

# Parameter-efficient Prompt Tuning and Hierarchical Textual Guidance for Few-shot Whole Slide Image Classification

## Supplementary Material

### 1. Experiments

Our code is available at <https://github.com/Jayanie/HIPSS>.

#### 1.1. Datasets

We evaluate our method on three publicly available computational pathology datasets containing Whole Slide Images (WSIs) and slide-level labels: Camelyon16 [2], TCGA-Lung<sup>1</sup>, UBC-OCEAN [1], spanning breast, lung, and ovarian cancer types respectively. Camelyon16 dataset consists of 399 WSIs, including 239 normal slides and 160 slides with breast cancer metastases. The test set includes 129 WSIs, while the remaining 270 WSIs are used for training and validation. TCGA-Lung dataset comprises 1,042 WSIs from two lung cancer subtypes: 530 lung adenocarcinoma (LUAD) and 512 lung squamous cell carcinoma (LUSC). The test set includes 214 WSIs, with the remaining 828 WSIs used for training and validation. UBC-OCEAN dataset contains 538 WSIs across five ovarian cancer subtypes: 99 clear cell carcinoma (CC), 124 endometrioid carcinoma (EC), 222 high-grade serous carcinoma (HGSC), 47 low-grade serous carcinoma (LGSC), and 46 mucinous carcinoma (MC). The test set includes 108 WSIs, and the remaining 322 WSIs are used for training and validation. We constructed 3 different training and validation folds, while keeping the test sets fixed and independent throughout all experiments for each dataset.

#### 1.2. Pathology Query Prompt Formulation

We leveraged a large language model (LLM), GPT-4o [3] to generate detailed textual descriptions of the WSIs. We formulated two structured queries to use with the LLM considering the hierarchical structure of the WSIs. The first query was designed to obtain pathological and visual characteristics of the entire WSIs in high resolution. The second query was devised to obtain descriptive visual characteristics of  $4096 \times 4096$  regions in WSIs which consists of  $256 \times 256$  patches. This enables the textual descriptions to explicitly focus on different resolutions of the WSIs belonging to the same class. Given that the Camelyon16 and TCGA-Lung datasets consists of 2 classes, we construct 4 distinct query sets of corresponding LLM-generated descriptions for each dataset. On the other hand, UBC-OCEAN dataset comprises 5 classes, for which we formulated 10 query and LLM-generated description pairs to account for the hierar-

chical structure of WSIs. Supplementary Fig. 1 and Fig. 2 include the queries and LLM-generated descriptions for Camelyon16 and TCGA-Lung datasets respectively. The figures Fig. 3, Fig. 4, and Fig. 5 presents the queries and LLM-generated descriptions for UBC-OCEAN dataset.

### 2. Results

#### 2.1. Ablation Studies

Supplementary Table 1 presents the impact of the components in HIPSS, considering the hierarchical textual guidance based WSI representation learning module and SSF based prompt learning approach. The experimental results exhibit a clear performance progression as each component is introduced, validating the individual and combined contributions of the proposed modules. Summarized results of the ablation studies is included in Table 2 in main paper.

#### 2.2. Impact of varying tuning depth of text encoder with SSF

We investigate how the number of text encoder blocks which the SSF is applied influences the model performance. Supplementary Table 2 reports the mean AUC values across Camelyon16 and UBC-OCEAN datasets as the number of tuned text encoder blocks  $d_s$  varies, where  $0 < d_s \leq L$  and  $L = 12$ . These results are illustrated in Fig. 2 in the main paper.

#### 2.3. Selection of the types of layers to tune with SSF in text encoder

From the distinct types of layers in a text encoder block including layer normalization (LN), multi-head attention (Attn), and multi-layer perceptron (MLP), we selected to tune LN and MLP layers with the Scaling and Shifting Features (SSF). This selection was based on preliminary empirical results that we obtained, considering the computational efficiency. Supplementary Table 3 shows the total number of trainable parameters when we select LN + MLP vs. LN + MLP + Attn. LN + MLP + Attn introduces 4.3% and 14.1% additional parameters compared to only tuning LN + MLP with SSF across Camelyon16/ TCGA-Lung and UBC-OCEAN datasets respectively. Therefore, we opted only to tune layer normalization and multi-layer perceptron layers in the text encoder blocks using SSF to maintain the parameter efficiency of our method.

