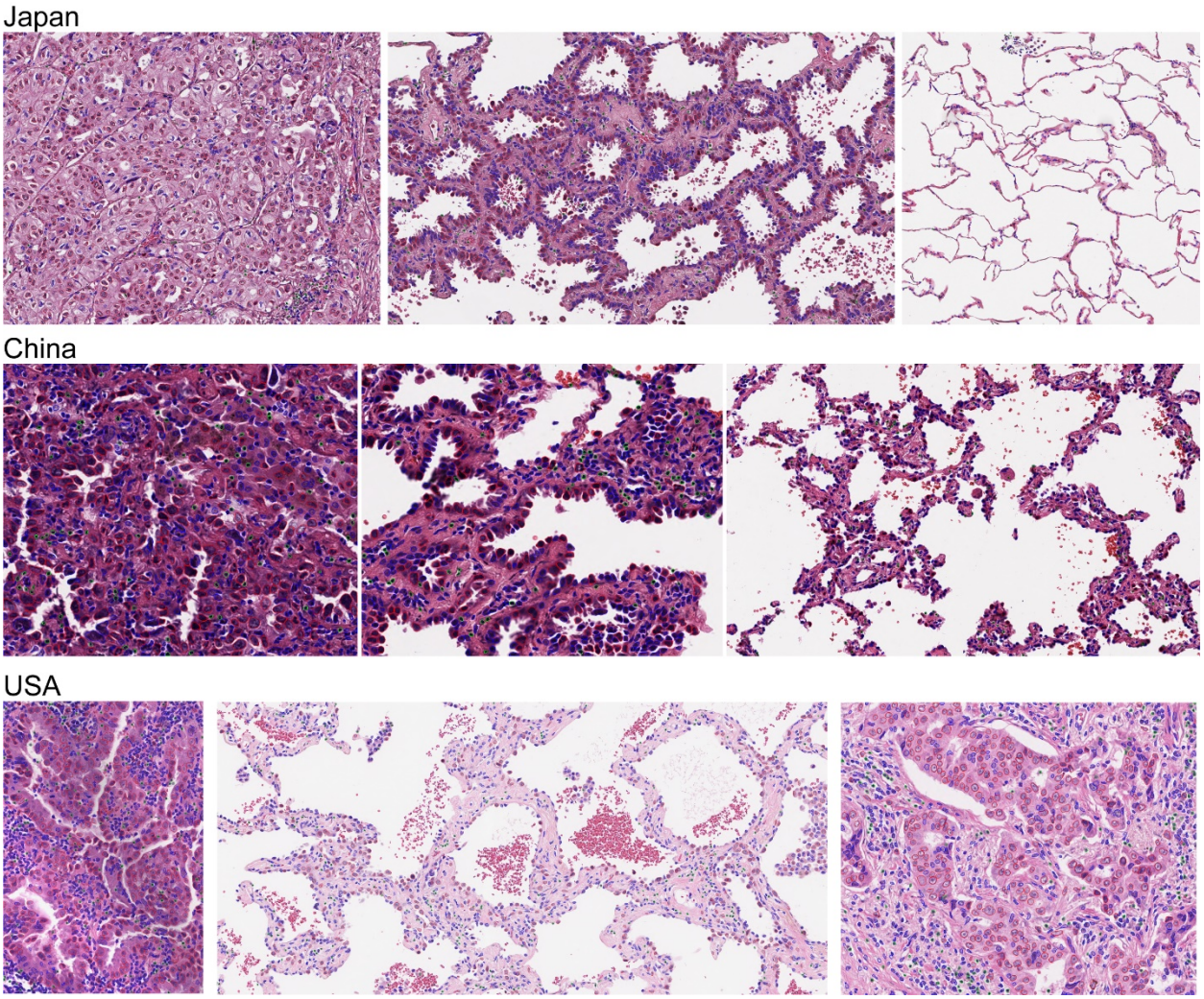
Slides No.	Normal-ROIs	AAH-ROIs	AIS-ROIs	MIA-ROIs	ADC-ROIs
Normal (9)	45	0	0	0	0
AAH (42)	77	125	0	0	0
AIS (28)	21	0	94	0	0
MIA (29)	0	0	0	98	0
ADC (54)	0	4	0	0	205
Total (162)	143	129	94	98	205

Supp Table 3. Extracted nine pathomic features and their descriptions.

