

Teaching pathology foundation models to accurately predict gene expression with parameter efficient knowledge transfer

Shi Pan^{1,*}, Jianan Chen^{2,*}, and Maria Secrier¹

¹ UCL Genetics Institute, University College London, London, UK

² UCL Cancer Institute, University College London, London, UK
{shi.pan, jianan.c, m.secrier}@ucl.ac.uk
<https://secrierlab.github.io/>

Abstract. Gene expression profiling provides critical insights into cellular heterogeneity, biological processes, and disease mechanisms. There has been an increasing interest in computational approaches that can predict gene expression directly from digitalized histopathology images. While image foundation models have shown promise in a variety of pathology downstream analysis, their performances on gene expression prediction are still limited. Explicitly incorporating information from the transcriptomic models can help image models address domain shift, yet the fine-tuning and alignment of foundation models can be expensive. In this work, we propose Parameter Efficient Knowledge trAnsfer (PEKA), a novel framework that leverages Block-Affine Adaptation and integrates knowledge distillation and structure alignment losses for cross-modal knowledge transfer. We evaluated PEKA for gene expression prediction using multiple spatial transcriptomics datasets (comprising 206,123 image tiles with matched gene expression profiles) that included various types of tissue. PEKA achieved at least 5% performance improvement over baseline foundation models while also outperforming alternative parameter-efficient fine-tuning strategies. We have released the code, datasets and aligned models at Github to facilitate broader adoption and further development for parameter efficient model alignment.

Keywords: Digital Pathology · Foundation Model · Spatial Transcriptomics.

1 Introduction

Gene expression profiling provides critical insights into cellular heterogeneity, biological processes, and disease mechanisms. However, single-cell RNA sequencing (scRNA-seq) [11] and spatial transcriptomics (ST) [13] remain costly, time-consuming, and are not routinely implemented in clinical settings [4]. This has led

* These authors contributed equally

to an increasing interest in computational approaches that can predict gene expression directly from digitalized histopathological images [18], which are widely used in routine clinical workflows and often considered the gold standard for diagnostics.

Early convolutional neural networks (CNN) and transformers for predicting gene expression were trained on paired bulk RNA-seq gene expression matrix and whole-slide images [8, 15]. Recently, pathology foundation models [1, 16, 17, 22], which employ self-supervised pre-training on large quantities of whole-slide images (WSI), have been revolutionizing the field of digital pathology and demonstrated superior performance in various downstream tasks including gene expression prediction.

Despite these advances, the embeddings extracted using image foundation models predominantly capture morphological patterns, which may not optimally align with the underlying biological signals. Apart from pathology findings directly correlating with observable morphological features, for example, mitosis and apoptosis [15], the vast majority of gene expression changes occur without manifesting visible alterations in cellular morphology. This means histopathology-based foundation models may not naturally emphasize the intrinsic dimensions that can link to the transcriptomic domain. It can be considered as an example of domain shift [3], which partly explains the performance bottleneck encountered when attempting to predict gene expression from histopathology foundation model feature spaces. To overcome this limitation, we need to explicitly incorporate information from the transcriptomic domain to guide the adaptation of image features toward dimensions that are more relevant for gene expression prediction.

Single-cell transcriptomic foundation models, such as scFoundation and scGPT [2, 5], encode rich biological signals from the transcriptome and have emerged as powerful tools for modeling gene expression data. To address the modality-shift and enhance the molecular understanding, we introduce Parameter Efficient Knowledge trAnsfer (PEKA), a novel framework that leverages Block-Affine Adaptation [10] and integrates knowledge distillation and structure alignment losses for cross-modal knowledge transfer. PEKA selectively amplifies dimensions in the image feature space that are most relevant to transcriptomic information, effectively bridging morphological features with their molecular underpinnings.

We evaluated PEKA in multiple datasets (comprising 206,123 image tiles with matched gene expression profiles) that encompassed various types of tissue, including breast, liver, kidney, and lung cancer subsets from the HEST dataset [7]. Our experiments demonstrate the framework’s robustness and generalizability across different histological contexts. Notably, PEKA demonstrates consistent performance gain over various baseline models, achieving at least 5% performance improvement while also outperforming alternative parameter-efficient fine-tuning strategies in gene expression prediction tasks. We have released the code, related datasets, and aligned models at Github to facilitate broader adoption and further development for parameter efficient model alignment.