<sup>1</sup><https://portal.gdc.cancer.gov>

## 2.4. Impact of varying attention weight refinement factor values in hierarchical textual guidance

We propose to modify the attention weights in WSI representation learning approach using a text-guided attention weights refinement factor ( $\lambda$ ). Supplementary Table 5 demonstrates the how the performance varies based on the attention weight refinement factor values in each dataset. Fig. 3(a) in the main paper shows the mean AUC averaged across the three datasets.

## 2.5. Impact of varying threshold values in hierarchical textual guidance

Supplementary Table 6 presents the how the performance of each dataset dataset is affected by the attention weight threshold value ( $\alpha$ ) in hierarchical textual guidance based WSI representation learning module. Fig. 3(b) in the main paper shows the mean AUC averaged across the three datasets.

## 2.6. Impact of text-guided attention refinement in a hierarchical manner

Text-guided refinement of attention weights can be applied in the region encoder, the WSI encoder, or jointly across both stages in HIPSS method. As shown in Supplementary Table 7, applying it at both stages yields the best performance.

## 2.7. Impact of varying region sizes in WSI representation learning

In our method each WSI was segmented into non-overlapping regions of size  $4096 \times 4096$ , which were later partitioned into  $256 \times 256$  instances. Since the spatial extent of tumor regions is unknown during training, the optimal region size should be empirically determined. We observed that using either excessively large or overly small region sizes leads to a drop in the performance as demonstrated in Supplementary Table 4.

## 2.8. Comparison with hard instance filtering

Supplementary Table 8 shows the performance gains over existing hard instance filtering approach, highlighting the effectiveness of soft textual guidance in WSI representation learning. Summarized results of this comparison is included in Table 3 in main paper.

## References

- [1] Ali Bashashati, Hossein Farahani, OTTA Consortium, Anthony Karnezis, Ardalan Akbari, Sirim Kim, Ashley Chow, Sohier Dane, Allen Zhang, and Maryam Asadi. Ubc ovarian cancer subtype classification and outlier detection (ubc-ocean). <https://kaggle.com/competitions/UBC-OCEAN>, 2023. Kaggle. 1
- [2] Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes Van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, Meyke Hermsen, Quirine F Manson, Maschenka Balkenhol, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22): 2199–2210, 2017. 1
- [3] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. 1

Table 1. Ablation Studies on 3 datasets. Mean AUC values are reported. Components are denoted as H: hierarchical WSI representation learning, T: text-guided attention refinement strategy, and S: SSF-based prompt learning in text encoder.