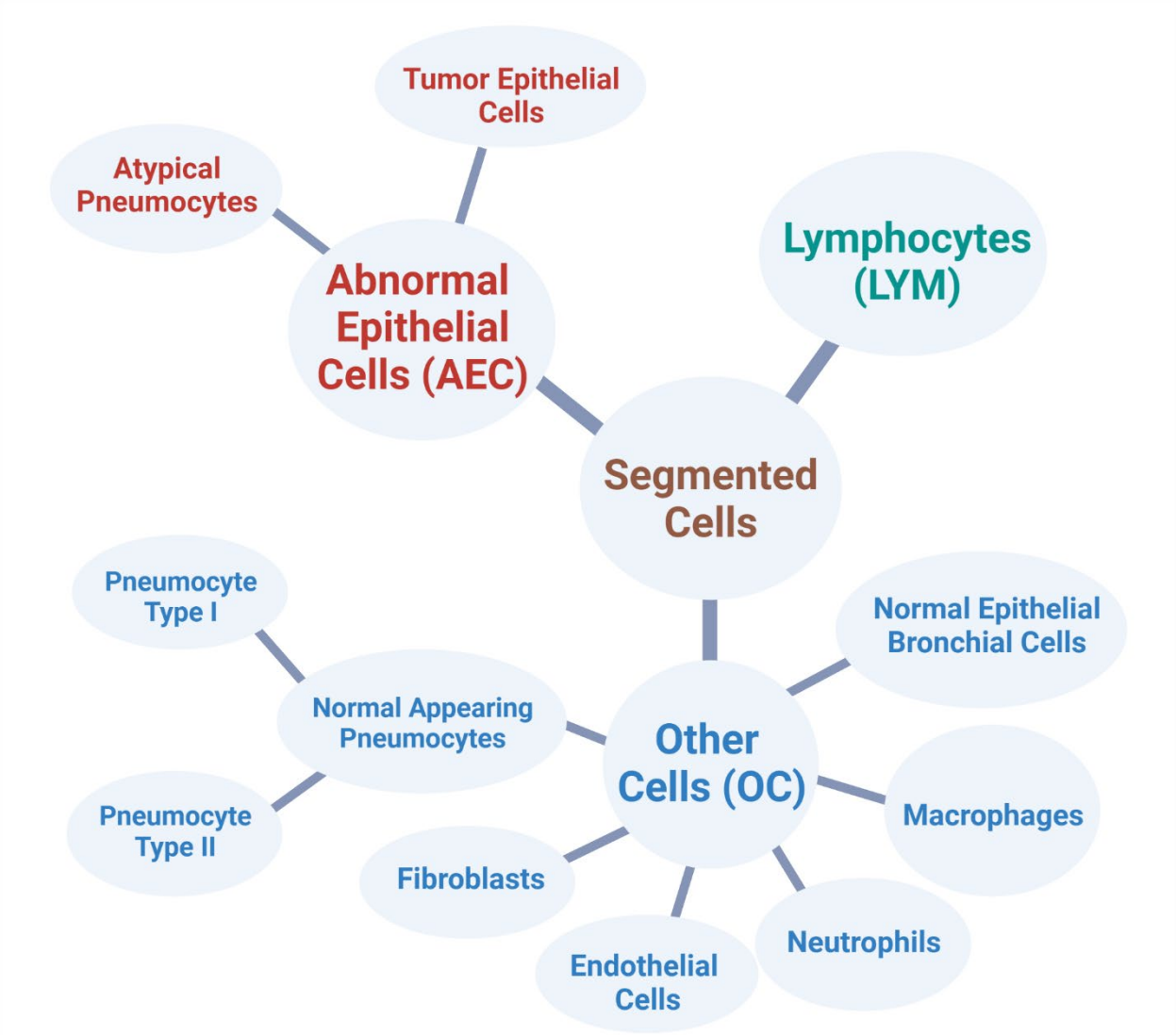
Pathomic Features	**Description**
AEC-Proportion	AEC cell number divided by the total number of cells within an ROI
LYM-Proportion	LYM cell number divided by the total number of cells within an ROI
AEC-Density	AEC cell number divided by the ROI area (μm^2)
LYM-Density	LYM cell number divided by the ROI area (μm^2)
Altieri-Entropy	spatial mutual information and spatial residual entropy among all 3 cell subtypes within an ROI
AEC-Contrast	the contrast property of ROI embedded AEC-map's GLCM
AEC-Energy	the energy property of ROI embedded AEC-map's GLCM
LYM-Contrast	the contrast property of ROI embedded LYM-map's GLCM
LYM-Energy	the energy property of ROI embedded LYM-map's GLCM

