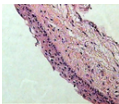


(A) Single-scale VQA

Question:

Based on the cellular morphology, which layer of stratified squamous epithelium is most prevalent on the left side of the image?



- (A) Spinous layer (C) Granular layer
(B) Basal layer (D) Keratin layer

Ground truth: (A) **Source:** PathMMU

(B) Cross-Scale VQA with Text-Only Shortcut Solution

Question: Which explanation best connects the low-magnification pattern and the high-magnification atypical nuclear findings in this case?

- (A) Crowded epithelial cell populations with irregular nuclei reduce intercellular spacing, producing architectural distortion at low magnification
(B) Stromal collagen deposition compresses otherwise orderly epithelial cells, altering the low-magnification outline
(C) Superficial epithelial flattening leads to apparent irregularity without corresponding nuclear changes
(D) Muscular layer thickening alters surface epithelial alignment despite preserved nuclear morphology

Ground truth: (A)

Text-only infer  Gemini 3 Pro

"The correct option is A. I can answer this without the images because pathology follows a strict cause-and-effect logic: "distorted" or "thickened" pattern at low magnification must be physically caused by "crowded" and "atypical" cells at high magnification. Option A is the only choice that provides this biological link, [...]"

(C) Our PathScale-VQA: Shortcut-Resistant Cross-Scale Pathological VQA

Cross-scale Semantic Reasoning

Task:  Localization

Question: Which option best characterizes the compartmental context of the structures shown in this set of images?

Options:

- (A) Cells are predominantly within peribronchiolar connective tissue, with partial collapse of adjacent alveolar spaces
(B) Cells are predominantly within alveolar airspaces (intra-alveolar), with partial collapse of adjacent alveolar spaces
(C) Cells are predominantly within the alveolar septal interstitium, with partial collapse of adjacent alveolar spaces
(D) Cells are predominantly within perivascular connective tissue, with partial collapse of adjacent alveolar spaces

Ground truth: (C)



Low-mag



Mid-mag



High-mag

Text-only infer 

"This question can't be answered correctly without the images because all four options describe plausible, [...]. Without seeing the exact coordinates of the cells, specifically whether they are hugging the airways (peribronchiolar) [...] there is no biological "rule" to favor one location over the others."

Cross-scale Visual Grounding

Task:  Origin Grounding

Question: The image above shows a 40× magnified patch extracted from the 10× patch below. Each option shows the same 10× patch with a different masked region. Which masked region corresponds to the location of the 40× patch?

Options:

(A)



Masked Low-mag 1

(B)



Masked Low-mag 2

(C)



Masked Low-mag 3

(D)



Masked Low-mag 4

Ground truth
(D)

Mid-mag



Text-only infer 

"This question cannot be answered without the images because it requires a direct visual comparison of textures, structures, and landmarks between the two magnification levels. Without the visual data, there is no text or coordinate information available to deduce the correct answer. Please upload the images!"