Dataset	H	T	S	16-shot	8-shot	4-shot	2-shot	1-shot
Camelyon16	×	×	×	0.7820 $\pm$ 0.0455	0.7195 $\pm$ 0.0314	0.6086 $\pm$ 0.0587	0.5969 $\pm$ 0.0431	0.5706 $\pm$ 0.0426
	✓	×	×	0.8666 $\pm$ 0.0085	0.7516 $\pm$ 0.0143	0.6903 $\pm$ 0.0170	0.6758 $\pm$ 0.0099	0.6433 $\pm$ 0.0169
	✓	✓	×	0.8724 $\pm$ 0.0086	0.8365 $\pm$ 0.0167	0.7552 $\pm$ 0.0173	0.6867 $\pm$ 0.0063	0.6512 $\pm$ 0.0089
	×	×	✓	0.8853 $\pm$ 0.0267	0.8166 $\pm$ 0.0056	0.7798 $\pm$ 0.0033	0.7303 $\pm$ 0.0092	0.7071 $\pm$ 0.0070
	✓	×	✓	0.9006 $\pm$ 0.0116	0.8429 $\pm$ 0.0065	0.8034 $\pm$ 0.0139	0.7738 $\pm$ 0.0086	0.7217 $\pm$ 0.0081
	✓	✓	✓	<b>0.9105</b> $\pm$ 0.0104	<b>0.8586</b> $\pm$ 0.0067	<b>0.8159</b> $\pm$ 0.0165	<b>0.8066</b> $\pm$ 0.0069	<b>0.7429</b> $\pm$ 0.0087
TCGA-Lung	×	×	×	0.8074 $\pm$ 0.0166	0.7709 $\pm$ 0.0118	0.6832 $\pm$ 0.0281	0.6541 $\pm$ 0.0166	0.5965 $\pm$ 0.0297
	✓	×	×	0.8487 $\pm$ 0.0305	0.7876 $\pm$ 0.0066	0.7420 $\pm$ 0.0148	0.6971 $\pm$ 0.0122	0.6202 $\pm$ 0.0197
	✓	✓	×	0.8661 $\pm$ 0.0154	0.8183 $\pm$ 0.0111	0.7451 $\pm$ 0.0093	0.7134 $\pm$ 0.0133	0.6311 $\pm$ 0.0123
	×	×	✓	0.8781 $\pm$ 0.0469	0.7952 $\pm$ 0.0072	0.7757 $\pm$ 0.0098	0.7387 $\pm$ 0.0117	0.6801 $\pm$ 0.0197
	✓	×	✓	0.9072 $\pm$ 0.0051	0.8296 $\pm$ 0.0046	0.7842 $\pm$ 0.0029	0.7508 $\pm$ 0.0061	0.7203 $\pm$ 0.0034
	✓	✓	✓	<b>0.9097</b> $\pm$ 0.0155	<b>0.8745</b> $\pm$ 0.0312	<b>0.8285</b> $\pm$ 0.0205	<b>0.7616</b> $\pm$ 0.0181	<b>0.7217</b> $\pm$ 0.0141
UBC-OCEAN	×	×	×	0.8432 $\pm$ 0.0132	0.8425 $\pm$ 0.0068	0.7857 $\pm$ 0.0107	0.7037 $\pm$ 0.0143	0.6244 $\pm$ 0.0237
	✓	×	×	0.9165 $\pm$ 0.0064	0.8564 $\pm$ 0.0075	0.8095 $\pm$ 0.0068	0.7429 $\pm$ 0.0108	0.6885 $\pm$ 0.0089
	✓	✓	×	0.9347 $\pm$ 0.0074	0.8735 $\pm$ 0.0049	0.8233 $\pm$ 0.0101	0.7745 $\pm$ 0.0088	0.6983 $\pm$ 0.0132
	×	×	✓	0.9467 $\pm$ 0.0036	0.8881 $\pm$ 0.0019	0.8330 $\pm$ 0.0032	0.8563 $\pm$ 0.0023	0.8278 $\pm$ 0.0054
	✓	×	✓	0.9517 $\pm$ 0.0023	0.9498 $\pm$ 0.0022	0.9320 $\pm$ 0.0056	0.8734 $\pm$ 0.0030	0.8498 $\pm$ 0.0019
	✓	✓	✓	<b>0.9540</b> $\pm$ 0.0138	<b>0.9516</b> $\pm$ 0.0121	<b>0.9305</b> $\pm$ 0.0131	<b>0.8797</b> $\pm$ 0.0152	<b>0.8542</b> $\pm$ 0.0146

Table 2. Variation of AUC values based on number of tuned blocks ( $d_s$ ) in the text encoder with SSF. Datasets are denoted as C16: Camelyon16, UBC: UBC-OCEAN. Bold: highest results.