2 Methods

2.1 Problem Formulation

Given a cell image $\mathbf{X}_{img} \in \mathbb{R}^{H \times W \times C}$ and its corresponding gene expression data $\mathbf{X}_{seq} \in \mathbb{R}^G$, conventional approaches directly employ a pretrained foundation model $F(\cdot)$ for prediction:

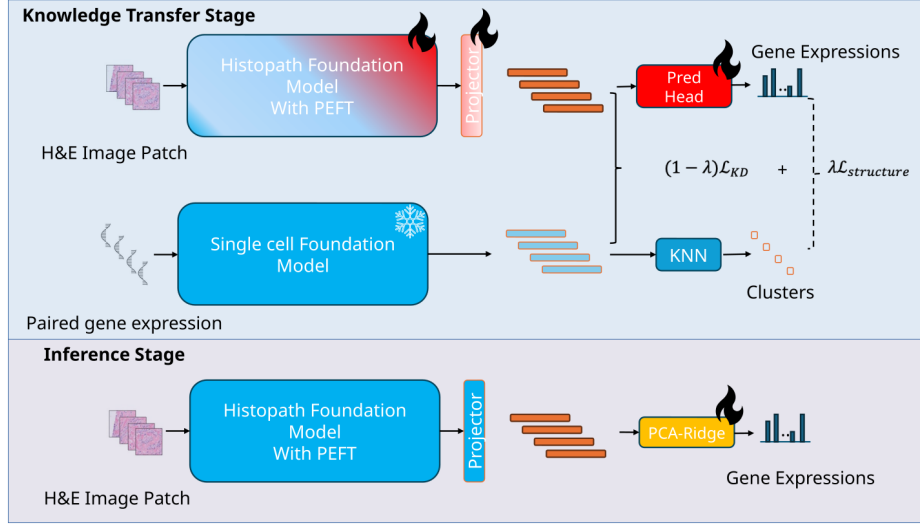


Fig. 1. PEKA framework for knowledge transfer between histopathology and transcriptomic foundation models.

$$P(\mathbf{X}_{seq}|\mathbf{X}_{img}) = h(F(\mathbf{X}_{img})) \quad (1)$$

where $h(\cdot)$ denotes a task-specific prediction head. However, this direct mapping often struggles to capture complex gene expression patterns because of the significant domain discrepancy between the pretrained representation space and the gene expression prediction task. We suggest that this limitation primarily stems from the inherent modality gap between image representations and gene expression patterns, as these models are predominantly optimized for general visual understanding and may lack understanding of molecular features.

2.2 Parameter Efficient Knowledge trAnsfer(PEKA)

To address this challenge, we propose PEKA (Parameter Efficient Knowledge trAnsfer), a plug-and-play framework that enhances the histopathology foundation model’s capability in gene expression modeling through efficient knowledge transfer and parameter adaptation (**Fig. 1**). From an information-theory

perspective, our approach bridges the modality gap by maximizing mutual information between the embedding spaces of different models. PEKA achieves this through two key components: knowledge transfer-based modality alignment and parameter-efficient fine-tuning. Specifically, PEKA pipeline consists of two stages: a Knowledge Transfer Stage where H&E image patches are processed by a histopathology foundation model enhanced with parameter-efficient fine-tuning (PEFT), while matched gene expression data is processed by a single-cell foundation model. The training objective combines direct gene expression prediction (\mathcal{L}_{KD}) with structural alignment of feature spaces (\mathcal{L}_{struct}), where cluster relationships from transcriptomic embeddings guide the adaptation of image features. In the Inference Stage, only the adapted histopathology model with the PEFT projector is used to process image patches, followed by a 256-dimensional PCA-Ridge regression model to predict gene expression profiles. This approach enables efficient transfer of transcriptomic domain knowledge while modifying around 5% of the original model parameters.