Supplementary Figures

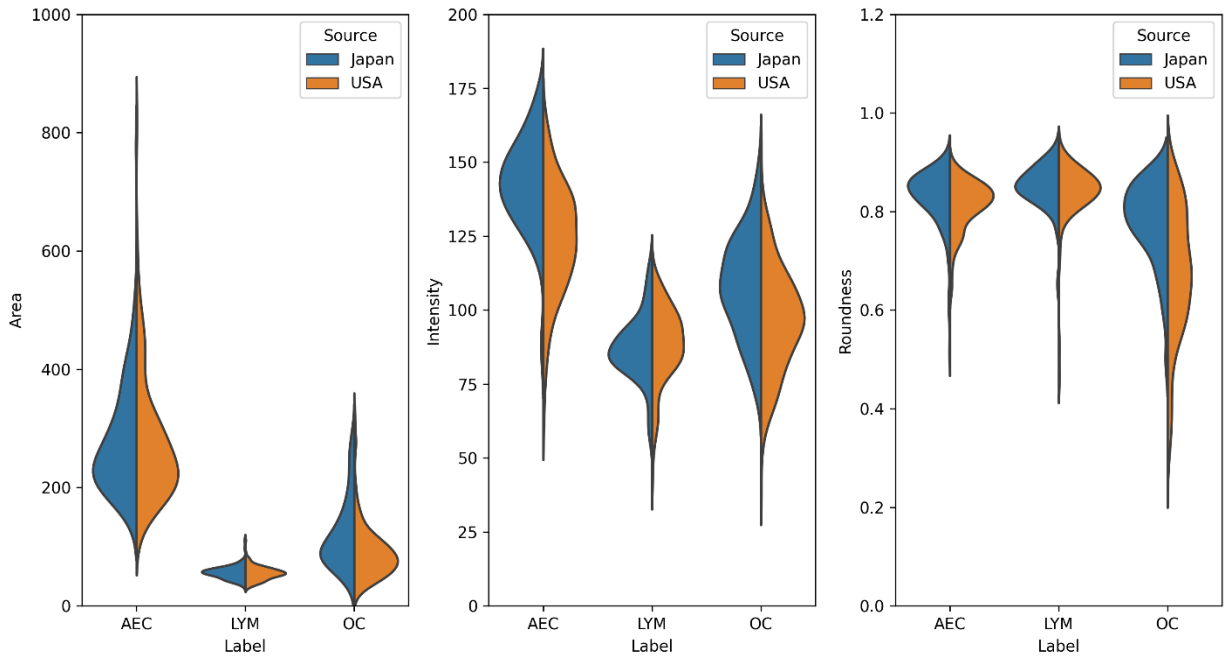




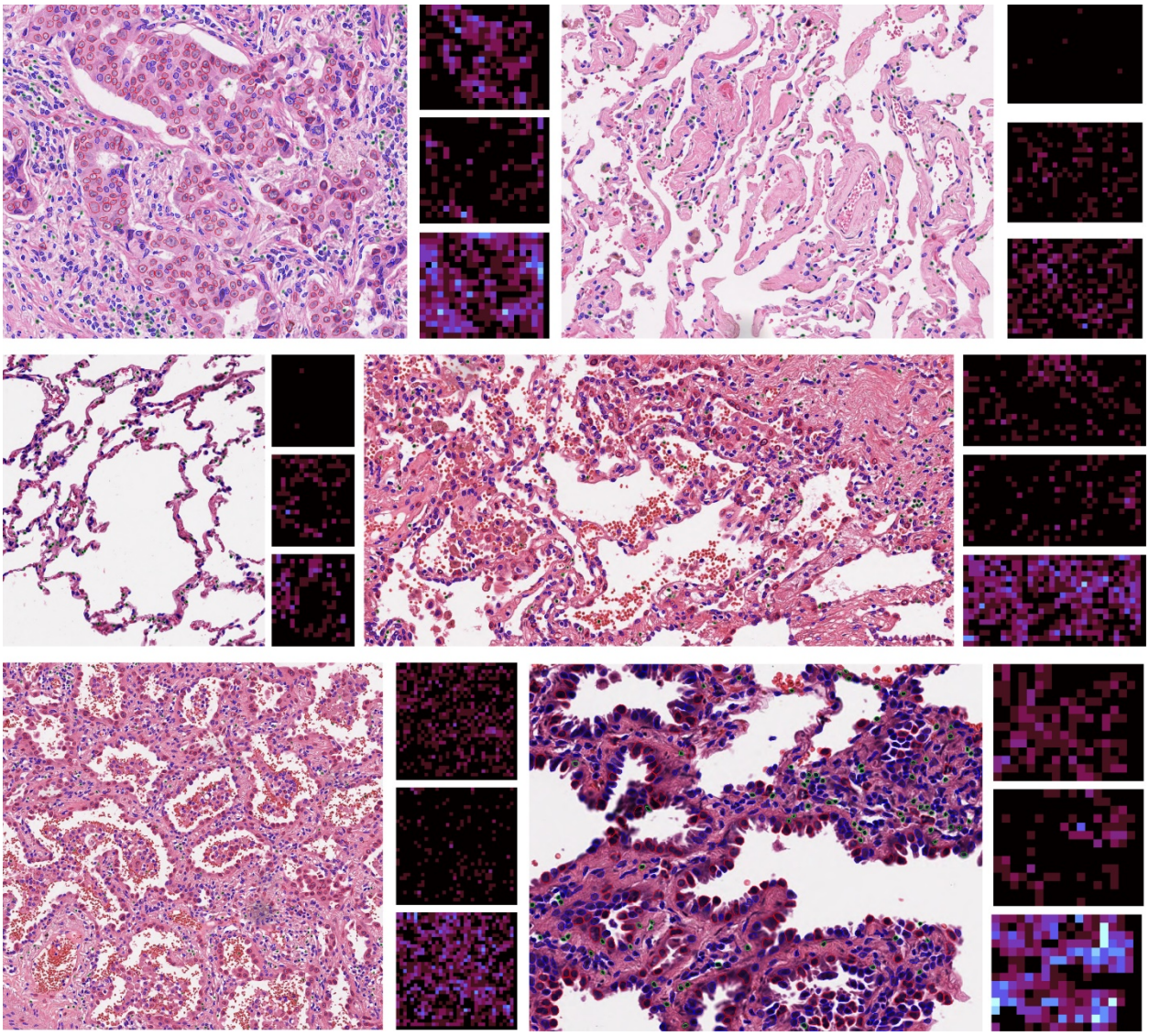
Supp Figure 2. Exemplary HoVer-Net segmented nuclei overlaying regions of interest (ROI). ROIs are randomly selected from three cohorts (Japan, China, and USA), and the nuclei contour are overlaid.