Dataset	$k$ -shot	Number of Tuned Blocks in Text Encoder with SSF ( $d_s$ )											
		1	2	3	4	5	6	7	8	9	10	11	12
C16	16-shot	0.8836	<b>0.9105</b>	0.8913	0.8848	0.8899	0.8671	0.8647	0.8631	0.8533	0.8482	0.8418	0.8279
	8-shot	0.8388	<b>0.8586</b>	0.8407	0.8379	0.8368	0.8369	0.8394	0.8309	0.8257	0.8275	0.8264	0.8222
	4-shot	0.8021	<b>0.8159</b>	0.7861	0.7822	0.7816	0.7686	0.7516	0.7501	0.7299	0.7235	0.7072	0.7012
	2-shot	0.7667	<b>0.8066</b>	0.7610	0.6840	0.6617	0.6272	0.6186	0.6136	0.6087	0.6114	0.6019	0.5998
	1-shot	0.7123	<b>0.7429</b>	0.6891	0.6527	0.6470	0.6099	0.5562	0.5761	0.5510	0.5505	0.5335	0.5366
UBC	16-shot	0.9425	0.9439	0.9412	0.9472	0.9491	0.9469	0.9480	<b>0.9540</b>	0.9495	0.9494	0.9435	0.9441
	8-shot	0.9387	0.9425	0.9406	0.9479	0.9486	0.9470	0.9451	<b>0.9516</b>	0.9447	0.9415	0.9433	0.9415
	4-shot	0.9155	0.9127	0.9148	0.9194	0.9199	0.9177	0.9271	<b>0.9305</b>	0.9261	0.9102	0.9149	0.9038
	2-shot	0.8154	0.8415	0.8504	0.8601	0.8652	0.8698	0.8702	<b>0.8797</b>	0.8716	0.8717	0.8707	0.8653
	1-shot	0.8003	0.8234	0.8309	0.8450	0.8486	0.8447	0.8497	<b>0.8542</b>	0.8478	0.8459	0.8441	0.8341

Table 3. Computational efficiency comparison between the two variations considering the types of text encoder layers which SSF is applied. Total number of trainable parameters (in millions) is reported. The increment in trainable parameters compared to the final HIPSS method is indicated by  $\uparrow$ . Types of layers are denoted as LN: layer normalization, MLP: multi-layer perceptron, and Attn: multi-head attention.

Types of Layers	Dataset	
	Camelyon16/TCGA-Lung	UBC-OCEAN
LN + MLP	<b>0.2844</b>	<b>0.3489</b>
LN + MLP + Attn	0.2967 ( $\uparrow$ 4.3%)	0.3981 ( $\uparrow$ 14.1%)

Table 4. Variation of AUC values based on different region sizes in WSI representation learning module on Camelyon16 dataset. Region size is increased from  $512 \times 512$  to considering the entire WSI as the region.

Region Size	2-shot	1-shot
$512 \times 512$	0.7423 $\pm$ 0.0089	0.7099 $\pm$ 0.0148
$2048 \times 2048$	0.8019 $\pm$ 0.0163	0.7414 $\pm$ 0.0152
$4096 \times 4096$	<b>0.8066</b> $\pm$ 0.0069	<b>0.7429</b> $\pm$ 0.0087
$8192 \times 8192$	0.7677 $\pm$ 0.0081	0.7221 $\pm$ 0.0108
WSI	0.7445 $\pm$ 0.0101	0.7128 $\pm$ 0.0086

Table 5. Variation of AUC values based on the attention weights refinement factor ( $\lambda$ ) in the soft hierarchical textual guidance strategy.

Dataset	$\lambda$	16-shot	8-shot	4-shot	2-shot	1-shot
C16	2	0.8795	0.8397	0.7949	0.7783	0.7229
	5	0.8622	0.8469	0.8013	0.7835	0.7334
	10	0.9105	0.8586	0.8159	0.8066	0.7429
	20	0.8672	0.8454	0.7986	0.7757	0.7248
TCGA	2	0.8543	0.7814	0.7508	0.5821	0.5611
	5	0.8561	0.7891	0.7257	0.6621	0.6567
	10	0.9097	0.8745	0.8285	0.7616	0.7217
	20	0.8563	0.8174	0.7661	0.6851	0.5993
UBC	2	0.9347	0.9249	0.8939	0.8482	0.8260
	5	0.9426	0.9289	0.8941	0.8536	0.8224
	10	0.9540	0.9516	0.9305	0.8797	0.8542
	20	0.9288	0.9263	0.8247	0.7943	0.7425

Table 6. Variation of AUC values based on the threshold values ( $\alpha$ ) in the soft hierarchical textual guidance strategy.