Knowledge Transfer with Parameter-Efficient Adaptation Given a pre-trained gene expression encoder as the teacher model $T(\cdot)$ and the adapted image encoder as the student model $S(\cdot)$, we formulate the knowledge transfer process as follows:

$$S(\mathbf{X}_{img}) = F(\mathbf{X}_{img}; \theta_F + \Delta W) \quad (2)$$

where θ_F represents the parameters of the Foundation Model and ΔW denotes the task-specific parameter updates. The teacher model $T(\cdot)$ provides domain-specific supervision signals to guide the adaptation of the student model. To achieve efficient adaptation while preserving the Foundation Model’s general capabilities, we explore various parameter-efficient fine-tuning (PEFT) strategies. The general form of parameter adaptation can be expressed as:

$$\theta_{adapted} = \theta_F + \Delta \mathbf{W} = \theta_F + (\mathbf{W}_g \mathbf{B}_g + \mathbf{B}_g) \quad (3)$$

where $\Delta \mathbf{W}$ represents the adaptive weight matrix. To achieve efficient adaptation while preserving structural information, we extend the original low rank adaptation framework by applying block-wise grouping, where $\mathbf{W}_g \in \mathbb{R}^{n \times m}$ is reshaped into blocks of size $b \times b$. These blocks are then strategically grouped to significantly reduce the number of trainable parameters without compromising performance. This grouped block structure yields a closed-form solution equivalent to the Bone formulation [10], which demonstrates superior performance in bridging the modality gap while maintaining parameter efficiency.

2.3 Optimization target

The training objective of PEKA comprises three components: knowledge distillation loss, structure alignment loss, and regularization term. The complete loss function can be formulated as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{KD} + \lambda_2 \mathcal{L}_{struct} \quad (4)$$

where λ_1 and λ_2 are weights balancing different losses. Each component serves a specific purpose:

- **Knowledge Distillation Loss** (\mathcal{L}_{KD}): Facilitates the transfer of transcriptomic domain knowledge from the teacher model to the student model:

$$\mathcal{L}_{KD} = D_{KL}(\text{softmax}(T(\mathbf{X}_{seq})/\tau) \parallel \text{softmax}(F(\mathbf{X}_{img}; \theta_F + \Delta W)/\tau)) \quad (5)$$

where τ is the temperature parameter controlling the softness of probability distributions.

- **Structure Alignment Loss** (\mathcal{L}_{struct}): Preserves the structural relationships in gene expression space:

$$\mathcal{L}_{struct} = \text{CrossEntropy}(S(\mathbf{X}_{img}), l) \quad (6)$$

where $l = \text{KNN}(\mathbf{X}_{seq})$ generates pseudo-labels by applying k-nearest neighbor clustering in the feature space from teacher model $T(\cdot)$. These structure-aware labels capture the intrinsic neighborhood relationships in the transcriptomic embedding space. By optimizing the cross-entropy loss between the student model’s predictions and these pseudo-labels, we encourage the image representations to preserve the structural topology of the teacher model $T(\cdot)$.

This training approach ensures effective knowledge transfer while maintaining structural consistency between modalities. In fact, the knowledge distillation process of PEKA results a maximization process of the mutual information between image and transcriptomic domain. Single cell large language models (scLLMs) serve as ideal teacher models for PEKA due to their comprehensive understanding of gene expression patterns across diverse cellular contexts. Among available scLLMs, we selected scFoundation [5] as our teacher model due to its extensive pretraining on over 50 million human single-cell transcriptomic profiles and shows good performance in multiple downstream tasks. The embeddings from scFoundation serve as supervision signals during the knowledge distillation process, guiding our image-based student model to learn meaningful representations that align with biologically relevant gene expression patterns. By distilling knowledge from scFoundation, our approach benefits from its extensive pre-training while adapting this knowledge to the spatial context provided by the imaging modality.