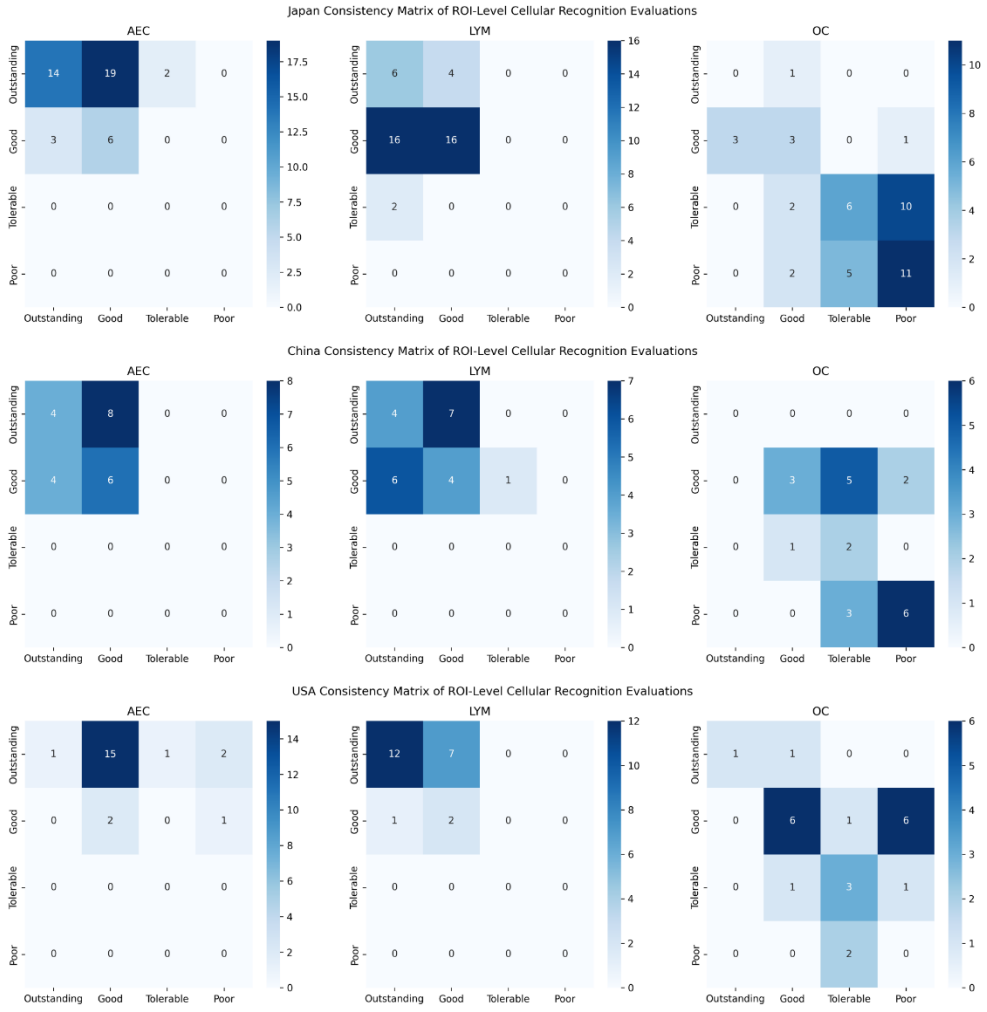
Supp Figure 3. A tree-structured taxonomy of cell subtypes. The segmented cells are split into three categories, including abnormal epithelial cell (AEC), lymphocyte (LYM), and other cell (OC).



Supp Figure 4. Comparison of cellular properties among cell subtypes annotated on two cohorts (Japan and USA). Cellular properties, cell area, mean intensity, and cell roundness are displayed from left to right. The cellular properties between the two datasets are generally consistent. Meanwhile, there are pronounced differences among three cell subtypes on three compared properties.



Supp Figure 5. Exemplary region of interest (ROI) embedded maps. Cells inside one ROI are classified into three types, and then three embedded maps (AEC-Map, LYM-Map, and OC-Map) are generated accordingly. For every group of four images, the left is the ROI, the top-right is its AEC-Map, the middle right is its LYM-Map, and the bottom right is its OC-Map. The embedded maps are grayscale images and are color-coded via colormap MAGMA.



Supp Figure 6. The consistency matrices of two pathologists' visual estimation of ROI-level cell recognition precision. From top to bottom, cover the evaluations on ROIs randomly selected from Japan, China, and USA cohorts. From left to right, cover three cell subtypes, including abnormal epithelial cell (AEC), lymphocyte (LYM), and other cell (OC). The majority of inconsistencies occur in neighboring recognition categories. Most AEC and LYM are recognized as "Outstanding" and "Good." In contrast, OC appears more often under the categories of "Good," "Tolerable," and "Poor."