Dataset	$\alpha$	16-shot	8-shot	4-shot	2-shot	1-shot
C16	0.1	0.8732	0.8385	0.7899	0.7793	0.7209
	0.2	0.9105	0.8586	0.8159	0.8066	0.7429
	0.5	0.8465	0.8371	0.7881	0.7744	0.7192
	0.8	0.8547	0.8375	0.7925	0.7779	0.7208
TCGA	0.1	0.8501	0.8427	0.7384	0.6062	0.5663
	0.2	0.9097	0.8745	0.8285	0.7616	0.7217
	0.5	0.8519	0.7682	0.7512	0.5577	0.5643
	0.8	0.8519	0.7684	0.7516	0.5477	0.5641
UBC	0.1	0.9384	0.9257	0.9130	0.8668	0.8336
	0.2	0.9540	0.9516	0.9305	0.8797	0.8542
	0.5	0.9382	0.9364	0.8935	0.8472	0.8259
	0.8	0.9381	0.9361	0.8932	0.8468	0.8256

Table 7. Performance comparison with text-guided attention weights refinement applied to region encoder, WSI encoder or both stages on Camelyon16 dataset. Mean AUC values are reported.

Method	16-shot	8-shot	4-shot	2-shot	1-shot
Only Region Encoder	0.8149 $\pm$ 0.0299	0.7992 $\pm$ 0.0092	0.7748 $\pm$ 0.0151	0.6947 $\pm$ 0.0083	0.6891 $\pm$ 0.0066
Only WSI Encoder	0.8615 $\pm$ 0.0395	0.8352 $\pm$ 0.0042	0.7983 $\pm$ 0.0058	0.7384 $\pm$ 0.0049	0.7247 $\pm$ 0.0048
Both (Ours)	<b>0.9105<math>\pm</math>0.0104</b>	<b>0.8586<math>\pm</math>0.0067</b>	<b>0.8159<math>\pm</math>0.0165</b>	<b>0.8066<math>\pm</math>0.0069</b>	<b>0.7429<math>\pm</math>0.0087</b>

Table 8. Performance comparison with hard instance filtering on UBC-OCEAN dataset. Mean AUC values are reported.

Method	16-shot	8-shot	4-shot	2-shot	1-shot
Hard instance filtering	0.9341 $\pm$ 0.0042	0.9298 $\pm$ 0.0035	0.8756 $\pm$ 0.0067	0.8588 $\pm$ 0.0054	0.8399 $\pm$ 0.0052
HIPSS (Ours)	<b>0.9540<math>\pm</math>0.0138</b>	<b>0.9516<math>\pm</math>0.0121</b>	<b>0.9305<math>\pm</math>0.0131</b>	<b>0.8797<math>\pm</math>0.0152</b>	<b>0.8542<math>\pm</math>0.0146</b>

<p style="text-align: center;"><b>Normal Lymph Node WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of normal lymph node in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Uniform small lymphocytes densely packed. Well-formed follicles in cortex, lymphocyte cords in medulla. Thin-walled blood vessels throughout. No atypical cells or architectural distortions.</p>	<p style="text-align: center;"><b>Normal Lymph Node WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of normal lymph node in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Normal lymph node regions show well-organized tissue with dense, uniform lymphocytes featuring round nuclei and minimal cytoplasm, indicative of active immune function. Depending on the location, well-formed follicles with germinal centers, interfollicular areas, or medullary cords may be present. The architecture remains intact with minimal pleomorphism and no signs of infiltration or disruption. Supporting structures like thin-walled blood vessels and reticular fibers help maintain node integrity. These features collectively reflect a benign, reactive immune environment typical of healthy lymphoid tissue in whole slide images.</p>
<p style="text-align: center;"><b>Metastatic Lymph Node WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of metastatic lymph node in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Large, pleomorphic cancer cells interspersed in lymphoid tissue. Cells show irregular nuclei, prominent nucleoli, abundant cytoplasm. Atypical cell arrangements. Desmoplasia, increased mitoses, potential necrosis. Abnormal blood vessels present.</p>	<p style="text-align: center;"><b>Metastatic Lymph Node WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of metastatic lymph node in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> In metastatic lymph node regions, malignant epithelial cells appear densely packed with pleomorphism, prominent nucleoli, and frequent mitoses. Tumor cells form disorganized nests or gland-like structures infiltrating the stroma, often showing desmoplastic reaction. Tumor budding—small clusters or single cells detaching from the main mass—is common and indicates invasiveness. Lymphovascular invasion may be present, with tumor cells in or near vessels. The microenvironment is heterogeneous, featuring immune cells and stromal components, reflecting active tumor-stroma interactions that support metastatic progression.</p>