3 Experiments and Results

3.1 Dataset Descriptions

The HEST dataset contains 1,229 whole slide images (WSI) with paired spatial transcriptomic (ST) profiles [7]. To mitigate the effect of different ST technologies and types of diseases, and ensure a clean comparison, we selected subsets

from the HEST dataset to construct four different benchmarking datasets. The datasets contain Visium ST data of Homo Sapiens with breast cancer (n=30,414 pairs of image tiles and gene expression data), kidney cancer (n=73,813), liver cancer (n=37,168), and lung cancer (n=64,728), respectively. We applied 5-fold cross-validation to compare the performance of our model with other methods.

Training Settings We applied the same quality control steps of scFoundation to filter out low-quality and damaged cells [5]. Following the designs in HEST, we calculated the top 50 highly-variable-genes (HVG) in each dataset as targets for evaluating models, and employed a 256-dimensional PCA followed by a ridge regression classifier to predict gene expressions from foundation model (FM) embeddings. We compared our PEKA model against backbone FMs including Resnet-50 trained on ImageNet [21], CtrPath [20], UNI [1] and Hoptimus0 [14]. We also compared our method with different strategies for parameter-efficient finetuning, including Low-Rank Adaptation (LoRA) [6], AdaLoRA [23].

During the Knowledge Transfer Stage of PEKA, the adapter parameters and MLP weights are updated while keeping the backbone frozen. In the inference stage, we freeze the entire adapted model (backbone with PEFT) and only train a PCA-Ridge regression head for gene expression prediction.

Evaluation metric Pearson correlation coefficient (PCC) is used as the evaluation metric.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (7)$$

where, x and y are the predicted and ground truth gene expression values respectively, \bar{x} and \bar{y} are their respective means, and n is the total number of samples.

Hyper parameters We ran knowledge transfer experiments for 50 epochs, using an Adam optimizer with a learning rate of 0.0001. We used $r = 256$, $\alpha = 32$, and a dropout rate of 0.1 for the low-rank adaptation of the models. We set $\lambda_1 = \lambda_2 = 0.5$ for the alignment of models. The backbone we selected: H-optimus-0 (H0) is a large foundation model, which includes 1.1B parameters. Using PEKA, we achieve parameter-efficient and data-efficient knowledge transfer by only tuning around 5% of the parameters in the imaging backbone. The alignment takes 12 GPU hours on a single V100 GPU.

Knowledge transfer results We trained our models by transferring knowledge from a scFoundation model (teacher) to a H0 image foundation model (student) using our parameter efficient knowledge transfer (PEKA) approach.

We compared our model with other foundation models in the four datasets derived from HEST. Our method significantly outperforms baseline models, with

Table 1. PEKA significantly outperforms Resnet, CtransPath, UNI and H0 in Pearson correlation (PCC) for gene expression prediction between the imaging and expression embedding spaces.

Model	HEST-Breast	HEST-Kidney	HEST-Liver	HEST-Lung
ResNet-50	0.554	0.587	0.554	0.696
CTransPath	0.424	0.410	0.489	0.592
UNI	0.617	0.673	0.576	0.708
Hoptimous0	0.624	0.698	0.594	0.720
PEKA (ours)	0.698	0.755	0.654	0.774