Figure 1. Queries and LLM-Generated Descriptions for Camelyon16 Dataset.

<p style="text-align: center;"><b>LUAD WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of lung adenocarcinoma (LUAD) in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Clear cytoplasm, round or oval nuclei, prominent nucleoli, rich vascularity, irregular blood vessels, intratumoral septa, and heterogeneity.</p>	<p style="text-align: center;"><b>LUAD WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of lung adenocarcinoma (LUAD) in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Lung adenocarcinoma regions, tumor cells form irregular glandular or acinar structures with crowded, hyperchromatic nuclei and prominent nucleoli. Cytoplasm is often pale or vacuolated, with evidence of mucin production in some areas. Patches may show single-layered or pseudostratified cells, cellular atypia, and frequent mitoses. Some regions display stromal desmoplasia, inflammatory infiltrates, or disrupted alveolar spaces. Variability across patches reflects the heterogeneous architecture and invasive nature of adenocarcinoma at the microscopic level.</p>
<p style="text-align: center;"><b>LUSC WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of lung squamous cell carcinoma (LUSC) in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Squamous cell differentiation, round structures with eosinophilic cytoplasm, distinct cell borders and abundant cytoplasm, enlarged nuclei, irregular nuclear shape, increased chromatin density.</p>	<p style="text-align: center;"><b>LUSC WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of lung squamous cell carcinoma (LUSC) in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Lung squamous cell carcinoma regions, individual tumor cells show dense eosinophilic cytoplasm, enlarged hyperchromatic nuclei, and prominent nucleoli. Keratinization, including partial keratin pearls, and intercellular bridges may appear in select patches. Cellular pleomorphism and frequent mitotic figures are common. Some patches reveal necrotic debris or inflammatory cells, while others show tightly packed tumor nests or isolated cells invading stroma. Overall, patches reflect heterogeneity in squamous differentiation and tumor aggression at the cellular scale.</p>

Figure 2. Queries and LLM-Generated Descriptions for TCGA-Lung Dataset.