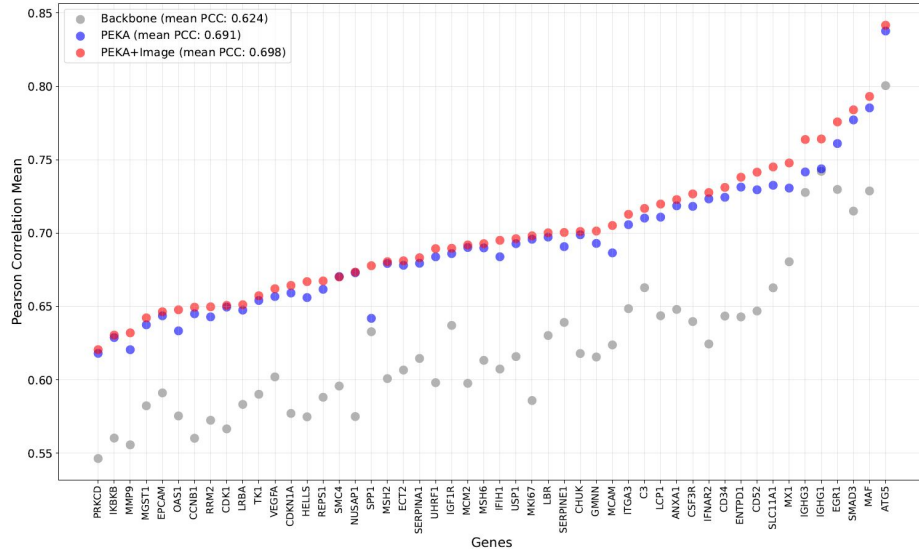


Fig. 2. PCC of top 50 highly-variable-genes in the breast cancer dataset.

a 5-7% absolute increase in PCC compared to the second-best performing model (**Table. 1**). Surprisingly, CTransPath, which is pretrained on a large cohort of pathology images, has much lower PCC compared to Resnet-50 pretrained on ImageNet. This observation shows the potential risk of foundation models overfitting to one modality, and highlights the importance bridging different modalities with knowledge transfer algorithms. Using the PCC of top-50 highly variable genes in the HEST-Breast dataset as an example, we show that PEKA improves gene expression prediction for all of the genes by a large margin (**Fig. 2**). Concatenating the embeddings of the aligned model with the embeddings of the pathology image foundation model also consistently improved prediction accuracy, suggesting that the aligned model captured extra information complementary to imaging through the knowledge transfer process.

Table 2. PEKA consistently outperforms other parameter efficient fine-tuning methods correlation coefficient across backbones and datasets. Values in the table are Pearson correlation coefficients between predicted gene expressions and ground truth.

Backbone	UNI				Hoptimus0			
Method	None	LoRA	AdaLoRA	PEKA	None	LoRA	AdaLoRA	PEKA
Breast	0.617	0.668	0.643	0.688	0.624	0.687	0.658	0.698
Kidney	0.673	0.711	0.686	0.729	0.698	0.734	0.715	0.755
Liver	0.576	0.626	0.608	0.654	0.594	0.601	0.586	0.640
Lung	0.708	0.752	0.734	0.774	0.720	0.738	0.716	0.774

Comparison with other PEFT methods We compared our method with three other settings of parameter-efficient fine-tuning (PEFT) including **None** (No fine-tuning), **LoRA** (Low Rank Adaptation) [6] and **AdaLoRA** [23]. In all four datasets, we found that all PEFT strategies significantly improves PCC of gene expressions compared to **None**. Our approach always provides the largest performance gain, with **LoRA** being the second.

4 Conclusions

In conclusion, we propose PEKA, a parameter-efficient knowledge transfer workflow for aligning foundation models trained on different modalities. Foundation models fine-tuned using PEKA achieved state-of-the-art performance in predicting gene expressions from pathology image tiles across four different tissue types. From our results, PEKA can effectively bridge the modality gap between histopathology images and gene expression data, enabling more accurate molecular predictions from morphological features. We believe PEKA and the aligned model will serve as powerful tools for researchers and clinicians to extract valuable transcriptomic insights from routinely available histopathology

images. Furthermore, PEKA provides a strong baseline for future advancements in cross-modal knowledge transfer, particularly in the medical domain, where paired multi-modal data are scarce. In the future, we plan to leverage the data-efficiency and parameter-efficiency of PEKA to align other modalities such as proteomics [12], metabolomics [9], or epigenomics [19] across cancer types and evaluate its utility in precision oncology applications, for example cancer subtyping and survival analysis.

Acknowledgments. Shi Pan and Maria Secrier were supported by a UKRI Future Leaders Fellowship (MR/T042184/1, MR/Y034031/1). Work in MS’s lab was supported by a BBSRC equipment grant (BB/R01356X/1) and a Wellcome Institutional Strategic Support Fund (204841/Z/16/Z). Jianan Chen was supported by the Tom Prince Cancer Trust, and the Marie Curie Postdoctoral Fellowship (101210483).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. *Nature Medicine* **30**(3), 850–862 (2024)
2. Cui, H., Wang, C., Maan, H., Pang, K., Luo, F., Duan, N., Wang, B.: scgpt: toward building a foundation model for single-cell multi-omics using generative ai. *Nature Methods* **21**(8), 1470–1480 (2024)
3. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering* **69**(3), 1173–1185 (2021)
4. Gulati, G.S., D’Silva, J.P., Liu, Y., Wang, L., Newman, A.M.: Profiling cell identity and tissue architecture with single-cell and spatial transcriptomics. *Nature Reviews Molecular Cell Biology* pp. 1–21 (2024)
5. Hao, M., Gong, J., Zeng, X., Liu, C., Guo, Y., Cheng, X., Wang, T., Ma, J., Zhang, X., Song, L.: Large-scale foundation model on single-cell transcriptomics. *Nature methods* **21**(8), 1481–1491 (2024)
6. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: Lora: Low-rank adaptation of large language models. *ICLR* **1**(2), 3 (2022)
7. Jaume, G., Doucet, P., Song, A., Lu, M.Y., Almagro Pérez, C., Wagner, S., Vaidya, A., Chen, R., Williamson, D., Kim, A., et al.: Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems* **37**, 53798–53833 (2025)
8. Jaume, G., Oldenburg, L., Vaidya, A., Chen, R.J., Williamson, D.F., Peeters, T., Song, A.H., Mahmood, F.: Transcriptomics-guided slide representation learning in computational pathology. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9632–9644 (2024)
9. Johnson, C.H., Ivanisevic, J., Siuzdak, G.: Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews Molecular cell biology* **17**(7), 451–459 (2016)
10. Kang, J.: Bone: Block affine transformation as parameter efficient fine-tuning methods for large language models. *arXiv preprint arXiv:2409.15371* (2024)
11. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., Teichmann, S.A.: The technology and biology of single-cell rna sequencing. *Molecular cell* **58**(4), 610–620 (2015)
12. Miao, Z., Humphreys, B.D., McMahon, A.P., Kim, J.: Multi-omics integration in the age of million single-cell data. *Nature Reviews Nephrology* **17**(11), 710–724 (2021)
13. Rao, A., Barkley, D., França, G.S., Yanai, I.: Exploring tissue architecture using spatial transcriptomics. *Nature* **596**(7871), 211–220 (2021)
14. Saillard, C., Jenatton, R., Llinares-López, F., Mariet, Z., Cahané, D., Durand, E., Vert, J.P.: H-optimus-0 (2024), <https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0>
15. Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., Kamoun, A., Sefta, M., Toldo, S., Zaslavskiy, M., et al.: A deep learning model to predict rna-seq expression of tumours from whole slide images. *Nature communications* **11**(1), 3877 (2020)
16. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Liu, S., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., et al.: Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778* (2023)
17. Vorontsov, E., Bozkurt, A., Casson, A., Shaikovski, G., Zelechowski, M., Severson, K., Zimmermann, E., Hall, J., Tenenholtz, N., Fusi, N., et al.: A foundation

- model for clinical-grade computational pathology and rare cancers detection. *Nature medicine* **30**(10), 2924–2935 (2024)
18. Wang, C., Chan, A.S., Fu, X., Ghazanfar, S., Kim, J., Patrick, E., Yang, J.Y.: Benchmarking the translational potential of spatial gene expression prediction from histology. *Nature Communications* **16**(1), 1544 (2025)
 19. Wang, K.C., Chang, H.Y.: Epigenomics: technologies and applications. *Circulation research* **122**(9), 1191–1199 (2018)
 20. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis* **81**, 102559 (2022)
 21. Wightman, R., Touvron, H., Jégou, H.: Resnet strikes back: An improved training procedure in timm. *arXiv preprint arXiv:2110.00476* (2021)
 22. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al.: A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**(8015), 181–188 (2024)
 23. Zhang, Q., Chen, M., Bukharin, A., Karampatziakis, N., He, P., Cheng, Y., Chen, W., Zhao, T.: Adalora: Adaptive budget allocation for parameter-efficient fine-tuning. *arXiv preprint arXiv:2303.10512* (2023)