<p style="text-align: center;"><b>CC WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of clear cell carcinoma (CC) in a whole slide image</p>  <p><b>LLM-Generated Description:</b> A whole slide image of clear-cell ovarian carcinoma typically reveals tumor regions composed of large, polygonal cells with abundant clear or glycogen-rich cytoplasm, giving a distinct glassy appearance. Nuclei are often atypical, centrally located, and show prominent nucleoli. The tumor architecture may display tubulocystic, papillary, or solid patterns, often surrounded by fibrous or hyalinized stroma. Hobnail cells—epithelial cells with apical nuclei protruding into glandular lumina—are frequently present. Hemorrhage, necrosis, and inflammatory infiltrates can also be observed. The clear cytoplasm sharply contrasts with the dense hematoxylin-stained nuclei, making the tumor areas visually distinct from adjacent benign tissue on hematoxylin and eosin-stained slides.</p>	<p style="text-align: center;"><b>CC WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of clear cell carcinoma (CC) in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Clear-cell ovarian carcinoma regions often observe a mosaic of heterogeneous tumor architecture. Many patches feature large, polygonal cells with clear cytoplasm and prominent central nuclei, giving a translucent appearance. Hobnail cells protrude into gland-like spaces, and tubulocystic or papillary structures are intermittently present. Dense fibrous stroma and hemorrhagic or necrotic areas contrast sharply with tumor regions. Some patches may include inflammatory infiltrates or blood vessels. The clear cells' cytoplasm contrasts with hematoxylin-stained nuclei, creating a distinctive visual pattern across neighboring patches.</p>
<p style="text-align: center;"><b>EC WSI Description - Entire WSI</b></p> <p><b>Query:</b> Describe the visual characteristics of endometrioid carcinoma (EC) in a whole slide image</p>  <p><b>LLM-Generated Description:</b> A whole slide image of endometrioid carcinoma reveals glandular structures resembling endometrial tissue, often arranged in back-to-back or cribriform patterns. The tumor cells are columnar to cuboidal with eosinophilic cytoplasm and nuclei showing varying degrees of atypia, from mild to moderate, depending on grade. Nucleoli are typically visible, and mitotic figures may be present. Squamous differentiation and mucinous areas are sometimes observed. The architecture is generally more organized than in high-grade tumors, with well-formed glands dominating the landscape. On hematoxylin and eosin staining, the glands are outlined in pink cytoplasm, with darker-stained nuclei providing contrast. Necrosis is uncommon, and the surrounding stroma may show inflammation or fibrosis.</p>	<p style="text-align: center;"><b>EC WSI Description - Region Level</b></p> <p><b>Query:</b> Describe the visual characteristics of endometrioid carcinoma (EC) in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>  <p><b>LLM-Generated Description:</b> Endometrioid carcinoma regions, many patches display well-formed glandular structures arranged in back-to-back or cribriform patterns. Tumor cells appear columnar or cuboidal with eosinophilic cytoplasm and round to oval nuclei showing mild to moderate atypia. Nucleoli are often visible, and mitotic figures may appear in scattered patches. Some patches may show squamous differentiation or mucinous features. The surrounding stroma is typically fibrotic or inflamed, providing structural contrast. On hematoxylin and eosin staining, pink cytoplasm and dark nuclei create clear boundaries, resulting in a patterned, moderately organized appearance across adjacent patches.</p>

Figure 3. Queries and LLM-Generated Descriptions for clear cell carcinoma (CC) and endometrioid carcinoma (EC) classes in UBC-OCEAN Dataset.



Figure 4. Queries and LLM-Generated Descriptions for high-grade serous carcinoma (HGSC) and low-grade serous carcinoma (LGSC) in UBC-OCEAN Dataset.



MC WSI Description - Entire WSI	MC WSI Description - Region Level
<p><b>Query:</b> Describe the visual characteristics of mucinous carcinoma (MC) in a whole slide image</p>	<p><b>Query:</b> Describe the visual characteristics of mucinous carcinoma (MC) in 4096 x 4096 regions consists of 256 x 256 patches in a whole slide image</p>
	
<p><b>LLM-Generated Description:</b> A whole slide image of mucinous carcinoma displays tumor cells arranged in glandular, cystic, papillary structures, surrounded by abundant extracellular mucin pools. The cells are typically columnar with basally located nuclei and pale, mucin-filled cytoplasm. Nuclear atypia ranges from mild to moderate, and mitotic activity is generally low. Tumor architecture may include multilocular cystic areas or floating tumor clusters within mucin lakes. Psammoma bodies are rare, but inflammatory cells or hemorrhage may be present in the stroma. On hematoxylin and eosin staining, mucin appears as light bluish or pale areas, contrasting with the darker-stained nuclei and outlining the glandular contours, producing a bubbly, gelatinous appearance across the slide.</p>	<p><b>LLM-Generated Description:</b> Mucinous carcinoma regions, many patches reveal glandular or cystic structures embedded in large pools of extracellular mucin. Tumor cells appear columnar with pale, mucin-rich cytoplasm and basally aligned nuclei showing mild to moderate atypia. Some patches contain floating tumor clusters within mucin lakes, while others show more solid or papillary arrangements. On hematoxylin and eosin staining, mucin stains light blue to pale violet, creating low-density regions that contrast with the darker nuclei. The overall visual texture across patches is heterogeneous, with bubbly mucin-filled areas interspersed with more cellular or fibrotic zones.</p>

Figure 5. Queries and LLM-Generated Descriptions for mucinous carcinoma (MC) class in UBC-OCEAN Dataset.